

РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА
И ГОСУДАРСТВЕННОЙ СЛУЖБЫ
ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ

АКАДЕМИЧЕСКИЙ УЧЕБНИК

Уильям Г. Грин

ЭКОНОМЕТРИЧЕСКИЙ АНАЛИЗ

Книга 1





РАНХиГС

**РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА
И ГОСУДАРСТВЕННОЙ СЛУЖБЫ
ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ**

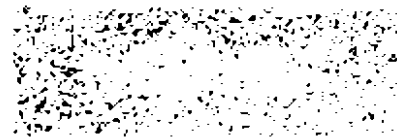
William H. Greene

**ECONOMETRIC
ANALYSIS**

Seventh edition

Prentice Hall

2012



© 2016. Все права защищены.

Уильям Г. Грин

ЭКОНОМЕТРИЧЕСКИЙ АНАЛИЗ

Книга 1

*Перевод с английского под научной редакцией
С.С. Синельникова и М.Ю. Турунцевой*

Рекомендуется Российской академией народного хозяйства и государственной службы при Президенте Российской Федерации в качестве учебника для студентов, обучающихся по экономическим направлениям и специальностям, а также для студентов бакалавриата, магистратуры, аспирантов, преподавателей экономических факультетов вузов. (Основание – приказ Министерства образования и науки №130 от 22 февраля 2012 г.)



Москва • 2016

УДК 330.4
ББК 65.05
Г85

Перевод с английского:

А. В. Ходырев (главы 1–4), А. С. Степанов (главы 5–6, 12–13),
С. С. Синельников (главы 7–8), Б. Н. Гафаров (главы 9, 17),
Ю. В. Набатова (главы 10–11), В. В. Громов (главы 14–16),
Ю. Ю. Пономарёв (главы 18–19), Е. В. Синельникова (главы 20–21),
А. В. Божечкова (приложения)

Грин, Уильям Г.

Г85 Эконометрический анализ. Книга 1 / Уильям Грин; пер. с англ.; под науч. ред. С. С. Синельникова и М. Ю. Турунцевой. — М. : Издательский дом «Дело» РАНХиГС, 2016. — 760 с. (Академический учебник).

ISBN 978-5-7749-0959-9 (общ.)

ISBN 978-5-7749-1157-8 (кн. 1)

«Эконометрический анализ» является учебником по эконометрике магистерского уровня, в нем рассмотрены все основные разделы, включаемые в современные курсы эконометрики. Он представляет собой попытку соединить введение в область эконометрики и профессиональную литературу для студентов и аналитиков, изучающих социальные науки и специализирующихся на прикладной эконометрике и ее теоретических основах. Книга дает широкий обзор различных областей эконометрики, что позволяет читателю перейти от их изучения к решению практических задач в одной или нескольких областях. Читатель имеет возможность ознакомиться и с базовыми понятиями различных областей эконометрики, представленных в книге, и с теми, которые ему необходимы на практике.

УДК 330.43
ББК 65.05

ISBN 978-5-7749-0959-9 (общ.)

ISBN 978-5-7749-1157-8 (кн. 1)

Authorized translation from the English language edition, entitled ECONOMETRIC ANALYSIS, 7th Edition; ISBN 0131395386; by GREENE, WILLIAM H.; published by Pearson Education, Inc.; publishing as Prentice Hall; Copyright © 2012 Pearson Education Limited

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from Pearson Education, Inc. RUSSIAN language edition published by DELO PUBLISHERS. Copyright © 2013

Лицензированный перевод английского издания под названием ECONOMETRIC ANALYSIS, 7th Edition; ISBN 0131395386; под авторством GREENE, WILLIAM H., опубликованного Pearson Education, Inc. под маркой Prentice Hall; Copyright © 2012 Pearson Education Limited

Все права защищены. Ни одна часть настоящей книги не может быть распространена или передана ни в каком виде и никакими средствами, электронными или механическими, включая фотокопирование, запись или любые информационно-поисковые системы, без разрешения от Pearson Education, Inc. Издание на русском языке выпущено Издательским домом «Дело». Copyright © 2013

© ФГБОУ ВО «Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации», 2016

Оглавление

| | |
|--|------|
| Предисловие к русскому изданию | xiii |
| Предисловие | xiv |
| Глава 1. Эконометрика | 1 |
| 1.1. Введение | 1 |
| 1.2. Парадигма эконометрики | 1 |
| 1.3. Практическая эконометрика | 4 |
| 1.4. Эконометрическое моделирование | 5 |
| 1.5. План книги | 9 |
| 1.6. Предварительные замечания | 11 |
| 1.6.1. Численные примеры | 11 |
| 1.6.2. Компьютерное обеспечение и воспроизводимость результатов .. | 11 |
| 1.6.3. Соглашения об обозначениях | 12 |
| Глава 2. Модель линейной регрессии | 13 |
| 2.1. Введение | 13 |
| 2.2. Модель линейной регрессии | 15 |
| 2.3. Предположения модели линейной регрессии | 19 |
| 2.3.1. Линейность модели регрессии | 19 |
| 2.3.2. Полный ранг | 23 |
| 2.3.3. Регрессия | 25 |
| 2.3.4. Сферические шоки | 27 |
| 2.3.5. Процесс порождения данных для регрессоров | 28 |
| 2.3.6. Нормальность | 29 |
| 2.3.7. Независимость | 29 |
| 2.4. Заключение | 30 |
| Глава 3. Метод наименьших квадратов | 32 |
| 3.1. Введение | 32 |
| 3.2. Регрессия наименьших квадратов | 32 |
| 3.2.1. Вектор коэффициентов метода наименьших квадратов | 33 |
| 3.2.2. Приложение: уравнение инвестиций | 34 |
| 3.2.3. Алгебраические свойства оценки наименьших квадратов | 37 |
| 3.2.4. Проекция | 37 |
| 3.3. Раздельная и частичная регрессии | 38 |
| 3.4. Частичная регрессия и частные коэффициенты корреляции | 42 |
| 3.5. Качество приближения и анализ разброса | 46 |
| 3.5.1. Скорректированный R-квадрат и меры качества подгонки | 50 |
| 3.5.2. R-квадрат и константа в модели | 52 |
| 3.5.3. Сравнение моделей | 52 |
| 3.6. Линейные преобразования регрессий | 54 |
| 3.7. Заключение | 55 |
| Глава 4. Оценки методом наименьших квадратов | 60 |
| 4.1. Введение | 60 |
| 4.2. Почему именно метод наименьших квадратов? | 61 |
| 4.2.1. Условия ортогональности генеральной совокупности | 61 |
| 4.2.2. Предиктор, минимизирующий среднеквадратичную ошибку ... | 62 |
| 4.2.3. Линейная несмещенная оценка с минимальной дисперсией ... | 63 |
| 4.3. Свойства метода наименьших квадратов на конечных выборках | 65 |

| | |
|---|-----|
| 4.3.1. Несмещенность оценки | 65 |
| 4.3.2. Смещение, вызванное пропущенными значимыми переменными. | 66 |
| 4.3.3. Включение лишних переменных | 68 |
| 4.3.4. Дисперсия оценки наименьших квадратов | 69 |
| 4.3.5. Теорема Гаусса–Маркова | 70 |
| 4.3.6. Особенности стохастических регрессоров | 71 |
| 4.3.7. Оценки дисперсии методом наименьших квадратов | 72 |
| 4.3.8. Предположение о нормальности | 73 |
| 4.4. Свойства оценок наименьших квадратов на больших выборках | 74 |
| 4.4.1. Состоятельность оценки наименьших квадратов коэффициента β | 74 |
| 4.4.2. Асимптотическая нормальность оценок наименьших квадратов | 77 |
| 4.4.3. Состоятельность s^2 и построение оценки для $Asy. Var[\mathbf{b}]$ | 78 |
| 4.4.4. Асимптотическое распределение функций от \mathbf{b} : дельта-метод | 79 |
| 4.4.5. Асимптотическая эффективность | 82 |
| 4.4.6. Оценка максимального правдоподобия | 86 |
| 4.5. Интервальные оценки | 88 |
| 4.5.1. Построение доверительного интервала для коэффициента линейной регрессии | 89 |
| 4.5.2. Построение доверительных интервалов для больших выборок | 91 |
| 4.5.3. Доверительные интервалы для линейных комбинаций коэффициентов: разложение Охака | 93 |
| 4.6. Предсказание и прогнозирование | 94 |
| 4.6.1. Доверительные интервалы для предсказаний. | 95 |
| 4.6.2. Предсказание y , если уравнение регрессии описывает логарифм y | 96 |
| 4.6.3. Доверительный интервал для предсказания y в случаях, когда уравнение регрессии описывает логарифм y | 97 |
| 4.6.4. Прогнозирование | 102 |
| 4.7. Проблемы в данных | 103 |
| 4.7.1. Мультиколлинеарность | 104 |
| 4.7.2. Предварительное оценивание. | 107 |
| 4.7.3. Метод главных компонент. | 108 |
| 4.7.4. Пропущенные значения и пополнение данных | 110 |
| 4.7.5. Ошибки измерения | 114 |
| 4.7.6. Влиятельные наблюдения и выбросы | 116 |
| 4.8. Заключение и выводы | 120 |
| Глава 5. Тестирование гипотез и выбор спецификации. | 127 |
| 5.1. Введение | 127 |
| 5.2. Методология тестирования гипотез | 127 |
| 5.2.1. Ограничения и гипотезы | 128 |
| 5.2.2. Вложенные модели | 129 |
| 5.2.3. Процедуры тестирования — методология Неймана–Пирсона | 130 |
| 5.2.4. Размер, мощность и состоятельность теста | 131 |
| 5.2.5. Методологическая дилемма: байесовское тестирование против классического | 132 |
| 5.3. Два подхода к тестированию гипотез. | 133 |
| 5.4. Тест Вальда. | 135 |
| 5.4.1. Тестирование гипотез о коэффициенте | 135 |
| 5.4.2. F-статистика и отклонение метода наименьших квадратов | 138 |
| 5.5. Тестирование ограничений с использованием показателей качества регрессии | 142 |
| 5.5.1. Оценка наименьших квадратов с ограничениями | 143 |
| 5.5.2. Потеря в качестве подгонки оценки наименьших квадратов с ограничениями | 144 |

| | |
|--|------------|
| 5.5.3. Тестирование значимости регрессии | 148 |
| 5.5.4. Вывод ограничений и замечание об использовании R^2 | 149 |
| 5.6. Ошибки, не являющиеся нормально распределенными, и асимптотические тесты | 150 |
| 5.7. Тестирование нелинейных ограничений | 155 |
| 5.8. Выбор между невложенными моделями | 158 |
| 5.8.1. Тестирование невложенных гипотез | 159 |
| 5.8.2. Принцип охвата | 160 |
| 5.8.3. Полная модель — J-тест | 161 |
| 5.9. Тестирование спецификации модели | 162 |
| 5.10. Построение модели — подход от общего к частному | 164 |
| 5.10.1. Критерии выбора модели | 165 |
| 5.10.2. Выбор модели | 166 |
| 5.10.3. Классический подход к выбору модели | 166 |
| 5.10.4. Байесовское усреднение моделей | 167 |
| 5.11. Заключение и выводы | 169 |
| Глава 6. Функциональная форма и структурный сдвиг | 177 |
| 6.1. Введение | 177 |
| 6.2. Использование бинарных переменных | 177 |
| 6.2.1. Бинарные переменные в регрессии | 178 |
| 6.2.2. Случай нескольких фиктивных переменных | 180 |
| 6.2.3. Случай нескольких групп | 182 |
| 6.2.4. Пороговые эффекты и индикаторные переменные | 184 |
| 6.2.5. Эффекты воздействия и регрессия «разности разностей» | 185 |
| 6.3. Нелинейность в переменных | 188 |
| 6.3.1. Кусочно-линейная регрессия | 189 |
| 6.3.2. Функциональные формы | 190 |
| 6.3.3. Эффект взаимодействия | 192 |
| 6.3.4. Выявление нелинейности | 193 |
| 6.3.5. Внутренне линейные модели | 197 |
| 6.4. Моделирование и тестирование структурного сдвига | 200 |
| 6.4.1. Различные векторы параметров | 201 |
| 6.4.2. Недостаточное число наблюдений | 202 |
| 6.4.3. Изменение части коэффициентов | 203 |
| 6.4.4. Тесты на структурное изменение при различных дисперсиях | 204 |
| 6.4.5. Тестирование стабильности модели при помощи теста на предсказательную силу | 209 |
| 6.5. Заключение и выводы | 210 |
| Глава 7. Нелинейная, полупараметрическая и непараметрическая модели регрессии | 217 |
| 7.1. Введение | 217 |
| 7.2. Нелинейные регрессионные модели | 218 |
| 7.2.1. Предположения нелинейной регрессионной модели | 219 |
| 7.2.2. Нелинейная оценка наименьших квадратов | 221 |
| 7.2.3. Асимптотические свойства оценки нелинейным методом наименьших квадратов | 223 |
| 7.2.4. Проверка гипотез и ограничения на параметры | 226 |
| 7.2.5. Примеры | 229 |
| 7.2.6. Вычисление оценки нелинейным методом наименьших квадратов | 241 |
| 7.3. Медианная и квантильная регрессии | 244 |
| 7.3.1. Оценка методом наименьших абсолютных отклонений | 246 |
| 7.3.2. Модели квантильных регрессий | 249 |
| 7.4. Частично линейная регрессия | 253 |
| 7.5. Непараметрическая регрессия | 255 |
| 7.6. Заключение и выводы | 259 |

| | |
|--|-----|
| Глава 8. Эндогенность и метод инструментальных переменных..... | 264 |
| 8.1. Введение | 264 |
| 8.2. Предположения расширенной модели..... | 269 |
| 8.3. Оценка | 272 |
| 8.3.1. Метод наименьших квадратов | 272 |
| 8.3.2. Оценка методом инструментальных переменных..... | 272 |
| 8.3.3. Причины, приводящие к методу инструментальных переменных | 274 |
| 8.3.4. Двухшаговый метод наименьших квадратов | 278 |
| 8.4. Два теста на спецификацию модели..... | 283 |
| 8.4.1. Тесты Хаусмана и Ву на спецификацию модели..... | 284 |
| 8.4.2. Тест на сверхидентифицирующие ограничения..... | 289 |
| 8.5. Ошибка измерения | 291 |
| 8.5.1. Недооценка метода наименьших квадратов | 291 |
| 8.5.2. Оценка методом инструментальных переменных..... | 294 |
| 8.5.3. Прокси-переменные..... | 294 |
| 8.6. Оценка нелинейным методом инструментальных переменных | 299 |
| 8.7. Слабые инструменты | 303 |
| 8.8. Естественные эксперименты и поиск причинно-следственных связей..... | 306 |
| 8.9. Заключение и выводы | 310 |
| Глава 9. Обобщенная регрессионная модель и гетероскедастичность | 313 |
| 9.1. Введение | 313 |
| 9.2. Неэффективность оценок методами наименьших квадратов и инструментальных переменных | 315 |
| 9.2.1. Свойства МНК в конечных выборках..... | 315 |
| 9.2.2. Асимптотические свойства МНК | 316 |
| 9.2.3. Устойчивая оценка асимптотической ковариационной матрицы | 318 |
| 9.2.4. Метод инструментальных переменных | 320 |
| 9.3. Эффективное оценивание обобщенным методом наименьших квадратов | 321 |
| 9.3.1. Обобщенный метод наименьших квадратов (ОМНК, GLS)..... | 322 |
| 9.3.2. Доступный обобщенный метод наименьших квадратов (ДОМНК, FGLS)..... | 324 |
| 9.4. Гетероскедастичность и взвешенный метод наименьших квадратов | 326 |
| 9.4.1. Обычный метод наименьших квадратов | 328 |
| 9.4.2. Неэффективность обычного метода наименьших квадратов | 328 |
| 9.4.3. Оцененная ковариационная матрица \mathbf{b} | 329 |
| 9.4.4. Оценка подходящей ковариационной матрицы для метода наименьших квадратов | 331 |
| 9.5. Тесты на гетероскедастичность | 334 |
| 9.5.1. Общий тест Уайта | 335 |
| 9.5.2. LM-тест Бройша–Пэгана/Годфри | 336 |
| 9.6. Взвешенный метод наименьших квадратов..... | 338 |
| 9.6.1. Взвешенный метод наименьших квадратов с известной матрицей Ω | 338 |
| 9.6.2. Оценка при условии, что матрица Ω содержит неизвестные параметры | 340 |
| 9.7. Приложения | 341 |
| 9.7.1. Мультипликативная гетероскедастичность | 341 |
| 9.7.2. Групповая гетероскедастичность | 344 |
| 9.8. Заключение и выводы | 347 |
| Глава 10. Системы уравнений | 354 |
| 10.1. Введение | 354 |
| 10.2. Системы внешне не связанных уравнений..... | 356 |

| | |
|---|------------|
| 10.2.1. Обобщенный метод наименьших квадратов | 357 |
| 10.2.2. Системы внешне не связанных уравнений с одинаковыми регрессорами | 359 |
| 10.2.3. Допустимый обобщенный метод наименьших квадратов (ДОМНК, FGLS) | 360 |
| 10.2.4. Проверка гипотез | 361 |
| 10.2.5. Тест на спецификацию для систем внешне не связанных уравнений | 362 |
| 10.2.6. Объединенная модель | 364 |
| 10.3. Обобщенные модели внешне не связанных регрессий. | 369 |
| 10.4. Нелинейные системы уравнений | 370 |
| 10.5. Системы уравнений спроса: вырожденные системы | 374 |
| 10.5.1. Функция издержек Кобба-Дугласа | 375 |
| 10.5.2. Гибкие функциональные формы: транслогарифмическая функция издержек. | 379 |
| 10.6. Модели одновременных уравнений | 383 |
| 10.6.1. Системы уравнений | 384 |
| 10.6.2. Общие замечания для моделей линейных одновременных уравнений | 387 |
| 10.6.3. Проблема идентификации | 391 |
| 10.6.4. Оценка одного уравнения и проверка гипотез | 397 |
| 10.6.5. Методы оценки систем уравнений | 400 |
| 10.6.6. Тестирование при наличии слабых инструментальных переменных | 405 |
| 10.7. Заключение и выводы | 408 |
| Глава 11. Модели панельных данных. | 417 |
| 11.1. Введение | 417 |
| 11.2. Модели панельных данных | 418 |
| 11.2.1. Общие подходы к моделированию панельных данных | 420 |
| 11.2.2. Структуры модели. | 421 |
| 11.2.3. Расширения | 422 |
| 11.2.4. Сбалансированные и несбалансированные панели | 423 |
| 11.2.5. Доброкачественные панельные данные | 423 |
| 11.3. Объединенная регрессионная модель | 424 |
| 11.3.1. Оценка объединенной модели методом наименьших квадратов | 425 |
| 11.3.2. Устойчивая оценка ковариационной матрицы | 425 |
| 11.3.3. Кластеризация и стратификация | 428 |
| 11.3.4. Устойчивое оценивание с использованием группового среднего значения | 430 |
| 11.3.5. Оценка в первых разностях. | 432 |
| 11.3.6. Внутри- и межгрупповые оценки. | 433 |
| 11.4. Модель с фиксированными эффектами. | 437 |
| 11.4.1. Оценки методом наименьших квадратов | 438 |
| 11.4.2. Асимптотики при малых T | 440 |
| 11.4.3. Проверка значимости групповых эффектов | 441 |
| 11.4.4. Фиксированные временные и групповые эффекты | 441 |
| 11.4.5. Постоянные во времени переменные и декомпозиция вектора фиксированных эффектов | 443 |
| 11.5. Случайные эффекты | 449 |
| 11.5.1. Оценка методом наименьших квадратов. | 453 |
| 11.5.2. Обобщенный метод наименьших квадратов. | 454 |
| 11.5.3. Допустимый обобщенный метод наименьших квадратов при неизвестной матрице Σ | 455 |
| 11.5.4. Проверка гипотез для случайных эффектов | 457 |
| 11.5.5. Тест на спецификацию Хаусмана для модели со случайными эффектами | 461 |

| | |
|--|------------|
| 11.5.6. Расширение модели с ненаблюдаемыми эффектами: подход Мундлака | 463 |
| 11.5.7. Расширение моделей с фиксированными и случайными эффектами: подход Чемберлена | 465 |
| 11.6. Несферические шоки и устойчивые оценки ковариационной матрицы | 470 |
| 11.6.1. Устойчивые оценки модели с фиксированными эффектами | 470 |
| 11.6.2. Гетероскедастичность в модели со случайными эффектами | 471 |
| 11.6.3. Автокорреляция в моделях панельных данных | 472 |
| 11.6.4. Кластерные (и панельные) устойчивые ковариационные матрицы для оценок с фиксированными и случайными эффектами | 473 |
| 11.7. Пространственная автокорреляция | 475 |
| 11.8. Эндогенность | 481 |
| 11.8.1. Оценка инструментальных переменных Хаусмана–Тейлора | 481 |
| 11.8.2. Состоятельная оценка динамических моделей панельных данных: ГУ-оценка Андерсона–Хсяо | 486 |
| 11.8.3. Эффективная оценка динамических моделей панельных данных: метод Ареллано–Бонда | 489 |
| 11.8.4. Нестационарные временные ряды и модели панельных данных | 498 |
| 11.9. Нелинейная регрессия с панельными данными | 501 |
| 11.9.1. Устойчивая ковариационная матрица для нелинейного метода наименьших квадратов | 501 |
| 11.9.2. Фиксированные эффекты | 503 |
| 11.9.3. Случайные эффекты | 505 |
| 11.10. Системы уравнений | 506 |
| 11.11. Неоднородность параметров | 508 |
| 11.11.1. Модель со случайными коэффициентами | 508 |
| 11.11.2. Иерархическая линейная модель | 511 |
| 11.11.3. Неоднородность параметров и динамические модели панельных данных | 514 |
| 11.12. Заключение и выводы | 519 |
| Глава 12. Подходы к оцениванию в эконометрике | 527 |
| 12.1. Введение | 527 |
| 12.2. Параметрические методы оценки и тестирование статистических гипотез | 529 |
| 12.2.1. Классический подход к оценке, основанный на функции правдоподобия | 530 |
| 12.2.2. Моделирование совместных распределений копула-функциями | 532 |
| 12.3. Полупараметрические методы оценки | 536 |
| 12.3.1. ОММ-оценка в эконометрике | 537 |
| 12.3.2. Оценка методом максимального эмпирического правдоподобия | 537 |
| 12.3.3. Оценка методом наименьших абсолютных отклонений и квантильная регрессия | 539 |
| 12.3.4. Ядерные методы оценки | 540 |
| 12.3.5. Сравнение параметрического и полупараметрического анализа | 541 |
| 12.4. Непараметрические методы оценки | 543 |
| 12.4.1. Ядерная оценка плотности | 543 |
| 12.5. Свойства оценок | 547 |
| 12.5.1. Статистические свойства оценок | 547 |
| 12.5.2. Экстремальные оценки | 548 |
| 12.5.3. Предположения для асимптотических свойств экстремальных оценок | 549 |
| 12.5.4. Асимптотические свойства оценок | 552 |

| | |
|---|-----|
| 12.5.5. Тестирование гипотез | 553 |
| 12.6. Заключение и выводы | 554 |
| Глава 13. Оценка методом наименьшего расстояния и обобщенным методом моментов | 556 |
| 13.1. Введение | 556 |
| 13.2. Состоятельная оценка: метод моментов | 558 |
| 13.2.1. Случайные выборки и оценка параметров распределений | 559 |
| 13.2.2. Асимптотические свойства оценки метода моментов | 563 |
| 13.2.3. Заключение — метод моментов | 566 |
| 13.3. Оценка методом наименьшего расстояния | 566 |
| 13.4. Оценка обобщенным методом моментов (GMM) | 572 |
| 13.4.1. Оценивание на основе условий ортогональности | 572 |
| 13.4.2. Обобщение метода моментов | 574 |
| 13.4.3. Свойства GMM-оценки | 579 |
| 13.5. Тестирование гипотез в контексте GMM | 584 |
| 13.5.1. Тестирование обоснованности моментных ограничений | 585 |
| 13.5.2. GMM-аналоги для тестов Вальда, множителей Лагранжа и отношения правдоподобия | 586 |
| 13.6. GMM-оценка эконометрических моделей | 588 |
| 13.6.1. Линейные модели для одного уравнения | 589 |
| 13.6.2. Нелинейные модели регрессии | 595 |
| 13.6.3. Системы внешне не связанных уравнений | 599 |
| 13.6.4. Модели одновременных уравнений с гетероскедастичностью | 600 |
| 13.6.5. GMM-оценка динамических моделей панельных данных | 604 |
| 13.7. Заключение и выводы | 616 |
| Глава 14. Метод максимального правдоподобия | 620 |
| 14.1. Введение | 620 |
| 14.2. Функция правдоподобия и идентификация параметров | 620 |
| 14.3. Эффективное оценивание: принцип максимального правдоподобия | 623 |
| 14.4. Свойства оценок максимального правдоподобия | 625 |
| 14.4.1. Условия регулярности | 627 |
| 14.4.2. Свойства регулярных распределений | 627 |
| 14.4.3. Уравнение правдоподобия | 629 |
| 14.4.4. Информационное равенство | 630 |
| 14.4.5. Асимптотические свойства оценки методом максимального правдоподобия | 630 |
| 14.4.6. Оценка асимптотической дисперсии оценки методом максимального правдоподобия | 635 |
| 14.5. Условные функции правдоподобия, эконометрические модели и GMM-оценка | 637 |
| 14.6. Тестирование гипотез и спецификации модели и показатели качества подгонки модели | 639 |
| 14.6.1. Критерий отношения правдоподобия | 641 |
| 14.6.2. Критерий Вальда | 643 |
| 14.6.3. Критерий множителей Лагранжа | 645 |
| 14.6.4. Применение критериев, основанных на правдоподобии | 647 |
| 14.6.5. Сравнение моделей и вычисление качества подгонки модели | 649 |
| 14.6.6. Тест Вуонга и информационный критерий Кульбака–Лейблера | 651 |
| 14.7. Двухшаговая оценка методом максимального правдоподобия | 653 |
| 14.8. Оценка методом квазimaxимального правдоподобия и устойчивые асимптотические ковариационные матрицы | 661 |
| 14.8.1. Максимальное правдоподобие и GMM-оценки | 661 |
| 14.8.2. Максимальное правдоподобие и M-оценки | 662 |

| | |
|--|-----|
| 14.8.3. «Сэндвич»-оценки | 664 |
| 14.8.4. Кластерные оценки | 665 |
| 14.9. Модели, оцениваемые методом максимального правдоподобия | 668 |
| 14.9.1. Нормальная модель линейной регрессии | 669 |
| 14.9.2. Обобщенная регрессионная модель | 673 |
| 14.9.3. Модель внешне не связанных регрессий | 683 |
| 14.9.4. Модели одновременных уравнений | 691 |
| 14.9.5. Оценка моделей нелинейной регрессии методом максимального правдоподобия | 692 |
| 14.9.6. Модели панельных данных | 700 |
| 14.10. Модели скрытых классов и модели смеси распределений с конечным числом значений параметра | 719 |
| 14.10.1. Модель смеси распределений | 720 |
| 14.10.2. Измеренная и неизмеренная гетерогенность | 722 |
| 14.10.3. Предсказание принадлежности к классу | 723 |
| 14.10.4. Условная модель скрытого класса | 724 |
| 14.10.5. Определение числа классов | 727 |
| 14.10.6. Модель панельных данных | 727 |
| 14.11. Заключение и выводы | 732 |

Предисловие к русскому изданию

Я рад представить вашему вниманию русский перевод 7-го издания учебника «Эконометрический анализ». В октябре 2015 г. ресурс Google Scholar опубликовал список 100 работ из всех областей знания, которые наиболее широко цитируются. «Эконометрический анализ» с почти 50 000 цитирований занял в нем 34-е место. Специалисты различных научных направлений (экономисты, социологи, политологи, инженеры, биостатистики и т.д.) используют эту книгу в своих исследованиях. Такое широкое признание и популярность во многих странах очень приятны. Учебник уже переведен на французский, испанский, японский, китайский и турецкий языки. Я надеюсь, что российские читатели найдут его полезным для своей работы, что он поможет студентам узнать больше об эконометрических методах и научиться лучше применять их к моделированию различных экономических ситуаций. Как учебник, «Эконометрический анализ» включает все основные разделы эконометрического анализа. Я уверен, что студенты смогут найти те разделы и подходы, которые наиболее полезны и интересны им. Многие исследовательские центры в России используют эконометрику в своей работе. Они успешно применяют эконометрические методы в экономических исследованиях, а также в экономико-политическом консалтинге. Среди этих центров можно отметить Институт прикладных экономических исследований РАНХиГС, Институт экономической политики им. Е.Т. Гайдара, Центр экономических и финансовых исследований и разработок, НИУ «Высшая школа экономики» и многие другие. Я надеюсь, что «Эконометрический анализ» окажется полезным для всех исследователей в их будущей научной работе и внесет вклад в формирование новых идей и научных разработок. Я хотел бы выразить признательность Издательскому дому «Дело» и всем тем, кто готовил книгу к публикации: всем переводчикам (Сергею Синельникову, Юлии Набатовой, Владимиру Громову, Александру Ходыреву, Булату Гафарову, Елене Синельниковой, Александру Степанову, Юрию Пономарёву, Александре Божечковой); научным редакторам Сергею Синельникову и Марине Турунцевой; редактору Фаине Морозовой и верстальщице Надежде Кильдишевой. (Мы хотим поблагодарить всех за помощь в исправлении некоторых опечаток, которые присутствуют в английском издании и исправлены здесь.) В течение нескольких последних лет Издательский дом «Дело» очень активно занимается переводом и изданием учебников по экономике и эконометрике. Я очень рад, что «Эконометрический анализ» стал одним из них. Я надеюсь, что он вместе с другими переводными учебниками по эконометрике восполнит нехватку русскоязычных учебников по эконометрике и поможет российским студентам и исследователям более глубоко изучить эконометрические методы.

Уильям Грин
Нью-Йорк
Февраль 2016

Предисловие

Эконометрический анализ

Учебник «Эконометрический анализ» представляет собой всеобъемлющее введение в область эконометрики. Эта область постоянно расширяется, об этом свидетельствует список журналов, посвященных если не полностью, то по крайней мере частично эконометрике: *The Journal of Applied Econometrics*, *The Journal of Econometrics*, *The Econometrics Journal*, *Econometric Theory*, *Econometric Reviews*, *Journal of Business and Economic Statistics*, *Empirical Economics*, *Foundations and Trends in Econometrics*, *The Review of Economics and Statistics* и *Econometrica*. Построение учебника таким образом, чтобы изложить эконометрику на магистерском уровне, становится все более амбициозной задачей. Тем не менее я верю в то, что можно успешно решить эту задачу в рамках одного учебника. В данном учебнике я попытался представить на начальном уровне достаточное число разделов эконометрики, чтобы студент мог свободно перейти от изучения эконометрики к практике или к более углубленному изучению каких-либо ее областей. Помимо этого учебник может служить в некотором смысле мостом между введением в область эконометрики и профессиональной литературой для студентов и аналитиков, изучающих социальные науки.

Новое в этом издании

Настоящее седьмое издание представляет серьезный пересмотр более ранних изданий «Эконометрического анализа». Среди наиболее очевидных изменений можно выделить:

- переструктуризацию материала первых глав книги, обычно изучаемых в течение первого семестра, включая следующие:
 - разделы, касающиеся тестирования гипотез и проверки спецификации модели, представлены в одной главе;
 - новые результаты по прогнозированию;
 - большее внимание методу инструментальных переменных и проблеме эндогенности;
 - дополнительные результаты, касающиеся базовой модели панельных данных;
- новые, более подробные приложения и примеры;
- большее внимание конкретным областям применения там, где рассматриваются более продвинутые разделы эконометрики;
- новые разделы, касающиеся методов, основанных на имитационном моделировании данных, особенно бутстрэпа (самонастройки) и метода Монте-Карло;
- некоторые примеры, объясняющие эффект взаимодействия;

- специальные приложения, в том числе квантильная регрессия;
- новые приложения в моделировании дискретного выбора;
- новый материал по проблеме эндогенности и ее последствий для структуры модели.

Седьмое издание «Эконометрического анализа»

У книги две цели. Первая — познакомить студентов с прикладной эконометрикой, в том числе с основными методами оценки модели линейной регрессии и некоторыми моделями, которые используются в ситуации, когда модель линейной регрессии оказывается недостаточной или неподходящей. Современное программное обеспечение существенно облегчило сложный процесс моделирования, но понимание основ теории также очень важно. Вторая цель состоит в том, чтобы дать студентам возможность получить хорошую теоретическую подготовку, чтобы они могли узнать уже изученные здесь модели в новых вариантах и воспринимали эти новые варианты как естественные, вписывающиеся в общие принципы расширения уже известных методов. Эта книга содержит очень много теоретических материалов, таких как GMM, оценка максимального правдоподобия и асимптотические результаты регрессионных моделей. Учебник предназначен для первого курса магистратуры по социальным наукам. Предшествующие, уже изученные курсы должны включать математический анализ, математическую статистику и введение в эконометрику на уровне, соответствующем, скажем, учебникам Гуджарати (Gujarati, 2002) «Basic Econometrics», Стока и Уотсона (Stock, Watson, 2006; 2015) «Introduction to Econometrics», Кеннеди (Kennedy, 2008) «Guide to Econometrics» или Вулдриджа (Wooldridge, 2009) «Introductory Econometrics: A Modern Approach». Я предполагаю, например, что читатель уже знаком с основами методологии эконометрического анализа, в том числе с фундаментальной ролью экономических и статистических предположений, различиями межобъектных данных, временных рядов и панельных данных, с основными этапами оценки моделей, проверки гипотез, статистических выводов и прогнозирования в модели множественной линейной регрессии. Самодостаточные (для наших целей) разделы по матричной алгебре, математической статистике и статистической теории, используемые в книге, можно найти в приложениях A – D. Я использую матричную алгебру на протяжении всей книги. Это может вызвать сложности у некоторых читателей при знакомстве с книгой, но матричная алгебра является незаменимым для наших целей инструментом, и я надеюсь, что читатель согласится, что это лишь средство для достижения этих целей, а не самоцель. Использование матриц позволяет представлять многочисленные результаты в едином виде без знака суммирования. Вся необходимая теория, касающаяся матричной алгебры, представлена в приложении A. Приложение E и глава 15 содержат описание численных методов, которые будут полезны для практикующих эконометристов (и для нас в последних главах книги). Современные компьютерные программы сделали оценивание более сложных нелинейных моделей такой же рутинной, как оценивание методом наименьших квадратов. Я включил пять глав о методах оценки, используемых в современных исследованиях, и пять глав о приложениях в микро- и макроэконометрике. Нелинейные модели, применяемые в этих областях,

в настоящее время являются главными методами, используемыми исследователями и публикуемыми в прикладной эконометрической литературе. Как следствие эта книга содержит большое количество материала, который будет выходить за пределы большинства программ первых курсов эконометрики. И, повторю снова, я включил это в надежде на то, что мне удастся заложить основу для изучения профессиональной литературы в этих областях. В основе появления всех семи изданий этого учебника лежит одна главная цель — чтобы подавляющим большинством его читателей стали пользователи, а не эконометрики, разрабатывающие теоретический аппарат. Поэтому я считаю недостаточным просто излагать теорию оценивания, проверки гипотез и эконометрического анализа. И хотя изощренные теории являются чрезвычайно важными, приложения также необходимы. Для доказательства этого я привожу сотни числовых примеров. Цель написания этой книги и моих продолжающихся усилий при ее обновлении — показать читателям, как проводить эконометрический анализ. Я невозмутимо принимаю нелестные оценки одного из критиков, который однажды сравнил эту книгу с «руководством пользователя по эконометрике».

Книга устроена следующим образом: часть I начинается с формального изучения эконометрики, с ее фундаментальной идеи — модели множественной линейной регрессии. Оценка и статистические выводы на основе линейного метода наименьших квадратов анализируются в главах 2–6. Модель нелинейной регрессии вводится в главе 7 вместе с квантильной, полупараметрической и непараметрической регрессиями как расширение уже знакомой линейной модели. Оценка методом инструментальных переменных рассмотрена в главе 8. В части II представлены три основных расширения регрессионной модели. В главе 9 рассмотрены последствия ослабления одного из основных предположений линейной модели — гомоскедастичности и автокоррелированности случайных ошибок и вводится модель обобщенной регрессии. Акцент здесь делается на гетероскедастичности; автокорреляция упоминается, но детальное ее рассмотрение в контексте моделей временных рядов откладывается до главы 20. В главе 10 рассматриваются системы регрессионных уравнений как подход к одновременному моделированию нескольких случайных величин, а в практическом отношении в качестве расширения модели обобщенной линейной регрессии. И наконец, методы анализа панельных данных, в первую очередь модели с фиксированными и случайными эффектами, представлены в главе 11. Вторая половина книги посвящена разделам, которые расширяют линейную регрессионную модель в нескольких направлениях. Начиная с главы 12 мы перейдем к более сложным методам анализа, которые современные исследователи используют при анализе данных «в реальном мире». В главах 12–16 части III представлены различные методологии оценивания. Глава 12 представляет собой обзор различий между параметрическими, полупараметрическими и непараметрическими методами. Основным приложением полупараметрических оценок в современной литературе является оценка обобщенным методом моментов (ОММ, GMM), представленным в главе 13. Эта методика дает основу для многих современных эконометристов. Оценка максимального правдоподобия рассматривается в главе 14. Метод Монте-Карло и методы, основанные на симулировании данных, такие как бутстрэп (са-

монастройка), ставшие основной частью современных исследований, разрабатываются в главе 15. И наконец, байесовские методы вводятся в главе 16. В частях IV и V развиваются два основных подполя эконометрических методов: микроэконометрика, имеющая, как правило, дело с межобъектными и панельными данными, и макроэконометрика, которая обычно ассоциируется с методами анализа временных рядов. В части IV, в главах 17–19, рассмотрены модели дискретного выбора, усеченные и цензурированные выборки, выборочный отбор, влияние воздействия в эксперименте и модели для числа событий. В части V, в главах 20 и 21, мы рассмотрим два раздела анализа временных рядов: модели серийной корреляции и модели регрессий для нестационарных данных — обычные методы макроэкономического анализа.

Обновления и изменения структуры учебника

Я существенно перестроил начало книги, чтобы сделать более естественной последовательность разделов для магистерского курса эконометрики. Глава 4 теперь полностью посвящена точечному и интервальному оцениванию, в том числе предсказанию и прогнозированию. Свойства метода наименьших квадратов в конечных выборках, а затем и асимптотические свойства рассмотрены подробно. Все, что касается проверки гипотез и тестирования спецификации модели, теперь находится в главе 5, а не является фрагментами нескольких глав, как в шестом издании. Я также поставил материал об инструментальных переменных намного раньше (глава 8), сразу же после полного описания линейной регрессионной модели, а не так, как это было в шестом издании, после введения обобщенной регрессионной модели. Это согласуется с большим значением этого метода в последних приложениях. Очень большое число других изменений также очевидно. Глава 7 теперь содержит ряд продвинутых расширений модели линейной регрессии, в том числе нелинейной, квантильной, частично линейной и непараметрической регрессии. Здесь также можно проанализировать различия между параметрическими, полупараметрическими и непараметрическими методами. Одна заметная модификация — сокращение длинной главы о системах линейных одновременных уравнений. Некоторые материалы из этой главы появляются в других местах. Двухшаговый метод наименьших квадратов теперь появляется вместе с методом инструментальных переменных. Остальные разделы этой главы, которые имели меньшее значение в недавних исследованиях, таких как условия ранга и порядка для идентификации линейных моделей и оценки 3SLS и FIML, были удалены или значительно сокращены и помещены в других разделах книги. Материал, касающийся моделей дискретного выбора, также был изменен. В главе 17 теперь подробно описаны модели дискретного выбора и модели со случайной функцией полезности, а затем рассмотрены в вариантах моделей бинарного выбора. Анализ продолжается в главе 18 для моделей неупорядоченного множественного выбора, затем для моделей упорядоченного выбора и, наконец, для моделей числа событий. В последней главе этой части изучаются модели для непрерывных переменных в контексте конкретных механизмов порождения данных и поведения. Я добавил новый материал и несколько различных примеров и приложений во множество мест. Разделы, которые были расширены или ко-

торым было уделено больше внимания, включают эффекты в экспериментах, бутстрэп (самонастройку), оценки, основанные на имитационном моделировании, устойчивые методы оценки, пропущенные и ошибочные данные и различные новые методы анализа дискретного выбора в микроэконометрике. Я также добавил или расширил материал, касающийся методов, интерес к которым возник не так давно, таких как квантильная регрессия и модели стохастической границы. Отмечу несколько конкретных изменений и обновлений: говоря общими словами, я немного усилил акцент на устойчивых (робастных) методах оценки. Обсуждение тестов на спецификацию модели появляется в нескольких местах в соответствии с присутствующей в литературе тенденцией изучать более внимательно слабости сильно параметризованных моделей. Вот некоторые новые приложения:

- Глава 15 по оценке на основе имитационного моделирования была значительно расширена. В настоящее время она включает намного больше материала об использовании бутстрэпа при расчете стандартных ошибок и доверительных интервалов. Подход Крински и Робб (Krinsky, Robb, 1986) для асимптотических статистических выводов также присутствует в ней.
- Большое внимание уделено недавним работам о понимании эффектов взаимодействия в нелинейных моделях. В главе 7 можно найти приложение модели с эффектами взаимодействия в нелинейной (экспоненциальной) модели регрессии. Этот вопрос вновь обсуждается в главе 17.
- В качестве упражнения, которое будет вызовом для студентов при изучении теории асимптотических распределений, я добавил детальное доказательство результата Мерфи и Топеля (Murphy, Topel, 2002) для двухшаговой оценки в главе 14.
- Обсуждение источников и влияния эндогенности появляется в различных местах, например, приложение оценки с весами, обратно пропорциональными вероятности при истощении выборки, встречается в главе 17. Седьмое издание является серьезным пересмотром «Эконометрического анализа» с точки зрения и изложения материала, и появления новых идей и трактовок. Я надеюсь, что читатели найдут изменения полезными.

Программное обеспечение и базы данных

Существует много компьютерных программ, которые широко используются для расчетов, описанных в этой книге. Все они были написаны эконометристами или статистиками и в общем регулярно обновляются, включая новые методы в области прикладной эконометрики. Ниже представлена выборка из наиболее широко используемых пакетов и их домашних страниц, на которых вы можете найти информацию о них:

EViews www.eviews.com (QMS, Irvine, CA); Gauss www.aptech.com (Aptech Systems, Kent, WA); LIMDEP www.limdep.com (Econometric Software, Plainview, NY); MATLAB www.mathworks.com (Mathworks, Natick, MA); NLOGIT www.nlogit.com (Econometric Software, Plainview, NY); R www.r-project.org/ (The R Project for Statistical Computing); RATS www.estima.com (Estima, Evanston, IL);

SAS www.sas.com (SAS, Cary, NC); Shazam econometrics.com (Northwest Econometrics Ltd., Gibsons, Canada); Stata www.stata.com (Stata, College Station, TX); TSP www.tspintl.com (TSP International, Stanford, CA).

Более полный список программных пакетов, используемых для эконометрического анализа, можно найти на веб-сайте ресурсов <http://www.oswego.edu/economic/econsoftware.htm>. За небольшим исключением расчеты, описанные в этой книге, могут быть осуществлены в любом из перечисленных пакетов. Я использовал NLOGIT для вычислений в примерах. В учебнике нет инструкции по использованию какого-либо конкретного программного пакета или языка программирования. (На веб-сайте автора учебника приведены некоторые коды и данные для повторения численных примеров.) Многие авторы давали коды, написанные в RATS, LIMDEP/NLOGIT, EViews, SAS или Stata, для некоторых наших примеров, в том числе документацию для своих компьютерных программ. Сейчас существует немало томов, посвященных решению конкретных эконометрических задач в конкретных пакетах, например приложение к книге Кэмерона и Триведи (Cameron, Trivedi, 2009) по микроэконометрике. Базы данных, используемые в примерах, также доступны на веб-сайте учебника <http://pages.stern.nyu.edu/wgreene/Text/econometricanalysis.htm>. В учебнике эти базы данных называются «Таблица Fn.m», например таблица F4.1. «F» относится к приложению F в конце текста, которое содержит описание данных. Фактические данные размещены в общем ASCII и конвертируемых форматах на веб-сайте вместе с другими дополнительными материалами учебника. В настоящее время есть тысячи интересных веб-сайтов, содержащих программное обеспечение, данные, исследования и комментарии по эконометрике. Было бы безнадежно пытаться привести все это здесь. Один из сайтов с кодами/данными, который особенно хорошо структурирован и направлен на читателей этого учебника, — это архив данных Journal of Applied Econometrics. В нем архивируются все неконфиденциальные базы данных, используемые в публикациях журнала с 1988 г. (с некоторыми пробелами до 1995 г.). Этот полезный сайт можно найти по адресу: <http://qed.econ.queensu.ca/jae/>. В нескольких примерах в тексте использовались базы данных JAE. Когда мы делали это, то отправляли читателя к веб-сайту JAE, а также воспроизводили результаты сами. Остальные журналы стали просить своих авторов предоставлять коды и данные, чтобы стимулировать исследователей воспроизводить уже полученные результаты. Другим хорошим и легким с точки зрения навигации сайтом с агрегированными данными по экономике США является www.economagic.com.

Благодарности

Мне приятно выразить свою признательность тем, кто повлиял на эту работу. Я по-прежнему благодарен покойным Артуру Голдбергеру и Арнольду Зелнеру, а также Деннису Айгнеру, Биллу Беккеру и Лоритсу Кристенсену за их поддержку и помощь. После семи изданий этой книги число людей, которые значительно улучшили ее своими комментариями, критикой и поддержкой, стало слишком большим, чтобы поблагодарить каждого из них в отдельности. Я благодарен за помощь и надеюсь, что все они увидят свой вклад в это издание. Я хотел бы отметить всех рецензентов моей рабо-

ты, чье внимательное прочтение значительно улучшило книгу в этой редакции: Скотт Эткинсон, Университет штата Джорджия; Бади Балтаги, Сиракузский университет; Нил Бек, Нью-Йоркский университет; Уильям Е. Беккер, Университет Индианы; Эрик Дж. Беласко, Технический университет Техаса; Анил Бера, Университет штата Иллинойс; Джон Беркетт, Университет Род-Айленда; Леонард Карлсон, Университет Эмори; Фрэнк Чалупка, Университет штата Иллинойс в Чикаго; Крис Корнуэлл, Университет штата Джорджия; Крейг Депкен II, Университет Техаса в Арлингтоне; Фрэнк Диболд, Университет Пенсильвании; Эдвард Дуайер, Университет Клемсона; Майкл Эллис, Университет Уэсли; Мартин Эванс, Джорджтаунский университет; Ваагн Галстян, Тринити-колледж Дублина; Пол Глевве, Университет Миннесоты; Эд Гринберг, Вашингтонский университет в Сент-Луисе; Мигель Херсе, Университет Северной Каролины; Джозеф Хилбе, Государственный университет Аризоны; д-р Уве Йенсен, Университет Кристиана-Альбрехта; Рао Кадыяла, Университет Пердью; Уильям Лотт, Университет штата Коннектикут; Томас Л. Марш, Университет штата Вашингтон; Эдвард Матис, Университет Вилланова; Мэри Мак-Гарви, Университет Небраски-Линкольна; Эд Мельник, Нью-Йоркский университет; Тэд Мирер, Государственный университет Нью-Йорка в Олбани; Сирил Паше, Женевский университет; Пол Рууд, Калифорнийский университет в Беркли; Шерри Рейн, Федеральная корпорация по страхованию вкладов; Терри Г. Сикс, Университет Северной Каролины в Гринсборо; Дональд Снайдер, Университет штата Калифорния в Лос-Анджелесе; Стивен Стерн, Университет Вирджинии; Хьюстон Стокс, Университет штата Иллинойс в Чикаго; Дмитриос Томакос, Колумбийский университет; Поль Вахтель, Нью-Йоркский университет; Мэри Бет Уокер, Университет штата Джорджия; Марк Уотсон, Гарвардский университет; Кеннет Вест, Университет штата Висконсин. Мои многочисленные беседы с Брюсом Маккалоу из Университета Дрексела улучшили Приложение Е и одновременно увеличили мое понимание численного анализа. Я особенно благодарен Яну Кивьету из Университета Амстердама, который подверг микроскопическому анализу третье издание учебника и внес десятки предложений, почти все они учтены в данном издании. Профессор Педро Бакао из Университета Коимбры в Португалии и Марк Страхан из «Сэнд-Хилл Эконометрикс» сделали то же самое в шестом издании. Многие близкие мне люди, особенно моя семья, а также многие другие поддерживали меня все эти годы. Ничто не было более приятно, чем полученные по почте письма от читателей со всего мира, которые разделяли мой энтузиазм в этой захватывающей области и в этой работе, научившей их и меня эконометрике с момента первого издания в 1990 г. Наконец, я хотел бы поблагодарить многих людей в издательстве «Прентис Холл», которые работали над этой книгой вместе со мной: Адриен Д'Амброзио, Джил Колонговски, Карлу Томпсон, Нэнси Фентон, Элисон Юсден, Джо Ветере и Марту Уэзерил, а также команду «Макмиллан Паблишинг Солюшнс».

Уильям Грин

Август 2010

ГЛАВА 1

Эконометрика

1.1. Введение

Эта книга представляет собой вводный обзор эконометрической науки. Мы обсудим ключевые идеи, определяющие методологию эконометрического исследования, и рассмотрим множество моделей, инструментов и методов, которыми эконометристы пользуются для анализа данных. В этой главе описываются основные идеи, составляющие парадигму эконометрики. В разделе 1.2 определим круг вопросов, изучаемых эконометрикой, и обсудим, какую роль играет теория при решении эконометрических задач на практике. В разделе 1.3 описываются основные виды задач, к которым применяются эконометрические методы. В разделе 1.4 вводится понятие эконометрического моделирования и описывается классическое приложение: кейнсианская функция потребления. В разделе 1.5 дается краткий план книги. В разделе 1.6 мы обращаем внимание читателей на некоторые особенности изложения материала, например на использование в тексте численных примеров и математических обозначений.

1.2. Парадигма эконометрики

В первом номере журнала *Econometrica* Рагнар Фриш [Ragnar Frisch (1933)] писал об Эконометрическом обществе (Econometric Society) следующее:

Его основной целью будет распространение знаний, способствующих объединению количественного, теоретического и эмпирического подходов к экономическим задачам и соответствующих стандартам точности и конструктивности, принятым в естественных науках. Однако существует множество разнообразных методов количественного анализа экономических задач и ни один из них в отдельности не следует путать с эконометрикой в целом. Таким образом, эконометрика — это не то же самое, что, например, экономическая статистика. Не совпадает эконометрика и с тем, что обычно называют общей экономической теорией, хотя значительная часть этой теории имеет ярко выраженный численный характер. Также нельзя считать эконометрику синонимом [*sic*] приложения математических методов к экономике. Опыт показал, что каждый из этих трех подходов — статистический, теоретико-экономический и математический — необходим, но сам по себе не является достаточным для полного понимания численных соотношений в современной экономике. Мощным

методом их описания будет только *объединение* всех трех подходов. Это объединение и составляет эконометрику.

Создание Эконометрического общества было реакцией на беспрецедентный рост темпов накопления статистической информации. Стала ясна необходимость установления базовых принципов, с помощью которых можно было бы как-то упорядочить эту массу данных. Ни ключевые подходы, ни основные цели не изменились с момента публикации приведенной выше цитаты. Эконометристы связывают свою деятельность с применением методов математической статистики и различных инструментов статистических выводов к практическому измерению величин и соотношений, которые постулируются экономической теорией.

Важнейшая роль эконометрических методов в экономике со временем только усиливалась. Нобелевскую премию по экономике многократно вручали эконометристам: впервые ее получил Рагнар Фриш [Ragnar Frisch] (автор приведенной выше цитаты) в 1969 г., а позже Лоуренс Клейн [Lawrence Klein] в 1980 г., Трюгве Ховельмо [Trygve Haavelmo] в 1989 г., Джеймс Хекман [James Heckman] и Дэниел Макфадден [Daniel McFadden] в 2000 г., а также Роберт Энгл [Robert Engle] и Клайв Грейнджер [Clive Granger] в 2003 г. Премия 2000 г. выделялась тем, что отмечала работу двух ученых, соединивших теорию, лежащую в основе поведенческой экономики, с эконометрическим моделированием.

Пример 1.1. Модели поведенческой экономики и нобелевские лауреаты

Теоретическим фундаментом новаторских работ Джеймса Хекмана (James Heckman) и Дэна МакФаддена (Dan McFadden) была теория максимизации полезности. Работа Хекмана начинается со стандартной модели домохозяйства, максимизирующего полезность путем балансирования потребления и досуга. Обычная модель базового учебника определяет спрос на досуг, который затем превращается в функцию предложения на рынке труда. Если домашнее производство (работа, выполняемая внутри домохозяйства без выхода на внешний формальный рынок труда) учитывается в модели, то число желаемых «часов» досуга может оказаться отрицательным. Определяющей переменной служит «резервная» заработная плата — уровень заработной платы, при котором у членов домохозяйства возникает желание участвовать в формальном рынке труда. Со стороны спроса на рынке труда фирмы предлагают рыночный уровень заработной платы, зависящий от таких параметров, как возраст работника, уровень его образования и опыт. Что можно сказать о поведении на рынке предложения труда, основываясь на наблюдаемых рыночных зарплатах, перечисленных переменных и наблюдаемых отработанных часах на формальном рынке труда? Меньше, чем могло бы показаться на первый взгляд: интуитивно ясно, что наши наблюдения упускают половину рынка, поскольку данные об активности на формальном рынке труда не являются случайными для всего населения.

Наблюдения Хекмана относительно этого неявного урезания выборки зарплат и часов работы перевернули всю область исследования рынков

труда. Интерпретации подобных наблюдений в других социальных науках также стали важными направляющими исследований. Методы изучения влияния различных мер, как, например, инициатив в области образования, обучения, борьбы с безработицей, медицинского страхования, создания рынков, финансового регулирования и многих других, теперь в значительной степени основываются на новаторских идеях Хекмана о том, что если факт участия служит частью изучаемого явления, то исследователь должен изучать эффект влияния различных факторов как на само участие, так и на его исход. Мы рассмотрим литературу о проблемах, возникающих из-за отбора наблюдений, и о влиянии различных программ и воздействий в главе 18.

Описания теорий спроса на товары, создающего полезность, приводимые в учебниках, будучи основаны на непрерывных переменных, обычно ничего не говорят о дискретных решениях, которые ежедневно принимают экономические агенты: какую выбрать марку товара, покупать ли какой-то дорогой товар вроде холодильника или автомобиля, как добираться до работы, покупать или снимать жилье, где жить, за какого кандидата голосовать и т.д. Однако модель «случайной полезности», определенной на тех вариантах, между которыми моделируется выбор потребителя, дает прочную теоретическую базу для исследования. В число важных переменных, как обычно, входят доход и относительные цены товаров. Что мы можем узнать о структуре предпочтений потребителей из сделанного ими выбора? Какие предположения необходимо сделать в отношении этих предпочтений, чтобы иметь возможность делать статистические выводы? Какого вида статистические модели могут позволить нам строить выводы о предпочтениях потребителей? Работа МакФаддена о том, как люди выбирают способ, которым будут добираться до работы, и о том, какая теоретическая база лежит в основе подобного моделирования, направляла эмпирические исследования дискретного выбора потребителей в течение нескольких десятилетий. Модели дискретного выбора МакФаддена будут рассмотрены нами в главе 18.

Связь между поведенческими моделями и современной эконометрической практикой становится все более сильной. Можно провести полезное разделение между *микроэконометрикой* и *макроэконометрикой*. Первая обычно изучает межобъектные либо панельные данные, касающиеся отдельных потребителей, фирм и других агентов на микроуровне. Исследователи активно используют теоретический инструментарий микроэкономики, включая теории максимизации полезности, максимизации прибыли и рыночного равновесия. Исследования направлены на тонкие и сложные вопросы, часто требующие нетривиальных формулировок. Вот некоторые приложения:

- Каково возможное влияние на предложение на рынке труда отрицательных подоходных налогов? [Ashenfelter and Heckman (1974).]
- Влияет ли учеба в элитном колледже на ожидаемое увеличение будущего пожизненного дохода, достаточного для оправдания более высоких затрат на обучение? [Kreuger and Dale (1999) and Kreuger (2000).]

- Дает ли какие-либо положительные результаты создание добровольной программы обучения? Могут ли эти результаты быть точно измерены? [Angrist (2001).]
- Повышает ли успеваемость уменьшение числа учеников в классе? [Hanushek (1999), Hoxby (2000), Angrist and Lavy (1999).]
- Побуждает ли наличие медицинской страховки людей к тому, чтобы они чаще обращались к врачу: возможно ли измерить риск недобросовестного поведения? [Riphahn et al. (2003).]

Макроэконометристы исследуют в основном данные временных рядов, причем, как правило, описывающих агрегированные данные, такие как уровень цен, предложение денег, обменные курсы, объем производства, уровень инвестиций, экономический рост и т.д. Граница между микро- и макроэконометрикой не является строгой. Так, например, ниже будет рассматриваться лежащее где-то между этими категориями приложение, описывающее траты муниципальных объединений. Значительная часть финансовой эконометрики изучает длинные временные ряды (а иногда очень большие панели), но при этом фокусируется на моделях, объясняющих поведение отдельных индивидов. Анализ рыночных уровней доходности или поведения обменных курсов не является частью исключительно микро- или макроэконометрики. (В этой книге финансовая эконометрика рассматриваться не будет. Читателям, интересующимся этой областью, мы можем порекомендовать известную книгу Campbell, Lo, and Mackinlay (1997) или Tsay (2005), где больше внимания уделяется анализу временных рядов.) Исследователи, занимающиеся построением макроэкономических моделей, обычно опираются на взаимодействие экономических агентов с властями, определяющими экономическую политику, например, по вопросам:

- Приводит ли к снижению реального выпуска в экономике США монетарная политика, сильно ориентированная на контроль над инфляцией? [Cecchetti and Rich (2001).]
- Привело ли сокращение федеральных налогов в США в 2001 г. (самое сильное в истории) к усилению или ослаблению тогдашней экономической рецессии? Или оно вообще не повлияло на ее ход?

Изучение любого из этих вопросов будет начинаться с рассмотрения формальной модели процесса, порождающего наблюдаемые данные.

1.3. Практическая эконометрика

Еще одну границу можно провести между *теоретической* и *прикладной эконометрикой*. Теоретики разрабатывают новые техники построения оценок и проверки гипотез, а также изучают возможные последствия применения тех или иных методов в ситуациях, когда предположения, используемые этими методами, не выполняются для имеющихся данных. Прикладные эконометристы пользуются этими методами и анализируют данные («реальные» и придуманные). Граница опять не является четкой: практики

часто создают новые методы специально для той проблемы, которой занимаются. Данная книга содержит много эконометрической теории, но направлена в сторону прикладной эконометрики. Мы попытались сделать обзор общеупотребительных методов, иногда довольно сложных, которые активно используются в реальных исследованиях.

Прикладные эконометрические методы используются для оценки важных величин, анализа результатов изменения экономической политики, изучения рынков и экономического поведения индивидов, проверки гипотез и прогнозирования. Последнее представляет собой отдельную науку или даже искусство и служит темой для широкого класса исследований. В этой книге мы рассмотрим некоторые вопросы прогнозирования, но в основном будем заниматься оценкой и анализом моделей. В тех случаях, когда такое разделение имеет смысл, будем стараться продемонстрировать и микро-, и макроэконометрические методы и приложения. Также следует выделить *анализ временных рядов* (не являющийся основной темой этой книги) и те методы, которые в основном используют данные временных рядов. Анализ временных рядов, как и прогнозирование, является самостоятельной быстроразвивающейся дисциплиной, которой посвящен большой объем литературы в различных областях. Хотя мы и будем использовать некоторые методы анализа временных рядов, но не будем уделять слишком большого внимания обоснованию этих методов.

1.4. Эконометрическое моделирование

Эконометрический анализ обычно начинается с формулировки теоретического утверждения. Возьмем, например, классическую модель одного из современников Фриша.

Пример 1.2. Кейнсианская функция потребления

Процитируем работу Кейнса (1936) *General Theory of Employment, Interest and Money*:

Мы будем называть склонностью к потреблению функциональную зависимость f между уровнем дохода X и расходами на потребление C при этом уровне дохода, т.е. $C = f(X)$. Сумма, которую общество затрачивает на потребление, зависит (i) частично от уровня дохода, (ii) частично от других наблюдаемых условий окружающего мира и (iii) частично от субъективных потребностей, психологических особенностей и привычек индивидов, из которых состоит это общество. Базовый психологический закон, на который мы можем полагаться с полной уверенностью как априорно исходя из наших знаний о человеческой природе, так и из опыта, состоит в том, что люди, как правило, увеличивают потребление при росте дохода, но не на всю величину роста дохода. Значит... dC/dX положительна и меньше единицы.

Однако ясно, что помимо кратких периодов быстрого изменения дохода более высокий абсолютный уровень доходов будет, как правило, увеличивать разницу между доходом и потреблением.... По этим причинам доля дохода, сохраняемая в виде сбережений, будет, скорее всего, увеличиваться при росте реального дохода.

Следовательно, теория предполагает наличие зависимости между потреблением и доходом: $C = f(X)$, а во втором абзаце утверждает, что предельная склонность к потреблению (MPC, marginal propensity to consume), dC/dX , находится между нулем и единицей¹. Последний абзац содержит утверждение о том, что средняя склонность к потреблению (APC, average propensity to consume), C/X , убывает при росте дохода, т.е. $d(C/X)/dX = (MPC - APC)/X < 0$. Следовательно, $MPC < APC$. Наиболее часто в качестве функции потребления используют линейную функцию $C = \alpha + X\beta$, которая удовлетворяет «законам» Кейнса при значениях β между нулем и единицей и значениях α , больших нуля.

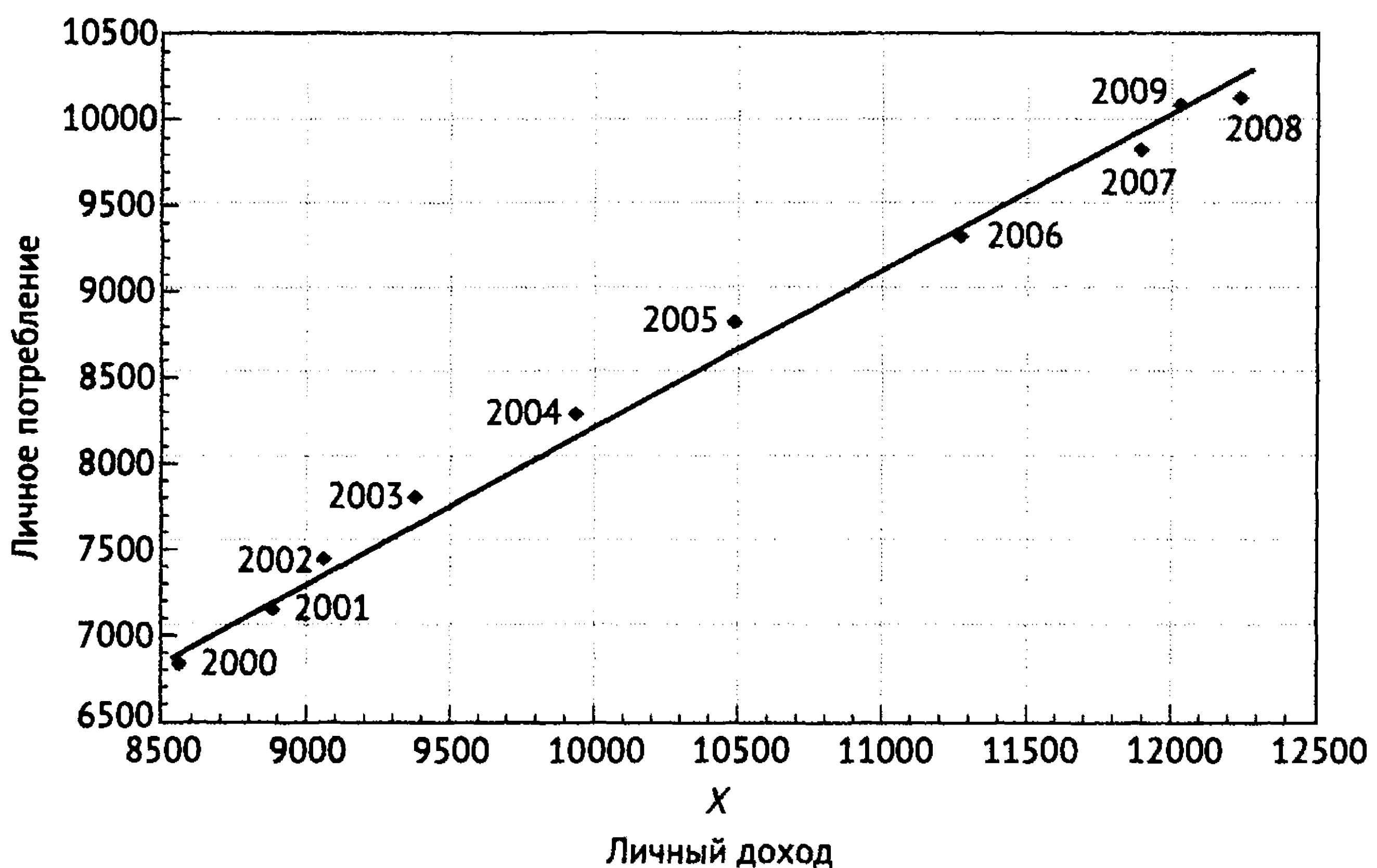


Рис. 1.1. Агрегированное потребление и доходы в США, 2000–2009 гг.

Эти теоретические предположения являются базой эконометрического исследования. При наличии подходящих данных мы могли бы изучить, соответствует ли эта теория наблюдаемым «фактам». Например, можно было бы проверить, будет ли линейная функция удовлетворительным описанием зависимости между потреблением и реальным доходом и, если это так, является ли величина α положительной и лежит ли β между нулем и единицей. Можно также исследовать, (1) является ли зависимость неизменной во времени или изменяется от одного поколения к следующему (изменение средней склонности к сбережению, равной $1 - APC$, может свидетельствовать о существенном изменении экономического поведения потребителей); (2) существуют ли систематические различия в поведении потребителей в различных странах и, если это так, чем они

могут объясняться; (3) существуют ли другие факторы, включение которых могло бы улучшить способность модели объяснить связь между уровнями потребления и дохода. Например, на рис. 1.1 показаны агрегированные данные об уровне личного потребления и о личных доходах в постоянных ценах в США за десятилетие 2000–2009 гг. (см. табл. F1.1 в приложении.)

Кажется, по крайней мере на первый взгляд, что имеющиеся фактические данные укладываются в теорию. Зависимость, хоть и приблизительно, похожа на линейную, точка пересечения прямой, лучше всего приближающей данные, с осью абсцисс положительна, а наклон меньше единицы, хоть и ненамного. (Впрочем, если оценивать линейную регрессию методом наименьших квадратов, то свободный член окажется отрицательным, а не положительным.)

Экономические теории, подобные описанной выше кейнсианской теории, обычно точны и конкретны. Модели спроса, производства, предложения на рынке труда, потребительского выбора, результатов повышения уровня образования, дохода и заработной платы, инвестиций, рыночного равновесия и агрегированного уровня потребления задают точные, *детерминированные* соотношения. Определяются зависимые и независимые переменные и вид связывающей их функции, а также, как правило, даются качественные предсказания относительно направления влияния, происходящего при изменении независимых переменных. Модель представляет собой упрощение действительности. Модель будет описывать наиболее заметные аспекты рассматриваемого соотношения, но может упускать имеющие место влияния, которые исследователь счел маловажными.

Корреляции между экономическими переменными можно легко обнаружить с помощью описательных статистик и методов, подобных линейной регрессии. Конечной целью большинства эконометрических моделей служит обнаружение более глубоких и сложных причинно-следственных связей с помощью структурных поведенческих моделей. Такой, например, является модель поведения «репрезентативного потребителя», предложенная Кейнсом для объяснения поведения макроэкономических переменных, таких как доход и уровень потребления. Построенная Хекманом модель предложения на рынке труда (см. пример 1.1) также определяется в терминах поведения отдельного индивида. В работе Berry, Levinsohn and Pakes (1995) авторы строят модель равновесного ценообразования на автомобильном рынке схожим образом.

Ни одна модель не может претендовать на то, чтобы охватить все случайные аспекты экономической действительности. Значит, необходимо включать в модели случайные компоненты. Как следствие наблюдаемые значения переменной будут отличаться не только из-за разных значений факторов, напрямую включенных в модель, но и из-за случайности, содержащейся в поведении людей, и из-за влияния массы факторов, не включенных в модель. Введение в детерминированную модель случайного «возмущения»

¹Современные экономисты, как правило, не настолько уверены в своих теориях. Современные приложения обычно исходят из базовых принципов и аксиом поведенческой теории, а не из простых наблюдений.

необходимо не только для того, чтобы скрыть неадекватность модели, но и для того, чтобы при последующем анализе убедиться, что этот случайный необъясненный фактор действительно не поддается объяснению. Если это не так, то модель и в самом деле является неадекватной. [В примере, приведенном выше, оценка свободного члена в линейной регрессии методом наименьших квадратов оказывается отрицательной. Вызвано ли это неадекватностью теории или случайными флуктуациями данных? Возможно также, что теория в целом верна, но между 1936 г., когда Кейнс ее формулировал, и периодом 2000–2009 гг., когда собирали данные, произошли какие-то изменения. Также возможно, что метод наименьших квадратов не подходит для оценки этой модели и плохой результат (отрицательный свободный член) вызван именно этим.] Случайный элемент придает модели статистические свойства. Мы считаем, что наблюдаемые значения исследуемых переменных получены как выборка из некоторого случайного процесса. При наличии достаточно определенной стохастической структуры и подходящего набора данных анализ сводится к выводу свойств распределения вероятностей. Для этого в нашем распоряжении имеется инструментарий математической статистики.

Любой модели (или теории) нельзя доверять целиком и полностью, если она не включает абсолютно все возможности. Но любую модель можно подвергнуть внимательному изучению и, если обнаружатся противоречия, отвергнуть. Детерминированная теория может быть разрушена одним противоречащим наблюдением. Дополнение модели случайными элементами превращает ее из точного утверждения в вероятностное описание ожидаемых исходов, откуда следует важный вывод. Теперь модель может быть опровергнута только при накоплении некоторой «критической массы» наблюдений, не укладывающихся в теорию. Какой именно должна быть «критическая масса» — вопрос, ответ на который субъективен. Таким образом, вероятностные модели, с одной стороны, менее точны, но с другой — более устойчивы².

Эконометрические методы используются в разных областях: в политике, социологии [см., напр., Long (1997) или DeMaris (2004)], экономике здравоохранения, медицинских исследованиях (что делать с людьми, отказавшимися от лечения до завершения испытания?), экономике окружающей среды, экономической географии, транспортном строительстве и многих других. Методы, описанные в этой книге, широко используются во всех этих областях.

Процесс эконометрического анализа начинается с формулировки какой-либо теоретической зависимости. Вначале мы будем оптимистично считать, что нам доступны точные измерения всех переменных верно сформулированной модели. Если на всех этапах выполнены эти идеальные условия, то анализ можно провести без всяких осложнений. К сожалению, так бывает редко. Среди прочих трудностей нам могут встретиться такие:

- Данные могут быть измерены с ошибками или соответствовать переменным теоретической модели лишь приблизительно. Так бывает, напри-

²См. работу Keuzenkamp, Magnus (1995), в которой приводится подробное обсуждение проверки гипотез в эконометрике.

мер, при работе с «процентными ставками».

- Некоторые переменные могут быть вообще ненаблюдаемы. К ним относятся, например, «ожидания».
- Теория может давать лишь общие черты верной модели или не давать даже этого; в этом случае мы будем вынуждены выбирать из длинного списка возможных моделей.
- Предполагаемые стохастические свойства случайных членов в модели могут явным образом нарушаться, что ставит под сомнение применяемые методы оценивания и статистические выводы.
- Некоторые важные переменные могут отсутствовать в модели.
- Условия сбора данных были такими, что полученная выборка систематически неверно представляет репрезентативную выборку из той группы (генеральной совокупности), которая представляет интерес в исследовании.

Дальнейшие этапы эконометрического анализа состоят в борьбе с этими проблемами и попытках понять, какая же информация содержится в несовершенной выборке. При этом используются методы математической статистики и экономической теории. Результатом этого процесса является эконометрическая модель.

1.5. План книги

Эконометрика — большая и продолжающаяся увеличиваться область знаний. И довольно сложно написать курс для начинающих ее изучение. Цель этой книги — сначала выработать некоторый инструментарий, а потом научиться применять его в приложениях. Мы рассмотрим широкий, но не всеобъемлющий набор приложений; большинство читателей наверняка столкнется с некоторыми из них на практике. Мы будем пытаться приводить достаточно общие результаты, чтобы читатели могли применять разобранные здесь методы в других ситуациях и приложениях.

Один из вариантов состоит в том, чтобы сгруппировать области изучения по данным, к которым обычно применяются рассматриваемые методы: межобъектные данные (кросс-секционные данные, перекрестные выборки), панельные данные, дискретные данные или данные временных рядов. Также можно было бы сразу разделить микро- и макроэконометрику³. Однако в каждой из этих областей необходим некоторый общий инструментарий: модель множественной регрессии, использование моментных условий при

³Методы микроэконометрики очень хорошо и на несколько более продвинутом уровне, чем в этой книге, изложены в работе Cameron, Trivedi (2005). Пока, насколько нам известно, нет всеобъемлющего пособия по макроэконометрике, включающего и теорию, и приложения. Из более специальных работ можно привести книгу Bardsen, G., Eitrheim, Ø., Jansen, E., and Nymoен, R. «The Econometrics of Macroeconomic Modelling» (Oxford University Press, 2005) и обзорные статьи Wallis, K. «Macroeconometric Models» // Macroeconomic Policy: Iceland in an Era of Global Integration / M. Gudmundsson, T.T. Herbertsson, and G. Zoega, eds., p. 399–414 (Reykjavik: University of Iceland Press, 2000); см. также: http://ecomod.net/conferences/ecomod2001/papers_web/Wallis_Iceland.pdf.

оценивании, инструментальные переменные (IV, instrumental variables), оценивание методом максимального правдоподобия. Учитывая это обстоятельство, мы решили построить эту книгу следующим образом: в первой ее половине разрабатываются фундаментальные понятия, необходимые для любого приложения. Понятие множественной регрессии и особенно модель линейной регрессии служат основой для большинства приложений, даже если сама линейная модель и не является окончательной эмпирической спецификацией. Эта часть книги заканчивается методами IV-оценивания и общим обзором методов моделирования панельных данных. Панельные данные используют широкий набор методов современной эконометрики, включая опять IV-оценивание, моделирование гетерогенности и широкий ассортимент различных обобщений линейной модели. Во второй половине книги рассматриваются различные темы. В части III дается обзор методов оценивания. Части IV и V посвящены соответственно микро- и макроэконометрике. Общий план книги выглядит так:

I. Регрессионные модели

Главы 2–6 посвящены модели множественной линейной регрессии. Мы поговорим о спецификациях таких моделей, об оценке их параметров и о статистических выводах из результатов оценивания. В этой части обсуждаются такие понятия, как *методы оценки, робастный анализ, функциональная форма и основные принципы спецификации моделей.*

II. Обобщенная регрессия, инструментальные переменные и панельные данные

В главе 7 модель регрессии распространяется на нелинейные функциональные формы. В главе 8 вводится метод инструментальных переменных. Главы 9 и 10 представляют обобщенную модель регрессии и системы регрессионных уравнений. Эта часть заканчивается обсуждением методов работы с панельными данными в главе 11.

III. Методы оценивания

В главах 12–16 приведены общие результаты, касающиеся таких методов оценивания, как GMM, метод максимального правдоподобия, и симуляционных методов. В главах 12 и 16 представлены различные формы оценивания, включая не- и полупараметрические методы, а также байесовские методы.

IV. Методы микроэконометрики

Главы 17–19 посвящены микроэконометрике, моделям дискретного выбора и моделям с ограниченными зависимыми переменными, а также изучению данных о событиях: сколько событий данного типа происходит при данных обстоятельствах и когда они происходят. Методы, которым посвящены эти главы, лучше всего подходят к межобъектным и панельным данным.

V. Методы макроэконометрики

В главах 20 и 21 мы будем говорить об анализе временных рядов и о макроэконометрике.

VI. Дополнительный материал

В приложениях с А по Е излагаются необходимые базовые сведения из используемых в эконометрике областей матричной алгебры, теории вероятностей, статистического оценивания и асимптотической теории.

В приложении Е приведены некоторые результаты, связанные с вычислениями. Во всех приложениях даются краткие обзоры инструментария, необходимого в эконометрике. Поскольку мы предполагаем, что читатель уже обладает некоторыми познаниями в этих областях, эти разделы приведены лишь для освежения памяти и удобства ссылок. Мы не предполагаем, что эти разделы могут заменить курс по любой из включенных дисциплин. Наша цель — дать краткое изложение результатов, явно используемых в основных частях книги. Данные, используемые в численных примерах, приведены в приложении F. Также данные и другие дополнительные материалы можно скачать со страницы автора по адресу: <http://pages.stern.nyu.edu/~wgreene/Text/>.

1.6. Предварительные замечания

Прежде чем приступить к изложению, отметим некоторые особенности представления информации в книге.

1.6.1. Численные примеры

В данной книге приводится множество численных примеров. Большинство из них являются либо независимыми упражнениями, либо фрагментами опубликованных исследований. Как правило, цель этих примеров — проиллюстрировать небольшое приложение метода или модели. При желании читатель может самостоятельно воспроизвести полученные результаты с помощью прилагаемых данных. Обычно это не требует воспроизведения всей опубликованной работы. В примерах будут разбираться в основном приложения, лишь ограниченно связанные с первоначальной работой, но зато акцентирующие внимание на определенном методе или модели. Например, в работе Riphahn, Wambach, Million (2003) приведен очень полезный и удобный (при сравнительно больших размерах) массив данных, на котором читатель может изучить различные проблемы экономики здравоохранения. В примерах также часто предлагаются пути дальнейшего исследования, иногда вновь на основе опубликованных работ.

1.6.2. Компьютерное обеспечение и воспроизводимость результатов

Как уже отмечалось в предисловии, для проведения описываемых в этой книге вычислений можно использовать целый ряд современных мощных компьютерных программ. В большинстве случаев приводимые примеры можно воспроизвести в любом современном пакете, как в высокоуровневых интегрированных средах (как, например, *NLOGIT*, *Stata* или *SAS*), так и путем написания собственных программ на таких языках, как *R*, *MatLab* или *Gauss*. Единственным исключением здесь являются упражнения, основанные на симуляции. Поскольку каждый пакет использует собственный генератор случайных чисел, результаты нельзя воспроизвести в точности (если, конечно, вы не пользуетесь той же программой, что и мы). В таких случаях, впрочем, разница в получаемых результатах будет равна небольшому

случайному колебанию. Основные результаты и свойства, рассматриваемые в этих приложениях, должны быть воспроизводимы в любом из упомянутых программных пакетов. Мы еще вернемся к вопросу воспроизводимости результатов по ходу книги, например в разделе 15.2, где будем говорить о методах генерирования случайных чисел для оценок, основанных на симуляциях.

1.6.3. Соглашения об обозначениях

Мы будем пользоваться векторной и матричной записью и преобразованиями на протяжении всей книги. Будут использоваться следующие формы записи: скалярные переменные будут обозначаться строчными наклонными буквами, как, например, y или x_{nK} ; векторы-столбцы скалярных значе-

ний — жирными строчными буквами, например, $\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$, \mathbf{x} или \mathbf{b} . Раз-

меры вектора-столбца всегда записываются в той же форме, что и размеры матрицы с одним столбцом, т.е., например, $K \times 1$ или $n \times 1$ и т.п. Матрицы всегда будут обозначаться жирными заглавными буквами, как, например,

матрица размером $n \times K$, $\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1K} \\ x_{21} & x_{22} & \cdots & x_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nK} \end{bmatrix}$.

Отдельные элементы матрицы всегда будут обозначаться нижними индексами, причем первый будет соответствовать номеру строки, а второй — столбца. Транспонирование вектора или матрицы будет обозначаться штрихом. Вектор-строка получается транспонированием вектора-столбца. Так, $\beta' = [\beta_1, \beta_2, \dots, \beta_K]$. Произведение строки и столбца всегда будет записываться в виде $\beta' \mathbf{x} = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K$. Элементы матрицы \mathbf{X} образуют набор векторов. Разлагая на столбцы, получаем $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K]$ — каждый столбец есть вектор размера $n \times 1$. Единственным, к сожалению, неизбежным источником путаницы здесь может быть обозначение, используемое для строки матрицы \mathbf{X} . Элементы i -й строки \mathbf{X} образуют вектор-строку $\mathbf{x}'_i = [x_{i1}, x_{i2}, \dots, x_{iK}]$. Если матрица \mathbf{X} обозначает матрицу данных, мы будем индексом i обозначать наблюдения (строки), а k — переменные (столбцы). Как мы уже отметили, к сожалению, при этом может казаться, что \mathbf{x}_i , полученный транспонированием \mathbf{x}'_i , будет i -м столбцом \mathbf{X} , что является конфликтом наших правил записи. Однако нам ничего другого не остается, и мы будем следовать этим правилам, всегда обозначая за \mathbf{x}'_i вектор-строку, полученную из i -й строки матрицы \mathbf{X} . Обсуждение использующихся в этой книге результатов из матричной алгебры приведено в приложении А. Особенную важность представляет результат раздела А.2.7 о суммировании и об элементах матрицы произведения матриц $\mathbf{X}'\mathbf{X}$.

ГЛАВА 2

Модель линейной регрессии

2.1. Введение

Эконометрика занимается *построением моделей*. В начале нашего обсуждения рассмотрим такой интересный вопрос: а что такое модель? Описание модели обычно начинают с наблюдения или предположения о том, что одна переменная вызывается другой или ее изменение связано с изменениями другой или с какого-либо другого качественного высказывания о наличии связей между переменной и одной или несколькими **ковариатами**, которые предположительно связаны с интересующей нас переменной. Модель может содержать общее высказывание поведенческого плана, например, что пользование услугами здравоохранения зависит среди прочего от оценки потребителями собственного здоровья, демографических показателей, например уровня дохода, возраста и уровня образования, а также от наличия у них медицинской страховки и ее вида. Это высказывание может быть представлено как вербально, так и в графическом виде, например в виде блок-схемы или **графа**, описывающего направления ожидаемых причинно-следственных связей. Эконометрическая модель редко возникает сразу в виде набора уравнений. Обычно она начинается с *идеи* о наличии какой-то зависимости. Следующим естественным шагом для эконометриста является перевод этой идеи в набор уравнений, причем надо иметь в виду, что какую-то часть этих уравнений можно было бы далее использовать для получения ответов на вопросы о поведении интересующей исследователя переменной. В нашем примере более четким вопросом, ответ на который можно пытаться получить из формальной модели, мог бы быть, например, вопрос о зависимости между использованием услуг системы здравоохранения и медицинской страховкой. Точнее, является ли эта зависимость «положительной», т.е. склонен ли при прочих равных более полно застрахованный потребитель «потреблять больше услуг системы здравоохранения», или «отрицательной»? Или же можно задать более точным вопросом: насколько больше (или меньше)? Эта глава и несколько последующих описывают инструментарий, которым эконометристы пользуются при построении моделей для поиска ответов на подобные вопросы с использованием данных и эконометрических методов.

С точки зрения чисто статистической исследователь может думать о переменной y (условно «уровень спроса на услуги здравоохранения H ») и векторе ковариат x (доход I , страховка T) как о переменных, имеющих совместное распределение $p(H, I, T)$. Будучи записанной в такой форме, зависимость не представляется особенно интересной: какой статистический процесс порождает спрос на услуги здравоохранения, уровень дохода и тип страхования? Однако нужно учесть, что $p(H, I, T) = p(H|I, T)p(I, T)$ и,

таким образом, модель раскладывается на два процесса, один из которых определяет совместное распределение дохода и уровня страхования в генеральной совокупности, а второй — распределение спроса на услуги здравоохранения при конкретных уровне дохода и типе страхования. При этой форме записи особенный интерес представляет условное распределение $p(H|I, T)$, в то время как $p(I, T)$, распределение уровня дохода и типа страхования, вторично или даже вообще малоинтересно. (С другой стороны, при этом же подходе можно также изучать условный спрос на страхование при данном уровне дохода $p(T|I)$.) Продолжая размышлять в том же ключе, можно понять, что исследователя обычно интересует не совместное распределение всех переменных модели, а **условные распределения** одной переменной относительно других.

Понятие условного распределения дает нам отправную точку для размышлений о связи между целевой переменной y и набором переменных x , которые, как мы подозреваем, могут быть связаны с целевой переменной. Здесь возникает вопрос, возвращающий нас к дискуссии о том, что есть модель. Какие именно аспекты условного распределения могут представлять интерес? Строящий модель исследователь, думая о свойствах условного распределения, часто склонен уделять внимание прежде всего математическому ожиданию, думая в первую очередь о значении $E[y|x]$ или о **функции регрессии**, что приводит нас к основной теме этой главы. В примере, приведенном выше, было бы естественно взять в качестве y число посещений врача, как это будет сделано в примере, который будет часто использоваться в дальнейшем. Если бы мы изучали уровень дохода I , который часто имеет сильно скошенное распределение, то среднее может не представлять большого интереса. Скорее всего, интересной статистикой была бы **условная медиана** для людей заданного возраста, $M[I|x]$. С другой стороны, говоря об уровнях дохода для различных возрастов, еще больший интерес могли бы представлять другие квантили, скажем, 20%-й уровень или черта бедности, определенная как, например, 5%-й квантиль. Наконец, для исследования в области финансов, где целевой переменной обычно является доходность актива, средние представляют малый интерес. По крайней мере, при некоторых подобных исследованиях средние вообще не представляют интереса; наиболее интересным объектом являются дисперсии, и в особенности условные дисперсии.

Начиная говорить о модели линейной регрессии, важно понимать, что такое модель. Пока мы будем заниматься в основном условными средними, поскольку зачастую исследователя интересует именно этот аспект. Поняв, как можно анализировать функции регрессии, сможем пользоваться этим аппаратом для изучения других свойств распределений, например квантилей и дисперсий. **Модель линейной регрессии** является самым полезным из инструментов, доступных эконометристу. Несмотря на то что в современных исследованиях линейные регрессии все чаще являются лишь отправной точкой для полного анализа, именно с линейных регрессий начинаются почти все эмпирические исследования. Исследователи обычно рассматривают связи между переменными именно через призму линейных регрессий. В этой главе мы обсудим модель линейной регрессии и детально разберем основные предположения этой модели. В нескольких последующих главах

мы обсудим более сложные спецификации, а также сложности, которые могут возникать при применении описываемых здесь техник, основанных на разбираемых простых моделях.

2.2. Модель линейной регрессии

Модель множественной линейной регрессии используется при описании связи между **зависимой переменной** и одной или более **независимыми переменными**. В общем виде модель записывается следующим образом:

$$\begin{aligned} y &= f(x_1, x_2, \dots, x_K) + \varepsilon = \\ &= x_1\beta_1 + x_2\beta_2 + \dots + x_K\beta_K + \varepsilon, \end{aligned} \quad (2-1)$$

где y — зависимая, или **объясняемая**, переменная, а x_1, \dots, x_K — независимые, или **объясняющие**, переменные. Вид $f(x_1, x_2, \dots, x_K)$ определяется из теоретических предпосылок. Эта функция обычно называется (**теоретическим**) **уравнением регрессии** y на x_1, \dots, x_K , что подразумевает, что она описывает зависимость y от x_1, \dots, x_K в генеральной совокупности. В этой формулировке y — **регрессанд**, а $x_k, k=1, \dots, K$ — **регрессоры**, или ковариаты. Лежащая в основе модели теория должна определить зависимую и независимые переменные модели. Не всегда ясно, как именно их нужно определять: например, уравнение спроса $quantity = \beta_1 + price \times \beta_2 + income \times \beta_3 + \varepsilon$ и обратное уравнение спроса $price = \gamma_1 + quantity \times \gamma_2 + income \times \gamma_3 + u$ являются верными описаниями рынка. При моделировании часто полезно думать в терминах «автономного изменения». Можно представить, что независимые переменные будут меняться вне ограничений, накладываемых моделью, но изменения зависимой переменной должны быть вызваны либо изменениями независимых переменных, либо экзогенными факторами¹.

Член ε представляет собой **случайное возмущение**, названное так, поскольку оно вносит помехи в устойчивую зависимость между переменными. Возмущение может возникать по разным причинам, в первую очередь из-за того, что никакая сколь угодно сложная модель не может включить все реально существующие факторы, влияющие на зависимую переменную. Положительный или отрицательный внутренний эффект этих факторов входит в случайное возмущение. В эмпирических исследованиях на возмущение могут влиять и другие факторы. Самый заметный из них — это возможная ошибка измерения. Достаточно просто рассуждать о предполагаемых связях между точными значениями переменных, но совершенно иное дело получить точные значения этих переменных. Например, сложность получения разумных показателей прибылей, процентных ставок, капитала или, еще сложнее, амортизации капитала является часто обсуждаемой темой в эмпирической литературе. В некоторых случаях у теоретической переменной может вообще не быть наблюдаемого аналога. Интересным примером

¹Применяя этот подход к уравнению спроса, можно подумать, что единственной независимой переменной является уровень дохода, а цена и объем продаж — зависимые переменные. Такой подход имеет смысл — на рынке цена и объем продаж действительно определяются одновременно и изменяются только из-за экзогенных факторов.

здесь может служить литература, посвященная модели потребления при постоянном уровне дохода [см., например, Friedman (1957)].

Мы предполагаем, что каждое наблюдение в выборке $(y_i, x_{i1}, x_{i2}, \dots, x_{iK})$, $i = 1, \dots, n$ порождено случайным процессом, описываемым уравнением

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \dots + x_{iK}\beta_K + \varepsilon_i.$$

Наблюдаемое значение y_i складывается из детерминированной части и случайного члена ε_i . Наша цель — оценить величины неизвестных параметров модели, исследовать соответствие теоретической модели и имеющихся данных, исследовать на этих данных, выполняются ли предположения теории, и, возможно, предсказать с помощью модели значения переменной y . Наши дальнейшие действия существенно зависят от предположений относительно случайного процесса, породившего имеющиеся у нас наблюдения.

Пример 2.1. Кейнсианская функция потребления

В примере 1.2 рассматривалась модель потребления, предложенная Кейнсом в его *General Theory* (1936). Теория о том, что уровни потребления C и дохода X связаны, представляется вполне соответствующей наблюдаемым фактам, представленным на рис. 1.1 и 2.1. (Данные содержатся в табл. F2.1.)

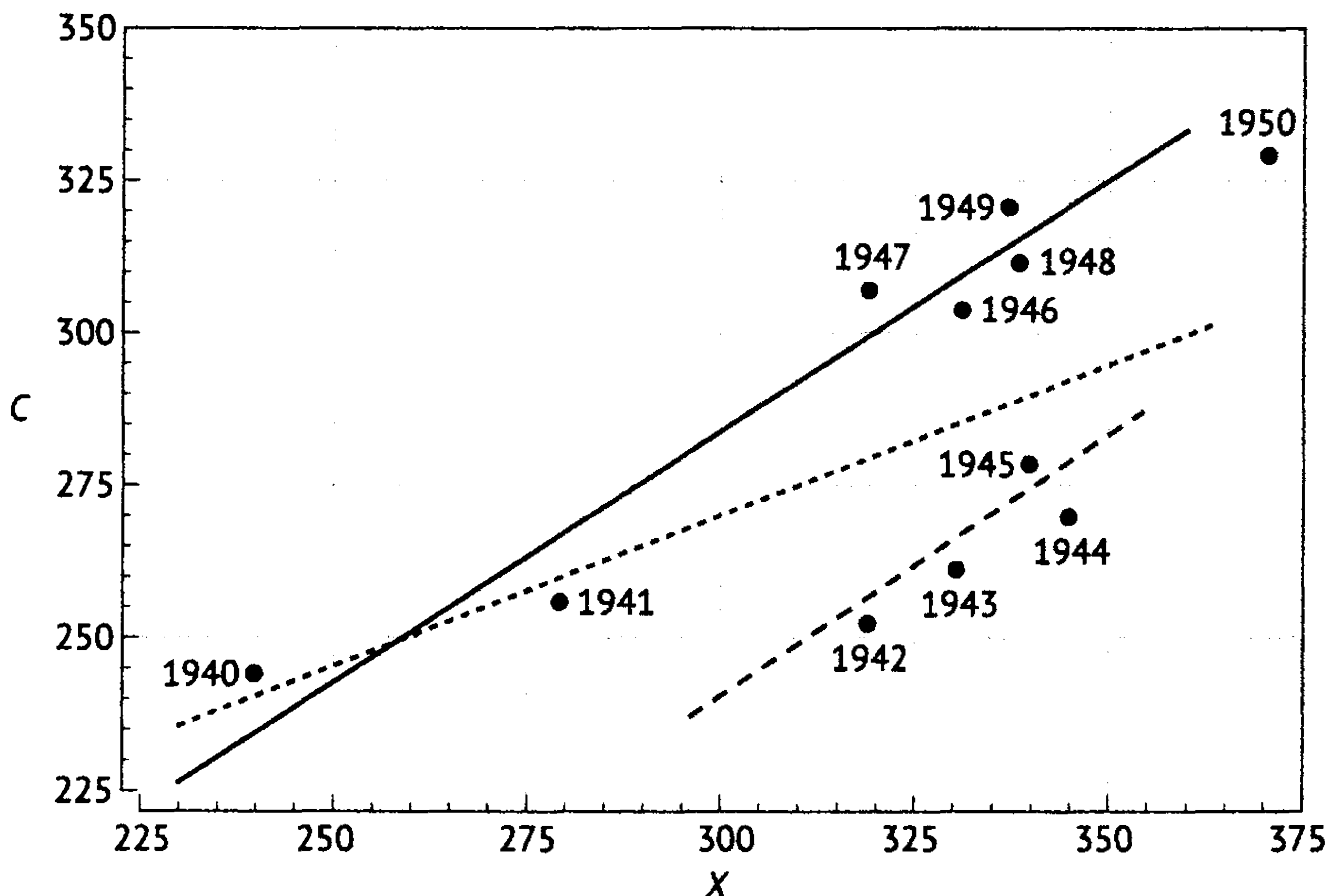


Рис. 2.1. Потребление в США в 1940–1950 гг.

Конечно, линейная функция является всего лишь приближением. Даже если пропустить аномальные наблюдения военных лет, потребление и доход нельзя связать каким-либо простым **детерминированным** образом. Линейная модель $C = \alpha + \beta X$ предназначена в первую очередь для выделения некоторых важных свойств этой части экономики. Попытка описать все факторы, влияющие на эту связь, была бы обречена на провал. Следующий шаг — включить в модель случайность, содержащуюся

в реальных переменных. Поэтому запишем $C = f(X, \varepsilon)$, где ε есть случайный элемент. Здесь важно избежать соблазна воспринимать ε как универсальную «ловушку» для всех недостатков модели. Кажется, что модель с ε адекватно описывает данные без наблюдений военных лет, но для объяснения наблюдений 1942–1945 гг. явно не хватает чего-то систематического. Потребление в эти годы не могло подняться до уровня, исторически соответствующего уровням дохода, из-за ограничений военного времени. Модель, претендующая на объяснение уровня потребления в этот период, должна включать влияние этих факторов.

Остается понять, каким образом случайный член должен быть включен в уравнение. Наиболее частый подход состоит в том, чтобы считать его *аддитивным*. Это значит, что уравнение нужно переписать в стохастических терминах в виде $C = \alpha + \beta X + \varepsilon$. Это уравнение представляет собой эмпирический аналог теоретической модели Кейнса. Но как быть с «аномальным» периодом введения ограничений? Если мы проигнорируем нашу интуицию и попытаемся построить линейное приближение ко всей выборке (в следующей главе подробно описывается, как это сделать), то получим пунктирную линию рисунка. Однако эта линия очевидно смещена вследствие ограничений военного времени. Более подходящей для этих данных спецификацией, включающей как случайную составляющую, так и особые условия 1942–1945 гг., была бы линия, сдвинутая вниз в этот период, т.е. $C = \alpha + \beta X + d_{waryears} \delta_w + \varepsilon$, где новая переменная $d_{waryears}$ равна единице в 1942–1945 гг. и нулю для остальных наблюдений, а $\delta_w < 0$.

Одной из наиболее полезных черт модели множественной регрессии является возможность выделить независимые влияния разных независимых переменных на зависимую переменную. В примере 2.2 описывается одна часто встречающаяся модель.

Пример 2.2. Зарплата и образование

Ряд исследований последних лет был посвящен вопросу о связи между уровнем образования и зарплатой. Можно было бы ожидать, что более высокий уровень образования (*education*) приводит в среднем к более высокому заработку (*earnings*). Простая модель регрессии

$$earnings = \beta_1 + \beta_2 education + \varepsilon,$$

однако, не учитывает того факта, что у большинства людей доход в зрелом возрасте превышает доход в молодости независимо от уровня их образования. Значит, β_2 будет преувеличивать предельный эффект повышения уровня образования. Если возраст и образование имеют положительную корреляцию, то модель будет считать повышение дохода результатом исключительно повышения уровня образования. Спецификацию можно улучшить включением влияния возраста (*age*):

$$earnings = \beta_1 + \beta_2 education + \beta_3 age + \varepsilon.$$

Часто отмечают, что доход, как правило, растет медленнее в более поздние годы трудовой жизни, чем в начале карьеры. Чтобы учесть и это обстоятельство, модель можно дополнить следующим образом:

$$\text{earnings} = \beta_1 + \beta_2 \text{education} + \beta_3 \text{age} + \beta_4 \text{age}^2 + \varepsilon.$$

Мы ожидаем, что β_3 будет положительным, а β_4 — отрицательным. Важным свойством этой модели является то, что она позволяет нам провести мысленный эксперимент, который, возможно, был бы неосуществим на реальных данных. В данном примере мы могли бы сравнить заработок двух людей одинакового возраста, имеющих разный уровень образования, хотя выборка, возможно, и не содержит ни одной такой пары. Как нужно измерять уровень образования в такой модели — вопрос сложный. В исследовании Ashenfelter and Krueger (1994), сравнивающем заработки близнецов и применяющем как раз такую спецификацию модели, используется интересный подход. [Вообще исследования, использующие данные о близнецах или братьях и сестрах, часто встречаются при изучении связи уровня образования и доходов. Можно привести еще два таких исследования: Ashenfelter and Zimmerman (1997) и Bonjour, Cherkas, Haskel, Hawkes and Spector (2003).] Эта работа будет подробно рассматриваться в подразделе 8.5.3.

Эксперимент, заложенный в построенной нами модели заработной платы, состоит в сравнении заработков двух одинаковых людей, отличающихся только уровнем образования. В этой интерпретации эффект образования будет равен $\partial E[\text{Earnings} | \text{Age}, \text{Education}] / \partial \text{Education} = \beta_2$. Можно, впрочем, возразить, что по-настоящему интересным было бы измерение ненаблюдаемого влияния повышения уровня образования на заработок конкретного индивида. Чтобы провести такой эксперимент в реальности, нужно было бы измерить его заработок дважды: один раз в реальной ситуации, Education_i , а второй — в гипотетической (hypothetical, counterfactual), где уровень его образования равен $\text{Education}_i + 1$. Если мы будем рассматривать Education в этом примере как **воздействие** (treatment), то целью эксперимента будет изучение **влияния воздействия на подвергшегося воздействию**. Попытка сделать подобный вывод из реальных данных, сравнивая двух «идентичных в других отношениях» индивидов, будет предпринята нами в главе 19.

Большой объем литературы посвящен следующему интересному вопросу: уровень образования в этой модели нельзя считать полностью независимым. Высокомотивированные индивиды, скорее всего, будут получать более высокий уровень образования (например, поступать в колледж или магистратуру). Но исходя из тех же качеств они также будут в среднем более склонны выбирать занятия, приводящие к более высокому уровню дохода. Если это так, то неясно, свидетельствует ли положительное β_2 о том, что доход связан с уровнем образования, или же отражает некоторую общую причину, влияющую на обе переменные модели, но которую мы не включили в наше уравнение регрессии? Мы вернемся к этому вопросу в главе 19².

2.3. Предположения модели линейной регрессии

Модель линейной регрессии состоит из набора предположений о том, как выборка возникает из «порождающего ее процесса». Теория предполагает детерминированную связь между зависимой и независимыми переменными. Предположения, описывающие вид модели и связи между ее частями и приводящие к используемым процедурам оценивания и статистических выводов, приведены в табл. 2.1.

Таблица 2.1. Предположения модели линейной регрессии

-
- A1. Линейность:** $y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \dots + x_{iK}\beta_K + \varepsilon_i$. Модель определяет линейную зависимость между y и x_1, \dots, x_K .
- A2. Полный ранг:** не существует точной линейной зависимости между независимыми переменными модели. Это предположение будет необходимо для построения оценок параметров модели.
- A3. Экзогенность независимых переменных:** $E[\varepsilon_i | x_{j1}, x_{j2}, \dots, x_{jK}] = 0$. Утверждается, что математическое ожидание случайного возмущения наблюдения i в выборке не является функцией значений независимых переменных как этого, так и других наблюдений. Это означает, что независимые переменные не содержат информации, которая могла бы помочь предсказать значения ε_i .
- A4. Гомоскедастичность и отсутствие автокорреляции:** каждый шок ε_i имеет одну и ту же конечную дисперсию σ^2 и не коррелирует ни с одним из других случайных возмущений ε_j . Это предположение ограничивает общность модели, и в дальнейших главах мы будем пытаться его ослабить.
- A5. Порождение данных:** данные в $(x_{j1}, x_{j2}, \dots, x_{jK})$ представляют собой комбинацию случайных и постоянных величин. Для наших целей наибольшую важность представляют предположение A3 о независимости математического ожидания возмущений и предположение A4 о независимости дисперсии возмущений. Весь анализ будет проводиться в терминах условных распределений при данной выборке \mathbf{X} , так что то, являются ли элементы \mathbf{X} постоянными или случайными величинами, получаемыми из какого-то случайного процесса, не может повлиять на конечный результат. Позже, в более сложных постановках, нам будет необходимо более точно описывать возможные связи между ε_i и x_j .
- A6. Нормальность:** шоки распределены нормально. Это предположение также является временным упрощением, которое мы будем пытаться ослабить, предварительно изучив его следствия.

2.3.1. Линейность модели регрессии

Пусть столбец x_k содержит n наблюдений случайной величины x_k , $k = 1, \dots, K$. Соберем из этих столбцов матрицу \mathbf{X} размером $n \times K$. Обычно предполагают, что первый столбец \mathbf{X} состоит из единиц и, таким образом,

²В этой модели есть еще одна ловушка с практической точки зрения. Межобъектные данные о более высоких доходах людей зрелого возраста могут быть сформированы процессом, носящим, скажем, макроэкономический характер (эффект когорты), и с этой точки зрения внутренне отличаться от данных более молодых и низкооплачиваемых индивидов. Вообще говоря, нельзя делать выводы о доходах молодых людей в выборке *если бы они были старше* путем сравнения их со старшим поколением в выборке. Похожая проблема возникает при изучении эффектов воздействия, к которым мы обратимся в главе 19.

β_1 является свободным членом в уравнении регрессии. Пусть y — n наблюдений y_1, \dots, y_n , а ϵ — столбец из n значений случайных возмущений (шоков). Модель (2-1) применительно к n наблюдениям можно записать в виде

$$y = x_1\beta_1 + \dots + x_K\beta_K + \epsilon \quad (2-2)$$

или в виде предположения 1:

$$\boxed{y = X\beta + \epsilon.} \quad (2-3)$$

Замечание об обозначениях

Здесь и в дальнейшем, чтобы избежать путаницы и излишнего усложнения формул, мы будем жирным обозначать столбец или строку X . Что именно — будет ясно из контекста. В (2-2) x_k обозначает k -й столбец X . Нижние индексы j и k будут использоваться для нумерации столбцов (переменных). Часто будет удобно говорить об отдельных наблюдениях из (2-3), и это мы будем записывать так:

$$y_i = x_i'\beta + \epsilon_i. \quad (2-4)$$

Нижние индексы i и t обычно будут обозначать строки (наблюдения) матрицы X . В (2-4) x_i — столбец, полученный транспонированием i -й строки матрицы X .

Нас будет интересовать в основном построение оценок и формулировка статистических выводов о векторе параметров модели β . Обратим внимание, что модель простой регрессии в примере 2.1 является частным случаем модели множественной регрессии с матрицей X , состоящей из двух столбцов, первый из которых состоит из единиц. Предположение о линейности модели регрессии предполагает наличие аддитивного шока. Чтобы регрессия была линейной в описываемом здесь смысле, она должна иметь вид (2-1) либо в исходном виде, либо после подходящих преобразований. Например, модель

$$y = Ax^\beta e^\epsilon$$

является линейной (нужно взять логарифм от обеих частей), в то время как модель

$$y = Ax^\beta + \epsilon$$

линейной не является. Наблюдаемые значения зависимой переменной, таким образом, складываются из двух частей: детерминированной $\alpha + \beta x$ и случайной переменной ϵ . Здесь важно подчеркнуть, что ни одно из этих слагаемых не поддается прямому наблюдению, поскольку α и β неизвестны.

Предположение линейности не является таким сильным ограничением, как кажется на первый взгляд. В контексте регрессий понятие *линейности* относится к тому, как в уравнение регрессии входят параметры и возмущения, что не всегда означает, что связь между переменными обязательно линейна. Например, уравнения $y = \alpha + \beta x + \epsilon$, $y = \alpha + \beta \cos(x) + \epsilon$, $y = \alpha + \beta/x +$

$+\varepsilon$ и $y = \alpha + \beta \ln x + \varepsilon$ линейны по некоторой функции x по использованному здесь определению. В этих примерах преобразования применялись только к x , но могли бы применяться и к y , как, например, в уравнении $y = Ax^\beta e^\varepsilon$, задающем линейную связь между логарифмами x и y , $\ln y = \alpha + \beta \ln x + \varepsilon$. Выбор функций ничем не ограничен. Эта черта модели используется во многих функциональных формах уравнения регрессии. Например, **логлинейная (или линейная в логарифмах) модель** записывается как

$$\ln y = \beta_1 + \beta_2 \ln x_2 + \beta_3 \ln x_3 + \dots + \beta_K \ln x_K + \varepsilon.$$

Эта модель известна также как уравнение **постоянной эластичности**, так как в нем эластичность y по x равна $\partial \ln y / \partial \ln x_k = \beta_k$ и не зависит от x_k . Логлинейная форма часто используется в моделях спроса и производства. Разные значения β дают существенно различные функции.

Пример 2.3. Рынок бензина в США

Данные о рынке бензина в США в 1953–2004 гг. приведены в табл. F2.2 приложения F. Мы используем эти данные, чтобы получить среди прочего оценки эластичностей спроса на этом рынке по доходу потребителей, цене и цене других товаров. Используя эти данные, можно поднять следующий интересный вопрос: что произойдет, если изменять значения некоторых переменных, оставляя значения остальных постоянными, как предлагалось в примере 2.2? Например, рассмотрим следующую урезанную модель потребления бензина на душу населения:

$$\ln(G/pop) = \beta_1 + \beta_2 \ln(Income/pop) + \beta_3 \ln price_G + \beta_4 \ln P_{newcars} + \beta_5 \ln P_{usedcars} + \varepsilon.$$

Эта модель позволяет получить оценки эластичностей спроса на бензин по доходу потребителей и по цене бензина, а также оценку эластичности спроса по ценам новых и подержанных автомобилей. Какой мы могли бы ожидать знак у β_4 ? Автомобили и бензин являются комплементарными товарами, так что при росте цен на новые автомобили при прочих равных условиях нужно было бы ожидать падения потребления бензина. Нужно ли? Если цены на новые автомобили растут, то потребители будут менее охотно их покупать; значит, старые автомобили будут использоваться дольше. Если старые автомобили потребляют больше бензина, то рост цен на автомобили приведет к росту, а не к падению спроса на бензин. Для ответа на этот вопрос нужно воспользоваться моделью множественной регрессии и данными по рынку бензина.

Полулогарифмическими моделями часто пользуются при моделировании темпов роста:

$$\ln y_t = \mathbf{x}'_t \boldsymbol{\beta} + \delta t + \varepsilon_t.$$

В этой модели самостоятельный (по крайней мере, необъясняемый этой моделью) мультипликативный темп роста в каждом периоде равен $\partial \ln y / \partial t =$

δ. Другие варианты конкретизации общего вида модели

$$f(y_t) = g(\mathbf{x}'_t \beta + \varepsilon_t)$$

дадут большой набор возможных функциональных форм, каждая из которых соответствует нашему определению линейности модели.

Модель линейной регрессии часто интерпретируют как приближение к какой-то настоящей неизвестной функции. (См. обсуждение в разделе А.8.1.) При такой интерпретации, однако, линейная модель даже с включением квадратичных членов представляется довольно ограниченной, поскольку такое приближение, скорее всего, будет полезно лишь в небольшом интервале значений независимых переменных. Транслогарифмическая модель, обсуждаемая в примере 2.4, показывает гораздо большую эффективность в качестве аппроксимирующей функции.

Пример 2.4. Транслогарифмическая модель

Авторы современных работ, изучающие спрос и производство, обычно пользуются **гибкими функциональными формами** уравнений регрессии. Гибкие функциональные формы используются в эконометрических исследованиях благодаря тому, что они позволяют исследователям моделировать нетривиальные особенности функции производства, такие как, например, эластичности замещения, которые являются функциями от вторых производных объема производства, стоимости или функции полезности. В линейной модели эти производные гарантированно равны нулю, а логлинейная модель (например, модель Кобба–Дугласа) ограничивает значения этих эластичностей плюс или минус единицей. Наиболее популярной гибкой функциональной формой является **транслогарифмическая модель**, которую можно воспринимать как приближение второго порядка к неизвестной функциональной форме. [Berndt and Christensen (1973).] К этой модели можно прийти, например, следующим образом. Сначала запишем $y = g(x_1, \dots, x_K)$. Тогда $\ln y = \ln g(\dots) = f(\dots)$. Поскольку мы можем рассмотреть тривиальное преобразование $x_k = \exp(\ln x_k)$, запишем эту функцию как функцию от логарифмов x -в. Итак, $\ln y = f(\ln x_1, \dots, \ln x_K)$.

Разложим эту функцию в ряд Тейлора (до члена второго порядка) в точке $\mathbf{x} = [1, 1, \dots, 1]'$ (в этой точке логарифмы всех переменных равны 0). Тогда

$$\begin{aligned} \ln y = & f(\mathbf{0}) + \sum_{k=1}^K [\partial f(\cdot) / \partial \ln x_k]_{\ln \mathbf{x}=\mathbf{0}} \ln x_k + \\ & + \frac{1}{2} \sum_{k=1}^K \sum_{l=1}^K [\partial^2 f(\cdot) / \partial \ln x_k \partial \ln x_l]_{\ln \mathbf{x}=\mathbf{0}} \ln x_k \ln x_l + \varepsilon. \end{aligned}$$

Эта модель предполагает, что шоки представляют собой как обычные факторы, так и ошибки приближения к неизвестной функции. Поскольку

функция и ее производные в точке 0 есть некоторые постоянные величины, мы используем их в качестве коэффициентов и запишем

$$\ln y = \beta_0 + \sum_{k=1}^K \beta_k \ln x_k + \frac{1}{2} \sum_{k=1}^K \sum_{l=1}^K \gamma_{kl} \ln x_k \ln x_l + \varepsilon.$$

Эта модель линейна согласно нашему определению, но может на самом деле имитировать значительную кривизну приближаемой функции. Интересной особенностью этой формулировки является то, что логлинейная модель оказывается ее частным случаем при $\gamma_{kl} = 0$. Еще эта модель дает интересную возможность проверить используемую теорию: если приближаемая функция предполагается непрерывной и дважды непрерывно дифференцируемой, то по теореме Юнга должно выполняться $\gamma_{kl} = \gamma_{lk}$. В главе 10 мы увидим, как такие проверки делаются на практике.

Несмотря на значительную гибкость, линейная модель подходит не для всех ситуаций, с которыми мы можем столкнуться на практике. В примере 14.10 и главе 18 мы рассмотрим регрессионную модель для числа посещений врача, предложенную во введении к этой главе. Подходящей для объяснения числа посещений врача является модель условного среднего $E\{y|\mathbf{x}\} = \exp(\mathbf{x}'\beta)$. Есть соблазн линеаризовать эту функцию напрямую с помощью логарифмов, поскольку $\ln E\{y|\mathbf{x}\} = \mathbf{x}'\beta$. Однако $\ln E\{y|\mathbf{x}\}$ не равно $E\{\ln y|\mathbf{x}\}$. В этой ситуации y может равняться нулю (и для большей части выборки равняется), так что $\mathbf{x}'\beta$ (который может быть отрицательным) не является удачной моделью ни для $\ln y$ (которого не существует), ни для неотрицательной величины y . Методы, описываемые в этой главе, не подходят для оценки параметров такой модели. Тем не менее для подобных нелинейных моделей существуют сравнительно несложные методы оценки. Мы будем подробно рассматривать их в главе 7.

2.3.2. Полный ранг

Предположение 2 говорит, что между переменными нет линейных зависимостей.

$$\mathbf{X} \text{ является матрицей размером } n \times K \text{ и имеет ранг } K. \quad (2-5)$$

Значит, \mathbf{X} имеет полный ранг по столбцам. Столбцы \mathbf{X} линейно независимы, и существует по крайней мере K наблюдений. [См. (A-42) и текст рядом.] Это предположение известно как **условие идентификации**. Чтобы понять необходимость этого условия, рассмотрим пример.

Пример 2.5. Короткий ранг

Предположим, что модель, оцениваемая на межобъектных данных, указывает, что потребление C связано с доходом следующим образом:

$$C = \beta_1 + \beta_2 \text{ nonlabor income} + \beta_3 \text{ salary} + \beta_4 \text{ total income} + \varepsilon,$$

где полный доход (total income) равен в точности сумме зарплаты (salary) и нетрудового дохода (nonlabor income). Видно, что в этой модели есть линейная зависимость между независимыми переменными. Предположим теперь, что

$$\begin{aligned}\beta'_2 &= \beta_2 + a, \\ \beta'_3 &= \beta_3 + a, \\ \beta'_4 &= \beta_4 - a,\end{aligned}$$

где a — любое число. Тогда одинаковые значения правой части получаются при значениях параметров β'_2, β'_3 и β'_4 и при значениях β_2, β_3 и β_4 . Очевидно, построить оценку параметров этой модели невозможно.

Если число наблюдений меньше K , то X не может иметь **полного ранга**. Таким образом, мы можем сделать (избыточное) предположение, что n , по крайней мере, не меньше K .

В модели с одной переменной и постоянным членом предположение о полноте ранга означает, что регрессор x должен иметь некоторый разброс. Если x постоянен, то все наши наблюдения будут лежать на одной вертикальной прямой. В этой ситуации остальные предположения не будут нарушены; скорее всего, это дефект выборки. В этой ситуации есть вероятность, что мы в принципе могли бы получить выборку, в которой были бы разные значения x , но так случилось, что не получили ее. Таким образом, модель, вообще говоря, может быть применима, но имеющиеся данные не позволяют нам проверить это.

Пример 2.6. Модель, не поддающаяся оценке

В примере 3.4 мы рассмотрим модель ценообразования на картины Моне. Теоретики и наблюдатели имеют целый набор моделей ценообразования на аукционах картин. Пусть один (наивный) исследователь этого вопроса предлагает следующую модель:

$$\begin{aligned}\ln Price &= \beta_1 + \beta_2 \ln Size + \beta_3 \ln Aspect Ratio + \beta_4 \ln Height + \varepsilon \\ &= \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon,\end{aligned}$$

где $Size = Width \times Height$ и $Aspect Ratio = Width/Height$. Простая арифметика показывает, что эта модель страдает от тех же недостатков, что и модель потребления примера 2.5: здесь $x_2 - x_4 = x_3 + x_4$. Значит, эта модель, как и предыдущая, не поддается оцениванию: она неидентифицируема. Здесь полезно, так сказать, взглянуть на проблему с другой стороны. В линейной модели все переменные должны иметь возможность изменяться линейно независимо друг от друга. Но здесь, несмотря на то что любые две из трех переменных могут принимать произвольные значения, все три переменные связаны линейной зависимостью. Эта модель, или теория, сама по себе вполне разумна. Покупатели картин могут действительно учитывать все три обстоятельства при оценке картины Моне. Но по наблюдаемым данным определить это, по крайней мере в рамках данной линейной модели, нельзя.

2.3.3. Регрессия

Предполагается, что случайный шок в каждом наблюдении имеет условное математическое ожидание, равное нулю, что записывается в виде

$$E[\varepsilon_i | \mathbf{X}] = 0. \quad (2-6)$$

Для всей выборки предположение 3 записывается в виде

$$E[\varepsilon | \mathbf{X}] = \begin{bmatrix} E[\varepsilon_1 | \mathbf{X}] \\ E[\varepsilon_2 | \mathbf{X}] \\ \vdots \\ E[\varepsilon_n | \mathbf{X}] \end{bmatrix} = \mathbf{0}. \quad (2-7)$$

Здесь есть тонкое место, которое наверняка заметил внимательный читатель. В (2-7) в левой части написано, что математическое ожидание каждого ε_i условно по всем наблюдениям \mathbf{x}_i равно нулю. Фактически это предположение об условном математическом ожидании говорит, что никакое наблюдение из \mathbf{x} не дает никакой информации об ожидаемом значении шока. Можно представить (например, при изучении временных рядов), что, несмотря на то что \mathbf{x}_i не дает никакой информации о $E[\varepsilon_i | \cdot]$, \mathbf{x}_j из другого наблюдения, например в следующем периоде, дает эту информацию. Наше предположение на данном этапе состоит в том, что информации о $E[\varepsilon_i | \cdot]$ нет ни в каком \mathbf{x}_j . Позднее, когда мы будем обобщать нашу модель, изучим возможные последствия нарушения этого предположения. [Wooldridge (1995).] Мы также будем предполагать, что значения случайных шоков не содержат информации друг о друге, т.е.

$$E[\varepsilon_i | \varepsilon_1, \dots, \varepsilon_{i-1}, \varepsilon_{i+1}, \dots, \varepsilon_n] = 0.$$

Резюмируя, пока мы предполагаем, что все случайные шоки являются случайной выборкой из одного и того же распределения.

Нулевое условное математическое ожидание влечет равенство безусловного математического ожидания нулю, поскольку

$$E[\varepsilon_i] = E_{\mathbf{x}}[E[\varepsilon_i | \mathbf{X}]] = E_{\mathbf{x}}[0] = 0.$$

Так как для всех ε_i $\text{Cov}[E[\varepsilon_i | \mathbf{X}], \mathbf{X}] = \text{Cov}[\varepsilon_i, \mathbf{X}]$, из предположения 3 следует, что $\text{Cov}[\varepsilon_i, \mathbf{X}] = 0$ для всех i . Обратное неверно: из того, что $E[\varepsilon_i] = 0$, не следует $E[\varepsilon_i | \mathbf{x}_i] = 0$. Это различие иллюстрирует пример 2.7.

Пример 2.7. Ненулевое условное среднее шоков

На рис. 2.2 показано важное различие между $E[\varepsilon_i] = 0$ и $E[\varepsilon_i | x_i] = 0$. Среднее всех возмущений в выборке равно нулю, но среднее на некоторых интервалах значений x явно отличается от нуля. Такая картина в наблюдаемых данных является сигналом о том, что предположение о линейности регрессии нужно подвергнуть сомнению. В данном случае истинной функцией условного среднего (которую исследователь не мог

знать заранее) на самом деле является $E[y|x] = 1 + \exp(1,5x)$. Данные выборки показывают, что линейная модель не подходит для них. Эта возможность будет далее рассматриваться в модели в примере 6.6.

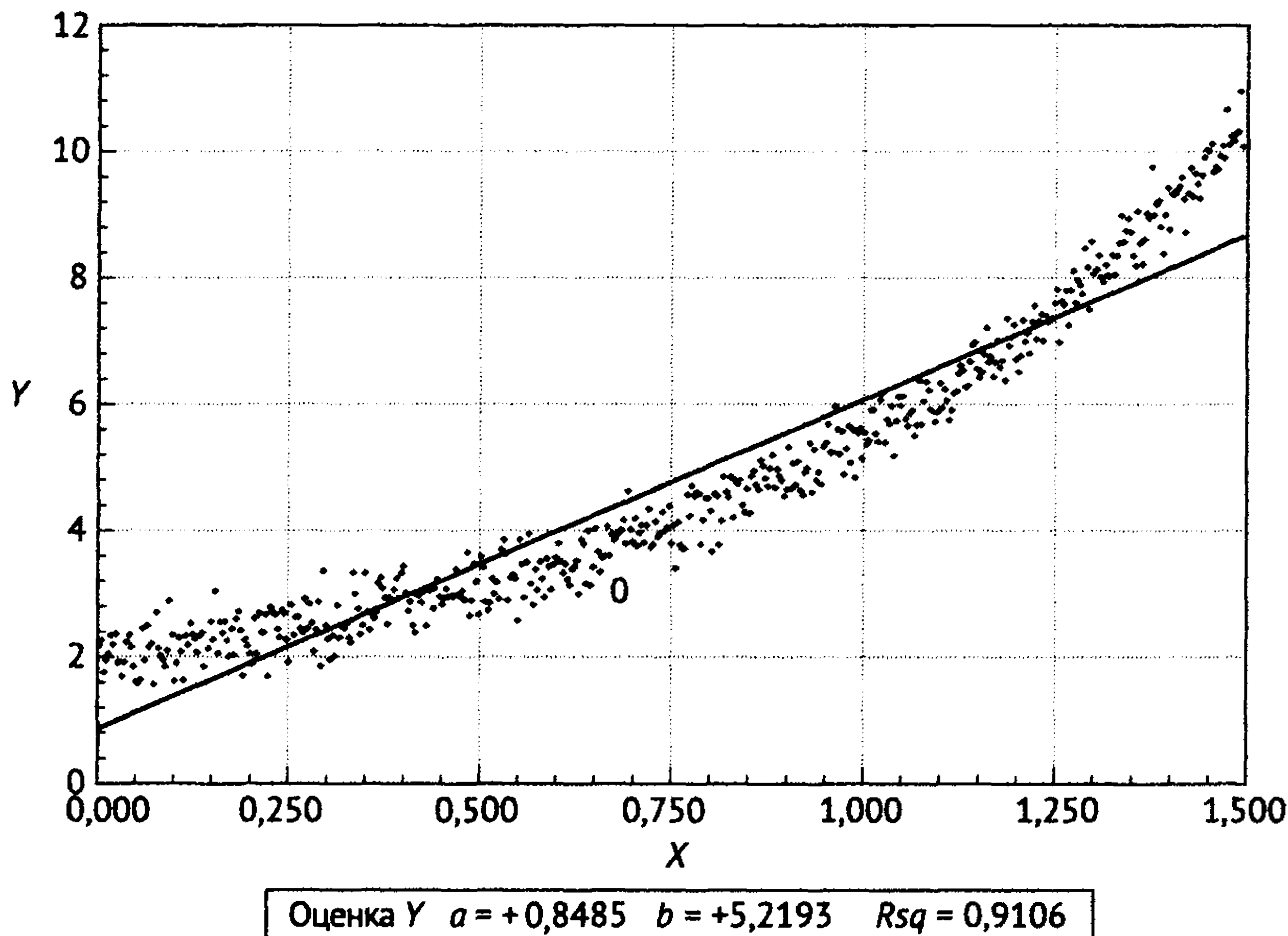


Рис. 2.2. Шоки с ненулевыми условными математическими ожиданиями и нулевым безусловным математическим ожиданием

В большинстве случаев предположение о нулевом безусловном математическом ожидании не является существенным ограничением. Возьмем модель с двумя переменными и предположим, что среднее ε равно $\mu \neq 0$. Тогда $\alpha + \beta x + \varepsilon$ совпадает с $(\alpha + \mu) + \beta x + (\varepsilon - \mu)$. Обозначая $\alpha' = \alpha + \mu$ и $\varepsilon' = \varepsilon - \mu$, мы получаем исходную модель. Подобная модель встретится нам в главе 19, при обсуждении функций производственной границы (или граничной производственной функции). Но если исходная модель не содержит постоянного члена, то предположение $E[\varepsilon_i] = 0$ может быть содержательным. Возникает подозрение, что использование моделей без константы может привести к проблемам. Как правило, модели регрессий должны включать константу, если только отсутствие константы не следует из теории³. Можно утверждать, что если мы по каким-то причинам предполагаем, что математическое ожидание шоков может быть отлично от нуля, то это нужно учитывать в содержательной части регрессии, а в шоках оставлять только неизвестную часть ε :

$$E[y | X] = X\beta. \quad (2-8)$$

³Модели, описывающие первые разности переменных, часто специфицируются без константы. Рассмотрим $y_t - y_{t-1}$. Если в правой части присутствует константа α , то y_t является функцией αt , т.е. имеет временной тренд. Моделям с временным трендом уделяется отдельное место в литературе об анализе временных рядов. Мы вернемся к этой теме в главе 21.

Предположения 1 и 3 вместе определяют *модель линейной регрессии*. **Регрессия** y на \mathbf{X} есть условное математическое ожидание $E[y | \mathbf{X}]$, так что без предположения 3 $\mathbf{X}\beta$ не является функцией условного среднего.

Оставшиеся предположения более полно определяют свойства случайных шоков в модели и формулируют условия получения выборки \mathbf{x} .

2.3.4. Сферические шоки

Четвертое предположение касается дисперсий и ковариаций случайных шоков:

$$\text{Var}[\varepsilon_i | \mathbf{X}] = \sigma^2 \quad \text{для } i = 1, \dots, n$$

и

$$\text{Cov}[\varepsilon_i, \varepsilon_j | \mathbf{X}] = 0 \quad \text{для всех } i \neq j.$$

Ситуация, когда дисперсия шоков постоянна, называется **гомоскедастичностью**. Возьмем модель, связывающую прибыль фирм в некоторой индустрии, например, с их размером. Даже при учете размера измеренные в долларах прибыли больших фирм будут иметь большую дисперсию, чем прибыли малых. Предположение о гомоскедастичности выполняться не будет. Данные о расходах, получаемые из опросов домохозяйств, часто обладают существенной гетероскедастичностью даже при учете уровней дохода и размеров домохозяйств.

Отсутствие корреляций между наблюдениями обычно называют **отсутствием автокорреляции**. При взгляде на рис. 2.1 возникает подозрение, что шоки в разных наблюдениях не полностью независимы. Даже при ограниченном числе наблюдений видно, что в среднем за каждым возмущением следует возмущение того же знака. Такая «инерция» — это как раз то, что называют **автокорреляцией**, и на данном этапе мы будем предполагать ее отсутствие. Методам работы с экономическими данными, для которых характерна автокорреляция, посвящен большой объем литературы, и мы будем рассматривать такие данные в главе 20. Обратите внимание, что отсутствие автокорреляции не означает отсутствия корреляции между y_i и y_j . Предположение состоит в том, что *отклонения* наблюдений от ожидаемых значений не скоррелированы между собой.

Из этих двух предположений следует, что

$$E[\varepsilon\varepsilon' | \mathbf{X}] = \begin{bmatrix} E[\varepsilon_1\varepsilon_1 | \mathbf{X}] & E[\varepsilon_1\varepsilon_2 | \mathbf{X}] & \cdots & E[\varepsilon_1\varepsilon_n | \mathbf{X}] \\ E[\varepsilon_2\varepsilon_1 | \mathbf{X}] & E[\varepsilon_2\varepsilon_2 | \mathbf{X}] & \cdots & E[\varepsilon_2\varepsilon_n | \mathbf{X}] \\ \vdots & \vdots & \vdots & \vdots \\ E[\varepsilon_n\varepsilon_1 | \mathbf{X}] & E[\varepsilon_n\varepsilon_2 | \mathbf{X}] & \cdots & E[\varepsilon_n\varepsilon_n | \mathbf{X}] \end{bmatrix} =$$

$$= \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix},$$

что мы и записываем в предположении 4:

$$E[\epsilon\epsilon' | \mathbf{X}] = \sigma^2 \mathbf{I}. \quad (2-9)$$

Применяя формулу разложения дисперсии (В-69), мы получаем

$$\text{Var}[\epsilon] = E[\text{Var}[\epsilon | \mathbf{X}]] + \text{Var}[E[\epsilon | \mathbf{X}]] = \sigma^2 \mathbf{I}.$$

Снова нужно подчеркнуть, что это предположение касается информации о дисперсиях и ковариациях возмущений, содержащейся в значениях независимых переменных. Пока мы предполагаем, что подобной информации нет. Это предположение также будет устранено впоследствии, когда мы еще немного продвинемся в модели регрессии. Мы также предполагаем, что сами значения шоков не содержат информации о дисперсиях и ковариациях. Пока это не очень важно, но это оказывается существенным в моделях, касающихся временных рядов. Модели вида $\text{Var}[\epsilon_t | \epsilon_{t-1}] = \sigma^2 + \alpha \epsilon_{t-1}^2$, так называемые модели GARCH (см. главу 20), не нарушают нашего предположения об условных дисперсиях, но в них предполагается, что $\text{Var}[\epsilon_t | \epsilon_{t-1}] \neq \text{Var}[\epsilon_t]$.

Случайные шоки (возмущения), удовлетворяющие условиям гомоскедастичности и отсутствия автокорреляции, иногда называют **сферическими шоками (возмущениями)**⁴.

2.3.5. Процесс порождения данных для регрессоров

Часто предполагают, что x_i неслучайны, как бывает в экспериментах, когда исследователь сам выбирает значения регрессоров и наблюдает y_i . Так происходит, например, в сельскохозяйственных экспериментах, где y_i — урожай, а x_i — количество удобрений и воды. Предположение о **неслучайности регрессоров** на данном этапе будет математическим упрощением. С его помощью мы можем применять простые методы математической статистики, рассматривая вектор x_i как известный вектор постоянных в распределении y_i . После такого упрощения предположения 3 и 4 можно сделать безусловными и они будут заключаться просто в том, что функция распределения ϵ_i не включает никаких постоянных из \mathbf{X} .

Исследователи, занимающиеся общественными науками, редко имеют возможность работать с экспериментальными данными, поэтому их модели редко строятся на предположении о неслучайности регрессоров. Ясно, например, что в любой макроэкономической модели нельзя так по-разному относиться к агрегированным данным. Как правило, нужно считать, что значения x_i случайны, равно как и y_i , поэтому при формулировке модели мы должны предполагать, что x_i — случайный вектор. Таким образом, наши предположения должны касаться процесса, порождающего наблюдения x_i . Если рассматривать x_i как случайный вектор, то предположения 1–4 становятся утверждениями о виде совместного распределения y_i и x_i . Наиболее значимым фактором, определяющим свойства наших оценок и статистических тестов, будет природа регрессоров и то, как мы видим процесс

⁴Этот термин также используется для описания многомерного нормального распределения; см. (В-95). Если $\Sigma = \sigma^2 \mathbf{I}$ в формуле многомерной нормальной плотности, то уравнение $f(\mathbf{x}) = c$ задает «шар» с центром μ и радиусом σ в n -мерном пространстве. Слово *сферический* используется независимо от того, предполагается ли нормальность распределения; иногда, впрочем, явно употребляют термин «сферическое нормальное» распределение.

выборки, порождающий эти регрессоры. Основным здесь является предположение 3 (отсутствие корреляции между X и ϵ). Однако такой подход тоже не идеален: X может содержать неслучайные элементы (например, константу, временной тренд или дамми-переменные, обозначающие временные периоды). Таким образом, ситуация не очень ясна. Однако есть простой способ избежать излишних трудностей: мы будем считать, что X может содержать смесь констант и случайных величин, а математические ожидания и дисперсии ϵ_i не зависят ни от каких элементов X .

| | |
|--|--------|
| X может быть как фиксированным, так и случайным. | (2-10) |
|--|--------|

2.3.6. Нормальность

Удобно предположить, что шоки **распределены нормально** и имеют нулевые средние и постоянную дисперсию. Таким образом, предположения 3 и 4 мы дополняем предположением о нормальности распределения.

| | |
|--|--------|
| $\epsilon X \sim N[0, \sigma^2 I]$. | (2-11) |
|--|--------|

В случае если источник возникновения шоков ϵ соответствует нашему описанию выше, будет, хотя бы приблизительно, приложима центральная предельная теорема, а значит, в большинстве ситуаций предположение о нормальности имеет смысл. Полезным следствием предположения 6 является то, что ϵ_i не только не коррелируют между собой, но еще и статистически независимы. [См. третий пункт в разделе В.9 (В-97) и (В-99).] **Предположение о нормальности** часто считают ненужным и даже излишним дополнением к классической модели регрессии. Однако это предположение представляется разумным в большинстве случаев, кроме ситуаций, когда явно предполагается другое распределение, как, например, в модели стохастической границы в главе 19.

Условие нормальности не является необходимым для получения многих результатов, используемых при анализе регрессий, однако это предположение позволяет получить ряд точных статистических результатов. Оно оказывается полезным при построении статистических тестов и доверительных интервалов, как мы увидим в разделе 4.5 и главе 5. Позже мы сможем ослабить это предположение, сохранив при этом большую часть полученных статистических результатов. (См. разделы 4.4 и 5.6.)

2.3.7. Независимость

Слово «независимость» в этой главе использовалось в нескольких разных смыслах.

В разделе 2.2 переменные из правой части уравнения модели называются независимыми. Здесь понятие независимости относится к источнику изменений. В контексте рассматриваемой модели источники изменения независимых переменных лежат за пределами описываемого процесса. Так,

в нашем примере из введения о связи дохода и пользования услугами здравоохранения мы предложили теорию, связывающую изменение дохода и изменение объема пользования услугами здравоохранения. Но мы не пытались объяснить изменение дохода в выборке; предполагается, что доход изменится по причинам, лежащим вне данной модели.

Предположение 3, $E[\varepsilon_i | \mathbf{X}] = 0$, называют **независимостью в среднем** (mean independence). Из него следует, что изменения шоков не объясняются изменениями независимых переменных. В подразделе 2.3.4 мы также предположили, что возмущения не коррелируют между собой. (Предположение A4 в табл. 2.1.) Отсюда также следует, что $E[\varepsilon_i | \varepsilon_j] = 0$ при $i \neq j$ — возмущения независимы в среднем между собой. Условная нормальность возмущений, сформулированная в подразделе 2.3.6 (предположение A6), влечет **статистическую независимость** (statistical independence) возмущений друг от друга, что является более сильным утверждением, чем независимость в среднем.

Наконец, в подразделе 2.3.2 говорится о **линейной независимости** (linear independence) столбцов матрицы \mathbf{X} . Здесь используется алгебраическое понятие независимости, связанное с рангом матрицы \mathbf{X} по столбцам. В данном случае это нужно интерпретировать так: переменные модели должны иметь возможность изменяться без сохранения какой-либо линейной зависимости. В примере 2.6 мы видели, что логарифмы площади, отношения сторон и высоты картины не могут изменяться независимо. Последствие этого для построения модели состоит в том, что если переменные не могут независимо изменяться, то их невозможно изучать в модели линейной регрессии, которая предполагает, что значение переменной можно изменять при сохранении значений остальных переменных. Здесь есть некоторая неясность: в примере 2.2 мы включили в модель возраст и квадрат возраста. Ответ заключается в том, что, хоть эти две переменные и связаны функциональной зависимостью, между ними нет линейной зависимости.

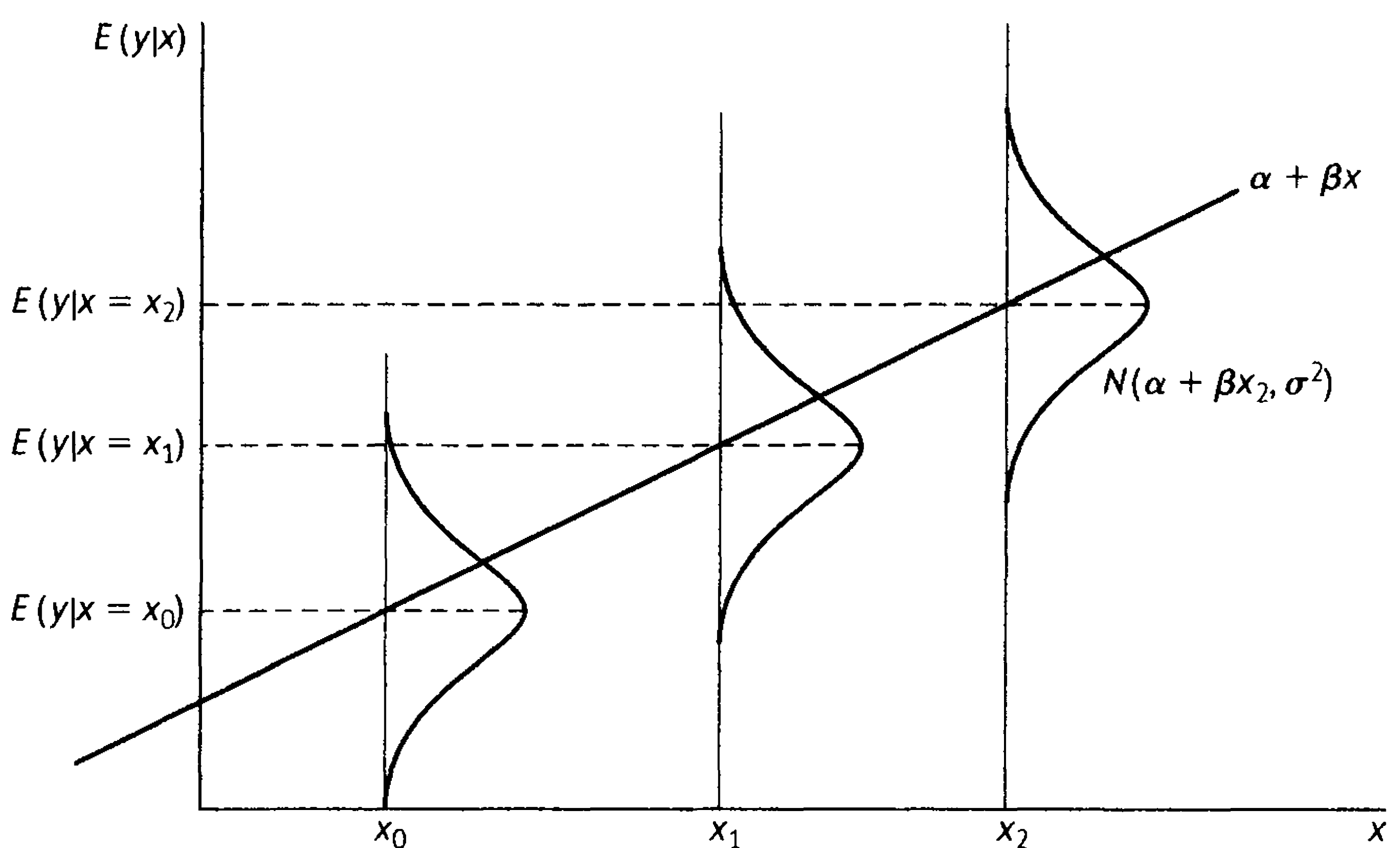


Рис. 2.3. Классическая модель регрессии

2.4. Заключение

В этой главе была введена модель линейной регрессии, являющаяся основной для построения эконометрических моделей. Все предположения классической модели регрессии изображены на рис. 2.3 для случая двух переменных.

Основные термины и понятия

| | |
|---|---|
| Autocorrelation — Автокорреляция | Linear independence — Линейная независимость |
| Central limit theorem — Центральная предельная теорема | Linear regression model — Модель линейной регрессии |
| Conditional median — Условная медиана | Loglinear model — Логлинейная модель (линейная в логарифмах модель) |
| Conditional variation — Условная дисперсия | Mean independence — Независимость в среднем |
| Constant elasticity — Постоянная эластичность | Multiple linear regression model — Модель множественной линейной регрессии |
| Counterfactual — Гипотетический | Nonautocorrelation — Отсутствие автокорреляции |
| Covariate — Ковариата | Nonstochastic regressors — Неслучайные регрессоры |
| Dependent variable — Зависимая переменная | Normality — Нормальность |
| Deterministic relationship — Детерминистическая зависимость | Normally distributed — Нормально распределенный |
| Disturbance — Возмущение, шок | Path diagram — Граф |
| Exogeneity — Экзогенность | Population regression equation — Теоретическое уравнение регрессии |
| Explained variable — Объясняемая переменная | Regressand — Регрессанд |
| Explanatory variable — Объясняющая переменная | Regression — Регрессия |
| Flexible functional form — Гибкая функциональная форма | Regressor — Регрессор |
| Full rank — Полный ранг | Second-order effects — Эффекты второго порядка |
| Heteroscedasticity — Гетероскедастичность | Semilog — Полулогарифмический |
| Homoscedasticity — Гомоскедастичность | Spherical disturbances — Сферические возмущения |
| Identification condition — Условие идентификации | Translog model — Транслогарифмическая модель |
| Impact of treatment on the treated — Эффект воздействия на подвергшегося воздействию | |
| Independent variable — Независимая переменная | |

ГЛАВА 3

Метод наименьших квадратов

3.1. Введение

В главе 2 модель линейной регрессии определялась через набор свойств распределения, из которого получены данные. Существует несколько различных подходов к оценке параметров модели. По ряду причин как теоретического, так и практического характера, которые мы рассмотрим в нескольких последующих главах, метод наименьших квадратов долгое время оставался наиболее популярным. Более того, в большинстве случаев, даже если найден более предпочтительный метод оценки, метод наименьших квадратов все равно используется для сравнения, а зачастую и сам используемый метод оказывается некоторой модификацией метода наименьших квадратов. В этой главе мы начнем изучать этот метод и вначале представим ряд полезных результатов алгебраического характера.

3.2. Регрессия наименьших квадратов

Объектом оценки являются неизвестные параметры стохастической зависимости $y_i = \mathbf{x}'_i \boldsymbol{\beta} + \varepsilon_i$. Необходимо различать параметры распределения $\boldsymbol{\beta}$ и ε_i и их выборочные аналоги, которые мы будем обозначать \mathbf{b} и e_i . **Теоретическая регрессия** (population regression) имеет вид $E[y_i | \mathbf{x}_i] = \mathbf{x}'_i \boldsymbol{\beta}$, а наша оценка $E[y_i | \mathbf{x}_i]$ записывается в виде

$$\hat{y}_i = \mathbf{x}'_i \mathbf{b}.$$

Шок, соответствующий i -му наблюдению, равен

$$\varepsilon_i = y_i - \mathbf{x}'_i \boldsymbol{\beta}.$$

Для любого значения \mathbf{b} мы будем оценивать ε_i при помощи **остатка** (residual):

$$e_i = y_i - \mathbf{x}'_i \mathbf{b}.$$

Из этих определений получаем

$$y_i = \mathbf{x}'_i \boldsymbol{\beta} + \varepsilon_i = \mathbf{x}'_i \mathbf{b} + e_i.$$

Эти уравнения применительно к регрессии с двумя переменными представлены на рис. 3.1.

Характеристика распределения (population quantity) $\boldsymbol{\beta}$ есть вектор неизвестных параметров распределения вероятности y_i , оценку которых мы будем пытаться построить на основе данных выборки (y_i, \mathbf{x}_i) , $i = 1, \dots, n$.

Это есть статистическая задача. Однако полезно начать с чисто алгебраической задачи выбора вектора \mathbf{b} , при котором подобранная прямая $\mathbf{x}'\mathbf{b}$ максимально близка к нашим наблюдениям. Мера близости определяется **критерием подгонки** (fitting criterion). Несмотря на то что предлагались разные критерии, чаще всего используют **метод наименьших квадратов**¹.

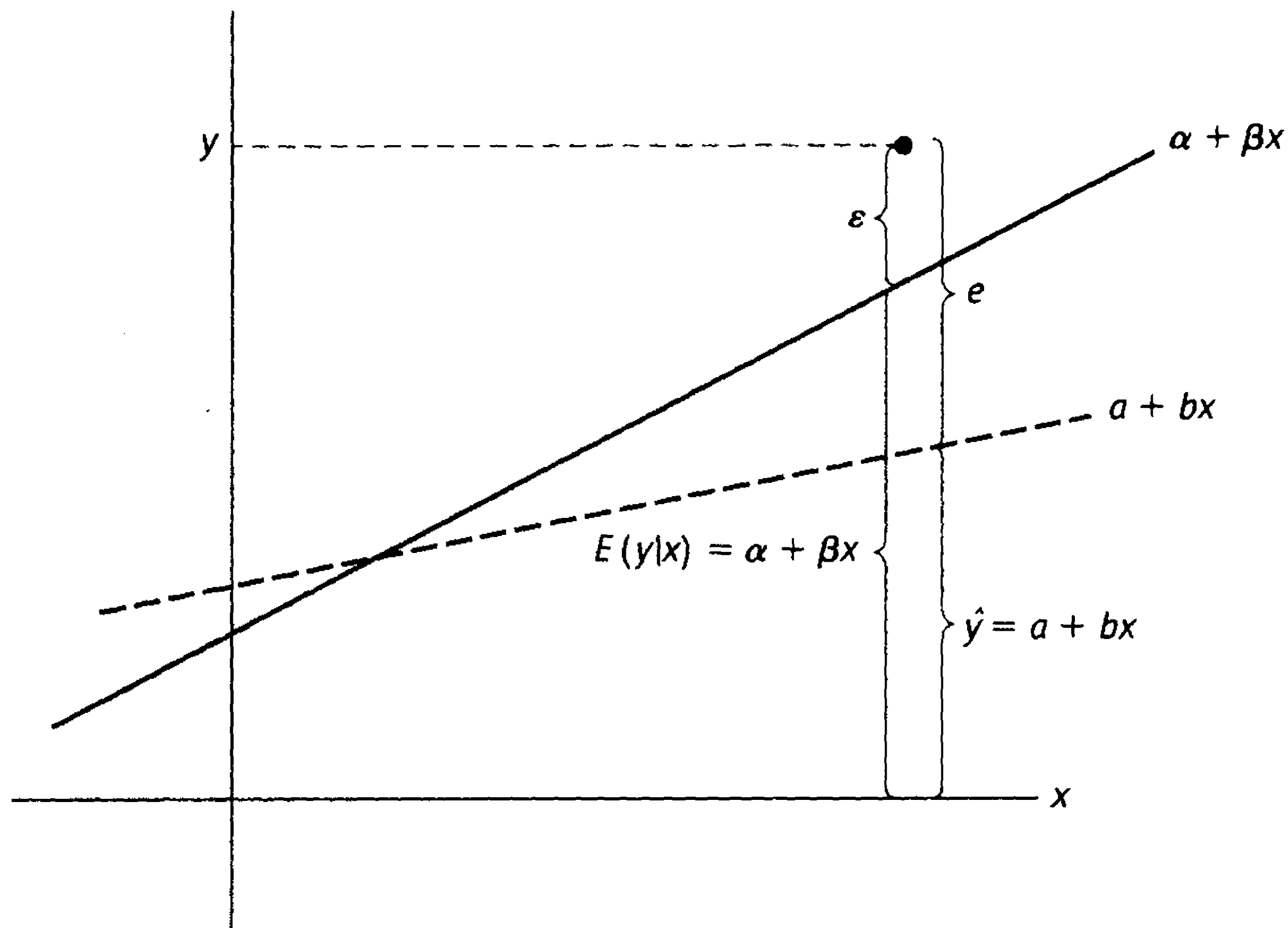


Рис. 3.1. Теоретическая и выборочная регрессии

3.2.1. Вектор коэффициентов метода наименьших квадратов

Вектор коэффициентов метода наименьших квадратов минимизирует сумму квадратов остатков:

$$\sum_{i=1}^n e_{i0}^2 = \sum_{i=1}^n (y_i - \mathbf{x}'_i \mathbf{b}_0)^2, \quad (3-1)$$

где \mathbf{b}_0 обозначает некоторый вектор коэффициентов. В матричных терминах минимизация суммы квадратов в (3-1) требует выбора такого \mathbf{b}_0 , что

$$\text{Minimize}_{\mathbf{b}_0} S(\mathbf{b}_0) = \mathbf{e}'_0 \mathbf{e}_0 = (\mathbf{y} - \mathbf{X}\mathbf{b}_0)'(\mathbf{y} - \mathbf{X}\mathbf{b}_0). \quad (3-2)$$

Раскрывая скобки, получаем

$$\mathbf{e}'_0 \mathbf{e}_0 = \mathbf{y}'\mathbf{y} - \mathbf{b}'_0 \mathbf{X}'\mathbf{y} - \mathbf{y}'\mathbf{X}\mathbf{b}_0 + \mathbf{b}'_0 \mathbf{X}'\mathbf{X}\mathbf{b}_0, \quad (3-3)$$

или

$$S(\mathbf{b}_0) = \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\mathbf{b}_0 + \mathbf{b}'_0 \mathbf{X}'\mathbf{X}\mathbf{b}_0.$$

¹ Нам еще необходимо показать, что наилучшее с точки зрения метода наименьших квадратов приближение данных прямой линией дает оценки с хорошими статистическими свойствами. Интуитивно кажется, что это и действительно справедливо. К этим вопросам мы вернемся в главе 4.

Необходимым условием минимизации является ²

$$\frac{\partial S(\mathbf{b}_0)}{\partial \mathbf{b}_0} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\mathbf{b}_0 = \mathbf{0}^2. \quad (3-4)$$

Пусть \mathbf{b} — решение этого уравнения. Тогда, преобразовывая (3-4), получаем, что \mathbf{b} удовлетворяет **нормальным уравнениям метода наименьших квадратов** (least squares normal equations):

$$\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{y}. \quad (3-5)$$

Если существует матрица, обратная к $\mathbf{X}'\mathbf{X}$ (что следует из предположения о полноте ранга A_2 из раздела 2.3), то решением будет

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \quad (3-6)$$

Чтобы это решение минимизировало сумму квадратов остатков, матрица

$$\frac{\partial^2 S(\mathbf{b}_0)}{\partial \mathbf{b}_0 \partial \mathbf{b}_0'} = 2\mathbf{X}'\mathbf{X}$$

должна быть положительно определенной. Пусть $q = \mathbf{c}'\mathbf{X}'\mathbf{X}\mathbf{c}$ для некоторого произвольного ненулевого вектора \mathbf{c} . Тогда

$$q = \mathbf{v}'\mathbf{v} = \sum_{i=1}^n v_i^2, \quad \text{где } \mathbf{v} = \mathbf{X}\mathbf{c}.$$

Если не все элементы \mathbf{v} равны нулю, то q положительно. Но если вектор \mathbf{v} нулевой, то, поскольку \mathbf{v} является линейной комбинацией столбцов \mathbf{X} , нарушается предположение о полноте ранга \mathbf{X} . Поскольку \mathbf{c} произвольно, q положительно для всех ненулевых \mathbf{c} , значит, $2\mathbf{X}'\mathbf{X}$ положительно определена. Таким образом, если \mathbf{X} имеет полный ранг, то минимум суммы квадратов остатков достигается, он единственный и равен \mathbf{b} .

3.2.2. Приложение: уравнение инвестиций

Чтобы проиллюстрировать вычисления метода наименьших квадратов в случае множественной регрессии, рассмотрим пример, основанный на макроэкономических данных из табл. F3.1 приложения. Чтобы оценить параметры уравнения инвестиций, мы сначала переведем данные инвестиций (Investment) и ВНП (GNP) из табл. F3.1 к реальному (Real) исчислению, разделив их на ИПЦ (CPI), а затем нормируем их так, чтобы значения были даны в триллионах долларов. Другие переменные регрессии — временной тренд (1, 2, ...), процентная ставка и темп инфляции, вычисляемый как процентное изменение ИПЦ. Полученные матрицы данных приведены в табл. 3.1. Рассмотрим сначала регрессию реальных инвестиций на константу, временной тренд и реальный ВНП, обозначив их за x_1 , x_2 и x_3 . (По причинам, которые мы будем обсуждать в главе 21, это не очень удачно специфицированное

²См. соответствующие результаты из курса математического анализа для матриц и векторов в приложении А.8.

уравнение для этих переменных. Однако оно вполне подойдет нам для простого численного примера.) Подставляя эти переменные в (3-5), получаем

$$\begin{aligned} b_1 n + b_2 \sum_i T_i + b_3 \sum_i G_i &= \sum_i Y_i, \\ b_1 \sum_i T_i + b_2 \sum_i T_i^2 + b_3 \sum_i T_i G_i &= \sum_i T_i Y_i, \\ b_1 \sum_i G_i + b_2 \sum_i T_i G_i + b_3 \sum_i G_i^2 &= \sum_i G_i Y_i. \end{aligned}$$

Решение можно получить следующим образом. Сначала разделим первое уравнение на n и преобразуем его так, чтобы

$$\begin{aligned} b_1 &= \bar{Y} - b_2 \bar{T} - b_3 \bar{G} = \\ &= 0,20333 - b_2 \times 8 - b_3 \times 1,2873. \end{aligned} \quad (3-7)$$

После подстановки этого результата в остальные два уравнения и ряда преобразований получаем набор из двух уравнений:

$$\begin{aligned} b_2 \sum_i (T_i - \bar{T})^2 + b_3 \sum_i (T_i - \bar{T})(G_i - \bar{G}) &= \sum_i (T_i - \bar{T})(Y_i - \bar{Y}), \\ b_2 \sum_i (T_i - \bar{T})(G_i - \bar{G}) + b_3 \sum_i (G_i - \bar{G})^2 &= \sum_i (G_i - \bar{G})(Y_i - \bar{Y}). \end{aligned} \quad (3-8)$$

Это показывает, что оценки коэффициентов наклона (при переменных) можно получить из сумм квадратов и попарных произведений переменных, записанных в отклонениях от своего среднего. Обозначая строчными буквами величины, записанные в отклонениях от среднего, получаем оценки метода наименьших квадратов b_2 и b_3 в виде

$$\begin{aligned} b_2 &= \frac{\sum_i t_i y_i \sum_i g_i^2 - \sum_i g_i y_i \sum_i t_i g_i}{\sum_i t_i^2 \sum_i g_i^2 - (\sum_i g_i t_i)^2} = \frac{1,6040(0,359609) - 0,066196(9,82)}{280(0,359609) - (9,82)^2} = -0,0171984, \\ b_3 &= \frac{\sum_i g_i y_i \sum_i t_i^2 - \sum_i t_i y_i \sum_i t_i g_i}{\sum_i t_i^2 \sum_i g_i^2 - (\sum_i g_i t_i)^2} = \frac{0,066196(280) - 1,6040(9,82)}{280(0,359609) - (9,82)^2} = 0,653723. \end{aligned}$$

Имея эти оценки, оценку b_1 можно получить из (3-7): $b_1 = -0,500639$.

Подумаем, что было бы, если бы мы оценивали регрессию только на константу и ВВП, без временного тренда. Корреляция, наблюдаемая в данных, будет частично объясняться, поскольку и инвестиции, и ВВП явно имеют временные тренды. Посмотрим, как этот эффект проявляется в вычислениях. Обозначая за b_{yg} коэффициент при x в парной регрессии y на x и константу, имеем

$$b_{yg} = \frac{\sum_i g_i y_i}{\sum_i g_i^2} = 0,184078. \quad (3-9)$$

Теперь разделим числитель и знаменатель выражения для b_3 на $\sum_i t_i^2 \sum_i g_i^2$. Преобразовывая полученную формулу, используя определение выборочной корреляции между G и T , $r_{gt}^2 = (\sum_i g_i t_i)^2 / (\sum_i g_i^2 \sum_i t_i^2)$, и определив таким же образом b_{yt} и b_{tg} , мы получаем

$$b_{yg \cdot t} = \frac{b_{yg}}{1 - r_{gt}^2} - \frac{b_{yt} b_{tg}}{1 - r_{gt}^2} = 0,653723. \quad (3-10)$$

Таблица 3.1. Матрицы данных

| Реальные инвестиции (V) | Константа (I) | Тренд (T) | ВВП (G) | Ставка инфляции (R) | (P) |
|-------------------------|---------------|-----------|---------|---------------------|------|
| 0,161 | 1 | 1 | 1,058 | 5,16 | 4,40 |
| 0,172 | 1 | 2 | 1,088 | 5,87 | 5,15 |
| 0,158 | 1 | 3 | 1,086 | 5,95 | 5,37 |
| 0,173 | 1 | 4 | 1,122 | 4,88 | 4,99 |
| 0,195 | 1 | 5 | 1,186 | 4,50 | 4,16 |
| 0,217 | 1 | 6 | 1,254 | 6,44 | 5,75 |
| 0,199 | 1 | 7 | 1,246 | 7,83 | 8,82 |
| y = 0,163 | X = 1 | 8 | 1,232 | 6,25 | 9,31 |
| 0,195 | 1 | 9 | 1,298 | 5,50 | 5,21 |
| 0,231 | 1 | 10 | 1,370 | 5,46 | 5,83 |
| 0,257 | 1 | 11 | 1,439 | 7,46 | 7,40 |
| 0,259 | 1 | 12 | 1,479 | 10,28 | 8,64 |
| 0,225 | 1 | 13 | 1,474 | 11,77 | 9,31 |
| 0,241 | 1 | 14 | 1,503 | 13,42 | 9,44 |
| 0,204 | 1 | 15 | 1,475 | 11,02 | 5,99 |
| 0,204 | 1 | 15 | 1,475 | 11,02 | 5,99 |

Примечание. Дальнейшие результаты основаны на этих значениях. Результаты, которые можно получить, введя в компьютерную программу данные табл. F3.1 и преобразовав их самостоятельно, могут незначительно отличаться от приводимых здесь.

(Использованную в левой части выражения запись $b_{yg \cdot t}$ надо интерпретировать как наклон линии регрессии y на g «при наличии t ».) Коэффициент **множественной регрессии** отличается от коэффициента простой регрессии тем, что требуется поправка, отвечающая за влияние переменной t на Y и G . В качестве еще более наглядного примера этого эффекта можно рассмотреть следующую ситуацию. В регрессии реальных инвестиций на временной тренд мы получим положительную оценку коэффициента $b_{yt} = 1,604/280 = 0,0057286$, отражающую восходящий тренд, присутствующий в данных. Однако в множественной регрессии, учитывающей влияние ВВП на реальные инвестиции, коэффициент при временном тренде будет равен $-0,0171984$, указывая на нисходящий временной тренд. Общий результат для регрессии с тремя переменными, где x_1 — константа, такой:

$$b_{y2 \cdot 3} = \frac{b_{y2} - b_{y3}b_{32}}{1 - r_{23}^2}. \quad (3-11)$$

Из этой формулы ясно, что величины $b_{y2 \cdot 3}$ и b_{y2} могут существенно различаться. Даже их знаки могут оказаться различными.

На практике вам никогда не придется строить оценки вручную или с помощью калькулятора. При оценке регрессий с более чем тремя независимыми переменными нельзя обойтись без инструментов матричной алгебры (и без компьютера). Рассмотрим, например, расширенную модель инвестиций, в которую в дополнение к константе, временному тренду и ВВП входят процентная ставка и темп инфляции. Применение метода наименьших квадратов требует решения системы из пяти нормальных уравнений. Обо-

значая за X и y матрицы данных, запишем нормальные уравнения (3-5) в виде

$$\begin{bmatrix} 15,000 & 120,00 & 19,310 & 111,79 & 99,770 \\ 120,000 & 1240,0 & 164,30 & 1035,9 & 875,60 \\ 19,310 & 164,30 & 25,218 & 148,98 & 131,22 \\ 111,79 & 1035,9 & 148,98 & 953,86 & 799,02 \\ 99,770 & 875,60 & 131,22 & 799,02 & 716,67 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{bmatrix} = \begin{bmatrix} 3,05 \\ 26,004 \\ 3,9926 \\ 23,521 \\ 20,732 \end{bmatrix}.$$

Решение будет иметь вид

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (-0,50907, -0,01658, 0,67038, -0,002326, -0,00009401)'.$$

3.2.3. Алгебраические свойства оценки наименьших квадратов

Нормальные уравнения имеют вид

$$\mathbf{X}'\mathbf{X}\mathbf{b} - \mathbf{X}'\mathbf{y} = -\mathbf{X}'(\mathbf{y} - \mathbf{X}\mathbf{b}) = -\mathbf{X}'\mathbf{e} = 0. \quad (3-12)$$

Таким образом, для каждого столбца x_k матрицы X выполняется $x_k'e = 0$. Если первый столбец X состоит из единиц (обозначим такой вектор \mathbf{i}), то отсюда можно вывести три следствия:

1. Сумма остатков метода наименьших квадратов равна нулю. Это следствие выводится из того, что $\mathbf{x}'_1\mathbf{e} = \mathbf{i}'\mathbf{e} = \sum_i e_i = 0$.
2. Среднее арифметическое полученных приближений зависимой переменной равно среднему арифметическому значений в выборке. Из первого нормального уравнения следует, что $\bar{y} = \bar{x}'\mathbf{b}$.
3. Среднее арифметическое полученных приближений зависимой переменной равно среднему арифметическому значений в выборке. Этот вывод следует из первого следствия, поскольку $\hat{y} = \mathbf{X}\mathbf{b}$.

Важно отметить, что ни одно из этих следствий не обязано выполняться при отсутствии постоянного члена.

3.2.4. Проекции

Вектор остатков метода наименьших квадратов равен

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{b}. \quad (3-13)$$

Подставляя значение \mathbf{b} из (3-6), получаем

$$\mathbf{e} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y} = \mathbf{M}\mathbf{y}. \quad (3-14)$$

Матрица \mathbf{M} размера $n \times n$, определенная в (3-14), крайне важна для изучения регрессий. Можно показать, что \mathbf{M} симметрична ($\mathbf{M} = \mathbf{M}'$) и идемпотентна ($\mathbf{M} = \mathbf{M}^2$). Из (3-13) можно видеть, что матрица \mathbf{M} , будучи умноженной слева на \mathbf{y} , дает вектор остатков метода наименьших квадратов в регрессии \mathbf{y} на \mathbf{X} . (В дальнейшем мы будем называть эту матрицу **генератором остатков**.) Следовательно,

$$\mathbf{M}\mathbf{X} = 0. \quad (3-15)$$

Один из способов интерпретации этого результата состоит в том, что если оценивать регрессию \mathbf{X} на \mathbf{X} , то будет получена точная оценка, а остатки будут нулевыми.

Наконец, из (3-13) следует, что $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$, что является выборочным аналогом (2-3). (См. также рис. 3.1.) Оценивание методом наименьших квадратов разбивает вектор \mathbf{y} на две части: на приближенные значения $\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{b}}$ и на остатки \mathbf{e} . [См. раздел А3.7, в особенности (А-54).] Поскольку $\mathbf{M}\mathbf{X} = \mathbf{0}$, эти две компоненты ортогональны. Из (3-13) следует, что

$$\hat{\mathbf{y}} = \mathbf{y} - \mathbf{e} = (\mathbf{I} - \mathbf{M})\mathbf{y} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{P}\mathbf{y}. \quad (3-16)$$

Матрица \mathbf{P} — **матрица проекции**, или просто **проектор** (projection matrix). Это матрица, полученная из \mathbf{X} , такая, что при умножении вектора \mathbf{y} на матрицу \mathbf{P} слева получаются подобранные с помощью метода наименьших квадратов значения \mathbf{y} при оценке регрессии на \mathbf{X} . Эти значения также являются **проекцией** вектора \mathbf{y} на гиперплоскость, порожденную столбцами \mathbf{X} . (См. разделы А3.5 и А3.7.) Непосредственно проверяется, что \mathbf{P} , как и \mathbf{M} , симметрична и идемпотентна. Из предыдущих результатов также следует, что \mathbf{M} и \mathbf{P} ортогональны:

$$\mathbf{P}\mathbf{M} = \mathbf{M}\mathbf{P} = \mathbf{0}.$$

Как видно из (3-15),

$$\mathbf{P}\mathbf{X} = \mathbf{X}.$$

Объединяя (3-14) и (3-16), мы видим, что оценивание методом наименьших квадратов разделяет вектор \mathbf{y} на две ортогональные части:

$$\mathbf{y} = \mathbf{P}\mathbf{y} + \mathbf{M}\mathbf{y} = \text{проекция} + \text{остаток}.$$

На рис. 3.2 этот результат проиллюстрирован для случая двух переменных. Серая плоскость порождена столбцами \mathbf{X} . Проекция и остаток — перпендикулярные штрихованные линии. Здесь можно также воспользоваться теоремой Пифагора применительно к сумме квадратов:

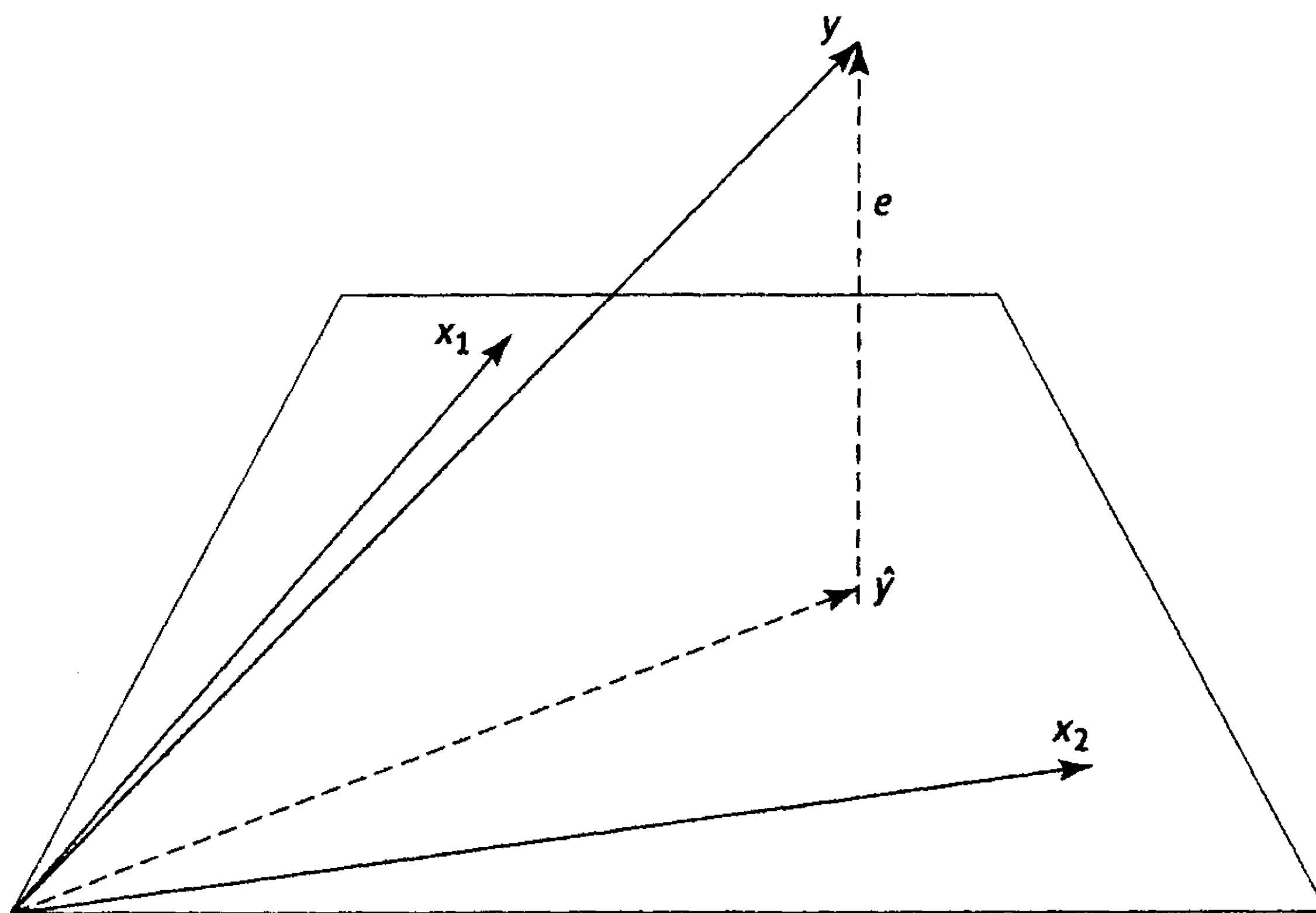
$$\mathbf{y}'\mathbf{y} = \mathbf{y}'\mathbf{P}'\mathbf{P}\mathbf{y} + \mathbf{y}'\mathbf{M}'\mathbf{M}\mathbf{y} = \hat{\mathbf{y}}'\hat{\mathbf{y}} + \mathbf{e}'\mathbf{e}.$$

При работе с уравнениями, связанными с методом наименьших квадратов, часто бывают полезны следующие эквивалентные формулы для суммы квадратов остатков:

$$\begin{aligned} \mathbf{e}'\mathbf{e} &= \mathbf{y}'\mathbf{M}'\mathbf{M}\mathbf{y} = \mathbf{y}'\mathbf{M}\mathbf{y} = \mathbf{y}'\mathbf{e} = \mathbf{e}'\mathbf{y}, \\ \mathbf{e}'\mathbf{e} &= \mathbf{y}'\mathbf{y} - \mathbf{b}'\mathbf{X}'\mathbf{X}\mathbf{b} = \mathbf{y}'\mathbf{y} - \mathbf{b}'\mathbf{X}'\mathbf{y} = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\mathbf{b}. \end{aligned}$$

3.3. Раздельная и частичная регрессии

При использовании модели множественной регрессии исследователя часто интересует лишь один или несколько коэффициентов. Возьмем, например, модель заработной платы, рассматриваемую в примере 2.2. Нас интересует только связь между уровнем образования и зарплатой, но мы тем не

Рис. 3.2. Проекция y на плоскость, порожденную столбцами X

менее вынуждены включить в модель еще и возраст. Здесь мы рассматриваем вопрос о том, какие вычисления нужно провести для получения коэффициентов лишь при некоторых переменных множественной регрессии (например, коэффициента при уровне образования в вышеупомянутой модели).

Предположим, что в регрессии участвуют два набора переменных, X_1 и X_2 , т.е. модель имеет вид

$$y = X\beta + \varepsilon = X_1\beta_1 + X_2\beta_2 + \varepsilon.$$

Какой вид имеет алгебраическое решение для b_2 ? Запишем **нормальные уравнения**:

$$\begin{aligned} (1) \quad & \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} X_1'y \\ X_2'y \end{bmatrix}. \\ (2) \quad & \end{bmatrix} \end{aligned} \quad (3-17)$$

Решение можно получить с помощью блочной обратной матрицы из (А-74). Также можно напрямую воспользоваться формулами (1) и (2) в (3-17), решив их относительно b_2 . Сначала решим (1), выразив b_1 :

$$b_1 = (X_1'X_1)^{-1}X_1'y - (X_1'X_1)^{-1}X_1'X_2b_2 = (X_1'X_1)^{-1}X_1'(y - X_2b_2). \quad (3-18)$$

Таким образом, b_1 состоит из коэффициентов регрессии y на X_1 и поправочного вектора. Сейчас мы сделаем небольшое отступление и обсудим важный результат, содержащийся в (3-18). Предположим, что $X_1'X_2 = 0$. Тогда $b_1 = (X_1'X_1)^{-1}X_1'y$, т.е. совпадает с вектором коэффициентов в регрессии y на X_1 . Общий результат сформулирован в следующей теореме.

Теорема 3.1. Ортогональная частичная регрессия

При оценке методом наименьших квадратов множественной регрессии y на два ортогональных множества переменных X_1 и X_2 векторы коэффициентов при переменных из каждого множества можно получить, оценивая регрессию y отдельно на переменные из X_1 и отдельно на переменные из X_2 .

Доказательство. Предпосылка теоремы состоит в том, что $X_1'X_2 = 0$ в нормальных уравнениях (3-17). Подставляя это в (3-18), получаем $b_1 = (X_1'X_1)^{-1}X_1'y$ и аналогичный результат для b_2 .

Если множества переменных X_1 и X_2 неортогональны, то полученные в (3-17) и (3-18) решения относительно b_1 и b_2 несколько сложнее, чем простые регрессии в теореме 3.1. Более общее решение описывается следующей теоремой, впервые опубликованной в первом томе журнала *Econometrica*³.

Теорема 3.2. Теорема Фриша–Во–Ловелла

При оценке методом наименьших квадратов регрессии переменной y на два набора переменных, X_1 и X_2 , подвектор b_2 есть множество коэффициентов, получаемых при оценке регрессии остатков регрессии y на X_1 на набор остатков регрессий каждого из столбцов матрицы X_2 на X_1 .

Доказательство. Начнем доказательство теоремы 3.2 с уравнения (2) в (3-17), т.е.

$$X_2'X_1b_1 + X_2'X_2b_2 = X_2'y.$$

Подставим сюда выражение для b_1 из (3-18). Получаем

$$X_2'X_1(X_1'X_1)^{-1}X_1'y - X_2'X_1(X_1'X_1)^{-1}X_1'X_2b_2 + X_2'X_2b_2 = X_2'y.$$

Отсюда получаем решение:

$$\begin{aligned} b_2 &= [X_2'(I - X_1(X_1'X_1)^{-1}X_1')X_2]^{-1} [X_2'(I - X_1(X_1'X_1)^{-1}X_1')y] = \\ &= (X_2'M_1X_2)^{-1}(X_2'M_1y). \end{aligned} \quad (3-19)$$

Матрицы в круглых скобках представляют собой «генераторы остатков», определенные в уравнении (3-14). В данном случае они определяются для регрессии на столбцы X_1 . Значит, M_1X_2 есть матрица остатков: каждый столбец M_1X_2 есть вектор остатков в регрессии соответствующего столбца X_2

³Теорема была сформулирована во введении к статье следующим образом: «Включение тренда в регрессию никогда не даст лучших результатов, чем выделение тренда из отдельных переменных, поскольку оба метода по определению приводят к идентичным результатам». То есть Фриша и Во (Frisch and Waugh (1933)) интересовало отсутствие разницы между оценкой регрессии y на временной тренд t и переменную x и оценкой регрессии остатков от y при регрессии на константу и временной тренд на вычисленные так же остатки от x . Приводимая формулировка теоремы в терминах матриц была сформулирована позднее Ловеллом в работе Lovell (1963).

на переменные X_1 . Учитывая, что матрица M_1 , как и M , симметрична и идемпотентна, можно переписать (3-19) в виде

$$b_2 = (X_2^{*'} X_2^*)^{-1} X_2^{*'} y^*, \quad (3-20)$$

где

$$X_2^* = M_1 X_2 \quad \text{и} \quad y^* = M_1 y.$$

Этот результат является фундаментальным для анализа регрессий.

Эти действия часто называют **выделением** (partialing out, netting out) эффекта X_1 . По этой же причине коэффициенты множественной регрессии иногда называют **коэффициентами частичной** (partial) **регрессии**. Приложение этой теоремы к предложенной в начале этой главы задаче вычисления отдельного коэффициента производится следующим образом. Возьмем регрессию y на переменные X и дополнительную переменную z . Обозначим коэффициенты как b и c .

Вывод 3.2.1. Отдельные коэффициенты регрессии

Коэффициент при z в множественной регрессии y на $W = [X, z]$ можно записать в виде $c = (z' M z)^{-1} (z' M y) = (z^{'} z^*)^{-1} z^{*'} y^*$ где z^* и y^* — векторы остатков метода наименьших квадратов из регрессий z и y на X ; $z^* = M z$ и $y^* = M y$, где M определялась формулой (3-14).*

Это есть частный случай теоремы 3.2 с X_1 , равным X , и X_2 , равным z .

Говоря в терминах примера 2.2, мы можем получить коэффициент при уровне образования в множественной регрессии, сначала оценив регрессии заработка и уровня образования на возраст (или возраст и квадрат возраста), а затем использовать остатки этих регрессий в простой парной регрессии. Классическое применение этого результата, сделанное в работе Frisch and Waugh (1933) (авторы которой и были авторами самой теоремы), состоит в том, что при анализе временных рядов оценка регрессии с временным трендом дает те же результаты, что и регрессия, использующая переменные с устраненным трендом (детрендрованные переменные)⁴.

В качестве примера рассмотрим случай, когда X_1 является постоянным членом (т.е. i — первый столбец X состоит из единиц). Тогда b_2 — это коэффициенты наклона в регрессии с постоянным членом. Используя теорему 3.2, получаем вектор остатков для любой переменной из X_2 :

$$\begin{aligned} x^* &= x - X_1 (X_1' X_1)^{-1} X_1' x \\ &= x - i(i'i)^{-1} i' x \\ &= x - i(1/n) i' x \\ &= x - i \bar{x} \\ &= M^0 x. \end{aligned} \quad (3-21)$$

(См. раздел A5.4, где мы выводим этот результат алгебраически.) Значит, в этом случае остатки есть отклонения от выборочного среднего. Таким образом, каждый столбец $M_1 X_2$ — это первоначальная переменная, записанная

⁴Как в нашем примере про инвестиции.

в отклонениях от своего выборочного среднего. Этот общий результат может быть записан как следующий вывод.

Вывод 3.2.2. Регрессия с постоянным членом

Коэффициенты наклона (коэффициенты при «экономических» переменных) множественной регрессии, содержащей постоянный член, могут быть получены путем замены всех переменных их отклонениями от выборочного среднего и построения оценки коэффициентов регрессии переменной y , записанной в новом виде, на объясняющие переменные, также записанные в виде отклонений.

[Этот результат использовался в (3-8).] После получения коэффициентов при X_2 как получить коэффициенты при X_1 (коэффициент при константе)? Можно, конечно, повторить всю процедуру, поменяв местами X_1 и X_2 . Но есть и более простой способ. Мы уже нашли b_2 . Значит, мы можем воспользоваться (3-18) и найти b_1 . Если X_1 — просто столбец единиц, то первое уравнение приводит к известной формуле

$$b_1 = \bar{y} - \bar{x}_2 b_2 - \dots - \bar{x}_K b_K,$$

[чем мы воспользовались в (3-7)].

Теорему 3.2 и выводы 3.2.1 и 3.2.2 вместе можно использовать для интерпретации отдельной регрессии в том случае, если модель содержит постоянную. По теореме 3.1 если столбцы X ортогональны, т. е. $x'_k x_m = 0$ для любых столбцов k и m , то коэффициенты в регрессии y на X при $X = [x_1, x_2, \dots, x_K]$ равны $x'_k y / x'_k x_k$. Если регрессия содержит постоянную, то мы можем вычислить коэффициенты множественной регрессии путем построения регрессии y в отклонениях на столбцы X тоже в отклонениях от их средних. В этом случае «ортогональность» столбцов означает, что выборочные ковариации (и корреляции) переменных равны нулю. Получаем новую теорему.

Теорема 3.3. Ортогональная регрессия

Если множественная регрессия y на X содержит постоянную и переменные регрессии не коррелируют между собой, то коэффициенты наклона в множественной регрессии равны коэффициентам парных регрессий y на константу и каждую переменную по отдельности.

Доказательство. Этот результат следует из теорем 3.1 и 3.2.

3.4. Частичная регрессия и частные коэффициенты корреляции

Множественная регрессия может применяться для проведения мысленного эксперимента, который, возможно, нельзя провести в реальности, анализа *ceteris paribus* (при прочих равных условиях), который часто встреча-

ется в экономических рассуждениях. Возвращаясь к примеру 2.2, уравнение регрессии, связывающей заработок с возрастом и уровнем образования, позволяет нам провести мысленный эксперимент, сравнивающий заработки двух людей одинакового возраста с разными уровнями образования, *даже если выборка не содержит ни одной такой пары*. Именно это свойство регрессий подразумевается под названием *частные коэффициенты регрессии*. Чтобы провести такой эксперимент, нужно сначала оценить регрессии заработка и уровня образования на возраст и вычислить остатки этих регрессий. По построению возраст совершенно никак не будет объяснять разброс этих остатков. Значит, корреляция между такими «очищенными» заработком и уровнем образования не будет зависеть от возраста (она «очищена» от его влияния).

Тот же подход можно применить к изучению коэффициентов корреляции между переменными. Продолжая наш пример, с какой долей уверенности мы можем утверждать, что корреляция между заработком и уровнем образования отражает прямую связь между ними, а не их совместную положительную зависимость от возраста? Чтобы ответить на этот вопрос, используются **частные коэффициенты корреляции**, способ расчета которых похож на способ вычисления отдельных коэффициентов регрессии. В нашем примере частный коэффициент корреляции заработка и уровня образования, исключаящий эффект возраста, вычисляется так:

1. y_* = остатки в регрессии заработка на константу и возраст.
2. z_* = остатки в регрессии уровня образования на константу и возраст.
3. Частная корреляция r_{yz}^{*2} равна простой корреляции y_* и z_* .

Может показаться, что здесь требуется большой объем вычислений. С помощью вывода 3.2.1 можно записать векторы остатков в пунктах 1 и 2 в виде $y_* = My$ и $z_* = Mz$, где $M = I - X(X'X)^{-1}X'$ — генератор остатков, определенный в (3-14). Мы будем предполагать, что X содержит константу, так что векторы остатков y_* и z_* имеют нулевые средние. Тогда квадрат частного коэффициента корреляции равен

$$r_{yz}^{*2} = \frac{(z_*'y_*)^2}{(z_*'z_*)(y_*'y_*)}.$$

Есть также еще одно удобное соотношение. После оценки множественной регрессии t -статистику из (5-13), применяемую для проверки гипотезы о равенстве коэффициента нулю (т.е. последний столбец табл. 4.1), можно использовать для вычисления:

$$r_{yz}^{*2} = \frac{t_z^2}{t_z^2 + \text{число степеней свободы}}, \quad (3-22)$$

где **число степеней свободы** равно $n - (K + 1)$. Доказательство этого неочевидного интуитивно результата может быть полезно в качестве иллюстрации к некоторым результатам, касающимся частичной регрессии. Мы будем пользоваться двумя полезными алгебраическими результатами метода наименьших квадратов. Первый дает способ вычислить отдельный диагональный элемент матрицы, обратной **матрице моментов**, такой как $(X'X)^{-1}$.

Теорема 3.4. Диагональные элементы матрицы, обратной к матрице моментов

Пусть WPI — блочная матрица $[X, z]$, т.е. матрица, состоящая из K столбцов X и дополнительного столбца z . Последний диагональный элемент $(W'W)^{-1}$ равен $(z'Mz)^{-1} = (z'_*z_*)^{-1}$, где $z_* = Mz$ и $M = I - X(X'X)^{-1}X'$.

Доказательство. Эта теорема есть приложение формулы обращения блочной матрицы (А-74), где $A_{11} = X'X$, $A_{12} = X'z$, $A_{21} = z'X$ и $A_{22} = z'z$. Обратим внимание, что эта теорема обобщает результат раздела А2.8, где X содержал только константу, т.е. столбец i .

Мы можем воспользоваться теоремой 3.4, чтобы получить (3-22). Пусть c и u обозначают соответственно коэффициент при z и вектор остатков множественной регрессии y на $W = [X, z]$.

Тогда по определению квадрат t -статистики в (3-22) равен

$$t_z^2 = \frac{c^2}{\left[\frac{u'u}{n-(K+1)} \right] (W'W)^{-1}_{K+1, K+1}},$$

где $(W'W)^{-1}_{K+1, K+1}$ — это $K+1$ -й (последний) диагональный элемент матрицы $(W'W)^{-1}$. (Член в скобках входит в формулу (4-17). Мы пока пользуемся только алгебраическим результатом.) Теорема утверждает, что этот элемент равен $(z'_*z_*)^{-1}$. Из вывода 3.2.1 мы знаем, что $c^2 = [(z'_*y_*)/(z'_*z_*)]^2$. Для удобства обозначим $DF = n - (K + 1)$. Тогда

$$t_z^2 = \frac{(z'_*y_*/z'_*z_*)^2}{(u'u/DF)/z'_*z_*} = \frac{(z'_*y_*)^2 DF}{(u'u)(z'_*z_*)}.$$

Отсюда следует, что формула (3-22) эквивалентна:

$$\frac{t_z^2}{t_z^2 + DF} = \frac{\frac{(z'_*y_*)^2 DF}{(u'u)(z'_*z_*)}}{\frac{(z'_*y_*)^2 DF}{(u'u)(z'_*z_*)} + DF} = \frac{\frac{(z'_*y_*)^2}{(u'u)(z'_*z_*)}}{\frac{(z'_*y_*)^2}{(u'u)(z'_*z_*)} + 1} = \frac{(z'_*y_*)^2}{(z'_*y_*)^2 + (u'u)(z'_*z_*)}.$$

Разделив числитель и знаменатель на $(z'_*z_*)(y'_*y_*)$, мы получаем

$$\begin{aligned} \frac{t_z^2}{t_z^2 + DF} &= \frac{(z'_*y_*)^2 / (z'_*z_*)(y'_*y_*)}{(z'_*y_*)^2 / (z'_*z_*)(y'_*y_*) + (u'u)(z'_*z_*) / (z'_*z_*)(y'_*y_*)} = \\ &= \frac{r_{yz}^{*2}}{r_{yz}^{*2} + (u'u) / (y'_*y_*)}. \end{aligned} \quad (3-23)$$

Теперь нам нужно воспользоваться еще одной теоремой для преобразования $u'u$. Нужный нам результат формулируется как теорема 3.5.

Теорема 3.5. *Изменение суммы квадратов при включении в регрессию дополнительной переменной*

Если $e'e$ — сумма квадратов остатков регрессии y на X , а $u'u$ — сумма квадратов остатков при регрессии y на X и z , то

$$u'u = e'e - c^2(z'_*z_*) \leq e'e, \quad (3-24)$$

где c — коэффициент при z в удлиненной регрессии y на $[X, z]$, а $z_* = Mz$ — вектор остатков регрессии z на X .

Доказательство. В удлиненной регрессии y на X и z вектор остатков равен $u = y - Xd - zc$. Отметим, что если не выполняется равенство $X'z = 0$, то d не будет равно $b = (X'X)^{-1}X'y$. (См. подраздел 4.3.2.) Более того, если не выполняется равенство $c = 0$, то u не будет равно $e = y - Xb$. Согласно выводу 3.2.1, $c = (z'_*z_*)^{-1}(z'_*y_*)$.

Из (3-18) следует, что коэффициенты при X в этой удлиненной регрессии равны

$$d = (X'X)^{-1}X'(y - zc) = b - (X'X)^{-1}X'zc.$$

Подставляя это выражение для d в выражение для u , мы получаем

$$u = y - Xb + X(X'X)^{-1}X'zc - zc = e - Mzc = e - z_*c.$$

Значит,

$$u'u = e'e + c^2(z'_*z_*) - 2c(z'_*e).$$

Однако $e = My = y_*$, а $z'_*e = z'_*y_* = c(z'_*z_*)$. Подставляя эти равенства в выражение для $u'u$ выше, мы получаем утверждение теоремы.

Возвращаясь к нашему рассуждению, мы получаем, что $e'e = y'_*y_*$ и $c^2(z'_*z_*) = (z'_*y_*)^2 / (z'_*z_*)$. Таким образом,

$$\frac{u'u}{y'_*y_*} = \frac{y'_*y_* - (z'_*y_*)^2 / z'_*z_*}{y'_*y_*} = 1 - r_{yz}^{*2}.$$

Подставляя эти выражения в знаменатель выражения (3-23), мы выводим желаемый результат.

Пример 3.1. Частные корреляции

В данных из приложения в подразделе 3.2.2 простые корреляции между инвестициями и регрессором r_{yk} и частные корреляции r_{yk}^* между инвестициями и четырьмя регрессорами (при заданных значениях остальных переменных) приведены в табл. 3.2. Как видно из таблицы, нет четкой связи между простыми и частными корреляциями. Стоит обратить внимание на знаки коэффициентов. Знаки частных коэффициентов корреляции равны знакам соответствующих коэффициентов регрессии, три из которых отрицательны. Все простые коэффициенты корреляции положительны из-за «скрытых» временных эффектов.

Таблица 3.2. Корреляция между инвестициями и другими переменными

| | Простая корреляция | Частная корреляция |
|-------------------|--------------------|--------------------|
| Временной тренд | 0,7496 | -0,9360 |
| ВВП | 0,8632 | 0,9680 |
| Процентная ставка | 0,5871 | -0,5167 |
| Темп инфляции | 0,4777 | -0,0221 |

3.5. Качество приближения и анализ разброса

Используемый нами критерий качества приближения, сумма квадратов остатков, является мерой близости линии регрессии к данным. Однако легко видеть, что сумма квадратов остатков может произвольно изменяться при умножении всех значений y на некоторую константу. Поскольку приближенные значения, получаемые из регрессии, определяются значениями x , нас, скорее, интересует вопрос, насколько хорошим объяснением изменения y являются изменения x . На рис. 3.3 показаны три возможных случая для модели линейной регрессии. Мера качества приближения, которую мы сейчас опишем, определяется как критерием качества приближения, так и мерой ковариации между y и x .

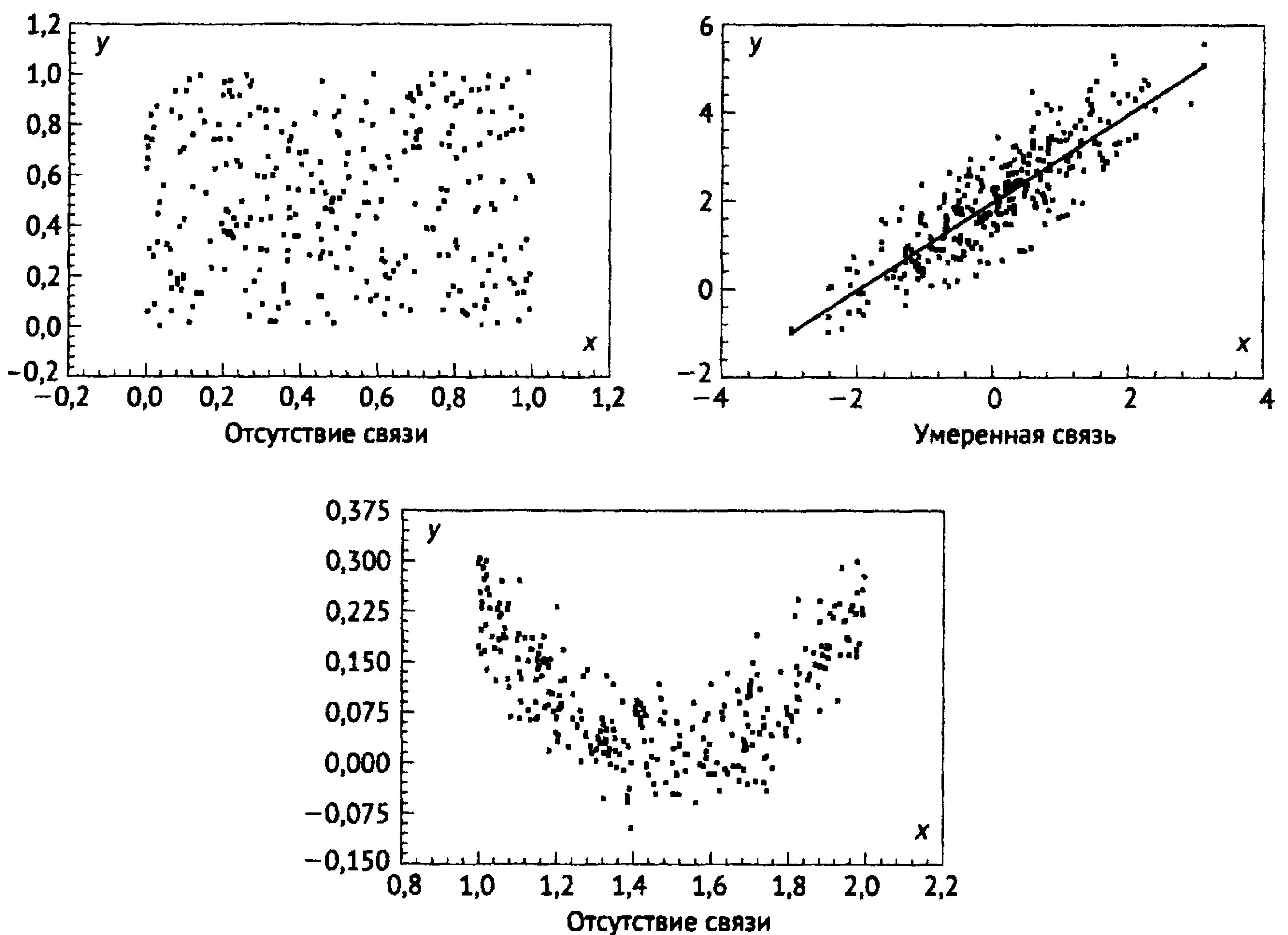


Рис. 3.3. Выборка

Дисперсия зависимой переменной определяется с помощью отклонений значений этой переменной от среднего ($y_i - \bar{y}$). Полная вариация y — это сумма квадратов отклонений:

$$SST = \sum_{i=1}^n (y_i - \bar{y})^2.$$

В контексте уравнения регрессии можно записать

$$y = \mathbf{Xb} + \mathbf{e} = \hat{y} + \mathbf{e}.$$

Для отдельного наблюдения имеем

$$y_i = \hat{y}_i + e_i = \mathbf{x}_i' \mathbf{b} + e_i.$$

Если регрессия содержит константу, то сумма остатков будет равна нулю, а среднее предсказанных значений \hat{y}_i будет равно среднему настоящих значений y_i в выборке. Вычитая из обеих частей \bar{y} и применяя результат 2 из подраздела 3.2.3, получаем

$$y_i - \bar{y} = \hat{y}_i - \bar{y} + e_i = (\mathbf{x}_i - \bar{\mathbf{x}})' \mathbf{b} + e_i.$$

Рисунок 3.4 иллюстрирует эти вычисления в случае регрессии с двумя переменными. Интуитивно понятно, что регрессия будет хорошо описывать данные в том случае, когда отклонения y от среднего в большей степени объясняются отклонениями x от среднего, чем остатками. Поскольку сумма членов этого разложения равна нулю, для оценки воспользуемся суммой квадратов. Для полной выборки имеем

$$\mathbf{M}^0 \mathbf{y} = \mathbf{M}^0 \mathbf{X} \mathbf{b} + \mathbf{M}^0 \mathbf{e},$$

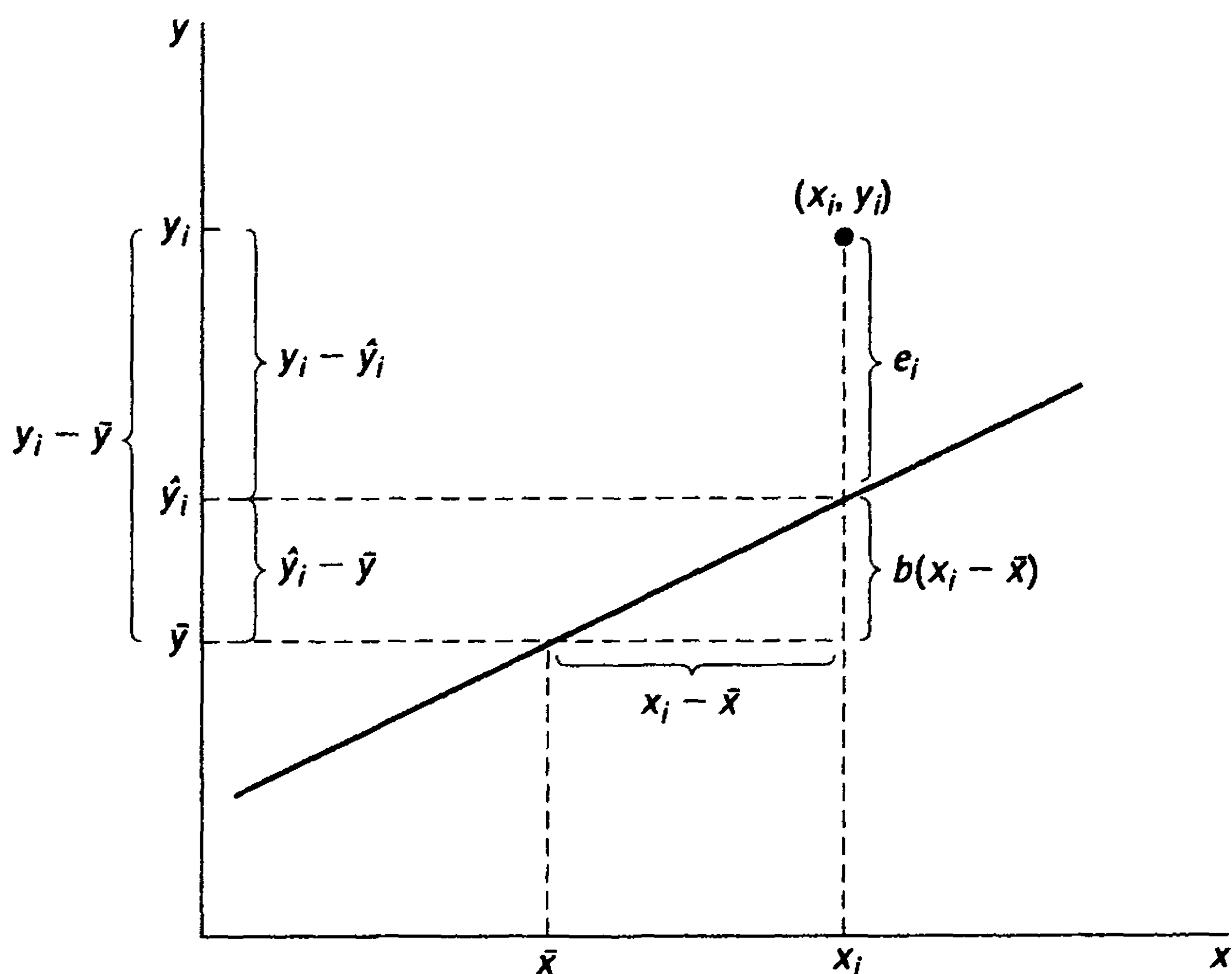


Рис. 3.4. Разложение y_i

где M^0 — идемпотентная матрица размером $n \times n$, приводящая наблюдения к отклонениям от средних значений. (См. (3-21) и раздел А2.8.) Столбец M^0X , соответствующий постоянному члену, равен нулю, а так как среднее остатков тоже равно нулю, то $M^0e = e$. Далее, поскольку $e'M^0X = e'X = 0$, сумма квадратов равна

$$y'M^0y = b'X'M^0Xb + e'e.$$

Перепишем это как:

Полная (total) сумма (sum of) квадратов (squares) = Сумма квадратов регрессии (regression) + Сумма квадратов ошибок (error), т. е.

$$SST = SSR + SSE. \quad (3-25)$$

(Отметим, что это же разложение используется в конце подраздела 3.2.4.)

Теперь мы можем получить меру того, насколько хорошо линия регрессии подгоняет наблюдаемые данные:

$$\text{коэффициент детерминации: } \frac{SSR}{SST} = \frac{b'X'M^0Xb}{y'M^0y} = 1 - \frac{e'e}{y'M^0y}. \quad (3-26)$$

Коэффициент детерминации обозначают R^2 . Как мы показали, он должен лежать между 0 и 1 и измерять часть вариации значений y , объясняемую изменениями регрессоров. Он равен нулю, если регрессия есть горизонтальная линия, т. е. все элементы b , кроме константы, равны нулю. В этом случае предсказываемые значения y всегда равны \bar{y} , так что изменения значений x никак не сказываются на предсказываемых значениях y . В этом случае регрессоры x не имеют объясняющей силы. В другом крайнем случае, при $R^2 = 1$, все значения x и y лежат в одной гиперплоскости (на прямой в случае регрессии с двумя переменными) и все остатки равны нулю. Если все наблюдения лежат на одной вертикальной прямой, то R^2 не имеет смысла и не может быть вычислен.

Регрессионный анализ часто используется для прогнозирования. В этом случае нас интересует, с какой точностью модель регрессии предсказывает изменения зависимой переменной. Имея это в виду, полезно иметь еще один способ вычисления R^2 . Запишем

$$b'X'M^0Xb = \hat{y}'M^0\hat{y},$$

но $\hat{y} = Xb$, $y = \hat{y} + e$, $M^0e = e$ и $X'e = 0$, так что $\hat{y}'M^0\hat{y} = \hat{y}'M^0y$. Умножим $R^2 = \hat{y}'M^0\hat{y}/y'M^0y = \hat{y}'M^0y/y'M^0y$ на $1 = \hat{y}'M^0y/\hat{y}'M^0\hat{y}$ и получим

$$R^2 = \frac{[\sum_i (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})]^2}{[\sum_i (y_i - \bar{y})^2][\sum_i (\hat{y}_i - \bar{\hat{y}})^2]}, \quad (3-27)$$

что есть квадрат коэффициента корреляции между наблюдаемыми значениями y и значениями, предсказываемыми регрессией.

Пример 3.2. Качество подгонки функции потребления

Данные, представленные на рис. 2.1, приведены в табл. F2.1. Для этих данных с y в качестве C и x в качестве X мы получаем $\bar{y} = 273,2727$, $\bar{x} = 323,2727$, $S_{yy} = 12618,182$, $S_{xx} = 12300,182$, $S_{xy} = 8423,182$ и $SST = 12618,182$, $b = 8423,182/12300,182 = 0,6848014$, $SSR = b^2 S_{xx} = 5768,2068$ и $SSE = SST - SSR = 6849,975$. Тогда $R^2 = b^2 S_{xx}/SST = 0,457135$. Как видно из рисунка, получается удовлетворительное, но все-таки не очень хорошее приближение этого агрегированного временного ряда. С другой стороны, ясно, что на качество отрицательно влияет то, что модель не учитывает аномальных военных лет. Полученное значение R^2 относится к модели, соответствующей пунктирной линии на рисунке. Если мы просто пропустим наблюдения за 1942–1945 гг. и проведем те же вычисления для оставшихся наблюдений, то полученная линия (жирная на рисунке) будет иметь $R^2 = 0,93697$. Если дополнить наблюдения переменной WAR , равной 1 в 1942–1945 гг. и нулю в остальных случаях, то получим модель, обозначенную двойной линией, а ее R^2 повысится до 0,94639.

Вычисление R^2 можно отразить таблицей дисперсионного анализа или таблицей анализа разброса (analysis of variance), такой как табл. 3.3.

Таблица 3.3. Дисперсионный анализ

| | Исходная запись | Степени свободы | Средний квадрат |
|--------------------------|----------------------|---|--------------------------|
| Регрессия | $b'X'y - n\bar{y}^2$ | $K - 1$ (предполагая наличие постоянного члена) | |
| Остаток | $e'e$ | $n - K$ | s^2 |
| Всего | $y'y - n\bar{y}^2$ | $n - 1$ | $S_{yy}/(n - 1) = s_y^2$ |
| Коэффициент детерминации | | $R^2 = 1 - e'e/(y'y - n\bar{y}^2)$ | |

Пример 3.3. Анализ разброса для уравнения инвестиций

Анализ разброса для уравнения инвестиций из подраздела 3.2.2 показан в табл. 3.4.

Таблица 3.4. Анализ разброса для уравнения инвестиций

| | Исходная запись | Степени свободы | Средний квадрат |
|-----------|--------------------------------------|-----------------|-----------------|
| Регрессия | 0,0159025 | 4 | 0,003976 |
| Остаток | 0,0004508 | 10 | 0,00004508 |
| Всего | 0,016353 | 14 | 0,0011681 |
| | $R^2 = 0,0159025/0,016353 = 0,97245$ | | |

3.5.1. Скорректированный R^2 -квадрат и меры качества подгонки

При использовании R^2 для оценки **точности подгонки** (goodness of fit) возникает ряд проблем. Первая касается числа степеней свободы, использованных при оценке параметров. [См. (3-22) и табл. 3.3.] R^2 никогда не уменьшается при добавлении переменных к уравнению регрессии. Это можно легко понять с помощью уравнения (3-23). Фактически мы сравниваем регрессию y на \mathbf{P} , имеющую сумму квадратов остатков $e'e$, с регрессией y на \mathbf{X} и дополнительную переменную z , которая имеет сумму квадратов остатков $u'u$. Векторы остатков равны $z_* = \mathbf{M}z$ и $y_* = \mathbf{M}y = e$, откуда следует, что $e'e = (y_*'y_*)$. Пусть c — коэффициент при z в удлиненной регрессии. Тогда $c = (z_*'z_*)^{-1}(z_*'y_*)$, и, подставляя это в (3-24), получаем

$$u'u = e'e - \frac{(z_*'y_*)^2}{(z_*'z_*)} = e'e(1 - r_{yz}^{*2}), \quad (3-28)$$

где r_{yz}^* — частная корреляция y и z при заданных \mathbf{X} . Разделим обе части равенства на $y'M^0y$. Из (3-26) $u'u/y'M^0y$ есть $(1 - R_{Xz}^2)$ для регрессии на \mathbf{X} и z , а $e'e/y'M^0y$ есть $(1 - R_X^2)$. Преобразовывая, получаем следующий результат.

Теорема 3.6. *Изменение R^2 при добавлении переменной к регрессии*
 Пусть R_{Xz}^2 есть коэффициент детерминации в регрессии y на \mathbf{X} и дополнительную переменную z , R_X^2 — в регрессии y только на \mathbf{X} , а r_{yz}^* — частная корреляция между y и z при данном \mathbf{X} . Тогда

$$R_{Xz}^2 = R_X^2 + (1 - R_X^2) r_{yz}^{*2}. \quad (3-29)$$

Таким образом, R^2 удлиненной регрессии не может быть меньше. Возникает соблазн воспользоваться этим, включая в модель дополнительные переменные: R^2 будет возрастать, стремясь к 1⁵. **Скорректированный R^2** (скорректированный на число степеней свободы), учитывающий эту возможность, записывается так⁶:

$$\bar{R}^2 = 1 - \frac{e'e/(n - K)}{y'M^0y/(n - 1)}. \quad (3-30)$$

При вычислениях полезна следующая связь между R^2 и \bar{R}^2 :

$$\bar{R}^2 = 1 - \frac{n - 1}{n - K}(1 - R^2).$$

Скорректированный R^2 может убывать при добавлении независимой переменной. В действительности \bar{R}^2 может даже принимать отрицательные значения. Можно рассмотреть экстремальный случай: если выборочная корреляция x и y равна нулю, то скорректированный R^2 будет равен $-1/(n - 2)$.

⁵У этого метода есть недостаток: точность оценки коэффициентов при включении дополнительных переменных уменьшается. Это будет показано в главе 4.

⁶Использование этой меры часто рекомендуют из-за несмещенности обеих величин, входящих в дробь. Впрочем, это отношение не является несмещенной оценкой какого-либо параметра распределения, так что этот довод не очень убедителен.

[В этом контексте название «скорректированный R -squared» представляется не совсем подходящим: из (3-30) можно видеть, что \bar{R}^2 не является квадратом какой-либо величины.] Рост или уменьшение \bar{R}^2 зависят от того, компенсирует ли улучшение качества приближения при добавлении переменной уменьшение коэффициента из-за уменьшения числа степеней свободы. Общий результат (доказательство которого остается в качестве упражнения) выглядит следующим образом.

Теорема 3.7. *Изменение \bar{R}^2 при добавлении переменной к регрессии*
В множественной регрессии \bar{R}^2 будет уменьшаться (или возрастать) при удалении переменной x из регрессии, если квадрат t -статистики для этой переменной больше (меньше) 1.

Мы показали, что R^2 не будет убывать при добавлении переменной к регрессии. Теперь мы попробуем обобщить этот результат. Изменение суммы квадратов остатков при добавлении к регрессии переменных X_2 равно

$$e'_{1,2}e_{1,2} = e'_1e_1 - b'_2X'_2M_1X_2b_2,$$

где нижний индекс 1 обозначает регрессию только на переменные X_1 , а индекс 1,2 обозначает регрессию на все переменные. Вектор коэффициентов b_2 — это коэффициенты при X_2 в множественной регрессии y на X_1 и X_2 . [См. определения b_2 и M_1 в (3-19) и (3-20).] Следовательно,

$$R^2_{1,2} = 1 - \frac{e'_1e_1 - b'_2X'_2M_1X_2b_2}{y'M^0y} = R^2_1 + \frac{b'_2X'_2M_1X_2b_2}{y'M^0y},$$

что больше R^2_1 при ненулевом векторе b_2 . (M_1X_2 не равно нулю, если только X_2 не является линейной функцией X_1 , а в этом случае нельзя было бы построить оценку регрессии на X_1 и X_2 .) Дальнейшее преобразование этого уравнения дает

$$R^2_{1,2} = R^2_1 + \frac{y'M_1y}{y'M^0y} \frac{b'_2X'_2M_1X_2b_2}{y'M_1y}.$$

Но $y'M_1y = e'_1e_1$, так что первый множитель равен $1 - R^2_1$. Вторым множителем равен **множественной корреляции** в регрессии M_1y на M_1X_2 или частной корреляции (учитывающей эффект X_1) в регрессии y на X_2 . Приводя подобные члены, получаем

$$R^2_{1,2} = R^2_1 + (1 - R^2_1)r^2_{y2.1}.$$

[Это аналог формулы (3-29) для случая многих переменных.]

Таким образом, R^2 можно увеличивать до желаемой величины простым добавлением регрессоров. Это свойство является причиной того, чтобы использовать скорректированный R^2 , определяемый формулой (3-30), вместо R^2 при выборе из нескольких моделей. Поскольку \bar{R}^2 включает поправку, отвечающую за уменьшение числа степеней свободы, и при этом все равно реагирует на улучшение качества приближения, можно взять модель с максимальным \bar{R}^2 . Есть мнение, что скорректированный R^2 недостаточно сильно «наказывает» за уменьшение числа степеней свободы⁷. В связи с этим пред-

⁷См., например, работу Amemiya (1985, p. 50–51).

лагалось сравнивать модели (которые мы обозначаем с помощью индекса j), используя, например,

$$\tilde{R}_j^2 = 1 - \frac{n + K_j}{n - K_j} (1 - R_j^2),$$

который минимизирует **предсказательный критерий** (prediction criterion) Амемии или информационные критерии Акаике и Шварца, которые приводятся в (5-43) и (5-44)⁸.

3.5.2. R -квадрат и константа в модели

Вторая сложность с R^2 касается постоянного члена в модели. Доказательство того, что $0 \leq R^2 \leq 1$, предполагает, что матрица \mathbf{X} содержит столбец единиц. Если это не так, то (1) $\mathbf{M}^0 \mathbf{e} \neq \mathbf{e}$, и (2) $\mathbf{e}' \mathbf{M}^0 \mathbf{X} \neq 0$, и член $2\mathbf{e}' \mathbf{M}^0 \mathbf{X} \mathbf{b} + \mathbf{y}' \mathbf{M}^0 \mathbf{y} = (\mathbf{M}^0 \mathbf{X} \mathbf{b} + \mathbf{M}^0 \mathbf{e})' (\mathbf{M}^0 \mathbf{X} \mathbf{b} + \mathbf{M}^0 \mathbf{e})$ не может быть исключен из уравнения, приводящего к (3-25). Значит, при вычислении

$$R^2 = 1 - \frac{\mathbf{e}' \mathbf{e}}{\mathbf{y}' \mathbf{M}^0 \mathbf{y}}$$

будет получен непредсказуемый результат. Он никогда не может быть больше, но может быть намного меньше того значения, которое было бы получено в регрессии с константой. Может даже получиться отрицательное значение. Компьютерные программы по-разному вычисляют R^2 в этом случае. Другой подход к вычислению R^2 ,

$$R^2 = \frac{\mathbf{b}' \mathbf{X}' \mathbf{M}^0 \mathbf{y}}{\mathbf{y}' \mathbf{M}^0 \mathbf{y}},$$

также приводит к проблемам. Полученное значение тоже будет отличаться от значения, полученного в регрессии с константой: здесь может получиться значение R^2 больше 1. Некоторые программы обходят проблему, приводя в этом случае «третий» R^2 : квадрат выборочной корреляции значений y и значений, предсказываемых регрессией. Этот подход может давать обманчивый результат. Если регрессия содержит константу, то, как мы видели, все три способа дают один и тот же результат. Последний способ дает значение между 0 и 1 даже при отсутствии постоянного члена. Однако это значение не равно доле объясненной вариации. С другой стороны, это значение может все же быть полезно при сравнении моделей. При использовании компьютерных программ нужно понимать, как вычисляется R^2 . Некоторые пакеты выдают предупреждение в случаях, когда R^2 вычисляется для регрессий без константы или при оценке методом, отличным от метода наименьших квадратов.

3.5.3. Сравнение моделей

Значение $R^2 = 0,94639$, полученное нами для функции потребления в примере 3.2, кажется высоким. Но является ли оно на самом деле высоким?

⁸Большинство авторов и компьютерных программ приводят логарифмы именно этих критериев.

К сожалению, не существует абсолютных критериев для сравнения. Вообще говоря, при изучении агрегированных данных временных рядов зачастую получают высокие коэффициенты детерминации, подобные этому. Для исследований, использующих межобъектные данные, высокими считаются и значения R^2 около 0,5. Даже коэффициент, равный 0,2, иногда можно считать успехом. Ответ на вопрос о том, качественно ли регрессия приближает данные, зависит от постановки задачи.

Сложно что-либо сказать о сравнительном качестве регрессий в разных контекстах или на разных данных, даже если данные предположительно порождаются одним и тем же процессом. Даже в одной и той же ситуации нужно убедиться, что используется одинаковая основа для сравнения. Так, например, надо выяснить, как именно вычисляются значения зависимой переменной. К примеру, часто возникает вопрос: какая модель лучше описывает данные — линейная или логлинейная? На этот вопрос, к сожалению, нельзя ответить прямым сравнением. R^2 линейной модели отличается от R^2 логлинейной модели. Вариация y отличается от вариации $\ln y$. R^2 логлинейной модели обычно оказывается больше, но это не значит, что логлинейная модель действительно лучше описывает данные.

Нужно подчеркнуть, что R^2 отражает линейную связь между x и y . Например, на рис. 3.3 показаны данные, которые могут быть порождены моделью

$$y_i = \alpha + \beta(x_i - \gamma)^2 + \varepsilon_i.$$

(Константа γ позволяет x иметь среднее, отличное от нуля.) Связь y и x в этой модели является нелинейной, и линейная регрессия не даст качественного приближения.

Упомянем еще одно обстоятельство. R^2 интерпретируется как объясненная доля изменчивости только в том случае, когда при оценке параметров регрессии использовался метод наименьших квадратов. Выражение

$$y_i - \bar{y} = (\hat{y}_i - \bar{y}) + e_i$$

будет всегда верным независимо от того, как были вычислены \hat{y}_i . Значит, можно использовать $\hat{y}_i = \exp(\widehat{\ln y}_i)$, взятую из логлинейной модели для вычисления сумм квадратов в обеих частях. Однако член с перекрестным произведением сокращается, только если при оценке использовался метод наименьших квадратов, а модель содержит константу. Поэтому мы игнорировали этот член при расчете R^2 в логлинейной модели. Только в той ситуации, когда метод наименьших квадратов применяется для оценки линейной регрессии с константой, мы можем интерпретировать R^2 как долю изменчивости y , объясняемую изменчивостью x . Эквивалентное вычисление можно провести, не приводя переменные к отклонениям от среднего, если регрессия не содержит постоянного члена. Однако в регрессиях без константы мы столкнемся с другими трудностями алгебраического характера. Так, значение R^2 будет изменяться при добавлении константы ко всем y , хоть и ясно, что никакого существенного изменения связей между переменными при этом не происходит. Нужно быть крайне осторожным (и даже скептически настроенным) при вычислении и интерпретации мер качества подгонки регрессий, не содержащих постоянного члена.

3.6. Линейные преобразования регрессий

В качестве последнего способа применения инструментария, разработанного в этой главе, рассмотрим еще один чисто алгебраический результат, полезный для понимания особенностей вычисления линейных моделей. Пусть в регрессии y на X столбцы X были подвергнуты некоторому линейному преобразованию. На практике типичным линейным преобразованием может быть, например, изменение используемых единиц измерения: изменение типа валюты, перевод часов в минуты, миль в километры. В примере 3.4 демонстрируется несколько более сложный случай.

Пример 3.4. Оценка произведений искусства

Первая теория, объясняющая аукционные цены картин Моне, утверждает, что цена определяется размерами (шириной W и высотой H) картины:

$$\begin{aligned}\ln P &= \beta_1(1) + \beta_2 \ln W + \beta_3 \ln H + \varepsilon \\ &= \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon.\end{aligned}$$

Вторая теория говорит, что покупатели интересуются скорее площадью и отношением сторон картины:

$$\begin{aligned}\ln P &= \gamma_1(1) + \gamma_2 \ln(WH) + \gamma_3 \ln(W/H) + \varepsilon \\ &= \gamma_1 z_1 + \gamma_2 z_2 + \gamma_3 z_3 + u.\end{aligned}$$

Видно, что $z_1 = x_1$, $z_2 = x_2 + x_3$ и $z_3 = x_2 - x_3$. В матричном виде $Z = XP$, где

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & -1 \end{bmatrix}.$$

Эффект преобразования линейной регрессии y на X по сравнению с регрессией y на Z описывается следующей теоремой.

Теорема 3.8. Преобразованные переменные

В линейной регрессии на $Z = XP$, где P — невырожденная матрица, преобразующая столбцы X , коэффициенты будут равны $P^{-1}b$, где b — вектор коэффициентов линейной регрессии y на X , а величина R^2 сохраняется.

Доказательство. Коэффициенты равны

$$\begin{aligned}d &= (Z'Z)^{-1}Z'y = [(XP)'(XP)]^{-1}(XP)'y = (P'X'XP)^{-1}P'X'y = \\ &= P^{-1}(X'X)^{-1}P^{-1}P'X'y = P^{-1}b.\end{aligned}$$

Вектор остатков равен $u = y - Z(P^{-1}b) = y - XPP^{-1}b = y - Xb = e$. Поскольку остатки совпадают, числитель $1 - R^2$ сохраняется и знаменатель также неизменен. Это завершает доказательство теоремы.

Эта теорема дает полезный практический алгебраический результат. Например, существенно упрощается самый простой пример: замена единиц измерения. Если независимая переменная умножается на константу p , то коэффициент регрессии должен быть умножен на $1/p$. Нет никакой необходимости снова проводить оценку.

3.7. Заключение

В этой главе описывалось чисто алгебраическое упражнение: как приблизить множество точек прямой (гиперплоскостью) так, чтобы минимизировать сумму квадратов расстояний. Мы сначала рассмотрели саму эту задачу применительно к набору из n наблюдений и K переменных. Затем исследовали некоторые свойства решения, включая природу матриц проекции и генераторов остатков, и получили несколько важных алгебраических результатов, касающихся вычисления остатков и сумм их квадратов. Мы также исследовали разницу между простой и множественной регрессиями, определив «частные коэффициенты регрессии» и «частные коэффициенты корреляции». Теорема Фриша–Во–Ловелла является фундаментальным результатом и полезным инструментом, позволяющим получить подвектор вектора коэффициентов регрессии в явном виде. Мы рассмотрели ряд особенностей раздельной регрессии, включая то, как качество приближения регрессионной модели изменяется при добавлении или удалении переменных. В конце мы описали и исследовали часто используемую меру качества подгонки данных регрессией.

Основные термины и понятия

| | |
|--|--|
| Adjusted — Скорректированный R^2 | Least squares normal equations — Нормальные уравнения метода наименьших квадратов |
| Analysis of variance — Дисперсионный анализ | Moment matrix — Матрица моментов |
| Bivariate regression — Парная регрессия | Multiple correlation — Множественная корреляция |
| Coefficient of determination — Коэффициент детерминации | Multiple regression — Множественная регрессия |
| Degrees of Freedom — Степени свободы | Netting out — Выделять эффект |
| Disturbance — Шок, возмущение | Normal equations — Нормальные уравнения |
| Fitting criterion — Критерий подгонки | Orthogonal regression — Ортогональная регрессия |
| Frisch–Waugh thm — Теорема Фриша–Во | Partial correlation coefficient — Коэффициент частной корреляции |
| Goodness of fit — Мера подгонки | |
| Least squares — Метод наименьших квадратов | |

| | |
|---|--|
| Partial regression coefficient — Коэффициент частной регрессии | Population regression — Теоретическая регрессия |
| Partialing out — Выделять эффект | Projection — Проекция |
| Partitioned regression — Частичная регрессия | Projection matrix — Проектор |
| Prediction criterion — Предсказательный критерий | Residual — Остаток |
| Population quantity — Характеристика распределения | Residual maker — Генератор остатков |
| | Total variation — Полная вариация |

Упражнения

1. Регрессия с двумя переменными

Для модели регрессии $y = \alpha + \beta x + \varepsilon$:

- Покажите, что из нормальных уравнений метода наименьших квадратов следует, что $\sum_i e_i = 0$ и $\sum_i x_i e_i = 0$;
- Покажите, что оценка постоянного члена будет равна $a = \bar{y} - b\bar{x}$;
- Покажите, что оценка для b равна $b = [\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})] / [\sum_{i=1}^n (x_i - \bar{x})^2]$;
- Докажите, что при этих значениях параметров достигается единственный минимум суммы квадратов. Для этого покажите, что диагональные элементы матрицы вторых производных суммы квадратов по параметрам положительны, а ее определитель равен $4n[(\sum_{i=1}^n x_i^2) - n\bar{x}^2] = 4n[\sum_{i=1}^n (x_i - \bar{x})^2]$ и положителен всегда, кроме случая, когда все значения x одинаковы.

2. Изменение суммы квадратов

Пусть b — это полученный с помощью метода наименьших квадратов вектор коэффициентов регрессии y на X , а c — другой вектор размером $K \times 1$. Докажите, что разность сумм квадратов остатков равна

$$(y - Xc)'(y - Xc) - (y - Xb)'(y - Xb) = (c - b)'X'X(c - b).$$

Покажите, что эта разность положительна.

3. Частная теорема Фриша-Во

При оценке коэффициентов регрессии y на константу и X мы для вычисления коэффициентов регрессии при X можем сначала преобразовать y , вычтя из каждого значения выборочное среднее, так же преобразовать каждый столбец X , а затем оценить регрессию преобразованных y на преобразованные X (без константы). Получим ли мы такой же результат, если преобразуем только y ? А если только X ?

4. Генераторы остатков

Чему равно произведение матриц $M_1 M$, где M_1 определяется формулой (3-19), а M — формулой (3-14)?

5. Добавление наблюдения

Выборка состоит из n наблюдений X_n и y_n . Оценка методом наименьших квадратов, основанная на этой выборке, равна $b_n = (X_n' X_n)^{-1} X_n' y_n$.

Получено еще одно наблюдение, x_s и y_s . Докажите, что оценка методом наименьших квадратов, вычисленная после включения этого наблюдения, будет равна

$$\mathbf{b}_{n,s} = \mathbf{b}_n + \frac{1}{1 + \mathbf{x}'_s(\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{x}_s} (\mathbf{X}'_n \mathbf{X}_n)^{-1} \mathbf{x}_s (y_s - \mathbf{x}'_s \mathbf{b}_n).$$

Заметьте, что последний множитель равен e_s , остатку при предсказании y_s с помощью оценок коэффициентов, построенных по \mathbf{X}_n и \mathbf{b}_n . Отсюда можно сделать вывод, что новые данные изменяют результат оценки только в том случае, когда новые наблюдения y не в точности равны их предсказываемым значениям.

6. Удаление наблюдения

В случае если в одном из наблюдений отсутствуют значения нескольких независимых переменных, часто поступают следующим образом: пропущенные значения заменяют нулями, а набор переменных дополняют еще одной переменной, равной 1 для этого наблюдения и 0 для остальных. Покажите, что эта «стратегия» дает такую же оценку коэффициентов, как и отбрасывание этого наблюдения, но изменяет R^2 . Рассмотрите частный случай, когда \mathbf{X} состоит из константы и одной переменной. Покажите, что замена пропущенных значений x средними значениями по всей выборке даст такие же результаты, что и добавление новой переменной.

7. Оценка спроса на группу товаров

Обозначим Y полные расходы на потребительские товары повседневного спроса и длительного пользования и услуги, а E_d , E_n и E_s — расходы в каждой отдельной категории. По определению $Y = E_d + E_n + E_s$. Рассмотрим систему уравнений для расходов:

$$\begin{aligned} E_d &= \alpha_d + \beta_d Y + \gamma_{dd} P_d + \gamma_{dn} P_n + \gamma_{ds} P_s + \varepsilon_d, \\ E_n &= \alpha_n + \beta_n Y + \gamma_{nd} P_d + \gamma_{nn} P_n + \gamma_{ns} P_s + \varepsilon_n, \\ E_s &= \alpha_s + \beta_s Y + \gamma_{sd} P_d + \gamma_{sn} P_n + \gamma_{ss} P_s + \varepsilon_s. \end{aligned}$$

Докажите, что если все уравнения оцениваются с помощью метода наименьших квадратов, то сумма коэффициентов при Y будет равна 1, а сумма любого другого столбца коэффициентов будет равна 0.

8. Изменение скорректированного R^2

Докажите, что скорректированный R^2 , задаваемый (3-30), растет (убывает) при удалении переменной x_k из регрессии, если квадрат t -статистики при x_k в множественной регрессии меньше (больше) 1.

9. Регрессия без константы

Пусть множественная регрессия оценивается сначала с константой, а потом без. Ответ на вопрос о том, будет ли во втором случае получен более высокий R^2 , зависит среди прочего от метода вычисления R^2 . Какая регрессия даст более высокое значение, если используется (стандартная) формула $R^2 = 1 - (e'e/y'M^0y)$?

10. Три переменные, N , D и Y , имеют нулевые средние и единичные дисперсии. Четвертая переменная $C = N + D$. В регрессии C на Y коэффициент

наклона равен 0,8, в регрессии C на N — 0,5, в регрессии D на Y — 0,4. Какова сумма квадратов остатков в регрессии C на D ? Всего имеется 21 наблюдение, и при вычислении всех моментов используется делитель $1/(n - 1)$.

11. Используя матрицы сумм квадратов и попарных произведений, приведенные непосредственно перед подразделом 3.2.3, найдите коэффициенты в множественной регрессии реальных инвестиций на константу, реальный ВВП и процентную ставку. Вычислите R^2 .

12. В номере журнала *American Economic Review* за декабрь 1969 г. (с. 886–896) Натаниел Лефф (Nathaniel Leff) приводит следующие результаты оценки методом наименьших квадратов регрессии, связывающей накопления и возраст по данным из 74 стран за 1964 г.:

$$\ln S/Y = 7,3439 + 0,1596 \ln Y/N + 0,0254 \ln G - 1,3520 \ln D_1 - 0,3990 \ln D_2,$$

$$\ln S/N = 2,7851 + 1,1486 \ln Y/N + 0,0265 \ln G - 1,3438 \ln D_1 - 0,3966 \ln D_2,$$

где S/Y — доля накоплений в доходе, S/N — подушевые накопления, Y/N — подушевой доход, D_1 — процент населения моложе 15 лет, D_2 — процент населения старше 64 лет и G — темп роста подушевого дохода. Верны ли эти результаты? Поясните. [См. обсуждение в работах Goldberger (1973) и Leff (1973).]

Эмпирическое приложение

Данные, приведенные в табл. 3.5, взяты из статьи Купа и Тобиаса (Коор, Tobias (2004)), посвященной связи между уровнем дохода и характеристиками образования, способностей и семейного окружения. (См. табл. F3.2.) Их данные представляют собой панель из 2178 человек и 17 919 наблюдений. В таблице показаны данные за первый год (и неизменные во времени переменные) для первых 15 индивидов из выборки. Переменные определены в статье.

Таблица 3.5. Подвыборка данных Купа и Тобиаса

| Человек | Образование | Зарплата | Опыт | Способности | Образование матери | Образование отца | Братья и сестры |
|---------|-------------|----------|------|-------------|--------------------|------------------|-----------------|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 1 | 13 | 1,82 | 1 | 1,00 | 12 | 12 | 1 |
| 2 | 15 | 2,14 | 4 | 1,50 | 12 | 12 | 1 |
| 3 | 10 | 1,56 | 1 | -0,36 | 12 | 12 | 1 |
| 4 | 12 | 1,85 | 1 | 0,26 | 12 | 10 | 4 |
| 5 | 15 | 2,41 | 2 | 0,30 | 12 | 12 | 1 |
| 6 | 15 | 1,83 | 2 | 0,44 | 12 | 16 | 2 |
| 7 | 15 | 1,78 | 3 | 0,91 | 12 | 12 | 1 |
| 8 | 13 | 2,12 | 4 | 0,51 | 12 | 15 | 2 |
| 9 | 13 | 1,95 | 2 | 0,86 | 12 | 12 | 2 |
| 10 | 11 | 2,19 | 5 | 0,26 | 12 | 12 | 2 |
| 11 | 12 | 2,44 | 1 | 1,82 | 16 | 17 | 2 |
| 12 | 13 | 2,41 | 4 | -1,30 | 13 | 12 | 5 |

Окончание табл. 3.5

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|----|----|------|---|-------|----|----|---|
| 13 | 12 | 2,07 | 3 | -0,63 | 12 | 12 | 4 |
| 14 | 12 | 2,20 | 6 | -0,36 | 10 | 12 | 2 |
| 15 | 12 | 2,12 | 3 | 0,28 | 10 | 12 | 3 |

Пусть X_1 — константа, образование, опыт и способности (собственные свойства наблюдаемого). Пусть X_2 содержит данные об образовании матери и отца и числе братьев и сестер (свойства семьи). Пусть y — заработная плата.

- Найдите оценки метода наименьших квадратов коэффициентов регрессии y на X_1 .
- Найдите оценки метода наименьших квадратов коэффициентов регрессии y на X_1 и X_2 .
- Оцените регрессии каждой из переменных X_2 на все переменные X_1 и вычислите остатки этих регрессий. Обозначьте матрицу остатков этих регрессий X_2^* . Каковы их средние значения? Объясните матрицу.
- С помощью (3-26) вычислите R^2 для регрессии y на X_1 и X_2 . Повторите вычисления в случае, когда константа не входит в X_1 . Как изменится R^2 ?
- Найдите скорректированный R^2 для полной регрессии с константой. Интерпретируйте полученный результат.
- Возвращаясь к результату пункта с: оцените регрессию y на X_1 и X_2^* . Как отличаются полученные результаты от результатов регрессии y на X_1 и X_2 ? Здесь требуется сравнить оценки метода наименьших квадратов регрессии y на X_1 и $M_1 X_2$ и регрессии y на X_1 и X_2 . Выведите этот результат аналитически. (Полученные численные значения, разумеется, должны соответствовать аналитическому результату.)

ГЛАВА 4

Оценки методом наименьших квадратов

4.1. Введение

В предыдущей главе линейное приближение данных методом наименьших квадратов было представлено как чисто алгебраическое упражнение. В этой главе мы подробно рассмотрим свойства этого метода как способа **оценивания** параметров модели линейной регрессии (табл. 4.1). Мы начнем раздел 4.2 с вопроса, ранее поднятого в сноске 1 в главе 3: зачем вообще нужно пользоваться методом наименьших квадратов? Затем перейдем к изучению свойств этой оценки. Существуют другие способы получения оценки для β . Например, можно было бы взять значения коэффициентов, минимизирующие сумму модулей остатков. Ответ на вопрос выбора метода оценивания должен основываться на таких **статистических свойствах** оценок, как несмещенность, состоятельность, эффективность, а также на свойствах их выборочных распределений. В разделе 4.3 рассматриваются свойства **оценок в конечных выборках**, такие как несмещенность. Конечно, выборочные свойства оценок методом наименьших квадратов не зависят от размера выборки. Линейная модель — одна из немногих, в которых можно сформулировать утверждения о свойствах оценок, применимые ко всем конечным выборкам. Для большинства моделей можно говорить лишь о свойствах достаточно больших выборок. В этом случае приходится считать конечные выборки приближением к большим выборкам и пользоваться их свойствами. Поэтому раздел 4.4 будет посвящен изучению свойств оценок регрессионных моделей методом наименьших квадратов на больших выборках, или их **асимптотических свойств**¹.

Изучая свойства оценки, мы обычно в первую очередь интересуемся **точечными оценками**, т. е. тем, как максимально эффективно использовать информацию, содержащуюся в выборке, для получения наилучшей оценки параметров модели. **Интервальное оценивание**, которым мы займемся в разделе 4.5, позволяет находить оценки, явно учитывающие неточность, всегда присутствующую при использовании случайных выборок для определения свойств генеральной совокупности. В разделе 4.5 мы также рассмотрим некоторые приложения интервальных оценок параметров распределений и функций этих параметров. Одно из наиболее широко используемых приложений интервальных оценок — использование модели для предсказания значений зависимой переменной и указание неточностей этих прогнозов. В разделе 4.6 рассматриваются прогнозы, использующие оцененную регрессионную модель.

¹ При этом мы будем пользоваться результатами, описывающими асимптотические распределения. Читателю рекомендуется ознакомиться с приложением D перед началом изучения этого материала.

При исследовании предполагается, что имеющиеся данные удовлетворяют предположениям модели. В разделе 4.7 мы рассмотрим ряд проблем, возникающих при изучении неэкспериментальных данных. Предположение А2 о полноте ранга X обычно считается данностью. В подразделе 2.3.2 мы видели, что при нарушении этого условия построение оценки не представляется возможным ни при каком размере выборки. **Мультиколлинеарность**, или близость реальных данных к нарушению этого условия, рассматривается в подразделах 4.7.1–4.7.3. Пропущенные данные могут существенно помешать изучению выборки. Благоприятный случай, когда пропущены случайные наблюдения, рассматривается в подразделе 4.7.4. Более сложный случай неслучайных пропусков будет рассмотрен в главе 18. Наконец, проблема ошибок в наблюдениях описана в подразделе 4.7.5.

Таблица 4.1. Предположения классической модели линейной регрессии

-
- А1. Линейность:** $y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \dots + x_{iK}\beta_K + \varepsilon_i$.
- А2. Полный ранг:** Матрица данных размером $n \times K$, X , имеет полный ранг по столбцам.
- А3. Экзогенность независимых переменных:** $E[\varepsilon_i | x_{j1}, x_{j2}, \dots, x_{jK}] = 0$, $i, j = 1, \dots, n$. Корреляция между шоками и независимыми переменными отсутствует.
- А4. Гомоскедастичность и отсутствие автокорреляции:** Все шоки ε_i имеют одинаковую дисперсию, σ^2 , и не коррелируют между собой условно по X .
- А5. Случайные или неслучайные данные:** $(x_{i1}, x_{i2}, \dots, x_{iK})$ $i = 1, \dots, n$.
- А6. Нормальное распределение шоков:** Шоки ε_i распределены нормально.
-

4.2. Почему именно метод наименьших квадратов?

Одна из причин широкого применения метода наименьших квадратов — это простота необходимых вычислений. Впрочем, есть и другие причины для использования этого метода. Во-первых, метод наименьших квадратов реализует естественный подход к оцениванию, явно используя все условия модели. Во-вторых, метод наименьших квадратов дает оптимальный линейный предиктор зависимой переменной, даже если истинное распределение зависимой переменной не описывается линейной моделью. Таким образом, метод наименьших квадратов обладает некоторой устойчивостью, которой лишены другие методы. В-третьих, при некоторых специальных условиях метод наименьших квадратов наиболее оптимально в некотором разумном смысле использует имеющиеся данные. Мы рассмотрим все эти соображения по очереди.

4.2.1. Условия ортогональности генеральной совокупности

Пусть x — вектор независимых переменных в регрессионной модели, причем в соответствии с предположением А5 данные могут быть как стохастическими, так и нестохастическими. Предположение А3 гласит, что распределение ошибок стохастически ортогонально независимым переменным,

т.е. $E[\varepsilon | \mathbf{x}] = 0$. Следовательно, $\text{Cov}[\mathbf{x}, \varepsilon] = \mathbf{0}$. Поскольку (по закону повторного математического ожидания — теорема В.1) $E_{\mathbf{x}}\{E[\varepsilon | \mathbf{x}]\} = E[\varepsilon] = 0$, то можно записать это в виде

$$E_{\mathbf{x}}E_{\varepsilon}[\mathbf{x}\varepsilon] = E_{\mathbf{x}}E_y[\mathbf{x}(y - \mathbf{x}'\beta)] = \mathbf{0}$$

или

$$E_{\mathbf{x}}E_y[\mathbf{x}y] = E_{\mathbf{x}}[\mathbf{x}\mathbf{x}']\beta. \quad (4-1)$$

(Поскольку правая часть не является функцией y , можно брать математическое ожидание только по \mathbf{x} .) Возьмем обычное уравнение для оценки методом наименьших квадратов: $\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}\mathbf{b}$. Поделим его на n и перепишем в виде суммы, получая

$$\left(\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i y_i\right) = \left(\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i'\right) \mathbf{b}. \quad (4-2)$$

Уравнение (4-1) описывает распределение, а (4-2) — его выборочный аналог. Предположим, что выполняются требования законов больших чисел (см. приложение D). Тогда суммы в левой и правой частях (4-2) являются оценками соответствующих частей (4-1). Таким образом, метод наименьших квадратов основывается на переносе теоретических соотношений между переменными на имеющуюся выборку. Мы вернемся к этому подходу построения оценок в главах 12 и 13 при рассмотрении оценок GMM.

4.2.2. Предиктор, минимизирующий среднеквадратичную ошибку

Разберем другой подход. Попробуем найти **оптимальный линейный предиктор** для y . Опустим предположения А6 и А1 (т. е. распределение не обязательно описывается линейной моделью). В качестве критерия оптимальности возьмем минимальность среднего квадрата ошибки и найдем предиктор y , минимизирующий среднеквадратичную ошибку, который обозначим $\mathbf{x}'\gamma$. Ожидаемый квадрат ошибки этого предиктора:

$$\text{MSE} = E_y E_{\mathbf{x}} [y - \mathbf{x}'\gamma]^2.$$

Это можно переписать в виде

$$\text{MSE} = E_{y,\mathbf{x}} \{y - E[y | \mathbf{x}]\}^2 + E_{y,\mathbf{x}} \{E[y | \mathbf{x}] - \mathbf{x}'\gamma\}^2.$$

Мы хотим найти γ , минимизирующее эту величину. Первый член не является функцией γ , так что нужно минимизировать только второй член. Заметим, что он не является функцией y , а значит, внешнее математическое ожидание можно было бы опустить. Однако оно нам скоро понадобится, поэтому пока оставим его. Необходимое условие представляет собой

$$\begin{aligned} \frac{\partial E_y E_{\mathbf{x}} \{[E(y | \mathbf{x}) - \mathbf{x}'\gamma]^2\}}{\partial \gamma} &= E_y E_{\mathbf{x}} \left\{ \frac{\partial [E(y | \mathbf{x}) - \mathbf{x}'\gamma]^2}{\partial \gamma} \right\} = \\ &= -2E_y E_{\mathbf{x}} \{ \mathbf{x} [E(y | \mathbf{x}) - \mathbf{x}'\gamma] \} = \mathbf{0}. \end{aligned}$$

Заметим, что в среднем преобразовании мы поменяли местами операции дифференцирования и взятия математического ожидания, поскольку область интегрирования не является функцией от γ . Наконец, мы приходим к эквивалентному условию:

$$E_y E_x [x E(y | x)] = E_y E_x [x x'] \gamma.$$

Левая часть представляет собой

$$\begin{aligned} E_x E_y [x E(y | x)] &= \text{Cov}[x, -E(y | x)] + \\ &+ E[x] E_x [E(y | x)] = \text{Cov}[x, y] + E[x] E[y] = E_x E_y [xy]. \end{aligned}$$

(Здесь мы воспользовались теоремой В.2.) Итак, необходимое условие для предиктора, минимизирующего среднее квадратичное отклонение, имеет вид

$$E_x E_y [xy] = E_x E_y [x x'] \gamma, \quad (4-3)$$

что совпадает с (4-1) и вновь приводит нас к методу наименьших квадратов. Предполагая, что эти математические ожидания существуют, их можно оценить суммами из (4-2); таким образом, независимо от вида функции условного среднего оценка методом наименьших квадратов дает линейный предиктор с наименьшим среднее квадратичным отклонением. Нам нужно еще установить необходимые предпосылки теоремы, но сформулировать ее удобно сейчас:

Теорема 4.1. *Предиктор, минимизирующий среднее квадратичное отклонение*

Если процесс, генерирующий выборку $(x_i, y_i)_{i=1, \dots, n}$, таков, что законы больших чисел применимы к оценкам (4-2) матриц из (4-1), то линейный предиктор y_i , минимизирующий среднее квадратичное отклонение, получается при помощи оценки линейной регрессии методом наименьших квадратов.

4.2.3. Линейная несмещенная оценка с минимальной дисперсией

Наконец, рассмотрим задачу нахождения **линейной несмещенной оценки**. Если мы будем искать оценку из этого класса, имеющую наименьшую дисперсию, то снова получим оценку OLS. Это предположение будет доказано в подразделе 4.3.5.

Этот факт не означает, что ни одна другая оценка не может быть предпочтительнее оценки методом наименьших квадратов. Мы рассматривали лишь линейные оценки. Сформулированный результат не учитывает, что могут рассматриваться допустимо малосмещенные оценки. Наконец, предположения модели могут быть неверны. Если А5 и А6 не столь принципиальны, то нарушение любого из первых четырех условий может сделать метод наименьших квадратов гораздо менее привлекательным методом оценивания, чем может показаться на данном этапе.

4.3. Свойства метода наименьших квадратов на конечных выборках

Оценка — это правило или формула, использующая выборочные данные, взятые из некоторой генеральной совокупности. Свойства оценки — это описание того, как эта оценка будет вести себя, будучи примененной к конкретному набору данных. Например, «несмещенность» означает, что в среднем оценка (правило) будет верно приближать оцениваемый параметр — она не будет систематически завышать или занижать его. Кажется совершенно неочевидным, как можно сделать подобный вывод, основываясь лишь на одной выборке из генеральной совокупности и ее дальнейшем анализе. Объяснение, традиционно даваемое в классической эконометрике, использует свойства самого правила оценивания. За этим лежит следующий мысленный эксперимент. Эконометрист представляет себе множество «повторных выборок» из генеральной совокупности и пытается описать поведение оценки на такой «выборке выборок». Соответствующая статистическая теория этой оценки является при этом основным инструментом. Проиллюстрируем это на примере 4.1.

Пример 4.1. Выборочное распределение оценки методом наименьших квадратов

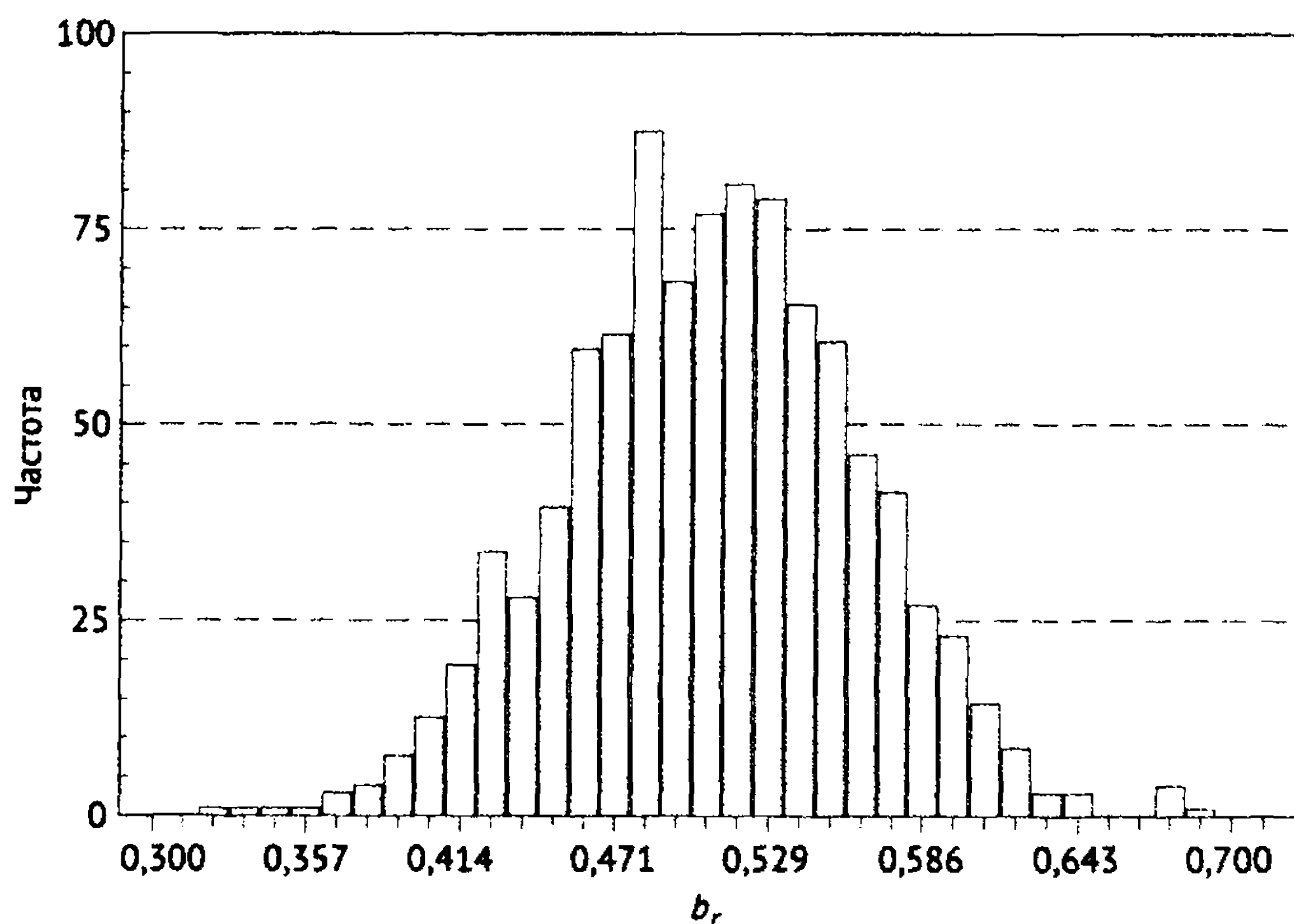


Рис. 4.1. Гистограмма для выборочных оценок коэффициентов наклона, полученных методом наименьших квадратов

Следующий эксперимент иллюстрирует природу выборочного распределения и как следствие несмещенности. Представим, что мы извлекли две выборки размером 10 000 переменных w_i и x_i из стандартного нормального распределения (среднее 0, дисперсия 1). Сгенерируем набор $\varepsilon_i = 0,5w_i$ и $y_i = 0,5 + 0,5x_i + \varepsilon_i$. Будем считать это нашей генеральной совокупностью. Из него мы извлечем 1000 случайных выборок по 100 наблюдений (y_i, x_i) и для каждой из них построим оценку коэффициента

наклона методом наименьших квадратов, где r — номер выборки, по формуле

$$b_r = [\sum_{j=1}^{100} (x_{ir} - \bar{x}_r) y_{ir}] / [\sum_{j=1}^{100} (x_{ir} - \bar{x}_r)^2].$$

Гистограмма на рис. 4.1 показывает результат такого эксперимента.

Обратите внимание, как близко среднее этого распределения к «истинному» значению 0,5, а также на то, что это распределение имеет значительную дисперсию, что отражает тот факт, что оценка коэффициента, как и любая другая статистика, вычисленная по выборке, есть случайная величина. Понятие несмещенности относится к средним значениям распределения оценок, полученных множеством последовательных выборок. Форма гистограммы также наводит на мысль, что оценка распределена нормально. Мы докажем этот факт в подразделе 4.3.8. (Результаты этого эксперимента могут быть воспроизведены с помощью любой программы построения регрессии, имеющей генератор случайных чисел и умеющей подсчитывать среднее случайной выборки из генеральной совокупности.)

4.3.1. Несмещенность оценки

Оценка метода наименьших квадратов является несмещенной на каждой выборке. Чтобы это доказать, запишем

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}. \quad (4-4)$$

Теперь возьмем условные математические ожидания относительно \mathbf{X} :

$$E[\mathbf{b} | \mathbf{X}] = \boldsymbol{\beta} + E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} | \mathbf{X}].$$

По предположению А3 второй член равен 0, поэтому

$$E[\mathbf{b} | \mathbf{X}] = \boldsymbol{\beta}. \quad (4-5)$$

Таким образом,

$$E[\mathbf{b}] = E_{\mathbf{X}}\{E[\mathbf{b} | \mathbf{X}]\} = E_{\mathbf{X}}[\boldsymbol{\beta}] = \boldsymbol{\beta}. \quad (4-6)$$

Этот результат можно интерпретировать так: для каждого набора наблюдений \mathbf{X} оценка методом наименьших квадратов имеет математическое ожидание, равное $\boldsymbol{\beta}$. Усредняя это по всем возможным \mathbf{X} , получаем, что безусловное математическое ожидание также равно $\boldsymbol{\beta}$.

Вы могли заметить, что в этом разделе весь анализ проводился условно относительно \mathbf{X} , т. е. по всей выборке, а в разделе 4.2 мы рассматривали условные распределения y_i по x_i . (Внимательные читатели также могли заметить, что в табл. 4.1 в предположении А3 мы обуславливали $E[\varepsilon_i | \cdot]$ по x_j , т. е. по всем i и j , по \mathbf{X} , а не только по x_i .) В разделе 4.2 мы предлагали рассматривать оценки методом наименьших квадратов в контексте совместного распределения случайной величины y и случайного вектора \mathbf{x} . Это справедливо, если данные представляют собой межобъектную выборку независимых наблюдений. В этом случае, как мы показали в подразделе 4.2.2, оценка методом наименьших квадратов является выборочным ана-

логом вектора γ — коэффициента наклона предиктора, минимизирующего среднеквадратичную ошибку, который, в свою очередь, является свойством распределения. В разделе 4.3 мы переходим к изучению самого процесса, генерирующего наблюдаемую выборку. Утверждение $E[\mathbf{b}|\mathbf{X}] = \beta$ лучше всего понимать в байесовском смысле: если наблюдался набор данных \mathbf{X} , то мы можем ожидать определенного поведения некоторой вычисляемой нами статистики, например коэффициент наклона, вычисленного методом наименьших квадратов, \mathbf{b} . Остаток этой главы, да и всей книги будет в большой степени посвящен изучению поведения статистики в следующем смысле: можем ли мы делать выводы из значений статистики, вычисленных для одной выборки, о значениях той же статистики, вычисленных на выборках, полученных таким же образом из того же распределения? Или же можно по значению статистики что-то узнать о свойствах распределения в целом? То есть можно воспринимать условие в $E[\mathbf{b}|\mathbf{X}]$ в обоих смыслах: и с чисто статистической точки зрения — как вывод относительно свойств оценки, и с точки зрения методологической — что можно узнать о распределении из конкретной конечной выборки.

4.3.2. Смещение, вызванное пропущенными значимыми переменными

Весь наш анализ в предыдущих разделах основывался на том, что известно верное уравнение регрессионной модели

$$\mathbf{y} = \mathbf{X}\beta + \epsilon. \quad (4-7)$$

Однако возможны различные **ошибки спецификации модели**, которые можно допустить при построении модели регрессии. К наиболее распространенным относятся **пропуск значимых переменных** и **включение излишних переменных**.

Предположим, что на самом деле уравнение регрессии имеет вид

$$\mathbf{y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \epsilon, \quad (4-8)$$

где две части \mathbf{X} имеют соответственно K_1 и K_2 столбцов. Если мы оцениваем регрессию \mathbf{y} на \mathbf{X}_1 , не включая \mathbf{X}_2 , то получим оценку

$$\mathbf{b}_1 = (\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1\mathbf{y} = \beta_1 + (\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1\mathbf{X}_2\beta_2 + (\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1\epsilon. \quad (4-9)$$

При взятии математического ожидания мы видим, что если не выполняется одно из соотношений, $\mathbf{X}'_1\mathbf{X}_2 = \mathbf{0}$ или $\beta_2 = \mathbf{0}$, то оценка \mathbf{b}_1 является смещенной. Широко известен следующий результат, называемый **формулой смещения из-за пропущенных переменных** (omitted variables formula):

$$E[\mathbf{b}_1 | \mathbf{X}] = \beta_1 + \mathbf{P}_{1.2}\beta_2, \quad (4-10)$$

где

$$\mathbf{P}_{1.2} = (\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1\mathbf{X}_2. \quad (4-11)$$

Каждый столбец матрицы $\mathbf{P}_{1.2}$ размером $K_1 \times K_2$ представляет собой столбец коэффициентов наклона, полученных методом наименьших квадратов регрессии соответствующего столбца \mathbf{X}_2 на столбцы \mathbf{X}_1 .

Пример 4.2. Пропущенные переменные

Если уравнение спроса оценивается без включения значимой переменной, отражающей доходы потребителей (*Income*), то (4-10) показывает смещение оценки эластичности цены. В качестве убедительного примера рассмотрим данные о ценах на бензин, которые мы использовали в примере 2.3. Пусть b — оценка. Имеем

$$E[b|price, income] = \beta + \frac{\text{Cov}[price, income]}{\text{Var}[price]} \gamma,$$

где γ — коэффициент при доходе.

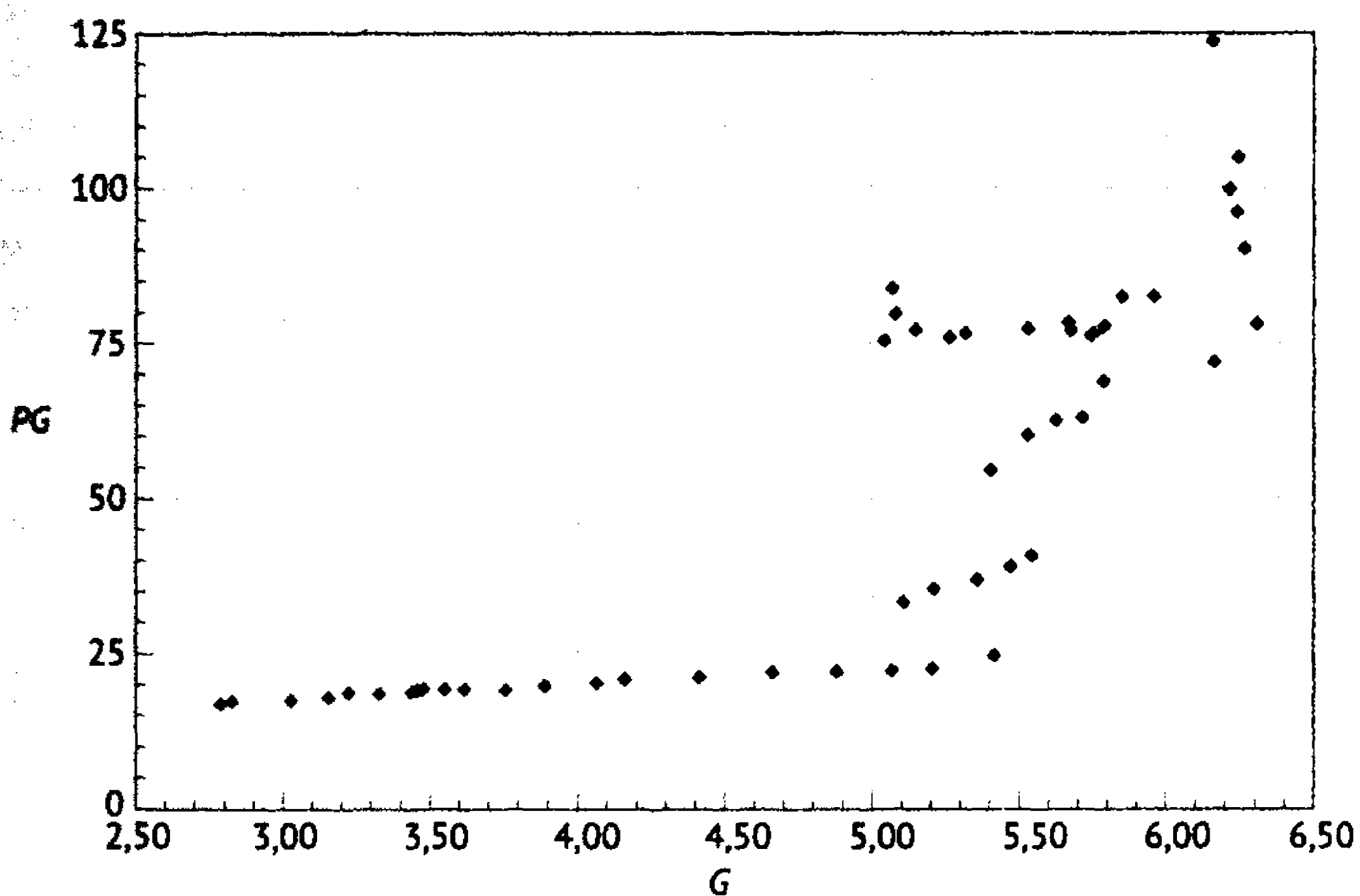


Рис. 4.2. Потребление бензина на душу населения в зависимости от цены, 1953–2004 гг.

По агрегированным данным нельзя сказать, будет ли γ положительным или отрицательным. Знак смещения b будет совпадать со знаком ковариации, поскольку $\text{Var}[price]$ и γ положительны для нормальных благ, каким является бензин. На рис. 4.2 изображен график потребления бензина на душу населения G/Pop в зависимости от индекса цен PG . График существенно отличается от ожидаемого. Однако если посмотреть непосредственно на данные (табл. F2.2 в приложении), то причины этого становятся яснее: при постоянных доходе $Income/Pop$ и ценах других товаров результаты были бы другими. Однако доходы растут и простые корреляции между G/Pop и $Income/Pop$ и между PG и $Income/Pop$ довольно велики — 0,938 и 0,934 соответственно. Чтобы увидеть, существует ли на самом деле ожидаемая связь между ценой и уровнем потребления, нам понадобится очистить наши данные от эффектов $Income/Pop$. Чтобы это сделать, мы воспользуемся результатом Фриша–Во из теоремы 3.2. В простой регрессии логарифма потребления бензина на душу населения на константу и логарифм индекса цен получается коэффициент 0,29904, у которого, как мы, вообще говоря, и ожидали, «неправильный» знак.

В регрессии логарифма потребления бензина на душу населения на константу, логарифм цены и логарифм дохода на душу населения оценка эластичности потребления относительно цены, $\hat{\beta}$, равна $-0,16949$, а оценка эластичности относительно дохода, $\hat{\gamma}$, равна $0,96595$. Здесь знаки коэффициентов соответствуют ожиданиям. Результат также соответствует известному наблюдению, что основным фактором, влиявшим на потребление бензина в это время (1953–2004), было не изменение цены, а рост доходов (выпуска).

В этом примере использовались всего одна переменная, включенная в регрессию, и всего одна пропущенная переменная. В таких случаях обычно легко определить знак смещения. Важно понимать, что если включено более одной переменной, то в формулу пропущенных переменных входит несколько коэффициентов регрессии, знаки которых соответствуют частным, а не простым корреляциям. Например, если бы в уравнение спроса из предыдущего примера входила цена похожего продукта, то по простой корреляции цены и дохода нельзя было бы определить знак смещения оценки эластичности. Потребовалось бы получить знак корреляции цены и дохода, очищенных от влияния этой цены. Это может оказаться непросто, и сложность задачи возрастает при росте числа переменных.

4.3.3. Включение лишних переменных

Если верным уравнением регрессии является

$$y = \mathbf{X}_1\beta_1 + \varepsilon, \quad (4-12)$$

а мы проводим оценивание так, будто верно (4-8) (т. е. включаем дополнительные переменные), то можно было бы предположить, что возникнут похожие проблемы. На самом деле это не так. Пропуск значимых переменных можно рассматривать как введение неверных условий на (4-8). Например, пропуск \mathbf{X}_2 эквивалентен оценке (4-8) при неверном условии $\beta_2 = 0$. Введение ложного ограничения дает смещенную оценку. Другая точка зрения состоит в том, что мы используем при оценивании неверную информацию. Предположим, однако, что наша ошибка заключается в том, что мы не использовали часть информации, которая является *верной*.

Включение в регрессию лишних переменных \mathbf{X}_2 эквивалентно тому, чтобы забыть наложить ограничение $\beta_2 = 0$ в (4-8) при оценивании. Но (4-8) не является неверным, оно лишь не включает условие $\beta_2 = 0$. Таким образом, не нужно отдельно доказывать, что OLS-оценка β в (4-8) не смещена при включении дополнительного условия; на самом деле мы это уже доказали. На основе наших предшествующих результатов можно заключить, что

$$E[\mathbf{b} | \mathbf{X}] = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \beta_1 \\ 0 \end{bmatrix}. \quad (4-13)$$

Так в чем же тогда заключается проблема? Казалось бы, нужно всегда включать в модель как можно больше переменных. С теоретической точки зрения проблема здесь заключается в том, что отказ от использования инфор-

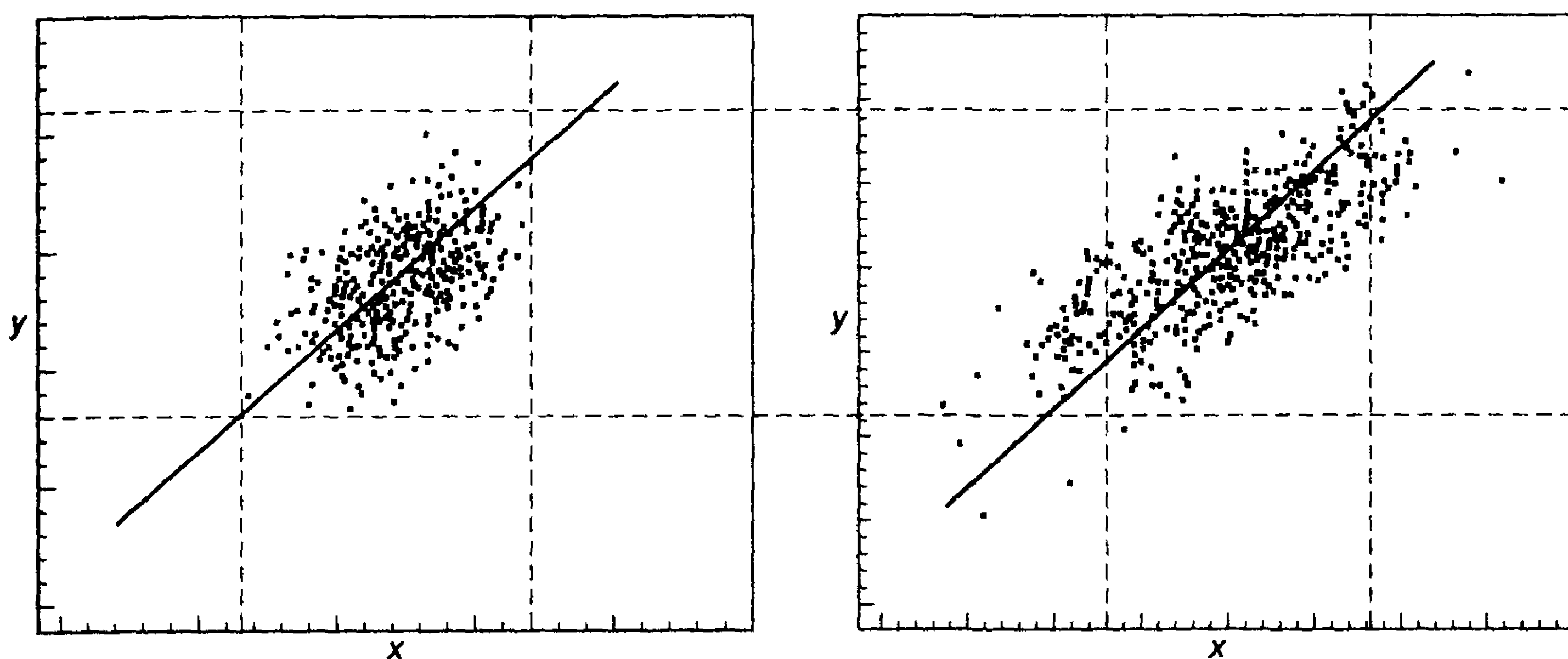


Рис. 4.3. Влияние увеличения дисперсии x при неизменных условной и безусловной дисперсиях y

мации имеет свою цену. В данном случае этой ценой оказывается уменьшение точности оценки. В подразделе 4.7.1 мы покажем, что матрица ковариаций укороченной регрессии (без X_2) никогда не будет больше матрицы ковариаций оценки, получаемой при включении лишних переменных². Рассмотрим пример с одной переменной: если x_2 сильно коррелирована с x_1 , то ошибочное включение x_2 в регрессию значительно увеличит дисперсию оценки β_1 .

4.3.4. Дисперсия оценки наименьших квадратов

Если рассматривать регрессоры как неслучайные величины, например, когда исследователь сам выбирает значения в X , то **выборочную дисперсию** оценки наименьших квадратов можно получить, обращаясь с X , как с матрицей констант. В противном случае мы можем рассматривать X как случайную величину, провести анализ условно относительно наблюдаемых X , а затем усреднить по X , как мы делали при выводе (4-6) из (4-5). С помощью (4-4) можно получить

$$b = (X'X)^{-1}X'(X\beta + \epsilon) = \beta + (X'X)^{-1}X'\epsilon. \quad (4-14)$$

Поскольку $b = \beta + A\epsilon$, где $A = (X'X)^{-1}X'$, b является линейной функцией от шоков, что мы будем называть **линейной оценкой**. Как мы уже видели, математическое ожидание второго слагаемого в (4-14) равно 0. Следовательно, независимо от распределения ϵ при условии выполнения остальных сделанных нами предположений b является линейной несмещенной оценкой β . По предположению $A4 \text{Var}[\epsilon|X] = \sigma^2 I$. Значит, матрица условных ковариаций вектора оценки коэффициентов методом наименьших квадратов равна

²Потери в точности не будет, если $X_1'X_2 = 0$, в этом случае X_2 не содержит никакой информации об X_1 . На практике такое случается крайне редко.

$$\begin{aligned}
\text{Var}[\mathbf{b} | \mathbf{X}] &= E[(\mathbf{b} - \beta)(\mathbf{b} - \beta)' | \mathbf{X}] \\
&= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon\varepsilon'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} | \mathbf{X}] \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\varepsilon\varepsilon' | \mathbf{X}]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\sigma^2\mathbf{I})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}.
\end{aligned} \tag{4-15}$$

Пример 4.3. Выборочная дисперсия оценки в модели парной регрессии. Предположим, что \mathbf{X} содержит только константу (столбец из одних единиц) и единственный регрессор x . Нижний правый элемент матрицы $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ равен

$$\text{Var}[b | \mathbf{x}] = \text{Var}[b - \beta | \mathbf{x}] = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Обратим особенное внимание на знаменатель дисперсии b . Чем больше дисперсия x , тем меньше дисперсия b . Например, рассмотрим задачу оценки коэффициентов наклона регрессий на рис. 4.3. Более точный результат будет получен для данных правой части рисунка.

4.3.5. Теорема Гаусса–Маркова

Получим общий результат для класса линейных несмещенных оценок β .

Теорема 4.2. Теорема Гаусса–Маркова

В модели линейной регрессии с матрицей регрессоров \mathbf{X} оценка методом наименьших квадратов $\hat{\mathbf{b}}$ имеет наименьшую дисперсию среди всех линейных несмещенных оценок параметра β . Для любого вектора констант \mathbf{w} линейной несмещенной оценкой с наименьшей дисперсией параметра $\mathbf{w}'\beta$ является $\mathbf{w}'\hat{\mathbf{b}}$.

Заметим, что доказательство теоремы не использует предположение А6 (нормальности распределения шоков). Необходимы только предположения А1–А4. Прямой подход к доказательству этой теоремы — это определить класс линейных несмещенных оценок ($\mathbf{b}_L = \mathbf{C}\mathbf{y}$, таких что $E[\mathbf{b}_L | \mathbf{X}] = \beta$) и потом найти в этом классе оценку с наименьшей дисперсией. Мы пойдем другим путем. Нам уже известно, что $\hat{\mathbf{b}}$ — линейная несмещенная оценка. Мы рассмотрим другие линейные несмещенные оценки и покажем, что любая другая линейная несмещенная оценка β имеет большую дисперсию.

Пусть $\mathbf{b}_0 = \mathbf{C}\mathbf{y}$ — линейная несмещенная оценка β , где \mathbf{C} — матрица. Поскольку \mathbf{b}_0 является несмещенной, то

$$E[\mathbf{C}\mathbf{y} | \mathbf{X}] = E[(\mathbf{C}\mathbf{X}\beta + \mathbf{C}\varepsilon) | \mathbf{X}] = \beta,$$

откуда следует, что $\mathbf{C}\mathbf{X} = \mathbf{I}$. Существует множество вариантов: например, возьмем первые K (или любые K) линейно независимых строк матрицы \mathbf{X} .

Тогда $\mathbf{C} = [\mathbf{X}_0^{-1} : \mathbf{0}]$, где \mathbf{X}_0^{-1} — матрица, обратная к матрице, образованной из K строк матрицы \mathbf{X} . Матрицу ковариаций вектора \mathbf{b}_0 можно получить, заменив $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ на \mathbf{C} в (4-14); в результате получим $\text{Var}[\mathbf{b}_0 | \mathbf{X}] = \sigma^2\mathbf{C}\mathbf{C}'$. Теперь пусть $\mathbf{D} = \mathbf{C} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, так что $\mathbf{D}\mathbf{y} = \mathbf{b}_0 - \mathbf{b}$. Тогда

$$\text{Var}[\mathbf{b}_0 | \mathbf{X}] = \sigma^2[(\mathbf{D} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')(\mathbf{D} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')'].$$

Мы знаем, что $\mathbf{C}\mathbf{X} = \mathbf{I} = \mathbf{D}\mathbf{X} + (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})$, так что $\mathbf{D}\mathbf{X}$ должно быть равно $\mathbf{0}$. Значит,

$$\text{Var}[\mathbf{b}_0 | \mathbf{X}] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} + \sigma^2\mathbf{D}\mathbf{D}' = \text{Var}[\mathbf{b} | \mathbf{X}] + \sigma^2\mathbf{D}\mathbf{D}'.$$

Поскольку квадратичная форма, соответствующая матрице $\mathbf{D}\mathbf{D}'$, равна $\mathbf{q}'\mathbf{D}\mathbf{D}'\mathbf{q} = \mathbf{z}'\mathbf{z} \geq 0$, условная матрица ковариаций \mathbf{b}_0 равна условной матрице ковариаций \mathbf{b} плюс некоторая неотрицательно определенная матрица. Значит, каждая квадратичная форма, соответствующая матрице $\text{Var}[\mathbf{b}_0 | \mathbf{X}]$, больше, чем квадратичная форма, соответствующая матрице $\text{Var}[\mathbf{b} | \mathbf{X}]$, откуда получаем первую часть теоремы.

Доказательство второй части теоремы следует из предшествующего рассуждения: дисперсия $\mathbf{w}'\mathbf{b}$ есть квадратичная форма, соответствующая матрице $\text{Var}[\mathbf{b} | \mathbf{X}]$. Аналогично предыдущему, рассматривая произвольный вектор \mathbf{b}_0 , можно получить, что каждый b_k является линейной несмещенной оценкой с наименьшей дисперсией для β_k . (Для этого нужно рассмотреть \mathbf{w} со всеми элементами, кроме k -го, равными нулю.) Утверждение теоремы является более общим, чем этот факт, поскольку он применим, вообще говоря, к любой линейной комбинации элементов β .

4.3.6. Особенности стохастических регрессоров

Предыдущий анализ проводился условно относительно наблюдаемой выборки. Безусловный результат удобно получить, проведя анализ условно по \mathbf{b} , а затем, «усреднив» (т. е. проинтегрировав по всему множеству условных распределений), по \mathbf{X} . Смысл этого состоит в том, что если мы установили несмещенность условно относительно произвольного \mathbf{X} , то можем усреднить по всем \mathbf{X} и получить безусловный результат. Мы уже пользовались этим подходом для доказательства несмещенности \mathbf{b} в подразделе 4.3.1; теперь мы применим его к условной дисперсии.

Условная дисперсия \mathbf{b} равна

$$\text{Var}[\mathbf{b} | \mathbf{X}] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}.$$

Для получения безусловной дисперсии применим (В-69):

$$\text{Var}[\mathbf{b}] = E_{\mathbf{X}}[\text{Var}[\mathbf{b} | \mathbf{X}]] + \text{Var}_{\mathbf{X}}[E[\mathbf{b} | \mathbf{X}]].$$

Второе слагаемое равно нулю, поскольку $E[\mathbf{b} | \mathbf{X}] = \beta$ для всех \mathbf{X} , поэтому

$$\text{Var}[\mathbf{b}] = E_{\mathbf{X}}[\sigma^2(\mathbf{X}'\mathbf{X})^{-1}] = \sigma^2 E_{\mathbf{X}}[(\mathbf{X}'\mathbf{X})^{-1}].$$

Вывод, сделанный нами в предыдущем разделе, несколько изменяется: чтобы получить матрицу ковариаций, нужно заменить величину $(\mathbf{X}'\mathbf{X})^{-1}$ на

ее математическое ожидание, что несколько изменяет интерпретацию результата. Безусловную дисперсию \mathbf{b} можно описать только в терминах усредненного поведения \mathbf{X} , так что необходимо сделать некоторые предположения о дисперсиях и ковариациях регрессоров. Мы вернемся к этому вопросу в разделе 4.4.

В подразделе 4.3.5 мы показали, что

$$\text{Var}[\mathbf{b} | \mathbf{X}] \leq \text{Var}[\mathbf{b}_0 | \mathbf{X}]$$

для любой линейной несмещенной оценки $\mathbf{b}_0 \neq \mathbf{b}$ и конкретной выборки \mathbf{X} . Но если это неравенство верно для каждого \mathbf{X} , то оно должно быть верно и для

$$\text{Var}[\mathbf{b}] = E_{\mathbf{X}}[\text{Var}[\mathbf{b} | \mathbf{X}]].$$

То есть если неравенство выполняется для каждого конкретного \mathbf{X} , то оно должно выполняться и для среднего значения \mathbf{X} .

Таким образом, полученные нами свойства оценок методом наименьших квадратов — несмещенность и теорема Гаусса–Маркова — верны как условно для конкретной имеющейся выборки, так для случайной выборки из генеральной совокупности.

Теорема 4.3. *Теорема Гаусса–Маркова (продолжение)*

В модели линейной регрессии оценка наименьших квадратов \mathbf{b} является линейной несмещенной оценкой β с наименьшей дисперсией независимо от того, является \mathbf{X} случайной или неслучайной матрицей, если при этом сохраняются остальные предположения модели.

4.3.7. Оценки дисперсии методом наименьших квадратов

Если требуется построить тест для гипотезы о векторе β или доверительные интервалы для него, то нам понадобится выборочная оценка его матрицы ковариаций $\text{Var}[\mathbf{b} | \mathbf{X}] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. Необходимо оценить параметр генеральной совокупности σ^2 . Поскольку σ^2 — это математическое ожидание ε_i^2 , а e_i — это оценка ε_i , по аналогии можно предположить, что

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n e_i^2$$

является естественной оценкой σ^2 . Однако остатки e_i дают не очень хорошую оценку своих теоретических аналогов: $e_i = y_i - \mathbf{x}_i' \mathbf{b} = \varepsilon_i - \mathbf{x}_i' (\mathbf{b} - \beta)$. Эта оценка может быть искаженной, поскольку β не наблюдается в явном виде. В правой части возникает математическое ожидание квадрата и второе слагаемое может иметь ненулевое математическое ожидание.

Остатки равны

$$\mathbf{e} = \mathbf{M}\mathbf{y} = \mathbf{M}[\mathbf{X}\beta + \boldsymbol{\varepsilon}] = \mathbf{M}\boldsymbol{\varepsilon},$$

поскольку $\mathbf{M}\mathbf{X} = \mathbf{0}$ [см. (3-15).] Наш метод оценивания σ^2 будет основан на следующем представлении суммы квадратов остатков:

$$\mathbf{e}'\mathbf{e} = \boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}. \quad (4-16)$$

Математическое ожидание этой квадратичной формы равно

$$E[e'e | \mathbf{X}] = E[\epsilon' \mathbf{M} \epsilon | \mathbf{X}].$$

Скаляр $\epsilon' \mathbf{M} \epsilon$ представляет собой матрицу размером 1×1 , поэтому ее след совпадает с ней. Используя результат о циклических перестановках (А-94), получаем

$$E[\text{tr}(\epsilon' \mathbf{M} \epsilon) | \mathbf{X}] = E[\text{tr}(\mathbf{M} \epsilon \epsilon') | \mathbf{X}].$$

Поскольку \mathbf{M} является функцией от \mathbf{X} , имеем

$$\text{tr}(\mathbf{M} E[\epsilon \epsilon' | \mathbf{X}]) = \text{tr}(\mathbf{M} \sigma^2 \mathbf{I}) = \sigma^2 \text{tr}(\mathbf{M}).$$

След \mathbf{M} равен

$$\text{tr}[\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] = \text{tr}(\mathbf{I}_n) - \text{tr}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}] = \text{tr}(\mathbf{I}_n) - \text{tr}(\mathbf{I}_K) = n - K.$$

Таким образом,

$$E[e'e | \mathbf{X}] = (n - K)\sigma^2,$$

так что естественная оценка, рассмотренная ранее, смещена к нулю, хотя смещение и уменьшается с ростом размера выборки. Несмещенная оценка σ^2 имеет вид

$$s^2 = \frac{e'e}{n - K}. \quad (4-17)$$

Эта оценка является также несмещенной безусловно, поскольку

$$E[s^2] = E_{\mathbf{X}} E[s^2 | \mathbf{X}] = E_{\mathbf{X}}[\sigma^2] = \sigma^2.$$

Стандартная ошибка регрессии — это s , квадратный корень s^2 . Используя s^2 , можно вычислить

$$\text{Est. Var}[\mathbf{b} | \mathbf{X}] = s^2(\mathbf{X}'\mathbf{X})^{-1}.$$

Далее мы будем обозначать выборочную оценку выборочной дисперсии оценки $\text{Est. Var}[\cdot]$. Квадратный корень k -го диагонального элемента этой матрицы, $\{[s^2(\mathbf{X}'\mathbf{X})^{-1}]_{kk}\}^{1/2}$, называется **стандартной ошибкой** оценки b_k , хотя часто его называют просто стандартной ошибкой b_k .

4.3.8. Предположение о нормальности

Пока что наш анализ регрессионных моделей был **полупараметрическим** (см. раздел 12.3). Мы нигде не пользовались предположением А6 (см. табл. 4.1) о нормальности ϵ . Это предположение нужно при построении доверительных интервалов. Из (4-4) следует, что \mathbf{b} является линейной функцией вектора шоков ϵ . Если допустить, что ϵ имеют многомерное нормальное распределение, то мы можем использовать результаты раздела В10.2, а также уже вычисленные вектор средних и ковариационную матрицу и получить

$$\mathbf{b} | \mathbf{X} \sim N[\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}]. \quad (4-18)$$

Это дает многомерное нормальное распределение, так что каждый элемент $\mathbf{b} | \mathbf{X}$ распределен нормально:

$$b_k | \mathbf{X} \sim N[\beta_k, \sigma^2 (\mathbf{X}'\mathbf{X})_{kk}^{-1}]. \quad (4-19)$$

Подобную картину мы наблюдали на рис. 4.1 в примере 4.1.

Распределение \mathbf{b} является условным по \mathbf{X} . Нормальное распределение \mathbf{b} на конечной выборке — это следствие нашего предположения о нормальности распределения возмущений. В отсутствие этого предположения (или другого похожего условия на распределение ϵ) мы не могли бы определить точное распределение \mathbf{b} , ни условное, ни безусловное. В разделе 4.4 мы сможем получить приближенно нормальное распределение независимо от того, нормально или нет распределение \mathbf{b} и случайны ли или нет регрессоры.

4.4. Свойства оценок наименьших квадратов на больших выборках

Пользуясь только предположениями A1–A4 из табл. 4.1, мы установили следующие свойства оценок наименьших квадратов \mathbf{b} и s^2 параметров β и σ^2 в **конечных выборках**:

- $E[\mathbf{b} | \mathbf{X}] = E[\mathbf{b}] = \beta$ — оценка коэффициента является несмещенной;
- $E[s^2 | \mathbf{X}] = E[s^2] = \sigma^2$ — оценка дисперсии шоков является несмещенной;
- $\text{Var}[\mathbf{b} | \mathbf{X}] = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}$ и $\text{Var}[\mathbf{b}] = \sigma^2 E[(\mathbf{X}'\mathbf{X})^{-1}]$;
- теорема Гаусса–Маркова: для любого постоянного вектора \mathbf{w} линейной несмещенной оценкой $\mathbf{w}'\beta$ с наименьшей дисперсией (MVLUE, minimum variance linear unbiased estimate) является $\mathbf{w}'\mathbf{b}$.

В этой модели можно также непосредственно вывести асимптотические свойства оценок (или свойства на больших выборках). Предположение о нормальности, A6, здесь перестает быть важным, и мы откажемся от него до обсуждения оценок максимального правдоподобия в подразделе 4.4.6 и главе 14.

4.4.1. Состоятельность оценки наименьших квадратов коэффициента β

Несмещенность является полезным свойством для первоначального изучения плюсов некоторой оценки. Знание о несмещенности оценки дает исследователю уверенность в том, что оценка не будет систематически заниженной или завышенной. Однако в качестве стратегии оценивания это знание имеет два недостатка. Во-первых, оценки (за исключением обсуждаемых в этой главе оценок коэффициентов методом наименьших квадратов) редко оказываются несмещенными. Практически всегда в моделях более сложных, чем множественная регрессия, можно лишь надеяться, что оценка будет улучшаться в том смысле, что ее смещение будет становиться все меньше по мере увеличения информации (при росте размера выборки). Таким образом, нам нужен более широкий набор средств для хорошего

эконометрического анализа. Во-вторых, несмещенность не дает гарантии, что качество оценки улучшается с ростом размера выборки. Выборочные средние, вычисленные по выборкам размеров 2, 100 или 10 000, будут несмещенными оценками среднего генеральной совокупности, и, если использовать лишь критерий несмещенности, нельзя сказать, что какая-то из этих оценок лучше. Естественно, хотелось бы надеяться, что большая выборка лучше маленькой (а соответственно очень большая намного лучше или даже идеальна). Свойство **состоятельности** в обоих отношениях значительно лучше, чем свойство несмещенности.

Для начала мы не будем специфицировать механизм, порождающий матрицу \mathbf{X} : матрица \mathbf{X} может быть любой комбинацией постоянных и случайных переменных, порожденных независимо от ε . Мы сделаем два важных допущения. Первое является модификацией предположения А5 из табл. 4.1:

А5а. $(\mathbf{x}_i, \varepsilon_i) \ i = 1, \dots, n$ – последовательность независимых наблюдений.

Второе касается поведения данных в больших выборках:

$$\text{plim}_{n \rightarrow \infty} \frac{\mathbf{X}'\mathbf{X}}{n} = \mathbf{Q} \text{ – положительно определенная матрица.} \quad (4-20)$$

[Мы вскоре вернемся к (4-20).] Оценку методом наименьших квадратов можно записать в виде

$$\mathbf{b} = \boldsymbol{\beta} + \left(\frac{\mathbf{X}'\mathbf{X}}{n} \right)^{-1} \left(\frac{\mathbf{X}'\boldsymbol{\varepsilon}}{n} \right). \quad (4-21)$$

Если определена \mathbf{Q}^{-1} , то

$$\text{plim } \mathbf{b} = \boldsymbol{\beta} + \mathbf{Q}^{-1} \text{plim} \left(\frac{\mathbf{X}'\boldsymbol{\varepsilon}}{n} \right),$$

поскольку обратная матрица является непрерывной функцией исходной. (Здесь мы воспользовались теоремой D.14.) Нам нужен предел по вероятности последнего члена. Пусть

$$\frac{1}{n} \mathbf{X}'\boldsymbol{\varepsilon} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \varepsilon_i = \frac{1}{n} \sum_{i=1}^n \mathbf{w}_i = \bar{\mathbf{w}}. \quad (4-22)$$

Тогда

$$\text{plim } \mathbf{b} = \boldsymbol{\beta} + \mathbf{Q}^{-1} \text{plim } \bar{\mathbf{w}}.$$

Из предположения об экзогенности А3 получаем, что

$$E[\mathbf{w}_i] = E_{\mathbf{x}}[E[\mathbf{w}_i | \mathbf{x}_i]] = E_{\mathbf{x}}[\mathbf{x}_i E[\varepsilon_i | \mathbf{x}_i]] = \mathbf{0},$$

так что точное математическое ожидание равно $E[\bar{\mathbf{w}}] = \mathbf{0}$. Для неслучайных элементов \mathbf{x}_i равенство математических ожиданий нулю можно получить из одномерных распределений ε_i . Теперь рассмотрим дисперсию. Согласно (B-70), $\text{Var}[\bar{\mathbf{w}}] = E[\text{Var}[\bar{\mathbf{w}} | \mathbf{X}]] + \text{Var}[E[\bar{\mathbf{w}} | \mathbf{X}]]$. Второй член равен нулю, поскольку $E[\varepsilon_i | \mathbf{x}_i] = 0$. Чтобы вычислить первый член, воспользуемся $E[\varepsilon \varepsilon' | \mathbf{X}] = \sigma^2 \mathbf{I}$, откуда

$$\text{Var}[\bar{\mathbf{w}} | \mathbf{X}] = E[\bar{\mathbf{w}} \bar{\mathbf{w}}' | \mathbf{X}] = \frac{1}{n} \mathbf{X}' E[\varepsilon \varepsilon' | \mathbf{X}] \mathbf{X} \frac{1}{n} = \left(\frac{\sigma^2}{n} \right) \left(\frac{\mathbf{X}'\mathbf{X}}{n} \right).$$

Таким образом,

$$\text{Var}[\bar{\mathbf{w}}] = \left(\frac{\sigma^2}{n}\right) E\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right).$$

Дисперсия будет сходиться к нулю, если математическое ожидание в скобках сходится к постоянной матрице, так что все выражение при росте n будет определяться первым скалярным членом. Предположения (4-20) для этого будет достаточно. (Вообще говоря, математическое ожидание может расходиться при сходящемся пределе по вероятности, но этот случай мало важен для практических целей.) Тогда отсюда следует, что

$$\lim_{n \rightarrow \infty} \text{Var}[\bar{\mathbf{w}}] = 0 \cdot \mathbf{Q} = \mathbf{0}. \quad (4-23)$$

Поскольку среднее $\bar{\mathbf{w}}$ тождественно равно нулю, а его дисперсия стремится к нулю, то $\bar{\mathbf{w}}$ сходится в среднеквадратичном к нулю, а значит, $\text{plim } \bar{\mathbf{w}} = \mathbf{0}$. Значит,

$$\text{plim } \frac{\mathbf{X}'\boldsymbol{\varepsilon}}{n} = \mathbf{0} \quad (4-24)$$

и тогда

$$\text{plim } \mathbf{b} = \boldsymbol{\beta} + \mathbf{Q}^{-1} \cdot \mathbf{0} = \boldsymbol{\beta}. \quad (4-25)$$

Таким образом, при выполнении предположений A1–A4 и дополнительного предположения (4-20) \mathbf{b} является **состоятельной оценкой** $\boldsymbol{\beta}$ в линейной модели регрессии.

При изучении временных рядов, включающих разнообразные временные тренды, часто возникают ситуации, при которых указанные предположения являются слишком строгими. В табл. 4.2 приводится набор более слабых условий **Гренандера**³, которые, как правило, верны для большинства ситуаций. Эти условия гарантируют «доброкачественность» («корректное поведение», «well-behaved») матрицы данных при росте размера выборки⁴.

Таблица 4.2. Условия Гренандера для доброкачественных данных (Grenander Conditions for Well-Behaved Data)

G1. Для каждого столбца матрицы \mathbf{X} , \mathbf{x}_k , из условия $d_{nk}^2 = \mathbf{x}'_k \mathbf{x}_k$ следует, что $\lim_{n \rightarrow \infty} d_{nk}^2 = +\infty$. Следовательно, \mathbf{x}_k не сходится к последовательности нулей. Суммы квадратов продолжают расти при росте размера выборки. Ни одна переменная не обращается в ноль.

G2. $\lim_{n \rightarrow \infty} x_{ik}^2 / d_{nk}^2 = 0$ для всех $i = 1, \dots, n$. Это условие означает, что $\mathbf{x}'_k \mathbf{x}_k$ не будет в большой степени определяться одним-единственным наблюдением и вклад каждого конкретного наблюдения уменьшается при росте n .

G3. Пусть \mathbf{R}_n — матрица выборочных корреляций столбцов матрицы \mathbf{X} , кроме столбца констант, если он присутствует. Тогда $\lim_{n \rightarrow \infty} \mathbf{R}_n = \mathbf{C}$, где \mathbf{C} — положительно определенная матрица. Это означает, что условие полноты ранга всегда будет выполняться. Мы уже предполагали, что на конечных выборках \mathbf{X} имеет полный ранг. Этим условием мы гарантируем, что наше предположение не будет нарушено при росте размера выборки.

³Judge et al. (1985, p. 162).

⁴Более подробное исследование подобных условий можно найти в White (2001).

4.4.2. Асимптотическая нормальность оценок наименьших квадратов

Как критерий качества оценки состоятельность важнее несмещенности. Поскольку мы стремимся ослабить некоторые условия исходной модели, в частности А6 (нормальность шоков), теряем полученный нами результат о нормальном распределении оценки, который нужен для построения доверительных интервалов в разделе 4.5. Кажется, что за построение более общей модели приходится жертвовать этим результатом. В этом разделе мы покажем, что нормальное распределение шоков не является необходимым условием для получения свойств распределения оценки, которые нам необходимы для статистических выводов, включая построение доверительных интервалов и проверку гипотез. Оказывается, при достаточно широких естественных условиях порождения данных распределения в больших выборках дадут вполне надежный фундамент для статистических выводов в модели регрессии. (Этот подход мы будем использовать и далее в книге при рассмотрении более тонких методов оценивания.)

Для получения асимптотических свойств оценок методом наименьших квадратов мы воспользуемся результатами раздела D.3. Чтобы пользоваться различными вариантами центральных предельных теорем, нам придется дополнить условие А3 (отсутствие корреляций) предположением о **независимости** наблюдений. Из (4-21) следует, что

$$\sqrt{n}(\mathbf{b} - \boldsymbol{\beta}) = \left(\frac{\mathbf{X}'\mathbf{X}}{n} \right)^{-1} \left(\frac{1}{\sqrt{n}} \right) \mathbf{X}'\boldsymbol{\varepsilon}. \quad (4-26)$$

Поскольку обратная матрица является непрерывной функцией исходной, то $\text{plim}(\mathbf{X}'\mathbf{X}/n)^{-1} = \mathbf{Q}^{-1}$. Значит, если предельное распределение случайного вектора из (4-26) существует, то оно совпадает с предельным распределением

$$\left[\text{plim} \left(\frac{\mathbf{X}'\mathbf{X}}{n} \right)^{-1} \right] \left(\frac{1}{\sqrt{n}} \right) \mathbf{X}'\boldsymbol{\varepsilon} = \mathbf{Q}^{-1} \left(\frac{1}{\sqrt{n}} \right) \mathbf{X}'\boldsymbol{\varepsilon}. \quad (4-27)$$

Значит, нам нужно найти предельное распределение

$$\left(\frac{1}{\sqrt{n}} \right) \mathbf{X}'\boldsymbol{\varepsilon} = \sqrt{n}(\bar{\mathbf{w}} - E[\bar{\mathbf{w}}]), \quad (4-28)$$

где $E[\bar{\mathbf{w}}] = \mathbf{0}$ [см. (4-22).] Для нахождения предельного распределения $\sqrt{n}\bar{\mathbf{w}}$ воспользуемся многомерным вариантом центральной предельной теоремы, известным как теорема Линдберга–Феллера (D19.A)⁵. Заметим, что $\bar{\mathbf{w}}$ равно среднему n независимых случайных векторов $\mathbf{w}_i = \mathbf{x}_i\varepsilon_i$ со средними $\mathbf{0}$ и дисперсиями

$$\text{Var}[\mathbf{x}_i\varepsilon_i] = \sigma^2 E[\mathbf{x}_i\mathbf{x}_i'] = \sigma^2 \mathbf{Q}_i. \quad (4-29)$$

Дисперсия $\sqrt{n}\bar{\mathbf{w}}$ равна

$$\sigma^2 \bar{\mathbf{Q}}_n = \sigma^2 \left(\frac{1}{n} \right) [\mathbf{Q}_1 + \mathbf{Q}_2 + \cdots + \mathbf{Q}_n]. \quad (4-30)$$

⁵Заметим, что теоремой Линдберга–Леви воспользоваться не удастся, так как величина $\text{Var}[\mathbf{w}_i]$ необязательно является постоянной.

Если эта сумма не определяется единственным слагаемым и регрессоры являются доброкачественными, что в нашем случае означает выполнение (4-20), то

$$\lim_{n \rightarrow \infty} \sigma^2 \bar{\mathbf{Q}}_n = \sigma^2 \mathbf{Q}. \quad (4-31)$$

Таким образом, можно применить центральную предельную теорему Линдберга–Феллера к вектору $\sqrt{n} \bar{\mathbf{w}}$ аналогично тому, как это делается в разделе D.3 с $\sqrt{n} \bar{x}$ (в одномерном случае). Теперь у нас есть все, что нужно для строгого доказательства. Если $[\mathbf{x}_i \varepsilon_i]$, $i = 1, \dots, n$ являются независимыми векторами со средними, равными \mathbf{P} , и дисперсиями $\sigma^2 \mathbf{Q}_i < \infty$ и выполняется (4-20), то

$$\left(\frac{1}{\sqrt{n}} \right) \mathbf{X}' \boldsymbol{\varepsilon} \xrightarrow{d} N[\mathbf{0}, \sigma^2 \mathbf{Q}]. \quad (4-32)$$

Отсюда следует, что

$$\mathbf{Q}^{-1} \left(\frac{1}{\sqrt{n}} \right) \mathbf{X}' \boldsymbol{\varepsilon} \xrightarrow{d} N[\mathbf{Q}^{-1} \mathbf{0}, \mathbf{Q}^{-1} (\sigma^2 \mathbf{Q}) \mathbf{Q}^{-1}]. \quad (4-33)$$

После необходимых преобразований получаем

$$\sqrt{n}(\mathbf{b} - \boldsymbol{\beta}) \xrightarrow{d} N[\mathbf{0}, \sigma^2 \mathbf{Q}^{-1}]. \quad (4-34)$$

Используя методы раздела D.3, получаем **асимптотическое распределение \mathbf{b}** .

Теорема 4.4. *Асимптотическое распределение \mathbf{b} в случае независимых наблюдений*

Если $\{\varepsilon_i\}$ распределены независимо с нулевым средним и конечной дисперсией σ^2 , а x_{ik} удовлетворяют условиям Гренандера, то

$$\mathbf{b} \stackrel{a}{\sim} N \left[\boldsymbol{\beta}, \frac{\sigma^2}{n} \mathbf{Q}^{-1} \right]. \quad (4-35)$$

На практике необходимо оценить $(1/n) \mathbf{Q}^{-1}$ с помощью $(\mathbf{X}' \mathbf{X})^{-1}$, а σ^2 — с помощью $\mathbf{e}' \mathbf{e} / (n - K)$.

Если ε распределены нормально, то условие (4-18), т.е. нормальность \mathbf{b}/X , выполняется для любой выборки, а значит, и асимптотически. Важный результат, полученный только что, гласит, что при доброкачественных регрессорах и независимых наблюдениях **асимптотическая нормальность** оценки методом наименьших квадратов не зависит от нормальности распределения шоков, а является следствием центральной предельной теоремы. В следующих разделах мы рассмотрим и другие, более общие случаи.

4.4.3. Состоятельность s^2 и построение оценки для $\text{Asy. Var}[\mathbf{b}]$

Чтобы продолжить обсуждение асимптотических свойств \mathbf{b} , нам понадобится оценка $\text{Asy. Var}[\mathbf{b}] = (\sigma^2/n) \mathbf{Q}^{-1}$ ⁶. При условии (4-20) достаточно огра-

⁶См. работу McCallum (1973) для ознакомления с полезными замечаниями о выводе асимптотической ковариационной матрицы оценки наименьших квадратов.

ничиться s^2 , поэтому наша цель — доказать состоятельность s^2 как оценки σ^2 . Из

$$s^2 = \frac{1}{n - K} \boldsymbol{\varepsilon}' \mathbf{M} \boldsymbol{\varepsilon}$$

следует, что

$$s^2 = \frac{1}{n - K} [\boldsymbol{\varepsilon}' \boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\varepsilon}] = \frac{n}{n - k} \left[\frac{\boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}}{n} - \left(\frac{\boldsymbol{\varepsilon}' \mathbf{X}}{n} \right) \left(\frac{\mathbf{X}' \mathbf{X}}{n} \right)^{-1} \left(\frac{\mathbf{X}' \boldsymbol{\varepsilon}}{n} \right) \right].$$

Скалярный множитель, очевидно, сходится к 1. Применяя (4-20), (4-24) (дважды) и правило произведения **пределов по вероятности** (теорема D.14), можно убедиться, что второе слагаемое в скобках сходится к 0. Таким образом, остается

$$\overline{\varepsilon^2} = \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2.$$

В нашем частном случае, когда ε_i^2 независимы и имеют конечное математическое ожидание, равное σ^2 , для того чтобы их среднее сходилось к $\sigma^2 = E[\varepsilon_i^2]$ почти наверное, требуется совсем немного. По теореме Маркова (D.8) необходимо, чтобы существовали конечные $E[|\varepsilon_i^2|^{1+\delta}]$, и, таким образом, минимальным необходимым предположением является существование у ε_i конечных моментов чуть большего порядка, чем 2. Если дополнительно предположить, что все ε_i распределены одинаково, то по теореме Хинчина (D.5) или следствию из D.8 достаточно существования конечных моментов второго и менее порядков. **Сходимость в среднеквадратичном** требует условия $E[\varepsilon_i^4] = \phi_\varepsilon < \infty$. При этом слагаемые нашей суммы независимы, имеют математическое ожидание σ^2 и дисперсию $\phi_\varepsilon - \sigma^4$. Таким образом, при достаточно слабых условиях первый член в скобках сходится по вероятности к σ^2 , что приводит нас к ответу

$$\text{plim } s^2 = \sigma^2$$

и по правилу произведения

$$\text{plim } s^2 (\mathbf{X}' \mathbf{X} / n)^{-1} = \sigma^2 \mathbf{Q}^{-1}.$$

Подходящей оценкой асимптотической ковариационной матрицы \mathbf{b} является

$$\text{Est. Asy. Var}[\mathbf{b}] = s^2 (\mathbf{X}' \mathbf{X})^{-1}.$$

4.4.4. Асимптотическое распределение функций от \mathbf{b} : дельта-метод

Теорему D.22 можно распространить на функции от оценки методом наименьших квадратов. Пусть $\mathbf{f}(\mathbf{b})$ — это набор из J непрерывных линейных или нелинейных, непрерывно дифференцируемых функций от оценки \mathbf{b} и пусть

$$\mathbf{C}(\mathbf{b}) = \frac{\partial \mathbf{f}(\mathbf{b})}{\partial \mathbf{b}'},$$

где \mathbf{C} — матрица размером $J \times K$, j -я строка которой представляет собой вектор из производных j -й функции по \mathbf{b}' . По теореме Слуцкого (D.12)

$$\text{plim } \mathbf{f}(\mathbf{b}) = \mathbf{f}(\boldsymbol{\beta})$$

и

$$\text{plim } \mathbf{C}(\mathbf{b}) = \frac{\partial \mathbf{f}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} = \boldsymbol{\Gamma}.$$

Раскладывая в ряд Тейлора до первого члена,

$$\mathbf{f}(\mathbf{b}) = \mathbf{f}(\boldsymbol{\beta}) + \boldsymbol{\Gamma} \times (\mathbf{b} - \boldsymbol{\beta}) + \text{члены более высоких порядков.}$$

Членами высоких порядков в больших выборках можно пренебречь, если $\text{plim } \mathbf{b} = \boldsymbol{\beta}$. Значит, асимптотическое распределение функции в левой части совпадает с таковым для правой части. Тогда математическое ожидание этого асимптотического распределения равно $\text{plim } \mathbf{f}(\mathbf{b}) = \mathbf{f}(\boldsymbol{\beta})$, а асимптотическая ковариационная матрица равна $\{\boldsymbol{\Gamma}[\text{Asy. Var}(\mathbf{b} - \boldsymbol{\beta})]\boldsymbol{\Gamma}'\}$, что приводит нас к следующей теореме.

Теорема 4.5. *Асимптотическое распределение функций от \mathbf{b}*

Если $\mathbf{f}(\mathbf{b})$ — набор непрерывных непрерывно дифференцируемых функций от \mathbf{b} , для которых $\boldsymbol{\Gamma} = \partial \mathbf{f}(\boldsymbol{\beta})/\partial \boldsymbol{\beta}'$ и выполняется теорема 4.4, то

$$\mathbf{f}(\mathbf{b}) \overset{a}{\sim} N \left[\mathbf{f}(\boldsymbol{\beta}), \boldsymbol{\Gamma} \left(\frac{\sigma^2}{n} \mathbf{Q}^{-1} \right) \boldsymbol{\Gamma}' \right]. \quad (4-36)$$

На практике оценкой ковариационной матрицы будет

$$\text{Est. Asy. Var}[\mathbf{f}(\mathbf{b})] = \mathbf{C}[s^2(\mathbf{X}'\mathbf{X})^{-1}]\mathbf{C}'.$$

Если некоторые из функций нелинейны, то несмещенность \mathbf{b} не влечет несмещенности $\mathbf{f}(\mathbf{b})$. Однако из (4-25) можно получить, что $\mathbf{f}(\mathbf{b})$ является состоятельной оценкой $\mathbf{f}(\boldsymbol{\beta})$, а ее асимптотическая ковариационная матрица легко вычисляется.

Пример 4.4. *Нелинейные функции от параметров: дельта-метод*

Динамический вариант модели спроса на бензин из примера 2.3 будет использован для разделения кратко- и долгосрочных эффектов изменений доходов и цен. Модель имеет вид

$$\begin{aligned} \ln(G/Pop)_t = & \beta_1 + \beta_2 \ln P_{G,t} + \beta_3 \ln(Income/Pop)_t + \beta_4 \ln P_{nc,t} + \\ & + \beta_5 \ln P_{uc,t} + \gamma \ln(G/Pop)_{t-1} + \varepsilon_t, \end{aligned}$$

где P_{nc} и P_{uc} — индексы цен новых и подержанных машин соответственно. В этой модели краткосрочные эластичности спроса по цене и доходу равны β_2 и β_3 соответственно. Долгосрочные эластичности равны $\phi_2 = \beta_2/(1 - \gamma)$ и $\phi_3 = \beta_3/(1 - \gamma)$. Чтобы оценить долгосрочные эластичности, оценим параметры методом наименьших квадратов, а затем вычислим эти нелинейные функции от оценок параметров. Дельта-метод позволит нам оценить стандартные отклонения.

В табл. 4.3 приведены оценки параметров модели, стандартные отклонения и t -статистики. В частности, приведены краткосрочные эластичности. Оценки долгосрочных эластичностей равны

$$f_2 = b_2 / (1 - c) = -0,069532 / (1 - 0,830971) = -0,411358$$

и

$$f_3 = 0,164047 / (1 - 0,830971) = 0,970522.$$

Чтобы оценить стандартные отклонения, вычислим частные производные наших функций по шести параметрам модели:

$$\begin{aligned} \mathbf{g}'_2 &= \frac{\partial \phi_2}{\partial \beta'} = [0, 1/(1 - \gamma), 0, 0, 0, \beta_2 / (1 - \gamma)^2] = \\ &= [0, 5,91613, 0, 0, 0, -2,43365], \\ \mathbf{g}'_3 &= \frac{\partial \phi_3}{\partial \beta'} = [0, 1/(1 - \gamma), 0, 0, 0, \beta_3 / (1 - \gamma)^2] = \\ &= [0, 0, 5,91613, 0, 0, 5,74174], \end{aligned}$$

Таблица 4.3. Результаты оценки регрессии уравнения спроса

| Сумма квадратов остатков: | | | | 0,0127352 | |
|---|-------------|--------------------|--------------|--------------|--------------------|
| Стандартная ошибка регрессии: | | | | 0,0168227 | |
| R^2 на основе 51 наблюдения | | | | 0,9951081 | |
| Переменная | Коэффициент | Стандартная ошибка | Статистика | | |
| Constant | -3,123195 | 0,99583 | -3,136 | | |
| $\ln P_G$ | -0,069532 | 0,01473 | -4,720 | | |
| $\ln Income / Pop$ | 0,164047 | 0,05503 | 2,981 | | |
| $\ln P_{nc}$ | -0,178395 | 0,05517 | -3,233 | | |
| $\ln P_{uc}$ | 0,127009 | 0,03577 | 3,551 | | |
| $\ln G / Pop$ в предыдущий момент | 0,830971 | 0,04576 | 18,158 | | |
| Оцененная матрица ковариаций для \mathbf{b} ($e - n =$ умножить на 10^{-n}) | | | | | |
| Constant | $\ln P_G$ | $\ln(Income/Pop)$ | $\ln P_{nc}$ | $\ln P_{uc}$ | $\ln(G/Pop)_{t-1}$ |
| 0,99168 | | | | | |
| -0,0012088 | 0,00021705 | | | | |
| -0,052602 | 1,62165e-5 | 0,0030279 | | | |
| 0,0051016 | -0,00021705 | -0,00024708 | 0,0030440 | | |
| 0,0091672 | -4,0551e-5 | -0,00060624 | -0,0016782 | 0,0012795 | |
| 0,043915 | -0,0001109 | -0,0021881 | 0,00068116 | 8,57001e-5 | 0,0020943 |

С помощью (4-36) можно теперь оценить асимптотические дисперсии долгосрочных эластичностей, вычислив $\mathbf{g}'_2 [s^2(\mathbf{X}'\mathbf{X})^{-1}] \mathbf{g}_2$ и $\mathbf{g}'_3 [s^2(\mathbf{X}'\mathbf{X})^{-1}] \mathbf{g}_3$. Получаем соответственно 0,023194 и 0,0263692. Асимптотические стандартные ошибки — это квадратные корни из этих чисел, равные 0,152296 и 0,162386.

4.4.5. Асимптотическая эффективность

Мы пока не выводили аналог теоремы Гаусса–Маркова для больших выборок. Это значит, что мы пока не знаем, являются ли оценки методом наименьших квадратов в каком-либо смысле оптимальными в асимптотике. Теорема Гаусса–Маркова дает условия на конечную выборку, при выполнении которых оценка оказывается оптимальной. Требования линейности и несмещенности оценки, однако, существенно ограничивают общность теоремы. В этом подразделе мы попытаемся построить аналог этой теоремы для класса состоятельных (хотя, возможно, смещенных) линейных оценок. Позже мы займемся и нелинейными оценками.

Для изучения этих классов оценок нам понадобится новый критерий качества оценки.

Определение 4.1. Асимптотическая эффективность

Оценка называется асимптотически эффективной, если она состоятельна, асимптотически нормальна и имеет ковариационную матрицу не большую, чем у любой другой состоятельной, асимптотически нормальной оценки.

Две оценки можно сравнивать по их асимптотическим дисперсиям. Сложность при сравнении двух состоятельных оценок заключается в том, что обе оценки стремятся к истинному значению параметра при росте размера выборки. Более того, обычно оказывается (как в примере 4.5), что они к тому же имеют одинаковую скорость сходимости, т.е. величины их асимптотических дисперсий имеют одинаковый порядок по размеру выборки, например $O(1/n)$. В этой ситуации иногда можно сравнить дисперсии двух оценок на выборках одинакового размера. В следующем примере мы рассмотрим оценку, минимизирующую модуль отклонения, как альтернативу оценке метода наименьших квадратов.

Пример 4.5. Метод наименьших квадратов против метода наименьших абсолютных отклонений: сравнение методом Монте-Карло

Ранее (в разделе 4.2) мы видели, что метод наименьших квадратов, несмотря на ряд своих плюсов, не единственный возможный метод построения оценок в линейной модели. В качестве альтернативы можно рассмотреть оценки LAD (least absolute deviations) — оценки, минимизирующие сумму модулей отклонений. (Более подробно эти оценки будут рассмотрены в подразделе 7.3.1.) Эти оценки строятся следующим образом:

$$\mathbf{b}_{\text{LAD}} = \text{вектор, минимизирующий } \sum_{i=1}^n |y_i - \mathbf{x}'_i \mathbf{b}_0|,$$

в отличие от оценок метода наименьших квадратов, которые строятся как

$$\mathbf{b}_{\text{LS}} = \text{вектор, минимизирующий } \sum_{i=1}^n (y_i - \mathbf{x}'_i \mathbf{b}_0)^2.$$

Предположим, что модель регрессии имеет вид

$$y_i = \mathbf{x}'_i \boldsymbol{\beta} + \varepsilon_i,$$

причем распределение ε_i имеет нулевое условное среднее, постоянную дисперсию σ^2 и нулевую условную медиану, т.е. распределение симметрично; также потребуем, чтобы $\text{plim}(1/n)\mathbf{X}'\boldsymbol{\varepsilon} = \mathbf{0}$. Это обычные предположения линейной модели, но с заменой требования нормальности распределения требованием его симметричности. При таких условиях оценка \mathbf{P}_{LS} состоятельна и асимптотически нормальна, а ее асимптотическая ковариационная матрица описывается теоремой 4.4 и будет обозначаться $\sigma^2 \mathbf{A}$. В ряде работ (Koenker, Bassett (1978, 1982), Huber (1987), Rogers (1993), Koenker (2005)) было показано, что в таких условиях оценка \mathbf{b}_{LAD} также является состоятельной. Хорошей оценкой асимптотической дисперсии \mathbf{b}_{LAD} будет величина $(1/2)^2 [1/f(0)]^2 \mathbf{A}$, где $f(0)$ — плотность распределения ε в нуле, т.е. в его медиане. Таким образом, мы можем сравнить эти две оценки, сравнивая их асимптотические дисперсии. Отношение асимптотической дисперсии k -го элемента \mathbf{b}_{LAD} и асимптотической дисперсии соответствующего члена \mathbf{b}_{LS} равно

$$q_k = \text{Var}(b_{k,LAD})/\text{Var}(b_{k,LS}) = (1/2)^2 (1/\sigma^2) [1/f(0)]^2.$$

Если ε на самом деле распределены нормально и имеют нулевое математическое ожидание и медиану, то

$$f(\varepsilon) = (2\pi\sigma^2)^{-1/2} \exp(-\varepsilon^2/(2\sigma^2)),$$

так что $f(0) = (2\pi\sigma^2)^{-1/2}$, и в этом случае $q_k = \pi/2$. Итак, если ошибки распределены нормально, то оценки LAD асимптотически менее эффективны в $\pi/2 = 1,573$ раза.

Оценки LAD оказываются полезны, когда нормальность распределения ошибок предполагать нельзя. В результате большого количества исследований было установлено, что преимущества LAD-оценок особенно значительны при маленьких размерах выборок и если распределение ε имеет более толстые хвосты, чем нормальное, т.е. более вероятны большие по модулю значения y_i . При росте размера выборки оценки можно ожидать, что оценки LS восстановят свое преимущество. Мы рассмотрим этот аспект с помощью маленького исследования методом Монте-Карло.

В примерах 2.6 и 3.4 мы наблюдали интересный аспект ценообразования на рынке произведений искусства: оказывается, большие картины на аукционах продаются в среднем дороже, чем маленькие. Таблица F4.1 в приложении содержит информацию о продажной цене, ширине и высоте 430 картин Моне. Все эти картины были проданы на аукционах по ценам от 10 тыс. до 33 млн долл. Результаты оценки линейной регрессии логарифма цены на константу, логарифм площади и отношение сторон холста приведены в верхней строке табл. 4.4. Это основной результат, на который мы опираемся в нашем исследовании.

Таблица 4.4. Оцененные уравнения для цен на произведения искусства

| Полная выборка | Константа | | Log Area | | Aspect Ratio | |
|-------------------|-----------|---------------------------|----------|---------------------------|--------------|---------------------------|
| | Среднее | Стандартное отклонение | Среднее | Стандартное отклонение | Среднее | Стандартное отклонение |
| LS | -8,42653 | 0,61184 | 1,33372 | 0,09072 | | 0,12753 |
| LAD | -7,62436 | 0,89055 | 1,20404 | 0,13626 | -0,21260 | 0,13628 |
| R = 10 | | | | | | |
| LS | -9,39384 | 6,82900 | 1,40481 | 1,00545 | 0,39446 | 2,14847 |
| LAD | -8,97714 | 10,24781 | 1,34197 | 1,48038 | 0,35842 | 3,04773 |
| R = 50 | | | | | | |
| LS | -8,73099 | 2,12135 | 1,36735 | 0,30025 | | 0,52222 |
| LAD | -8,91671 | 2,51491 | 1,38489 | 0,36299 | -0,06129 | 0,63205 |
| R = 100 | | | | | | |
| LS | -8,36163 | 1,32083 | 1,32758 | 0,17836 | -0,17357 | 0,28977 |
| LAD | -8,05195 | 1,54190 | 1,27340 | 0,21808 | -0,20700 | 0,29465 |

Чтобы продемонстрировать разницу в поведении оценок LS и LAD, мы проведем следующее исследование методом Монте-Карло⁷. Сделаем 100 выборок без возвращения размера R из нашего набора из 430 наблюдений. Для каждой из этих ста выборок построим оценки $\mathbf{b}_{LS,r}$ и $\mathbf{b}_{LAD,r}$. Затем усредним оценки каждого метода и построим их выборочную дисперсию⁸. Выборочная дисперсия 100 наборов результатов соответствует понятию «разброс в повторяющихся выборках». Мы проделаем этот эксперимент для $R = 10, 50$ и 100 . Поскольку вся выборка достаточно велика, можно воспользоваться результатами регрессии, полученными с использованием всей выборки, как истинными значениями параметров. Приведенные в табл. 4.1 стандартные отклонения для оценок LAD, построенных для всей выборки, получены с помощью **бутстрэпа** (бутстрэпирования, bootstrap). Короче говоря, процедура состоит в следующем: мы берем B — у нас $B = 100$ — выборок размером n (430) с возвращением из общей выборки размера n . По каждой из этих выборок строится оценка LAD. Оценка дисперсии оценки LAD получается вычислением среднеквадратичного отклонения этих B -оценок от оценки, построенной по первоначальной выборке (не среднее этих B -оценок). Этот метод будет подробно описан в разделе 15.4. Если предположения модели верны, то должно наблюдаться следующее:

1. Поскольку обе оценки состоятельны, полученные средние должны быть близки к результатам регрессии всей выборки и стремиться к ним при росте R .
2. При росте R выборочные дисперсии оценок должны убывать.
3. Вообще говоря, стандартные отклонения оценок LAD должны быть больше стандартных отклонений оценок LS.

4. При малых R оценка LAD должна быть лучше по отношению к соответствующей оценке OLS, чем при больших R , т. е. преимущества LS растут с ростом размера выборки.

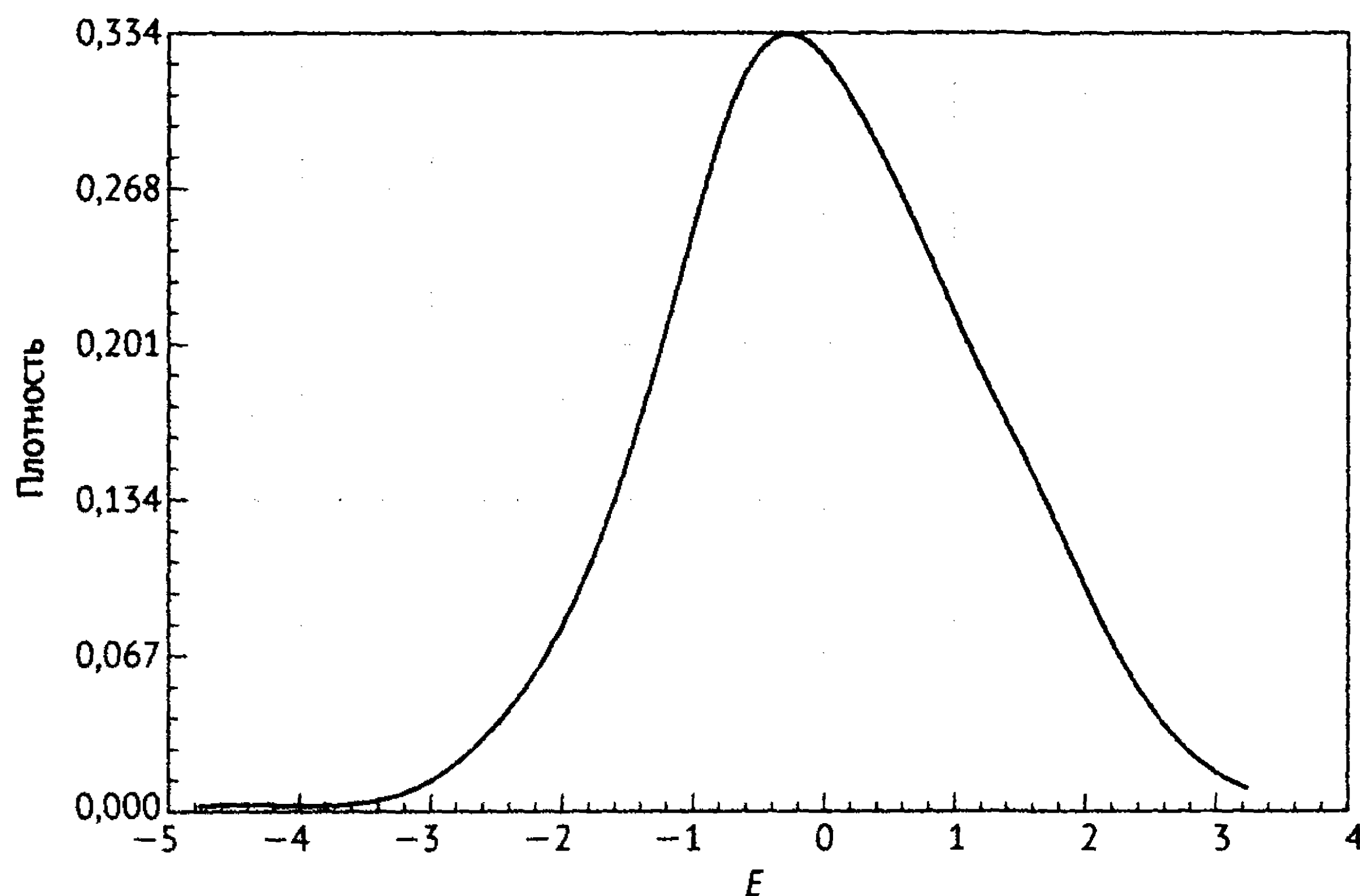


Рис. 4.4. Оценка ядерной плотности для остатков метода наименьших квадратов

Оценка ядерной плотности распределения остатков метода наименьших квадратов показана на рис. 4.4. Наблюдается некоторая скошенность в распределении, и поэтому главное предположение нашего эксперимента может быть не совсем верным. Результаты эксперимента представлены в табл. 4.4. Плюсы асимптотического подхода хорошо видны из столбца коэффициента при $\log Area$. Уменьшение стандартного отклонения по мере роста R является свидетельством состоятельности обеих оценок. Для каждой пары результатов (LS, LAD) видно, что оцененное стандартное отклонение LAD-оценки больше в 1,2–1,4 раза, что и было ожидаемо. При нормальном распределении мы бы ожидали, что отношение будет равно $\sqrt{1,573} = 1,254$.

⁷В методе Монте-Карло используется генератор случайных чисел, поэтому встает вопрос о воспроизводимости результатов. Это исследование было проведено с помощью пакета NLOGIT, и оно может быть воспроизведено. Программу можно скачать с веб-сайта этой книги. Качественно те же результаты, пусть и не с теми же конкретными числовыми значениями, можно получить в любом статистическом пакете, который имеет возможность создания случайной выборки из набора данных.

⁸Обратите внимание, что размер выборки R может быть довольно велик по сравнению с размером массива данных 430 для каждого повторения. Однако нет необходимости корректировать из-за этого дисперсии в табл. 4.4. Мы вычисляем не дисперсию по выборке размера R из генеральной совокупности размера 430, а дисперсию по выборке размера 100 из множества всех выборок размера R . Существует немногим менее 10^2 различных выборок из 10 наблюдений. Количество выборок из 50 или 100 наблюдений, по существу, бесконечно.

4.4.6. Оценка максимального правдоподобия

Мы видели две причины для применения метода наименьших квадратов: во-первых, теорема 4.1 показывает, что оценка эта является выборочным аналогом линейного предиктора y , минимизирующего среднеквадратичное отклонение; во-вторых, теорема 4.2 Гаусса–Маркова показывает, что при выполнении условий модели эта оценка является несмещенной линейной оценкой β , имеющей наименьшую дисперсию. Ни один из этих двух результатов не использовал предположение A_6 (нормальность распределения ε). Возникает естественный вопрос: зачем вообще вводилось это предположение? На то есть две причины. Во-первых, предположение нормальности будет использоваться для построения доверительных интервалов в разделах 4.5 и 4.6. Однако в подразделе 4.4.2 было показано, что ЦПТ позволяет делать статистические выводы исходя из асимптотической нормальности, даже если шоки не распределены нормально. Кажется, что предположение A_6 становится излишним, что было бы во многом верным, если бы не вторая причина.

Если шоки распределены нормально, то оценка наименьших квадратов также является **оценкой максимального правдоподобия** (maximum likelihood estimator, MLE). Оценки максимального правдоподобия будут подробно рассмотрены в главе 14, а здесь будет приведено лишь их краткое описание. В конечном счете мы приходим к тому, что, будучи MLE, оценка LS оказывается *асимптотически эффективной в классе всех состоятельных асимптотически нормальных оценок*. Это утверждение служит аналогом теоремы Гаусса–Маркова для больших выборок и известно как граница Крамера–Рао. Обе эти теоремы схожи тем, что указывают на то, что оценки LS наиболее эффективны среди оценок из некоторого класса; различие в том, о каком именно классе идет речь:

- теорема Гаусса–Маркова: линейные несмещенные оценки;
- MLE: состоятельные и асимптотически нормальные оценки при условии нормальности распределения шоков.

Ни один из этих классов не вложен в другой. Заметим, например, что результат, касающийся MLE, не опирается ни на линейность, ни на несмещенность. Теорема Гаусса–Маркова, наоборот, не требует ни нормальности, ни состоятельности. Теорема Крамера–Рао описывает асимптотические свойства. Важным вопросом исследования является эффективность оценки. Результаты, касающиеся эффективности, позволяют определить, как лучше всего пользоваться имеющимся набором данных для статистических выводов. Вообще говоря, обычно трудно говорить об эффективности оценки без упоминания конкретных альтернатив для сравнения. Теорема Гаусса–Маркова представляет собой важный результат для линейной модели. Однако у нее нет аналогов при любом другом моделировании, поэтому, как только мы отходим от линейной модели, нам требуются другие способы сравнения оценок. Принцип максимального правдоподобия позволяет исследователю определять асимптотическую эффективность оценок, но только при наличии дополнительного предположения о распределениях всех наблюдаемых случайных величин. Пример 4.6 показывает, что \mathbf{b} является MLE-оценкой

в модели регрессии с нормально распределенными шоками. В примере 4.7 рассмотрен случай, когда распределение шоков отлично от нормального и как следствие \mathbf{b} менее эффективна, чем MLE.

Пример 4.6. MLE с нормально распределенными возмущениями

Если шоки распределены нормально, условные распределения $y_i|\mathbf{x}_i$ также нормальны и имеют средние $\mathbf{x}'_i\boldsymbol{\beta}$ и дисперсии σ^2 , так что плотность $y_i|\mathbf{x}_i$ имеет вид

$$f(y_i|\mathbf{x}_i) = \frac{\exp\left[-\frac{1}{2}(y_i - \mathbf{x}'_i\boldsymbol{\beta})^2\right]}{\sqrt{2\pi\sigma^2}}.$$

Логарифм функции правдоподобия, вычисленной по выборке из n независимых наблюдений, равен логарифму плотности совместного распределения наблюдаемых случайных величин. Если выборка случайна, то это будет произведение плотностей каждой из величин, так что логарифм функции правдоподобия по имеющимся данным, обозначаемый как $\ln L(\boldsymbol{\beta}, \sigma^2|\mathbf{y}, \mathbf{X})$, будет равен сумме логарифмов плотностей. После некоторых преобразований мы придем к

$$\ln L(\boldsymbol{\beta}, \sigma^2|\mathbf{y}, \mathbf{X}) = -(n/2)[\ln \sigma^2 + \ln 2\pi + (1/\sigma^2)\frac{1}{n}\sum_i (y_i - \mathbf{x}'_i\boldsymbol{\beta})^2].$$

Значения $\boldsymbol{\beta}$ и σ^2 , максимизирующие эту функцию, являются оценками метода MLE для $\boldsymbol{\beta}$ и σ^2 . Как мы увидим далее в главе 14, функции от данных, максимизирующие эту сумму, будут представлять собой вектор коэффициентов метода наименьших квадратов \mathbf{b} и усредненный квадрат остатков $\mathbf{e}'\mathbf{e}/n$. Также в главе 14 будет получен следующий результат:

$$\text{Asy. Var}[\hat{\boldsymbol{\beta}}_{ML}] = -E[\partial^2 \ln L / \partial \boldsymbol{\beta} \partial \boldsymbol{\beta}']^{-1} = \sigma^2 E[(\mathbf{X}'\mathbf{X})^{-1}],$$

который в точности совпадает с тем, что мы получили в подразделе 4.3.6. Таким образом, оценка наименьших квадратов является оценкой максимального правдоподобия. Она состоятельна, асимптотически (и точно) нормально распределена и при нормальном распределении шоков по теореме 14.4 асимптотически эффективна.

Важно отметить, что свойства оценок MLE зависят от того, каким предполагается распределение наблюдаемых случайных величин. Если предположить распределение ε , отличное от нормального, то \mathbf{b} не будет оценкой MLE и может не быть эффективной. Рассмотрим следующий пример.

Пример 4.7. Модель гамма-регрессии

В работе Green (1980a) рассматривается оценивание регрессионной модели с асимметрично распределенными шоками:

$$y = (\alpha + \sigma\sqrt{P}) + \mathbf{x}'\boldsymbol{\beta} + (\varepsilon - \sigma\sqrt{P}) = \alpha^* + \mathbf{x}'\boldsymbol{\beta} + \varepsilon^*,$$

где ε имеет гамма-распределение (как в разделе В4.5 приложения [см. (В-39)]), а $\sigma = \sqrt{P}/\lambda$ — это стандартное отклонение шоков. В этой модели ковариационная матрица для оценки коэффициентов методом LS (не включая константу) равна

$$\text{Asy. Var}[\mathbf{b} | \mathbf{X}] = \sigma^2(\mathbf{X}'\mathbf{M}^0\mathbf{X})^{-1},$$

а для оценки MLE (отличной от LS)⁹—

$$\text{Asy. Var}[\hat{\beta}_{ML}] \approx [1 - (2/P)]\sigma^2(\mathbf{X}'\mathbf{M}^0\mathbf{X})^{-1}.$$

За исключением параметра, соответствующего асимметрии, результат совпадает с аналогом для метода наименьших квадратов. Однако нами показывается, что эта оценка, учитывающая асимметрию, более эффективна асимптотически.

На эту модель немного похожа модель стохастической границы (stochastic frontier model), которая будет рассмотрена в главе 18. В обеих этих моделях распределение шоков асимметрично. Оценки MLE вычислены способом, учитывающим этот момент, в то время как оценка наименьших квадратов не различает симметричных наблюдений выше и ниже линии регрессии. Это различие и дает большую эффективность оценок MLE в обеих моделях.

4.5. Интервальные оценки

Цель получения интервальных оценок состоит в том, чтобы вместе с оценкой параметра еще и явно показывать степень неопределенности этой оценки. Обычно результатом интервального оценивания параметра θ оказывается выражение вида

$$\hat{\theta} \pm \text{возможное выборочное отклонение.} \quad (4-37)$$

(Здесь мы предполагаем, что интересующий нас интервал симметричен относительно $\hat{\theta}$.) Поскольку величина возможного выборочного отклонения, вообще говоря, должна отражать степень неуверенности в значении параметра, рассмотрим естественные ограничения. Например, мы можем быть абсолютно уверены, что истинное значение лежит в интервале $\hat{\theta} \pm \infty$. Это, разумеется, не слишком информативное ограничение. Другое очевидное ограничение состоит в том, что наша уверенность в том, что истинное значение параметра лежит в $\hat{\theta} \pm 0$, равна 0%. В самом деле вероятность того, что наша оценка в точности совпадет с оцениваемым значением, равна нулю. Нужно выбрать какой-то конкретный уровень α — обычно берут 0,05 или 0,01 — и построить интервал вида (4-13), такой, что истинное значение параметра лежит в нем с вероятностью (уровнем доверия) $100(1 - \alpha)$. Далее

⁹С помощью матрицы данные приводятся к отклонениям от выборочного среднего. (См. раздел А.2.8 приложения.) В этой модели P должно быть больше 2.

мы рассмотрим процедуру построения такого интервала, а потом применим ее для решения трех часто встречающихся задач — построения интервальной оценки параметра линейной регрессии, построения оценки функции от параметров модели и предсказания значения зависимой переменной при заданных значениях независимых переменных. Вначале мы будем исходить из предположения А6 о независимости распределений шоков, затем несколько ослабим это требование и будем полагаться на асимптотическую нормальность оценок.

4.5.1. Построение доверительного интервала для коэффициента линейной регрессии

Из (4-18) следует, что $\mathbf{b}|\mathbf{X} \sim N[\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}]$. Отсюда вытекает, что для любого элемента \mathbf{b} , скажем b_k , верно

$$b_k \sim N[\beta_k, \sigma^2 S^{kk}],$$

где S^{kk} — k -й диагональный элемент матрицы $(\mathbf{X}'\mathbf{X})^{-1}$. Нормируя эту величину, мы получаем, что

$$z_k = \frac{b_k - \beta_k}{\sqrt{\sigma^2 S^{kk}}} \quad (4-38)$$

имеет стандартное нормальное распределение. Заметим, что z_k , являясь функцией от b_k, β_k, σ^2 и S^{kk} , тем не менее имеет распределение, не зависящее от параметров модели или наблюдаемых данных: z_k является **базисной (или основной) статистикой** (pivotal statistic). Если нам нужно построить 95%-й доверительный интервал, то необходимо воспользоваться тем, что $\text{Prob}[-1,96 \leq z_k \leq 1,96]$. С помощью простых преобразований получаем

$$\text{Prob}\left[b_k - 1,96\sqrt{\sigma^2 S^{kk}} \leq \beta_k \leq b_k + 1,96\sqrt{\sigma^2 S^{kk}}\right] = 0,95. \quad (4-39)$$

Заметим, что здесь определяется вероятность того, что случайный интервал $b_k \pm$ возможное выборочное отклонение содержит β_k , а не вероятность того, что β_k лежит в конкретном интервале. Если нужно получить интервал с доверительной вероятностью α , отличной от 95%, то нужно заменить 1,96 в (4-39) на $z_{(1-\alpha/2)}$. (Через $z_{(1-\alpha/2)}$ мы обозначаем такое число, что для случайной величины z , имеющей стандартное нормальное распределение, $\text{Prob}[z \geq z_{(1-\alpha/2)}] = 1 - \alpha/2$. Таким образом, $z_{0,975} = 1,96$, что соответствует $\alpha = 0,05$.)

Требуемый доверительный интервал был получен в (4-39), однако величина σ^2 неизвестна, поэтому воспользоваться этим интервалом нельзя. Кажется естественным воспользоваться здесь полученной нами ранее оценкой s^2 . В самом деле величина

$$\frac{(n-K)s^2}{\sigma^2} = \frac{\mathbf{e}'\mathbf{e}}{\sigma^2} = \left(\frac{\boldsymbol{\varepsilon}}{\sigma}\right)' \mathbf{M} \left(\frac{\boldsymbol{\varepsilon}}{\sigma}\right) \quad (4-40)$$

является идемпотентной квадратичной формой от стандартного нормального вектора $(\boldsymbol{\varepsilon}/\sigma)$. Значит, она имеет распределение хи-квадрат с числом степеней свободы, равным $\text{rank}(\mathbf{M}) = \text{trace}(\mathbf{M}) = n - K$. (См. доказательство этого факта в разделе В11.4.) Случайная величина хи-квадрат в (4-40)

и нормально распределенная случайная величина в (4-38) являются независимыми. Чтобы это доказать, достаточно показать, что

$$\left(\frac{\mathbf{b} - \boldsymbol{\beta}}{\sigma}\right) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\frac{\boldsymbol{\varepsilon}}{\sigma}\right)$$

не зависит от $(n - K)s^2/\sigma^2$. В разделе В.11.7 (теорема В.12) показывается, что достаточным условием независимости линейной формы $\mathbf{L}\mathbf{x}$ и идемпотентной квадратичной формы $\mathbf{x}'\mathbf{A}\mathbf{x}$ от стандартного нормального вектора \mathbf{x} является равенство $\mathbf{L}\mathbf{A} = \mathbf{0}$. Принимая $\boldsymbol{\varepsilon}/\sigma$ за \mathbf{x} , для нашего случая получаем условие $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{M} = \mathbf{0}$. Это условие выполняется, как мы видели в (3-15). Следующий общий результат является основным при выводе многих тестовых статистик в регрессионном анализе.

Теорема 4.6. Независимость \mathbf{b} и s^2

Если величина $\boldsymbol{\varepsilon}$ распределена нормально, то оценка коэффициентов \mathbf{b} , полученная с помощью метода наименьших квадратов, и вектор остатков \mathbf{e} (а следовательно, и функции этого вектора, в том числе s^2) распределены независимо.

Таким образом, отношение

$$t_k = \frac{(b_k - \beta_k)/\sqrt{\sigma^2 S^{kk}}}{\sqrt{[(n - K)s^2/\sigma^2]/(n - K)}} = \frac{b_k - \beta_k}{\sqrt{s^2 S^{kk}}} \quad (4-41)$$

имеет t -распределение с $(n - K)$ степенями свободы¹⁰. Мы можем воспользоваться статистикой t_k для проверки гипотез или построения доверительных интервалов для отдельных элементов $\boldsymbol{\beta}$.

Результат (4-41) отличается от (4-38) тем, что вместо σ^2 используется s^2 , а также базисным распределением статистики, имеющей t -распределение с $(n - K)$ степенями свободы вместо стандартного нормального. Таким образом, доверительный интервал для β_k можно построить как

$$\text{Prob}\left[b_k - t_{(1-\alpha/2), [n-K]} \sqrt{s^2 S^{kk}} \leq \beta_k \leq b_k + t_{(1-\alpha/2), [n-K]} \sqrt{s^2 S^{kk}}\right] = 1 - \alpha, \quad (4-42)$$

где $t_{(1-\alpha/2), [n-K]}$ — соответствующее критическое значение t -распределения. Здесь распределение пивотальной статистики зависит от размера выборки (через число степеней свободы $(n - K)$), но вновь не зависит от данных. Преимущество равенства (4-42) в том, что в нем не используются неизвестные параметры. На его основе можно построить доверительный интервал для β_k .

¹⁰См. (В-36) в разделе В.4.2. Это есть отношение стандартной нормально распределенной величины к квадратному корню из случайной величины, имеющей распределение χ^2 , деленной на число ее степеней свободы.

Пример 4.8. Доверительный интервал для эластичности спроса на бензин по доходу

Используя данные о рынке бензина из примеров 4.2 и 4.4, мы оценили следующее уравнение спроса (по 52 наблюдениям):

$$\ln(G/Pop) = \beta_1 + \beta_2 \ln P_G + \beta_3 \ln(Income/Pop) + \beta_4 \ln P_{nc} + \beta_5 \ln P_{uc} + \varepsilon.$$

OLS-оценки параметров модели, стандартные ошибки и t -статистики приведены в табл. 4.5.

Таблица 4.5. Результаты регрессии уравнения спроса

| Сумма квадратов остатков: | 0,120871 | | |
|-------------------------------|-------------|--------------------|-----------------|
| Стандартная ошибка регрессии: | 0,050712 | | |
| R^2 | 0,958443 | | |
| Переменная | Коэффициент | Стандартная ошибка | t -статистика |
| Константа | -21,21109 | 0,75322 | -28,160 |
| $\ln P_G$ | -0,021206 | 0,04377 | -0,485 |
| $\ln Income/Pop$ | 1,095874 | 0,07771 | 14,102 |
| $\ln P_{nc}$ | -0,373612 | 0,15707 | -2,379 |
| $\ln P_{uc}$ | 0,02003 | 0,10330 | 0,194 |

Чтобы построить доверительный интервал для эластичности по доходу, нам необходимо критическое значение t -распределения с $n - K = 52 - 5 = 47$ степенями свободы. 95%-е критическое значение равно 2,012. Следовательно, 95%-й доверительный интервал для $\beta_3 = 1,095874 \pm 2,012(0,07771) = [0,9395; 1,2522]$.

4.5.2. Построение доверительных интервалов для больших выборок

Если распределение шоков не является нормальным, то подход, использованный в предыдущем разделе и основанный на этом предположении, не может быть применен. Однако асимптотические результаты, полученные в разделе 4.4, позволяют применить альтернативный подход. Основываясь на нашем выводе теоремы 4.4 и (4-35), мы можем утверждать, что предельным распределением статистики

$$z_n = \frac{\sqrt{n}(b_k - \beta_k)}{\sqrt{\sigma^2 Q^{kk}}}$$

является стандартное нормальное распределение, где $\mathbf{Q} = [\text{plim}(\mathbf{X}'\mathbf{X}/n)]^{-1}$, а Q^{kk} — k -й диагональный элемент матрицы \mathbf{Q} . По теореме Слуцкого (D.16) σ^2 можно заменить ее состоятельной оценкой s^2 и получить статистику с тем же предельным распределением. И разумеется, мы оцениваем \mathbf{Q} как $(\mathbf{X}'\mathbf{X}/n)^{-1}$. Отсюда получаем в точности (4-41), т. е. при соблюдении условий из раздела 4.4 t -статистика в (4-41) сходится к стандартной нормальной, да-

же если распределение шоков не является нормальным. Значит, чтобы воспользоваться свойствами асимптотического распределения \mathbf{b} , нужно строить доверительные интервалы в соответствии с (4-42), но при этом пользоваться критическими значениями стандартного нормального распределения (например, 1,96), а не t -распределения. На практике если число степеней свободы в (4-42) велико, например больше 100, то t -распределение неотличимо от стандартного нормального, так что в любом случае можно пользоваться асимптотическим результатом. При меньших размерах выборки для большей осторожности на практике можно посоветовать использовать критические значения t -статистики (а не стандартного нормального распределения) даже при отсутствии предположения нормальности распределения возмущений. В случае примера 4.8, где выборка состояла из 52 наблюдений, мы строили доверительный интервал для эластичности по доходу, используя критическое значение 2,012 из таблицы распределения для случайной величины, распределенной по t -закону с 47 степенями свободы. Если бы мы решили воспользоваться асимптотическим результатом, то нам следовало бы воспользоваться критическим значением стандартного нормального распределения 1,96. Осторожный исследователь, впрочем, счел бы такую оценку излишне оптимистичной и выбрал бы значение 2,012.

Пример 4.9. *Доверительный интервал, основанный на аксиоматическом подходе*

В примере 4.4 мы рассматривали уравнения спроса на бензин в динамическом виде:

$$\ln(G/Pop)_t = \beta_1 + \beta_2 \ln P_{G,t} + \beta_3 \ln(Income/Pop) + \dots + \gamma \ln(G/POP)_{t-1} + \varepsilon_t.$$

В этой модели долгосрочные цена и эластичность по доходу равны $\theta_P = \beta_2/(1-\gamma)$ и $\theta_I = \beta_3/(1-\gamma)$. Мы построили оценки этих нелинейных функций с помощью метода наименьших квадратов и дельта-метода (теорема 4.5). Полученные точечные оценки равны $-0,411358$ и $0,970522$ соответственно. Оценки стандартных отклонений асимптотического распределения — $0,152296$ и $0,162386$. Для построения доверительных интервалов для θ_P и θ_I мы воспользуемся асимптотическим, а не конечно-выборочным распределением. Тогда доверительные интервалы имеют вид

$$\hat{\theta}_P = -0,411358 \pm 1,96(0,152296) = [-0,709858; -0,112858]$$

и

$$\hat{\theta}_I = 0,970523 \pm 1,96(0,162386) = [0,652246; 1,288800].$$

При размере выборки, равном 51, можно возразить, что использование критического значения предельного нормального распределения излишне оптимистично. В этом случае использование критического значения для t -распределения с $51 - 6 = 45$ степенями свободы даст несколько более широкий интервал. Например, для эластичности по доходу получится интервал $0,970523 \pm 2,014(0,162386) = [0,643460; 1,297585]$. Заметим, однако, что это чисто практическая поправка. На самом деле статистика,

основанная на *асимптотическом* стандартном распределении, не будет иметь *t*-распределение с 45 степенями свободы.

4.5.3. Доверительные интервалы для линейных комбинаций коэффициентов: разложение Охака

При условии нормального распределения отклонений оценка методом наименьших квадратов \mathbf{b} распределена нормально со средним β и ковариационной матрицей $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. В примере 4.8 было показано, как можно воспользоваться этим фактом для построения доверительного интервала для одного из элементов β . Этот результат можно расширить при построении доверительного интервала для линейной функции параметров. Для этого часто используют **разложение Охака и Блаиндера** (Оахаса (1973) и Blinder (1973))¹¹.

Пусть \mathbf{w} — вектор известных констант размером $K \times 1$. Тогда линейная комбинация $c = \mathbf{w}'\mathbf{b}$ распределена нормально и имеет среднее $\gamma = \mathbf{w}'\beta$ и дисперсию $\sigma_c^2 = \mathbf{w}'[\sigma^2(\mathbf{X}'\mathbf{X})^{-1}]\mathbf{w}$, которую можно оценить как $s_c^2 = \mathbf{w}'[s^2(\mathbf{X}'\mathbf{X})^{-1}]\mathbf{w}$. Имея эти величины, можно, воспользовавшись полученными результатами, построить доверительный интервал для γ :

$$\text{Prob}[c - t_{(1-\alpha/2), [n-k]}s_c \leq \gamma \leq c + t_{(1-\alpha/2), [n-k]}s_c] = 1 - \alpha. \quad (4-43)$$

Этот общий результат можно использовать, например, для суммы или разности коэффициентов.

Рассмотрим пример из работы Охака (1973). При исследовании предложения на рынке труда были отдельно оценены регрессии для выборок из n_m мужчин и n_f женщин. Использовались уравнения регрессий

$$\ln \text{wage}_{m,i} = \mathbf{x}'_{m,i}\beta_m + \varepsilon_{m,i}, \quad i = 1, \dots, n_m$$

и

$$\ln \text{wage}_{f,j} = \mathbf{x}'_{f,j}\beta_f + \varepsilon_{f,j}, \quad j = 1, \dots, n_f.$$

В число регрессоров входили социодемографические переменные (такие, как, например, возраст) и переменные, описывающие человеческий капитал (такие, как образование, опыт работы). Мы заинтересованы в сравнении этих двух регрессий, а в особенности в наличии или отсутствии дискриминации при определении зарплат. Охака предложил подход к сравнению этих регрессий. Для двух любых векторов характеристик

$$\begin{aligned} E[\ln \text{wage}_{m,i} | \mathbf{x}_{m,i}] - E[\ln \text{wage}_{f,j} | \mathbf{x}_{f,i}] &= \mathbf{x}'_{m,i}\beta_m - \mathbf{x}'_{f,j}\beta_f \\ &= \mathbf{x}'_{m,i}\beta_m - \mathbf{x}'_{m,i}\beta_f + \mathbf{x}'_{m,i}\beta_f - \mathbf{x}'_{f,j}\beta_f \\ &= \mathbf{x}'_{m,i}(\beta_m - \beta_f) + (\mathbf{x}_{m,i} - \mathbf{x}_{f,j})'\beta_f. \end{aligned}$$

Второй член этого разложения связан с различиями в человеческом капитале и является естественным объяснением различий в уровне заработной

¹¹См., напр., Bourignon et al. (2002).

платы, если рынок труда реагирует на эти характеристики так, как мы обычно предполагаем. Первый член показывает разницу в логарифмах заработной платы, не объясняемую разницей в человеческом капитале, т. е. если зафиксировать эти характеристики на уровне x_m , то первый член будет отражать влияние других факторов. Охака предложил вычислить это разложение в средних векторов регрессоров, \bar{x}_m и \bar{x}_f , и OLS-оценках коэффициентов, b_m и b_f . Если в уравнения регрессий входят постоянные, то эти действия эквивалентны рассмотрению $\ln y_m - \ln y_f$.

Нас интересует доверительный интервал для первого члена разложения, при построении которого нужно дважды применить наш результат. Будем считать два вектора выборочных средних известными. Предполагая, что две имеющиеся выборки независимы, наши оценки, b_m и b_f , независимы и имеют средние β_m и β_f и ковариационные матрицы $\sigma_m^2(\mathbf{X}'_m \mathbf{X}_m)^{-1}$ и $\sigma_f^2(\mathbf{X}'_f \mathbf{X}_f)^{-1}$. Ковариационная матрица разности равна сумме этих двух матриц. Мы построим доверительный интервал для $\bar{x}'_m d$, где $d = b_m - b_f$. Оцененная ковариационная матрица равна

$$\text{Est. Var}[d] = s_m^2(\mathbf{X}'_m \mathbf{X}_m)^{-1} + s_f^2(\mathbf{X}'_f \mathbf{X}_f)^{-1}. \quad (4-44)$$

Теперь можно применить полученный ранее результат, а также построить доверительный интервал для второго члена: нужно взять $w = \bar{x}_m - \bar{x}_f$ и применить тот же результат к величине $w'b_f$.

4.6. Предсказание и прогнозирование

После оценки параметров модели ее часто применяют для предсказания значений зависимой переменной. Мы будем различать предсказание и прогнозирование, основываясь на различии между межобъектными данными и временными рядами. **Предсказание** (как для межобъектных данных, так и для временных рядов) использует модель регрессии для вычисления подобранных (fited) (или предсказанных) значений зависимой переменной для наблюдений как из выборки, так и вне ее. Результаты, касающиеся предсказания, будут применимы к межобъектным и панельным данным, а также к временным рядам. **Прогнозирование** отличается тем, что временной структуре данных будет отводиться особая роль, причем зачастую будут использоваться лагированные зависимые переменные и шоки, скоррелированные с шоками предыдущих временных периодов. Здесь обычно будут прогнозироваться будущие значения зависимой переменной. Важная разница между предсказанием и прогнозированием (в нашем понимании значений этих слов) заключается в том, что при предсказании обычно исследуется сценарий, определенный исследователем. Например, в приведенном ниже примере мы будем предсказывать цену картин Моне; при этом мы можем пытаться предсказать цену как гипотетической картины по заданным размеру и отношению сторон, так и картины из имеющейся выборки. При изучении временных рядов мы обычно пытаемся спрогнозировать событие (например, реальный объем инвестиций в следующем году), основываясь не на гипотетическом сценарии развития экономики, а на наилучшей имеющейся оценке экономической обстановки в следующем году. Мы будем

пользоваться термином **предсказание ex post** (или **прогноз ex post**) в случаях, когда данные, используемые в уравнении регрессии при построении предсказания (или прогноза), либо наблюдаемы, либо построены экспериментально. Это будет первой из рассмотренных нами в этом разделе ситуаций. **Прогноз ex ante** (в контексте временных рядов) требует предварительной оценки будущих значений независимых переменных. В одном из примеров этой главы объем реальных инвестиций прогнозируется с помощью модели, в которую входят реальный ВВП и индекс цен. Чтобы спрогнозировать реальные инвестиции в этой модели, нужно сначала построить прогноз для реального ВВП и индекса цен. Прогнозирование ex ante будет кратко рассмотрено в этой главе, а затем более подробно в главе 20.

4.6.1. Доверительные интервалы для предсказаний

Пусть нам нужно предсказать значение y^0 по вектору регрессоров \mathbf{x}^0 . Фактическое значение имеет вид

$$y^0 = \mathbf{x}^{0'}\boldsymbol{\beta} + \varepsilon^0.$$

Из теоремы Гаусса–Маркова получаем, что

$$\hat{y}^0 = \mathbf{x}^{0'}\mathbf{b} \quad (4-45)$$

является линейной несмещенной оценкой с минимальной дисперсией для $E[y^0|\mathbf{x}^0] = \mathbf{x}^{0'}\boldsymbol{\beta}$. **Ошибка предсказания** равна

$$e^0 = \hat{y}^0 - y^0 = (\mathbf{b} - \boldsymbol{\beta})'\mathbf{x}^0 + \varepsilon^0.$$

Дисперсия ошибки этой оценки равна

$$\text{Var}[e^0|\mathbf{X}, \mathbf{x}^0] = \sigma^2 + \text{Var}[(\mathbf{b} - \boldsymbol{\beta})'\mathbf{x}^0|\mathbf{X}, \mathbf{x}^0] = \sigma^2 + \mathbf{x}^{0'}[\sigma^2(\mathbf{X}'\mathbf{X})^{-1}]\mathbf{x}^0. \quad (4-46)$$

Если уравнение регрессии содержит константу, то это эквивалентно

$$\text{Var}[e^0|\mathbf{X}, \mathbf{x}^0] = \sigma^2 \left[1 + \frac{1}{n} + \sum_{j=1}^{K-1} \sum_{k=1}^{K-1} (x_j^0 - \bar{x}_j)(x_k^0 - \bar{x}_k)(\mathbf{Z}'\mathbf{M}^0\mathbf{Z})^{jk} \right], \quad (4-47)$$

где \mathbf{Z} — $K - 1$ столбцов \mathbf{X} без константы, $\mathbf{Z}'\mathbf{M}^0\mathbf{Z}$ — матрица сумм квадратов и произведений для столбцов матрицы \mathbf{X} , выраженных в отклонениях от их средних [см. (3-21)], а верхний индекс jk обозначает элемент jk обратной матрицы. Из этого результата следует, что ширина доверительного интервала для предсказания зависит от расстояния между элементами вектора \mathbf{x}^0 и «центром» данных. Интуитивно это достаточно ясно: чем более удалены новые данные от нашего накопленного опыта, тем меньше точность наших предсказаний. Рисунок 4.5 демонстрирует этот эффект для случая парной регрессии. Обратим внимание, что дисперсия предсказания состоит из трех частей. Вторая и третья уменьшаются при увеличении выборки (т.е. при росте n). Но первый член, σ^2 , является постоянным: это значит, что никакой объем данных не даст абсолютно точного предсказания.

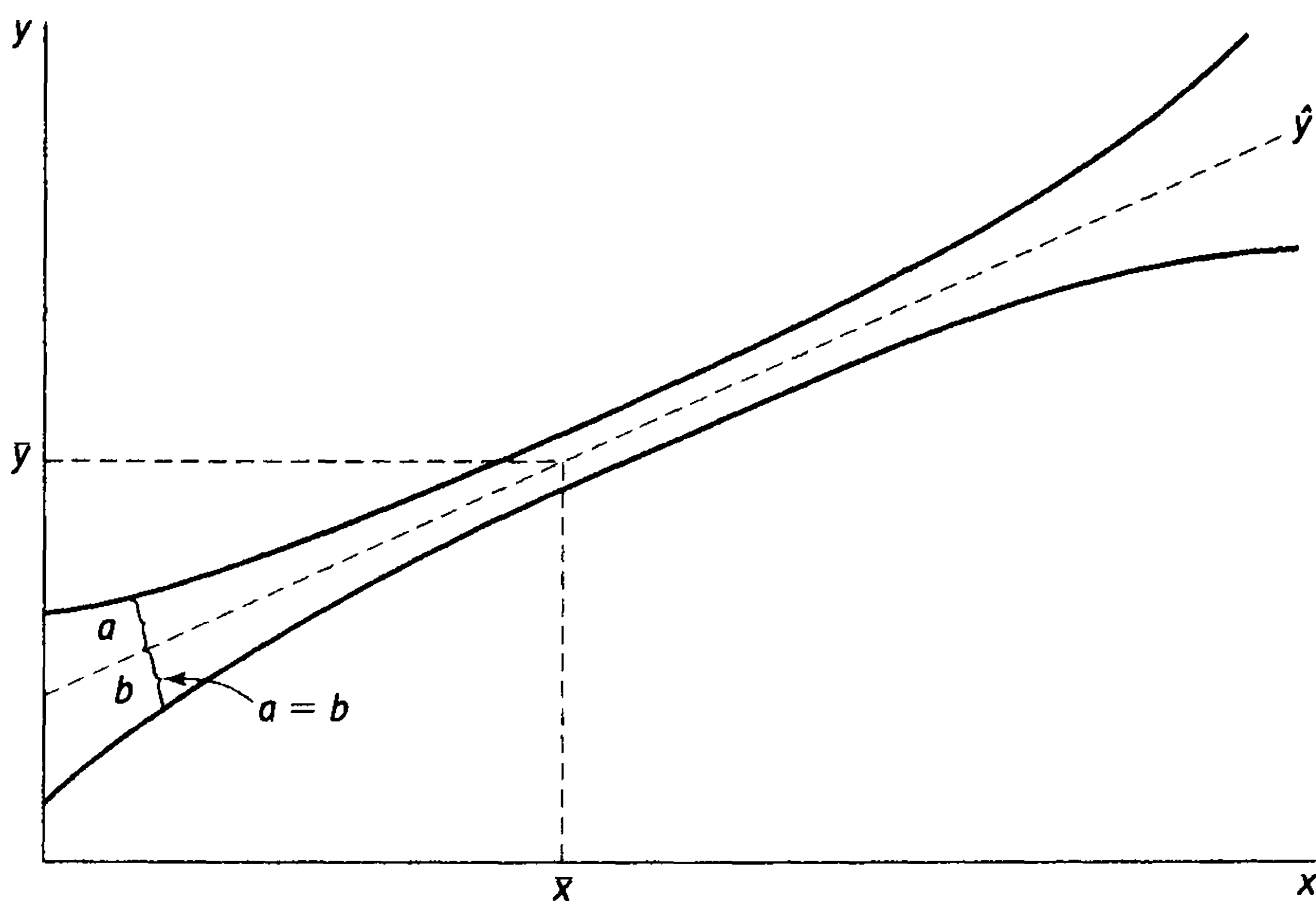


Рис. 4.5. Доверительные интервалы для предсказаний

Дисперсию предсказаний можно оценить, используя s^2 вместо σ^2 . Доверительный интервал для предсказания y^0 может быть построен как

$$\text{интервал предсказания} = \hat{y}^0 \pm t_{(1-\alpha/2), [n-K]} se(e^0), \quad (4-48)$$

где $t_{(1-\alpha/2), [n-K]}$ — подходящее критическое значение для уровня значимости $100(1-\alpha)\%$ из таблицы для t -распределения с $n-K$ степенями свободы, а $se(e^0)$ — квадратный корень оценки дисперсии предсказания.

4.6.2. Предсказание y , если уравнение регрессии описывает логарифм y

Часто используют регрессионные модели, описывающие не саму переменную, а некоторую функцию от нее. В примере 4.5 мы описывали цены на картины Моне с помощью уравнения

$$\ln Price = \beta_1 + \beta_2 \ln Area + \beta_3 AspectRatio + \varepsilon.$$

(Area — произведение ширины и высоты картины, а AspectRatio — высота, деленная на ширину.) Логарифмическая форма удобна тем, что коэффициент равен эластичности зависимой переменной по независимой, т. е. в этой модели $\beta_2 = \partial E[\ln Price | \ln Area, AspectRatio] / \partial \ln Area$. Однако уравнение в такой форме хуже подходит для целей предсказания, чем простая регрессия цены. Естественно было бы взять предиктор для

$$\ln y^0 = \mathbf{x}^0 \mathbf{b}$$

в виде

$$\hat{y}^0 = \exp(\mathbf{x}^0 \mathbf{b}).$$

Проблема, однако, состоит в том, что $E[y|\mathbf{x}^0]$ не равно $\exp(E[\ln y|\mathbf{x}^0])$. Правильная функция условного математического ожидания здесь будет равна

$$\begin{aligned} E[y|\mathbf{x}^0] &= E[\exp(\mathbf{x}^{0'}\boldsymbol{\beta} + \varepsilon^0)|\mathbf{x}^0] = \\ &= \exp(\mathbf{x}^{0'}\boldsymbol{\beta})E[\exp(\varepsilon^0)|\mathbf{x}^0]. \end{aligned}$$

Второй член, вообще говоря, не равен $\exp(E[\varepsilon^0|\mathbf{x}^0]) = 1$. В случае если $\varepsilon^0|\mathbf{x}^0$ распределены нормально с нулевым средним дисперсией σ^2 , точный результат равен $E[\exp(\varepsilon^0)|\mathbf{x}^0] = \exp(\sigma^2/2)$. (См. раздел В4.4.) При нормальном распределении возмущений получаем предсказание для условного математического ожидания:

$$\hat{y}^0 = \exp(\mathbf{x}^{0'}\mathbf{b} + s^2/2) > \exp(\mathbf{x}^{0'}\mathbf{b}), \quad (4-49)$$

откуда видно, что наш наивный метод будет систематически занижать (недооценивать) y . Однако это не единственная возможная интерпретация этого результата. Из неравенства следует, что наивный метод будет систематически занижать условное математическое ожидание, но необязательно сами значения переменной. Интересный вопрос состоит в том, является ли условное математическое ожидание хорошим предиктором для экспоненты зависимой переменной логарифмической регрессии. Условная медиана, возможно, представляет больший интерес, особенно в случае финансовых переменных, таких как доход, расходы или цена картины. Если распределение переменной в логарифмической регрессии симметрично (как в случае нормального распределения возмущений), то экспонента будет иметь «толстый хвост» в положительном направлении, а среднее будет больше медианы, причем, возможно, намного. В этих случаях медиана является предпочтительной оценкой для центра распределения. Для того чтобы оценить медиану, а не среднее, нужно вернуться к наивному предиктору $\hat{y}^0 = \exp(\mathbf{x}^{0'}\mathbf{b})$.

Итак, с учетом вышесказанного попробуем оценить $E[\exp(y)|\mathbf{x}^0]$. Если мы хотим обойтись без предположения о нормальности, то остается определить, что использовать в качестве $E[\exp(\varepsilon^0)|\mathbf{x}^0]$. В работе Duan (1983) была предложена состоятельная оценка (основанная на предположении о постоянстве математического ожидания, т.е. о гомоскедастичности регрессии)

$$\hat{E}[\exp(\varepsilon^0)|\mathbf{x}^0] = h^0 \frac{1}{n} \sum_{i=1}^n \exp(e_i), \quad (4-50)$$

где e_i — остатки оценки OLS в первоначальной регрессии (с логарифмом). Предложенная **размытая оценка** для предсказания y^0 имеет вид

$$\hat{y}^0 = h^0 \exp(\mathbf{x}^{0'}\mathbf{b}).$$

4.6.3. Доверительный интервал для предсказания y в случаях, когда уравнение регрессии описывает логарифм y

В (4-48) мы построили доверительный интервал для предсказания $\ln y|\mathbf{x}^0$ в логлинейной модели $\ln y = \mathbf{x}'\boldsymbol{\beta} + \varepsilon$:

$$[\ln \hat{y}_{LOWER}^0, \ln \hat{y}_{UPPER}^0] \left[\mathbf{x}^{0'}\mathbf{b} - t_{(1-\alpha/2), [n-K]} se(e^0), \mathbf{x}^{0'}\mathbf{b} + t_{(1-\alpha/2), [n-K]} se(e^0) \right].$$

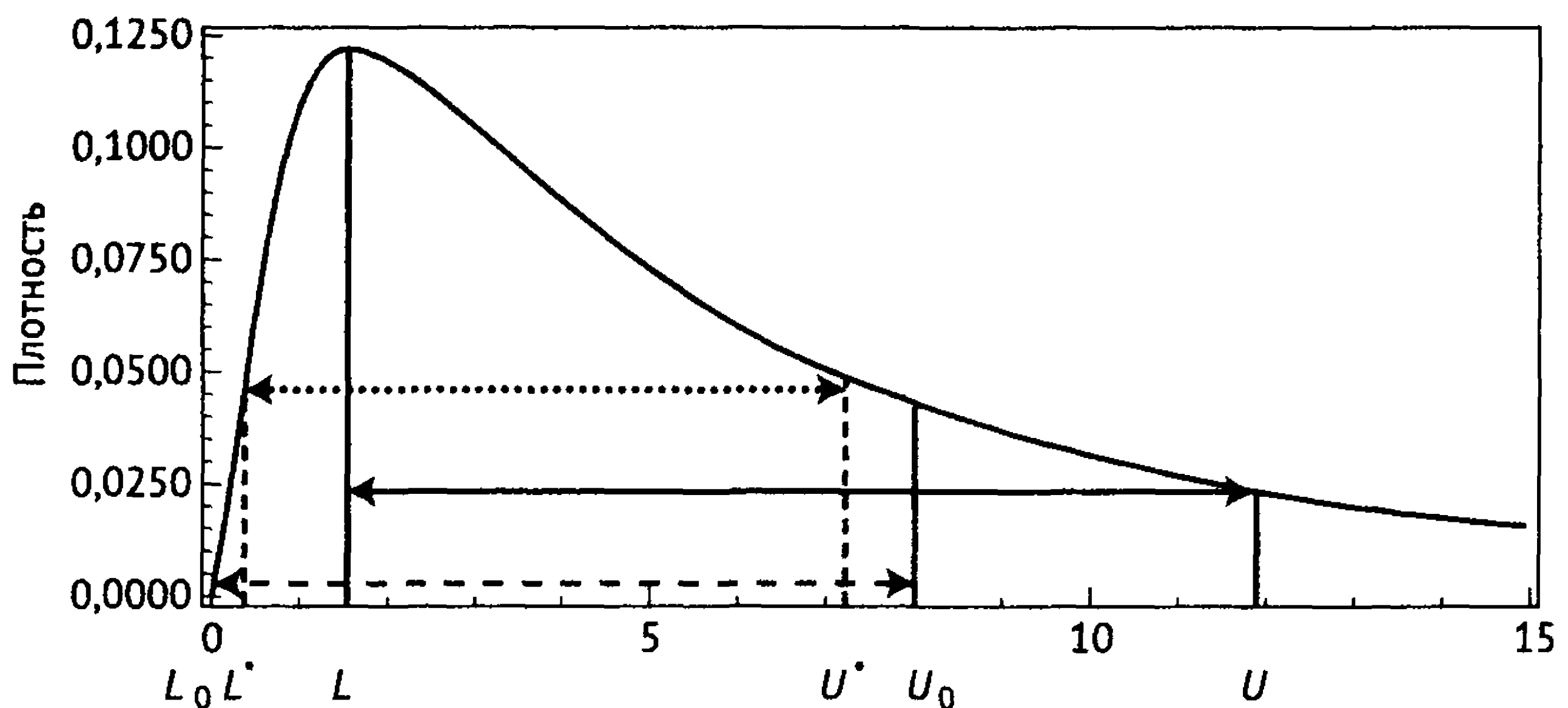


Рис. 4.6. Логнормальное распределение цен картин Моне

Для заданного α , равного, скажем, 0,05, эти значения дают 0,025 и 0,975 квантили распределения $\ln y | \mathbf{x}^0$. Если мы хотим оценить эти квантили для распределения $y | \mathbf{x}^0$, а не $\ln y | \mathbf{x}^0$, то нужно воспользоваться

$$[\hat{y}_{LOWER}^0, \hat{y}_{UPPER}^0] = \left\{ \exp \left[\mathbf{x}^{0'} \mathbf{b} - t_{(1-\alpha/2), [n-K]} se(e^0) \right], \right. \\ \left. \exp \left[\mathbf{x}^{0'} \mathbf{b} + t_{(1-\alpha/2), [n-K]} se(e^0) \right] \right\} \quad (4-51)$$

Это вытекает из следующего факта: если $\text{Prob}[\ln y \leq \ln L] = 1 - \alpha/2$, то $\text{Prob}[y \leq L] = 1 - \alpha/2$. Это означает, что естественная оценка может быть использована при оценке квантилей распределения исходной переменной. Однако если требуется построить как можно более узкую интервальную оценку $y | \mathbf{x}^0$, то этот подход не является оптимальным. Если распределение y асимметрично, как, например, в случае логлинейной модели с нормально распределенными шоками, то наивная интервальная оценка несколько шире, чем надо. На рис. 4.6 видно, из-за чего возникает эта ситуация. Пусть (L, U) — доверительный интервал предсказания, построенный по (4-51). Тогда вероятности слева от L и справа от U равны $\alpha/2$. Рассмотрим интервал с концами $L_0 = 0$ и U_0 . По построению площадь (т. е. вероятность) между L_0 и L равна площади между U_0 и U . Но поскольку плотность намного выше в L , расстояние $(0, U_0)$, длина штрихованного интервала, явно меньше длины (L, U) . Сумма вероятностей «хвостов» все равно равна α , так что мы получаем более узкий доверительный интервал для предсказания. (4-51) можно еще улучшить, взяв $(0, U_0)$, где U_0 равно $\exp[\mathbf{x}^{0'} \mathbf{b} + t_{(1-\alpha), [n-K]} se(e^0)]$ (т. е. мы берем только правый «хвост»). Однако, несмотря на то что это улучшение, оно, как мы сейчас покажем, не является наилучшим.

Рассмотрим задачу нахождения наиболее узкого доверительного интервала для предсказания. Рассмотрим это как задачу оптимизации:

$$\text{Minimize } (L, U) : I = U - L \text{ при ограничении } F(L) + [1 - F(U)] = \alpha,$$

где F — функция распределения случайной величины y (не $\ln y$). Таким образом, мы ищем кратчайший интервал, такой что сумма вероятностей «хво-

стов» равна желаемому α (обычно 0,05). Запишем лагранжиан:

$$\text{Minimize } (L, U, \lambda) : I^* = U - L + \lambda[F(L) + (1 - F(U)) - \alpha].$$

Найдем решения, приравнивая частные производные к нулю:

$$\begin{aligned} \partial I^* / \partial L &= -1 + \lambda f(L) = 0, \\ \partial I^* / \partial U &= 1 - \lambda f(U) = 0, \\ \partial I^* / \partial \lambda &= F(L) + [1 - F(U)] - \alpha = 0, \end{aligned}$$

где $f(L) = F'(L)$ и $f(U) = F'(U)$ — производные функции распределения, т. е. плотности распределения в L и U соответственно. Третье уравнение описывает ограничение на сумму вероятностей «хвостов», но не требует их равенства. Складывая два первых уравнения, получаем $\lambda[f(L) - f(U)] = 0$, что при $\lambda \neq 0$ означает, что решение можно найти, найдя такие (L^*, U^*) , что сумма вероятностей «хвостов» равна α , а плотности равны между собой. Посмотрев на рис. 4.6, можно увидеть, что искомое решение — (L^*, U^*) , где $0 < L^* < L$ и $U^* < U_0$. Это наиболее короткий интервал, и он короче как $[0, U_0]$, так и $[L, U]$.

Это рассуждение применимо к любому распределению, как симметричному, так и нет. Для симметричного распределения мы, очевидно, получим симметричный интервал из (4-51). Верное решение мы получим и в случае асимметричного распределения. В байесовском подходе эквивалентом этому при изучении условного распределения параметра при известных данных является **интервал с наибольшей апостериорной плотностью**. (См. подраздел 16.4.2.) На практике для этого вычисления необходимо знать (или явно предполагать) распределение $y|x^0$, как, например, логнормальное. Обычно размытая оценка используется специально для того, чтобы избежать необходимости делать такое предположение. Также нужно помнить, что простой формулы для построения этого интервала не существует даже в случае логнормального распределения. Наилучшим способом, пожалуй, будет обычный поиск по сетке, поскольку при этом вычисления в каждой точке весьма просты. Главным в приведенном выше рассуждении является то, что наивную интервальную оценку можно значительно улучшить, даже просто используя интервал.

Пример 4.10. Оценка цен на произведения искусства

В примере 4.5 мы рассматривали интересную особенность рынка картин Моне: на аукционах большие картины обычно стоят дороже маленьких. В этом примере мы рассмотрим эту гипотезу с эмпирической точки зрения. Таблица F4.1 содержит данные о 430 результатах аукционных продаж картин Моне, включая данные о размерах картин и некоторых других переменных, которые мы рассмотрим в последующих примерах. Рисунок 4.7 показывает гистограмму цен продаж (в млн долл.). На рис. 4.8 представлена гистограмма логарифмов цен.

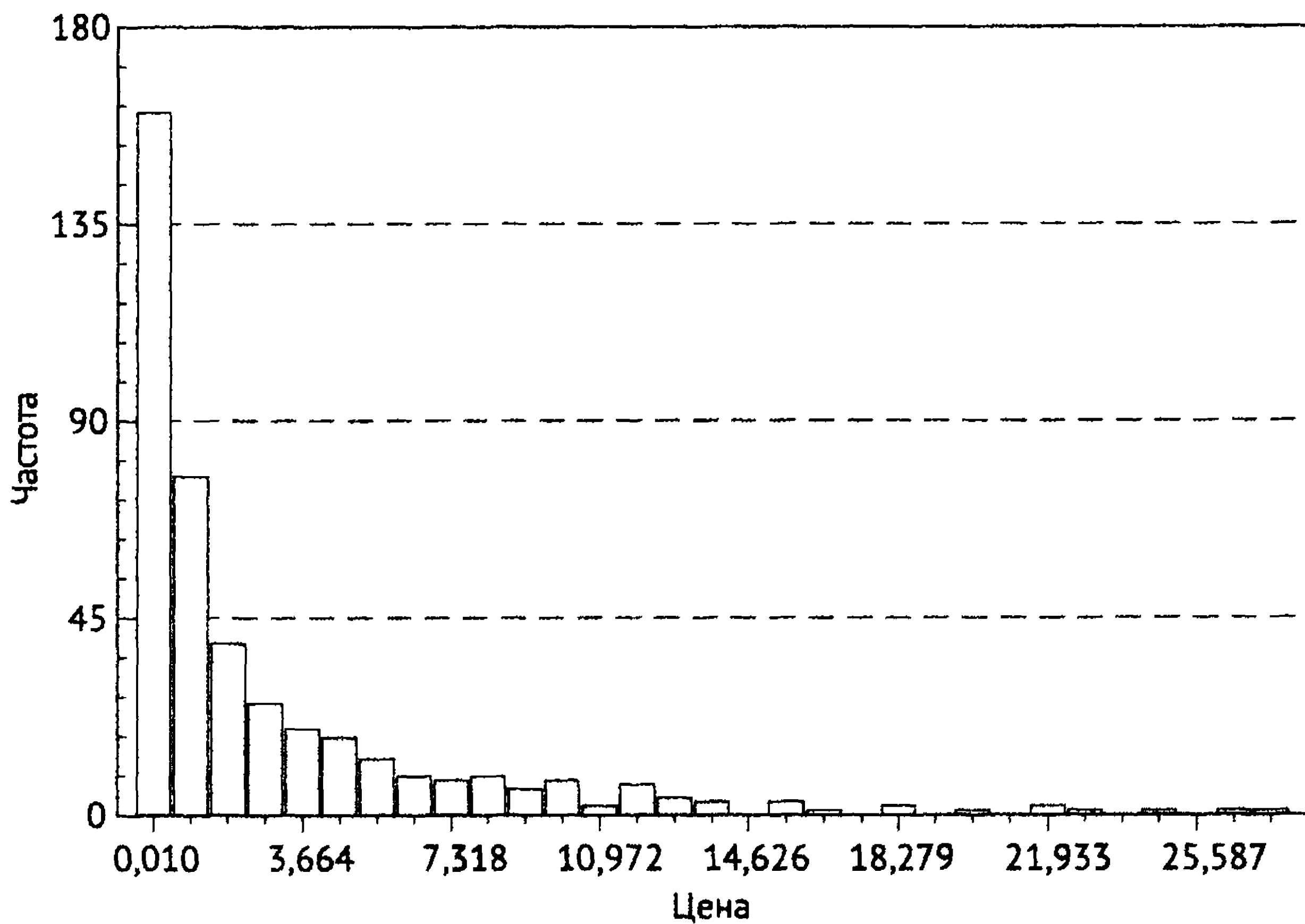


Рис. 4.7. Гистограмма цен продаж 430 картин Моне, млн долл.

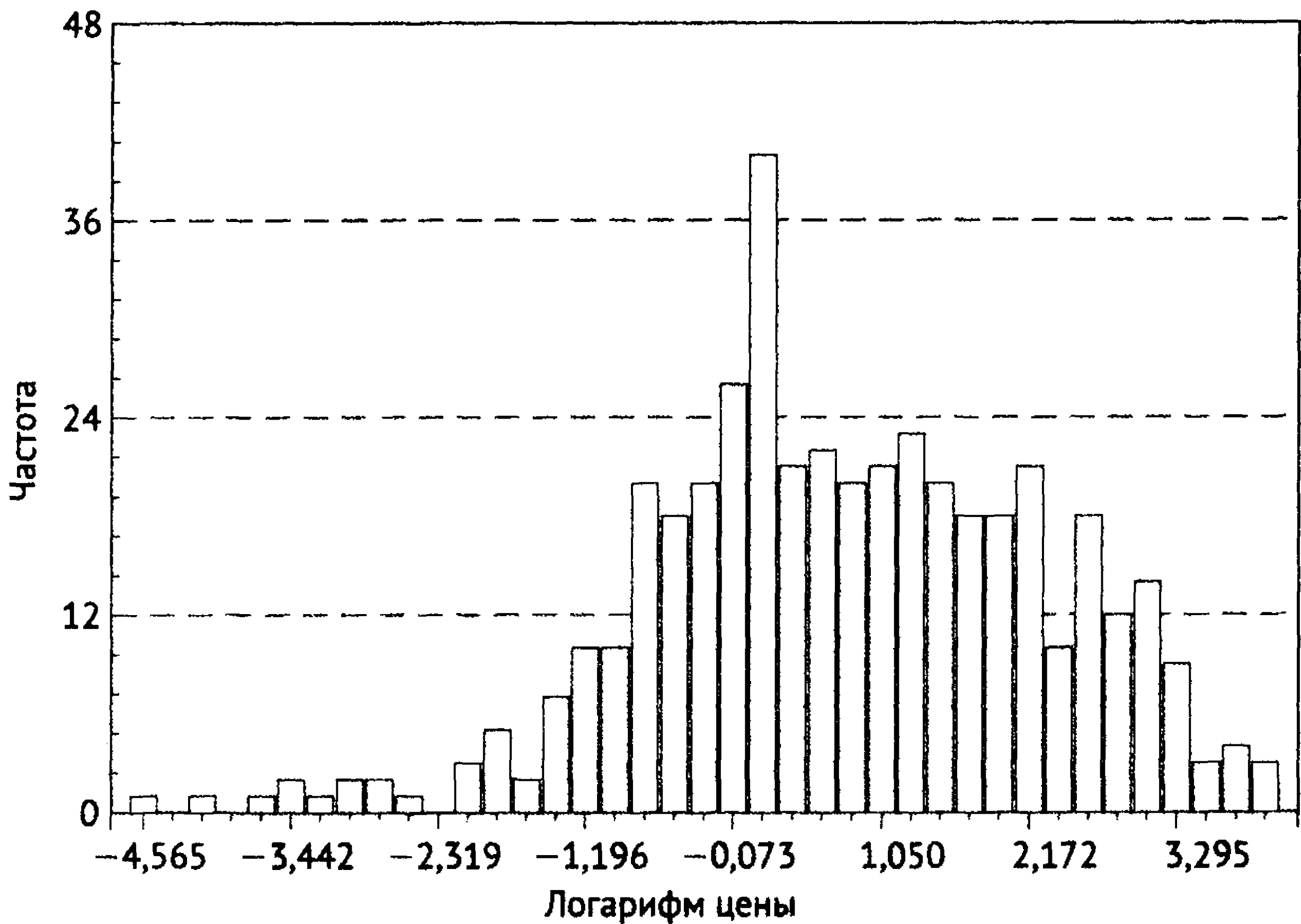


Рис. 4.8. Гистограмма логарифмов цен картин Моне

Результаты линейной регрессии $\ln Price$ на $\ln Area$ (произведение ширины и высоты) и $Aspect Ratio$ (высота, деленная на ширину) приведены в табл. 4.6.

Мы будем применять регрессионную модель для предсказания цены одной из картин — изображения моста Чаринг Кросс 1903 г., которая была продана за 3 522 500 долл. Картина имеет размер 25,6 дюйма в

высоту и 31,9 дюйма в ширину. (Это наблюдение 59 из выборки.) Логарифм площади равен $\ln(25,6 \times 31,9) = 6,705198$, а соотношение сторон — $25,6/31,9 = 0,802508$. Предсказанием логарифма цены будет

$$\ln P|x^0 = -8,42653 + 1,33372(6,705198) - 0,16537(0,802508) = 0,383636.$$

Таблица 4.6. Оцененное уравнение для логарифма цены

| | |
|------------------------------|-----------|
| Среднее логарифма цены | 0,33274 |
| Сумма квадратов остатков | 519,17235 |
| Стандартная ошибка регрессии | 1,10266 |
| R-квадрат | 0,33620 |
| Скорректированный R-квадрат | 0,33309 |
| Количество наблюдений | 430 |

| Переменная | Коэффициент | Стандартная ошибка | t | Среднее X |
|--------------------|-------------|--------------------|--------|-----------|
| Постоянная | -8,42653 | 0,61183 | -13,77 | 1,00000 |
| <i>ln Area</i> | 1,33372 | 0,09072 | 14,70 | 6,68007 |
| <i>AspectRatio</i> | -0,16537 | 0,12753 | -1,30 | 0,90759 |

Оцененная асимптотическая ковариационная матрица

| | Коэффициент | <i>ln Area</i> | <i>AspectRatio</i> |
|--------------------|-------------|----------------|--------------------|
| Постоянная | 0,37434 | -0,05429 | -0,00974 |
| <i>ln Area</i> | -0,05429 | 0,00823 | -0,00075 |
| <i>AspectRatio</i> | -0,00974 | -0,00075 | 0,01626 |

Заметим, что средний логарифм цены равен 0,33274, поэтому на основании данных о ее размерах мы ожидаем, что эта картина будет продана примерно на 5% дороже среднего. Оценка дисперсии предсказания вычисляется с помощью (4-47): $s_p = 1,104027$. Выборка достаточно велика, чтобы для получения 95%-го доверительного интервала можно было использовать критическое значение стандартного нормального распределения 1,96. Доверительный интервал для предсказания логарифма цены, таким образом, равен

$$0,383636 \pm 1,96(1,104027) = [-1,780258; 2,547529].$$

Если мы заинтересованы в предсказании самой цены, то наивный предиктор цены даст значение $\exp(0,383636) = \$1,476411\text{М}$, что намного ниже реальной цены $\$3,5225\text{М}$. Чтобы найти размытую оценку, возьмем среднее экспонент остатков, которое равно 1,813045. Используя этот подход, мы получаем точечную оценку цены, равную $1,813045 \times 1,47641 = \$2,660844\text{М}$, что лучше, но все-таки еще далеко от реального значения. По всей видимости, эта картина была продана по цене, существенно превышающей ожидаемую на основе данных.

Чтобы построить интервальную оценку для цены, сначала воспользуемся наивным подходом: возьмем экспоненты от границ соответствующего интервала для логарифма цены; при этом мы получим 95%-й

доверительный интервал для предсказания [\$0,168595M, \$12,77503M]. С помощью метода из подраздела 4.6.3 можно сузить этот интервал до [0,021261; 9,027543] длиной \$9M, что меньше исходной длины в \$12,5M. Этот интервал дает вероятности «хвостов» 0,00063 слева и 0,04937 справа. Ниже мы описываем алгоритм поиска этого интервала.

Поиск по сетке оптимального доверительного интервала [LO, UO]

$$\mathbf{x}^0 = (1, \log(25,6 \times 31,9), 25,6/31,9)';$$

$$\hat{\mu}^0 = \mathbf{x}^{0'}\mathbf{b}, \hat{\sigma}_p^0 = \sqrt{s^2 + \mathbf{x}^{0'}[s^2(\mathbf{X}'\mathbf{X})^{-1}]\mathbf{x}^0}.$$

Доверительный интервал для $\log P|\mathbf{x}^0$: [Lower, Upper] = $[\hat{\mu}^0 - 1,96\hat{\sigma}_p^0, \hat{\mu}^0 + 1,96\hat{\sigma}_p^0]$.

Наивный доверительный интервал для Price $|\mathbf{x}^0$: L1 = exp(Lower); U1 = exp(Upper).

Первоначальное значение $L = 0,168595$, LO = 0,168595.

Поиск оптимального интервала по сетке с шагом $\Delta = 0,005$ (выбран ad hoc).

Уменьшаем LO и ищем соответствующее значение UO, пока не получим то же значение плотности:

(*) LO = LO - Δ = новое LO;

$$f(\text{LO}) = [LO\hat{\sigma}_p^0\sqrt{2\pi}]^{-1} \exp\left[-\frac{1}{2}\left(\frac{\ln LO - \hat{\mu}^0}{\hat{\sigma}_p^0}\right)^2\right];$$

$F(\text{LO}) = \Phi\left(\frac{\ln(\text{LO}) - \hat{\mu}^0}{\hat{\sigma}_p^0}\right)$ = вероятность левого «хвоста»;

UO = exp($\hat{\sigma}_p^0\Phi^{-1}[F(\text{LO}) + 0,95] + \hat{\mu}^0$) = следующее значение UO;

$$f(\text{UO}) = [UO\hat{\sigma}_p^0\sqrt{2\pi}]^{-1} \exp\left[-\frac{1}{2}\left(\frac{\ln UO - \hat{\mu}^0}{\hat{\sigma}_p^0}\right)^2\right];$$

$1 - F(\text{UO}) = 1 - \Phi\left(\frac{\ln(\text{UO}) - \hat{\mu}^0}{\hat{\sigma}_p^0}\right)$ = вероятность правого «хвоста».

Сравниваем $f(\text{LO})$ и $f(\text{UO})$. Если не равны, возвращаемся к (*). Если равны — алгоритм завершен.

4.6.4. Прогнозирование

В предшествующих разделах мы считали \mathbf{x}^0 известными ex post или абсолютно точно спрогнозированными ex ante. Если \mathbf{x}^0 нужно прогнозировать (прогноз ex ante), то формулу для дисперсии прогноза (4-46) нужно изменить, чтобы учесть возможную ошибку при прогнозировании \mathbf{x}^0 по аналогии с использованием σ^2 при оценке дисперсии предсказания для учета ε^0 . Это существенно осложняет вычисления. Многие авторы считают построение оценки в таких случаях неосуществимым. Начиная с работы Фельдштайна (Feldstein (1971)), вывод аналитических результатов, касающихся построения верной оценки дисперсии прогноза, проводился только в частных случаях. Единственным точно известным качественным результатом остается тот факт, что (4-46) занижает оценку дисперсии. Маккаллоу (McCullough (1996)) описывает альтернативный подход к построению стандартных ошибок прогнозов, основанный на методе бутстрэпа (самонастройки). (См. главу 15.)

Предлагались различные способы оценки точности прогнозов¹². Большинство этих способов рассматривает прогнозы ex post, т. е. случаи, когда

¹²Theil (1961) и Fair (1984).

независимые переменные прогнозировать не нужно. Есть две меры, основанные на остатках прогнозирования: **квадратный корень из среднеквадратичной ошибки** (root mean squared error, RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{n^0} \sum_i (y_i - \hat{y}_i)^2}$$

и **средний модуль ошибки**, или **средняя абсолютная ошибка** (mean absolute error, MAE),

$$\text{MAE} = \frac{1}{n^0} \sum_i |y_i - \hat{y}_i|,$$

где n^0 — число периодов прогнозирования. (Заметим, что обе эти меры, как и другие ниже, опираются на прошлое, т. е. вычисляются с использованием результатов прошлых наблюдений независимых переменных.) Для этих статистик, очевидно, актуальна проблема масштаба: умножение значений зависимой переменной на константу умножает статистику на ту же константу. Несколько мер, являющихся инвариантными к проблеме масштаба, основаны на **U-статистике Тейла** (Theil)¹³:

$$U = \sqrt{\frac{(1/n^0) \sum_i (y_i - \hat{y}_i)^2}{(1/n^0) \sum_i y_i^2}},$$

Эта мера связана с R^2 , но не ограничена отрезком $[0, 1]$. Большие значения этой статистики свидетельствуют о плохом качестве прогноза. Альтернативная мера основана на изменениях y :

$$U_{\Delta} = \sqrt{\frac{(1/n^0) \sum_i (\Delta y_i - \Delta \hat{y}_i)^2}{(1/n^0) \sum_i (\Delta y_i)^2}},$$

где $\Delta y_i = y_i - y_{i-1}$ и $\Delta \hat{y}_i = \hat{y}_i - y_{i-1}$, или в процентном отношении $\Delta y_i = (y_i - y_{i-1})/y_{i-1}$ и $\Delta \hat{y}_i = (\hat{y}_i - y_{i-1})/y_{i-1}$. Эти меры отражают способность оценки обнаруживать изменения в характере поведения данных.

4.7. Проблемы в данных

До сих пор мы предполагали, что имеющиеся данные, X и y , правильно измерены и удовлетворяют предположениям табл. 2.1 и различным теоретическим допущениям о переменных. Теперь мы рассмотрим несколько случаев отклонения жизненных, неэкспериментальных данных от этих предположений. Нарушение предположений обычно имеет последствия для качества оценок параметров модели, причем, к сожалению, плохие. Мы рассмотрим следующие ситуации:

- **Мультиколлинеарность**: предположение о полном ранге, A2, хоть и выполняется, но очень близко к нарушению. («Близко» здесь понятие относительное, а иногда и зависящее от интерпретации.) Мультиколлинеарность ведет к потере точности оценки, но не к систематическому смещению.

¹³Theil (1961).

- Пропущенные значения: пропуски в данных \mathbf{X} и/или y могут быть безвредны. Во многих случаях исследователь может (и должен) их просто игнорировать и пользоваться только полными наблюдениями. В других случаях, если данные пропущены по причинам, имеющим отношение к исследуемым переменным, игнорирование этой проблемы может привести к несостоятельности полученной оценки.
- Ошибки измерения: данные зачастую лишь приблизительно соответствуют теоретическим построениям, отраженным в модели, частым примером являются данные об индивидуальном доходе или образовании. Ошибка измерения всегда вредна. Наиболее благоприятен случай, когда ошибка касается зависимой переменной. В этом случае при некоторых дополнительных разумных предположениях нужно просто считать оценку менее точной по сравнению с оценкой, которую можно было бы получить из (к сожалению, лишь гипотетических) точных данных. Ошибка в регрессорах более опасна — она ведет к систематическому смещению в оценке, которое трудно компенсировать.

4.7.1. Мультиколлинеарность

В теореме Гаусса–Маркова утверждается, что среди всех линейных несмещенных оценок оценка OLS обладает наименьшей дисперсией. Этот результат полезен, но он ничего не говорит о малости значения дисперсии в абсолютном смысле. Возьмем, например, модель с двумя объясняющими переменными и константой. Для обоих коэффициентов

$$\text{Var}[b_k | \mathbf{X}] = \frac{\sigma^2}{(1 - r_{12}^2) \sum_{i=1}^n (x_{ik} - \bar{x}_k)^2} = \frac{\sigma^2}{(1 - r_{12}^2) S_{kk}}, \quad k = 1, 2. \quad (4-52)$$

Если переменные полностью (совершенно) коррелированы, то дисперсия бесконечна. Случай наличия точной линейной зависимости регрессоров является провалом предположений модели, а не данных. Более часто встречается ситуация, когда переменные сильно, но не точно (совершенно) скоррелированы. В этом случае модель удовлетворяет своим предположениям, но могут возникнуть существенные статистические проблемы. Проблема, возникающая перед исследователем в случаях, когда зависимые переменные сильно, но не полностью коррелированы, имеет следующие симптомы:

- малые изменения в данных приводят к большим изменениям в оценках параметров;
- коэффициенты имеют очень большие стандартные ошибки и низкие уровни значимости, несмотря на то что они совместно значимы и R^2 регрессии достаточно высок;
- коэффициенты имеют «неправильный» знак или чрезмерную абсолютную величину.

Пусть для удобства матрица данных \mathbf{X} содержит константу и $K - 1$ других переменных, выраженных в отклонениях от средних. Пусть x_k — k -я переменная, а $\mathbf{X}_{(k)}$ обозначает все остальные переменные (включая константу).

Тогда в обратной матрице $(\mathbf{X}'\mathbf{X})^{-1}$ k -й диагональный элемент равен

$$\begin{aligned} (\mathbf{x}'_k \mathbf{M}_{(k)} \mathbf{x}_k)^{-1} &= [\mathbf{x}'_k \mathbf{x}_k - \mathbf{x}'_k \mathbf{X}_{(k)} (\mathbf{X}'_{(k)} \mathbf{X}_{(k)})^{-1} \mathbf{X}'_{(k)} \mathbf{x}_k]^{-1} \\ &= \left[\mathbf{x}'_k \mathbf{x}_k \left(1 - \frac{\mathbf{x}'_k \mathbf{X}_{(k)} (\mathbf{X}'_{(k)} \mathbf{X}_{(k)})^{-1} \mathbf{X}'_{(k)} \mathbf{x}_k}{\mathbf{x}'_k \mathbf{x}_k} \right) \right]^{-1} \\ &= \frac{1}{(1 - R_{k.}^2) S_{kk}}, \end{aligned} \quad (4-53)$$

где $R_{k.}^2$ — это R^2 регрессии x_k на все остальные переменные. В модели множественной регрессии дисперсия k -го коэффициента оценки метода наименьших квадратов равна σ^2 , умноженному на это отношение. Отсюда следует, что чем сильнее переменная скоррелирована с другими переменными модели (в совокупности), тем больше ее дисперсия. В экстремальном случае, когда \mathbf{x}_k можно представить как линейную комбинацию остальных переменных, т. е. $R_{k.}^2 = 1$, дисперсия становится бесконечной. Результат

$$\text{Var}[b_k | \mathbf{X}] = \frac{\sigma^2}{(1 - R_{k.}^2) \sum_{i=1}^n (x_{ik} - \bar{x}_k)^2} \quad (4-54)$$

выявляет три составляющие, влияющие на точность оценки коэффициента методом наименьших квадратов:

- При прочих равных чем выше корреляция x_k с другими переменными, тем больше дисперсия оценки (из-за мультиколлинеарности).
- При прочих равных чем выше выборочная дисперсия x_k , тем ниже дисперсия оценки. Этот результат продемонстрирован на рис. 4.3.
- При прочих равных чем лучше приближение регрессии в целом, тем меньше дисперсия. Это происходит из-за уменьшения σ^2 . Этот результат еще необходимо вывести, но его можно предположить, посмотрев на рис. 4.3 и представив картину, где все точки сдвинуты к линии регрессии.

Поскольку экспериментальные данные никогда не окажутся ортогональными ($R_{k.}^2 = 0$), мультиколлинеарность всегда будет присутствовать. Когда же она становится проблемой? Иными словами, когда дисперсии наших оценок становятся столь велики, что нам нужно об этом беспокоиться? Некоторые компьютерные пакеты при оценке регрессии возвращают для каждой переменной диагностическую величину **множитель наполнения дисперсии** (variance inflation factor, VIF), $1/(1 - R_{k.}^2)$. Можно видеть, что VIF переменной отражает увеличение $\text{Var}[b_k]$ из-за того, что эта переменная не ортогональна остальным переменным модели. Другая мера, специально направленная на матрицу \mathbf{X} , — это **число обусловленности** (condition number) $\mathbf{X}'\mathbf{X}$, равное квадратному корню отношения наибольшего собственного значения матрицы $\mathbf{X}'\mathbf{X}$ к наименьшему (после такой нормировки столбцов, чтобы у каждого из них была единичная длина). Значения больше 20 считаются признаком наличия проблемы мультиколлинеарности [Belsley, Kuh, Welsch (1980)] (Число обусловленности матрицы данных в примере 4.11 превышает 15 000!)

Пример 4.11. Мультиколлинеарность в данных Лонгли

Данные табл. F4.2 были собраны в работе Лонгли (Longley (1967)) для оценки точности вычисления OLS-оценок компьютерными программами. (Эта база данных и сейчас для этого часто применяется.) Данные Лонгли известны исключительно высокой степенью мультиколлинеарности. Рассмотрим, например, данные за последний год. Последнее наблюдение не кажется сколько-нибудь выделяющимся. Однако результаты, представленные в табл. 4.7, демонстрируют сильнейший эффект, который исключение одного этого наблюдения оказывает на результат регрессии на константу и остальные переменные. Последний коэффициент увеличивается на 600%, а третий — на 800%.

Таблица 4.7. Результаты Лонгли: зависимая переменная — занятость

| | 1947–1961 | Variance Inflation | 1947–1962 |
|--------------|------------|--------------------|------------|
| Constant | 1 459 415 | | 1 169 087 |
| Year | –721,756 | 143,4638 | –576,464 |
| GNP deflator | –181,123 | 75,6716 | –19,7681 |
| GNP | 0,0910678 | 132,467 | 0,0643940 |
| Armed Forces | –0,0749370 | 1,55319 | –0,0101453 |

Предлагалось несколько различных стратегий для обнаружения мультиколлинеарности и борьбы с нею¹⁴. Считая, что мультиколлинеарность является следствием недостатка информации, можно было бы рекомендовать найти дополнительные данные. Однако здесь можно возразить, что если у исследователя был бы доступ к дополнительным данным, то их нужно было использовать до появления проблемы мультиколлинеарности. Тем не менее увеличение объема информации необязательно означает увеличение числа наблюдений. Очевидный практический способ (и наиболее часто используемый) — это отбросить переменные, которые, по всей видимости, вызывают эффект мультиколлинеарности, т. е. наложить на регрессию дополнительное (и, возможно, неверное) условие, что «проблемные» переменные отсутствуют в модели.

Поступая таким образом, мы можем столкнуться с проблемами спецификации модели, которые рассмотрим в подразделе 4.7.2. Если отброшенная переменная на самом деле должна входить в модель (т. е. соответствующий коэффициент β_k не равен 0), то оценки остальных коэффициентов окажутся смещенными, причем, возможно, достаточно сильно. С другой стороны, включение в модель чрезмерного числа переменных (overfitting) — т. е. попытка оценить слишком большую модель — является распространенной ошибкой, так что отбрасывание переменных из перегруженной модели может иметь некоторые преимущества.

¹⁴Описание обычных способов обнаружения мультиколлинеарности можно найти в работе (Hill, Adkins (2001)).

Методы обнаружения мультиколлинеарности можно рассматривать как способ различить «плохие данные» и «плохую модель». Однако на самом деле проблема возникает из-за несоответствия данных предварительно сформулированным ожиданиям. Если оказывается, что мультиколлинеарность отрицательно влияет на качество оценки, то предполагается, что в отсутствие этой проблемы все переменные будут значимы и будут иметь нужный знак. Конечно, так бывает не всегда. Если данные говорят, что переменная не важна в контексте данной модели, то при наличии сколь угодно убедительных теоретических построений рано или поздно нужно задуматься, насколько нам на самом деле важна эта теория. Предлагаемые методы «борьбы» с мультиколлинеарностью зачастую представляют собой способы притягивания теории к данным.

4.7.2. Предварительное оценивание

Часто при столкновении с проблемой мультиколлинеарности трудно противостоять соблазну просто исключить неудобные переменные из регрессии. Эта «стратегия» ставит исследователя перед дилеммой. Рассмотрим модель регрессии

$$\mathbf{y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \varepsilon.$$

Если взять регрессию y только на \mathbf{X}_1 , то получится смещенная оценка:

$$E[\mathbf{b}_1|\mathbf{X}] = \beta_1 + \mathbf{P}_{1.2}\beta_2.$$

Ковариационная матрица этой оценки равна

$$\text{Var}[\mathbf{b}_1|\mathbf{X}] = \sigma^2(\mathbf{X}'_1\mathbf{X}_1)^{-1}.$$

(Обратим внимание, что разброс здесь идет около среднего, равного $E[\mathbf{b}_1|\mathbf{X}]$, а не β_1 .) Если β_2 не равно 0, то дисперсия оценки $\mathbf{b}_{1.2}$ около его среднего β_1 в регрессии y на $(\mathbf{X}_1, \mathbf{X}_2)$ будет равна

$$\text{Var}[\mathbf{b}_{1.2}|\mathbf{X}] = \sigma^2(\mathbf{X}'_1\mathbf{M}_2\mathbf{X}_1)^{-1},$$

где

$$\mathbf{M}_2 = \mathbf{I} - \mathbf{X}_2(\mathbf{X}'_2\mathbf{X}_2)^{-1}\mathbf{X}'_2,$$

или

$$\text{Var}[\mathbf{b}_{1.2}|\mathbf{X}] = \sigma^2[\mathbf{X}'_1\mathbf{X}_1 - \mathbf{X}'_1\mathbf{X}_2(\mathbf{X}'_2\mathbf{X}_2)^{-1}\mathbf{X}'_2\mathbf{X}_1]^{-1}.$$

Сравним эти две ковариационные матрицы. Проще сравнивать обратные матрицы. [См. результат (A-120).] Таким образом,

$$\{\text{Var}[\mathbf{b}_1|\mathbf{X}]\}^{-1} - \{\text{Var}[\mathbf{b}_{1.2}|\mathbf{X}]\}^{-1} = (1/\sigma^2)\mathbf{X}'_1\mathbf{X}_2(\mathbf{X}'_2\mathbf{X}_2)^{-1}\mathbf{X}'_2\mathbf{X}_1,$$

что дает неотрицательно определенную матрицу. Следовательно, дисперсия \mathbf{b}_1 не выше дисперсии $\mathbf{b}_{1.2}$ (так как обратная величина не меньше соответствующей обратной). Несмотря на то что оценка \mathbf{b}_1 смещена, ее дисперсия не превышает дисперсию несмещенной оценки. В реалистичной ситуации (когда $\mathbf{X}'_1\mathbf{X}_2$ не равно 0) ее дисперсия на самом деле будет строго меньше. В случае простой регрессии от двух переменных (представленных в отклонениях от своих средних) можно получить полезное соотношение. Имеем $\text{Var}[\mathbf{b}_1|\mathbf{X}] = \sigma^2/S_{11}$, где $S_{11} = \sum_{i=1}^n (x_{i1} - \bar{x}_1)^2$ и $\text{Var}[b_{1.2}|\mathbf{X}] = \sigma^2/[S_{11}(1 - r_{12}^2)]$, где r_{12}^2 — квадрат корреляции между x_1 и x_2 .

Результат предыдущего параграфа ставит исследователя перед проблемой, часто возникающей при поиске подходящей модели. Если исследователь сталкивается с переменной, которая, по всей видимости, должна участвовать в модели, но ее включение приводит к мультиколлинеарности, то можно либо отбросить ее, либо смириться с тем, что коэффициенты при этой и других переменных будут неточны. Возникает проблема выбора между оценками b_1 и $b_{1.2}$. На практике исследователи обычно строят третью оценку. Проблемная переменная включается условно: если при включении ее t -статистика оказывается достаточно велика, то она сохраняется; если нет — отбрасывается. Эта третья оценка обычно называется **предварительной** (pretest estimator). То, что известно об этих оценках, не внушает энтузиазма. Разумеется, они являются смещенными. Степень смещения зависит от неизвестных параметров. Аналитические результаты свидетельствуют о том, что такая оценка более смещена, чем обе исходные, как раз в тех случаях, когда наиболее велик соблазн ее использовать. [Judge et al. (1985).] Из всего этого можно сделать вывод, что корректная методология оценивания не должна включать стратегий, опирающихся на ad hoc-методы борьбы с мультиколлинеарностью.

4.7.3. Метод главных компонент

Для борьбы с мультиколлинеарностью предлагалась [Gurmu, Rilstone, Stern (1999)] следующая конструкция: нужно оценивать регрессию на несколько (скажем L) **главных компонент**, представляющих собой линейные комбинации K исходных переменных [Johnson, Wichern (2005, Chapter 8).] (Этот метод иллюстрируется в примере 4.12.) Основное возражение против этого метода состоит в том, что если вместо верной регрессии $y = \mathbf{X}\beta + \varepsilon$ мы оцениваем регрессию y на небольшой набор линейных комбинаций столбцов \mathbf{X} , то неясно, что именно мы оцениваем. Для набора $L < K$ главных компонент, если \mathbf{d} — оценка коэффициентов регрессии y на $\mathbf{Z} = \mathbf{X}\mathbf{C}_L$, то $E[\mathbf{d}] = \boldsymbol{\delta} = \mathbf{C}'_L\boldsymbol{\beta}$. (Доказательство предлагается читателю в качестве упражнения.) Если в экономической задаче параметр $\boldsymbol{\beta}$ имеет понятную интерпретацию, то маловероятно, что $\boldsymbol{\delta}$ имеет ясный экономический смысл. (Например, как интерпретировать эластичность по цене минус удвоенная эластичность по доходу?)

Этот традиционный подход предостерегает исследователя от использования алгебраических методов, результаты которых приводят к неинтерпретируемым смесям коэффициентов. Однако иногда модель строится так, что смеси наблюдаемых переменных могут иметь ясный смысл. Например, исследователя может интересовать модель, в которой участвуют неясно определенные «способности». Вместо них у исследователя может иметься набор результатов, полученных при прохождении ряда тестов, ни один из которых не имеет собственной интерпретации в контексте рассматриваемой модели. В этом случае отсутствующую переменную «способности» можно заменить некоторой комбинацией имеющихся результатов тестов. Другой естественный случай применения этого подхода рассматривается в примере 4.12.

¹⁵Тестирование мнения «а кто его знает» сейчас может быть проверено на открытой

Пример 4.12. Предсказание успеха фильма

Предсказание размера сборов кинофильмов — любимое упражнение эконометристов. [Litman (1983), Ravid (1999), De Vany (2003), De Vany, Walls (1999, 2002, 2003) и Simonoff, Sparrow (2000).] Обычно для предсказания используется уравнение вида

$$\text{Box Office Receipts} = f(\text{Budget, Genre, MPAА Rating, Star Power, Sequel, etc.}) + \varepsilon.$$

Коэффициент детерминации таких уравнений часто достигает 0,4 и более. Однако, несмотря на такую объяснительную силу, обычное мнение в Голливуде — «а кто его знает». Успех фильма часто случаен, и немногие берутся предсказывать его с какой-либо уверенностью¹⁵. В работе Версачи (Versaci (2009)) в традиционную модель была добавлена новая переменная «шум в интернете». Шум в Интернете здесь определяется как имеющий отношение к фильму трафик и общая заинтересованность на популярных сайтах, таких как RottenTomatoes.com, IMDb.com, Fandango.com и traileraddict.com. Каждый из них, измеренный по отдельности, не определяет меру шума в интернете. Но в совокупности данные об активности на этих веб-сайтах (скажем, за три недели до выхода фильма в прокат) могут быть важным признаком будущего успеха. Данные работы Версачи (Versaci (2009)) (табл. F4.3) содержат информацию о 62 фильмах, вышедших на экраны в 2009 г., включая переменные, описывающие 4 меры шума в интернете, измеренные за три недели до выхода фильма в прокат:

$buzz_1$ = число просмотров трейлера фильма на traileraddict.com;

$buzz_2$ = число комментариев о фильме на ComingSoon.net;

$buzz_3$ = сумма числа голосов «can't wait» и «don't care» на Fandango.com;

$buzz_4$ = процент голосов «can't wait» на Fandango.

Мы собрали из этих данных одну главную компоненту следующим образом: сначала вычисляются логарифмы $buzz_1 - buzz_3$, чтобы убрать эффекты масштаба. Затем переменные нормируются так, чтобы z_k равнялась первоначальной переменной минус ее среднее \bar{z}_k , деленной на стандартное отклонение s_k . Пусть Z — полученная матрица размером 62×4 из (z_1, z_2, z_3, z_4) . Тогда $V = (1/61)Z'Z$ — выборочная матрица корреляций. Пусть c_1 — собственный вектор V , соответствующий наибольшему собственному значению. Первая главная компонента (описывающая наибольшую долю вариации четырех переменных) равна Zc_1 . (Корни равны 2,4142, 0,7742, 0,4522, 0,3585, так что первая главная компонента описывает 2,4142/4, или 60,3% разброса. В табл. 4.8 представлены результаты оценивания регрессии на выборке из 62 фильмов 2009 г. Оказывается, что переменная «шум в интернете» существенно увеличивает предсказательную силу модели. При добавлении этой переменной R^2

недавно (в апреле 2010) интернет-площадке биржевого типа, где торгуются фьючерсы на будущий успех фильмов (а продюсеры могут захеджировать свои ставки). См. обсуждение на <http://www.cantorexchange.com/>. Эта торговая площадка была создана компанией Cantor Fitzgerald, Inc. после покупки популярного развивающего интернет-сайта *Hollywood Stock Exchange*.

регрессии почти удваивается, с 0,34 до 0,58. Как мы увидим в главе 5, «шум» является еще и очень значимым предиктором успеха.

Таблица 4.8. Результаты регрессии для успеха фильма

| | Модель с шумом в Интернете | | | Традиционная модель | | |
|------------|----------------------------|---------------|-------|---------------------|---------------|-------|
| | | | | | | |
| $e'e$ | | 22,30215 | | | 35,66514 | |
| R^2 | | 0,58883 | | | 0,34247 | |
| Переменная | Коэффициент | Станд. ошибка | t | Коэффициент | Станд. ошибка | t |
| Постоянная | 15,4002 | 0,64273 | 23,96 | 13,5768 | 0,68825 | 19,73 |
| ACTION | -0,86932 | 0,29333 | -2,96 | -0,30682 | 0,34401 | -0,89 |
| COMEDY | -0,01622 | 0,25608 | -0,06 | -0,03845 | 0,32061 | -0,12 |
| HORROR | 0,37460 | 0,37109 | 1,01 | 1,02644 | 0,44008 | 2,33 |
| G | 0,38440 | 0,55315 | 0,69 | 0,25242 | 0,69196 | 0,36 |
| PG | 0,53359 | 0,29976 | 1,78 | 0,32970 | 0,37243 | 0,89 |
| PG13 | 0,21505 | 0,21885 | 0,98 | 0,07176 | 0,27206 | 0,26 |
| LOGBUDGT | 0,26088 | 0,18529 | 1,41 | 0,70914 | 0,20812 | 3,41 |
| SEQUEL | 0,27505 | 0,27313 | 1,01 | 0,64368 | 0,33143 | 1,94 |
| STARPOWR | 0,00433 | 0,01285 | 0,34 | 0,00648 | 0,01608 | 0,40 |
| BUZZ | 0,42906 | 0,07839 | 5,47 | | | |

4.7.4. Пропущенные значения и пополнение данных

Вследствие различных причин в реальных наборах данных зачастую встречаются пропуски. Чаще всего такое случается с данными опросов населения, где респонденты могут отвечать не на все вопросы. Во временных рядах данные могут быть пропущены из-за того, что их нельзя получить с нужной частотой: например, модель может описывать месячные события, а некоторые переменные могут быть измерены только ежеквартально. В панельных данных пропуски могут возникать из-за **истощения выборки**, т.е. отсева участников исследования. Особенно часто это наблюдается в медицинских исследованиях, где индивиды могут прекращать свое участие в исследовании, возможно, вследствие успеха или неуспеха изучаемого метода лечения.

В зависимости от того, по каким причинам данные оказались пропущены, подходы к проблеме могут различаться. Данные могут быть недоступны по неизвестным исследователю причинам, не связанным с полнотой других наблюдений или значениями переменных в выборке. Это наименее неблагоприятная ситуация. Если дело обстоит так, то полные наблюдения дают выборку, которой можно пользоваться, а единственный вопрос состоит в том, нельзя ли сохранить какую-то информацию из неполных наблюдений. Согласно работе Грилихеса (Griliches (1986)) это называют **пренебрежимым случаем** (ignorable case), поскольку, если нас мало заботит эффективность получаемой оценки, мы можем просто пренебречь неполными наблюдениями и закрыть глаза на имеющуюся проблему. В работах Рубина (Rubin (1976, 1987)) и Литтла и Рубина (Little, Rubin (1987, 2002)) этот случай назван **пол-**

ностью случайными пропусками (missing completely at random, MCAR). Второй случай, широко изучаемый в эконометрической литературе, заключается в том, что пропуски в данных не случайны, а систематически связаны с изучаемым явлением. Это чаще всего происходит в данных опросов, особенно когда данные, входящие в выборку, выбирали или определяли респонденты¹⁶. Например, если опрос нужен для исследования характера трат, а люди с высоким доходом пытаются скрыть свой уровень достатка, то пропущенные данные будут означать нечто большее, чем просто недостаток информации. Другим примером могут быть данные клинических испытаний. В этих (наихудших) случаях подвыборка, состоящая только из полных наблюдений, будет качественно отличаться от случайной выборки, взятой из генеральной совокупности. В этом случае пропущенные данные называют **пропущенными не случайно** (not missing at random, NMAR). Мы будем рассматривать этот случай в главе 19 в контексте **отбора наблюдений** (sample selection), а пока отложим эту тему.

Промежуточный случай заключается в том, что полные наблюдения содержат информацию, касающуюся пропущенных данных, с помощью которых инференция может быть улучшена. В этой ситуации **случайных пропусков** (missing at random, MAR) неполные наблюдения снова можно проигнорировать, поскольку, в отличие от случая NMAR, использование подвыборки, состоящей исключительно из полных наблюдений, не приведет к смещению, если процесс, порождающий пропуски, не связан с оцениваемой моделью [Allison (2002).] Это, конечно, маловероятно, если пропуски наблюдаются в значениях зависимой переменной в регрессии. Игнорируя неполные наблюдения в ситуации MAR, в отличие от MCAR мы теряем информацию, а значит, жертвуем эффективностью оценки. Исследователи используют для заполнения пропусков различные методы **пополнения данных** (data imputation). Самый простой вариант в данной ситуации — это когда пропуски встречаются в регрессорах. В этом случае полезно рассмотреть отдельно простую и множественную регрессии. В первом случае X содержит два столбца: столбец единиц (для свободного члена) и столбец, содержащий пропуски на месте отсутствующих данных. **Метод нулевого порядка** (zero-order method) заменяет пропущенные x на \bar{x} , вычисленные по имеющимся данным, не меняет оценки и эквивалентен оценке, построенной только по полным наблюдениям. (См. упражнение 7 к главе 3.) При этом, однако, R^2 несколько уменьшается. Альтернативным подходом является **модифицированный метод нулевого порядка** (modified zero-ordered method), который состоит в заполнении пропусков нулями и добавлении переменной, принимающей значение 1 для пропущенных наблюдений и 0 в остальных случаях¹⁷. Мы оставляем в качестве упражнения доказательство того факта, что этот метод алгебраически эквивалентен замене пропущенных наблюдений на \bar{x} . Также можно попробовать подбирать x вместо пропущенных

¹⁶Широкие опросы мнений американцев по вопросам пола, собранные в работах Ann Landers (1984, passim) и Shere Hite (1987), — широко известный пример исследований, безусловно испорченных вследствие смещений, вызванных самоотбором. Последнее исследование подверглось резкой критике за то, что в нем пытались выдать за мнение населения в целом мнение группы людей, готовых отвечать на задававшиеся вопросы. Выводы первого исследования были более аккуратными.

¹⁷Maddala (1977a, p. 202).

данных, построив регрессию x на y по полным данным.

Свойства такой оценки изучены мало, но имеющиеся данные говорят не в ее пользу¹⁸.

Те же методы можно использовать и в случае множественной регрессии. Снова возникает соблазн заменить пропущенные значения \mathbf{x}_k простыми средними, построенными по полным наблюдениям, или предсказаниями, построенными с помощью регрессий \mathbf{x}_k на другие переменные модели. В большинстве случаев эти методы можно описывать в следующем образом: в местах пропусков «истинное» значение x_{ik} заменяется неточным приближением, которое можно записать в виде $\hat{x}_{ik} = x_{ik} + u_{ik}$, т. е. в терминах **ошибок измерения** (measurement error). Вообще говоря, оценки метода наименьших квадратов при наличии ошибок измерения смещены и несостоятельны (этот вопрос будет исследоваться в главе 8.) Остается вопрос: будет ли это смещение мало? Интуиция подсказывает, что размер смещения будет определяться двумя свойствами имеющихся данных: (а) насколько хороши наши приближения к x_{ik} , т. е. насколько велика дисперсия ошибки измерения u_{ik} по сравнению с дисперсией истинных данных x_{ik} , и (б) насколько большая часть данных нуждается в пополнении.

Метод регрессий заменяет пропущенные значения \mathbf{x}_k предсказанием, полученным из регрессии \mathbf{x}_k на другие экзогенные переменные, т. е. заменяет пропущенные x_{ik} приближениями, основанными на регрессионной модели. В байесовской постановке в приложениях, требующих применения ненаблюдаемых переменных (как, например, в нашем примере для модели бинарного выбора в главе 17), применяется техника **расширения данных** (data augmentation), когда ненаблюдаемые данные считают «параметрами» модели, оцениваемыми вместе со структурными параметрами, такими как β в регрессионной модели. Основываясь на этом подходе, исследователи (Rubin (1987) и Allison (2002)) предлагают использовать нечто похожее в классической постановке оценивания. В этой технике используется пополнение данных, похожее на предлагавшееся выше, но расширенное так, чтобы учесть возможную изменчивость при оценивании регрессии, использованной для подстановки пропущенных значений. Для иллюстрации этого рассмотрим случай, когда независимая переменная \mathbf{x}_k является нормальной, т. е. выбирается из непрерывного распределения, имеющего среднее, дисперсию и определенные совместные распределения с другими переменными модели. Пополнение будет включать следующие формальные шаги:

1. Используя максимально возможный объем содержащейся в выборке информации (полные данные), оцениваем линейную регрессию \mathbf{x}_k на другие переменные модели (а также, возможно, вне модели, если такая информация доступна), \mathbf{Z}_k , и получаем вектор коэффициентов \mathbf{d}_k , асимптотическую ковариационную матрицу \mathbf{A}_k и оценку дисперсии шоков s_k^2 .
2. Для пополнения возьмем значение из оцененного выше асимптотического нормального распределения \mathbf{d}_k , т. е. $\mathbf{d}_{k,m} = \mathbf{d}_k + \mathbf{v}_k$, где \mathbf{v}_k — вектор, выбранный из многомерного нормального распределения со средним 0 и ковариационной матрицей \mathbf{A}_k .

¹⁸Affi, Elashoff (1966, 1967) и Haitovsky (1968). В работе Griliches (1986) рассматриваются некоторые другие возможности.

3. Для каждого пропущенного значения x_k , которое мы хотим заполнить, вычисляем $x_{i,k,m} = \mathbf{d}'_{k,m} \mathbf{z}_{i,k} + s_{k,m} u_{i,k}$, где $s_{k,m} = s_k$, деленное на случайное значение, полученное из выборки распределения χ^2 с числом степеней свободы, равным числу степеней свободы использованной для пополнения регрессии.

После этих шагов мы имеем результат, похожий на описанный выше, когда пропущенные значения оценивались с помощью регрессии. Приведенная здесь процедура, однако, несколько сложнее. После этого нужно построить оценку коэффициентов регрессии, пользуясь как полными данными, так и пополненными наблюдениями. Отсюда получается вектор коэффициентов \mathbf{b}_m и оценка ковариационной матрицы \mathbf{V}_m . Это одна итерация процедуры пополнения. Техника **множественного пополнения** (multiple imputation) состоит в повторении этой последовательности шагов M раз. Оценки параметров и соответствующей асимптотической ковариационной матрицы равны

$$\hat{\beta} = \bar{\mathbf{b}} = \frac{1}{M} \sum_{m=1}^M \mathbf{b}_m,$$

$$\hat{\mathbf{V}} = \bar{\mathbf{V}} + \mathbf{B} = \frac{1}{M} \sum_{m=1}^M \mathbf{V}_m + \left(1 + \frac{1}{M}\right) \left(\frac{1}{M-1}\right) \sum_{m=1}^M (\mathbf{b}_m - \bar{\mathbf{b}}) (\mathbf{b}_m - \bar{\mathbf{b}})'$$

Мнения исследователей относительно эффективности и применимости множественного пополнения расходятся. Эта процедура не устраняет ошибки измерения в пополненных значениях. Для того чтобы вывести, что множественное повторение итераций метода приведет к усреднению и исчезновению этой ошибки, нужно принять очень сильные допущения. Были изучены различные техники для работы с данными в случае, когда межобъектные данные являются совместно нормально распределенными переменными, как в рассмотренном нами выше случае. Однако обычно нужно работать с данными опросов населения, где пропуски встречаются в ответах на качественные вопросы типа «да/нет». Эффективность применения теории к бинарным, порядковым и прочим переменным, имеющим качественный смысл, изучена мало.

Несколько проще работать в случае наличия пропусков значений зависимой переменной y_i . Снова нужно, чтобы y_i была по крайней мере MAR, а механизм, порождающий пропуски, не был связан с параметрами модели. Если данные о x_i полны для всех наблюдений, то можно было бы применить следующую двухшаговую процедуру: (1) построить оценку \mathbf{b}_c для параметра β , используя только полные наблюдения, \mathbf{X}_c и y_c , а затем (2) заполнить пропуски в y_m предсказаниями $\hat{y}_m = \mathbf{X}_m \mathbf{b}_c$ и заново оценить регрессию. Мы оставляем в качестве упражнения для читателя (упражнение 17 к этой главе) доказательство того, что полученная оценка будет совпадать с \mathbf{b}_c . Однако получаемая на втором шаге оценка дисперсии s^2 будет занижать истинное значение σ^2 (интуитивно это ясно: мы добавляем наблюдения, идеально соответствующие оценке) [Plamondon, Trivedi (2005, Chapter 27).] Таким образом, этот метод не является особенно полезным. Недостаток этого метода заключается в использованном механизме пополнения пропущенных значений y_i . В последнее время предлагаются различные методы, улучшающие эту процедуру, как правило, путем использования рандомизированной версии

$\hat{y}_m = \mathbf{X}_m \mathbf{b}_c + \hat{\varepsilon}_m$, где $\hat{\varepsilon}_m$ случайно выбираются из нормального распределения с нулевым средним и оцененной дисперсией $s^2[\mathbf{I} + \mathbf{X}_m(\mathbf{X}'_c \mathbf{X}_c)^{-1} \mathbf{X}'_m]$. (Оценка ковариационной матрицы соответствует $\mathbf{P}_m \mathbf{b}_c + \varepsilon_m$.) Это определяет одну итерацию. Снова оценив β с помощью пополненных данных, можно повторно пополнить данные, уже пользуясь новым $\hat{\beta}$, потом опять оценить \mathbf{b} и т. д. Процесс можно повторять до тех пор, пока оценка коэффициентов не прекратит изменяться. (Здесь нужно сделать важное замечание. В каждой итерации следует использовать одни и те же значения случайной переменной. В противном случае нет гарантии, что итерационный процесс сойдется.)

Вообще говоря, о свойствах оценок, основанных на использовании предсказаний для заполнения пропусков в значениях y , известно немного. Имеющиеся результаты получены в основном из моделирования с использованием конкретного набора данных или механизма, порождающего пропуски. Результаты этих симуляций метода Монте-Карло обычно сложно обобщить. В целом, по всей видимости, можно заключить, что при исследовании регрессий с единственным уравнением пополнение y ведет к смещению в оценках, которое сложно определить. Единственный ясный результат состоит в том, что пополнения чаще оказываются полезны, если доля пропусков в наборе данных мала, причем чем меньше, тем лучше.

4.7.5. Ошибки измерения

В целом ряде случаев наблюдаемые данные представляют собой не точные значения, используемые в теоретической модели, а результаты их измерения с некоторой погрешностью. Примерами могут служить доход, уровень образования, навыки, здоровье, процентные ставки, производство, капитал и т. п. Ошибки при измерении переменных модели обычно приводят к негативным последствиям для оценок метода наименьших квадратов. Методы решения этой проблемы сложны и часто требуют чрезмерно сильных допущений. В этом разделе мы кратко опишем суть проблемы. Более полное обсуждение проблемы ошибки измерения и обсуждение наиболее часто используемого решения (инструментальных переменных), а также некоторых приложений мы отложим до раздела 8.5.

Удобно будет разделить обсуждение ошибок измерения в зависимой переменной и в регрессорах. В последнем случае также удобно сначала рассмотреть случай модели с одним регрессором, а затем расширить его до случая множественной регрессии. Рассмотрим модель, описывающую ожидаемый уровень дохода населения:

$$I^* = \mathbf{x}'\beta + \varepsilon, \quad (4-55)$$

где I^* — полный ожидаемый доход. Предположим, что вместо этого измерен уровень заработанного дохода I . Неясно, как I связана с I^* . Обычно предполагают, что ошибка измерения аддитивна, т. е. $I = I^* + w$. Подставляя это равенство вместо I в (4-55), получаем

$$\begin{aligned} I &= \mathbf{x}'\beta + \varepsilon + w \\ &= \mathbf{x}'\beta + v, \end{aligned} \quad (4-56)$$

что, кажется, несколько сложнее исходной регрессии, хотя и очень на нее похоже. Если w и x некоррелированы, то это действительно так. Если w — гомоскедастичная ошибка с нулевым средним, не коррелированная с x , то вся разница между моделями (4-55) и (4-56) состоит в том, что дисперсия шоков в (4-56) равна $\sigma_w^2 + \sigma_\varepsilon^2 > \sigma_\varepsilon^2$. В остальном это две регрессии, и в обоих случаях β может быть состоятельно оценена с помощью метода наименьших квадратов. Ошибка измерения проявляет себя через уменьшение точности оценки, поскольку асимптотическая дисперсия оценки из (4-56) равна $(\sigma_v^2/n)[\text{plim}(\mathbf{X}'\mathbf{X}/n)]^{-1}$ в отличие от $(\sigma_\varepsilon^2/n)[\text{plim}(\mathbf{X}'\mathbf{X}/n)]^{-1}$ в случае, если β оценивается с помощью (4-55). Также ошибка измерения приводит к ухудшению качества приближения регрессии. Обратите внимание, что R^2 в регрессии (4-55) равен

$$R_*^2 = 1 - (\mathbf{e}'\mathbf{e}/n)/(\mathbf{I}^*\mathbf{M}^0\mathbf{I}^*/n).$$

Числитель сходится к σ_ε^2 , в то время как знаменатель сходится к полной дисперсии I^* , которая может достигать $\sigma_\varepsilon^2 + \beta'Q\beta$, где $Q = \text{plim}(\mathbf{X}'\mathbf{X}/n)$. Таким образом,

$$\text{plim}R_*^2 = \beta'Q\beta/[\sigma_\varepsilon^2 + \beta'Q\beta].$$

Аналогичный показатель для (4-56), R^2 , отличается только тем, что σ_ε^2 в знаменателе заменяется на $\sigma_v^2 > \sigma_\varepsilon^2$. Следовательно,

$$\text{plim}R_*^2 - \text{plim}R^2 > 0.$$

Таким образом, качество приближения, полученного оценкой регрессии (4-56), будет или, по крайней мере, ожидается хуже, чем качество приближения, полученного оценкой регрессии (4-55). (Вывод выше касался асимптотического приближения, так что это неравенство может не выполняться в конкретной конечной выборке.)

Мы показали, какие последствия будет иметь ошибка при измерении зависимой переменной. Заметим, что если ошибка неаддитивна, если она коррелирована с x или обладает другими «плохими» свойствами вроде гетероскедастичности, то эти результаты не будут верными, а о последствиях ошибки измерения в этих случаях ничего сказать с уверенностью нельзя. Существует ли здесь «решение», также неясно. Рассуждения, приведенные выше, показывают, что предпочтительнее было бы иметь точные значения зависимой переменной. При их отсутствии не лучше ли использовать какую-нибудь прокси-переменную? К сожалению, I и так является прокси, поэтому, если нет доступной I' с меньшей дисперсией ошибки измерения, ничего нельзя поделать. С другой стороны, кажется, что вывод довольно безобиден. Выборка содержит меньше информации, чем нам хотелось бы, но имеющейся информации достаточно для построения состоятельной оценки β и для построения соответствующих статистических выводов на этой информации.

Случай, когда с ошибкой измерены независимые переменные, значительно сложнее. Для простоты мы сохраним обозначения I и I^* для наблюдаемой и теоретической переменных соответственно. Возьмем простую парную регрессию

$$y = \beta_1 + \beta_2 I^* + \varepsilon,$$

где y — точно измеренная зависимая переменная, а независимая переменная описывается все тем же уравнением ошибки измерения $I = I^* + w$. Подставляя I в уравнение и преобразовывая, получаем

$$\begin{aligned} y &= \beta_1 + \beta_2 I + (\varepsilon - \beta_2 w) \\ &= \beta_1 + \beta_2 I + v. \end{aligned} \quad (4-57)$$

На первый взгляд кажется, что мы снова пришли к (4-56). К сожалению, это не так, поскольку $\text{Cov}[I, v] = \text{Cov}[I^* + w, \varepsilon - \beta_2 w] = -\beta_2 \sigma_w^2$. Так как регрессор в (4-57) коррелирован с шоками, оценка методом наименьших квадратов в этом случае не является состоятельной. В этой ситуации можно вывести еще некоторые факты, которые мы приведем здесь без доказательства (оно будет приведено в разделе 8.5). В этом случае

$$\text{plim } b_2 = \beta_2 [\sigma_*^2 / (\sigma_*^2 + \sigma_w^2)],$$

где σ_*^2 — безусловная дисперсия I^* . Масштабирующий множитель меньше единицы, поэтому оценка методом наименьших квадратов смещена к нулю. Чем больше дисперсия ошибки измерения, тем больше это смещение. (Оно называется **недооценкой метода наименьших квадратов** (least squares attenuation).) Теперь предположим, что модель содержит более одного регрессора:

$$y = \mathbf{x}'\beta_1 + \beta_2 I^* + \varepsilon.$$

В этом случае мы не сможем привести почти никаких полезных результатов. Можно вывести следующие общие утверждения (доказательство мы снова отложим до раздела 8.5):

1. Оценка β_2 с помощью метода OLS все равно смещена к 0.
2. Несмотря на то что все переменные из \mathbf{x} измерены точно, все компоненты оценки β_1 смещены, причем направления смещений неизвестны.

Существует два типа решения «проблемы наличия ошибок измерения». Если доступна некоторая дополнительная информация о параметрах модели, то иногда удастся получить масштабирующие множители (с помощью **метода моментов** (method of moments)) и скомпенсировать смещение. В качестве очевидного примера: если σ_w^2 в (4-57) была бы известна, то параметр σ_*^2 можно было бы вывести из $\text{Var}[I] = \sigma_*^2 + \sigma_w^2$ и получить множитель, компенсирующий смещение. Такой информации обычно нет. Второй подход, применяемый во многих приложениях, — это метод инструментальных переменных. Этот метод в приложении к проблеме ошибок измерения подробно рассматривается в разделе 8.5.

4.7.6. Влиятельные наблюдения и выбросы

На рис. 4.9 показаны данные о ценах продажи картин Моне, которые использовались в примере 4.10. Выделены две точки. Точка, обозначенная «I» и обведенная в квадрат, соответствует самой маленькой картине в выборке.

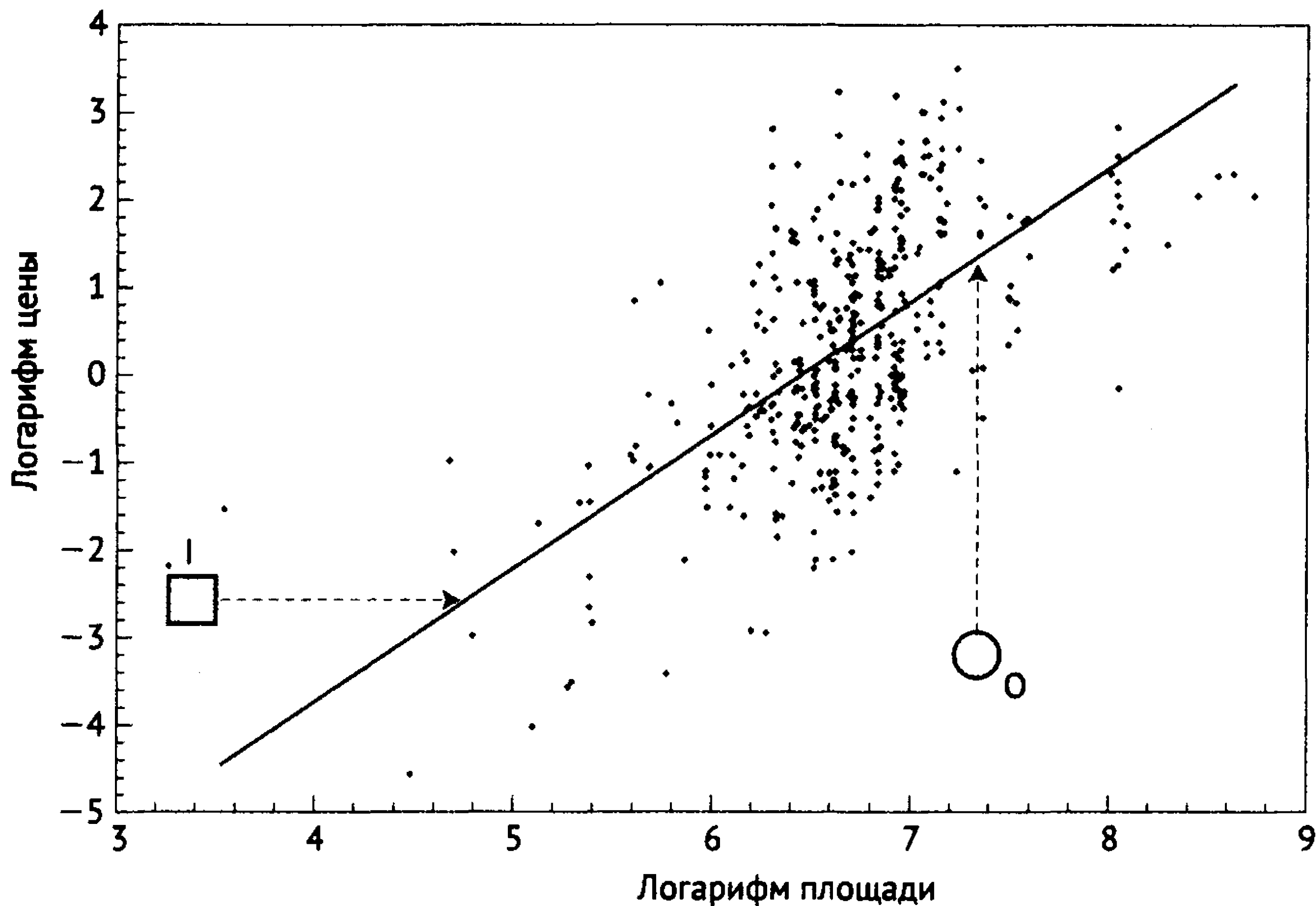


Рис. 4.9. Диаграмма рассеяния логарифмов цены и площади картин Моне

Точка, обозначенная «O» и обведенная в кружок, соответствует неожиданно дешевой (сравнительно с предсказанием регрессии) картине. (Это не самая дешевая картина в выборке, а та, предсказание регрессии для которой оказалось наиболее ошибочным.) Поскольку метод наименьших квадратов основан на квадратах отклонений, оценка, скорее всего, сильно зависит от экстремальных значений вроде этих, особенно если выборка невелика.

«Влиятельное наблюдение» — это такое, которое может иметь значительное влияние на оценки коэффициентов регрессии, полученные при помощи метода наименьших квадратов. Для простой регрессии, показанной на рис. 4.9, в работе Belsley, Kuh и Welsh (1980) была определена мера влиятельности наблюдения i :

$$h_i = \frac{1}{n} + \frac{(x_i - \bar{x}_n)^2}{\sum_{j=1}^n (x_j - \bar{x}_n)^2}, \quad (4-58)$$

где \bar{x}_n и сумма в знаменателе дроби вычисляются без наблюдения i . (Эта мера основывается на разности \mathbf{b} и $\mathbf{b}_{(i)}$, где последнее вычисляется без наблюдения i . Мы скоро к этому вернемся.) Предлагается считать наблюдение влиятельным, если $h_i > 2/n$. Нужно принять решение: отбрасывать это наблюдение или нет. Нужно отметить, что наблюдения с большим «рычагом» необязательно являются «выбросами» (которые мы определим ниже), поскольку анализ проводится условно по x_i . Чтобы это проиллюстрировать, наблюдение «I» на рисунке было бы помечено, даже если бы лежало в точности на линии регрессии — здесь источник влиятельности лежит в числителе второго члена h_i , который не связан с расстоянием до прямой. В нашем примере «влиятельное наблюдение» — результат того, что Моне решил написать маленькую картину. Суть здесь в том, что у нас отсутствует теория, объясняющая экстремальные значения x_i . Устранение таких наблюдений —

это чисто алгебраическое действие, которое приведет к тому, что линия регрессии будет строиться по значениям x_i , наиболее близким к средним.

Изменение вектора коэффициентов множественной линейной регрессии при добавлении к выборке наблюдения равно

$$\mathbf{b} - \mathbf{b}_{(i)} = \Delta \mathbf{b} = \frac{1}{1 + \mathbf{x}'_i \left(\mathbf{X}'_{(i)} \mathbf{X}_{(i)} \right)^{-1} \mathbf{x}_i} \left(\mathbf{X}'_{(i)} \mathbf{X}_{(i)} \right)^{-1} \mathbf{x}_i (y_i - \mathbf{x}'_i \mathbf{b}_{(i)}), \quad (4-59)$$

где оценка \mathbf{b} построена с учетом наблюдения i , а оценка $\mathbf{b}_{(i)}$ — без наблюдения i и $\mathbf{X}_{(i)}$ не включает наблюдение i . (См. упражнение 5 к главе 3.) Трудно выделить характеристики наблюдения, которые сильно влияли бы на это изменение. Мера влиятельности

$$\begin{aligned} h_{ii} &= \mathbf{x}'_i \left(\mathbf{X}'_{(i)} \mathbf{X}_{(i)} \right)^{-1} \mathbf{x}_i \\ &= \frac{1}{n} + \sum_{j=1}^{K-1} \sum_{k=1}^{K-1} (x_{i,j} - \bar{x}_{n,j}) (x_{i,k} - \bar{x}_k) \left(\mathbf{Z}'_{(i)} \mathbf{M}^0 \mathbf{Z}_{(i)} \right)^{jk} \end{aligned} \quad (4-60)$$

обычно используется для определения влиятельных наблюдений [Belsley, Kuh, Welsh (1980) и Cook (1977).] В данном случае критерием было бы $h_{ii} > 2(K-1)/n$. Квадраты отклонений элементов \mathbf{x}_i от средних значений переменных входят в h_{ii} , так что эта мера снова опирается на удаленность наблюдений от средних значений. (Для примера см. выражение для дисперсии прогноза в подразделе 4.6.1.)

Вообще говоря, «выброс» — это наблюдение, которое кажется не подчиняющимся правилам, определяемым моделью. Так, возможно, оно возникает из другого процесса порождения данных. Точка «О» на рис. 4.9 кажется именно такой. Выбросы возникают по разным причинам. Самое простое объяснение — это ошибки в данных. Если предположить, что данные верны, то нужно как-то определить, какое наблюдение является выбросом. Необычно большие остатки являются очевидным кандидатом для подобного выбора. Но поскольку распределение шоков предполагает наличие некоторого количества значительных отклонений, просто убирать неудобные наблюдения будет явной ошибкой. С другой стороны, можно подозревать, что такие наблюдения выбраны из другого распределения. «Стандартизированные» (studentized) остатки строятся с учетом этого факта: оцениваются значения коэффициентов и дисперсия остатков без учета наблюдения i для каждого наблюдения в выборке, а затем модифицированные остатки стандартизируются. i -й стандартизированный остаток равен

$$e(i) = \frac{e_i}{\sqrt{(1 - h_{ii})}} \bigg/ \sqrt{\frac{\mathbf{e}'\mathbf{e} - e_i^2/(1 - h_{ii})}{n - 1 - K}}, \quad (4-61)$$

где \mathbf{e} — вектор остатков для полной выборки, построенный по \mathbf{b} , включающий остаток e_i для наблюдения i . Вообще говоря, этот остаток имеет t -распределение с $n - 1 - K$ степенями свободы (или асимптотически стандартное нормальное распределение). Наблюдения с большими стандартизированными остатками, скажем, большими 2,0, можно считать выбросами.

Этот метод определения выбросов создает ряд осложнений. Во-первых, если мы придерживаемся мнения, что выбросы являются наблюдениями, выбранными из другого распределения, то проблемой будет то, что этот метод не использует никаких априорных предположений о том, какие именно наблюдения считаются выбросами. С теоретической точки зрения это означает, что мы не уверены в спецификации модели. Во-вторых, если выборка содержит значительный процент выбросов, то трудно сказать, каковы будут свойства оценки, построенной по урезанной выборке. В нижеследующем приложении эта процедура удаляет 4,7% выборки (20 наблюдений). Наконец, обычно оказывается, что наблюдения, не являвшиеся выбросами в исходной выборке, становятся выбросами в урезанной выборке. Неясно, как нужно действовать в такой ситуации. (В данных о ценах картин Моне первая процедура стандартизации остатков удаляет 20 наблюдений. После 16 итераций размер выборки стабилизируется на уровне 316 из первоначальных 430 наблюдений, т. е. размер выборки уменьшается на 26,5%.) В табл. 4.9 показаны первоначальные результаты (из табл. 4.6) и результаты оценки регрессии на уменьшенной выборке, из которой были убраны 20 выбросов. Поскольку 430 наблюдений являются сравнительно большой выборкой, ожидается довольно небольшое изменение результатов.

Таблица 4.9. Оценки уравнений для логарифма цены

| | | |
|------------------------------|-----------|-----------|
| Число наблюдений | 430 | 410 |
| Среднее логарифма цены | 0,33274 | 0,36043 |
| Сумма квадратов остатков | 519,17235 | 383,17982 |
| Стандартная ошибка регрессии | 1,10266 | 0,97030 |
| R-квадрат | 0,33620 | 0,39170 |
| Скорректированный R-квадрат | 0,33309 | 0,38871 |

Трудно сделать какой-либо теоретический вывод из результатов этого упражнения. В маленьких выборках обычно требуется осторожность и подробный анализ данных. Если сразу есть подозрение, что процесс, порождающий данные, склонен давать большие величины наблюдений, то можно попробовать использовать другой метод оценивания, например метод наименьших абсолютных отклонений, или вообще другую спецификацию модели, учитывающую подобные возможности. Так, например, идея о том, что выборка может содержать наблюдения, сгенерированные каким-то другим процессом, лежит в основе класса латентных моделей, который мы будем обсуждать в главах 14 и 18.

4.8. Заключение и выводы

В этой главе мы рассмотрели ряд свойств оценок методом наименьших квадратов, применимых ко всем выборкам, включая несмещенность и эффективность в классе несмещенных оценок. Формальные предположения линейной модели являются основой результатов, полученных в этой главе. Все они могут нарушаться в случаях более общих, чем рассмотренный здесь. Например, в большинстве ситуаций, которые мы будем рассматривать в будущем, оценка может иметь смещение, но оно уменьшается с ростом выборки. Для построения доверительных интервалов и проверки гипотез предположение о нормальности распределения возмущений слишком ограничительно, поэтому необходимо расширить модель так, чтобы она допускала распределения шоков, отличные от нормального. Эти и другие расширения модели, применимые в больших выборках, были рассмотрены в разделе 4.4. Важнейшими результатами здесь были состоятельность оценки и способ получения ковариационных матриц и асимптотического распределения, позволяющие строить доверительные интервалы и проверять гипотезы. Статистические выводы в виде интервальных оценок параметров модели и значений зависимой переменной были рассмотрены в разделах 4.5 и 4.6. К этому вопросу мы вернемся в главе 5, где будем говорить о проверке гипотез и выборе модели.

Наконец, мы рассмотрели ряд практических проблем, возникающих при использовании несовершенных данных для построения оценок и анализа регрессионной модели, включая такие проблемы, как мультиколлинеарность, пропущенные данные, ошибки измерения и выбросы.

Основные термины и понятия

| | |
|--|---|
| Assumptions — Предположения | Efficient scale — Эффективный объем производства |
| Asymptotic covariance matrix — Асимптотическая ковариационная матрица | Estimator — Оценка |
| Asymptotic distribution — Асимптотическое распределение | Ex ante forecast — Прогноз ex ante |
| Asymptotic efficiency — Асимптотическая эффективность | Ex post forecast — Прогноз ex post |
| Asymptotic normality — Асимптотическая нормальность | Finite sample properties — Свойства в конечных выборках |
| Asymptotic properties — Асимптотические свойства | Gauss–Markov theorem — Теорема Гаусса–Маркова |
| Attrition — Истощение выборки | Grenander conditions — Условия Гренандера |
| Bootstrap — Бутстрэп (самонастройка) | Highest posterior density interval — Интервал с наибольшей апостериорной вероятностью |
| Condition number — Число обусловленности | Identification — Идентификация |
| Confidence interval — Доверительный интервал | Ignorable case — Пренебрежимый случай |
| Consistency — Состоятельность | Inclusion of superfluous (irrelevant) variables — Включение несущественных (ненужных) переменных |
| Consistent estimator — Состоятельная оценка | Indicator — Индикатор |
| Data imputation — Пополнение данных | |

- Interval estimation** — Интервальное оценивание
- Least squares attenuation** — Недооценка метода наименьших квадратов
- Lindeberg-Feller Central Limit Theorem** — Центральная предельная теорема Линдберга-Феллера
- Linear estimator** — Линейная оценка
- Linear unbiased estimator** — Линейная несмещенная оценка
- Maximum likelihood estimator** — Оценка методом максимального правдоподобия
- Mean absolute error** — Средняя абсолютная ошибка
- Mean square convergence** — Сходимость в среднеквадратичном
- Mean squared error** — Средний квадрат ошибки
- Measurement error** — Ошибка измерения
- Method of moments** — Метод моментов
- Minimum mean squared error** — Минимальная среднеквадратичная ошибка
- Minimum variance linear unbiased estimator** — Линейная несмещенная оценка с наименьшей дисперсией
- Missing at random** — Случайные пропуски
- Missing completely at random** — Полностью случайные пропуски
- Missing observations** — Пропущенные наблюдения
- Modified zero-order regression** — Модифицированный метод нулевого порядка
- Monte Carlo study** — Исследования методом Монте-Карло
- Multicollinearity** — Мультиколлинеарность
- Not missing at random** — Неслучайные пропуски
- Oaxaca's and Blinder's decomposition** — Разложение Охака и Блаиндера
- Omission of relevant variables** — Пропуск существенных (значимых) переменных
- Optimal linear predictor** — Оптимальный линейный предиктор
- Orthogonal random variables** — Ортогональные случайные величины
- Panel data** — Панельные данные
- Pivotal statistic** — Базисная статистика
- Point estimation** — Точечная оценка
- Prediction error** — Ошибка предсказания
- Prediction interval** — Доверительный интервал для предсказания
- Prediction variance** — Дисперсия предсказания
- Pretest estimator** — Предварительная оценка
- Principal components** — Главные компоненты
- Probability limit** — Предел по вероятности
- Root mean squared error** — Квадратный корень из среднеквадратичной ошибки
- Sample selection** — Выборочная селективность
- Sampling distribution** — Выборочное распределение
- Sampling variance** — Выборочная дисперсия
- Semiparametric** — Полупараметрический
- Smearing estimator** — Размытая оценка
- Specification errors** — Ошибки спецификации
- Standard error** — Стандартная ошибка
- Standard error of the regression** — Стандартная ошибка регрессии
- Stationary process** — Стационарный процесс
- Statistical properties** — Статистические свойства
- Stochastic regressors** — Стохастические (случайные) регрессоры
- Theil statistic** — Статистика Тейла
- t ratio** — t -статистика
- Variance inflation factor** — Множитель наращения дисперсии
- Zero-order method** — Метод нулевого порядка

Упражнения

1. Пусть есть две независимые несмещенные оценки параметра θ , $\hat{\theta}_1$ и $\hat{\theta}_2$, с различными дисперсиями, v_1 и v_2 . Какая из линейных комбинаций вида

$\hat{\theta} = c_1\hat{\theta}_1 + c_2\hat{\theta}_2$ будет несмещенной оценкой с наименьшей дисперсией для θ ?

2. Рассмотрим парную регрессию $y_i = \beta x_i + \varepsilon_i$, с $E[\varepsilon | x] = 0$ и $E[\varepsilon^2 | x] = \sigma^2$.

а. Найдите линейную оценку β с минимальным среднеквадратичным отклонением. [Подсказка. Пусть оценка имеет вид $(\hat{\beta} = c'y)$. Выберите c , минимизирующее $\text{Var}(\hat{\beta}) + (E(\hat{\beta} - \beta))^2$. Ответ является функцией неизвестных параметров.]

б. Для оценки из пункта а) покажите, что отношение среднеквадратичной ошибки $\hat{\beta}$ к среднеквадратичной ошибке оценки методом наименьших квадратов b равно

$$\frac{\text{MSE}[\hat{\beta}]}{\text{MSE}[b]} = \frac{\tau^2}{(1 + \tau^2)}, \quad \text{где } \tau^2 = \frac{\beta^2}{[\sigma^2/\mathbf{x}'\mathbf{x}]}.$$

Заметим, что τ — теоретический аналог для t -статистики, проверяющей гипотезу о том, что $\beta = 0$, приведенной в (5-11). Как можно интерпретировать поведение этого отношения при $\tau \rightarrow \infty$?

3. Предположим, что классическая модель регрессии применима, но истинное значение константы равно 0. Сравните дисперсию оценки методом наименьших квадратов коэффициентов наклона, построенных по уравнению без постоянного члена, с соответствующей дисперсией оценки, построенной по уравнению с (излишним) постоянным членом.

4. Рассмотрим уравнение регрессии $y_i = \alpha + \beta x_i + \varepsilon_i$, где шоки ε_i имеют плотность $f(\varepsilon_i) = (1/\lambda) \exp(-\varepsilon_i/\lambda)$, $\varepsilon_i \geq 0$. Эта модель несколько необычна, поскольку все шоки в этой модели неотрицательны. Заметим, что $E[\varepsilon_i | x_i] = \lambda$, а $\text{Var}[\varepsilon_i | x_i] = \lambda^2$. Покажите, что оценка OLS коэффициента наклона является несмещенной, а оценка свободного члена смещена.

5. Докажите, что оценка OLS для свободного члена в классической модели регрессии является линейной несмещенной оценкой с наименьшей дисперсией.

6. Как монополист, максимизирующий прибыль, вы наблюдаете функцию спроса $Q = \alpha + \beta P + \varepsilon$. В прошлом вы устанавливали следующие цены и продавали следующие соответствующие количества товара:

| | | | | | | | | | | | | | | | |
|-----|----|----|----|----|----|----|----|----|----|----|---|----|----|----|----|
| Q | 3 | 3 | 7 | 6 | 10 | 15 | 16 | 13 | 9 | 15 | 9 | 15 | 12 | 18 | 21 |
| P | 18 | 16 | 17 | 12 | 15 | 15 | 4 | 13 | 11 | 6 | 8 | 10 | 7 | 7 | 7 |

Пусть ваши предельные издержки равны 10. Постройте с помощью метода наименьших квадратов 95%-й доверительный интервал для ожидаемого значения выпуска, максимизирующего прибыль.

7. Следующие выборочные моменты $x = [1, x_1, x_2, x_3]$ были вычислены на основе 100 наблюдений, полученных с помощью генератора случайных чисел:

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} 100 & 123 & 96 & 109 \\ 123 & 252 & 125 & 189 \\ 96 & 125 & 167 & 146 \\ 109 & 189 & 146 & 168 \end{bmatrix}, \quad *Xy = \begin{bmatrix} 460 \\ 810 \\ 615 \\ 712 \end{bmatrix}, \quad *y'y = 3924.$$

Истинная модель, описывающая эти данные, имеет вид $y = x_1 + x_2 + x_3 + \varepsilon$.

- a. Вычислите коэффициенты корреляции между регрессорами.
 - b. Вычислите МНК-оценки коэффициентов в регрессии y на константу x_1 , x_2 и x_3 .
 - c. Вычислите МНК-оценки коэффициентов в регрессиях y на константу x_1 и x_2 и на константу x_2 и x_3 .
 - d. Вычислите множитель наполнения дисперсии (VIF) для каждого регрессора.
 - e. Регрессоры очевидно коллинеарны. Из-за какой переменной возникает проблема?
8. Рассмотрим множественную регрессию y на K переменных \mathbf{X} и дополнительную переменную z . Докажите, что при выполнении условий А1–А6 классической модели регрессии истинное значение дисперсии OLS оценки коэффициентов при \mathbf{X} больше, когда z включена в оцениваемое уравнение регрессии, чем при ее отсутствии. Будет ли то же самое верно для выборочной оценки ковариационной матрицы? Почему? Предполагайте, что \mathbf{X} и z являются нестохастическими и коэффициент при z не равен 0.
 9. В классической нормальной модели регрессии $y = \mathbf{X}\beta + \varepsilon$ без константы и с K регрессорами, предполагая, что истинное значение коэффициента β равно нулю, вычислите точное ожидаемое значение величины $F[K, n - K] = (R^2/K) / [(1 - R^2)/(n - K)]$.
 10. Докажите, что $E[\mathbf{b}'\mathbf{b}] = \beta'\beta + \sigma^2 \sum_{k=1}^K (1/\lambda_k)$, где \mathbf{b} — OLS-оценка, а λ_k — собственное значение матрицы $\mathbf{X}'\mathbf{X}$.
 11. Для классической нормальной регрессии $y = \mathbf{X}\beta + \varepsilon$ без константы и с K регрессорами, чему равно $\text{plim } F[K, n - K] = \text{plim } \frac{R^2/K}{(1 - R^2)/(n - K)}$, если истинное значение β равно 0?
 12. Пусть e_i — i -й остаток метода наименьших квадратов в регрессии y на \mathbf{X} в классической модели регрессии и пусть ε_i — соответствующий шок. Докажите, что $\text{plim}(e_i - \varepsilon_i) = 0$.
 13. В простой модели регрессии $y_i = \mu + \varepsilon_i$, $\varepsilon_i \sim N[0, \sigma^2]$ докажите, что выборочное среднее состоятельно и асимптотически нормально распределено. Теперь рассмотрите альтернативную оценку

$$\hat{\mu} = \sum_i w_i y_i, \quad w_i = \frac{i}{(n(n+1)/2)} = \frac{i}{\sum_i i}.$$

Заметьте, что $\sum_i w_i = 1$. Докажите, что это состоятельная оценка μ , и выведите ее асимптотическую дисперсию. [Подсказка: $\sum_i i^2 = n(n+1)(2n+1)/6$.]

14. Возьмем выборку из n наблюдений, из которых n_c полных, а n_m неполных, в которых пропущено значение зависимой переменной y_i . Данные о значениях регрессоров x_i полны для всех n наблюдений, \mathbf{X}_c и \mathbf{X}_m . Мы хотим с помощью этих данных оценить параметры линейной регрессии $y = \mathbf{X}\beta + \varepsilon$. Рассмотрим следующую стратегию пополнения. Шаг 1: оценим линейную регрессию y_c на \mathbf{X}_c и вычислим \mathbf{b}_c . Шаг 2: с помощью \mathbf{X}_m

построим предсказания значений y_m по $X_m b_c$, затем оценим регрессию полной выборки наблюдений $(y_c, X_m b_c)$ на полную выборку регрессоров (X_c, X_m) .

- a. Покажите, что оценки векторов коэффициентов, полученные на первом и на втором шаге, совпадают.
 - b. Является ли оценка коэффициентов, полученная на втором шаге, несмещенной?
 - c. Покажите, что сумма квадратов остатков одинакова для обеих оценок.
 - d. Покажите, что оценка σ^2 , полученная на втором шаге, смещена вниз.
15. В (4-13) мы видели, что если лишние переменные X_2 добавляются к регрессии y на X_1 , то OLS-оценка вектора коэффициентов является несмещенной оценкой истинного вектора параметров $\beta = (\beta_1', 0')'$. Покажите, что в этой длинной регрессии $e'e/(n - K_1 - K_2)$ также является несмещенной оценкой σ^2 .
16. В подразделе 4.7.3 мы рассматривали регрессию y на набор главных компонент, а не на исходные данные. Для простоты допустим, что X не содержит постоянного члена и что K переменных измерены в своих отклонениях от среднего и стандартизированы путем деления на соответствующие стандартные отклонения. Рассмотрим регрессию y на L главных компонент, $Z = X C_L$, где $L < K$. Пусть d обозначает вектор коэффициентов. Уравнение регрессии имеет вид $y = X\beta + \epsilon$. Выше утверждалось, что $E[d] = C_L' \beta$. Докажите это утверждение.
17. В примере 4.10 была рассмотрена регрессионная модель, по которой можно предсказать цены на картины Моне. Самая дорогая картина была продана за \$33,0135М ($\log = 17,3124$). Высота и ширина этой картины равнялись 35 и 39,4 дюйма соответственно. Постройте с помощью этих данных и модели доверительные интервалы для предсказаний логарифма цены и самой цены этой картины.

Эмпирические приложения

1. В табл. F2.2 приведены данные по потреблению бензина в США с 1953 по 2004 г. Заметим, что потребление здесь измерено как величина суммарных расходов на бензин. Чтобы получить величину потребления на душу населения, поделите показатель GASEXP на величину GASP, умноженную на Pop. Другие переменные не нуждаются в преобразовании.
 - a. Оцените множественную регрессию душевого потребления бензина на доход на душу населения, цену бензина, другие цены и временной тренд. Приведите все полученные результаты. Соответствуют ли знаки полученных коэффициентов вашим ожиданиям?
 - b. Проверьте гипотезу о том, что с точки зрения спроса на бензин для потребителей не имеет значения, изменяются цены на новые или подержанные автомобили.
 - c. Оцените эластичность спроса на бензин по цене бензина, эластичность по доходу и перекрестную эластичность по цене общественного транспорта. Вычислите их в точке, соответствующей 2004 г.

- d. Оцените регрессию в логарифмах, чтобы значения коэффициентов давали непосредственно оценки эластичностей. (Не нужно использовать логарифм временного тренда.) Сравните полученные значения с оценками из предыдущего пункта. Какая спецификация кажется вам более предпочтительной?
- e. Вычислите простые корреляции между различными переменными цен. Существует ли проблема мультиколлинеарности для регрессий из пункта а или пункта d?
- f. Заметьте, что индекс цен на бензин нормализован таким образом, что он равен 100 в 2000 г., в то время как остальные цены отнормированы (примерно) к 1983 г. Как изменятся результаты регрессии пункта а, если нормализовать все индексы так, чтобы они имели значение 100 в 2004 г.? Как при этом изменятся результаты регрессии в пункте d?
- g. Это упражнение основано на модели, оцененной в пункте d. Нас интересуют изменения, происшедшие на рынке бензина в 1973 г. Во-первых, вычислите средние логарифмов потребления бензина на душу населения в 1953–1973 и 1974–2004 гг. и их разность. Если мы поделим выборку на эти две группы наблюдений, то, как показано в подразделе 4.5.3, можем разделить изменение в математическом ожидании логарифма потребления на изменение, связанное с регрессорами, и на изменение, связанное с изменениями параметров модели. Используя описанный в этом разделе подход Охаки–Блаиндера, вычислите разложение, разделив выборку на две части и отдельно оценив регрессию для каждой части. Используя полученные результаты, постройте доверительный интервал для величины изменений, вызванных структурным изменением рынка, т. е. изменением коэффициентов регрессии.
2. В работе Кристенсена и Грина (Christensen, Greene (1976)) была оценена обобщенная функция Кобба–Дугласа для функции затрат на производство электричества в виде

$$\ln C = \alpha + \beta \ln Q + \gamma \left[\frac{1}{2} (\ln Q)^2 \right] + \delta_k \ln P_k + \delta_l \ln P_l + \delta_f \ln P_f + \varepsilon,$$

где P_k , P_l и P_f обозначают цены единицы капитала, труда и топлива соответственно, Q — объем производства, а C — суммарные издержки. Для соответствия требованиям теории производства нужно потребовать, чтобы функция издержек была однородна первой степени по всем трем ценам. Для этого добавим условие $\delta_k + \delta_l + \delta_f = 1$ или $\delta_f = 1 - \delta_k - \delta_l$. Подстановка этого условия в функцию издержек после преобразований даст следующее оцениваемое уравнение:

$$\ln(C/P_f) = \alpha + \beta \ln Q + \gamma \left[\frac{1}{2} (\ln Q)^2 \right] + \delta_k \ln(P_k/P_f) + \delta_l \ln(P_l/P_f) + \varepsilon.$$

Целью этого обобщения было получение кривой средних совокупных издержек U-образного вида. [См. пример 6.6 для описания более раннего исследования Нерлова (Nerlove (1963)), которое послужило отправной точкой для этой работы.] Нас интересует **эффективный объем производства** (efficient scale), при котором кривая издержек достигает минимума. В этой точке $(\partial \ln C / \partial \ln Q)|_{Q=Q^*} = 1$ или $Q^* = \exp[(1 - \beta)/\gamma]$.

- a. В табл. F4.4 представлены данные по 158 фирмам из исследования Кристенсена и Грина (Christensen, Greene (1976)). Постройте оценку параметров функции издержек и оцените асимптотическую ковариационную матрицу по этим 158 наблюдениям.
- b. Обратите внимание, что из функции издержек нельзя непосредственно получить оценку δ_f . Постройте такую оценку с помощью результатов вашей регрессии и оцените асимптотическую стандартную ошибку.
- c. Постройте оценку для Q^* , используя результаты оценки вашей регрессии, а затем построьте доверительный интервал для оцененного эффективного объема производства.
- d. Используя исходные данные, найдите, где в выборке лежит эффективный объем производства. Таким образом, определите, сколько фирм из выборки достигло этого объема и насколько, по вашему мнению, велик этот объем по сравнению с размерами фирм в выборке. В работе Кристенсена и Грина (Christensen, Greene (1976)) этот вопрос был исследован путем вычисления того, какую часть общего выпуска выборки производят фирмы, не достигшие эффективного объема производства. (*Обратите внимание:* в выборке есть двойной подсчет — более 20 крупнейших фирм являются холдингами или объединениями, включающими другие фирмы из выборки. В рамках этого вычислительного упражнения это обстоятельство можно проигнорировать.)

ГЛАВА 5

Тестирование гипотез и выбор спецификации

5.1. Введение

Модель линейной регрессии используется для трех важных целей: оценивания и предсказания, которые освещались в предыдущей главе, и тестирования гипотез. В этой главе будут рассмотрены некоторые приложения тестирования гипотез с использованием линейной регрессионной модели. Мы начнем с методологической и статистической теории. Часть этой теории была представлена в главе 4 (включая идею основной (базисной) статистики в подразделе 4.5.1) и в приложении С.7. В разделе 5.2 мы обобщим методологию на случай тестирования на основе регрессионной модели. После изложения теории в разделах 5.3–5.7 будут рассмотрены некоторые приложения регрессионного анализа. Это обобщение будет касаться ограничений, накладываемых на параметры модели, таких как «релевантность» переменной (т.е. наличие при ней ненулевого коэффициента) или адекватность модели данным (т.е. соответствуют ли данные гипотезе о том, что все коэффициенты равны нулю). Основным интересом в обсуждении будут представлять линейные ограничения. Мы обратимся к нелинейным ограничениям в конце изложения, в разделе 5.7. В разделе 5.8 рассматривается более широкий класс гипотез, включающий такие, как выбор из двух моделей, например, между линейной и логлинейной. В каждом из приведенных случаев процедура тестирования разрешала сопоставление двух теорий происхождения данных; в разделах 5.2–5.7 — узкой и широкой моделей, а в разделе 5.8 примерно одинаковых моделей. В разделе 5.9 иллюстрируется конкретный **тест спецификации**, который в сущности представляет тест такого утверждения, как «модель корректна», против утверждения «модель неадекватна». Этот тест позволяет сравнить теорию, лежащую в основе модели, с «некоторой другой, неназванной теорией». Наконец, в разделе 5.10 представлены некоторые общие принципы и элементы стратегии тестирования модели и выбора спецификации.

5.2. Методология тестирования гипотез

Мы начинаем анализ с регрессионной модели как утверждения о том, что

$$y = X\beta + \epsilon. \quad (5-1)$$

В качестве конкретного приложения пример 4.6 иллюстрировал аукционные цены картин:

$$\ln Price = \beta_1 + \beta_2 \ln Size + \beta_3 AspectRatio + \varepsilon. \quad (5-2)$$

Некоторые вопросы могут возникнуть относительно «модели» в (5-2), в сущности, о переменных. Естественно предположить, что ценители изобразительного искусства будут интересоваться соотношением сторон, которое является элементом эстетической ценности картины. Но идея о том, что размер должен влиять на цену, противостоит природе искусства, особенно учитывая удивительно малые размеры некоторых наиболее известных картин, таких как *Мона Лиза* (30" в высоту и 21" в ширину) или *Постоянство памяти* Дали (лишь 9,5" в высоту и 13" в ширину). Скептик может усомниться в наличии $\ln Size$ в уравнении или, другими словами, ненулевого коэффициента β_2 . Чтобы разрешить эту проблему, можно поставить вопрос следующим образом: соответствует ли специфицированное уравнение данным, т. е. наблюдаемым ценам, по которым продаются картины? Чтобы ответить на этот вопрос, естественным представляется сначала оценить регрессию и затем рассмотреть оценку β_2 . «Тест» в этом случае будет состоять в том, равен ли β_2 нулю в регрессии, оцененной методом наименьших квадратов, или нет. Понимая, что наклон оценки линейной регрессии — это случайная величина, которая никогда не будет в точности равна нулю, даже если нулю на самом деле равен коэффициент β_2 , мы рассмотрим более мягкую формулировку — достаточно ли выборочная оценка близка к нулю для того, чтобы мы сделали вывод о том, что ее теоретический аналог на самом деле ноль, т.е. что ненулевое значение, которое мы наблюдаем, — это не что иное, как шум, вызванный выборочной вариацией. Без ответа остаются вопросы: как близка к нулю должна быть оценка, чтобы мы могли сделать этот вывод? в чем измерять эту близость? в какой степени мы можем быть уверены, что сделали правильный вывод? (Конечно, не полностью.) насколько вероятно, что правило принятия решения, которое мы выберем, приведет к неверному выводу? Этот раздел формализует эти идеи. После детального изложения методологии будет рассмотрено несколько численных примеров.

5.2.1. Ограничения и гипотезы

В рамках нашего подхода мы будем формулировать гипотезу в форме ограничения на модель. Таким образом, в рамках обсуждаемой здесь классической методологии модель является утверждением общего характера, а гипотеза — предложением, конкретизирующим это утверждение. В примере с изобразительным искусством в (5-2) таким дополнительным утверждением является $\beta_2 = 0$ — без комментариев относительно β_1 или β_3 . Мы определяем **нулевую гипотезу** как утверждение, конкретизирующее модель, и **альтернативную гипотезу** как более общее утверждение. В рассматриваемом примере более общая модель позволяет уравнению содержать $\ln Size$ и $AspectRatio$ — она допускает возможность того, что любой из коэффициентов может быть равен нулю, но не настаивает на этом. Нулевая гипотеза настаивает на том, что $\beta_2 = 0$, но не утверждает ничего относительно β_1 или β_3 . Формальные обозначения для фиксирования этой гипотезы выглядят как

$$\begin{aligned} \ln Price &= \beta_1 + \beta_2 \ln Size + \beta_3 AspectRatio + \varepsilon, \\ H_0: \beta_2 &= 0, \\ H_1: \beta_2 &\neq 0. \end{aligned} \quad (5-3)$$

Отметим, что нулевая и альтернативная гипотезы вместе взаимоисключающи и исчерпывающи. Третьей возможности нет; верна либо одна, либо другая, но не обе.

Начиная с этого момента, анализ будет посвящен сопоставлению нулевой гипотезы с данными. Данные могут убедить эконометриста отвергнуть нулевую гипотезу. В этот момент может показаться уместным принятие альтернативы. Однако в интересах поддержания гибкости методологии, т.е. открытости к новой информации, уместным выводом в данном случае будет отвержение или неотвержение нулевой гипотезы. Неотвержение нулевой гипотезы неэквивалентно ее принятию, хотя формулировка и может намекать на это. Принимая нулевую гипотезу, мы неявно отказываемся от дальнейшего продолжения исследования. Таким образом, традиционная, классическая методология оставляет возможность того, что дополнительная информация может поменять выводы. Методология нашего тестирования будет сформулирована таким образом, чтобы или отвергнуть H_0 : данные несовместимы с гипотезой с существенной долей уверенности, или не отвергать H_0 : данные совместимы с гипотезой.

5.2.2. Вложенные модели

Общий подход к тестированию гипотезы состоит в том, чтобы сформулировать статистическую модель, которая включает гипотезу в качестве ограничения на свои параметры. Теория имеет **тестируемые утверждения**, если из нее следуют какие-либо тестируемые ограничения модели. Рассмотрим, например, модель инвестиций, I_t ,

$$\ln I_t = \beta_1 + \beta_2 i_t + \beta_3 \Delta p_t + \beta_4 \ln Y_t + \beta_5 t + \varepsilon_t, \quad (5-4)$$

которая говорит о том, что инвесторы чувствительны к номинальной ставке процента i_t , темпу инфляции Δp_t , логарифму реального выпуска $\ln Y_t$ и другим факторам, растущим со временем, включаемым во временной тренд, t . Альтернативная теория говорит о том, что «инвесторы интересуются реальной процентной ставкой». В этом случае альтернативная модель имеет вид

$$\ln I_t = \beta_1 + \beta_2(i_t - \Delta p_t) + \beta_3 \Delta p_t + \beta_4 \ln Y_t + \beta_5 t + \varepsilon_t. \quad (5-5)$$

Несмотря на то что эта новая модель отражает вторую теорию, уравнение все же содержит номинальную процентную ставку и инфляцию. Теория не порождает тестируемые утверждения для нашей модели. Но рассмотрим более сильную гипотезу «инвесторы интересуются *только* реальной процентной ставкой». Получаемое уравнение

$$\ln I_t = \beta_1 + \beta_2(i_t - \Delta p_t) + \beta_4 \ln Y_t + \beta_5 t + \varepsilon_t \quad (5-6)$$

теперь ограничено; в контексте (5-4) тестируемое утверждение состоит в том, что $\beta_2 + \beta_3 = 0$. Более сильное утверждение подразумевает конкретное соотношение параметров уравнения, которое может подтверждаться, но может и опровергаться эмпирическими наблюдениями.

Описание тестируемых утверждений в предыдущем параграфе (верно) указывает на то, что тестируемые ограничения подразумевают, что только некоторые из возможных моделей, допустимых исходной спецификацией, будут действительны, т.е. совместимы с теорией. В предыдущем примере (5-4) специфицирует модель с пятью неограниченными параметрами $(\beta_1, \beta_2, \beta_3, \beta_4, \beta_5)$. Но (5-6) показывает, что только некоторые значения параметров совместимы с теорией, а именно те, для которых верно $\beta_3 = -\beta_2$. Это подмножество значений содержится внутри неограниченного множества. Поэтому модели называются **вложенными (nested)**. Рассмотрим другую гипотезу — «инвесторы не беспокоятся об инфляции». В этом случае меньший набор коэффициентов — это $(\beta_1, \beta_2, 0, \beta_4, \beta_5)$. И снова ограничения подразумевают подходящее **пространство параметров** «меньшее» (имеющее меньшее число измерений), чем неограниченное. Общий результат состоит в том, что гипотеза, специфицируемая ограниченной моделью, содержится внутри неограниченной модели. Теперь рассмотрим альтернативную пару моделей: модель₀: «инвесторов интересует только инфляция»; модель₁: «инвесторов интересует только номинальная процентная ставка». В этом случае два вектора параметров — это $(\beta_1, 0, \beta_3, \beta_4, \beta_5)$ в модель₀ и $(\beta_1, \beta_2, 0, \beta_4, \beta_5)$ в модель₁. В этом случае обе спецификации являются подмножествами неограниченной модели, но ни одна из них не может быть получена наложением ограничений на другую. Две эти модели **не вложены**. Пока что нас интересуют только вложенные модели. Невложенные модели рассматриваются в разделе 5.8.

5.2.3. Процедуры тестирования — методология Неймана–Пирсона

В примере, приведенном в (5-2), интуиция подсказывает подход к тестированию, основанный на сопоставлении данных и гипотезы. Основа методологии, предложенной в работе Неймана и Пирсона (Neuman, Pearson (1933)), предоставляет надежное основание для тестирования гипотез в постановке, рассматриваемой нами в этой главе. В целом исследователь следует логике, основанной на вопросе: какие данные позволят мне отвергнуть гипотезу? Учитывая формулировку гипотезы в подразделе 5.2.1, он эквивалентен вопросу о том, какого рода данные поддержат модель. Наблюдаемые данные делятся на **область отвержения** и **область принятия**. Процедура тестирования сводится к рассмотрению статистических доводов. Как только определился вид области отвержения, если наблюдаемые данные относятся к нему, гипотеза отвергается. Для того чтобы увидеть, как этот подход работает на практике, рассмотрим снова гипотезу о размере в уравнении цены на произведения искусства. Мы проверяем гипотезу о том, что β_2 равно нулю. Оценим наклон регрессии методом наименьших квадратов. Определим заранее, как далеко от нуля должна быть оценка β_2 , чтобы привести к отвержению нулевой гипотезы. Как только правило определено, сам тест работает механически. В частности, в этом случае b_2 далека от нуля, если $b_2 > \beta_2^{0+}$ или $b_2 < \beta_2^{0-}$. Если происходит одно из этих событий, гипотеза отвергается. Принципиальным является то, что правило определяется заранее.

5.2.4. Размер, мощность и состоятельность теста

Поскольку процедура тестирования определяется заранее и оценки коэффициентов в регрессии случайны, метод Неймана–Пирсона может привести к ошибкам в двух случаях. В контексте конкретных чисел выборочная регрессия, соответствующая (5-2), приведена в табл. 4.6. Оценка коэффициента при $\ln Area$ равна 1,33372 с оценкой стандартной ошибки, равной 0,09072. Предположим, правило для тестирования уже определено (мы формализуем его определение ниже) как следующее: если b_2 больше +1,0 или меньше, чем -1,0, то мы отвергаем гипотезу о том, что коэффициент равен нулю (и делаем вывод о том, что покупатели произведений искусства действительно интересуются размером картин). Таким образом, на основании этого правила мы действительно отвергнем гипотезу. Однако, поскольку b_2 — это случайная величина, возможны следующие ошибки:

- ошибка I рода: $\beta_2 = 0$, но мы отвергаем гипотезу. Нулевая гипотеза ошибочно отвергается;
- ошибка II рода: $\beta_2 \neq 0$, но мы не отвергаем гипотезу.

Нулевая гипотеза ошибочно не отвергается. Вероятность ошибки I рода называется **размером теста**. Размер теста — это вероятность того, что тест ошибочно отвергнет нулевую гипотезу. Как будет ясно позднее, исследователь определяет ее заранее. Единица минус вероятность ошибки II рода называется **мощностью теста**. Мощность теста — это вероятность того, что он правильно отвергнет неверную нулевую гипотезу. Мощность теста зависит от альтернативы. Она не подконтрольна исследователю. Возвращаясь к примеру, мы отвергнем гипотезу, если $|b_2| > 1$. Если коэффициент β_2 на самом деле равен 1,5, то на основе увиденных нами результатов мы скорее всего получим значение b_2 , превышающее 1,0. С другой стороны, если β_2 равен всего лишь 0,3, то не кажется вероятным, что мы получим выборочное значение, превышающее 1,0. Таким образом, снова мощность теста зависит от истинных значений, лежащих в основе данных. Идея мощности теста относится к его способности найти то, что он ищет.

Процедура тестирования **состоятельна**, если ее мощность стремится к 1,0 при стремлении размера выборки к бесконечности. Это качество легко отметить снова в контексте одного параметра, такого как рассматриваемый здесь. Поскольку метод наименьших квадратов дает состоятельную оценку, следовательно, с ростом выборки мы сможем узнать истинное значение коэффициента β_2 , так что нам будет известно, равен ли он нулю. Таким образом, в этом примере ясно, что с ростом выборки мы точно узнаем, нужно ли отвергнуть гипотезу. Для большей части этой книги мы можем пользоваться следующим правилом: процедура тестирования, относящаяся к параметрам модели, состоятельна, если она основана на состоятельной оценке этих параметров. Поскольку почти вся книга основана на состоятельных оценках, и, за исключением последних разделов этой главы, где наши тесты будут касаться параметров вложенных моделей, наши тесты будут состоятельными.

5.2.5. Методологическая дилемма: байесовское тестирование против классического

Как мы отметили ранее, методология тестирования Неймана–Пирсона, которую мы будем использовать, — это предложение вида «все или ничего». Мы определим правила тестирования заранее, соберем данные и либо отвергнем, либо не будем отвергать нулевую гипотезу. Третьего не дано. Это ставит исследователя перед двумя неудобными дилеммами. Во-первых, тестируемый исход, т. е. выборочные данные могут оказаться расположенными слишком близко к границе области отвержения. Рассмотрим наш пример. Если бы мы заранее решили отвергнуть нулевую гипотезу, если $b_2 > 1,00$, а выборочное значение составило 0,9999, будет сложно устоять перед соблазном все же отвергнуть гипотезу, в частности, если мы начали анализ, имея четкое убеждение, что нулевая гипотеза неверна. (То есть независимо от интуиции я убежден, что покупатели произведений искусства действительно интересуются размером.) Во-вторых, изложенная здесь методология не позволяет подключать другие исследования. Развиваем наш пример: если бы я был десятым исследователем, работающим с рынком произведений искусства, и предыдущие девять уверенно отвергли гипотезу о том, что $\beta_2 = 0$, мне было бы очень трудно не отвергать эту гипотезу, даже если бы мое исследование, основанное на моей процедуре тестирования, показывало, что я не должен этого делать.

Эта дилемма встроена в классическую методологию тестирования. Но есть и промежуточный вариант. Байесовская методология, которую мы обсудим в главе 16, не упирается в эту дилемму, поскольку байесовские исследователи никогда не приходят к точным выводам. Они только обновляют свои априорные представления. Таким образом, в первом случае, когда наблюдаемые данные близки к области отвержения, исследователь просто обновит априорное представление не столь убедительным доказательством, как хотелось бы. Но методология допускает такую возможность. Для второго случая нам доступно много информации в пользу отвержения H_0 . Потребуется серьезный десятый довод для изменения предыдущих девяти выводов. Результаты десятого исследования (постериорные результаты) будут включать не только текущие доказательства, но и все априорные данные.

5.3. Два подхода к тестированию гипотез

Общая линейная гипотеза — это набор J -ограничений, накладываемых на модель линейной регрессии:

$$y = X\beta + \varepsilon.$$

Ограничения записываются как

$$\begin{aligned} r_{11}\beta_1 + r_{12}\beta_2 + \cdots + r_{1K}\beta_K &= q_1; \\ r_{21}\beta_1 + r_{22}\beta_2 + \cdots + r_{2K}\beta_K &= q_2; \\ &\dots \\ r_{J1}\beta_1 + r_{J2}\beta_2 + \cdots + r_{JK}\beta_K &= q_J. \end{aligned} \tag{5-7}$$

В простейшем случае это одно ограничение на один коэффициент:

$$\beta_k = 0.$$

Более общий случай может быть записан в матричной форме:

$$\mathbf{R}\beta = \mathbf{q}. \quad (5-8)$$

Каждая строка \mathbf{R} — это коэффициенты в одном из ограничений. Обычно \mathbf{R} будет иметь лишь несколько строк и много нулей в каждой строке. Вот некоторые примеры:

1. Один из коэффициентов равен нулю, $\beta_j = 0$,

$$\mathbf{R} = [0 \ 0 \ \dots \ 1 \ 0 \ \dots \ 0] \text{ и } \mathbf{q} = 0.$$

2. Два коэффициента равны, $\beta_k = \beta_j$,

$$\mathbf{R} = [0 \ 0 \ 1 \ \dots \ -1 \ \dots \ 0] \text{ и } \mathbf{q} = 0.$$

3. Сумма нескольких коэффициентов равна единице, $\beta_2 + \beta_3 + \beta_4 = 1$,

$$\mathbf{R} = [0 \ 1 \ 1 \ 1 \ 0 \ \dots] \text{ и } \mathbf{q} = 1.$$

4. Несколько коэффициентов равны нулю, $\beta_1 = 0$, $\beta_2 = 0$ и $\beta_3 = 0$,

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \end{bmatrix} = [\mathbf{I} \ \mathbf{0}] \text{ и } \mathbf{q} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

5. Несколько линейных ограничений, $\beta_2 + \beta_3 = 1$, $\beta_4 + \beta_6 = 0$ и $\beta_5 + \beta_6 = 0$,

$$\mathbf{R} = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \text{ и } \mathbf{q} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}.$$

6. Все коэффициенты модели, за исключением константы, равны нулю:

$$\mathbf{R} = [\mathbf{0} : \mathbf{I}_{K-1}] \text{ и } \mathbf{q} = \mathbf{0}.$$

Матрица \mathbf{R} имеет K столбцов, чтобы сочетаться с β , J строк для общего числа в J ограничений и *полный ранг по строкам*, так что J должно быть меньше или равно K . Строки \mathbf{R} должны быть линейно независимы. Случай $J = K$ также должен быть исключен, несмотря на то что не нарушает рассмотренные условия. Если K коэффициентов удовлетворяют $J = K$ ограничениям, то \mathbf{R} квадратна и невырождена и $\beta = \mathbf{R}^{-1}\mathbf{q}$. Задача оценивания или инференции не стоит. Ограничение $\mathbf{R}\beta = \mathbf{q}$ накладывает J ограничений на K свободных параметров. Таким образом, с учетом наложенных ограничений остается на самом деле только $K - J$ свободных параметров.

Нам потребуется обобщить методы на случай нелинейных ограничений. В следующем примере гипотеза принимает форму $H_0: \beta_j/\beta_k = \beta_l/\beta_m$. **Общая нелинейная гипотеза** включает набор из J возможно нелинейных ограничений

$$\mathbf{c}(\beta) = \mathbf{q}, \quad (5-9)$$

где $\mathbf{c}(\beta)$ — это набор из J нелинейных функций β . Линейная гипотеза является частным случаем. Аналогом нашего требования для линейного случая является то, что снова J строго меньше K и матрица производных

$$\mathbf{G}(\beta) = \partial \mathbf{c}(\beta) / \partial \beta' \quad (5-10)$$

имеет полный ранг по строкам. Это означает, что ограничения **функционально независимы**. В линейном случае $\mathbf{G}(\beta)$ — это матрица констант \mathbf{R} , которую мы видели ранее, и функциональная независимость эквивалентна линейной независимости. Мы подробно рассмотрим нелинейные ограничения в разделе 5.7. На данный момент ограничимся рассмотрением общей линейной гипотезы.

Гипотеза, подразумеваемая ограничениями, записывается как

$$\begin{aligned} H_0: \mathbf{R}\beta - \mathbf{q} &= \mathbf{0}, \\ H_1: \mathbf{R}\beta - \mathbf{q} &\neq \mathbf{0}. \end{aligned}$$

Мы рассмотрим два подхода к тестированию гипотезы, тесты Вальда и тесты, основанные на качестве модели. Гипотеза характеризует популяцию. Если гипотеза верна, выборочная статистика должна воспроизводить это описание. Продолжая предыдущий пример, гипотеза утверждает, что некий коэффициент в регрессионной модели равен нулю. Если гипотеза верна, оценка коэффициента должна быть близка к нулю по крайней мере в пределах выборочной вариации. Тесты выглядят следующим образом:

- Тест Вальда: гипотеза утверждает, что $\mathbf{R}\beta - \mathbf{q}$ равен $\mathbf{0}$. Оценка методом наименьших квадратов \mathbf{b} является несмещенной и состоятельной оценкой β . Если гипотеза верна, то **выборочное отклонение** $\mathbf{Rb} - \mathbf{q}$ должно быть близко к нулю. Для примера с единственным коэффициентом, если верна гипотеза о том, что β_k равен нулю, то b_k должен быть близок к нулю. Тест Вальда измеряет, насколько близок $\mathbf{Rb} - \mathbf{q}$ к нулю.
- Тесты на основе качества модели: мы получаем наилучшее качество подгонки модели — самый высокий R^2 , используя метод наименьших квадратов и не накладывая никаких ограничений. Мы доказали это в главе 3. Теперь мы покажем, что сумма квадратов остатков никогда не уменьшится при наложении ограничений. Точнее, она будет увеличиваться, за исключением маловероятного специального случая. Например, когда мы накладываем ограничение $\beta_k = 0$, исключая x_k из модели, то должны ожидать падения R^2 . Эмпирическим приложением для тестирования гипотез будет мера того, насколько сильно падает R^2 при наложении ограничений.

ВАЖНАЯ ПРЕДПОСЫЛКА. Для получения тестовых статистик в этом разделе мы предполагаем, что ошибки распределены нормально. Как мы видели в главе 4, с этим предположением мы сможем получить точные распределения тестовых статистик. В разделе 5.6 рассмотрим последствия избавления от этой предпосылки и разработаем альтернативный набор результатов, который позволит нам продолжить анализ без нее.

5.4. Тест Вальда

Тест Вальда является наиболее часто используемой процедурой. Он нередко называется тестом значимости. Основопологающим принципом его работы служит оценка регрессии без ограничений с последующей оценкой того, совместимы ли результаты с гипотезой в пределах выборочной вариации.

5.4.1. Тестирование гипотез о коэффициенте

Простейшим случаем является тестирование равенства одного коэффициента какому-либо значению. Рассмотрим еще раз наш пример с рынком произведений искусства из раздела 5.2. Нулевая гипотеза имеет вид

$$H_0: \beta_2 = \beta_2^0,$$

где β_2^0 — это гипотетическое значение коэффициента, в данном случае нуль. **Расстоянием Вальда** между оценкой коэффициента и гипотетическим значением называется линейное расстояние (разность между этими значениями), измеренное в единицах стандартных отклонений. Таким образом, в этом случае расстояние между b_k и β_k^0 будет равно

$$W_k = \frac{b_k - \beta_k^0}{\sqrt{\sigma^2 S^{kk}}}. \quad (5-11)$$

Как мы видели в (4-38), W_k (которая раньше называлась z_k) имеет стандартное нормальное распределение в предположении, что $E[b_k] = \beta_k^0$. Теперь отметьте, что если $E[b_k]$ не равен β_k^0 , то W_k все так же имеет нормальное распределение, но среднее не равно нулю. В частности, если $E[b_k]$ равно β_k^1 , которое отличается от β_k^0 , то

$$E\{W_k | E[b_k] = \beta_k^1\} = \frac{\beta_k^1 - \beta_k^0}{\sqrt{\sigma^2 S^{kk}}}. \quad (5-12)$$

(Например, если гипотеза состоит в том, что $\beta_k = \beta_k^0 = 0$ и β_k не равен нулю, то ожидаемое значение $W_k = b_k / \sqrt{\sigma^2 S^{kk}}$ будет равно $\beta_k^1 / \sqrt{\sigma^2 S^{kk}}$, что не равно нулю.) Для целей использования W_k в процессе тестирования гипотез наша интерпретация состоит в том, что если β_k не равен β_k^0 , то b_k будет близок к β_k^0 в единицах стандартных отклонений. Следовательно, логика теста будет состоять в принятии решения о ложности H_0 — необходимости ее отвержения, если W_k «велика».

Прежде чем мы определим, что значит «велика», отметим, что предложенное здесь расстояние Вальда не может быть использовано, поскольку σ^2 неизвестно. Его оценка была получена с помощью s^2 . Воспользуемся снова результатами главы 4 и оценим W_k , используя выборочную оценку σ^2 . Тогда получаем

$$t_k = \frac{b_k - \beta_k^0}{\sqrt{s^2 S^{kk}}}. \quad (5-13)$$

Предполагая, что β_k действительно равен β_k^0 , т.е. что нулевая гипотеза верна, t_k имеет t -распределение с $n - K$ степенями свободы. [См. (4-41).] Теперь

мы можем описать процедуру тестирования. В ней заранее определяется, с какой степенью уверенности мы хотели бы сделать вывод — обычно выбранное значение равно 95%. На основании (5-13) мы можем сказать, что

$$\text{Prob}\{-t_{(1-\alpha/2),[n-K]}^* < t_k < +t_{(1-\alpha/2),[n-K]}^*\},$$

где $t_{(1-\alpha/2),[n-K]}^*$ — это соответствующее значение из таблицы t -распределения. По построению получение выборочного значения t_k , попавшего за пределы этого диапазона, маловероятно. Наша тестовая процедура подтверждает, что это настолько маловероятно, что мы могли бы сделать вывод о том, что это невозможно, в случае верности гипотезы, так что гипотеза должна быть ошибочной.

Типичный тест предполагает гипотезу о том, что параметр равен нулю, другими словами, это тест на релевантность переменной в регрессии. При построении тестовой статистики мы приравниваем β_k^0 нулю в (5-13) для получения стандартного « t -отношения»:

$$t_k = \frac{b_k}{s_{bk}}.$$

Эта статистика указывается в результатах регрессии в нескольких из наших предыдущих примеров, таких как 4.10, где приводятся результаты регрессии для модели (5-2). Эта статистика обычно обозначается как **t -ratio** (**t -отношение**) для оценки b_k . Если $|b_k|/s_{bk} > t_{(1-\alpha/2),[n-K]}$, где $t_{(1-\alpha/2),[n-K]}$ является $100(1 - \alpha/2)$ %-м критическим значением для t -распределения с $(n - K)$ степенями свободы, то нулевая гипотеза о том, что коэффициент равен нулю, отвергается и коэффициент (на самом деле соответствующая переменная) называется статистически значимым. Значение 1,96, которое соответствует 95%-му уровню значимости в большой выборке, часто используется в качестве порогового значения, когда таблица критических значений недоступна. t -значение для теста, проверяющего гипотезу о том, что коэффициент равен нулю, является стандартной частью результата регрессионного анализа в большинстве компьютерных программ.

Полезен другой подход к рассмотрению процедуры тестирования. Также основываясь на (4-39) и (5-13), мы построили доверительный интервал для β_k как $b_k \pm t^* s_k$. Мы можем рассматривать этот интервал как набор допустимых значений β_k с уровнем значимости $100(1 - \alpha)\%$ для выбранного нами значения α , обычно 5%. Доверительный интервал предоставляет удобный инструмент для тестирования гипотезы о β_k , поскольку мы можем просто задаться вопросом, содержится ли гипотетическое значение β_k^0 в этом диапазоне допустимых значений.

Пример 5.1. Удорожание произведений искусства

Результаты оценки регрессии для модели (5-3) на основе выборки из 430 сделок с картинами Моне представлены в табл. 4.6 в примере 4.10. Оценка коэффициента при $\ln Area$ составляет 1,33372 с оценкой стандартной ошибки в 0,09072. Расстояние между оцененным коэффициентом и нулем равно $1,33372/0,09072 = 14,70$. Поскольку это

существенно больше критического 95%-го значения, равного 1,96, мы отвергаем гипотезу о том, что коэффициент β_2 равен нулю; оказывается, что покупатели картин Моне интересуются размером. Напротив, коэффициент при *AspectRatio* равен $-0,16537$ с оценкой стандартной ошибки $0,12753$, так что соответствующее t -значение для тестирования гипотезы $H_0: \beta_3 = 0$ всего лишь $-1,30$. Поскольку это сильно меньше 1,96, мы делаем вывод, что покупателей произведений искусства (картин Моне) не интересует соотношение сторон картины. В качестве финального замечания рассмотрим другую (столь же загадочную) гипотезу — являются ли аукционные цены неэластичными $H_0: \beta_2 \leq 1$ или эластичными $H_1: \beta_2 > 1$ по отношению к площади. Это так называемый **односторонний тест**. Используя подход Неймана–Пирсона к формулировке теста, мы отвергнем нулевую гипотезу, если оценка коэффициента в достаточной степени превысит 1,0 (и не отвергнем, если она будет меньше или равна 1,0). Для поддержания размера теста на уровне 0,05 мы разместим всю площадь критической области (области отвержения) справа от 1,0; критическое значение из таблицы составляет 1,645. Тестовая статистика — это $(1,33372 - 1,0)/0,09072 = 3,679 > 1,645$, т. е. мы отвергнем и эту нулевую гипотезу.

Пример 5.2. Уравнение зарплаты

Таблица F5.1 из приложения содержит 753 наблюдения, использовавшихся в исследовании предложения труда замужних женщин (Mroz (1987)). Мы используем эти данные в качестве примера. Из 753 индивидов выборки 428 формально были участниками рынка труда. Для них мы оценим уравнение зарплаты в полулогарифмической форме, предложенной в примере 2.2:

$$\ln \text{earnings} = \beta_1 + \beta_2 \text{age} + \beta_3 \text{age}^2 + \beta_4 \text{education} + \beta_5 \text{kids} + \varepsilon,$$

где *earnings* — это *hourly wage* (почасовая зарплата), умноженная на *hours worked* (число отработанных часов), *education* (образование) измеряется в годах полученного образования и *kids* — это бинарная переменная, равная единице при наличии в домохозяйстве детей до 18 лет. (См. подробное описание данных в приложении F.) Результаты оценки регрессии представлены в табл. 5.1. Так, для оценки 5 параметров используются 428 наблюдений, t -статистика имеет $(428 - 5) = 423$ степени свободы. Для 95%-го уровня значимости при таком большом числе степеней свободы можно использовать значение стандартного нормального распределения, равное 1,96. По этому показателю все переменные статистически значимы и знаки соответствуют ожиданиям. Интерес представляет вопрос о том, оказывает ли влияние число детей (*kids*) на заработную плату, на отработанные часы или на оба показателя. Интерпретация оценки при переменной, характеризующей уровень образования, состоит в том, что дополнительный год обучения дает с 6,7%-й прирост зарплаты. Квадрат возраста предполагает, что для заданного уровня образования и размера

семьи зарплата растет до максимума у $-b_2/(2b_3)$, что составляет примерно 43 года, после чего она начинает падать. Некоторые замечания: (1) оценка регрессии только для выборки из индивидов, которые работали положительное число часов, не является безвредным механизмом построения выборки. Поскольку индивиды делали выбор (работать или не работать), вполне вероятно (этого стоит ожидать), что потенциальный уровень зарплаты (вместе с некоторыми другими аспектами, которые мы рассмотрим в главе 19) играл важную роль.

(2) Уравнение зарплаты представляет собой смесь уравнения предложения труда — часов, отработанных индивидом, и уравнения спроса на труд, заработная плата — это, надо полагать, принятое предложение. Поэтому неясно, какова точная природа этого уравнения. Скорее всего, это смесь уравнений сложной системы структурных уравнений. (См. обсуждение в примере 10.1.)

Таблица 5.1. Результаты оценки регрессии для уравнения зарплаты

| Сумма квадратов остатков: | 599,4582 | | | |
|--|-------------|--------------------|-----------------|----------|
| Стандартная ошибка регрессии: | 1,19044 | | | |
| R^2 на основе 428 наблюдений | 0,040995 | | | |
| Переменная | Коэффициент | Стандартная ошибка | t - отношение | |
| Константа | 3,24009 | 1,7674 | 1,833 | |
| Age | 0,20056 | 0,08386 | 2,392 | |
| Age ² | -0,0023147 | 0,00098688 | -2,345 | |
| Education | 0,067472 | 0,025248 | 2,672 | |
| Kids | -0,35119 | 0,14753 | -2,380 | |
| Оценка ковариационной матрицы для b ($e - n = \times 10^{-n}$) | | | | |
| Константа | Age | Age ² | Education | Kids |
| 3,12381 | | | | |
| -0,14409 | 0,0070325 | | | |
| 0,0016617 | -8,23237e-5 | 9,73928e-7 | | |
| -0,0092609 | 5,08549e-5 | -4,96761e-7 | 0,00063729 | |
| 0,026749 | -0,0026412 | 3,84102e-5 | -5,46193e-5 | 0,021766 |

5.4.2. F-статистика и отклонение метода наименьших квадратов

Теперь рассмотрим процедуру тестирования набора из J линейных ограничений, указанных в нулевой гипотезе,

$$H_0 : R\beta - q = 0,$$

против альтернативной гипотезы

$$H_1 : R\beta - q \neq 0.$$

Имея оценку \mathbf{b} , полученную методом наименьших квадратов, наш интерес сосредоточен на **векторе отклонений** $R\mathbf{b} - \mathbf{q} = \mathbf{m}$. Маловероятно, что вектор \mathbf{m} будет в точности равен $\mathbf{0}$. Статистический вопрос состоит в том, может

ли отклонение \mathbf{m} от $\mathbf{0}$ быть отнесено к выборочной ошибке или оно является значимым. Поскольку \mathbf{b} распределен нормально [см. (4-18)] и \mathbf{m} — это линейная функция от \mathbf{b} , то \mathbf{m} также распределен нормально. Если нулевая гипотеза верна, то $\mathbf{R}\beta - \mathbf{q} = \mathbf{0}$ и \mathbf{m} имеет вектор средних

$$E[\mathbf{m} | \mathbf{X}] = \mathbf{R}E[\mathbf{b} | \mathbf{X}] - \mathbf{q} = \mathbf{R}\beta - \mathbf{q} = \mathbf{0}$$

и ковариационную матрицу

$$\text{Var}[\mathbf{m} | \mathbf{X}] = \text{Var}[\mathbf{Rb} - \mathbf{q} | \mathbf{X}] = \mathbf{R}\{\text{Var}[\mathbf{b} | \mathbf{X}]\}\mathbf{R}' = \sigma^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'.$$

Мы можем построить тест для гипотезы H_0 на основе критерия Вальда. Условно относительно \mathbf{X} мы получаем:

$$\begin{aligned} W &= \mathbf{m}'\{\text{Var}[\mathbf{m} | \mathbf{X}]\}^{-1}\mathbf{m} \\ &= (\mathbf{Rb} - \mathbf{q})'[\sigma^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{Rb} - \mathbf{q}) \\ &= \frac{(\mathbf{Rb} - \mathbf{q})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{Rb} - \mathbf{q})}{\sigma^2} \\ &\sim \chi^2[J]. \end{aligned} \quad (5-14)$$

Если гипотеза верна, то статистика W имеет хи-квадрат-распределение с J степенями свободы¹. Как подсказывает интуиция, чем больше значение \mathbf{m} , т. е. выше степень несоответствия оценки ограничениям, тем больше статистика хи-квадрат. Следовательно, большое значение хи-квадрат будет аргументом против нулевой гипотезы.

Статистика хи-квадрат в (5-14) не может быть использована, поскольку параметр σ^2 неизвестен. Используя s^2 вместо σ^2 и поделив результат на J , мы получаем F -статистику с J и $n - K$ степенями свободы, которую можно использовать. Производя замену в (5-14), поделив на J и умножив и поделив на $n - K$, мы получаем

$$\begin{aligned} F &= \frac{W \sigma^2}{J s^2} \\ &= \left(\frac{(\mathbf{Rb} - \mathbf{q})'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{Rb} - \mathbf{q})}{\sigma^2} \right) \left(\frac{1}{J} \right) \left(\frac{\sigma^2}{s^2} \right) \left(\frac{(n - K)}{(n - K)} \right) \\ &= \frac{(\mathbf{Rb} - \mathbf{q})'[\sigma^2\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{Rb} - \mathbf{q})/J}{[(n - K)s^2/\sigma^2]/(n - K)}. \end{aligned} \quad (5-15)$$

Если $\mathbf{R}\beta = \mathbf{q}$, т. е. если нулевая гипотеза верна, то $\mathbf{Rb} - \mathbf{q} = \mathbf{Rb} - \mathbf{R}\beta = \mathbf{R}(\mathbf{b} - \beta) = \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}$. [См. (4-4).] Пусть $\mathbf{C} = [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']$. Так как

$$\frac{\mathbf{R}(\mathbf{b} - \beta)}{\sigma} = \mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\frac{\boldsymbol{\varepsilon}}{\sigma}\right) = \mathbf{D}\left(\frac{\boldsymbol{\varepsilon}}{\sigma}\right),$$

числитель F равен $[(\boldsymbol{\varepsilon}/\sigma)'\mathbf{T}(\boldsymbol{\varepsilon}/\sigma)]/J$, где $\mathbf{T} = \mathbf{D}'\mathbf{C}^{-1}\mathbf{D}$. Числитель — это W/J из (5-14). Он распределен как произведение $1/J$ и хи-квадрат $[J]$, как показано выше. В (4-16) мы получили, что $s^2 = \mathbf{e}'\mathbf{e}/(n - K) = \boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}/(n - K)$,

¹Эти вычисления являются приложением «квадратичной формы полного ранга» из раздела В.11.6. Отметьте, что, несмотря на то что хи-квадрат-распределение рассматривается условно относительно \mathbf{X} , оно также свободно от \mathbf{X} .

где \mathbf{M} — это идемпотентная матрица. Следовательно, знаменатель F равен $[(\boldsymbol{\varepsilon}/\sigma)' \mathbf{M}(\boldsymbol{\varepsilon}/\sigma)] / (n - K)$. Эта статистика распределена как произведение $1/(n - K)$ и хи-квадрат $[n - K]$. Следовательно, F -статистика — это отношение двух хи-квадрат-переменных, каждая из которых поделена на число своих степеней свободы. Поскольку $\mathbf{M}(\boldsymbol{\varepsilon}/\sigma)$ и $\mathbf{T}(\boldsymbol{\varepsilon}/\sigma)$ обе распределены нормально и их ковариация $\mathbf{T}\mathbf{M}$ равна $\mathbf{0}$, векторы квадратичной формы независимы. Числитель и знаменатель F являются функциями независимых случайных векторов и, следовательно, независимы. Это завершает доказательство того, что полученное распределение является F -распределением. [См. (B-35).] Сокращение σ^2 в (5-15) дает F -статистику для тестирования линейных гипотез:

$$F[J, n - K | \mathbf{X}] = \frac{(\mathbf{R}\mathbf{b} - \mathbf{q})' \{ \mathbf{R} [s^2(\mathbf{X}'\mathbf{X})^{-1}] \mathbf{R}' \}^{-1} (\mathbf{R}\mathbf{b} - \mathbf{q})}{J}. \quad (5-16)$$

Для тестирования одного линейного ограничения вида

$$H_0 : r_1\beta_1 + r_2\beta_2 + \dots + r_K\beta_K = \mathbf{r}'\boldsymbol{\beta} = q$$

(обычно некоторые r равны нулям) F -статистика равна

$$F[1, n - K] = \frac{(\sum_j r_j b_j - q)^2}{\sum_j \sum_k r_j r_k \text{Est. Cov}[b_j, b_k]}.$$

Если гипотеза состоит в том, что j -й коэффициент равен конкретному числу, то \mathbf{R} имеет единственную строку с 1 в j -м столбце и нулями в остальных, $\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'$ — это j -й диагональный элемент обратной матрицы и $\mathbf{R}\mathbf{b} - \mathbf{q}$ — это $(b_j - q)$. F -статистика в таком случае равна

$$F[1, n - K] = \frac{(b_j - q)^2}{\text{Est. Var}[b_j]}.$$

Рассмотрим альтернативный подход. Выборочная оценка $\mathbf{r}'\boldsymbol{\beta}$ равна

$$r_1 b_1 + r_2 b_2 + \dots + r_K b_K = \mathbf{r}'\mathbf{b} = \hat{q}.$$

Если \hat{q} существенно отличается от q , то мы заключаем, что данные выборки несовместимы с гипотезой. Естественно построить тест на основе

$$t = \frac{\hat{q} - q}{\text{se}(\hat{q})}. \quad (5-17)$$

Нам нужно оценить стандартную ошибку \hat{q} . Поскольку \hat{q} является линейной функцией от \mathbf{b} и у нас есть оценка ковариационной матрицы \mathbf{b} , $s^2(\mathbf{X}'\mathbf{X})^{-1}$, мы можем оценить дисперсию \hat{q} с помощью

$$\text{Est. Var}[\hat{q} | \mathbf{X}] = \mathbf{r}' [s^2(\mathbf{X}'\mathbf{X})^{-1}] \mathbf{r}.$$

Знаменатель t — это квадратный корень из этого числа. Другими словами, t — это расстояние между гипотетической функцией истинных коэффициентов и той же функцией наших оценок этих коэффициентов, выраженное в

единицах стандартной ошибки. Если гипотеза верна, то наши оценки должны отражать этот факт, по крайней мере, в пределах выборочной вариации. Таким образом, если абсолютное значение предыдущего t -значения больше, чем соответствующее критическое значение, гипотеза ставится под вопрос.

Между статистиками (5-16) и (5-17) существует полезное соотношение. Мы можем записать квадрат t -статистики как

$$t^2 = \frac{(\hat{q} - q)^2}{\text{Var}(\hat{q} - q | \mathbf{X})} = \frac{(\mathbf{r}'\mathbf{b} - q) \{ \mathbf{r}' [s^2 (\mathbf{X}'\mathbf{X})^{-1}] \mathbf{r} \}^{-1} (\mathbf{r}'\mathbf{b} - q)}{1}$$

Следовательно, для тестирования одного ограничения модуль t -статистики — это квадратный корень из F -статистики, которая может быть использована для тестирования этой гипотезы.

Пример 5.3. Уравнение инвестиций с ограничениями

В подразделе 5.2.2 предложена следующая теория поведения инвесторов: они заботятся только о реальной процентной ставке. Если бы инвесторы интересовались только реальной ставкой процента, одинаковые изменения в процентной ставке и темпе инфляции не оказывали бы независимого влияния на инвестиции. Нулевая гипотеза имеет вид

$$H_0: \beta_2 + \beta_3 = 0.$$

Оценки параметров уравнений (5-4) и (5-6), полученных с использованием квартальных данных за период между I кварталом 1950 г. и IV кварталом 2000 г. о реальных инвестициях, реальном ВВП, процентной ставке (90-дневной ставке процента по казначейским векселям) и инфляции, измеренной как разность логарифма ИПЦ, приведенные в таблице из приложения F5.2, представлены в табл. 5.2. (При вычислении разности логарифма ИПЦ одно наблюдение потеряно.) Для получения соответствующей тестовой статистики нам требуется стандартная ошибка $\hat{q} = b_2 + b_3$, которая составляет

$$\text{se}(\hat{q}) = [0,00319^2 + 0,00234^2 + 2(-3,718 \times 10^{-6})]^{1/2} = 0,002866.$$

t -статистика для теста, таким образом, равна

$$t = \frac{-0,00860 + 0,00331}{0,002866} = -1,845.$$

Таблица 5.2. Оценка уравнений инвестиций (в скобках — оценки стандартных ошибок)

| | β_1 | β_2 | β_3 | β_4 | β_5 |
|--------------|-----------|---|-----------|-----------|-----------|
| Model | -9,135 | -0,00860 | 0,00331 | 1,930 | -0,00566 |
| (5-4) | (1,366) | (0,00319) | (0,00234) | (0,183) | (0,00149) |
| | | $s = 0,08618, R^2 = 0,979753, e'e = 1,47052,$ | | | |
| | | $\text{Est. Cov}[b_2, b_3] = -3,718e-6$ | | | |
| Model | -7,907 | -0,00443 | 0,00443 | 1,764 | -0,00440 |
| (5-6) | (1,201) | (0,00227) | (0,00227) | (0,161) | (0,00133) |
| | | $s = 0,8670, R^2 = 0,979405, e'e = 1,49578$ | | | |

Используя 95%-е критическое значение для $t [203-5] = 1,96$ (значение для стандартного нормального распределения), мы заключаем, что сумма двух коэффициентов незначимо отличается от нуля, так что гипотеза не должна быть отвергнута.

Обычно есть более чем один способ для того, чтобы сформулировать регрессионную модель. Один из удобных способов параметризации ограничений состоит в такой его формулировке, при которой для тестирования гипотезы стандартные тестовые статистики, получаемые из регрессии, могли бы быть использованы без дальнейших подсчетов. В предыдущем примере мы могли записать регрессионную модель в том виде, в котором она специфицирована в (5-5). Тогда эквивалентный способ тестирования H_0 состоял бы в оценке инвестиционного уравнения со ставкой процента и инфляцией в качестве регрессоров и тестировании нашей теории путем тестирования гипотезы о том, что β_3 равен нулю, используя стандартную t -статистику, которая получается элементарным вычислением. Когда регрессия оценивается таким образом, $b_3 = -0,00529$ и оценка стандартной ошибки равна $0,00287$, приводя к t -статистике, равной $-1,844(!)$. (Упражнение: предположим, что номинальная процентная ставка — а не темп инфляции — была включена в качестве дополнительного регрессора. Как вы считаете, чему была бы равна оценка коэффициента и стандартной ошибки?)

Рассмотрим, наконец, тест для проверки совместной гипотезы:

$$\begin{aligned} \beta_2 + \beta_3 &= 0 && \text{(инвесторы рассматривают реальную процентную ставку),} \\ \beta_4 &= 1 && \text{(предельная склонность к инвестированию равна 1),} \\ \beta_5 &= 0 && \text{(временной тренд отсутствует).} \end{aligned}$$

Тогда

$$\mathbf{R} = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{q} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \text{и} \quad \mathbf{Rb} - \mathbf{q} = \begin{bmatrix} -0.0053 \\ 0.9302 \\ -0.0057 \end{bmatrix}.$$

Подставляя эти значения в F , получаем $F = 109,4$. Пятипроцентное критическое значение для $F[3, 198]$ равняется $2,65$. Таким образом, мы заключаем, что эти данные несовместимы с гипотезой. Результат не указывает, какое из ограничений оказывает наиболее значимое влияние на отвержение гипотезы. Если три ограничения тестируются по одному, t -статистики в (5-17) равны $-1,844$, $5,076$ и $-3,803$. Следовательно, на основе отдельных тестовых статистик стоит ожидать отвержения второй и третьей гипотез.

5.5. Тестирование ограничений с использованием показателей качества регрессии

Другой подход к тестированию основан на использовании характеристик качества регрессии. Вспомним, что вектор оценок \mathbf{b} выбирался, минимизируя сумму квадратов отклонений, $\mathbf{e}'\mathbf{e}$. Поскольку R^2 равен $1 - \mathbf{e}'\mathbf{e}/\mathbf{y}'\mathbf{M}^0\mathbf{y}$

и $y'M^0y$ постоянна и не включает \mathbf{b} , из этого следует, что \mathbf{b} выбирается, максимизируя R^2 . Возникает вопрос, приведет ли выбор других значений для коэффициентов наклона к значительной потере в качестве описания данных. Например, в уравнении инвестиций (5-4) интерес представляет вопрос о том, приводит ли к существенному ухудшению качества регрессии предположение о выполнении гипотезы (что инвесторы интересуются только реальной процентной ставкой) по сравнению с неограниченной регрессией. Для получения тестовой статистики мы сначала рассматриваем получение оценки методом наименьших квадратов при условии наличия ограничений, затем построим тестовую статистику на основе сравнения R^2 двух регрессий.

5.5.1. Оценка наименьших квадратов с ограничениями

Предположим, что мы явным образом наложили ограничения общей линейной гипотезы на регрессию. Оценка наименьших квадратов с ограничениями может быть получена как решение задачи минимизации:

$$\text{Minimize}_{\mathbf{b}_0} S(\mathbf{b}_0) = (\mathbf{y} - \mathbf{X}\mathbf{b}_0)'(\mathbf{y} - \mathbf{X}\mathbf{b}_0) \quad \text{при ограничении } \mathbf{R}\mathbf{b}_0 = \mathbf{q}. \quad (5-18)$$

Функция Лагранжа для этой задачи может быть записана как

$$L^*(\mathbf{b}_0, \boldsymbol{\lambda}) = (\mathbf{y} - \mathbf{X}\mathbf{b}_0)'(\mathbf{y} - \mathbf{X}\mathbf{b}_0) + 2\boldsymbol{\lambda}'(\mathbf{R}\mathbf{b}_0 - \mathbf{q})^2. \quad (5-19)$$

Решения \mathbf{b}_* и $\boldsymbol{\lambda}_*$ будут удовлетворять необходимым условиям

$$\frac{\partial L^*}{\partial \mathbf{b}_*} = -2\mathbf{X}'(\mathbf{y} - \mathbf{X}\mathbf{b}_*) + 2\mathbf{R}'\boldsymbol{\lambda}_* = \mathbf{0}, \quad (5-20)$$

$$\frac{\partial L^*}{\partial \boldsymbol{\lambda}_*} = 2(\mathbf{R}\mathbf{b}_* - \mathbf{q}) = \mathbf{0}.$$

Деление на 2 и раскрытие скобок приводят к уравнению для блочной матрицы:

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{R}' \\ \mathbf{R} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{b}_* \\ \boldsymbol{\lambda}_* \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{q} \end{bmatrix} \quad (5-21)$$

или

$$\mathbf{A}\mathbf{d}_* = \mathbf{v}.$$

Предполагая, что блочная матрица в скобках невырождена, оценка наименьших квадратов с ограничениями — это верхняя часть решения

$$\mathbf{d}_* = \mathbf{A}^{-1}\mathbf{v}. \quad (5-22)$$

Если помимо этого $\mathbf{X}'\mathbf{X}$ невырождена, то явное решение для \mathbf{b}_* и $\boldsymbol{\lambda}_*$ может быть получено с использованием формулы для блочной обратной матрицы (A-74)³:

$$\begin{aligned} \mathbf{b}_* &= \mathbf{b} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b} - \mathbf{q}) = \\ &= \mathbf{b} - \mathbf{C}\mathbf{m}, \end{aligned}$$

²Так как λ не ограничена, мы можем сформулировать ограничения в терминах 2λ . Удобство масштабирования можно увидеть в (5-20).

³Общее решение, приведенное для \mathbf{d}_* , может быть использовано, даже если $\mathbf{X}'\mathbf{X}$ вырождена. Например, предположим, что $\mathbf{X}'\mathbf{X}$ имеет размерность 4×4 и ранг 3. Тогда $\mathbf{X}'\mathbf{X}$ вырождена. Но если на β наложено параметрическое ограничение, то матрица в скобках размерности 5×5 может иметь ранг 5. Эта формулировка и набор связанных результатов приведены в работе Грина и Сикса (Greene, Seaks (1991)).

и (5-23)

$$\lambda_* = [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b} - \mathbf{q}).$$

Грин и Сикс (Greene, Seaks (1991)) показывают, что ковариационная матрица \mathbf{b}_* является произведением σ^2 и верхнего левого блока \mathbf{A}^{-1} . Опять же в обычном случае, в котором $\mathbf{X}'\mathbf{X}$ невырождена, может быть получена явная формула:

$$\text{Var}[\mathbf{b}_* | \mathbf{X}] = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} - \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}. \quad (5-24)$$

Таким образом,

$$\text{Var}[\mathbf{b}_* | \mathbf{X}] = \text{Var}[\mathbf{b} | \mathbf{X}] - \text{неотрицательно определенная матрица.}$$

Одна из интерпретаций такого сокращения в дисперсии заключается в том, сколько информации содержится в ограничениях.

Следует отметить, что явное решение для λ_* включает вектор отклонений $\mathbf{R}\mathbf{b} - \mathbf{q}$. Если оценка методом наименьших квадратов без ограничений удовлетворяет ограничению, множители Лагранжа будут равны нулю и \mathbf{b}_* будет равен \mathbf{b} . Конечно, это маловероятно. Решение \mathbf{b}_* с ограничениями равно решению \mathbf{b} без ограничений за вычетом слагаемого, которое отражает невозможность решения без ограничений удовлетворить ограничениям.

5.5.2. Потеря в качестве подгонки оценки наименьших квадратов с ограничениями

Для того чтобы получить тест, основанный на оценке наименьших квадратов с ограничениями, рассмотрим сначала один коэффициент, а затем обратимся к общему случаю с J линейными ограничениями. Рассмотрим изменение в качестве множественной регрессии при добавлении переменной z в модель, уже содержащую $K - 1$ переменных, \mathbf{x} . Как показано в разделе 3.5 (теорема 3.6) (3-29), влияние на качество модели будет выражаться как

$$R_{\mathbf{X}z}^2 = R_{\mathbf{X}}^2 + (1 - R_{\mathbf{X}}^2)r_{yz}^{*2}, \quad (5-25)$$

где $R_{\mathbf{X}z}^2$ — новый R^2 после включения z , $R_{\mathbf{X}}^2$ — исходный R^2 и r_{yz}^{*2} — частная корреляция между y и z с учетом влияния \mathbf{x} . Таким образом, как известно, качество улучшается (или, по крайней мере, не ухудшается). При выводе формулы для коэффициента частной корреляции между y и z в (3-22) мы получили удобный результат:

$$r_{yz}^{*2} = \frac{t_z^2}{t_z^2 + (n - K)}, \quad (5-26)$$

где t_z^2 — это квадрат t -статистики для тестирования гипотезы о том, что коэффициент при z равен нулю в множественной регрессии y на \mathbf{X} и z . Если мы решим (5-25) относительно r_{yz}^{*2} и (5-26) относительно t_z^2 и затем подставим первое решение во второе, получим результат

$$t_z^2 = \frac{(R_{\mathbf{X}z}^2 - R_{\mathbf{X}}^2)/1}{(1 - R_{\mathbf{X}z}^2)/(n - K)}. \quad (5-27)$$

Мы видели в конце подраздела 5.4.2, что для единственного ограничения, такого как $\beta_z = 0$,

$$F[1, n - K] = t^2[n - K],$$

что дает искомый результат, т.е. из (5-27) видно, что квадрат t -статистики (F -статистика) получается с использованием изменения в R^2 . Интерпретируя представленные выше преобразования как следствие *исключения z из регрессии*, мы видим, что был доказан результат для случая тестирования гипотезы о равенстве нулю отдельного коэффициента наклона. Но представленный результат имеет общий характер. Тестовая статистика для одного линейного ограничения равна квадрату t -статистики в (5-17). По построению можно отметить, что для одного ограничения F является мерой потери качества модели, которая происходит в результате наложения ограничения. Для получения этого результата перейдем к общему случаю J линейных ограничений, который будет включать одно ограничение в качестве частного случая.

Качество оценки коэффициентов методом наименьших квадратов с ограничениями не может быть лучше, чем в случае решения без ограничений. Пусть e_* равен $y - Xb_*$. Тогда, используя знакомый подход, имеем

$$e_* = y - Xb - X(b_* - b) = e - X(b_* - b).$$

Новая сумма квадратичных отклонений — это

$$e_*'e_* = e'e + (b_* - b)'X'X(b_* - b) \geq e'e.$$

(Средний член в выражении включает $X'e$, равный нулю.) Потеря в качестве тогда равна

$$e_*'e_* - e'e = (Rb - q)'[R(X'X)^{-1}R']^{-1}(Rb - q). \quad (5-28)$$

Это выражение возникает в числителе F -статистики в (5-7). Подставляя оставшиеся части выражения, мы получаем

$$F[J, n - K] = \frac{(e_*'e_* - e'e)/J}{e'e/(n - K)}. \quad (5-29)$$

Наконец, поделив числитель и знаменатель F на $\sum_i (y_i - \bar{y})^2$, получаем общий результат:

$$F[J, n - K] = \frac{(R^2 - R_*^2)/J}{(1 - R^2)/(n - K)}. \quad (5-30)$$

Эта форма имеет некоторую интуитивную привлекательность, состоящую в том, что характеристика различия качества двух моделей непосредственно включена в тестовую статистику. В качестве примера этого подхода рассмотрим совместный тест для проверки гипотезы о том, что все коэффициенты наклона в модели равны нулю, т.е. общее F -отношение, которое будет обсуждаться в подразделе 5.5.3, где $R_*^2 = 0$.

Для наложения набора **исключающих ограничений**, таких как $\beta_k = 0$, для одного или большего числа коэффициентов очевидный подход состоит в исключении переменных из регрессии и построении теста на основе сумм

квадратов остатков для регрессий с ограничениями и без них. F -статистика для тестирования гипотезы о том, что в подмножестве, например β_2 , все коэффициенты равны нулю, строится с использованием $\mathbf{R} = (\mathbf{0} : \mathbf{I})$, $\mathbf{q} = \mathbf{0}$, и $J = K_2 =$ числу элементов в β_2 . Матрица $\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'$ — это $K_2 \times K_2$, нижний правый блок полной обратной матрицы. Используя предыдущие результаты для блочных обратных матриц и результаты раздела 3.3, имеем

$$\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' = (\mathbf{X}'_2\mathbf{M}_1\mathbf{X}_2)^{-1}$$

и

$$\mathbf{R}\mathbf{b} - \mathbf{q} = \mathbf{b}_2.$$

Подставляя эти выражения в (5-28) получаем потерю в качестве, которая происходит при исключении набора переменных из регрессии:

$$e'_*e_* - e'e = \mathbf{b}'_2\mathbf{X}'_2\mathbf{M}_1\mathbf{X}_2\mathbf{b}_2.$$

Процедура для подсчета соответствующей F -статистики состоит в простом сравнении сумм квадратичных отклонений от «короткой» и «длинной» регрессий, которые мы видели ранее.

Пример 5.4. Производственная функция

Данные в таблице из приложения F5.3 использовались в нескольких исследованиях производственных функций⁴. Регрессия логарифма выпуска (добавленной стоимости) на константу и логарифма труда и капитала, оцениваемая методом наименьших квадратов, дает оценки производственной функции Кобба–Дугласа, представленные в табл. 5.3. Мы протестируем несколько гипотез на основе этих результатов. Обобщение модели Кобба–Дугласа — это *транслогарифмическая модель*⁵, которая записывается как

$$\ln Y = \beta_1 + \beta_2 \ln L + \beta_3 \ln K + \beta_4 \left(\frac{1}{2} \ln^2 L\right) + \beta_5 \left(\frac{1}{2} \ln^2 K\right) + \beta_6 \ln L \ln K + \varepsilon.$$

Как станет ясно позднее, в главе 10, эта модель отличается от модели Кобба–Дугласа тем, что в нее не вводится предпосылка о единичной эластичности замещения. Модель Кобба–Дугласа получается из нее при наложении ограничения $\beta_4 = \beta_5 = \beta_6 = 0$. Результаты оценки двух регрессий представлены в табл. 5.3. F -статистика для проверки гипотезы о том, что модель имеет форму Кобба–Дугласа, равна

$$F[3, 21] = \frac{(0,85163 - 0,67993)/3}{0,67993/21} = 1,768.$$

⁴База данных состоит из наблюдений по штатам в металлургическом производстве и производстве готовых металлических изделий (SIC 33). Изначально данные были собраны Хильдебрандом и Лью (Hildebrand, Liu (1957)) и затем использованы рядом авторов, в частности Эйгнером, Ловеллом и Шмидтом (Aigner, Lovell, Schmidt (1977)). 28-е наблюдение, использованное в исходном исследовании, неполно; мы использовали только оставшиеся 27.

⁵Берндт и Кристенсен (Berndt, Christensen (1973)). См. обсуждение в примере 2.4 и подразделе 10.5.2.

Таблица 5.3. Оценка производственной функции

| | | Транслогарифмическая | | | Кобба–Дугласа | |
|------------------------------|-----------------------|----------------------------|--------------------|-----------------------|----------------------------|------------------|
| Сумма квадратов остатков | | 0,67993 | | | 0,85163 | |
| Стандартная ошибка регрессии | | 0,17994 | | | 0,18837 | |
| R-квадрат | | 0,95486 | | | 0,94346 | |
| Скорректированный R-квадрат | | 0,94411 | | | 0,93875 | |
| Число наблюдений | | 27 | | | 27 | |
| Пере- менная | Коэффициент | Стан- дартная ошибка | t-отно- шение | Коэффи- циент | Стан- дартная ошибка | t-отно- шение |
| Константа | 0,944196 | 2,911 | 0,324 | 1,171 | 0,3268 | 3,582 |
| $\ln L$ | 3,61364 | 1,548 | 2,334 | 0,6030 | 0,1260 | 4,787 |
| $\ln K$ | -1,89311 | 1,016 | -1,863 | 0,3757 | 0,0853 | 4,402 |
| $\frac{1}{2} \ln^2 L$ | -0,96405 | 0,7074 | -1,363 | | | |
| $\frac{1}{2} \ln^2 K$ | 0,08529 | 0,2926 | 0,291 | | | |
| $\ln L \ln K$ | 0,31239 | 0,4389 | 0,712 | | | |
| Ковариационная матрица для | | | | | | |
| | Константа | $\ln L$ | $\ln K$ | $\frac{1}{2} \ln^2 L$ | $\frac{1}{2} \ln^2 K$ | $\ln L \ln K$ |
| Константа | 8,472 (0,1068) | | | | | |
| $\ln L$ | -2,388 (-0,01984) | 2,397 (0,01586) | | | | |
| $\ln K$ | -0,3313 (0,001189) | -1,231 (-0,009) | 1,033 (0,00728) | | | |
| $\frac{1}{2} \ln^2 L$ | -0,08760 | -0,6658 | 0,5231 | 0,5004 | | |
| $\frac{1}{2} \ln^2 K$ | -0,2332 | 0,03477 | 0,02637 | 0,1467 | 0,08562 | |
| $\ln L \ln K$ | 0,3635 | 0,1831 | -0,2255 | -0,2880 | -0,1160 | 0,1927 |

Критическое значение из таблицы для F -распределения равно 3,07, так что нет оснований отвергать гипотезу о том, что производственная функция может быть оценена в виде функции Кобба–Дугласа.

Гипотеза о постоянной отдаче от масштаба часто тестируется в исследованиях, касающихся оценки вида производственной функции. Эта гипотеза эквивалентна ограничению, состоящему в том, что сумма двух коэффициентов производственной функции Кобба–Дугласа равна единице. Для описанных выше данных получаем

$$F[1, 24] = \frac{(0,6030 + 0,3757 - 1)^2}{0,01586 + 0,00728 - 2(0,00961)} = 0,1157,$$

что существенно меньше, чем 95%-е критическое значение, составляющее 4,26. Оснований отвергнуть гипотезу нет; данные согласуются с гипотезой о постоянной отдаче от масштаба. Эквивалентными ограничениями для транслогарифмической модели были бы $\beta_2 + \beta_3 = 1$ и $\beta_4 + \beta_5 + 2\beta_6 = 0$. F -статистика с 2 и 21 степенями свободы равна 1,8991, что меньше, чем критическое значение 3,47. Снова гипотеза не отвергается.

В большинстве случаев, встречающихся на практике, можно включить соответствующие гипотезе ограничения непосредственно в регрессию и

оценить модель с ограничениями⁶. Например, ограничение $\beta_2 = 1$ в модели Кобба–Дугласа можно записать так:

$$\ln Y = \beta_1 + 1,0 \ln L + \beta_3 \ln K + \varepsilon$$

или так:

$$\ln Y - \ln L = \beta_1 + \beta_3 \ln K + \varepsilon.$$

Таким образом, модель с ограничениями оценивается с помощью регрессии $\ln Y - \ln L$ на константу и $\ln K$. Отдельного внимания требует использование этой регрессии для получения F -статистики. Если F -статистика получается с использованием сумм квадратов остатков [см. (5-29)], то проблем не возникнет. Однако если используется выражение (5-30), то может потребоваться учесть наличие в регрессии с ограничениями другой зависимой переменной. В предыдущей регрессии зависимой переменной в регрессии без ограничений была $\ln Y$, тогда как в регрессии с ограничениями $\ln Y - \ln L$. R^2 из регрессии с ограничениями составляет всего 0,26979, что дает F -статистику, равную 285,96, тогда как верное значение составляет 9,935. Если же мы оценим R_*^2 , используя правильный знаменатель, его значение составит 0,92006, что даст верное значение F -статистики.

Отметим, что коэффициент при $\ln K$ в транслогарифмической модели отрицателен. Можно заключить, что оценка эластичности выпуска по капиталу имеет неверный знак. Однако такой вывод будет неверен; в транслогарифмической модели эластичность выпуска по капиталу составляет

$$\frac{\partial \ln Y}{\partial \ln K} = \beta_3 + \beta_5 \ln K + \beta_6 \ln L.$$

Если мы подставим оценки коэффициентов и средние значения для $\ln K$ и $\ln L$ (не логарифмы средних), равные 7,44592 и 5,7637 соответственно, то результат составит 0,5425, что примерно соответствует нашим ожиданиям и не сильно отличается от значения 0,3757, полученного для модели Кобба–Дугласа. Оценка стандартной ошибки для этой линейной комбинации МНК-оценок получается как квадратный корень из

$$\text{Est. Var}[b_3 + b_5 \overline{\ln K} + b_6 \overline{\ln L}] = \mathbf{w}'(\text{Est. Var}[\mathbf{b}])\mathbf{w},$$

где $\mathbf{w} = (0, 0, 1, 0, \overline{\ln K}, \overline{\ln L})'$ и \mathbf{b} представляет полный вектор коэффициентов размерности 6×1 . Это значение равно 0,1122, что достаточно близко к предыдущей оценке, равной 0,0853.

5.5.3. Тестирование значимости регрессии

Нередко интерес представляет вопрос о том, значимо ли регрессионное уравнение в целом. Тогда это совместный тест для проверки гипотезы о том, что все коэффициенты, за исключением константы, равны нулю. Если все

⁶Это не так, когда ограничения нелинейны. Мы рассмотрим этот случай в главе 7.

коэффициенты наклона равны нулю, то коэффициент детерминации R^2 также равен нулю, поэтому мы можем построить тест для проверки этой гипотезы на основе R^2 . Центральный результат, необходимый для проведения теста, приведен в (5-30). Это специальный случай, в котором $R_*^2 = 0$, так что F -статистика, которая обычно приводится вместе с результатами множественной регрессии, представима в виде

$$F[K - 1, n - K] = \frac{R^2/(K - 1)}{(1 - R^2)/(n - K)}.$$

Если гипотеза о том, что $\beta_2 = \mathbf{0}$ (часть β , не включающая константу), верна и ошибки распределены нормально, то эта статистика имеет F -распределение с $K - 1$ и $n - K$ степенями свободы. Большие значения F выступают в пользу отвержения гипотезы. Следует отметить, что большое значение F вызвано большим значением R^2 . Логика теста состоит в том, что F -статистика — это мера потери качества описания данных (а именно всего R^2), что происходит, когда мы накладываем ограничение, состоящее в равенстве нулю всех наклонов. Если F велика, гипотеза отвергается.

Пример 5.5. *F-тест для уравнений доходов*

F -статистика для тестирования гипотезы о том, что четыре коэффициента наклона в уравнении доходов из примера 5.2 равны нулю, равна

$$F[4, 423] = \frac{0,040995/(5 - 1)}{(1 - 0,040995)/(428 - 5)} = 4,521,$$

что сильно превышает 95%-е критическое значение, составляющее 2,39. Отсюда следует вывод, что данные не согласуются с гипотезой о том, что все коэффициенты наклона в уравнении доходов равны нулю. Возможно, этого результата следовало ожидать, учитывая большие значения t -статистик, представленные ранее. Но так происходит не всегда. Можно привести пример того, как коэффициенты оказываются статистически значимыми по одному, но незначимыми в совокупности. Этот случай может рассматриваться как патологический и обратный более распространенному примеру, в котором ни один из коэффициентов незначим, тогда как R^2 имеет высокую значимость. Проблема состоит в том, что взаимное влияние между переменными может скрыть их индивидуальный вклад в качество регрессии, тогда как их совместный вклад может оказаться значимым.

5.5.4. Вывод ограничений и замечание об использовании R^2

Как правило, ограничения, накладываемые линейной гипотезой, можно выразить при помощи алгебраического выражения. Для этого нужно разбить матрицу \mathbf{R} на две части, одну с J столбцами и другую с $K - J$ столбцами, так что первый набор линейно независим. (Существует много способов это осуществить; в данный момент подойдет любой.) Затем, используя вектор β , разбитый и упорядоченный соответствующим образом, можно запи-

сать

$$\mathbf{R}\boldsymbol{\beta} = \mathbf{R}_1\boldsymbol{\beta}_1 + \mathbf{R}_2\boldsymbol{\beta}_2 = \mathbf{q}.$$

Если J столбцов матрицы \mathbf{R}_1 независимы, то

$$\boldsymbol{\beta}_1 = \mathbf{R}_1^{-1}[\mathbf{q} - \mathbf{R}_2\boldsymbol{\beta}_2].$$

Это указывает на то, что модель с ограничениями можно оценить прямо, используя преобразованное уравнение, а не через неудобную оценку с ограничениями, приведенную в (5-23). Проиллюстрируем это простым примером. Рассмотрим наложение ограничения о постоянной отдаче от масштаба на производственную функцию с двумя факторами производства:

$$\ln y = \beta_1 + \beta_2 \ln x_1 + \beta_3 \ln x_2 + \varepsilon.$$

Гипотеза линейной однородности может быть записана как $\beta_2 + \beta_3 = 1$ или $\beta_3 = 1 - \beta_2$. Простая подстановка этого ограничения в модель дает

$$\ln y = \beta_1 + \beta_2 \ln x_1 + (1 - \beta_2) \ln x_2 + \varepsilon$$

или

$$\ln y = \ln x_2 + \beta_1 + \beta_2(\ln x_1 - \ln x_2) + \varepsilon.$$

Можно получить оценку с ограничениями методом наименьших квадратов, оценив линейную регрессию $(\ln y - \ln x_2)$ на константу и $(\ln x_1 - \ln x_2)$. Однако тестовая статистика для этой гипотезы не может быть протестирована с использованием знакомого результата из (5-30), поскольку знаменатели в двух R^2 'х различаются. Статистика (5-30) может даже оказаться отрицательной. Уместным подходом будет использование эквивалентной, но более подходящей статистики (5-29), основанной на сумме квадратов остатков. Общий результат для этого примера состоит в том, что при использовании (5-30) нужно обращать внимание на то, что зависимая переменная в двух регрессиях должна быть одна и та же.

5.6. Ошибки, не являющиеся нормально распределенными, и асимптотические тесты

Теперь исследуем соотношение между выборочными тестовыми статистиками и данными в \mathbf{X} . Во-первых, рассмотрим стандартную t -статистику (4-41) для тестирования гипотезы $H_0 : \beta_k = \beta_k^0$:

$$t|\mathbf{X} = \frac{b_k - \beta_k^0}{\sqrt{s^2(\mathbf{X}'\mathbf{X})_{kk}^{-1}}}.$$

Условно относительно \mathbf{X} , если $\beta_k = \beta_k^0$ (т.е. при выполнении H_0), $t|\mathbf{X}$ имеет t -распределение с $(n - K)$ степенями свободы. Интерес, однако, представляет безусловное распределение t . Как мы видели, вектор \mathbf{b} распределен нормально только условно относительно \mathbf{X} ; безусловное распределение может не быть нормальным, поскольку зависит от \mathbf{X} (через условную дисперсию). Таким же образом ввиду наличия \mathbf{X} знаменатель t -статистики не

равен квадратному корню из переменной хи-квадрат, поделенной на количество степеней ее свободы, опять же кроме рассмотрения условно относительно \mathbf{X} . Но поскольку распределения $(b_k - \beta_k) / \sqrt{s^2(\mathbf{X}'\mathbf{X})_{kk}^{-1}} | \mathbf{X}$ и $[(n - K)s_2 / \sigma^2] | \mathbf{X}$ все так же являются независимыми, $N[0, 1]$ и $\chi^2[n - K]$ соответственно и не содержат \mathbf{X} . Тогда получаем удивительный результат, состоящий в том, что независимо от распределения \mathbf{X} или даже от того, является ли \mathbf{X} случайным или неслучайным, безусловные распределения t — это все то же t , несмотря на то что безусловное распределение b_k может быть ненормальным. Этот любопытный результат следует из того, что $f(t | \mathbf{X})$ не является функцией \mathbf{X} . Подобные рассуждения могут быть использованы для доказательства утверждения о том, что обычная F -статистика, используемая для тестирования линейных ограничений и обсуждавшаяся в предыдущем разделе, может быть использована независимо от того, является \mathbf{X} случайным или нет. Это очень мощный результат. Из него следует, что если ошибки распределены нормально, то можно проводить тесты и строить доверительные интервалы для параметров, не меняя процедур, независимо от того, являются ли регрессоры стохастическими, детерминированными или комбинацией двух случаев.

Распределения этих статистик выводятся из предположения о нормальности ϵ , но они не зависят от \mathbf{X} . Однако без предположения о нормальности точное распределение этих статистик зависит от данных и параметров и не относится к F , t или хи-квадрат. По крайней мере, на первый взгляд может показаться, что для тестов требуются новые наборы критических значений или даже тестовых статистик. В этом разделе будут рассмотрены результаты, обобщающие знакомые процедуры. Эти асимптотические результаты указывают на то, что хотя обычные t - и F -статистики все так же применимы в более общем случае без предположения о нормальности, но они рассматриваются как приближения, качество которых улучшается по мере увеличения выборки. Используя результаты раздела D.3 (посвященного асимптотическим распределениям) и некоторые асимптотические результаты для метода наименьших квадратов, мы можем построить ряд необходимых для вывода процедур, основанных на знакомых вычислениях.

Предполагая, что данные хорошие, асимптотическое распределение оценки коэффициента методом наименьших квадратов \mathbf{b} записывается как

$$\mathbf{b} \stackrel{a}{\sim} N \left[\beta, \frac{\sigma^2}{n} \mathbf{Q}^{-1} \right], \text{ где } \mathbf{Q} = \text{plim} \left(\frac{\mathbf{X}'\mathbf{X}}{n} \right). \quad (5-31)$$

Интерпретация состоит в том, что в случае отсутствия нормальности ϵ с ростом размера выборки n нормальное распределение все лучше приближает истинное, хотя и неизвестное в настоящий момент распределение \mathbf{b} . С ростом n распределение $\sqrt{n}(\mathbf{b} - \beta)$ сходится к нормальному распределению, что позволяет получить предыдущее приближение для конечной выборки. Этот результат основан для центральной предельной теореме и не требует нормальности распределения ошибок. Второй результат, который потребуется в дальнейшем, касается σ^2 :

$$\text{plim } s^2 = \sigma^2, \quad \text{где } s^2 = \mathbf{e}'\mathbf{e}/(n - K).$$

Используя эти результаты, мы можем получить асимптотические приближения наших тестовых статистик, которые покажут, как поступить в случае, когда выборка конечна, а ошибки не являются нормальными.

Выборочная статистика для тестирования гипотезы о том, что один из коэффициентов β_k равен конкретному значению β_k^0 , равна

$$t_k = \frac{\sqrt{n}(b_k - \beta_k^0)}{\sqrt{s^2(\mathbf{X}'\mathbf{X}/n)_{kk}^{-1}}}.$$

(Следует обратить внимание, что два \sqrt{n} сокращаются, приводя к знакомому результату.) При выполнении нулевой гипотезы и в случае нормальности распределения ошибок t_k распределено в точности как t с $n - K$ степенями свободы. [См. теорему 4.6 и начало этого раздела.] Однако если ϵ не распределены нормально, точное распределение этой статистики неизвестно. Из предыдущих результатов можно заключить, что знаменатель t_k сходится к $\sqrt{\sigma^2 \mathbf{Q}_{kk}^{-1}}$. Следовательно, если t_k имеет предельное распределение, то оно совпадает со статистикой, имеющей это же выражение в знаменателе. (См. пункт 3 теоремы D.16.), т.е. асимптотическое распределение t_k совпадает с

$$\tau_k = \frac{\sqrt{n}(b_k - \beta_k^0)}{\sqrt{\sigma^2 \mathbf{Q}_{kk}^{-1}}}.$$

Но так как $\tau_k = (b_k - E[b_k]) / (\text{Asy. Var}[b_k])^{1/2}$ из асимптотического нормального распределения (в условиях нулевой гипотезы $\beta_k = \beta_k^0$), то τ_k имеет стандартное нормальное асимптотическое распределение и является асимптотическим распределением нашей t -статистики. Таким образом, в качестве асимптотического приближения мы будем использовать стандартное нормальное распределение для аппроксимации истинного распределения тестовой статистики t_k и критические значения стандартного нормального распределения для тестирования.

Результат, представленный в предыдущем параграфе, действителен только в больших выборках. Для средних выборок он только предлагает t -распределение в качестве разумного приближения. Соответствующие критические значения *сходятся* к критическим значениям стандартного нормального в общем случае *сверху*, но в этом нельзя быть уверенным. В интересах консерватизма, т.е. контролируя вероятность ошибки I рода, следует использовать критические значения для t -распределения даже в отсутствие нормальности. Рассмотрим, например, использование стандартного нормального критического значения, равного 1,96, для двустороннего теста при проверке гипотезы, основанной на 25 степенях свободы. Номинальный размер этого теста 0,05. Реальный размер теста, однако, это истинная неизвестная вероятность $|t_k| > 1,96$, что равно 0,0612, если $t[25]$ -распределение правильное, и некоторое другое значение, если ошибки не распределены нормально. В конечном счете стандартный t -тест может быть использован в больших выборках. Так что мы не можем сказать точно, каков истинный размер теста на основе t -распределения, если не сделаем некоторых ограничивающих предположений об ошибках ϵ , но обычно t -распределение используется в качестве надежного приближения.

Для тестирования набора из J линейных ограничений будет использоваться аналогичный подход, основанный на анализе F -статистики. На первом шаге мы покажем, что при нормально распределенных ошибках JF сходится с ростом выборки к переменной хи-квадрат. Затем продемонстрируем, что этот результат на самом деле не зависит от нормальности ошибок; он основан на центральной предельной теореме. Наконец, как и прежде, рассмотрим соответствующие критические значения для этой тестовой статистики, надежные только для случая больших выборок.

F -статистика для тестирования обоснованности J линейных ограничений, $\mathbf{R}\beta - \mathbf{q} = \mathbf{0}$, приводится в (5-6). При нормально распределенных ошибках и выполнении нулевой гипотезы точное распределение этой статистики совпадает с $F[J, n - K]$. Чтобы увидеть, как F ведет себя при более общей постановке задачи, следует разделить числитель и знаменатель в (5-16) на σ^2 и слегка преобразовать дробь, так что

$$F = \frac{(\mathbf{R}\mathbf{b} - \mathbf{q})' \{ \mathbf{R}[\sigma^2(\mathbf{X}'\mathbf{X})^{-1}] \mathbf{R}' \}^{-1} (\mathbf{R}\mathbf{b} - \mathbf{q})}{J(s^2/\sigma^2)}. \quad (5-32)$$

Поскольку $\text{plim } s^2 = \sigma^2$ и $\text{plim}(\mathbf{X}'\mathbf{X}/n) = \mathbf{Q}$, знаменатель F сходится к J и выражение в скобках в числителе ведет себя так же, как и $(\sigma^2/n)\mathbf{R}\mathbf{Q}^{-1}\mathbf{R}'$. (См. теорему D16.3.) Таким образом, независимо от того, что это за распределение, если F имеет предельное распределение, то оно совпадает с предельным распределением

$$\begin{aligned} W^* &= \frac{1}{J}(\mathbf{R}\mathbf{b} - \mathbf{q})' [\mathbf{R}(\sigma^2/n)\mathbf{Q}^{-1}\mathbf{R}']^{-1} (\mathbf{R}\mathbf{b} - \mathbf{q}) \\ &= \frac{1}{J}(\mathbf{R}\mathbf{b} - \mathbf{q})' \{ \text{Asy. Var}[\mathbf{R}\mathbf{b} - \mathbf{q}] \}^{-1} (\mathbf{R}\mathbf{b} - \mathbf{q}). \end{aligned}$$

Это выражение равно произведению $(1/J)$ и статистики Вальда, основанной на асимптотическом распределении. Асимптотическое распределение W^* будет совпадать с произведением $(1/J)$ и хи-квадрат распределения с J степенями свободы. Из этого следует, что при нормально распределенных ошибках JF сходится к хи-квадрат-распределению с J степенями свободы. Доказательство поучительно. [См. White (2001, с. 76).]

Теорема 5.1. *Предельное распределение статистики Вальда*

Если $\sqrt{n}(\mathbf{b} - \beta) \xrightarrow{d} N[\mathbf{0}, \sigma^2\mathbf{Q}^{-1}]$ и если $H_0: \mathbf{R}\beta - \mathbf{q} = \mathbf{0}$ верна, то

$$W = (\mathbf{R}\mathbf{b} - \mathbf{q})' \{ \mathbf{R}s^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}' \}^{-1} (\mathbf{R}\mathbf{b} - \mathbf{q}) = JF \xrightarrow{d} \chi^2[J].$$

Доказательство: Поскольку \mathbf{R} представляет собой матрицу постоянных и $\mathbf{R}\beta = \mathbf{q}$, то

$$\sqrt{n}\mathbf{R}(\mathbf{b} - \boldsymbol{\beta}) = \sqrt{n}(\mathbf{R}\mathbf{b} - \mathbf{q}) \xrightarrow{d} N[\mathbf{0}, \mathbf{R}(\sigma^2\mathbf{Q}^{-1})\mathbf{R}']. \quad (1)$$

Для удобства запишем это уравнение как

$$\mathbf{z} \rightarrow N[\mathbf{0}, \mathbf{P}]. \quad (2)$$

В разделе А.6.11 мы определяем обратный корень из положительно определенной матрицы \mathbf{P} как некоторую матрицу \mathbf{T} такую, что $\mathbf{T}^2 = \mathbf{P}^{-1}$, и обозначаем \mathbf{T} как $\mathbf{P}^{-1/2}$. Затем, следуя той же логике, что и в (1) и (2), получаем, что

$$\text{если } \mathbf{z} \xrightarrow{d} N[\mathbf{0}, \mathbf{P}], \quad \text{то } \mathbf{P}^{-1/2}\mathbf{z} \xrightarrow{d} N[\mathbf{0}, \mathbf{P}^{-1/2}\mathbf{P}\mathbf{P}^{-1/2}] = N[\mathbf{0}, \mathbf{I}]. \quad (3)$$

Воспользуемся теперь теоремой D.21 для предельного распределения функции случайной величины. Сумма квадратов некоррелированных (т.е. независимых) стандартных нормальных случайных величин распределена как хи-квадрат. Таким образом, предельное распределение

$$(\mathbf{P}^{-1/2}\mathbf{z})'(\mathbf{P}^{-1/2}\mathbf{z}) = \mathbf{z}'\mathbf{P}^{-1}\mathbf{z} \xrightarrow{d} \chi^2(J). \quad (4)$$

Переставляя отдельные части выражения, мы показали, что предельное распределение

$$n(\mathbf{R}\mathbf{b} - \mathbf{q})'[\mathbf{R}(\sigma^2\mathbf{Q}^{-1})\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b} - \mathbf{q}) \quad (5)$$

является хи-квадрат-распределением с J степенями свободы. Отметим, что этот результат похож на результаты из раздела В.11.6. Наконец, если

$$\text{plim } s^2 \left(\frac{1}{n} \mathbf{X}'\mathbf{X} \right)^{-1} = \sigma^2\mathbf{Q}^{-1}, \quad (6)$$

то статистика, получаемая заменой $\sigma^2\mathbf{Q}^{-1}$ на $s^2(\mathbf{X}'\mathbf{X}/n)^{-1}$ в выражении (5), имеет такое же предельное распределение. n сокращаются, и остается та же самая статистика Вальда, которую мы встречали ранее. Этот шаг завершает доказательство.

Соответствующие критические значения для F -теста ограничений $\mathbf{R}\boldsymbol{\beta} - \mathbf{q} = \mathbf{0}$ сходятся сверху к произведению $1/J$ и критических значений распределения хи-квадрат, основанных на статистике Вальда (см. таблицы распределения хи-квадрат). Например, для тестирования $J = 5$ ограничений критическое значение из таблицы распределения хи-квадрат (таблица из приложения G.4) для 95%-го уровня значимости равно 11,07. Критические значения из таблицы для F -распределения (таблица из приложения G.5) равны $3,33 = 16,65/5$ для $n - K = 10$, $2,60 = 13,00/5$ для $n - K = 12$, $2,31 = 11,55/5$ для $n - K = 50$, $2,31 = 11,55/5$ для $n - K = 100$ и $2,214 = 11,07/5$ для больших $n - K$. Таким образом, с нормально распределенными ошибками при росте n F -тест может проводиться путем сравнения JF с критическими значениями из таблицы для распределения хи-квадрат.

Ключевой для наших целей результат состоит в том, что распределение статистики Вальда строится из распределения \mathbf{b} , которое асимптотически нормально, даже если ошибки нормально не распределены. Из этого следует, что соответствующая асимптотическая тестовая статистика распределена как хи-квадрат $= JF$. Опять же это следствие основано на центральной предельной теореме, а не на нормальном распределении ошибок. Каким же будет правильный подход при наличии малой или средней выборки? Как было показано ранее, критические значения для F -распределения сходятся сверху к произведению $(1/J)$ и значений для упомянутого выше распределения хи-квадрат. Как и прежде, нельзя сказать, что это утверждение будет верным в любом случае и для каждого набора данных и параметров. Однако в общем случае, за исключением специальных наборов данных и параметров, можно ожидать его выполнения. Следствием этого является то, что в отсутствие дополнительной, более точной спецификации модели F -статистика с критическими значениями из таблицы для F -распределения остается консервативным инструментом, который становится более точным с ростом выборки.

Упражнение 7 в конце этой главы предлагает другой допустимый в больших выборках подход к тестированию — **тест множителей Лагранжа**. Вектор множителей Лагранжа в (5-23) имеет вид $[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b} - \mathbf{q})$, т.е. он пропорционален вектору отклонений. В принципе тест для проверки гипотезы о том, что λ_* равен нулю, должен быть эквивалентен тесту для проверки нулевой гипотезы. Поскольку первая матрица в произведении имеет полный ранг, выражение может быть равно нулю, только если отклонение равно нулю. Тест Вальда для проверки гипотезы о том, что $\lambda_* = 0$, действительно является обоснованным в данной ситуации. Асимптотическое распределение статистики Вальда будет иметь распределение хи-квадрат с J степенями свободы. (Процедура рассматривается в упражнении 7.) Для набора исключаяющих ограничений $\beta_2 = 0$ есть простой способ провести этот тест. Статистика хи-квадрат, в таком случае имеющая K_2 степеней свободы, может быть получена как nR^2 в регрессии e_* (остатки в короткой регрессии) на полный набор независимых наблюдений.

5.7. Тестирование нелинейных ограничений

Предшествующее обсуждение в существенной степени основывалось на линейности регрессионной модели. Когда анализируются нелинейные функции параметров и нелинейные регрессионные модели, большинство результатов, связанных с точными распределениями, перестают работать.

Общая постановка задачи состоит в тестировании гипотезы, включающей нелинейную функцию регрессионных коэффициентов:

$$H_0: c(\beta) = q.$$

Сначала следует рассмотреть случай одного ограничения. Более общий случай, в котором $c(\beta) = q$, является набором ограничений, представляет собой простое обобщение. Аналогом тестовой статистики, использовавшейся

ранее, будет

$$z = \frac{c(\hat{\beta}) - q}{\text{оценка стандартной ошибки}}, \quad (5-33)$$

или ее квадрат, которые ранее были распределены как $t[n - K]$ и $F[1, n - K]$ соответственно. Вычисление отклонения в числителе не представляет сложности. Получение оценки выборочной дисперсии $c(t\hat{\beta}) - q$, однако, включает дисперсию нелинейной функции $\hat{\beta}$.

Результаты, требуемые для этой оценки, представлены в подразделах 4.4.4, В.10.3 и D.3.1. Линейное приближение $c(\hat{\beta})$ рядом Тейлора около истинного вектора параметров β имеет вид

$$c(\hat{\beta}) \approx c(\beta) + \left(\frac{\partial c(\beta)}{\partial \beta} \right)' (\hat{\beta} - \beta). \quad (5-34)$$

Тут приходится опираться на состоятельность, а не на несмещенность, поскольку в общем случае ожидаемое значение нелинейной функции не равно функции от ожидаемого значения. Если $\text{plim } \hat{\beta} = \beta$, то использование $c(\hat{\beta})$ в качестве оценки $c(\beta)$ обоснованно. (Требуемый здесь результат — это теорема Слуцкого.) Предполагая, что использование этого приближения здесь уместно, дисперсия нелинейной функции приблизительно равна дисперсии правой части, которая в таком случае равна

$$\text{Var}[c(\hat{\beta})] \approx \left(\frac{\partial c(\beta)}{\partial \beta} \right)' \text{Var}[\hat{\beta}] \left(\frac{\partial c(\beta)}{\partial \beta} \right). \quad (5-35)$$

Производные в выражении для дисперсии являются функциями неизвестных параметров. Поскольку они оцениваются, для оценки производных используются выборочные аналоги. В качестве оценки дисперсии оценки можно использовать $s^2(\mathbf{X}'\mathbf{X})^{-1}$. Наконец, основываясь на теореме D.22 из подраздела D.3.1 и берем стандартное нормальное распределение вместо t -распределения для тестовой статистики. Используя $g(\hat{\beta})$ для оценки $g(\beta) = \partial c(\beta)/\partial \beta$, можно проверить гипотезу тем же способом, что и ранее.

Пример 5.6. Долгосрочная предельная склонность к потреблению

Функция потребления, имеющая различные — краткосрочную и долгосрочную — предельные склонности к потреблению, может быть записана в виде

$$\ln C_t = \alpha + \beta \ln Y_t + \gamma \ln C_{t-1} + \varepsilon_t,$$

что является примером модели с **распределенными лагами**. В этой модели краткосрочная предельная склонность к потреблению (МРС) (эластичность, поскольку переменные в логарифмах) равна β , а долгосрочная МРС равна $\delta = \beta/(1 - \gamma)$. Рассмотрим тестирование гипотезы о том, что $\delta = 1$.

Квартальные данные об агрегированном потреблении и располагаемом доходе в США за период с 1950 по 2000 г. представлены в таблице из

приложения F5.2. Оценки уравнения для этих данных представлены ниже:

$$\ln C_t = 0,003142 + 0,07495 \ln Y_t + 0,9246 \ln C_{t-1} + e_t,$$

$$(0,01055) \quad (0,02873) \quad (0,02859)$$

$$R^2 = 0,999712, \quad s = 0,00874.$$

Оценки стандартных ошибок приведены в скобках. Также требуется $\text{Est. Asy. Cov}[b, c] = -0,0008207$. Оценка долгосрочной MPC тогда равна $d = b/(1 - c) = 0,07495/(1 - 0,9246) = 0,99403$. Для получения оценки дисперсии d потребуется

$$g_b = \frac{\partial d}{\partial b} = \frac{1}{1 - c} = 13,2626, \quad g_c = \frac{\partial d}{\partial c} = \frac{b}{(1 - c)^2} = 13,1834.$$

Оценка асимптотической дисперсии d равна

$$\begin{aligned} \text{Est. Asy. Var}[d] &= g_b^2 \text{Est. Asy. Var}[b] + g_c^2 \text{Est. Asy. Var}[c] + \\ &+ 2g_b g_c \text{Est. Asy. Cov}[b, c] = 13,2626^2 \times 0,02873^2 + \\ &+ 13,1834^2 \times 0,02859^2 + 2(13,2626)(13,1834)(-0,0008207) = 0,0002585. \end{aligned}$$

Ее квадратный корень равен 0,016078. Для тестирования гипотезы о том, что долгосрочная MPC больше или равна 1, следует использовать

$$z = \frac{0,99403 - 1}{0,016078} = -0,37131.$$

Поскольку используется асимптотическое приближение, обратимся к таблице для стандартного нормального распределения, а не для t . Гипотеза $\gamma = 1$ не отвергается.

Можно было отметить, что эта гипотеза могла быть протестирована с помощью линейного ограничения; если $\delta = 1$, то $\beta = 1 - \gamma$, или $\beta + \gamma = 1$. Оценка $q = b + c - 1 = -0,00045$. Оценка стандартной ошибки этой линейной функции равна $[0,02873^2 + 0,02859^2 - 2(0,0008207)]^{1/2} = 0,00118$. t -статистика для этого теста равна $-0,38135$, что примерно совпадает с предыдущим значением. Поскольку выборка в данном случае достаточно велика, этого следует ожидать. Однако в расчетах нет ничего, что обеспечивало бы этот результат. В меньшей выборке ответ мог получиться другим. Например, используя последние 11 лет наблюдений, t -статистики для двух гипотез равны 7,652 и 5,681. Тест Вальда неинвариантен относительно формулировки гипотезы. В пограничном случае можно было получить иной вывод. Такое **отсутствие инвариантности** не характеризует тесты отношения правдоподобия или множителей Лагранжа, обсуждаемые в главе 14. С другой стороны, оба этих теста требуют предположения о нормальности, тогда как статистика Вальда его не требует. Все это является примером своего рода баланса между более детальной спецификацией и мощностью соответствующей процедуры тестирования.

Обобщение в случае наличия более одной функции параметров получается аналогично. Пусть $c(\hat{\beta})$ — это набор J функций вектора оценки. Обозначим матрицу производных $c(\hat{\beta})$ размерностью $J \times K$ как

$$\hat{\mathbf{G}} = \frac{\partial c(\hat{\beta})}{\partial \hat{\beta}}. \quad (5-36)$$

Оценка асимптотической ковариационной матрицы этих функций тогда имеет вид

$$\text{Est. Asy. Var}[\hat{\mathbf{c}}] = \hat{\mathbf{G}} \{ \text{Est. Asy. Var}[t\hat{\beta}] \} \hat{\mathbf{G}}'. \quad (5-37)$$

j -я строка $\hat{\mathbf{G}}$ представляет собой K производных c_j по K элементам $\hat{\beta}$. Например, ковариационная матрица для оценок краткосрочной и долгосрочной предельной склонности к потреблению может быть получена с использованием

$$\mathbf{G} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1/(1-\gamma) & \beta/(1-\gamma)^2 \end{bmatrix}.$$

Статистика для тестирования J гипотез $c(\beta) = \mathbf{q}$ равна

$$W = (\hat{\mathbf{c}} - \mathbf{q})' \{ \text{Est. Asy. Var}[\hat{\mathbf{c}}] \}^{-1} (\hat{\mathbf{c}} - \mathbf{q}). \quad (5-38)$$

В больших выборках W имеет распределение хи-квадрат с числом степеней свободы, равным числу ограничений. Следует отметить, что для одного ограничения это значение равно квадрату статистики (5-33).

5.8. Выбор между невложенными моделями

Классические процедуры тестирования, приведенные выше, являются наиболее мощными для рассмотренных типов гипотез⁷. Несмотря на то что использование этих процедур, безусловно, желательно, требование спецификации гипотезы в виде $y = \mathbf{X}\beta + \epsilon$,

$$H_0 : \mathbf{R}\beta = \mathbf{q}$$

против

$$H_1 : \mathbf{R}\beta \neq \mathbf{q},$$

может вносить некоторые ограничения. Два популярных исключения: общая проблема определения того, какой из двух наборов регрессоров подходит лучше для описания данных, а также выбор между линейной и логлинейной моделями. В настоящий момент интерес представляет сравнение двух конкурирующих линейных моделей:

$$H_0 : y = \mathbf{X}\beta + \epsilon_0 \quad (5-39)$$

и

$$H_1 : y = \mathbf{Z}\gamma + \epsilon_1. \quad (5-40)$$

⁷Stuart, Ord (1989, ch. 27).

Классические процедуры, рассматривавшиеся до сих пор, не предлагают методов выбора более предпочтительной модели. Общей проблеме тестирования невложенных гипотез такого типа уделяется значительное внимание в теоретической литературе, она встречается в большом числе эмпирических приложений⁸.

5.8.1. Тестирование невложенных гипотез

Полезное различие между процедурами тестирования гипотез, обсуждавшимися в предыдущих разделах, и выбором модели, рассматриваемым здесь, заключается в асимметрии между нулевой и альтернативной гипотезами, являющейся частью классической процедуры тестирования⁹. Поскольку по построению классические процедуры рассматривают выборку на предмет наличия аргументов в пользу отвержения «нулевой» гипотезы, конкретная формулировка этой гипотезы может быть принципиальна для результата. К счастью, методология Неймана–Пирсона дает рекомендацию; нулевой гипотезе обычно соответствует самая узкая модель в рассматриваемом наборе. С другой стороны, классические процедуры никогда не приводят к точным выводам. Пока уровень значимости процедуры тестирования не настолько высок, чтобы исключить все альтернативы, всегда останется вероятность ошибки I рода. В результате нулевая гипотеза никогда не отвергается с убежденностью, а лишь с заданной степенью уверенности. Тесты для выбора спецификации, напротив, рассматривают конкурирующие гипотезы как равные. Естественной нулевой гипотезы в этом случае нет. Однако процесс заканчивается четким решением — при тестировании в виде (5-39а, б) одна из моделей будет отвергнута, а другая принята; затем анализ продолжится в рамках исключительно выбранной модели независимо от отвергнутой. Действительно, он и не может быть продолжен, пока одна из моделей не отвергнута. Частой практикой, например, в такой постановке для исследователя является тестирование с одной из моделей в качестве нулевой, а затем с другой. К сожалению, учитывая способ построения тестов, может сложиться ситуация, при которой ни одна из моделей не будет отвергнута; в любом случае необходимо проводить дальнейший анализ. Как будет показано, наука в этом вопросе несколько неточна.

Первые работы, посвященные тестированию невложенных гипотез, в частности работа Кокса (Cox (1961, 1962)), были написаны в рамках процедур, использующих выборочные правдоподобия и принцип максимального правдоподобия. Последние исследования были выстроены вокруг общей основы, получившей название **принцип охвата** [Mizon, Richard (1986)]. В сущности, принцип концентрирует внимание на вопросе о том, может ли выбранная модель объяснить особенности конкурирующих моделей, т.е. может ли эта модель охватить альтернативу. Третий подход основан на формулировке **полной модели**, которая включает обе конкурирующие модели

⁸Обзоры работ по данной тематике представлены в работах Уайта (White (1982а, 1983)), Гурьера и Монфора (Gourieroux, Monfort (1994)), Макалера (McAleer (1995)), Песарана и Уикса (Pesaran, Weeks (2001)). В обзоре Макалера описано много приложений, тогда как Гурьер и Монфор сосредоточены на теоретических аспектах вопроса.

⁹См. работу Грейнджера и Песарана (Granger, Pesaran (2000)) для обсуждения.

в качестве частных случаев. Когда это возможно, тест между моделями основывается, в сущности, на процедурах тестирования, похожих на классические. Далее будут рассмотрены тесты, использующие все три подхода.

5.8.2. Принцип охвата

Всеобъемлющим называется подход, при котором тестируется способность одной модели объяснить особенности другой. Модель 0 «охватывает» модель 1, если особенности модели 1 могут быть объяснены моделью 0, но обратное неверно¹⁰. Поскольку H_0 не может быть записана в качестве ограничения на H_1 , ни одна из рассмотренных до сих пор процедур не может быть использована. Одной из возможностей является искусственное вложение двух моделей. Пусть \bar{X} представляет набор переменных в X , не входящих в Z . Определим \bar{Z} аналогично по отношению к X , и пусть W будут переменными, которые входят в обе модели. Тогда H_0 и H_1 могут быть объединены в «супермодель»:

$$y = \bar{X}\bar{\beta} + \bar{Z}\bar{\gamma} + W\delta + \varepsilon.$$

В принципе H_1 отвергается, если обнаруживается, что $\bar{\gamma} = 0$, при помощи обыкновенного F -теста, тогда как H_0 отвергается, если обнаруживается что $\bar{\beta} = 0$. С этим подходом связаны две проблемы. Во-первых, δ остается смесью из частей β и γ и F не указывает на равенство нулю какой-либо из этих частей. Таким образом, этот тест на самом деле не позволяет различить между собой H_0 и H_1 ; он различает между собой H_1 и гибридную модель. Во-вторых, эта объединенная модель может иметь очень большое число регрессоров. В контексте временных рядов может возникнуть проблема мультиколлинеарности.

Рассмотрим альтернативный подход. Если H_0 верна, то y будет полностью объясняться X , за исключением случайного члена ε . Предположим, что затем следует попытка оценить γ при помощи регрессии y на Z . Какой бы вектор параметров не получился в результате оценки этой регрессии, назовем его c , если H_0 верна, мы должны оценить в точности тот же вектор коэффициентов в результате регрессии $X\beta$ на Z , поскольку ε_0 является случайным шумом при выполнении H_0 . Поскольку должен быть оценен вектор β , предположим, что вместо этого мы используем Xb и оцениваем c_0 . Тестирование предположения о том, что модель 0 «охватывает» модель 1, будет представлять тест для проверки гипотезы о том, что $E[c - c_0] = 0$. Несложно показать [см. книгу Дэвидсона и Маккиннона (Davidson, MacKinnon (2004, с. 671–672))], что тест может быть проведен с использованием стандартного F -теста для тестирования гипотезы $\gamma_1 = 0$ в дополненной регрессии:

$$y = X\beta + Z_1\gamma_1 + \varepsilon_1,$$

где Z_1 — это переменные в Z , которые не входят в X . (Можно показать, что \bar{Z} и Z_1 совпадают, а значит, совпадают и тесты.)

¹⁰См.: Дитон (Deaton (1982)), Дастур (Dastoor (1983)), Гурьер и др. (Gourieroux et al. (1983, 1995)) и особенно Мизон и Ричард (Mizon, Richard (1986)).

5.8.3. Полная модель — J -тест

Подход, основанный на полной модели, связан с функцией плотности как характеристики процесса порождения данных. Пусть $f_0(y_i | \text{данные}, \beta_0)$ будет предполагаемой функцией плотности для модели 0. Определим альтернативу соответственно как $f_1(y_i | \text{данные}, \beta_1)$. Тогда функция плотности для полной модели, включающей обе модели, имеет вид

$$f_c(y_i | \text{данные}, \beta_0, \beta_1) = \frac{[f_0(y_i | \text{данные}, \beta_0)]^{1-\lambda} [f_1(y_i | \text{данные}, \beta_1)]^\lambda}{\int [f_0(y_i | \text{данные}, \beta_0)]^{1-\lambda} [f_1(y_i | \text{данные}, \beta_1)]^\lambda dy_i},$$

где интеграл берется по области определения функции плотности. Оценка полной модели с последующим тестом $\lambda = 0$ или 1 используется для оценки обоснованности модели 0 или 1 соответственно¹¹.

J -тест, предложенный Дэвидсоном и Маккинном (Davidson, MacKinnon (1981)), является приложением этого принципа к линейной регрессионной модели, что может быть показано [см. Pesaran и Weeks (2001)]. Предложенная ими альтернатива предыдущей полной модели имеет вид

$$y = (1 - \lambda)X\beta + \lambda(Z\gamma) + \epsilon.$$

В этой модели тестирование ограничения $\lambda = 0$ будет тестом против альтернативы H_1 . Проблема состоит в том, что λ не может быть оценена отдельно в этой модели; это означало бы излишнее масштабирование регрессионных коэффициентов. J -тест Дэвидсона и Маккинна заключается в оценке γ при помощи оценки регрессии y на Z с последующей оценкой регрессии y на X и $Z\hat{\gamma}$, являющиеся значениями, предсказанными первой регрессией. Обоснованный, по крайней мере асимптотически, тест H_1 состоит в тестировании $H_0 : \lambda = 0$. Если H_0 истинна, то $\text{plim } \hat{\lambda} = 0$. Асимптотически отношение $\hat{\lambda}/\text{se}(\hat{\lambda})$ (т.е. обычная t -статистика) распределено согласно стандартному нормальному закону распределения, и при тестировании могут использоваться значения из стандартной таблицы. К сожалению, при тестировании H_0 против H_1 и наоборот возможны все четыре варианта (отвергнуть обе или какую-либо одну из гипотез, не отвергнуть ни одну). Эта проблема, однако, связана с конечностью выборки. Дэвидсон и Маккинн показали, что при $n \rightarrow \infty$, если H_1 верна, вероятность того, что $\hat{\lambda}$ будет значимо отличаться от 0, стремится к 1.

Пример 5.7. J -тест функции потребления

Гэвер и Гейзель (Gaver, Geisel (1974)) рассматривали две модели функции потребления:

$$H_0 : C_t = \beta_1 + \beta_2 Y_t + \beta_3 Y_{t-1} + \epsilon_{0t}$$

и

$$H_1 : C_t = \gamma_1 + \gamma_2 Y_t + \gamma_3 C_{t-1} + \epsilon_{1t}.$$

¹¹Сильва (Silva (2001)) демонстрирует приложение к выбору между логит- и пробит-моделями для случая моделей бинарного выбора.

Первая модель предполагает, что потребление реагирует на изменение в доходах в течение двух периодов, тогда как вторая — что изменение в доходах влияет на потребление на протяжении многих периодов. Квартальные данные по агрегированному реальному потреблению в США и реальному располагаемому доходу приведены в таблице из приложения F5.2. Здесь будет показано применение J -теста к этим данным и двум предложенным спецификациям. Во-первых, две модели оцениваются отдельно (используя наблюдения со II квартала 1950 г. по IV квартал 2000 г.). Регрессия C на константу, Y , лаг Y и предсказанные второй моделью значения дают оценку λ , равную 1,0145, с t -статистикой 62,861. Таким образом, H_0 должна быть отвергнута в пользу H_1 . Но, меняя H_0 и H_1 местами, получается оценка λ , равная $-10,677$, с t -статистикой $-7,188$. Таким образом, H_1 также отвергается¹².

5.9. Тестирование спецификации модели

Рассмотренные до сих пор тесты оценивали вложенные модели. Предполагалось, что одна из двух моделей верна. В разделе 5.8 мы расширили класс рассматриваемых моделей на случай двух невложенных моделей. При этом не предполагается, что любая из них обязательно является истинным процессом порождения данных; мы лишь пытаемся выяснить при помощи теста, какая из двух моделей находится ближе к истине. Тесты на спецификацию находятся между этими двумя подходами. Идея **теста на спецификацию модели** состоит в рассмотрении конкретной нулевой модели и альтернатив, которые не задаются явным образом в качестве ограничений на регрессионное уравнение. Полезным способом рассмотрения некоторых тестов на спецификацию является подход, при котором центральная модель $y = X\beta + \epsilon$ отождествляется с нулевой гипотезой, а альтернатива представляет собой, возможно, незаявленное обобщение этой модели. Тест Рамсея (Ramsey (1969)) **RESET-тест** рассматривает нелинейности в функциональной форме и является одним из примеров таких тестов. Один (являющийся неоднозначным) подход к анализу выглядит следующим образом:

- $H_0: y = X\beta + \epsilon$,
- $H_1: y = X\beta +$ более высокие степени x_k и другие переменные $+\epsilon$.

Естественным подходом было бы включение в модель квадратов, кубов и перекрестных произведений регрессоров с последующим тестированием до H_0 как ограничения на большую модель. Сложности состоят в том, что, во-первых, этот подход может слишком конкретно задавать форму альтернативной гипотезы и что, во-вторых, при большом числе регрессоров X он может быстро разрастись. Решение, предложенное Рамсеем, заключается во включении степеней $x'_i\beta$ в регрессию, используя предсказания метода наименьших квадратов, обычно следует добавить квадраты и, возможно, кубы регрессоров. Для этого нужна двухшаговая процедура, поскольку, чтобы

¹²Обсуждение такой возможности см. в работе Макалера, Фишера и Фолкера (McAleer, Fisher, Volker (1982)).

добавить $(\mathbf{x}'_i \mathbf{b})^2$ и $(\mathbf{x}'_i \mathbf{b})^3$, требуются коэффициенты. Предложение состоит в том, чтобы сначала оценить модель, используя метод наименьших квадратов, а затем, на втором шаге, квадраты (и кубы) предсказанных на первом шаге значений добавить в уравнение и переоценить его. Затем (асимптотический) тест Вальда используется для тестирования гипотез, соответствующих исходной модели.

В качестве общей стратегии такого рода спецификация предназначена для определения нарушений предпосылок нулевой модели. Очевидным преимуществом такого теста является то, что он допускает существенно большую степень общности, чем простая процедура тестирования ограничений, например равенства коэффициента нулю. Но такая общность дается существенной ценой:

1. Тест неконструктивен. Он не указывает, что должен делать исследователь, если модель, лежащая в основе нулевой гипотезы, отвергается. Это общая особенность тестов на спецификацию модели. Отвержение нулевой модели не указывает ни на какую конкретную альтернативу.

2. Поскольку альтернативная гипотеза не указывается, неясно, какова мощность теста против какой-либо конкретной альтернативы.

3. Для этого конкретного теста (возможно, не для некоторых других тестов на спецификацию модели, которые мы рассмотрим позднее), поскольку $\mathbf{x}'_i \mathbf{b}$ использует один \mathbf{b} для всех наблюдений, наблюдения являются коррелированными, тогда как в исходной модели предполагается, что они некоррелированы. Ввиду двухшаговой природы оценки неясно, какую ковариационную матрицу следует использовать для теста Вальда. Две другие сложности возникают в связи с этим тестом. Во-первых, неясно, куда сходятся коэффициенты, предполагая, что они сходятся куда бы то ни было. Во-вторых, дисперсия разности между $\mathbf{x}'_i \mathbf{b}$ и $\mathbf{x}'_i \beta$ — это функция \mathbf{x} , так что регрессия на втором шаге может оказаться гетероскедастичной. Это приводит к тому, что размер и мощность теста могут оказаться совершенно не теми, чего следовало ожидать.

Пример 5.8. Размер RESET-теста

Для изучения истинного размера RESET-теста в конкретном приложении при помощи метода Монте-Карло был проведен эксперимент. Результаты, представленные в табл. 4.6, приводят следующие оценки уравнения (5-2):

$$\ln Price = -8,42653 + 1,33372 \ln Area - 0,16537 Aspect Ratio + e,$$

где $sd(e) = 1,10266$.

Мы рассматриваем правую часть регрессии как нашу генеральную совокупность. Далее генерируем 5000 выборок по 430 наблюдениям (размер исходной выборки), используя один набор регрессионных коэффициентов и генерируя новую выборку ошибок для каждой реализации. Таким образом, для каждой реализации r есть новая выборка наблюдений $\ln Price_{it}$, для которой регрессионная часть используется много раз, а ошибки генерируются заново. Для каждой выборки оценивается коэффициент и предсказанное значение. Затем регрессия переоценивается с включением

в нее квадрата и куба предсказанного значения. Наконец, для каждой выборки оценивается статистика хи-квадрат и модель, лежащая в основе нулевой гипотезы, отвергается при значении этой статистики, превышающем 5,99, 95-й процентиль распределения хи-квадрат с двумя степенями свободы. **Номинальный размер** этого теста составляет 0,05. Таким образом, в выборках, состоящих из 100, 500, 1000 и 5000 наблюдений, модель, лежащая в основе нулевой гипотезы, должна отвергаться 5, 25, 50 и 250 раз. В рассматриваемом эксперименте значение хи-квадрат-статистики превысило 5,99 для 8, 31, 65 и 259 случаев соответственно, что указывает на то, что, по крайней мере, при достаточном числе повторений тест ведет себя, как и ожидается. Затем исследовалась мощность теста путем подключения к предсказаниям In Area, умноженного на 0,1. Невозможно определить точную мощность RESET-теста в выявлении этого нарушения нулевой модели. В эксперименте с 1000 повторений нулевая гипотеза была отвергнута 321 раз. Следует заключить, что процедура обладает мощностью в отношении выявления этого нарушения предположений модели.

5.10. Построение модели — подход от общего к частному

В последние 20 лет произошел сдвиг в подходе к построению моделей отчасти ввиду результатов, представленных в предыдущих двух разделах. Стараясь сохранить простоту спецификации, исследователи обычно начинают с маленькой спецификации, планомерно расширяя ее до интересующей, включая дополнительные переменные. Но на основе предыдущих результатов можно предположить, что любой критерий, который может быть использован для принятия решения о включении переменной в текущую спецификацию, будет испорчен смещениями, вызванными неполной спецификацией на начальных этапах. Пропуск важных переменных в общем случае оказывается худшей из двух возможных ошибок. Таким образом, **подход от частного к общему** имеет не много доводов в свою пользу. Основываясь на работе Хендри [например, Hendry (1995)] и последних достижениях в компьютерных технологиях, исследователи находятся в более комфортной ситуации, начиная поиск правильной спецификации с больших моделей, включающих много переменных и, возможно, сложные лаговые структуры. Привлекательная стратегия заключается в применении подхода **от общего к частному**, т.е. при сокращении модели до предпочтительной спецификации. [Этот процесс был полностью автоматизирован в компьютерной программе Hendry PCGets^(c). См., например, Хендри и Котцис (Hendry, Kotzis (2001)).] Конечно, такой подход должен учитывать два связанных между собой соображения. В регрессии «все в кучу», включающей все переменные, которые могут оказаться нужными, принятие фиксированной вероятности ошибки I рода, например 5%, гарантирует для достаточно большой модели, что некоторые переменные покажутся значимыми, скорее всего, «случайно». Во-вторых, проблемы, связанные с предварительным оцениванием и **пошаговым построением модели**, также связаны с риском итоговой

неправильной спецификации модели. В качестве одного, к сожалению, часто встречающегося примера можно привести ситуацию, когда используемые таким образом статистики приводят к необъяснимым лаговым структурам в динамических моделях с большим количеством лагов зависимой или независимой переменной.

5.10.1. Критерии выбора модели

В ходе предшествующего обсуждения было указано на ряд подходов к выбору модели, основанных на тестировании невложенных гипотез. Показатели качества модели и процедуры тестирования, основанные на сумме квадратов ошибок, такие как R^2 и тест Кокса (Сох (1961)), полезны, когда интерес представляет описание данных или прогнозирование зависимой переменной внутри выборки. Когда же построение модели нацелено на прогнозирование, внутривыборочные показатели необязательно оптимальны. Как было показано, R^2 не может падать после включения переменных в модель, так что для него характерна тенденция к «слишком хорошему» описанию выборки. Этот критерий может помешать выбрать наилучшую прогнозирующую модель, поскольку включение переменных в модель может увеличить дисперсию прогнозной ошибки (см. раздел 4.6), несмотря на улучшенное качество описания данных. Учитывая это соображение, **нормированный** или **скорректированный R^2** ,

$$\bar{R}^2 = 1 - \frac{n-1}{n-K}(1-R^2) = 1 - \frac{n-1}{n-K} \left(\frac{\mathbf{e}'\mathbf{e}}{\sum_{i=1}^n (y_i - \bar{y})^2} \right), \quad (5-41)$$

был предложен как мера качества, которая корректно штрафует потерю степеней свободы, вызванную включением дополнительных переменных в модель. Следует отметить, что \bar{R}^2 может упасть при включении переменной в модель, если сумма квадратов остатков падает недостаточно быстро. (Результат приводится в теореме 3.7; \bar{R}^2 не растет при включении переменной в модель, если t -статистика для параметра при этой переменной не превышает единицы по абсолютному значению.) Предпочтительность нормированного R^2 для оценки качества прогноза была продемонстрирована в литературе. [См. Diebold (2003), где демонстрируется, что простой R^2 имеет отрицательное смещение как мера дисперсии ошибки прогноза на шаг вперед за пределы выборки.]

Нормированный R^2 штрафует потерю степеней свободы, которая происходит при расширении модели. Однако возникает вопрос о том, достаточно ли велик штраф для того, чтобы убедиться в истинности модели, к которой приведет исследователя этот показатель (предполагая, что истинная модель находится среди рассматриваемых моделей) с ростом выборки. Было предложено два альтернативных показателя качества модели: **информационный критерий Акаике**:

$$\text{AIC}(K) = s_y^2(1-R^2)e^{2K/n} \quad (5-42)$$

и **информационный критерий Шварца**, или **байесовский**:

$$\text{BIC}(K) = s_y^2(1-R^2)n^{K/n}. \quad (5-43)$$

(s_y^2 не корректируется на число степеней свободы.) Оба показателя улучшаются (ухудшаются) с увеличением (уменьшением) R^2 , но при прочих равных сокращаются при увеличении числа переменных в модели. Как и \bar{R}^2 , эти показатели поощряют достижение заданного качества при меньшем числе параметров на одно наблюдение, K/n . Логарифмы обычно более удобны; показатели качества, приводимые большинством программных пакетов, в этом случае имеют вид

$$\text{AIC}(K) = \ln\left(\frac{\mathbf{e}'\mathbf{e}}{n}\right) + \frac{2K}{n}, \quad (5-44)$$

$$\text{BIC}(K) = \ln\left(\frac{\mathbf{e}'\mathbf{e}}{n}\right) + \frac{K \ln n}{n}. \quad (5-45)$$

Оба критерия качества имеют свои преимущества, ни один не превосходит другой. [См. Diebold (2003).] **Критерий Шварца** с более строгим штрафом за потерю степеней свободы будет склонен выбирать более простую модель. При прочих равных простота действительно имеет свои преимущества.

5.10.2. Выбор модели

Выше было приведено несколько возможных способов **выбора модели**, но в то же время было сделано несколько неудобных предложений. Особенности поиска спецификации, связанные с претестовой оценкой, основаны на знании «истины» исследователем и последствий того, что это знание не используется. Несмотря на общеизвестность последствий слепого поиска статистической значимости, предполагать, что истинная модель станет известна до начала анализа, представляется чрезмерно оптимистичным. Смещение, приведенное в выражении (4-10), достойно внимательного изучения. Но на практике любая информация о его величине предполагает, что мы знаем, какие переменные входят в X_2 , чего может не быть в действительности. Наряду с тем, что мы можем согласиться, что исследователь исключит доход из уравнения спроса на свой страх и риск, мы также можем посочувствовать исследователю, ищущему правильную спецификацию для своей прогнозной модели среди десятков альтернатив. Тесты для невложенных моделей должны освобождать исследователя от необходимости делать заявление, что представленный набор моделей содержит «истину». Но эта свобода дается ценой снижающейся мощности таких процедур. J -тест может дать возможность сделать точный выбор между двумя альтернативами, но он исключает третью возможность — того, что обе модели неверны. Тест Вонга (см. подраздел 14.6.6), конечно, допускает такую возможность, но его характеризует достаточно большая область неопределенности, что указывает на отсительно низкую мощность против большого количества альтернатив. Вывод состоит в том, что в области выбора модели предстоит сделать еще довольно много. Недавние комментарии представляют предложения с двух точек зрения — классической и байесовской.

5.10.3. Классический подход к выбору модели

Хансен (Hansen (2005)) приводит четыре недостатка рассмотренной выше методологии:

- 1) параметрический подход;
- 2) предположение об истинном процессе, порождающем данные;
- 3) оценка, основанная на качестве подгонки модели;
- 4) игнорирование неопределенности относительно модели.

Все эти аспекты рассматривались в предыдущих разделах. Точка зрения Хансена состоит в том, что приведенный анализ слишком узок и мешает развитию процедур поиска моделей.

Все представленные здесь процедуры выбора модели основаны на функции правдоподобия, которая требует конкретных предположений о распределении. Хансен, напротив, предлагает обратить внимание на полупараметрические структуры. В случае регрессионного анализа это приводит к рассмотрению оценок обобщенным методом моментов. Жертвами такого поворота будут тестовые статистики, основанные на распределении, такие как статистики Кокса и Вонга, и даже показатели AIC и BIC, которые являются преобразованиями функции правдоподобия. Однако здесь были предложены альтернативы [см., например, Hong, Preston, Shum (2000)]. Второе направление критики уже было озвучено. Предполагаемая «истинная» модель может быть слишком узкой. Вместо этого (утверждает Хансен) следует рассматривать спецификацию как приближение истинного процесса, порождающего данные, это в существенной степени расширяет поиск спецификации, направляя его в сторону модели, которая дает наилучшее приближение. Конечно, здесь следует поднять вопрос о том, что есть «лучшее». До этого момента мы рассматривали функцию правдоподобия, которая в случае классической регрессии может рассматриваться как возрастающая функция R^2 . Автор выступает за более «сфокусированный» информационный критерий (FIC), который напрямую рассматривает интересующие нас параметры, а не качество описания данных моделью. Каждое из этих предложений направлено на улучшение процесса выбора модели на основе знакомых критериев, таких как тестовые статистики, основанные на показателях качества и характеристиках модели.

Возможно, центральной проблемой является неопределенность относительно самой модели. Поиск истинной модели, вероятно, оказывает влияние на статистические выводы, сопоставимое с поиском спецификации, при условии заданной формы модели (см. подразделы 4.3.2 и 4.3.3). К сожалению, включение такого рода неопределенности в статистические процедуры остается нерешенной задачей. Хансен предлагает в качестве возможного направления методы байесовского усреднения моделей, обсуждаемые ниже, хотя и высказывает некоторый скептицизм относительно байесовских методов в целом.

5.10.4. Байесовское усреднение моделей

При наличии сомнений относительно уместности одной из двух моделей существует возможность сделать вывод о том, что «истинной» не является ни одна из них. Подход к тестированию, основанный на принципе «вправо или влево», загоняет исследователя в тупик. Байесовский подход к этому вопросу означает рассмотрение проблемы сравнения двух гипотез, а не обоснованности одной по отношению к другой. Выборочный эксперимент

начинается с набора априорных вероятностей об относительных преимуществах двух гипотез, который обобщается в «априорном отношении вероятностей» $P_{01} = \text{Prob}[H_0]/\text{Prob}[H_1]$. После сбора данных строится байесовский множитель, который обобщает вес аргументов в выборке в пользу той или иной модели. В результате анализа данных получается «постериорное отношение вероятностей» $P_{01} | \text{data} = \text{Bayes factor} \times P_{01}$. В конечном счете, *ex post*, ни одна модель не отбрасывается; просто наше представление об относительном правдоподобии моделей пересматривается с учетом данных. Конечно, такой подход оставляет открытым вопрос спецификации. Выбирая между моделями, как наилучшим образом использовать имеющуюся в наличии информацию? В недавних работах, посвященных **байесовскому усреднению моделей** [Hoeting et al. (1999)], был предложен ответ.

Приложение Райта (Wright (2003)) представляет интересную иллюстрацию. Новые методы, такие как байесовская векторная авторегрессия (BVAR), улучшили прогнозную силу эконометрических моделей. Сток и Уотсон (Stock, Watson (2001, 2004)) показывают, что существенное улучшение в качестве прогноза международной инфляции может быть получено путем усреднения большого числа прогнозов различных моделей и источников. Этот примечательный результат устойчив по отношению к странам и периодам. На его основе было озвучено две идеи. Во-первых, идея смешения различных моделей в существенной степени отражает дух четвертого аргумента Хансена. Во-вторых, следует отметить, что суть улучшения состоит не в качестве описания данных (аргумент 3), а в предсказательной силе модели. Сток и Уотсон отметили, что простое усреднение с равными весами по непонятным пока причинам приносит существенное улучшение. Райт предложил байесовское усреднение моделей в качестве более систематического подхода при выборе весов для получения еще лучшей предсказательной силы.

Лимер (Leamer (1978)), по всей видимости, первым предложил байесовское усреднение моделей в качестве способа комбинирования моделей. Позднее эта идея была рассмотрена Мином и Зельнером (Min, Zellner (1993)) в контексте прогнозирования роста выпуска, Доппельхофером и др. (Doppelhofer et al. (2000)) для межстрановых регрессий, Купом и Поттером (Koop, Potter (2004)) для макроэкономических предсказаний, а также рядом других исследователей. Предположим, что рассматривается M моделей, индексируемых при помощи индекса $m = 1, \dots, M$. Для простоты запишем m -ю модель в простой форме, $f_m(\mathbf{y} | \mathbf{Z}, \boldsymbol{\theta}_m)$, где $f(\cdot)$ — это функция плотности, \mathbf{y} и \mathbf{Z} — это данные и $\boldsymbol{\theta}_m$ — это вектор параметров модели m . Предположим также, что модель m^* истинна, но неизвестна исследователю. Исследователь имеет априорные представления π_m относительно вероятности, что модель m верна, так что π_m — это априорная вероятность того, что $m = m^*$. Постериорные вероятности для моделей равны

$$\Pi_m = \text{Prob}(m = m^* | \mathbf{y}, \mathbf{Z}) = \frac{P(\mathbf{y}, \mathbf{Z} | m)\pi_m}{\sum_{r=1}^M P(\mathbf{y}, \mathbf{Z} | r)\pi_r}, \quad (5-46)$$

где $P(\mathbf{y}, \mathbf{Z} | m)$ — это безусловное правдоподобие для m -й модели,

$$P(\mathbf{y}, \mathbf{Z} | m) = \int_{\boldsymbol{\theta}_m} P(\mathbf{y}, \mathbf{Z} | \boldsymbol{\theta}_m, m)P(\boldsymbol{\theta}_m)d\boldsymbol{\theta}_m, \quad (5-47)$$

где $P(\mathbf{y}, \mathbf{Z} | \theta_m, m)$ — это условное (относительно θ_m) правдоподобие m -й модели и $P(\theta_m)$ — это априорное представление исследователя относительно параметров m -й модели. Все это дает набор весов, альтернативный $\Pi_m = 1/M$, предложенному Стоком и Уотсоном. Пусть $\hat{\theta}_m$ обозначает байесовскую оценку (постериорное среднее) параметров модели m . (См. главу 16.) Каждая модель предоставляет апостериорную прогнозную плотность $f^*(\mathbf{y} | \mathbf{Z}, \hat{\theta}_m, m)$. Байесовская усредненная по моделям прогнозная плотность в таком случае будет равна

$$\bar{f}^* = \sum_{m=1}^M f^*(\mathbf{y} | \mathbf{Z}, \hat{\theta}_m, m) \Pi_m. \quad (5-48)$$

Точечный прогноз тогда будет равен взвешенному подобным образом значению прогнозов отдельных моделей.

Пример 5.9. Байесовское усреднение классических оценок

Многие исследователи относятся скептически к байесовским методам из-за явной произвольности спецификации априорных функций плотности неизвестных параметров. В контексте байесовского усреднения моделей от исследователя требуются априорные плотности не только для вероятностей моделей π_m , но и для параметров отдельных моделей θ_m . В своем исследовании Доппельхофер, Миллер и Сала-и-Мартин (Doppelhofer, Miller, Sala-i-Martin (2000)) исследовали правильность набора регрессоров для включения в уравнение долгосрочного макроэкономического роста (дохода). При рассмотрении 32 возможных кандидатов M для их приложения составляло 2^{32} (минус один, если исключить модель без регрессоров), или примерно 4 млрд. Формулировка такого числа априорных представлений была бы крайне оптимистична. Авторы предложили новый способ взвешивания большого подмножества (приблизительно 21 млн) из 2^M возможных (классических) регрессий. Веса формировались с использованием байесовской процедуры, однако взвешивались оценки, получаемые классическим методом наименьших квадратов. И хотя этот способ существенным образом сокращает затраты на оценивание моделей, он все же требует оценивания миллионов МНК-регрессий. [См. Sala-i-Martin (1997).] В результате была получена модель с 12 независимыми переменными.

5.11. Заключение и выводы

В данной главе был рассмотрен третий аспект приложения линейной регрессии — тестирование гипотез. Главным результатом здесь является F -статистика, которая может быть получена одним из двух способов. Во-первых, как мера степени отличия оценки модели без ограничений от оценки модели, лежащей в основе нулевой гипотезы. Во-вторых, как мера потери качества подгонки модели, происходящей в результате предположения, что

нулевая гипотеза верна. Затем мы обобщили F -статистику на случай больших выборок и рассмотрели ее асимптотические свойства, что позволило нам избавиться от предположения о нормальности остатков и обобщить ее на случай нелинейных ограничений.

Эта глава является последней из пяти глав, посвященных методологии, связанной с наиболее широко применяемым инструментом в эконометрике — классической регрессионной моделью. Глава 2 началась с формулировки регрессионной модели. Затем в главе 3 был описан процесс получения оценок параметров при помощи метода наименьших квадратов — исключительно алгебраическое упражнение. В главе 4 оценка методом наименьших квадратов была переформулирована в терминах оценки неизвестного вектора параметров, а также были представлены выборочные свойства этой оценки в больших и малых выборках. Глава 5 была посвящена построению и уточнению регрессионной модели с помощью статистических результатов для тестирования гипотез относительно свойств генеральной совокупности. В этой главе были рассмотрены более общие вопросы, связанные со спецификацией модели и выбором модели из набора конкурирующих альтернатив. Рассмотренные здесь концепции тесно связаны с одной из основ эконометрической парадигмы; в основе модели лежит некоторая теоретическая конструкция, набор истинных поведенческих соотношений, которые составляют модель. Только в этом смысле понятия смещения и смещенных оценок имеют какой-то смысл — «смещение» как понятие может быть описано только в отношении некоторой «модели», по сравнению с которой оценка может быть смещенной, т.е. требуется отправная точка. Это понятие представляет собой центральный момент при анализе спецификации, в ходе которого были рассмотрены последствия недостаточно хорошего (исключение переменных) и слишком хорошего (включение избыточных переменных) описания данных моделью. Глава (и наше обсуждение линейной регрессионной модели) завершилась рассмотрением процедур, используемых для выбора между конкурирующими спецификациями модели.

Основные термины и понятия

| | |
|---|---|
| Acceptance region — Область принятия | Encompassing principle — Принцип охвата |
| Adjusted R-squared — Нормированный R -квадрат | Exclusion restrictions — Исключающие ограничения |
| Akaike Information Criterion — Информационный критерий Акаике | Ex post forecast — Ex post-прогноз |
| Alternative hypothesis — Альтернативная гипотеза | Functionally independent — Функционально независимый |
| Bayesian model averaging — Байесовское усреднение моделей | General nonlinear hypothesis — Общая нелинейная гипотеза |
| Bayesian Information Criterion — Байесовский информационный критерий | General-to-simple strategy — Стратегия от общего к частному |
| Biased estimator — Смещенная оценка | Inclusion of superfluous variables — Включение избыточных переменных |
| Comprehensive model — Полная модель | J test — J -тест |
| Consistent — Состоятельный | Lack of invariance — Отсутствие инвариантности |
| Distributed lag — Распределенный лаг | |
| Discrepancy vector — Вектор отклонений | |

| | |
|--|---|
| Lagrange multiplier test — Тест множителей Лагранжа | Rejection region — Область отвержения |
| Linear restrictions — Линейные ограничения | Restricted least squares — Метод наименьших квадратов с ограничениями |
| Mean squared error — Среднеквадратичная ошибка | Root mean squared error — Корень квадратный из среднеквадратичной ошибки |
| Model selection — Выбор модели | Sample discrepancy — Выборочное отклонение |
| Nested — Вложенный | Schwarz criterion — Критерий Шварца |
| Nested models — Вложенные модели | Simple-to-general — От простого к сложному |
| Nominal size — Номинальный размер | Size of the test — Размер теста |
| Nonnested — Невложенный | Specification test — Тест на спецификацию модели |
| Nonnested models — Невложенные модели | Stepwise model building — Пошаговое построение модели |
| Nonnormality — Ненормальность | t ratio — t -статистика |
| Null hypothesis — Нулевая гипотеза | Testable implications — Тестируемые утверждения |
| One-sided test — Односторонний тест | Theil U statistic — U -статистика Тейла |
| Parameter space — Пространство параметров | Wald criterion — Критерий Вальда |
| Power of a test — Мощность теста | Wald distance — Расстояние Вальда |
| Prediction criterion — Критерий предсказания | Wald statistic — Статистика Вальда |
| Prediction interval — Предсказываемый интервал | Wald test — Тест Вальда |
| Prediction variance — Дисперсия предсказания | |

Упражнения

1. Оценка множественной регрессии y на константу x_1 и x_2 дает следующие результаты: $\hat{y} = 4 + 0,4x_1 + 0,9x_2$, $R^2 = 8/60$, $e'e = 520$, $n = 29$,

$$X'X = \begin{bmatrix} 29 & 0 & 0 \\ 0 & 50 & 10 \\ 0 & 10 & 80 \end{bmatrix}.$$

- Протестируйте гипотезу о том, что сумма коэффициентов наклона равна 1.
- Используя результаты из упражнения 1, протестируйте гипотезу о том, что коэффициент наклона при x_1 равен 0, оценив регрессию с ограничениями и сравнив две суммы квадратов отклонений.
- Анализируемая регрессионная модель имеет вид $y = X_1\beta_1 + X_2\beta_2 + \varepsilon$, где матрицы X_1 и X_2 имеют K_1 и K_2 столбцов соответственно. Накладывается ограничение $\beta_2 = 0$.
 - Используя (5-23), докажите, что оценка с ограничениями имеет вид $[b_{1*}, 0]$, где b_{1*} представляет собой вектор МНК-оценки в регрессии y на X_1 .
 - Докажите, что если наложить ограничение $\beta_2 = \beta_2^0$ для ненулевого вектора β_2^0 , то оценка с ограничениями β_1 равна $b_{1*} = (X_1'X_1)^{-1}X_1'(y - X_2\beta_2^0)$.

4. Выражение для вектора коэффициентов в модели с ограничениями (5-23) может быть записано в виде $\mathbf{b}_* = [\mathbf{I} - \mathbf{CR}]\mathbf{b} + \mathbf{w}$, где \mathbf{w} не включает \mathbf{b} . Чему равно \mathbf{C} ? Покажите, что ковариационная матрица МНК-оценки с ограничениями равна

$$\sigma^2(\mathbf{X}'\mathbf{X})^{-1} - \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}$$

и что эта матрица может быть записана как

$$\text{Var}[\mathbf{b} | \mathbf{X}] \{ [\text{Var}(\mathbf{b} | \mathbf{X})]^{-1} - \mathbf{R}'[\text{Var}(\mathbf{Rb} | \mathbf{X})]^{-1}\mathbf{R} \} \text{Var}[\mathbf{b} | \mathbf{X}].$$

5. Докажите, что МНК-оценка с ограничениями не может иметь бóльшую ковариационную матрицу, чем МНК-оценка без ограничений.
6. Докажите, что R^2 , связанный с МНК-оценкой с ограничениями, никогда не превышает R^2 , связанного с оценкой без ограничений. Сделайте вывод о том, что наложение ограничений никогда не улучшает качество регрессии.
7. Альтернативный способ протестировать гипотезу $\mathbf{R}\beta - \mathbf{q} = \mathbf{0}$ заключается в использовании теста Вальда для проверки гипотезы о том, что $\lambda_* = \mathbf{0}$, где λ_* определена в (5-23). Докажите что

$$\chi^2 = \lambda_*' \{ \text{Est. Var}[\lambda_*] \}^{-1} \lambda_* = (n - K) \left[\frac{\mathbf{e}'_*\mathbf{e}_*}{\mathbf{e}'\mathbf{e}} - 1 \right].$$

Заметьте, что дробь в квадратных скобках представляет отношение двух оценок σ^2 . Благодаря (5-28) и обсуждению выше мы знаем, что это отношение больше 1. Наконец, докажите что эта тестовая статистика эквивалентна статистике JF , где J равно числу тестируемых ограничений и F — это обычная F -статистика, приведенная в (5-16). Формально тест множителей Лагранжа требует, чтобы оценка дисперсии была основана на МНК с ограничениями. В таком случае тестовая статистика примет вид $LM = nJ / [(n - K)/F + J]$. (См. Godfrey (1988)).

8. Используйте тестовую статистику, выведенную в упражнении 7, для тестирования гипотезы из упражнения 1.
9. Докажите, что в условиях нулевой гипотезы о том, что $\mathbf{R}\beta = \mathbf{q}$, оценка

$$s_*^2 = \frac{(\mathbf{y} - \mathbf{Xb}_*)'(\mathbf{y} - \mathbf{Xb}_*)}{n - K + J},$$

где J — число ограничений, является несмещенной оценкой σ^2 .

10. Покажите, что оценка множественной регрессии y на константу x_1 и x_2 при наличии ограничений $\beta_1 + \beta_2 = 1$ эквивалентна оценке регрессии $y - x_1$ на константу $x_2 - x_1$.
11. Предположим, что истинная регрессионная модель приведена в (4-8). Результат в (4-10) показывает, что если или $\beta_{1.2}$ или β_2 не равны нулю, то регрессия y только на X_1 дает смещенную и несостоятельную оценку β_1 . Предположим, интерес представляет прогнозирование y , а не оценка параметров. Рассмотрим регрессию только на X_1 для оценки β_1 с помощью \mathbf{b}_1 (которая является смещенной). Будет ли прогноз y , полученный с использованием $\mathbf{X}_1\mathbf{b}_1$, также смещен? Предположим, что $E[\mathbf{X}_2 | \mathbf{X}_1]$ — линейная функция \mathbf{X}_1 . Обсудите ваши результаты в общем случае. Как называется на прогнозировании исключение переменных из регрессии?

12. Сравните среднеквадратичные ошибки b_1 и $b_{1.2}$ из подраздела 4.7.2. (Подсказка: сравнение зависит от данных и параметров модели, но для двух величин может быть получено компактное выражение.)
13. Функция логарифмического правдоподобия для линейной регрессионной модели с нормально распределенными ошибками представлена в примере 4.6. Покажите, что в оценках максимального правдоподобия \mathbf{b} для β и $e'e/n$ для σ^2 логарифм правдоподобия является возрастающей функцией R^2 для модели.
14. Покажите, что модель, соответствующая альтернативной гипотезе из примера 5.7, может быть записана как

$$H_1: C_t = \theta_1 + \theta_2 Y_t + \theta_3 Y_{t-1} + \sum_{s=2}^{\infty} \theta_{s+2} Y_{t-s} + \varepsilon_{it} + \sum_{s=1}^{\infty} \lambda_s \varepsilon_{t-s}.$$

В таком случае H_0 выглядит как ограничение на H_1 . Однако, поскольку ограничений бесконечное множество, это не сводит тест к стандартному тестированию ограничений, но тем не менее указывает на связь между двумя формулировками.

Эмпирические приложения

1. В приложении к главе 3 мы использовали 15 из 17 919 наблюдений из исследования взаимосвязи между заработной платой и образованием, способностями и характеристиками семьи, проведенного Купом и Тобиасом (Кoop, Tobias (2004)). (См. таблицу из приложения F3.2.) В этом упражнении мы будем использовать полный набор данных. Они могут быть получены в архиве данных *Journal of Applied Econometrics* на <http://www.econ.queensu.ca/jae/12004-v19.7/koop-tobias/>. Файл с данными состоит из двух частей. Первая часть файла содержит панель из 17 919 наблюдений для следующих переменных:

| | |
|-----------|---|
| Столбец 1 | <i>Person id (om 1 до 2,178)</i> — номер индивида |
| Столбец 2 | <i>Education</i> — образование |
| Столбец 3 | <i>Log of hourly wage</i> — логарифм почасовой заработной платы |
| Столбец 4 | <i>Potential experience</i> — опыт |
| Столбец 5 | <i>Time trend</i> — временной тренд |

Столбцы 2–5 содержат переменные, меняющиеся во времени. Вторая часть данных содержит постоянные во времени переменные для 2178 домохозяйств:

| | |
|-----------|--|
| Столбец 1 | <i>Ability</i> — способности |
| Столбец 2 | <i>Mother's education</i> — образование матери |
| Столбец 3 | <i>Father's education</i> — образование отца |
| Столбец 4 | Dummy variable for residence in a broken home — фиктивная переменная, равная 1, если человек живет в плохом доме |
| Столбец 5 | <i>Number of siblings</i> — число братьев или сестер. |

Для создания данных для этого упражнения необходимо совместить эти части файла. i -е наблюдение во втором файле будет воспроизведено T_i раз для набора из T_i наблюдений в первом файле. Переменная *person id* указывает на номера строк, которые должны содержать данные из второго файла. (Конкретные процедуры подготовки данных будут зависеть от используемого компьютерного пакета.) (Замечание. Мы не пытаемся воспроизвести результаты Купа и Тобиаса, мы только используем их интересные данные.) Пусть $X_1 = [\text{constant, education, experience, ability}]$ и пусть $X_2 = [\text{mother's education, father's education, broken home, number of siblings}]$.

- a. Оцените полную регрессию $\log \text{wage}$ на X_1 и X_2 и приведите все результаты.
 - b. Используйте F -тест для проверки гипотезы о том, что все коэффициенты, кроме постоянного члена, равны нулю.
 - c. Используйте F -статистику для проверки совместной гипотезы о том, что коэффициенты при четырех переменных, входящих в X_2 , равны нулю.
 - d. Используйте тест Вальда для проведения теста из пункта c.
2. Обобщенная функция издержек Кобба–Дугласа, рассмотренная в эмпирическом приложении 2 к главе 4, представляет специальный случай **транслогарифмической функции издержек**:

$$\begin{aligned} \ln C = \alpha &+ \beta \ln Q + \delta_k \ln P_k + \delta_l \ln P_l + \delta_f \ln P_f \\ &+ \phi_{kk} \left[\frac{1}{2} (\ln P_k)^2 \right] + \phi_{ll} \left[\frac{1}{2} (\ln P_l)^2 \right] + \phi_{ff} \left[\frac{1}{2} (\ln P_f)^2 \right] \\ &+ \phi_{kl} [\ln P_k][\ln P_l] + \phi_{kf} [\ln P_k][\ln P_f] + \phi_{lf} [\ln P_l][\ln P_f] \\ &+ \gamma \left[\frac{1}{2} (\ln Q)^2 \right] \\ &+ \theta_{Qk} [\ln Q][\ln P_k] + \theta_{Ql} [\ln Q][\ln P_l] + \theta_{Qf} [\ln Q][\ln P_f] + \varepsilon. \end{aligned}$$

Теоретическое требование линейной однородности по ценам факторов накладывает следующие ограничения:

$$\begin{aligned} \delta_k + \delta_l + \delta_f = 1, & \quad \phi_{kk} + \phi_{kl} + \phi_{kf} = 0, & \quad \phi_{kl} + \phi_{ll} + \phi_{lf} = 0, \\ \phi_{kf} + \phi_{lf} + \phi_{ff} = 0, & \quad \theta_{Qk} + \theta_{Ql} + \theta_{Qf} = 0. \end{aligned}$$

Заметим, что, несмотря на то что теория требует наложения ограничения линейной однородности, модель может быть оценена методом наименьших квадратов без ограничений. [Таким образом, можно «протестировать» теорию, протестировав правильность этих ограничений. См.: Christensen, Jorgenson, Lau (1975).] Мы повторим это упражнение в пункте b. Некоторые дополнительные ограничения были рассмотрены в исследовании Кристенсена и Грина (Christensen, Greene (1976)). Гипотеза гомотетичности производственной структуры добавляет дополнительные ограничения:

$$\theta_{Qk} = 0, \quad \theta_{Ql} = 0, \quad \theta_{Qf} = 0.$$

Однородность производственной структуры добавляет ограничение $\gamma = 0$. Гипотеза о равенстве всех эластичностей замещения в производственной структуре -1 накладывает шесть ограничений $\phi_{ij} = 0$

для всех i и j . Мы используем данные из раннего приложения для тестирования этих ограничений. Для целей этого упражнения обозначим за $\beta_1, \dots, \beta_{15}$ 15 параметров функции издержек, приведенной выше, в том порядке, в котором они входят в модель, начиная с первой строки и двигаясь слева направо и вниз.

- a. Выпишите матрицу \mathbf{R} и вектор \mathbf{q} в (5-8), которые требуются для наложения ограничений линейной однородности по ценам.
- b. «Протестируйте» теорию производства, используя все 158 наблюдений. Используйте F -тест для проверки ограничений линейной однородности. Заметьте, что вы можете использовать общую форму F -статистики в (5-16) для тестирования. Кристенсен и Грин наложили ограничения линейной однородности, включив их в модель. Вы можете сделать то же самое, поделив издержки и цену капитала и труда на цену топлива. Слагаемые с индексами f выпадают из модели, оставляя уравнение с 10 параметрами. Сравните суммы квадратов остатков двух моделей, чтобы провести тест. Конечно, тест может проводиться любым из способов и приведет к одному и тому же результату.
- c. Протестируйте гипотезу о гомотетичности производственной структуры в предположении о линейной однородности по ценам.
- d. Протестируйте гипотезу об обобщенной функции издержек Кобба–Дугласа из главы 4 против более общей транслогарифмической модели, предложенной здесь, снова (здесь и далее) предполагая линейную однородность по ценам.
- e. Простая функция Кобба–Дугласа записана в первой строке приведенной выше модели. Протестируйте гипотезу о модели Кобба–Дугласа против альтернативы о полной транслогарифмической модели.
- f. Протестируйте гипотезу обобщенной модели Кобба–Дугласа против гомотетичной транслогарифмической модели.
- g. Какую из предложенных функциональных форм вы считаете наиболее предпочтительной для этих данных?

3. Модель спроса на бензин, предложенная в пункте d эмпирического приложения 1 к главе 4, может быть записана как

$$\ln(G/Pop) = \alpha + \beta_P \ln P_g + \beta_I \ln (Income/Pop) + \gamma_{nc} \ln P_{nc} + \gamma_{uc} \ln P_{uc} + \gamma_{pt} \ln P_{pt} + \tau_{year} + \delta_d \ln P_d + \delta_n \ln P_n + \delta_s \ln P_s + \varepsilon.$$

- a. Протестируйте гипотезу о том, что агрегированные ценовые индексы не являются значимыми детерминантами спроса на бензин.
- b. Рассмотрим гипотезу о том, что микроэластичности являются фиксированными долями эластичностей по соответствующим показателям. Таким образом, для некоторого положительного θ (скорее всего, между 0 и 1) $\gamma_{nc} = \theta\delta_d$, $\gamma_{uc} = \theta\delta_d$, $\gamma_{pt} = \theta\delta_s$. Первые два ограничения обозначают простое линейное ограничение $\gamma_{nc} = \gamma_{uc}$. Взяв отношения, получаем, что первое (или второе) и третье ограничения

равносильны нелинейным ограничениям:

$$\frac{\gamma_{nc}}{\gamma_{pt}} = \frac{\delta_d}{\delta_s} \quad \text{или} \quad \gamma_{nc}\delta_s - \gamma_{pt}\delta_d = 0.$$

Опишите подробно, как бы вы тестировали обоснованность такого ограничения.

- c. Используя данные о рынке бензина из табл. F2.2, протестируйте два ограничения, предложенные здесь, отдельно и совместно.
4. *J*-тест в примере 5.7 проводится с использованием данных за более чем 50 лет. Будет оптимистично надеяться, что структура экономики не поменялась за 50 лет. Остается ли верным результат теста из примера 5.7, если провести его на данных только с 1980 по 2000 г.? Повторите вычисления на этом подмножестве данных.

ГЛАВА 6

Функциональная форма и структурный сдвиг

6.1. Введение

Этой главой мы завершим анализ модели линейной регрессии. Мы начнем с анализа различных видов функциональной формы регрессионной модели. Многие типы функций являются *линейными* в смысле определения из подраздела 2.3.1. При помощи различных преобразований зависимой и независимых переменных, бинарных переменных и различных комбинаций функций от переменных можно получить обширный класс моделей, которые все могут быть оценены с помощью линейного метода наименьших квадратов. В разделе 6.2 рассматривается использование бинарных переменных для учета нелинейности в модели. В разделе 6.3 мы расширим класс рассматриваемых моделей до моделей, линейных по параметрам. Используя логарифмы, квадратичные и перекрестные члены (произведения величин), можно рассмотреть большое число функциональных форм в рамках модели регрессии.

В разделе 6.4 о **структурном сдвиге** (structural change) рассматривается вопрос спецификации модели и тестирования наличия дискретного изменения в процессе порождения данных. В контексте временных рядов это относится к внезапным изменениям экономики, таким как важные события на финансовых рынках (например, мировой финансовый кризис 2007–2009 гг.) или изменениям на рынках ресурсов, таким как нефтяные шоки. В случае межобъектных данных регрессионная модель может быть модифицирована для учета дискретных различий между группами, таких как, например, различия в структуре предпочтений или в поведении мужчин и женщин на рынке труда.

6.2. Использование бинарных переменных

Одним из самых полезных приемов в регрессионном анализе является использование **бинарных переменных**, или **дамми-переменных** (**дамми**) (фиктивная переменная, binary variable, dummy variable, dummy). Дамми-переменная принимает значение один для некоторых наблюдений, указывая на наличие влияния или членства в группе, и значение ноль для остальных наблюдений. Бинарные переменные предоставляют удобную возможность учета дискретных сдвигов в регрессионной модели.

6.2.1. Бинарные переменные в регрессии

Фиктивные переменные обычно используются в уравнениях регрессии, включающих и другие количественные переменные. В уравнении доходов (earnings) в примере 5.2 мы включили переменную *Kids*, чтобы указать, были ли в домохозяйстве дети, предполагая, что этот факт оказывает важное влияние на поведение многих замужних женщин на рынке предложения труда. Оказывается, что результаты, полученные в примере 6.1, согласуются с этой гипотезой.

Пример 6.1. Фиктивная переменная в уравнении доходов

Таблица 6.1 воспроизводит оценку уравнения доходов, полученную в примере 5.2. Переменная *Kids* представляет собой фиктивную переменную, равную единице, если в домохозяйстве есть дети младше 18 лет, и нулю в противном случае. Поскольку это **полулогарифмическое уравнение**, значение коэффициента, равное $-0,35$, соответствует крайне сильному эффекту, который утверждает, что при прочих равных условиях доходы женщин с детьми примерно на треть меньше, чем женщин без детей. Это существенная разница, но она, безусловно, требует более детального рассмотрения. Вопрос о том, вызван ли этот результат особенностями рынка труда, влияющими на заработную плату, но не на количество отработанных часов, или же наоборот, остается открытым. Кроме того, поскольку мы рассматриваем неслучайную выборку, состоящую лишь из индивидов с положительными доходами, остается неясным, внес ли смещение в оценку коэффициента сам механизм формирования выборки.

Таблица 6.1. Оцененное уравнение доходов

$$\ln \text{earnings} = \beta_1 + \beta_2 \text{age} + \beta_3 \text{age}^2 + \beta_4 \text{education} + \beta_5 \text{kids} + \varepsilon$$

Сумма квадратов остатков: 599,4582

Стандартная ошибка регрессии: 1,19044

R^2 на 428 наблюдениях 0,040995

| Переменная | Коэффициент | Стандартная ошибка | t-статистика |
|------------------|-------------|--------------------|--------------|
| Константа | 3,24009 | 1,7674 | 1,833 |
| Age | 0,20056 | 0,08386 | 2,392 |
| Age ² | -0,0023147 | 0,00098688 | -2,345 |
| Education | 0,067472 | 0,025248 | 2,672 |
| Kids | -0,35119 | 0,14753 | -2,380 |

Дамми-переменные особенно полезны в логлинейных регрессиях. В модели вида

$$\ln y = \beta_1 + \beta_2 x + \beta_3 d + \varepsilon$$

коэффициент при дамми-переменной d указывает на мультипликативный сдвиг функции. Процентное изменение в $E[y|x,d]$, связанное с изменением в d , составляет

$$\% (\Delta E[y|x,d] / \Delta d) = 100\% \left\{ \frac{E[y|x,d=1] - E[y|x,d=0]}{E[y|x,d=0]} \right\} =$$

$$= 100\% \left\{ \frac{\exp(\beta_1 + \beta_2 x + \beta_3) E[\exp(\varepsilon)] - \exp(\beta_1 + \beta_2 x) E[\exp(\varepsilon)]}{\exp(\beta_1 + \beta_2 x) E[\exp(\varepsilon)]} \right\} =$$

$$= 100\% [\exp(\beta_3) - 1].$$

Пример 6.2. Ценность подписи

В примере 4.10 мы рассматривали взаимосвязь между логарифмом цены продажи и площадью поверхности для 430 продаж полотен Моне. Результаты оценки регрессий для этого примера включены в табл. 6.2. Эти результаты указывают на наличие сильной зависимости между площадью и ценой: значение коэффициента, равное 1,33372, указывает на высокоэластичную взаимосвязь, а t -статистика, равная 14,70, утверждает, что эта взаимосвязь статистически значима. Однако в модель не включена переменная (эффект) наличия подписи художника, влияющая на цену продажи. Из 430 сделок, рассмотренных в выборке, 77 представляли собой неподписанные картины. Результаты правой части табл. 6.2 включают фиктивную переменную наличия подписи. Результаты показывают крайне сильное влияние этой переменной. Из результатов оценки регрессии следует, что

$$E[\text{Price} | \text{Area}, \text{Aspect}, \text{Signature}] =$$

$$= \exp[-9,64 + 1,35 \ln \text{Area} - 0,08 \text{AspectRatio} + 1,28 \text{Signature} + 0,993^2/2].$$

Таблица 6.2. Оцененное уравнение для логарифма цены

$$\ln \text{price} = \beta_1 + \beta_2 \ln \text{Area} + \beta_3 \text{aspect ratio} + \beta_4 \text{signature} + \varepsilon$$

Значение $\ln \text{Price}$ 0,33274

Число наблюдений 430

| Сумма квадратов остатков | 519,17235 | 420,16787 | | | | |
|--------------------------------|-------------|--------------------|--------|-------------|--------------------|--------|
| Стандартная ошибка | 1,10266 | 0,99313 | | | | |
| R -квадрат | 0,33620 | 0,46279 | | | | |
| Скорректированный R -квадрат | 0,33309 | 0,45900 | | | | |
| Переменная | Коэффициент | Стандартная ошибка | t | Коэффициент | Стандартная ошибка | t |
| Константа | -8,42653 | 0,61183 | -13,77 | -9,64028 | 0,56422 | -17,09 |
| $\ln \text{Area}$ | 0,33372 | 0,09072 | 14,70 | 0,34935 | 0,08172 | 16,51 |
| Aspect ratio | -0,16537 | 0,12753 | -1,30 | -0,07857 | 0,11519 | -0,68 |
| Signature | 0,00000 | 0,00000 | 0,00 | 1,25541 | 0,12530 | 10,02 |

(См. раздел 4.6.) Сравнивая этот результат для картины такой же площади и такого же соотношения между сторонами, мы получаем, что согласно модели эффект от наличия подписи (signature) составляет

$$100\% \times (\Delta E[\text{Price}] / \text{Price}) = 100\% [\exp(1,26) - 1] = 252\%.$$

Наличие подписи увеличивает цену по сравнению с похожей картиной более чем вдвое. Оценка стандартной ошибки для этого коэффициента составляет 0,1253. С помощью дельта-метода мы можем получить оцененную стандартную ошибку $[\exp(b_3) - 1]$, которая равна квадратному корню из $[\exp(b_3)]^2 \times 0,1253^2$, т.е. 0,4417. Для процентной разницы в 252% мы получаем оценку стандартной ошибки, равную 44,17%.

На первый взгляд влияние размера картины, о котором мы говорили ранее, может быть объяснено наличием подписи. Если художник был склонен подписывать в основном более крупные картины, это объясняло бы противоречащее интуиции влияние размера. (Это являлось бы примером своего рода эффекта мультиколлинеарности.) Для регрессии, включающей и непрерывную переменную, и дамми, мы можем с легкостью подтвердить или опровергнуть это предложение. Средний размер для 77 сделок с неподписанными картинами составил 1 228,69 квадратного дюйма. Средний размер полотен для других 353 сделок был равен 878,016 квадратного дюйма. Действительно, обнаруживается существенная систематическая разница между размером подписанных и неподписанных картин, однако она обратна той, которую мы предполагали. Таким образом, мы обнаружили значимое влияние размера и подписи на аукционные цены картин Моне. *Aspect Ratio*, однако, представляется не оказывающим существенного влияния.

Для этой выборки характерна еще одна особенность, заслуживающая нашего внимания. В этих 430 сделках было продано лишь 387 различных картин. Некоторые сделки включали повторную продажу уже продававшейся картины. Предположение, что наблюдения являются независимыми реализациями случайной величины, нарушено, по крайней мере, для некоторых наблюдений. Такой вид «кластеризации» будет рассмотрен в главе 11 в ходе обсуждения панельных данных.

Исследователи часто включают фиктивную переменную в регрессию для того, чтобы учесть эффект, относящийся к единственному наблюдению. Например, при анализе временного ряда иногда включается фиктивная переменная, равная единице, только для единственного нетипичного года, например года крупной забастовки или года существенного изменения в государственной политике. (См. пример анализа функции спроса на деньги в Германии в подразделе 21.3.5.) Легко показать (мы рассмотрим это в качестве упражнения) крайне полезное следствие этого использования дамми-переменной.

Включение фиктивной переменной, принимающей значение один только для одного наблюдения, равносильно удалению этого наблюдения из выборки с точки зрения оценки коэффициентов методом наименьших квадратов и оценки дисперсии (но не с точки зрения R -квадрата).

6.2.2. Случай нескольких фиктивных переменных

Когда в данных присутствует несколько групп объектов, требуется включение целого набора бинарных переменных. Типичным примером такой ситуации является поправка на сезонные факторы при анализе макроэко-

номических данных. Мы можем записать функцию потребления для квартальных данных как

$$C_t = \beta_1 + \beta_2 x_t + \delta_1 D_{t1} + \delta_2 D_{t2} + \delta_3 D_{t3} + \varepsilon_t,$$

где x_t — это располагаемый доход. Отметим, что только три из четырех фиктивных переменных включаются в модель. Если бы была включена четвертая дамми, то сумма всех четырех фиктивных переменных была бы равна единице для каждого наблюдения, в точности повторяя константу, и мы бы имели случай совершенной мультиколлинеарности. Эта ситуация известна как **ловушка фиктивных переменных** (dummy variable trap). Чтобы ее избежать, мы исключаем дамми-переменную для четвертого квартала. (В зависимости от ситуации может оказаться удобнее включить четыре отдельные фиктивные переменные и исключить общую константу¹.) Любой из четырех кварталов (или 12 месяцев) может быть использован в качестве базового.

Это рассуждение представляет собой подход к *десезонализации* данных (или сезонной корректировке данных). Рассмотрим альтернативную формулировку:

$$C_t = \beta x_t + \delta_1 D_{t1} + \delta_2 D_{t2} + \delta_3 D_{t3} + \delta_4 D_{t4} + \varepsilon_t. \quad (6-1)$$

Используя результаты раздела 3.3 для частичной регрессии, мы можем сказать, что оценки этой множественной регрессии эквивалентны оценкам, получаемым в ходе следующей процедуры. Сначала мы оцениваем регрессии C и x на четыре фиктивные переменные, а затем используем остатки этих регрессий для последующей оценки регрессии очищенного от сезонности показателя потребления на очищенный от сезонности показатель дохода. Очевидно, что очистка от сезонности таким способом перед оценкой регрессии потребления на доход приводит к той же оценке коэффициента при доходе (и тому же вектору остатков), что и включение набора фиктивных переменных в регрессию.

Пример 6.3. Влияние жанра фильма на кассовые сборы кинофильмов

В примере 4.12 из табл. 4.8 представлены результаты оценки регрессии логарифма кассовых сборов для 62 фильмов, вышедших в прокат в 2009 г., на несколько переменных, в числе которых и набор дамми-переменных для определения жанра фильма: *Action*, *Comedy*, *Animated* или *Horror*. Оставшаяся категория включает «любой из оставшихся 9 жанров» в стандартном наборе из 13 жанров, который используется в подобных моделях. Четыре получившиеся оценки коэффициентов составили $-0,869$, $-0,016$, $-0,833$, $+0,375$ соответственно. Все это позволяет говорить о том, что, за исключением фильмов ужасов (*horror*), фильмы этих жанров обычно собирают в прокате существенно меньше денег, чем фильмы других видов. Отметим, что интерпретация коэффициента b непосредственно как оценки процентного изменения для категории, как мы поступили

¹Suits (1984) и Greene, Seaks (1991).

в примере 6.1, когда интерпретировали коэффициент $-0,35$ при *Kids* как указывающий на 35%-е изменение в доходе, — это аппроксимация, хорошо работающая в тех случаях, когда коэффициент b близок к нулю, но ухудшающаяся по мере удаления от нуля. Таким образом, значение $-0,869$, приведенное выше, не означает 87%-й разницы в сборах между фильмами жанра *Action* и другими. Используя формулу из примера 6.2, мы получаем разницу примерно в $[\exp(-0,869) - 1]$, или 58%.

6.2.3. Случай нескольких групп

Случай, при котором требуется несколько наборов фиктивных переменных, во многом напоминает уже рассмотренные. Однако есть одно важное исключение. Рассмотрим модель расходов штата на душу населения на образование y как функцию дохода штата на душу населения x . Предположим, что у нас есть наблюдения для всех $n = 50$ штатов за $T = 10$ лет. Регрессионная модель, которая позволяет ожидаемым расходам меняться как со временем, так и между штатами, будет выглядеть как

$$y_{it} = \alpha + \beta x_{it} + \delta_i + \theta_t + \varepsilon_{it}. \quad (6-2)$$

Как и прежде, необходимо исключить одну фиктивную переменную из каждого набора для избежания ловушки фиктивных переменных. В нашем примере, если включить 50 переменных для штатов и 10 переменных для периодов, все равно останется проблема «точной мультиколлинеарности»: сумма 50 переменных для штатов и 10 для периодов одинакова и равна 1. Одна из переменных в каждом наборе (или общая константа и одна из переменных в одном из наборов) должна быть исключена.

Пример 6.4. Анализ ковариации

Данные из таблицы приложения F6.1 были использованы при исследовании эффективности в предоставлении услуг авиаперелетов в работе Грина (Greene (2007a)). Отрасль авиалиний является одним из любимых объектов для анализа [см., например, работы Schmidt, Sickles (1984), Sickles, Good, Johnson (1986)] отчасти ввиду интереса к этому быстро меняющемуся в период дерегулирования рынку, отчасти ввиду доступности больших качественных массивов данных, собранных (уже не существующим) Комитетом гражданской авиации (Civil Aeronautics Board). Исходные данные состояли из 25 фирм, наблюдавшихся ежегодно на протяжении 15 лет (с 1970 по 1984) и составлявших, таким образом, «сбалансированную панель». Одни из этих фирм на протяжении этого периода подверглись слиянию, а в других проходили забастовки, это существенно сокращало число полных наблюдений. Исключение этих и других наблюдений с недостающими данными оставляет 10 полных наблюдений, из которых для нижеследующего примера мы выбрали шесть. Мы оценим уравнение издержек вида

$$\ln C_{i,t} = \beta_1 + \beta_2 \ln Q_{i,t} + \beta_3 \ln^2 Q_{i,t} + \beta_4 \ln P_{fuel i,t} + \beta_5 Loadfactor_{i,t} + \sum_{t=1}^{14} \theta_t D_t + \sum_{i=1}^5 \delta_i F_i + \varepsilon_{i,t}.$$

D_t и F_i — это дамми-переменные для годов и фирм соответственно. Мы исключили последние переменные в каждой группе. Оцениваемая модель для полной спецификации имеет вид

$$\begin{aligned} \ln C_{i,t} = & 13,56 + 0,8866 \ln Q_{i,t} + 0,01261 \ln^2 Q_{i,t} + 0,1281 \ln P_{fi,t} - \\ & - 0,8855 LF_{i,t} + \text{временные эффекты} + \\ & + \text{индивидуальные эффекты} + e_{i,t}. \end{aligned}$$

Временные эффекты (фиктивные переменные для лет) обнаруживают явную закономерность, как видно из рис. 6.1. Рассматривался период быстро растущих цен на топливо, так что можно ожидать отражения этого факта в издержках. Поскольку одна из дамми-переменных по годам исключена, эффект проиллюстрирован относительно базового года (1984).

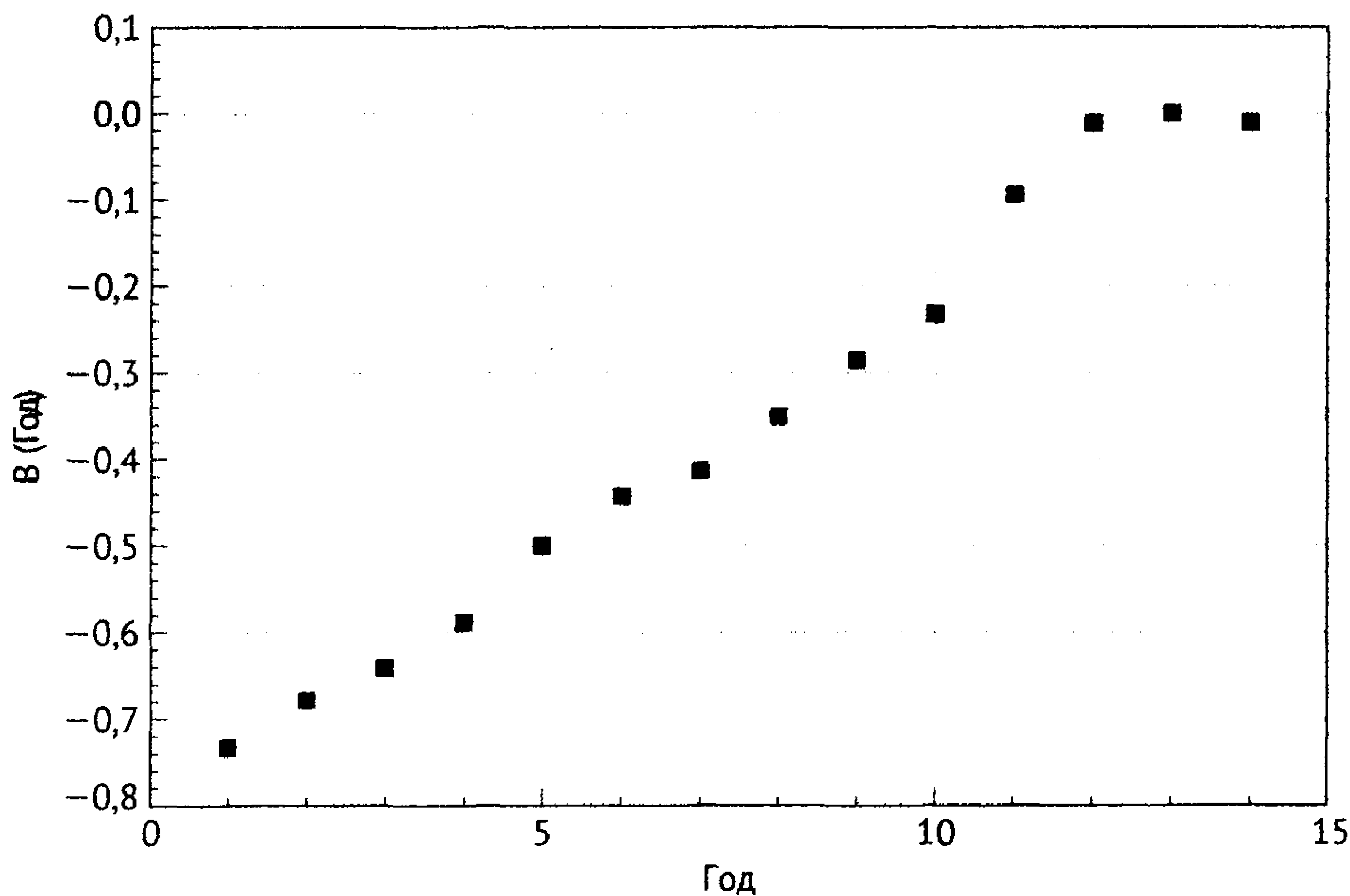


Рис. 6.1. Оценки коэффициентов при дамми-переменных, соответствующих году

Интерес представляет вопрос о том, значимы ли статистически индивидуальные эффекты (фиктивные переменные для фирм), временные эффекты, все эффекты или никакие из них. В табл. 6.3 представлены суммы квадратов остатков всех четырех регрессий. F -статистика для гипотезы об отсутствии индивидуальных эффектов составляет 65,94, что значительно превышает критическое значение. Статистика для временных эффектов составляет 2,61, что превышает критическое значение 1,84, но, пожалуй, в меньшей степени, чем можно было бы судить, глядя на

рис. 6.1. При отсутствии дамми-переменных для каждого года временной эффект, должно быть, в существенной степени отражается в эффекте влияния роста цены топлива.

Таблица 6.3. *F*-тесты на влияние фирмы и года

| Модель | Сумма квадратов | Ограничения | <i>F</i> | Ст. св. |
|-------------------------------|-----------------|-------------|----------|----------|
| Полная модель | 0,17257 | 0 | | |
| Только временные эффекты | 1,03470 | 5 | 65,94 | [5, 66] |
| Только индивидуальные эффекты | 0,26815 | 14 | 2,61 | [14, 66] |
| Нет эффектов | 1,27492 | 19 | 22,19 | [19, 66] |

6.2.4. Пороговые эффекты и индикаторные переменные

В большинстве приложений фиктивные переменные используются для того, чтобы учесть чисто качественные факторы, такие как членство в группе или учет конкретного временного периода. Однако в некоторых случаях фиктивные переменные представляют собой уровни некоторого фактора, который мог бы быть измерен непосредственно, если бы это было возможно. Например, образование — это как раз тот случай, когда мы обычно наблюдаем прохождение некоторых этапов, а не, скажем, годы образования. Предположим, например, что нас интересует регрессия вида

$$income = \beta_1 + \beta_2 age + effect\ of\ education + \varepsilon.$$

Данные об образовании могут состоять из наивысшего достигнутого уровня образования, такого как среднее образование (high school) (*HS*), бакалавр (undergraduate) (*B*), магистр (master) (*M*) или доктор наук (Ph.D) (*P*). Очевидно, неудовлетворительный подход здесь состоит в использовании переменной *E*, равной 0 для первой группы, 1 для второй, 2 для третьей и 3 для четвертой, т.е. рассмотрении уравнения в виде $income = \beta_1 + \beta_2 age + \beta_3 E + \varepsilon$. Проблема здесь состоит в том, что этот подход подразумевает один и тот же прирост дохода для каждого пройденного этапа: β_3 есть разница между доходами доктора наук и магистра, равно как магистра и бакалавра. Это маловероятно и излишне ограничивает регрессию. Более гибкая модель использовала бы три (или четыре) бинарные переменные, по одной для каждого уровня образования. Таким образом, можно было бы записать

$$income = \beta_1 + \beta_2 age + \delta_B B + \delta_M M + \delta_P P + \varepsilon.$$

Соотношение между коэффициентами и доходом для заданного возраста имеет вид

$$\text{Среднее образование: } E[income | age, HS] = \beta_1 + \beta_2 age,$$

$$\text{Бакалавр: } E[income | age, B] = \beta_1 + \beta_2 age + \delta_B,$$

$$\text{Магистр: } E[income | age, M] = \beta_1 + \beta_2 age + \delta_M,$$

$$\text{Доктор: } E[\text{income} | \text{age}, P] = \beta_1 + \beta_2 \text{age} + \delta_P.$$

Пусть наш интерес касается различий между, скажем, δ_P и δ_M , а также между δ_M и δ_B . Очевидно, их очень просто оценить. Альтернативным подходом может стать рассмотрение уравнения в другом виде, позволяющем выявить эти различия напрямую. Для этого требуется переопределить фиктивные переменные так, чтобы они были равны 1, если человеку присуждена соответствующая степень, даже если она для него не наивысшая. Так, для доктора наук все три бинарные переменные равны 1 и т.д. Определив переменные таким образом, получаем регрессию вида

$$\text{Средняя школа: } E[\text{income} | \text{age}, HS] = \beta_1 + \beta_2 \text{age},$$

$$\text{Бакалавр: } E[\text{income} | \text{age}, B] = \beta_1 + \beta_2 \text{age} + \delta_B,$$

$$\text{Магистр: } E[\text{income} | \text{age}, M] = \beta_1 + \beta_2 \text{age} + \delta_B + \delta_M,$$

$$\text{Доктор: } E[\text{income} | \text{age}, P] = \beta_1 + \beta_2 \text{age} + \delta_B + \delta_M + \delta_P.$$

Теперь вместо разницы между научной степенью и базовым уровнем в этой модели δ_P обозначает предельную ценность научной степени. Как именно специфицировать дамми-переменные в конкретном случае, является вопросом удобства. Все эти результаты могут быть получены из базового уравнения.

6.2.5. Эффекты воздействия и регрессия «разности разностей»

Исследователи многих научных направлений изучали эффект **воздействия** (treatment) на различного рода **реакции** (response). В качестве примеров можно привести влияние наличия университетского образования на доход на протяжении жизни [Dale, Krueger (2002)], влияние денежных переводов на здоровье детей [Gertler (2004)], влияние участия в программах повышения квалификации на доход [LaLonde (1986)] и сравнение состояний до и после изменения режима в макроэкономических моделях [Mankiw (2006)]. Эти примеры могут быть сформулированы в терминах модели регрессии, включающей единственную фиктивную переменную:

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + \delta D_i + \varepsilon_i,$$

где параметр сдвига δ измеряет влияние воздействия или изменения в политике (условно относительно \mathbf{x}) на объекты выборки. В простейшем случае при сравнении двух групп

$$y_i = \beta_1 + \beta_2 D_i + \varepsilon_i$$

мы получим $b_1 = (\bar{y} | D_i = 0)$, т.е. средний результат для тех, кто не испытал воздействия, и $b_2 = (\bar{y} | D_i = 1) - (\bar{y} | D_i = 0)$, разность средних между двумя группами. В исследовании Дэйл и Крюгера (Dale, Krueger (2002)) в модели сравнивались доходы студентов, обучавшихся в элитных университетах, и тех, кто там не обучался. Когда анализируется некоторое вмешательство, происходящее с течением времени, как в работе Крюгера (Krueger(1999)), в которой исследовался эксперимент STAR в Теннесси, в ходе которого показатели результативности школ наблюдали до и после введения правила,

вносящего изменение в число учеников в классе, фиктивной переменной для вмешательства служит индикатор периода D_t , равный 0 в период 1 и 1 в период 2. Величина β_2 измеряет изменение в реакции зависимой переменной, например эффективности школ, до и после вмешательства: $b_2 = \bar{y}_1 - \bar{y}_0$.

Предположение, что экспериментальная (подвергаемая воздействию) группа не меняется от периода 1 к периоду 2, делает данное сравнение менее убедительным. Усилить результат можно, включив в выборку группу **контрольных наблюдений** (control observations), которые не подвергаются воздействию. Изменение реакции в **экспериментальной группе** (treatment group) может быть затем соотнесено с изменением в **контрольной группе** (control group) в предположении, что это различие вызвано воздействием. Любопытное приложение эта стратегия часто находит в клинических испытаниях для медицинского вмешательства с учетом **эффекта плацебо**. Эффект плацебо — это спорный, но, по всей видимости, имеющий место исход некоторых клинических испытаний, в которых субъекты откликаются на воздействие, даже когда оно мнимо и, например, представляет собой сахарную или крахмальную таблетку при испытании лекарств. [См. работу Hróbjartsson, Gøtzsche, 2001.] Рассмотрение результатов подобных клинических испытаний проводится следующим образом. Субъекты, получающие плацебо, являются контрольной группой. Зависимая переменная (например, уровень холестерина) измеряется на базовом уровне для обеих групп. Экспериментальная группа получает лекарство, контрольная группа получает плацебо, и после этого зависимая переменная измеряется еще раз. Влияние измеряется как разность разностей:

$$E = [(\bar{y}_{exit}|treatment) - (\bar{y}_{baseline}|treatment)] - [(\bar{y}_{exit}|placebo) - (\bar{y}_{baseline}|placebo)].$$

Предположение состоит в том, что мера разности в разностях устойчива к эффекту плацебо, *если он имеет место*. Если же эффекта плацебо нет, то результат даже усиливается (предполагая, конечно, что результат вообще существует).

В общественных науках все более распространяются приложения, использующие модели эффекта воздействия с фиктивными переменными для оценки влияния дискретных изменений в политике². Одним из первых приложений было исследование Акта о развитии и подготовке рабочей силы (Manpower Development and Training Act, MDTA) в работе Ашенфельтера и Карда (Ashenfelter, Card (1985)). В простейшей форме эта модель включает наблюдение групп до и после вмешательства, где результирующая переменная — это переменная y , равная

$$y_{it} = \beta_1 + \beta_2 T_t + \beta_3 D_i + \beta_4 T_t \times D_i + \varepsilon_{it}. \quad t = 1, 2. \quad (6-3)$$

В этой модели T_t — это фиктивная переменная, равная нулю в период до воздействия и единице после него, и D_i равна единице для тех, кто был подвергнут «воздействию». Изменение реакции подвергшихся «воздействию» индивидов составит

²Обзоры литературы, посвященной эффектам воздействия, включая использование оценок «разности разностей», представлены в работах Imbens, Wooldridge (2009) и Millimet, Smith, Vytlačil (2008).

$$(y_{i2}|D_i = 1) - (y_{i1}|D_i = 1) = (\beta_1 + \beta_2 + \beta_3 + \beta_4) - (\beta_1 + \beta_3) = \beta_2 + \beta_4.$$

Для контрольной группы это будет равно

$$(y_{i2}|D_i = 0) - (y_{i1}|D_i = 0) = (\beta_1 + \beta_2) - (\beta_1) = \beta_2.$$

«Разность разностей» (difference in differences) запишется как

$$[(y_{i2}|D_i = 1) - (y_{i1}|D_i = 1)] - [(y_{i2}|D_i = 0) - (y_{i1}|D_i = 0)] = \beta_4.$$

В множественной регрессии y_{it} на константу, T , D и TD оценка методом наименьших квадратов для β_4 будет равна разности в изменениях средних:

$$\begin{aligned} b_4 &= (\bar{y}|D = 1, \text{ период } 2) - (\bar{y}|D = 1, \text{ период } 1) - \\ &\quad - (\bar{y}|D = 0, \text{ период } 2) - (\bar{y}|D = 0, \text{ период } 1) = \\ &= \Delta\bar{y}|_{\text{экспериментальная}} - \Delta\bar{y}|_{\text{контрольная}}. \end{aligned}$$

Эта регрессия называется оценкой «разности разностей», имея в виду этот результат.

Когда воздействие является результатом изменения в политике или события, которое происходит полностью вне контекста исследования, анализ обычно называется **естественным экспериментом** (natural experiment). Исследование Карда (Card (1990)), посвященное масштабной иммиграции в Майами в 1979 г. и обсуждаемое в примере 6.5, служит примером такого эксперимента.

Пример 6.5. *Естественный эксперимент: «лодочный мост» из Мэриэла*
Резкое изменение в политике может представлять естественный эксперимент. Пример, рассматриваемый в работе Карда (Card (1990)), посвящен «лодочному мосту» из города Мэриэль на Кубе в Майами (май–сентябрь 1980 г.), который увеличил предложение труда в Майами на 7%. Автор рассматривает влияние такого резкого изменения на рынке труда на заработную плату и занятость местного населения. В модели Майами сравнивается с сопоставимым городом — с Лос-Анджелесом. Пусть i обозначает индивида, а D — воздействие, которое для индивида будет эквивалентно «проживанию в городе, который подвергся иммиграции». Для индивидов в Майами или Лос-Анджелесе результирующая переменная имеет вид

$$(Y_i) = 1 \text{ для безработных, } 0 \text{ — для трудоустроенных.}$$

Пусть c обозначает город, а t — период до (1979) или после (1981) иммиграции. Тогда уровень безработицы в городе c в момент t равен $E[y_{i,0}|c, t]$, если иммиграции нет, и $E[y_{i,1}|c, t]$, если она наблюдается. Эти уровни предполагаются постоянными. В этом случае

$$\beta E[y_{i,0}|c, t] = \beta_t + \gamma_c \text{ без иммиграции,}$$

$$E[y_{i,1}|c, t] = \beta_t + \gamma_c + \delta \text{ с иммиграцией.}$$

Эффект иммиграции, оказываемый на уровень безработицы, измеряется как δ . Суть естественного эксперимента состоит в том, что

иммиграция происходит в Майами, но не происходит в Лос-Анджелесе и не является результатом действий жителей этих городов. Тогда

$$E[y_i|M, 79] = \beta_{79} + \gamma_M \text{ и } E[y_i|M, 81] = \beta_{81} + \gamma_M + \delta \text{ для Майами,}$$

$$E[y_i|L, 79] = \beta_{79} + \gamma_L \text{ и } E[y_i|L, 81] = \beta_{81} + \gamma_L \text{ для Лос – Анджелеса.}$$

Предполагается, что рост безработицы в обоих городах был бы постоянным при отсутствии иммиграции. Если ни один из городов не испытывал бы иммиграции, изменение в уровне безработицы составило бы

$$E[y_{i,0}|M, 81] - E[y_{i,0}|M, 79] = \beta_{81} - \beta_{79} \text{ для Майами,}$$

$$E[y_{i,0}|L, 81] - E[y_{i,0}|L, 79] = \beta_{81} - \beta_{79} \text{ для Лос – Анджелеса.}$$

Если бы миграция затронула оба города, то

$$E[y_{i,1}|M, 81] - E[y_{i,1}|M, 79] = \beta_{81} - \beta_{79} + \delta \text{ для Майами,}$$

$$E[y_{i,1}|L, 81] - E[y_{i,1}|L, 79] = \beta_{81} - \beta_{79} + \delta \text{ для Лос – Анджелеса.}$$

Миграция («воздействие») произошла только в Майами. Разность между разностями, которая определяет результат этого эксперимента, равна

$$\{E[y_{i,1}|M, 81] - E[y_{i,1}|M, 79]\} - \{E[y_{i,0}|L, 81] - E[y_{i,0}|L, 79]\} = \delta.$$

Автор анализировал изменения в уровнях занятости и заработной платы в двух городах на протяжении нескольких лет после строительства «лодочного моста». Эффект был удивительно мал, учитывая масштаб эксперимента в Майами.

Одной из важных проблем при анализе политики является измерение эффекта воздействия, при котором фиктивная переменная выступает результатом решения индивида. В примере клинического исследования, о котором говорилось выше, представители контрольной группы (по предположению) не знают о том, что они находятся в контрольной группе. Распределение воздействия экзогенно по отношению к эксперименту. В исследовании Крюгера и Дэйла (Krueger, Dale (2002)), напротив, включение в экспериментальную группу студентов элитного колледжа полностью добровольно и определяется индивидом. Ключевым моментом для анализа в таком случае служит принятие во внимание вполне допустимой ситуации, когда «фиктивная переменная воздействия» отражает скрытую мотивацию и инициативу участников, а не влияние самой программы. В этом и состоит основное преимущество естественных экспериментов: они наиболее точно (возможно, в полной мере) воспроизводят экзогенное распределение воздействия из клинических испытаний³. Некоторые подобные случаи будут рассмотрены нами в главах 8 и 19.

6.3. Нелинейность в переменных

В настоящий момент будет полезно записать линейную регрессионную модель в очень общей форме. Пусть $\mathbf{z} = z_1, z_2, \dots, z_L$ есть L независимых переменных, f_1, f_2, \dots, f_K — K линейно независимых функций от \mathbf{z} , $g(y)$ — на-

³Обсуждение этого подхода см. в работах Angrist, Krueger (2001) и Angrist, Pischke (2010).

блюдаемая функция от y . Пусть также выполнены все наши обычные предположения относительно шоков. Модель линейной регрессии может быть записана как

$$\begin{aligned} g(y) &= \beta_1 f_1(\mathbf{z}) + \beta_2 f_2(\mathbf{z}) + \dots + \beta_K f_K(\mathbf{z}) + \varepsilon \\ &= \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_K x_K + \varepsilon = \mathbf{x}'\boldsymbol{\beta} + \varepsilon. \end{aligned} \quad (6-4)$$

Используя логарифмы, экспоненты, обратные величины, трансцендентные функции, многочлены, произведения, дроби и т.д., эта «линейная» модель может быть оценена в большом числе случаев.

6.3.1. Кусочно-линейная регрессия

При рассмотрении данных о доходах в большой выборке индивидов различных возрастов ясно прослеживаются некоторые закономерности, касающиеся пороговых возрастов. В частности, доход растет с возрастом на всем рассматриваемом диапазоне, но темп роста может меняться в определенные моменты, такие, например, как 18 лет, когда типичный индивид заканчивает среднюю школу, и 22 года, когда он или она заканчивает университет. **Зависимость дохода от возраста** (time profile of income) может напоминать график, представленный на рис. 6.2. Основываясь на предшествующем обсуждении, мы можем оценить такую регрессионную модель, просто разделив выборку на три подвыборки. Однако такой подход игнорирует непрерывность предполагаемой зависимости. Результат будет больше похож на точечный график, чем на непрерывную функцию, которую мы ожидаем увидеть. Регрессия с ограничениями и то, что обычно называется **сплайн** (spline)-функцией, могут быть использованы для достижения желаемого результата⁴.

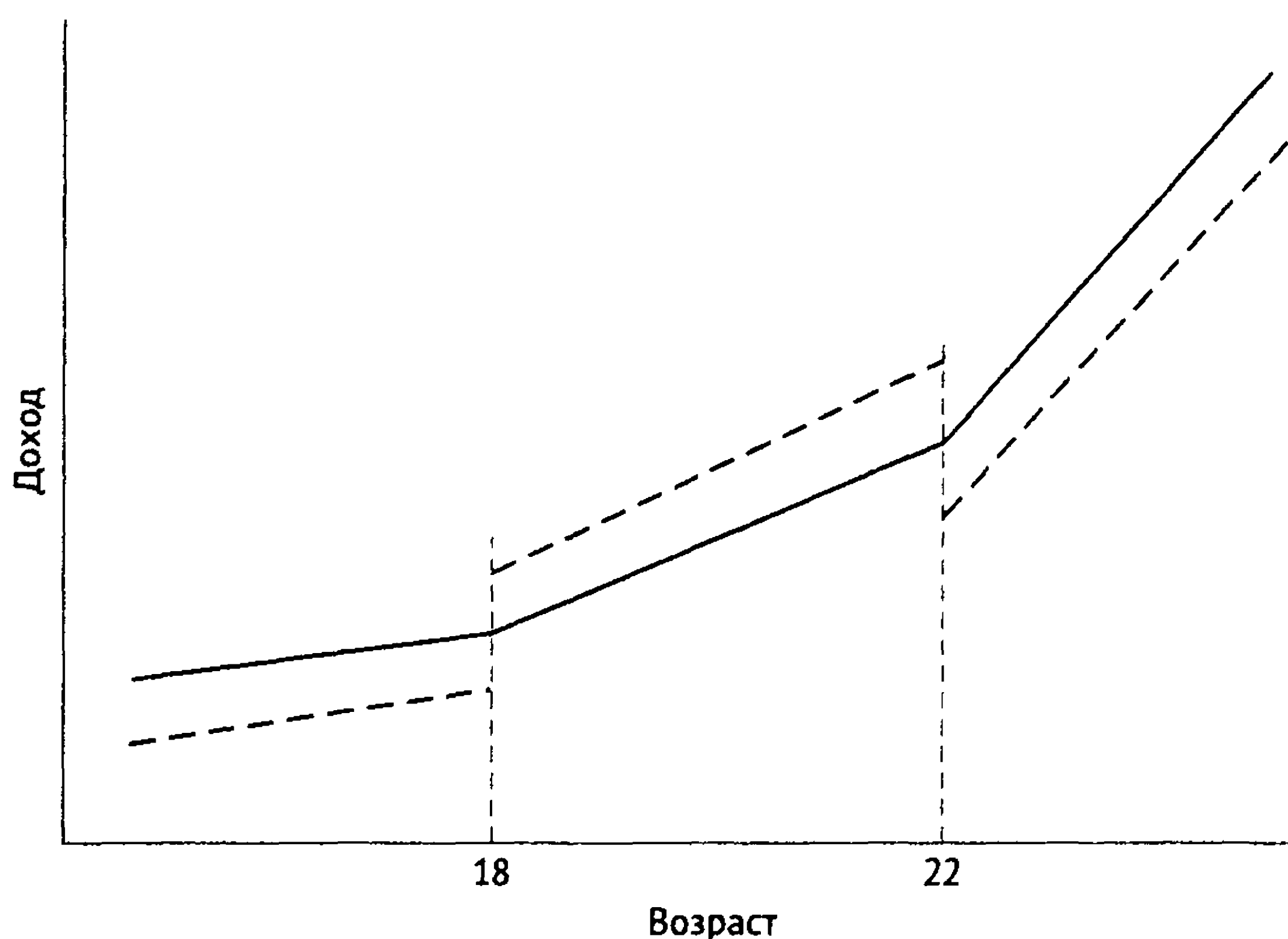


Рис. 6.2. Сплайн-функция

⁴Этому вопросу посвящена работа Пуарье (Poirier (1974)). Часто цитируемый пример приведен в работе Гарбера и Пуарье (Garber, Poirier (1974)).

Функция, которую мы хотим оценить, имеет вид

$$E[income | age] = \begin{cases} \alpha^0 + \beta^0 age, & \text{если } age < 18, \\ \alpha^1 + \beta^1 age, & \text{если } age \geq 18 \text{ и } age < 22, \\ \alpha^2 + \beta^2 age, & \text{если } age \geq 22. \end{cases}$$

Граничные значения 18 и 22 называются **узлами** (knots). Пусть

$$d_1 = 1, \quad \text{если } age \geq t_1^*, \quad d_2 = 1, \quad \text{если } age \geq t_2^*,$$

где $t_1^* = 18$ и $t_2^* = 22$. Для объединения всех трех уравнений мы используем

$$income = \beta_1 + \beta_2 age + \gamma_1 d_1 + \delta_1 d_1 age + \gamma_2 d_2 + \delta_2 d_2 age + \varepsilon.$$

Эта зависимость отражена пунктиром на рис. 6.2. Коэффициенты наклона у трех сегментов равны β_2 , $\beta_2 + \delta_1$ и $\beta_2 + \delta_1 + \delta_2$. Для того чтобы сделать функцию **кусочно-непрерывной** (piecewise continuous), мы потребуем, чтобы сегменты соединялись в узлах, т.е.

$$\beta_1 + \beta_2 t_1^* = (\beta_1 + \gamma_1) + (\beta_2 + \delta_1) t_1^*$$

и

$$(\beta_1 + \gamma_1) + (\beta_2 + \delta_1) t_2^* = (\beta_1 + \gamma_1 + \gamma_2) + (\beta_2 + \delta_1 + \delta_2) t_2^*.$$

Это линейные ограничения на коэффициенты. Приводя подобные члены, первое ограничение запишем как

$$\gamma_1 + \delta_1 t_1^* = 0 \quad \text{или} \quad \gamma_1 = -\delta_1 t_1^*.$$

Повторяя ту же операцию для второго ограничения и подставляя их в (6-3), мы получаем

$$income = \beta_1 + \beta_2 age + \delta_1 d_1 (age - t_1^*) + \delta_2 d_2 (age - t_2^*) + \varepsilon.$$

Оценки методом наименьших квадратов при ограничениях могут быть получены при помощи множественной регрессии, используя константу и переменные:

$$\begin{aligned} x_1 &= age, \\ x_2 &= age - 18, \quad \text{если } age \geq 18, \text{ и } 0 \text{ иначе} \end{aligned}$$

и

$$x_3 = age - 22, \quad \text{если } age \geq 22, \text{ и } 0 \text{ иначе.}$$

Мы можем проверить гипотезу о том, что наклон функции постоянен, проверяя совместно ограничения $\delta_1 = 0$ и $\delta_2 = 0$.

6.3.2. Функциональные формы

Часто используемой формой регрессионной модели является **логлинейная модель** (loglinear model):

$$\ln y = \ln \alpha + \sum_k \beta_k \ln X_k + \varepsilon = \beta_1 + \sum_k \beta_k x_k + \varepsilon.$$

В этой модели коэффициенты являются эластичностями:

$$\left(\frac{\partial y}{\partial X_k}\right)\left(\frac{X_k}{y}\right) = \frac{\partial \ln y}{\partial \ln X_k} = \beta_k. \quad (6-5)$$

В логлинейном уравнении изменения измеряются в терминах пропорций или процентов: β_k измеряет процентное изменение в y , связанное с однопроцентным изменением в X_k . Это позволяет исключить из рассмотрения единицы измерений при использовании регрессионной модели. Иногда применяется альтернативный подход, состоящий в измерении переменных и соответствующих изменений в стандартных отклонениях. Если данные «» перед оцениванием с помощью формулы $x_{ik}^* = (x_{ik} - \bar{x}_k)/s_k$ (и аналогично для y), то коэффициенты регрессии, полученные методом наименьших квадратов, измеряют изменения в единицах стандартных отклонений, а не в естественных единицах или процентах. (Обратите внимание, что из регрессии при этом пропадает константа.) Для того чтобы получить эти результаты, необязательно стандартизировать данные: умножение каждого коэффициента b_k в исходной регрессии на s_k/s_y приводит к тому же результату.

Гибридом линейной и логлинейной моделей является полулогарифмическое уравнение

$$\ln y = \beta_1 + \beta_2 x + \varepsilon. \quad (6-6)$$

Мы использовали такую форму для уравнения инвестиций в подразделе 5.2.2:

$$\ln I_t = \beta_1 + \beta_2 (i_t - \Delta p_t) + \beta_3 \Delta p_t + \beta_4 \ln Y_t + \beta_5 t + \varepsilon_t,$$

где логарифм инвестиций моделировался через уровень реальной процентной ставки, уровень цен и временной тренд. В таком полулогарифмическом уравнении с временным трендом $d \ln I / dt = \beta_5$ является средним темпом роста показателя I . Оцененное значение, равное $-0,00566$ (см. табл. 5.2), говорит о том, что на всем рассматриваемом периоде, после учета всех других факторов, средний темп роста инвестиций составил $-0,566\%$ в год.

Коэффициенты в полулогарифмической модели называются частными эластичностями или полуэластичностями: в (6-6) коэффициент β_2 равен $\partial \ln y / \partial x$. Эта форма естественна для моделей с фиктивными переменными, такими как уравнение доходов в примере 5.2. Коэффициент при переменной $Kids$, равный $-0,35$, говорит о том, что при прочих равных условиях доходы приблизительно на 35% ниже в домохозяйствах с детьми.

Квадратичное уравнение заработков в примере 6.1 представляет другой пример использования нелинейностей в переменных. Используя результаты примера 6.1, мы выяснили, что для женщин с 12 годами образования и имеющих детей в домохозяйстве график зависимости доходов от возраста выглядит так, как показано на рис. 6.3. Этот график вызывает вопрос при анализе доходов индивидов. Глядя на рис. 6.3, возникает соблазн заключить, что он демонстрирует траекторию заработков индивида в зависимости от его возраста. Однако ответ на этот вопрос не содержится в наших данных. Модель основана на межобъектных данных и демонстрирует заработки людей в разном возрасте. Рассуждение о том, каким образом этот график связан с ожидаемой траекторией доходов индивида, представляет собой отдельный, причем очень непростой вопрос.

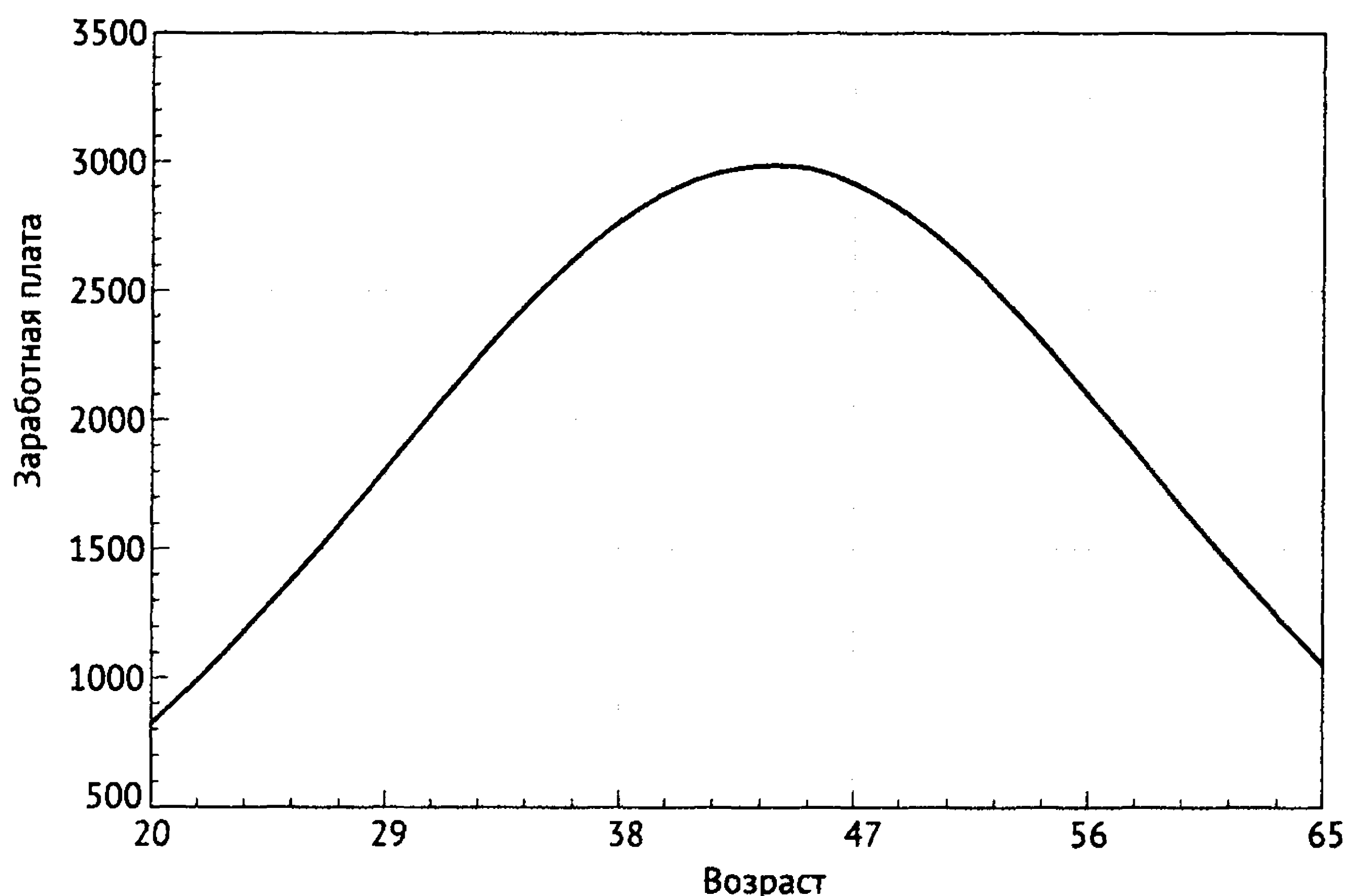


Рис. 6.3. График зависимости заработной платы от возраста

6.3.3. Эффект взаимодействия

Другой полезной формой регрессионной модели является модель с **компонентами взаимодействия**. Например, модель, связывающая тормозной путь D со скоростью S и влажностью дорожного покрытия W , может выглядеть так:

$$D = \beta_1 + \beta_2 S + \beta_3 W + \beta_4 SW + \varepsilon.$$

В этой модели

$$\frac{\partial E[D | S, W]}{\partial S} = \beta_2 + \beta_4 W,$$

откуда следует, что **предельный эффект** (marginal effect) от наличия более высокой скорости на тормозной путь увеличивается, когда дорожное покрытие влажное (предполагая, что коэффициент β_4 положителен). Если требуется построить доверительные интервалы или тестировать гипотезы относительно этих предельных эффектов, то необходимая стандартная ошибка может быть оценена из

$$\text{Var}\left(\frac{\partial \hat{E}[D | S, W]}{\partial S}\right) = \text{Var}[\hat{\beta}_2] + W^2 \text{Var}[\hat{\beta}_4] + 2W \text{Cov}[\hat{\beta}_2, \hat{\beta}_4],$$

а также аналогичным образом для $\partial E[D | S, W] / \partial W$. Требуется подставить некоторое значение вместо W . Естественным значением будет выборочное среднее, но в некоторых случаях можно предпочесть и другое — такое, как, например, максимальное значение W в выборке.

6.3.4. Выявление нелинейности

Если функциональная форма неизвестна заранее, то некоторые подходы могут помочь выявить любую нелинейность, а также предоставить о ней некоторую информацию на основе имеющихся данных. Например, если предполагаемая нелинейность относится к единственному регрессору в уравнении, то оценка квадратичного или кубического многочлена вместо линейной функции может учесть некоторую часть этой нелинейности. Выбрав несколько интервалов для рассматриваемого регрессора и получив различные оценки параметров на каждом из них, можно оценить кусочно-линейную аппроксимацию нелинейной функции.

Пример 6.6. Функциональная форма нелинейной функции издержек

В знаменитом исследовании экономии от масштаба в отрасли электроэнергетики США (Nerlove (1963)) автор проанализировал производственные издержки 145 американских генерирующих компаний. Это исследование привнесло несколько новшеств в микроэконометрику. Оно представляло собой одно из первых существенных изучений статистического анализа функции издержек. Теоретические выкладки в исследовании Нерлова впервые показали, как фундаментальная теория двойственности между функцией издержек и производственной функцией может быть использована для формулировки эконометрической модели. Наконец, Нерлов использовал несколько полезных приемов для уточнения своей базовой модели.

Основное внимание в статье направлено на изучение эффекта экономии от масштаба, обычно моделируемого как характеристика производственной функции. Автор выбрал функцию Кобба–Дугласа для моделирования выпуска как функции капитала (K), труда (L) и топлива (F):

$$Q = \alpha_0 K^{\alpha_K} L^{\alpha_L} F^{\alpha_F} e^{\varepsilon_i},$$

где Q — это выпуск, а ε_i отражает неизмеряемые различия между фирмами. Параметр экономии от масштаба — это $r = \alpha_K + \alpha_L + \alpha_F$. Значение параметра, равное 1, соответствует постоянной отдаче от масштаба. В своей работе Нерлов изучал широко распространенное предположение, что производители в этой отрасли получают существенную экономию от масштаба. Модель производства является логлинейной, поэтому, предполагая, что другие условия классической регрессионной модели выполняются, четыре параметра могут быть оценены с помощью метода наименьших квадратов. Однако автор отметил, что три фактора модели не могут рассматриваться в качестве экзогенных переменных. Для фирмы, которая оптимизирует издержки по факторам производства, спрос на топливо выглядел бы как $F^* = F^*(Q, P_K, P_L, P_F)$ (аналогичное верно и для труда и капитала), так что, конечно, предположения классической модели нарушены.

В действовавшей на тот момент системе регулирования государственные комиссии устанавливали цены, и фирмы удовлетворяли спрос, предъявляемый по этим регулируемым ценам. Таким образом, утверждалось, что

выпуск (так же как и цены факторов производства) экзогенен для фирмы, и, основываясь на позиции, обсуждаемой в работе Зельнера, Кменты и Дриза (Zellner, Kmenta, Dreze (1966)), Нерлов утверждал, что при равновесии отклонение издержек от долгосрочного оптимума не будет зависеть от выпуска. (Из этого факта выводится следствие, которое можно проверить эмпирически, и мы обсудим его в подразделе 19.2.4.) Таким образом, цель фирмы представляла собой минимизацию издержек при ограничении на производственную функцию. Тогда можно записать лагранжиан:

$$\text{Min}_{K,L,F} P_K K + P_L L + P_F F + \lambda(Q - \alpha_0 K^{\alpha_K} L^{\alpha_L} F^{\alpha_F}).$$

Решением этой минимизационной задачи являются три характеристики спроса на факторы и множитель (который измеряет предельные издержки). Подставляя решение обратно в формулу для общих издержек, получаем (в сущности линейную) логлинейную функцию издержек:

$$P_K K + P_L L + P_F F = C(Q, P_K, P_L, P_F) = r A Q^{1/r} P_K^{\alpha_K/r} P_L^{\alpha_L/r} P_F^{\alpha_F/r} e^{\varepsilon_i/r}$$

или

$$\ln C = \beta_1 + \beta_q \ln Q + \beta_K \ln P_K + \beta_L \ln P_L + \beta_F \ln P_F + u_i, \quad (6-7)$$

где $\beta_q = 1/(\alpha_K + \alpha_L + \alpha_F)$ — интересующий нас параметр, а $\beta_j = \alpha_j/r$, $j = K, L, F$. Таким образом, двойственность между функцией издержек и производственной функцией была использована для вывода из теоретических основ оцениваемого уравнения.

Осталась одна деталь. Параметры функции издержек должны в сумме давать единицу: $\beta_K + \beta_L + \beta_F = 1$, поэтому оценка должна производиться с учетом этого ограничения⁵. Это ограничение может быть учтено, если рассматривать регрессию $\ln(C/P_F)$ на константу, $\ln Q$, $\ln(P_K/P_F)$ и $\ln(P_L/P_F)$. Эта первая часть результатов представлена в верхней части табл. 6.4⁶.

⁵В контексте эконометрической модели из этого ограничения можно вывести следствие, которое можно протестировать, следуя определению, данному в главе 5. Но есть экономические основания требовать выполнения этого ограничения: оно использовалось при выводе функции издержек. В связи с этим неясно, какой смысл может иметь тестирование этого ограничения. На первый взгляд, если гипотеза о выполнении этого ограничения отвергается, наше исследование в этом месте должно быть прекращено, поскольку без него функция издержек не является допустимым представлением производственной функции. Мы еще раз столкнемся с этой дилеммой в другом виде в главе 10. К счастью, в рассматриваемом случае гипотеза не отвергается. (Она отвергается в примере в главе 10.)

⁶Читатели, желающие воспроизвести исследование Нерлова, должны отметить, что он использовал в своих вычислениях логарифм по основанию 10, а не натуральный логарифм. Практический совет: для преобразования натурального логарифма в логарифм по основанию 10 разделите первый на $\log_e 10 = 2,302585093$. Кроме того, хотя первые 145 строк данных табл. F6.2 приложения полностью совпадают с данными исходного исследования, только первая регрессия из приведенных в табл. 6.4 может быть воспроизведена с использованием этих данных. Результаты для групп 1–5 в таблице были пересчитаны заново и не совпадают с результатами оригинальной работы Нерлова (Nerlove (1963)). Аналогичным образом были пересчитаны результаты табл. 6.5, которые также не совпали с результатами исходного исследования.

Таблица 6.4. Функция издержек Кобба–Дугласа (в скобках указаны стандартные ошибки)

| | $\log Q$ | $\log P_L - \log P_F$ | $\log P_K - \log P_F$ | R^2 |
|-----------|-------------------|-----------------------|-----------------------|-------|
| Все фирмы | 0,721 (0,0174) | 0,593 (0,205) | -0,0085 (0,191) | 0,932 |
| Группа 1 | 0,400 | 0,615 | -0,081 | 0,513 |
| Группа 2 | 0,658 | 0,094 | 0,378 | 0,633 |
| Группа 3 | 0,938 | 0,402 | 0,250 | 0,573 |
| Группа 4 | 0,912 | 0,507 | 0,093 | 0,826 |
| Группа 5 | 1,044 | 0,603 | -0,289 | 0,921 |

Начальные оценки параметров функции издержек указаны в верхнем ряду табл. 6.4. Гипотеза о постоянной отдаче от масштаба может быть уверенно отвергнута. t -статистика составляет $(0,721 - 1)/0,0174 = -16,03$, и поэтому можно сделать вывод о том, что эта оценка значимо меньше единицы или, другими словами, r значимо превышает 1. Отметим, что коэффициент при цене капитала отрицателен. Согласно теории он должен быть равен величине α_K/r , которая (если только предельный продукт капитала неотрицателен) должна быть положительна. Автор рассматриваемого исследования объясняет это ошибкой измерения в переменной цены капитала. Аргумент кажется убедительным, но из него следует, что остальные коэффициенты также измерены плохо. [Метод, используемый в работе Кристенсена и Грина (Christensen and Greene (1976)), дал положительную оценку этой величины на тех же данных. См. подраздел 10.5.2.]

Ярко выраженная структура остатков, изображенных на рис. 6.4, а также некоторые наши рассуждения о предполагаемой форме производственной функции указывают на то, что в модели пропущены какие-то факторы⁷. Согласно теории, оцененная модель должна демонстрировать непрерывно убывающую кривую средних издержек, которая, в свою очередь, указывает на устойчивую экономию от масштаба при любом уровне выпуска. Это противоречит указанным соображениям об U-образной кривой средних издержек и представляется противоречащим данным. Обратите внимание на три кластера остатков на графике. Два подхода были использованы для расширения этой модели.

При делении выборки на пять групп по 29 фирм на основании выпуска и оценке отдельных регрессий для каждой группы в работе Нерлова (Nerlove (1963)) была оценена кусочная логлинейная модель. Результаты представлены в нижних строках табл. 6.4, причем в каждой следующей группе находятся все более крупные фирмы. Результаты убедительно показывают, что (лог)линейная спецификация функции издержек не является подходящей.

⁷Тест Дарбина–Уотсона на корреляцию в остатках (см. раздел 20.7) указывает автору на существенную автокорреляцию. Обычно используемые при анализе временных рядов статистика Дарбина–Уотсона и тесты на «автокорреляцию» могут послужить полезными инструментами при определении подходящей функциональной формы в модели для межобъектных данных. Для использования этого подхода необходимо отсортировать наблюдения относительно интересующей нас переменной (выпуска). Наличие кластеров остатков одного знака указывает на необходимость пересмотра предполагаемой функциональной формы.

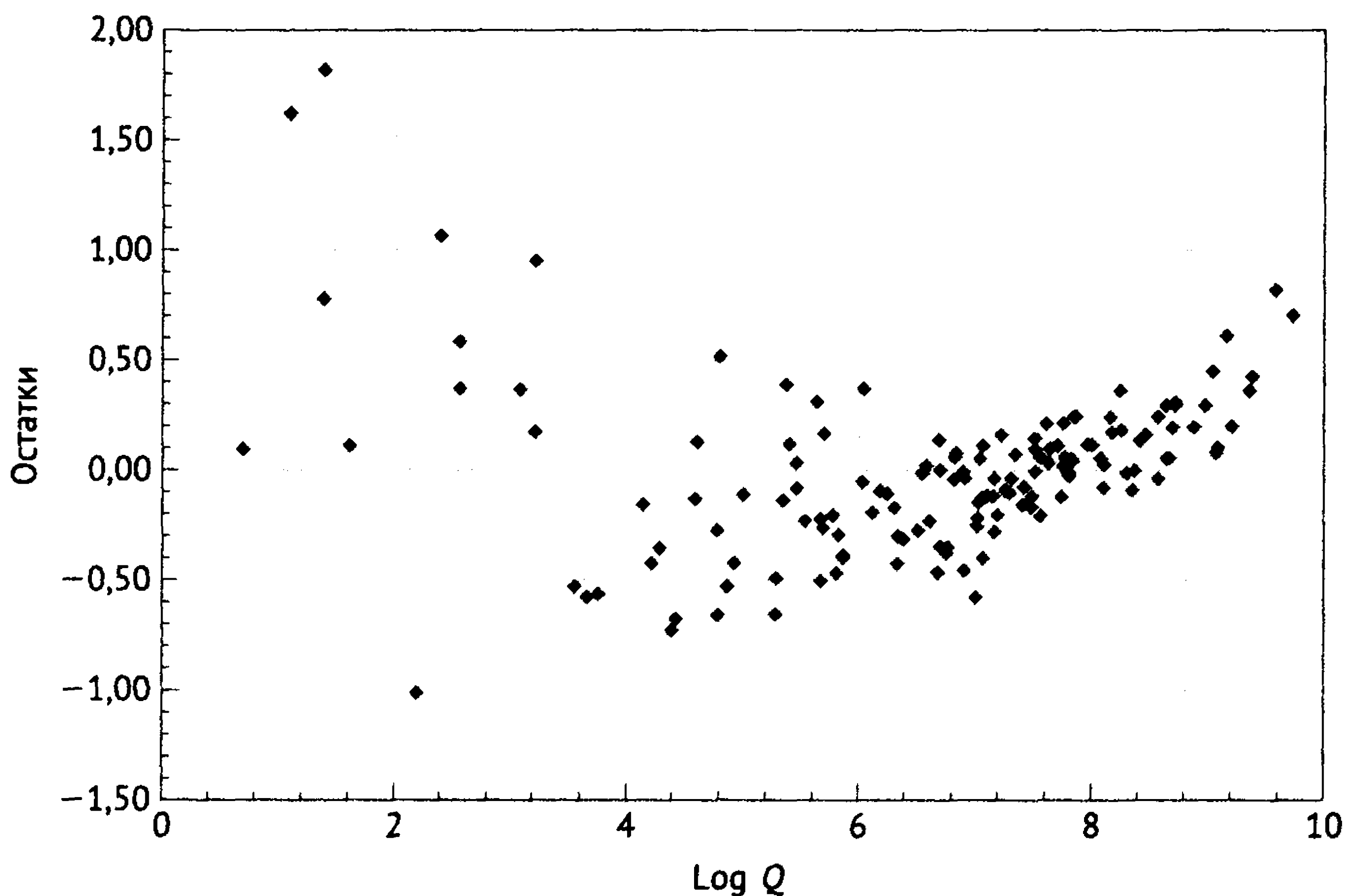


Рис. 6.4. Остатки от предсказанной функции издержек

Коэффициент при выпуске, который растет и затем превышает 1, соответствует U -образной функции издержек, как и предполагалось ранее.

Второй подход состоит в расширении функции издержек и включении в нее квадратичного члена логарифма выпуска. Этот подход соответствует значительно более общей модели и приводит к результатам, указанным в табл. 6.5. Снова t -тест уверенно демонстрирует, что повышение общности модели — это шаг в правильном направлении: $t = 0,051/0,00054 = 9,44$. Эластичность выпуска в этой квадратичной модели равна $\beta_q + 2\gamma_{qq} \log Q$ ⁸. Отдача от масштаба возрастает, когда это значение меньше 1, и постоянна, когда оно равно 1. Используя два значения из таблицы (0,152 и 0,0052 соответственно), можно определить, что эта функция и в самом деле приводит к U -образной кривой средних издержек с минимумом в точке $\ln Q = (1 - 0,152)/(2 \times 0,051) = 8,31$, или $Q = 4079$, что находится примерно посередине диапазона выпуска для доступной в исследовании выборки фирм.

Это исследование было дополнено в работе Кристенсена и Грина (Christensen, Greene (1976)). Используя те же данные, но более сложную (транслогарифмическую) функциональную форму и оценивая спрос на факторы одновременно с функцией издержек, авторы пришли к выводам, в целом соответствующим результатам работы Нерлова (Nerlove (1963)). Выбранная ими функциональная форма указала на то, что обобщенная модель в работе Нерлова (Nerlove (1963)), приведенная в табл. 6.5, несколько недооценила диапазон выпуска, для которого издержки производства единицы продукции продолжали бы убывать. Авторы также повторили исследование с вы-

⁸Нерлов некорректно измерял экономию от масштаба для этой функции как $1/(\beta_q + \delta \log Q)$, где β_q и δ — это коэффициенты при логарифме Q и $\log^2 Q$. Верное выражение выглядит как $1/[\partial \log C / \partial \log Q] = 1/[\beta_q + 2\delta \log Q]$. Эта неточность периодически обнаруживалась и в некоторых более поздних публикациях.

боркой из 123 фирм начиная с 1970 г. и пришли к похожим результатам. В этой выборке, однако, многие фирмы расширялись достаточно быстро, чтобы истощить доступную экономию от масштаба. Мы вернемся к рассмотрению данных 1970 г. при исследовании издержек производства в подразде 10.5.1.

Таблица 6.5. Логквадратичная функция издержек (в скобках приведены стандартные ошибки)

| | $\log Q$ | $\log^2 Q$ | $\log P_L - \log P_F$ | $\log P_K - \log P_F$ | R^2 |
|-----------|------------------|-------------------|-----------------------|-----------------------|-------|
| Все фирмы | 0,152 (0,062) | 0,051 (0,0054) | 0,481 (0,161) | 0,074 (0,150) | 0,96 |

Предшествующий пример иллюстрирует три полезных инструмента для определения неспецифицированной нелинейности и работы с ней: анализ ошибок, использование кусочно-линейной регрессии и многочленов для аппроксимации неизвестной регрессионной функции.

6.3.5. Внутренне линейные модели

Логлинейная модель иллюстрирует промежуточный случай нелинейной регрессионной модели. Уравнение, однако, является **внутренне линейным** (или линейным по сути) (intrinsically linear). Взяв логарифмы от $Y_i = \alpha X_i^{\beta_2} e^{\varepsilon_i}$, мы приходим к

$$\ln Y_i = \ln \alpha + \beta_2 \ln X_i + \varepsilon_i$$

или

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i.$$

Хотя это уравнение почти во всем линейно, следует отметить, что оно уже нелинейно по α . Записывая его через β_1 , мы получаем полностью линейную модель. Но интерес может представлять именно другая форма записи. Конечно, в принципе ничего не изменилось, поскольку β_1 — это просто $\ln \alpha$. Если параметр β_1 может быть оценен, то можно предложить очевидную оценку для α , $\hat{\alpha} = \exp(\beta_1)$.

Этот факт приводит нас к полезному свойству внутренне линейных моделей — свойству инвариантности. Используя процедуру нелинейного метода наименьших квадратов, описанную в следующей главе, мы можем оценить α и β_2 непосредственно, минимизируя функцию суммы квадратов ошибок:

$$\text{минимизировать по } (\alpha, \beta_2) : S(\alpha, \beta_2) = \sum_{i=1}^n (\ln Y_i - \ln \alpha - \beta_2 \ln X_i)^2. \quad (6-8)$$

Эта математическая задача представляет некоторую сложность ввиду наличия члена $\ln \alpha$. Однако эквивалентная линейная задача минимизации квадратов —

$$\text{минимизировать по } (\beta_1, \beta_2) : S(\beta_1, \beta_2) = \sum_{i=1}^n (y_i - \beta_1 - \beta_2 x_i)^2 - \quad (6-9)$$

легко решается с помощью обычной оценки методом наименьших квадратов, которую мы использовали до сих пор. Свойство инвариантности, которое мы здесь применяем, состоит в том, что получившиеся в обоих случаях результаты будут численно идентичны: мы получим идентичные результаты, оценивая α с помощью (6-8) и используя $\exp(\beta_1)$ из (6-9). Используя этот результат, мы можем обобщить определение линейности и включить некоторые дополнительные случаи, которые иначе были бы достаточно сложны.

Определение 6.1. Внутренняя линейность. Если в классической линейной регрессионной модели K параметры $\beta_1, \beta_2, \dots, \beta_K$ могут быть записаны как K взаимно однозначных, возможно, нелинейных функций от интересующих нас параметров $\theta_1, \theta_2, \dots, \theta_K$, то модель является внутренне линейной по вектору параметров θ .

Пример 6.7. Внутренне линейная регрессия

В подразделе 14.6.4 мы оценим с помощью метода максимального правдоподобия параметры модели

$$f(y|\beta, x) = \frac{(\beta + x)^{-\rho}}{\Gamma(\rho)} y^{\rho-1} e^{-y/(\beta+x)}.$$

В этой модели $E[y|x] = (\beta\rho) + \rho x$, что предполагает другой подход к оценке двух параметров. Эта функция является по сути линейной регрессионной моделью $E[y|x] = \beta_1 + \beta_2 x$, где $\beta_1 = \beta\rho$ и $\beta_2 = \rho$. Оценки этих параметров могут быть получены при помощи метода наименьших квадратов, а оценка β получена как b_1/b_2 . Так как это выражение является нелинейной функцией оцененных параметров, то для оценки стандартных ошибок нужно использовать дельта-метод. Используя данные из этого примера⁹, получаем, что оценки методом наименьших квадратов β_1 и β_2 равны $-4,1431$ ($23,734$) и $2,4261$ ($1,5915$) (в скобках указаны стандартные ошибки). Оценка ковариации равна $-36,979$. Оценка β равна $-4,1431/2,4261 = -1,7077$. Мы оцениваем выборочную дисперсию $\hat{\beta}$ как

$$\begin{aligned} \text{Est. Var}[\hat{\beta}] = & \left(\frac{\partial \hat{\beta}}{\partial b_1}\right)^2 \widehat{\text{Var}}[b_1] + \left(\frac{\partial \hat{\beta}}{\partial b_2}\right)^2 \widehat{\text{Var}}[b_2] + \\ & + 2\left(\frac{\partial \hat{\beta}}{\partial b_1}\right)\left(\frac{\partial \hat{\beta}}{\partial b_2}\right) \widehat{\text{Cov}}[b_1, b_2] = 8,6889^2. \end{aligned}$$

В табл. 6.6 сравниваются оценки параметров, полученные при помощи методов наименьших квадратов и максимального правдоподобия. Меньшие стандартные ошибки оценок максимального правдоподобия — это результат наличия неэффективных (равных) весов наблюдений, которые использует метод наименьших квадратов. Гамма-распределение обладает

⁹Данные приведены в таблице FC.1 приложения.

существенной скошенностью. Кроме того, как нам известно из результатов приложения С, это распределение принадлежит к экспоненциальному семейству. Мы выяснили, что для гамма-распределения достаточные статистики для плотности записываются как $\sum_i y_i$ и $\sum_i \ln y_i$. Эффективная оценка должна использовать оба этих выражения, тогда как оценка наименьших квадратов использует только первое из них.

Таблица 6.6. Оценки регрессии в гамма-модели: сравнение методов наименьших квадратов и максимального правдоподобия

| | β | | ρ | |
|-----------------------------------|---------|--------------------|--------|--------------------|
| | Оценка | Стандартная ошибка | Оценка | Стандартная ошибка |
| Метод наименьших квадратов | -1,708 | 8,689 | 2,426 | 1,592 |
| Метод максимального правдоподобия | -4,719 | 2,345 | 3,151 | 0,794 |

В понятии внутренней линейности по сути принципиальным моментом является взаимная однозначность. Если требуемые условия выполняются, то модель может быть оценена через функции β_1, \dots, β_K , а после их оценки из них уже можно вывести оценки интересующих нас параметров. Взаимная однозначность отображения является **условием идентификации** (identification condition). Если оно выполнено, то говорят, что интересующие нас параметры (θ) регрессии **точно идентифицируются** (exactly identified) через параметры β линейной модели. Прекрасный пример приведен в работах Кменты (Kmenta (1986, с. 515), Kmenta (1967)).

Пример 6.8. Производственная CES-функция

Производственная функция с постоянной эластичностью замещения может быть записана как

$$\ln y = \ln \gamma - \frac{\nu}{\rho} \ln[\delta K^{-\rho} + (1 - \delta)L^{-\rho}] + \varepsilon. \quad (6-10)$$

Разложение в ряд Тейлора этой функции около точки $\rho = 0$ имеет вид

$$\begin{aligned} \ln y &= \ln \gamma + \nu \delta \ln K + \nu(1 - \delta) \ln L + \rho \nu \delta(1 - \delta) \left\{ -\frac{1}{2} [\ln K - \ln L]^2 \right\} + \\ &+ \varepsilon' = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon', \end{aligned} \quad (6-11)$$

где $x_1 = 1$, $x_2 = \ln K$, $x_3 = \ln L$, $x_4 = -\frac{1}{2} \ln^2(K/L)$, а преобразования выглядят как

$$\begin{aligned} \beta_1 &= \ln \gamma, \quad \beta_2 = \nu \delta, \quad \beta_3 = \nu(1 - \delta), \quad \beta_4 = \rho \nu \delta(1 - \delta), \\ \gamma &= e^{\beta_1}, \quad \delta = \beta_2 / (\beta_2 + \beta_3), \quad \nu = \beta_2 + \beta_3, \quad \rho = \beta_4 (\beta_2 + \beta_3) / (\beta_2 \beta_3). \end{aligned} \quad (6-12)$$

Оценки параметров β_1 , β_2 , β_3 и β_4 могут быть получены при помощи метода наименьших квадратов. Оценки γ , δ , ν и ρ , получаемые из второй строки (6-12), совпадают с теми, какие мы получили бы с помощью нелинейного метода наименьших квадратов, использованного непосредственно для (6-11). (Однако, как показывает автор, они не совпадают с нелинейной оценкой наименьших квадратов для (6-10) ввиду использования разложения Тейлора для получения (6-11)). Для построения оценки асимптотической ковариационной матрицы $\theta' = [\gamma, \delta, \nu, \rho]$ воспользуемся дельта-методом. Матрица производных имеет вид

$$C = \frac{\partial \theta}{\partial \beta'} = \begin{bmatrix} e^{\beta_1} & 0 & 0 & 0 \\ 0 & \beta_3/(\beta_2 + \beta_3)^2 & -\beta_2/(\beta_2 + \beta_3)^2 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & -\beta_3\beta_4/(\beta_2^2\beta_3) & -\beta_2\beta_4/(\beta_2\beta_3^2) & (\beta_2 + \beta_3)/(\beta_2\beta_3) \end{bmatrix}.$$

Оценка ковариационной матрицы для $\hat{\theta}$ равна $\hat{C} [s^2(\mathbf{X}'\mathbf{X})^{-1}]\hat{C}'$.

Не все модели вида

$$y_i = \beta_1(\theta)x_{i1} + \beta_2(\theta)x_{i2} + \dots + \beta_K(\theta)x_{ik} + \varepsilon_i \quad (6-13)$$

являются внутренне линейными. Напомним, что для этого требуется условие взаимной однозначности функции (т. е. точная идентификация параметров). Например, модель

$$y_i = \alpha + \beta x_{i1} + \gamma x_{i2} + \beta\gamma x_{i3} + \varepsilon_i$$

не является линейной. Причина здесь состоит в том, что ее запись в форме (6-13) не учитывает условия равенства коэффициентов β_4 и $\beta_2\beta_3$, что является **нелинейным ограничением** (nonlinear restriction). В этой модели три параметра α , β и γ **переопределены** (overidentified) через четыре параметра β_1 , β_2 , β_3 и β_4 . Оценки методом наименьших квадратов без ограничений коэффициентов β_2 , β_3 и β_4 могут быть использованы для получения двух оценок каждого из интересующих параметров, и нет никакой гарантии, что они совпадут. Модели, не являющиеся внутренне линейными, обсуждаются в главе 7.

6.4. Моделирование и тестирование структурного сдвига

Одним из распространенных приложений F -теста является тестирование на наличие **структурного сдвига**¹⁰. Специфицируя регрессионную модель, мы предполагаем, что ее предпосылки выполняются для всех наблюдений в выборке. Однако легко представить процедуру тестирования гипотезы о том, что некоторые (или все) коэффициенты регрессии отличаются в

¹⁰Этот тест обычно называется **тестом Чоу** (Chow test), следуя работе Chow (1960).

разных подвыборках исходных данных. Для примера вернемся к данным о рынке бензина, который мы рассматривали в примерах 2.3, 4.2, 4.4, 4.8 и 4.9. Как демонстрирует рис. 6.5, этот рынок развивался вполне предсказуемо до нефтяного шока 1973 г., а затем был достаточно переменчив. Выделяются большие скачки цен в 1973 и 1980 гг., так же как и большой разброс в потреблении¹¹. Представляется маловероятным, что одна и та же регрессионная модель сможет описывать оба этих периода.

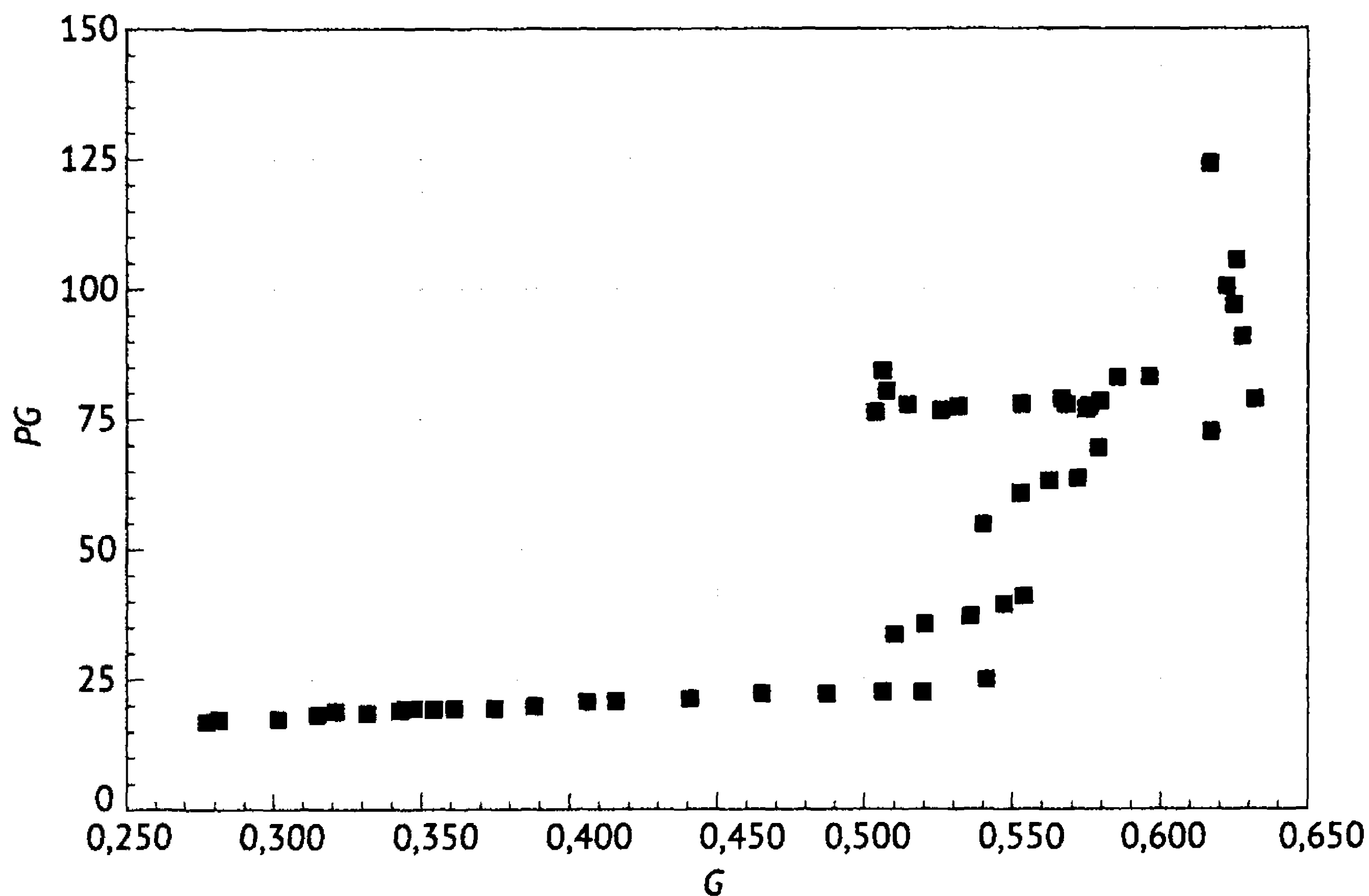


Рис. 6.5. Цены на бензин и подушевое потребление, 1953–2004 гг.

6.4.1. Различные векторы параметров

Данные о потреблении бензина охватывают два очень разных периода. До 1973 г. топливо имелось в изобилии и мировые цены на бензин были стабильны или снижались по крайней мере на протяжении двух десятилетий. Эмбарго 1973 г. было вызвано новым для этого рынка периодом, связанным с дефицитом, растущими ценами и повторяющейся нестабильностью. Вполне возможно, что соотношение, описываемое нашей регрессионной моделью, изменилось в 1974 г. Для тестирования подобной гипотезы можно действовать следующим образом. Обозначим первые 21 год из имеющихся данных по y и X как y_1 и X_1 , а оставшиеся — как y_2 и X_2 . Регрессия без ограничений, допускающая различные коэффициенты в этих двух периодах, выглядит как

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix}. \quad (6-14)$$

¹¹Схожая картина будет наблюдаться и в 2006 г.

Обозначая матрицы данных как y и X , мы получаем, что оценка методом наименьших квадратов без ограничений имеет вид

$$b = (X'X)^{-1}X'y = \begin{bmatrix} X_1'X_1 & 0 \\ 0 & X_2'X_2 \end{bmatrix}^{-1} \begin{bmatrix} X_1'y_1 \\ X_2'y_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \quad (6-15)$$

что соответствует оценке методом наименьших квадратов, полученной отдельно для каждого из двух уравнений. Таким образом, общая сумма квадратов остатков этой регрессии будет равна сумме двух сумм квадратов остатков двух отдельных регрессий:

$$e'e = e_1'e_1 + e_2'e_2.$$

Вектор коэффициентов при ограничениях может быть получен двумя путями. Формально ограничение $\beta_1 = \beta_2$ имеет вид $R\beta = q$, где $R = [I : -I]$ и $q = 0$. Соответственно общий результат, обсуждавшийся ранее, может быть применен непосредственно. Более легким путем является подстановка этого ограничения в модель напрямую. Если два вектора коэффициентов совпадают, то (6-14) может быть записано как

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \beta + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}$$

и оценка с ограничениями может быть получена простым объединением данных и оцениванием общей регрессии. В этом случае сумма квадратов остатков такой регрессии с ограничением, равная $e_*'e_*$, становится основой для теста. Тестовая статистика выписана в (5-29), где число ограничений J — это число столбцов в матрице X_2 , а число степеней свободы в знаменателе равно $n_1 + n_2 - 2k$.

6.4.2. Недостаточное число наблюдений

В некоторых случаях ряды данных недостаточно длинны для оценки той или иной отдельной регрессии при тестировании структурного сдвига. Например, можно предположить, что у потребителей ушел год или два на то, чтобы приспособиться к двум нефтяным шокам 1973 и 1979 гг., но рынок принципиально не изменился или же изменение было временным. Мы можем применить тот же тест, что и прежде, но теперь выделим только четыре года (1974, 1975, 1980 и 1981) для отдельного рассмотрения. Поскольку требуется оценить шесть коэффициентов, а наблюдений только четыре, отдельные модели оценить невозможно. В работе Фишера (Fisher (1970)) было показано, что в такой ситуации можно поступить следующим образом.

1. Оценить регрессию, используя все данные, и посчитать сумму квадратов остатков при ограничении $e_*'e_*$.
2. Оценить, используя более длинный (адекватный) подпериод (из n_1 наблюдений), сумму квадратов остатков без ограничений $e_1'e_1$. Последнее вычисление производится, считая, что, имея лишь $n_2 < K$ наблюдений, мы можем идеально подогнать модель для y_2 и ничего, таким образом, не прибавим к сумме квадратов остатков.

3. Затем вычисляем F -статистику, используя формулу

$$F [n_2, n_1 - K] = \frac{(\mathbf{e}'_* \mathbf{e}_* - \mathbf{e}'_1 \mathbf{e}_1) / n_2}{\mathbf{e}'_1 \mathbf{e}_1 / (n_1 - K)}. \quad (6-16)$$

Отметим, что у числителя имеется n_2 , а не K степеней свободы¹². Этот тест был назван **тестом Чоу на предсказательную силу** (predictive test), поскольку он эквивалентен расширению ограниченной модели на более короткий подпериод и проведению теста на основе ошибок прогноза модели на этом подпериоде.

6.4.3. Изменение части коэффициентов

Предложенная выше общая постановка задачи допускает различные варианты, позволяющие провести много тестов. Некоторые важные частные случаи можно рассмотреть на примере рынка бензина. Одним из возможных подходов к описанию ситуации на рынке может быть следующий. После нефтяного шока 1973 г. американцы просто уменьшили потребление бензина на фиксированную долю, но другие рыночные соотношения, такие как эластичность по доходу, остались на прежнем уровне. Этот случай привел бы к простому сдвигу вниз логлинейной регрессионной модели или, иными словами, лишь к сокращению значения постоянного члена. Таким образом, уравнение без ограничений имеет различные коэффициенты в двух подпериодах, тогда как уравнение с ограничениями представляет собой объединенную регрессию с различными постоянными членами. Матрицы регрессоров в этих двух случаях выглядят

$$\text{(без ограничений)} \mathbf{X}_U = \begin{bmatrix} \mathbf{i} & \mathbf{0} & \mathbf{W}_{\text{pre73}} & \mathbf{0} \\ \mathbf{0} & \mathbf{i} & \mathbf{0} & \mathbf{W}_{\text{post73}} \end{bmatrix}$$

и

$$\text{(при наличии ограничений)} \mathbf{X}_R = \begin{bmatrix} \mathbf{i} & \mathbf{0} & \mathbf{W}_{\text{pre73}} \\ \mathbf{0} & \mathbf{i} & \mathbf{W}_{\text{post73}} \end{bmatrix}.$$

Первые два столбца матрицы \mathbf{X}_U — это фиктивные переменные, которые обозначают подпериод, в который попадает наблюдение.

Другое возможное объяснение состоит в том, что константа и один или несколько коэффициентов наклона изменились, но другие параметры остались неизменными. Результаты примера 6.9 указывают на то, что постоянный член и эластичности по цене и доходу изменились гораздо сильнее, чем перекрестные ценовые эластичности и временной тренд. Тест Чоу для такого ограничения очень похож на случай изменения только в константе. Пусть \mathbf{Z} обозначает переменные, коэффициенты которых, как мы считаем, изменились, а \mathbf{W} обозначает переменные, коэффициенты которых, по нашему мнению, остались на прежнем уровне. Тогда матрица регрессоров для регрессии с ограничениями будет выглядеть как

$$\mathbf{X} = \begin{bmatrix} \mathbf{i}_{\text{pre}} & \mathbf{Z}_{\text{pre}} & \mathbf{0} & \mathbf{0} & \mathbf{W}_{\text{pre}} \\ \mathbf{0} & \mathbf{0} & \mathbf{i}_{\text{post}} & \mathbf{Z}_{\text{post}} & \mathbf{W}_{\text{post}} \end{bmatrix}. \quad (6-17)$$

¹²Один из способов объяснения этого состоит в том, что для получения идеальной подгонки модели требуется лишь $n_2 < K$ коэффициентов.

Как и прежде, вектор коэффициентов в модели без ограничений есть комбинация двух отдельных регрессий.

6.4.4. Тесты на структурное изменение при различных дисперсиях

Важным предположением, сделанным при использовании теста Чоу, является тот факт, что дисперсия ошибок постоянна для обеих (или всех) регрессий. В модели с ограничениями, если это условие не выполняется, первые n_1 элементов ε имеют дисперсию σ_1^2 , тогда как следующие n_2 имеют дисперсию σ_2^2 и т.д. Таким образом, модель с ограничениями является гетероскедастичной и наши результаты для классической регрессионной модели неприменимы. Как было показано в работах Шмидта и Сиклса (Schmidt, Sickles (1977)), Охтани и Тойода (Ohtani, Toyoda (1985)) и Тойода и Охтани (Toyoda, Ohtani (1986)), есть основания предполагать, что в подобной ситуации вероятность ошибки первого рода будет превышать выбранный нами уровень значимости. (Таким образом, мы будем считать большой F -статистику, которая на самом деле не превышает *соответствующего*, но неизвестного критического значения.) Конкретное влияние этого эффекта будет сложным образом зависеть от данных и от того, насколько различаются дисперсии.

Если размер выборки достаточно велик, то существует тест, обоснованный независимо от равенства дисперсий ошибок. Предположим, что $\hat{\theta}_1$ и $\hat{\theta}_2$ — это две состоятельные и асимптотически нормально распределенные оценки параметра, основанные на независимых выборках¹³, с асимптотическими ковариационными матрицами V_1 и V_2 . Тогда в условиях нулевой гипотезы о том, что истинные параметры совпадают, $\hat{\theta}_1 - \hat{\theta}_2$ имеет среднее 0 и асимптотическую ковариационную матрицу $V_1 + V_2$. В условиях нулевой гипотезы статистика Вальда

$$W = (\hat{\theta}_1 - \hat{\theta}_2)'(\hat{V}_1 + \hat{V}_2)^{-1}(\hat{\theta}_1 - \hat{\theta}_2) \quad (6-18)$$

имеет предельное распределение хи-квадрат с K степенями свободы. Тест на равенство нулю разности между параметрами может быть основан на этой статистике¹⁴. Этот тест может быть непосредственно применен для тестирования нашей гипотезы об общих векторах параметров в регрессиях. Большие значения статистики приводят к отвержению гипотезы.

В малой или средней выборке тест Вальда обладает неприятным свойством, что вероятность ошибки первого рода устойчиво превышает критическое значение, которое мы используем для его проведения. (То есть мы будем слишком часто отвергать нулевую гипотезу о том, что параметры в подвыборках равны.) Мы должны использовать большее критическое значение. В работе Охтани и Кобаяси (Ohtani, Kobayashi (1986)) был предложен «граничный» тест, который частично решает эту проблему¹⁵.

¹³Без требования независимости этот и некоторые другие похожие тесты никуда не годятся. Эта проблема представляет собой вариант известной проблемы Беренса–Фишера.

¹⁴Andrews, Fair (1988). Истинный размер предложенного теста не определен. Он зависит от природы альтернативной гипотезы. Если дисперсии различаются существенно, то предполагаемые критические значения могут оказаться ненадежными.

¹⁵Kobayashi (1986). Альтернативный, несколько более запутанный тест предложен в работе Jayatissa (1977). Дальнейшее обсуждение этого вопроса можно найти в работе Thursby (1982).

Таблица 6.7. Функции потребления бензина

| Коэффициенты | 1953–2004 | Весь период (с разными константами для подпе- риодов) | До шока | После шока |
|-----------------------|-----------|---|------------|-------------|
| Константа | -26,6787 | -24,9009 | -22,1647 | |
| Константа | | -24,8167 | | -15,3283 |
| $\ln Income/Pop$ | 1,6250 | 1,4562 | 0,8482 | 0,3739 |
| $\ln PG$ | -0,05392 | -0,1132 | -0,03227 | -0,1240 |
| $\ln PNC$ | -0,08343 | -0,1044 | 0,6988 | -0,001146 |
| $\ln PUC$ | -0,08467 | -0,08646 | -0,2905 | -0,02167 |
| Год | -0,01393 | -0,009232 | 0,01006 | 0,004492 |
| R^2 | 0,9649 | 0,9683 | 0,9975 | 0,9529 |
| Стандартная ошибка | 0,04709 | 0,04524 | 0,01161 | 0,01689 |
| Сумма квадратов | 0,101997 | 0,092082 | 0,00202244 | 0,007127899 |

Было отмечено, что размер теста Вальда (Wald test) может отличаться от используемого нами и что это отклонение будет функцией альтернативной гипотезы. Существует две общие постановки задачи, в которых тест такого типа может представлять интерес. При сравнении двух (возможно) различных генеральных совокупностей (как уравнения предложения труда мужчин и женщин) мы мало что можем сказать по поводу предложенной статистики, поскольку отсутствует конкретная информация об альтернативных гипотезах. Но большая работа над статистикой подобного рода может быть проделана в контексте временных рядов. В этом случае природа альтернативной гипотезы определена более конкретно.

Пример 6.9. Структурный сдвиг на рынке бензина

На рис. 6.5 приведен график цен и выпуска для американского рынка бензина в период с 1953 по 2004 г. Первые 21 наблюдение находятся внизу на графике и указывают на спокойное функционирование рынка. Оставшиеся наблюдения явно отражают последовавший беспорядок на этом рынке.

Мы будем использовать описанные тесты Чоу для изучения этого рынка. Модель, которую мы будем рассматривать, была предложена в примере 2.3, но здесь мы включим временной тренд:

$$\ln(G/Pop)_t = \beta_1 + \beta_2 \ln(Income/Pop)_t + \beta_3 \ln PG_t + \beta_4 \ln PNC_t + \beta_5 \ln PUC_t + \beta_6 t + \varepsilon_t.$$

Три цены в этом уравнении относятся к G , новым автомобилям и подержанным автомобилям. Переменная $Income/Pop$ представляет собой доход на душу населения, а G/Pop — это потребление бензина на душу населения. Временной тренд вычислен как $t = Year - 1952$, поэтому для первого наблюдения $t = 1$. Результаты регрессий для четырех функциональных форм представлены в табл. 6.7. Используя данные по всей выборке с 1953 по 2004 г. и для двух подпериодов, с 1953 по 1973 и с 1974 по 2004 г., мы получаем три оцененные регрессии: в первом и в двух последних столбцах. F -статистика для тестирования ограничения, что коэффициенты в двух уравнениях совпадают, равна

$$F[6, 40] = \frac{(0,101997 - (0,00202244 + 0,007127899))/6}{(0,00202244 + 0,007127899)/(21 + 31 - 12)} = 67,645.$$

Критическое значение из таблицы распределения равно 2,336, поэтому в соответствии с нашими ожиданиями мы отвергаем гипотезу о том, что векторы коэффициентов одинаковы для двух периодов. Используя полный набор из 52 наблюдений для оценивания модели, мы получаем сумму квадратов ошибок, равную $e^*e^* = 0,101997$. Когда $n_2 = 4$, наблюдения для 1974, 1975, 1980 и 1981 гг. удаляются из выборки, сумма квадратов падает до $e'e = 0,0973936$. F -статистика равна 0,496. Поскольку критическое значение для распределения $F[4, 48 - 6]$ равно 2,594, мы не отвергаем гипотезу о стабильности. На данный момент мы можем прийти к выводу, что, хотя что-то на рынке определенно изменилось, гипотеза временного неравновесия не является адекватным описанием ситуации.

Другой способ вычисления этой статистики может оказаться более удобным. Рассмотрим исходные данные со всеми 52 наблюдениями. Теперь добавим в регрессию четыре бинарные переменные: Y_{1974} , Y_{1975} , Y_{1980} и Y_{1981} . Каждая из них принимает значение один в единственный указанный год и значение ноль во все 51 оставшихся года. Затем мы оцениваем регрессию с исходными шестью переменными и этими четырьмя дополнительными дамми-переменными. Сумма квадратов остатков в этой регрессии равна 0,0973936 (в точности столько же, сколько получилось бы в случае удаления из выборки этих четырех наблюдений — см. упражнение 7 в главе 3), поэтому F -статистика для тестирования совместной гипотезы о том, что эти четыре коэффициента равны нулю, равна

$$F[4, 42] = \frac{(0,101997 - 0,0973936)/4}{0,0973936/(52 - 6 - 4)} = 0,496,$$

что совпадает со значением, полученным ранее. (См. подраздел 6.4.2 для обсуждения этого теста.)

F -статистика для тестирования ограничения о том, что все коэффициенты, за исключением константы, в этих двух уравнениях совпадают, рассчитывается на основе последних строк трех последних столбцов таблицы:

$$F[5, 40] = \frac{(0,092082 - (0,00202244 + 0,007127899))/5}{(0,00202244 + 0,007127899)/(21 + 31 - 12)} = 72,506.$$

Табличное критическое значение равно 2,449, так что эта гипотеза также отвергается. Данные говорят о том, что модели в этих двух периодах различаются существенно, а не простым изменением постоянного члена.

F -статистика, которая получается при оценке модели с ограничением, состоящим в том, что эластичности по цене автомобиля и коэффициент при временном тренде не меняются, равна

$$F[3, 40] = \frac{(0,01441975 - (0,00202244 + 0,007127899))/3}{(0,00202244 + 0,007127899)/(52 - 6 - 6)} = 7,678.$$

(Соответствующая регрессия с ограничением не приводится.) Критическое значение из таблицы F -распределения равно 2,839, так что и эта гипотеза отвергается. Следует, однако, обратить внимание на то, что это значение гораздо меньше полученных нами ранее. Этот факт указывает на то, что существенная часть различий в моделях между двумя подпериодами действительно объясняется изменениями в константе и эластичностях по цене и доходу.

Тестовая статистика в (6-18) для результатов регрессии из табл. 6.7 дает значение 502,34. 5%-е критическое значение из таблицы распределения хи-квадрат с шестью степенями свободы равно 12,59. Таким образом, на основании теста Вальда мы снова отвергаем гипотезу о том, что один и тот же вектор коэффициентов применим для описания обоих подпериодов: с 1953 по 1973 и с 1974 по 2004 г. Также стоит отметить, что статистика Вальда справедлива лишь в больших выборках, а наши выборки в 21 и 31 наблюдение едва ли соответствуют этому стандарту. Мы протестировали гипотезу о том, что регрессионная модель для рынка бензина изменилась в 1973 г. и на основании F -теста (теста Чоу) уверенно отвергли гипотезу о стабильности модели.

Пример 6.10. Доклад ВОЗ

Версия 2000 г. Доклада Всемирной организации здравоохранения (ВОЗ, World Health Organization, WHO) включала масштабное сопоставление мировых систем здравоохранения по странам. [См. World Health Organization (2000). См. также <http://www.who.int/whr/en/>.] В книге приводятся результаты многолетних исследований и содержатся тысячи страниц материала. Среди наиболее спорных и подвергшихся наиболее активному публичному обсуждению частей доклада есть глава, описывающая сравнение предоставления услуг здравоохранения в 191 стране, т.е. практически во всем мире [Evans et al. (2000a,b). Обзор обсуждения в популярной прессе см., например, в Hiltz (2000).] В исследовании была проанализирована эффективность предоставления услуг здравоохранения по двум показателям: стандартному и широко используемому — ожидаемой продолжительности здоровой жизни (Disability Adjusted Life Expectancy, DALE) и новому инновационному показателю, разработанному авторами исследования и объединяющему пять индикаторов (composite, COMP), который учитывал эффективность и справедливость в предоставлении услуг. Регрессионный анализ, использующий метод стохастической границы (см. подраздел 19.2.4), соотносил показатели здравоохранения с двумя важными факторами: образованием и расходами на здравоохранение (на душу населения). Для сравнения различных стран были проанализированы остатки.

В ходе работы исследователи ВОЗ использовали данные из табл. F6.3 приложения. (Они использовали панель данных с 1993 по 1997 г. Для этого примера мы выбрали данные 1997 г.) Данные ВОЗ были использованы многими исследователями в последующих работах. [См., например, работы Hollingsworth, Wildman (2002), Gravelle and al. (2002) и Greene (2004).] Регрессионная модель, использованная ВОЗ, содержала переменные DALE или COMP в левой части и расходы на здравоохранение, образование и квадрат образования в правой части уравнения. В работе Грина (Greene (2004)) в регрессию был добавлен ряд дополнительных переменных, таких как ВВП на душу населения, мера распределения дохода и показатели государственной эффективности и демократизации политической структуры, предоставляемые Всемирным банком (World Bank).

Среди аспектов исследования, вызвавших споры, обсуждался тот факт, что модель рассматривала в совокупности страны с сильно различающимися показателями. Вторым ярким аспектом получившихся результатов, предположенным в работе Хилтса (Hilts (2000)) и подтвержденным в работе Грина (Greene (2004)), был тот факт, что «эффективными» среди рассматриваемых стран оказались 30 относительно здоровых стран, входящих в ОЭСР (Организация экономического сотрудничества и развития, Organization for Economic Cooperation and Development, OECD), тогда как остальной мир в среднем выглядел гораздо хуже. Здесь мы рассмотрим этот аспект подробнее в отношении переменной DALE. Анализ переменной COMP оставляется читателю в качестве упражнения. Таблица 6.8 содержит оценки регрессионных моделей для переменной DALE для всей выборки, стран ОЭСР и остальных соответственно. На первый взгляд между двумя подгруппами не наблюдается очень больших различий. Сначала нами была протестирована совместная значимость дополнительных переменных: распределения дохода (GINI), ВВП на душу населения и т.д. Для каждой группы F -статистика составляет $\{(e'^*e^* - e'e)/7\}/\{e'e/(n-11)\}$. Эти F -статистики указаны в последней строке таблицы. Критические значения для $F[7,180]$ (вся выборка), $F[7, 19]$ (ОЭСР) и $F[7, 150]$ (не ОЭСР) равны 2,061, 2,543 и 2,071 соответственно. Мы заключаем, что дополнительные переменные вносят важный вклад в модель для стран, не входящих в ОЭСР (и для всех стран), но не для членов ОЭСР. Наконец, для тестирования на структурный сдвиг между членами ОЭСР и остальными странами мы вычислили

$$F[11, 169] = \frac{[7757,007 - (69,74428 + 7378,598)]/11}{(69,74428 + 7378,598)/(191 - 11 - 11)} = 0,637.$$

95%-е критическое значение для $F[11, 169]$ равно 1,846. Таким образом, мы не отвергаем гипотезу о том, что регрессионная модель одинакова для двух групп стран. Статистика Вальда, вычисленная по (6-18), говорит о другом. Значение этой статистики составляет 35,221. 95%-е критическое значение из таблицы хи-квадрат-распределения с 11 степенями свободы равно 19,675. На этом основании мы можем отвергнуть гипотезу о равенстве векторов коэффициентов.

Таблица 6.8. Результаты регрессий для продолжительности жизни

| | Все страны | | ОЭСР | | Не ОЭСР | |
|------------------------|------------|-----------|----------|-----------|-----------|-----------|
| Константа | 5,237 | 38,734 | 42,728 | 49,328 | 26,812 | 41,408 |
| Health | 0,00629 | -0,00180 | 0,00268 | 0,00114 | 0,00955 | -0,00178 |
| exp | | | | | | |
| Education | 7,931 | 7,178 | 6,177 | 5,156 | 7,0433 | 6,499 |
| Education ² | -0,439 | -0,426 | -0,385 | -0,329 | -0,374 | -0,372 |
| Gini | | -17,333 | | -5,762 | | -21,329 |
| coeff | | | | | | |
| Tropic | | -3,200 | | -3,298 | | -3,144 |
| Pop. | | -0,255e-4 | | 0,000167 | | -0,425e-4 |
| Dens. | | | | | | |
| Public | | -0,0137 | | -0,00993 | | -0,00939 |
| exp | | | | | | |
| GDP PC | | 0,000483 | | 0,000108 | | 0,000600 |
| Democracy | | 1,629 | | -0,546 | | 1,909 |
| Govt. | | 0,748 | | 1,224 | | 0,786 |
| Eff. | | | | | | |
| R ² | 0,6824 | 0,7299 | 0,6483 | 0,7340 | 0,6133 | 0,6651 |
| Станд. | 6,984 | 6,565 | 1,883 | 1,916 | 7,366 | 7,014 |
| ошибки | | | | | | |
| Сумма | 9 121,795 | 7 757,002 | 92,21064 | 69,74428 | 8 518,750 | 7 378,598 |
| квадра- | | | | | | |
| тов | | | | | | |
| N | | 191 | | 30 | | 161 |
| GDP/Pop | | 6 609,37 | | 18 199,07 | | 4 449,79 |
| F-тест | | 4,524 | | 0,874 | | 3,311 |

6.4.5. Тестирование стабильности модели при помощи теста на предсказательную силу

Тест, определенный в (6-16) в подразделе 6.4.2, эквивалентен проверке гипотезы $H_0: \beta_2 = \beta_1$ в «модели»

$$y_t = \mathbf{x}'_t \beta_1 + \varepsilon_t, \quad t = 1, \dots, T_1,$$

$$y_t = \mathbf{x}'_t \beta_2 + \varepsilon_t, \quad t = T_1 + 1, \dots, T_1 + T_2.$$

(Отметим, что дисперсия ошибок предполагается одинаковой в обоих подпериодах.) Альтернативная постановка (использованная в примере) выглядит как

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{X}_2 & \mathbf{I} \end{bmatrix} \begin{pmatrix} \beta \\ \gamma \end{pmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}.$$

Такая постановка означает, что

$$y_t = \mathbf{x}'_t \beta_1 + \varepsilon_t, \quad t = 1, \dots, T_1,$$

$$y_t = \mathbf{x}'_t \beta_2 + \gamma_t + \varepsilon_t, \quad t = T_1 + 1, \dots, T_1 + T_2.$$

Поскольку на γ_t не накладывается никаких ограничений, эта альтернативная постановка фактически говорит о том, что регрессионная модель для первых T_1 -периодов перестает работать во втором периоде (и на самом деле никакая систематическая модель не работает во втором периоде). Проверка гипотезы $\gamma = 0$ в такой постановке задачи, таким образом, будет являться тестом на стабильность модели. Коэффициенты, полученные методом наименьших квадратов для этой регрессии, могут быть найдены, если использовать следующую формулу для блочной обратной матрицы:

$$\begin{aligned} \begin{pmatrix} \mathbf{b} \\ \mathbf{c} \end{pmatrix} &= \begin{bmatrix} \mathbf{X}'_1\mathbf{X}_1 + \mathbf{X}'_2\mathbf{X}_2 & \mathbf{X}'_2 \\ \mathbf{X}_2 & \mathbf{I} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}'_1\mathbf{y}_1 + \mathbf{X}'_2\mathbf{y}_2 \\ \mathbf{y}_2 \end{bmatrix} \\ &= \begin{bmatrix} (\mathbf{X}'_1\mathbf{X}_1)^{-1} & -(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_2 \\ -\mathbf{X}_2(\mathbf{X}'_1\mathbf{X}_1)^{-1} & \mathbf{I} + \mathbf{X}_2(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_2 \end{bmatrix} \begin{bmatrix} \mathbf{X}'_1\mathbf{y}_1 + \mathbf{X}'_2\mathbf{y}_2 \\ \mathbf{y}_2 \end{bmatrix} \\ &= \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{c}_2 \end{pmatrix}, \end{aligned}$$

где \mathbf{b}_1 — это оценка методом наименьших квадратов, полученная по первым T_1 наблюдениям, а \mathbf{c}_2 — это $\mathbf{y}_2 - \mathbf{X}_2\mathbf{b}_1$. Ковариационная матрица для всего набора оценок представляет собой оценку s^2 , умноженную на матрицу в скобках. Два подвектора остатков в этой регрессии равны $\mathbf{e}_1 = \mathbf{y}_1 - \mathbf{X}_1\mathbf{b}_1$ и $\mathbf{e}_2 = \mathbf{y}_2 - (\mathbf{X}_2\mathbf{b}_1 + \mathbf{I}\mathbf{c}_2) = 0$, поэтому сумма квадратов остатков в этой регрессии составляет просто $\mathbf{e}'_1\mathbf{e}_1$. Это ровно та же сумма квадратов, что встречалась нам ранее в (6-16). Число степеней свободы для знаменателя равно $[T_1 + T_2 - (K + T_2)] = T_1 - K$, а число степеней свободы для числителя — это число элементов в векторе γ , равное T_2 . Регрессия с ограничением $\gamma = 0$ представляет собой объединенную модель, которая также совпадает с уже встречавшейся в (6-16). Отсюда следует, что F -статистика для тестирования нулевой гипотезы в этой модели — это в точности то, что мы рассматривали выше в (6-16). Именно поэтому этот тест называют тестом на предсказательную силу.

6.5. Заключение и выводы

В этой главе обсуждалась функциональная форма регрессионной модели. Мы рассмотрели использование фиктивных переменных и других преобразований для того, чтобы включить различные нелинейности в модель, а также другие нелинейные модели, в которых параметры нелинейной модели могут быть построены с помощью оценок, получаемых с помощью линейной регрессии. В заключительном разделе этой главы описаны тесты, позволяющие выяснить, менялась ли предполагаемая модель за рассматриваемый в выборке период или отличалась ли она для различных групп наблюдений.

Основные термины и понятия

| | |
|--|---|
| Binary variable — Бинарная переменная | Nonlinear restriction — Нелинейное ограничение |
| Test Chow — Тест Чоу | Overidentified — Сверхидентифицированный |
| Control group — Контрольная группа | Piecewise continuous — Кусочно-непрерывная |
| Control observations — Контрольные наблюдения | Placebo effect — Эффект плацебо |
| Difference in differences — Разность разностей | Predictive test — Тест на предсказательную силу |
| Dummy Variable — Фиктивная переменная, Дамми-переменная, Дамми, Бинарная переменная | Qualification indexes — Индексы квалификации |
| Dummy variable trap — Ловушка дамми-переменных | Response — Реакция |
| Exactly identified — Точная идентификация | Semilog equation — Полулогарифмическое уравнение |
| Identification conditions — Условие идентификации | Spline — Сплайн |
| Interaction terms — Компоненты взаимодействия | Structural change — Структурный сдвиг |
| Intrinsically linear — Внутренне линейная | Threshold effects — Граничные (пороговые) эффекты |
| Knots — Узлы | Time profile — Временной срез |
| Loglinear model — Логлинейная модель | Treatment — Воздействие |
| Marginal effect — Предельный эффект | Treatment group — Группа, подвергшаяся воздействию |
| Natural experiment — Естественный эксперимент | Wald test — Тест Вальда |

Упражнения

1. Оценивается регрессионная модель с $K = 16$ независимыми переменными на панельных данных, насчитывающих семь лет наблюдений. Суммы квадратов для семи отдельных регрессий и для объединенной регрессии приведены ниже. Модель для объединенных данных включает разные константы для каждого года. Проверьте гипотезу о том, что для каждого года коэффициенты совпадают.

| | 1954 | 1955 | 1956 | 1957 | 1958 | 1959 | 1960 | Вместе |
|------------|------|------|------|------|------|------|------|--------|
| Наблюдения | 65 | 55 | 87 | 95 | 103 | 87 | 78 | 570 |
| $e'e$ | 104 | 88 | 206 | 144 | 199 | 308 | 211 | 1425 |

2. *Обратная регрессия.* Стандартным подходом к статистическому выявлению дискриминации на рабочем месте является оценка регрессии

$$y = \alpha + \mathbf{x}'\beta + \gamma d + \varepsilon, \quad (1)$$

где y — ставка заработной платы, d — фиктивная переменная, указывающая принадлежность ($d = 1$) или непринадлежность ($d = 0$) к классу, по отношению к которому предполагается наличие дискриминации. Регрессоры \mathbf{x} включают факторы, характерные для конкретного вида ра-

боты, а также показатели квалификации индивида. Представляет интерес гипотеза $H_0: \gamma \geq 0$ против альтернативы $H_1: \gamma < 0$. Регрессия пытается ответить на следующий вопрос: «Оплачивается ли труд индивидов, занятых на этой работе и принадлежащих к рассматриваемому классу ($d = 1$), хуже, чем труд столь же квалифицированных индивидов, к этому классу не принадлежащих ($d = 0$)» Рассмотрим альтернативный подход. Обладают ли представители класса на той же работе, получающие ту же заработную плату, более высокой квалификацией? В таком случае это тоже может рассматриваться как форма дискриминации. Для изучения этого вопроса в работе Конвэй и Робертса (Conway, Roberts (1983)) предлагается следующая процедура:

- 1) Оценить (1) методом наименьших квадратов. Обозначить оценки как a , b и c .
- 2) Оценить набор **индексов квалификации** (qualification indexes),

$$q = ai + Xb. \quad (2)$$

Обратите внимание на отсутствие cd в правой части уравнения.

- 3) Оценить регрессию q на константу, y и d . Уравнение выглядит как

$$q = \alpha_* + \beta_* y + \gamma_* d + \varepsilon_*. \quad (3)$$

При проведении анализа предполагается, что $\gamma < 0, \gamma_* > 0$.

- a. Докажите, что независимо от теории оценки наименьших квадратов c и c_* связаны через соотношение

$$c_* = \frac{(\bar{y}_1 - \bar{y})(1 - R^2)}{(1 - P)(1 - r_{yd}^2)} - c, \quad (4)$$

где \bar{y}_1 = среднее значение y для наблюдений с $d = 1$,

\bar{y} = среднее значение y для всех наблюдений,

P = среднее значение d ,

R^2 = коэффициент детерминации для уравнения (1),

r_{yd}^2 = квадрат коэффициента корреляции между y и d .

[Подсказка: модель содержит постоянный член.] Таким образом, чтобы упростить преобразования, предположим, что все наблюдения измерены как отклонения от общих выборочных средних, и используем частичную регрессию для оценки коэффициентов (3). Далее для (2) используем результат оценки регрессии $y = ai + Xb + cd + e$, откуда $q = y - cd - e$. Здесь и далее мы опускаем постоянный член. Таким образом, в регрессии (3) вы оцениваете регрессию $[y - cd - e]$ на y и d .

- b. Будут ли выводы, полученные на основе выборки, обязательно соответствовать теории? [Подсказка: рассмотрите $c = 0$.]

Материалы, в которых обсуждается статья Конвэй и Робертса (Conway and Roberts (1983)), вышли в *Journal of Business and Economic Statistics* в апреле 1983 г.

3. *Обратная регрессия (продолжение)*. Это и следующее упражнение продолжают анализ из упражнения 2. В упражнении 2 интерес представляла конкретная фиктивная переменная, а регрессоры были измерены точно.

Здесь мы рассматриваем случай, в котором главный регрессор измерен с ошибкой. В статье (Kamlich, Polachek (1982)) подробно рассматривается этот случай.

Рассмотрим простую модель с ошибкой измерения в регрессорах:

$$y = \alpha + \beta x^* + \varepsilon, \quad x = x^* + u,$$

где u и ε не коррелированы, а регрессор x , измеренный с ошибкой, есть наблюдаемый аналог истинного x^* .

- a. Предположим что x^* , u и ε распределены нормально со средними μ^* , 0 и 0, дисперсиями $\sigma_{x^*}^2$, σ_u^2 и σ_ε^2 и нулевыми ковариациями. Выведите пределы по вероятности оценок, полученных при помощи метода наименьших квадратов, коэффициентов α и β .
 - b. В качестве альтернативы рассмотрите регрессию x на константу и y с последующей оценкой величины, обратной полученной оценке. Получите предел по вероятности этой оценки.
 - c. Устанавливают ли «прямая» и «обратная» оценки границы для истинного коэффициента?
4. *Обратная регрессия (продолжение)*. Предположим, что модель из упражнения 3 обобщается до модели $y = \beta x^* + \gamma d + \varepsilon$, $x = x^* + u$. Для удобства мы опустим постоянный член. Предположим, что x^* , ε и u независимы и распределены нормально с нулевыми средними. Предположим, что d — это случайная величина, принимающая значения один и ноль с вероятностями π и $1 - \pi$ в генеральной совокупности и не зависящая от всех остальных переменных в модели. Предшествующая модель, а также ее варианты встречались в литературе, посвященной рассматриваемой нами проблеме — дискриминации. Мы рассматриваем y как переменную «зарботной платы», x^* — как «квалификацию», а x — как некоторую несовершенную меру измерения, как, например, образование. Фиктивная переменная d указывает на принадлежность ($d = 1$) или отсутствие принадлежности ($d = 0$) к некоторому рассматриваемому классу. Гипотеза о дискриминации выглядит как $\gamma < 0$ против $\gamma \geq 0$.
- a. Чему равен предел по вероятности s — оценки наименьших квадратов параметра γ в регрессии y на x и d ? [Подсказки: важна независимость x^* и d . Также $\text{plim } d'd/n = \text{Var}[d] + E^2[d] = \pi(1 - \pi) + \pi^2 = \pi$. Эта небольшая модификация не затрагивает сути модели, но существенно упрощает преобразования.] Теперь предположим, что x^* и d зависимы. В частности, предположим, что $E[x^* | d = 1] = \mu^1$ и $E[x^* | d = 0] = \mu^0$. Повторите выкладки с учетом этого предположения.
 - b. Рассмотрим теперь регрессию x на y и d . Чему равен предел по вероятности коэффициента при d в этой регрессии в предположении, что x^* и d независимы?
 - c. Предположим, что x^* и d не являются независимыми, но γ на самом деле меньше нуля. Предполагая, что оба предшествующих уравнения все еще верны, что оценивается при помощи выражения $(\bar{y} | d = 1) - (\bar{y} | d = 0)$? Что оценивает это выражение, если γ на самом деле равен нулю?

Эмпирические приложения

1. В задаче 1 из главы 3 и в задаче 1 из главы 5 были рассмотрены данные работы Купа и Тобиаса (Коор, Tobias (2004)) о заработных платах, об образовании, о способностях и т.д. Мы продолжим здесь анализ этих данных. (Источник, расположение и форма данных уже обсуждались в задачах, упомянутых выше.) Рассмотрим модель

$$\begin{aligned} \ln Wage = & \beta_1 + \beta_2 Educ + \beta_3 Ability + \beta_4 Experience \\ & + \beta_5 Mother's\ education + \beta_6 Father's\ education + \\ & + \beta_7 Brokenhome + \beta_8 Siblings + \varepsilon. \end{aligned}$$

- a. Оцените полную регрессию методом наименьших квадратов и представьте ваши результаты. На основе ваших результатов скажите, чему равна оценка предельного значения, в \$/часах, дополнительного года образования для индивида с 12 годами образования, когда остальные переменные принимают свои средние значения, а *Broken-*

home = 0?

- b. Нас интересуют возможные нелинейности в эффекте влияния уровня образования на *lnWage*. (В исходной работе исследовался опыт работы. Как и прежде, в нашу задачу не входит воспроизведение результатов.) На гистограмме показателя образования видны значения от 9 до 20, высокий пик на 12 годах (окончание средней школы) и, что несколько удивительно, второй пик на 15: интуитивно можно было бы ожидать его на 16. Рассмотрите следующие фиктивные переменные, являющиеся характеристиками уровня образования:

$$\begin{aligned} HS &= 1, \text{ если } Educ \leq 12, 0 \text{ в противном случае,} \\ Col &= 1, \text{ если } Educ > 12 \text{ и } Educ \leq 16, 0 \text{ в противном случае,} \\ Grad &= 1, \text{ если } Educ > 16, 0 \text{ в противном случае.} \end{aligned}$$

- c. Замените переменную *Educ* в модели на (*Col*, *Grad*), сделав фиктивную переменную о среднем образовании (*HS*) базовой категорией, и снова оцените модель. Приведите все результаты. Как эти результаты изменились? На основе ваших результатов скажите, чему равна предельная ценность высшего образования? (На самом деле это предельная ценность 16 лет образования: в последние годы окончание высшего учебного заведения требовало в среднем несколько более четырех лет.) Чему равно предельное влияние на *lnWage* магистерской степени?
- d. Введение фиктивных переменных в пункте b приводит к потере достаточно большой части информации. Другой способ включения нелинейности в образование — это использовать соответствующую функцию. Добавьте $Educ^2$ в уравнение в пункте a и переоцените модель. Снова приведите все результаты. Какие можно предложить изменения? Проверьте гипотезу о том, что квадратичный член в уравнении является лишним, т.е. что коэффициент при нем равен нулю. На основе ваших результатов изобразите график логарифма зарплаты как функции от образования.

- е. Можно предположить, что ценность образования повышается с увеличением способностей человека. Мы можем проверить этот эффект, включив компоненту взаимодействия в уравнение. Добавьте переменную

$$Educ_Ability = Educ \times Ability$$

в базовую модель из пункта а. Чему теперь равна предельная ценность дополнительного года образования? Выборочное среднее показателя способностей равно 0,052374. Получите доверительный интервал для предельного влияния на $\ln Wage$ дополнительного года образования для человека со средними способностями.

- ф. Объедините модели из пунктов с и d. Добавьте $Educ^2$ и $Educ_Ability$ в базовую модель из пункта а и получите новую оценку. Как и прежде, приведите все результаты и опишите ваши выводы. Если мы определим «низкие способности» как способности ниже среднего, а «высокие способности» как способности выше среднего, то выборочные средние составят $-0,798563$ для 7864 индивидов с низкими способностями в выборке и $+0,717891$ для 10 055 индивидов с высокими способностями в выборке. Используя постановку задачи из пункта с, с помощью этой новой функциональной формы нарисуйте, опишите и сравните графики логарифмов зарплат для индивидов с низкими и высокими способностями.

2. (Обобщение задачи 1.) Здесь мы рассмотрим, применимы ли различные модели, рассмотренные в задаче 1, к индивидам, проживающим в «распавшихся семьях» (Broken homes). Используя результаты подразделов 6.4.1 и 6.4.4, проверьте гипотезу о том, что одна и та же модель (не включающая фиктивную переменную *Broken home*) применима к обоим группам индивидов: для тех, у кого *Broken home* = 0, и тех, у кого *Broken home* = 1.
3. В классическом исследовании Солоу (Solow (1957)) о влиянии технического прогресса на экономику США автор предложил следующую производственную функцию: $q(t) = A(t)f[k(t)]$, где $q(t)$ — это агрегированный выпуск за рабочий час, $k(t)$ — это агрегированное соотношение труда и капитала, а $A(t)$ — это индекс технологии. В работе рассматривались четыре статические модели, $q/A = \alpha + \beta \ln k$, $q/A = \alpha - \beta/k$, $\ln(q/A) = \alpha + \beta \ln k$ и $\ln(q/A) = \alpha + \beta/k$. Данные исследования с 1909 по 1949 г. приведены в табл. F6.4 приложения.
 - а. Используйте эти данные для оценки α и β для четырех функций, приведенных выше. (Замечание: ваши результаты не вполне совпадут с результатами работы Солоу (Solow (1957)). См. следующее упражнение для разрешения этого несовпадения.)
 - б. В вышеприведенном исследовании автор отмечает: «Диаграмма рассеяния q/A в зависимости от k изображена на рис. 4. Учитывая объем предварительной обработки, которой подверглись данные, качество подгонки знаменательно. Это, однако, не относится к группе

точек, которые, очевидно, находятся слишком высоко. Эти наблюдения относятся к семи последним годам рассматриваемого периода (1943–1949). По тому, как они расположились почти параллельно основному разбросу, есть соблазн заключить, что в 1943 г. агрегированная производственная функция просто сдвинулась».

Приведите диаграмму рассеяния q/A от k и проверьте приведенный выше результат.

- с. Оцените четыре модели, которые вы оценивали в предыдущем задании, включив фиктивную переменную для годов с 1943-го по 1949-й. Как меняются ваши результаты? (*Замечание: эти результаты совпадают с приводимыми автором исследования, хотя он и не приводит значения коэффициентов при фиктивной переменной.*)
- d. Автором был сделан следующий шаг: он предположил, что данные до 1943 г. принципиально отличались от данных начиная с 1943 г. Используйте тест Чоу для оценки различий между двумя периодами, включая все четыре функциональные формы. Обратите внимание на то, что с помощью фиктивной переменной вы можете провести тест, взяв компоненту взаимодействия между фиктивной переменной и любой функцией k , включенной в регрессию. Используйте F -тест для проверки гипотезы.
4. Данные о числе аварий, включающие информацию о числе поврежденных кораблей, типе корабля и периоде его постройки, приведены в табл. 6.9. Выделяются пять типов кораблей и четыре различных периода их постройки. Используйте F -тесты и регрессии с фиктивными переменными для проверки гипотез об отсутствии значимого «эффекта типа корабля» на ожидаемое число аварий. Затем используйте ту же процедуру для проверки наличия значимого «эффекта периода».

Таблица 6.9. Аварии с повреждением кораблей

| Тип корабля | Период постройки | | | |
|-------------|------------------|-----------|-----------|-----------|
| | 1960–1964 | 1965–1969 | 1970–1974 | 1975–1979 |
| A | 0 | 4 | 18 | 1 |
| B | 29 | 53 | 44 | 18 |
| C | 1 | 1 | 2 | 1 |
| D | 0 | 0 | 11 | 4 |
| E | 0 | 7 | 12 | 1 |

ГЛАВА 7

Нелинейная, полупараметрическая и непараметрическая модели регрессии¹

7.1. Введение

До этого момента акцент делался на **модели линейной регрессии**

$$y = x_1\beta_1 + x_2\beta_2 + \dots + \varepsilon. \quad (7-1)$$

Главы 2–5 были посвящены оцениванию параметров методом наименьших квадратов и выводу статистических свойств оценок. Эти свойства мы использовали для получения точечных и интервальных оценок, проверки гипотез и построения прогнозов. Изменения, которым посвящена глава 6, привели к рассмотрению несколько более общей линейной модели

$$y = f_1(\mathbf{x})\beta_1 + f_2(\mathbf{x})\beta_2 + \dots + \varepsilon. \quad (7-2)$$

По определению, которое мы хотим использовать в данной главе, эта модель все еще является линейной, так как параметры входят в модель линейным образом. Раздел 7.1 этой главы будет посвящен **нелинейной регрессионной модели** (включающей (7-1) и (7-2) как частные случаи):

$$y = h(x_1, x_2, \dots, x_P; \beta_1, \beta_2, \dots, \beta_K) + \varepsilon, \quad (7-3)$$

где функция условного среднего содержит P переменных и K параметров. Модель такой формы меняет функцию условного среднего с $E[y|\mathbf{x}, \beta] = \mathbf{x}'\beta$ на $E[y|\mathbf{x}] = h(\mathbf{x}, \beta)$ для более общих функций. Это дает возможность рассматривать гораздо большее число функциональных форм, чем может предложить линейная модель². Изменение в форме модели потребует от нас вывода альтернативного метода оценивания, **нелинейного метода наименьших квадратов**. Мы также изучим более детально интерпретацию параметров в нелинейных моделях. В частности, поскольку $\partial E[y|\mathbf{x}]/\partial \mathbf{x}$ более не равен β , мы заинтересованы в новой интерпретации вектора коэффициентов β .

Линейный и нелинейный методы наименьших квадратов используются для оценки параметров **функции условного среднего**, $E[y|\mathbf{x}]$. Как мы

¹В этой главе изучаются некоторые достаточно углубленные аспекты регрессионного моделирования и численного анализа. Она может быть пропущена при первом чтении без потери связности изложения.

²Полное обсуждение этого вопроса можно найти в Amemiya (1985). Другие важные источники — это Jennrich (1969), Malinvaud (1970) и особенно Goldfeld, Quandt (1971, 1972). Подробное обсуждение, содержащее авторскую позицию, изложено в Davidson, MacKinnon (1993).

увидели из примера 4.5, интерес могут представлять и другие соотношения между y и x , такие как **условная медиана**. В разделе 7.3 мы возвращаемся к этому аспекту, изучая функцию условной медианы и оценку наименьших абсолютных отклонений. В этом разделе также ослабляется ограничение, согласно которому коэффициенты модели должны быть одинаковы в разных частях распределения y (при заданном x). Оценка LAD измеряет параметры условной медианы, т.е. функции 50-го перцентиля. **Модель квантильной регрессии** позволяет параметрам регрессии меняться в зависимости от той части условного распределения, которую мы рассматриваем.

Модели, рассматривавшиеся до сих пор, являются по своей природе полупараметрическими и становятся менее параметрическими при переходе от раздела 7.2 к разделу 7.3. **Частично линейная регрессия**, изучаемая в разделе 7.2, расширяет (7-1) до вида $y = f(x) + z'\beta + \varepsilon$. Итогом подобных расширений становится модель, в которой связь y и x не обязана описываться конкретной параметризованной функцией. Используя большое число графических и ядерных методов, в разделе 7.5 мы будем изучать вопрос об анализе связи с помощью **непараметрической** регрессии, которая по сути требует чуть больше, чем выполнение соотношения $E[y|x] = h(x)$.

7.2. Нелинейные регрессионные модели

Общая форма нелинейной регрессионной модели имеет вид

$$y_i = h(x_i, \beta) + \varepsilon_i. \quad (7-4)$$

Очевидно, линейная модель является ее частным случаем. Более того, некоторые модели, кажущиеся нелинейными, как, например,

$$y = e^{\beta_1} x_1^{\beta_2} x_2^{\beta_3} e^{\varepsilon},$$

становятся линейными после некоторого преобразования, в данном случае логарифмирования. В этой главе нас будут интересовать модели, для которых не существует подобного преобразования. Рассмотрим следующий пример.

Пример 7.1. Производственная функция CES

В примере 6.8 мы рассматривали производственную функцию с постоянной эластичностью замещения факторов (Constant Elasticity of Substitution, CES):

$$\ln y = \ln \gamma - \frac{\nu}{\rho} \ln [\delta K^{-\rho} + (1 - \delta)L^{-\rho}] + \varepsilon. \quad (7-5)$$

Не существует преобразования, приводящего это уравнение к уравнению, линейному по параметрам. В примере 6.5 мы использовали приближение этой функции рядом Тейлора в окрестности точки $\rho = 0$ для того, чтобы получить внутренне линейное уравнение, к которому можно было бы применить метод наименьших квадратов. Тем не менее модель (7.5) нелинейна в том смысле, в котором мы понимаем линейность в этой главе.

В этом и следующем разделе мы обобщим предположения модели линейной регрессии, чтобы иметь возможность использовать нелинейные формы, как в примере 7.1. После этого выведем оценку нелинейным методом наименьших квадратов, изучим ее статистические свойства и рассмотрим, как полученная оценка может быть использована для проверки гипотез и построения прогнозов в модели.

7.2.1. Предположения нелинейной регрессионной модели

Далее нам потребуется несколько более формальное определение нелинейной модели регрессии. Для наших целей достаточно будет следующего определения, которое, как уже было замечено ранее, включает линейную модель как частный случай. Мы предположим, что существует распределение вероятностей или процесс порождения данных (DGP) для наблюдаемых y_i и вектор истинных параметров β , который является характеристикой этого DGP. Предположения нелинейной модели регрессии заключаются в следующем:

1. **Функциональная форма.** Функция условного среднего y_i при заданном \mathbf{x}_i имеет вид

$$E[y_i | \mathbf{x}_i] = h(\mathbf{x}_i, \beta), \quad i = 1, \dots, n,$$

где $h(\mathbf{x}_i, \beta)$ — непрерывная и дифференцируемая по β функция.

2. **Идентифицируемость параметров модели.** Вектор параметров в модели называется идентифицируемым (т.е. может быть оценен), если не существует ненулевого параметра $\beta^0 \neq \beta$ такого, что $h(\mathbf{x}_i, \beta^0) = h(\mathbf{x}_i, \beta)$ для всех \mathbf{x}_i . В линейной модели этому предположению соответствовало предположение о полном ранге, но простого отсутствия «мультиколлинеарности» среди переменных в \mathbf{x} недостаточно, чтобы это условие выполнялось для нелинейной модели регрессии. Этот факт демонстрирует пример 7.2.
3. **Нулевое среднее шоков.** Из предположения 1 следует, что мы можем записать

$$y_i = h(\mathbf{x}_i, \beta) + \varepsilon_i,$$

где $E[\varepsilon_i | h(\mathbf{x}_i, \beta)] = 0$. Утверждается, что шок в наблюдении i не коррелирован с функцией условного среднего для всех наблюдений в выборке. Это не совсем то же самое, что предположение о некоррелированности шоков и экзогенных переменных, которое также достаточно часто используется.

4. **Гомоскедастичность и отсутствие автокорреляции.** Как и в линейной модели, мы предполагаем условную гомоскедастичность $E[\varepsilon_i^2 | h(\mathbf{x}_j, \beta), j = 1, \dots, n] = \sigma^2$ конечно и постоянно

и отсутствие автокорреляции $E[\varepsilon_i \varepsilon_j | h(\mathbf{x}_i, \beta), h(\mathbf{x}_j, \beta),$

$$j = 1, \dots, n] = 0 \quad \text{для всех } j \neq i. \quad (7-6)$$

5. **Процесс порождения данных** для \mathbf{x}_i предполагается доброкачественным, таким что первые и вторые моменты стремятся к фиксированным конечным аналогам генеральной совокупности. Существенным является

предположение о строгой экзогенности процесса, порождающего x_i по отношению к процессу, порождающему ε_i . Данные x_i предполагаются «доброкачественными».

6. **Основная вероятностная модель.** Существует корректное распределение вероятностей, порождающее ε_i . На данном этапе мы предположим, что процесс порождает выборку из некоррелированных, одинаково (безусловно) распределенных случайных величин ε_i с нулевым средним и с дисперсией σ^2 условно относительно $h(x_i, \beta)$. Таким образом, на данном этапе форма нашей модели является **полупараметрической**. (См. раздел 12.3.) Мы не будем предполагать конкретного распределения для величин ε_i . Предположений 3 и 4 об условных моментах будет достаточно для получения результатов данной главы. В главе 14 мы полностью параметризуем модель, предположив, что шоки распределены нормально. Это позволит нам сделать более точные выводы относительно некоторых тестовых статистик и к тому же позволит рассмотреть некоторые обобщения регрессионной модели. Здесь это предположение не требуется.

Пример 7.2. Идентификация в транслогарифмической системе спроса

В работе Кристенсена, Джоргенсена и Ло (Christensen, Jorgenson, Lau (1975)) была предложена транслогарифмическая **косвенная функция полезности** для потребителя, распределяющего свой бюджет между K товарами:

$$\ln V = \beta_0 + \sum_{k=1}^K \beta_k \ln(p_k/M) + \sum_{k=1}^K \sum_{j=1}^K \gamma_{kj} \ln(p_k/M) \ln(p_j/M),$$

где V — косвенная полезность, p_k — цена k -го товара, а M — это доход. Полезность, прямая или косвенная, ненаблюдаема, и функция полезности не может быть использована для эмпирической модели. **Тождество Роя**, примененное к этой логарифмической функции, приводит к уравнению распределения бюджета для k -го товара в форме

$$S_k = - \frac{\partial \ln V / \partial \ln p_k}{\partial \ln V / \partial \ln M} = \frac{\beta_k + \sum_{j=1}^K \gamma_{kj} \ln(p_j/M)}{\beta_M + \sum_{j=1}^K \gamma_{Mj} \ln(p_j/M)} + \varepsilon, \quad k = 1, \dots, K,$$

где $\beta_M = \sum_k \beta_k$ и $\gamma_{Mj} = \sum_k \gamma_{kj}$. Никакое преобразование уравнения распределения бюджета не приведет к линейной модели. Это уравнение регрессии нелинейно по своей сущности. (Оно также является частью системы уравнений, но этот аспект мы пока игнорируем.) Несмотря на то что уравнение распределения сформулировано в терминах наблюдаемых величин, его все равно нельзя рассматривать в качестве эмпирической модели из-за **проблемы идентификации**. Если мы домножим каждый параметр в уравнении распределения на одну и ту же константу, то эта константа, появившаяся и в числителе, и в знаменателе, сократится и значение функции в уравнении не изменится. Указанную неопределенность можно разрешить, наложив условие нормировки $\beta_M = 1$. Заметим, что проблемы идентификации подобного характера не возникают в линейных моделях.

7.2.2. Нелинейная оценка наименьших квадратов

Нелинейная оценка наименьших квадратов определяется как точка минимума суммы квадратов:

$$S(\beta) = \frac{1}{2} \sum_{i=1}^n \varepsilon_i^2 = \frac{1}{2} \sum_{i=1}^n [y_i - h(\mathbf{x}_i, \beta)]^2. \quad (7-7)$$

Условия первого порядка для задачи минимизации выглядят как

$$\frac{\partial S(\beta)}{\partial \beta} = \sum_{i=1}^n [y_i - h(\mathbf{x}_i, \beta)] \frac{\partial h(\mathbf{x}_i, \beta)}{\partial \beta} = \mathbf{0}. \quad (7-8)$$

В линейной модели вектор частных производных был бы равен регрессорам \mathbf{x}_i . Далее мы будем называть производные функции условного среднего по параметрам псевдорегрессорами, $\mathbf{x}_i^0(\beta) = \mathbf{x}_i^0$. Мы получаем, что нелинейная оценка наименьших квадратов находится как решение

$$\frac{\partial S(\beta)}{\partial \beta} = \sum_{i=1}^n \mathbf{x}_i^0 \varepsilon_i = \mathbf{0}. \quad (7-9)$$

Эти уравнения для нелинейной регрессии представляют собой аналоги обычных уравнений метода наименьших квадратов (3-5). Для вычислений требуется применение итерационной процедуры решения (см. пример 7.3). Метод решения приведен в подразделе 7.2.6.

Из предположений 1 и 3 следует, что $E[\varepsilon_i | h(\mathbf{x}_i, \beta)] = 0$. Для линейной модели в связи с линейностью условного среднего получается, что ε_i и \mathbf{x}_i не коррелированы между собой. Тем не менее некоррелированность ε_i с некоторой нелинейной функцией от \mathbf{x}_i (функцией регрессии) необязательно влечет некоррелированность как с \mathbf{x}_i , так и с другой нелинейной функцией от \mathbf{x}_i . С другой стороны, результаты, касающиеся свойств оценки в этой модели, которые мы получим, формулируются не в терминах \mathbf{x}_i , а в терминах некоторых функций от \mathbf{x}_i (производных функции регрессии), поэтому с этой точки зрения нам не требуется предположение $E[\varepsilon | \mathbf{X}] = \mathbf{0}$.

Последний вывод не является бесспорным для произвольной модели. Динамические модели, столь популярные в современной литературе, значительно усложняют подобный анализ. Если можно предположить, что ε_i строго некоррелирует с любой предшествующей информацией в модели, включая предыдущие шоки, то, возможно, интерпретация, аналогичная случаю линейной модели, может оказаться подходящей. Однако результаты о сходимости, необходимые для вывода асимптотических свойств оценки, все равно придется усилить. Динамическая нелинейная регрессионная модель выходит за рамки рассматриваемого нами материала. Строгой независимости ε_i и \mathbf{x}_i будет достаточно для некоррелированности ε_i и любой функции от \mathbf{x}_i , но опять-таки для динамической модели это предположение может быть спорным. Некоторое обсуждение этого аспекта модели нелинейной регрессии может быть найдено в книгах Дэвидсона и Маккиннона (Davidson, MacKinnon (1993, 2004)).

Если шоки в нелинейной модели распределены нормально, то логарифм функции плотности нормального распределения для i -го наблюдения имеет вид

$$\ln f(y_i | \mathbf{x}_i, \beta, \sigma^2) = -(1/2) \{ \ln 2\pi + \ln \sigma^2 + [y_i - h(\mathbf{x}_i, \beta)]^2 / \sigma^2 \}. \quad (7-10)$$

В этом частном случае, как следует из пункта D.2 теоремы 14.2 (об оценке методом максимального правдоподобия), производные логарифма функции плотности по параметрам имеют нулевое среднее, т.е.

$$E \left[\frac{\partial \ln f(y_i | x_i, \beta, \sigma^2)}{\partial \beta} \right] = E \left[\frac{1}{\sigma^2} \frac{\partial h(x_i, \beta)}{\partial \beta} \varepsilon_i \right] = 0, \quad (7-11)$$

и в случае нормального распределения производные и шоки не коррелированы. Выполнение этого предположения в других случаях зависит от модели, но при разумных условиях мы будем считать это предположение выполненным. [Ruud (2000, с. 540).]

Для линейной модели **условие ортогональности** $E[\mathbf{x}_i \varepsilon_i] = 0$ дает возможность получить оценку наименьших квадратов как **GMM-оценку** в модели. (См. главу 13.) Условие ортогональности заключается в том, что регрессоры не коррелированы с шоками. Для нашей модели то же условие выполняется для первых производных функции условного среднего. Результат (7-11) задает моментное условие, из которого мы можем определить нелинейную оценку наименьших квадратов как GMM-оценку.

Пример 7.3. Условия первого порядка для нелинейной модели

Условия первого порядка для оценки параметров модели нелинейной регрессии,

$$y_i = \beta_1 + \beta_2 e^{\beta_3 x_i} + \varepsilon_i,$$

нелинейным методом наименьших квадратов [см. (7-13)] выглядят как

$$\frac{\partial S(\mathbf{b})}{\partial b_1} = - \sum_{i=1}^n [y_i - b_1 - b_2 e^{b_3 x_i}] = 0,$$

$$\frac{\partial S(\mathbf{b})}{\partial b_2} = - \sum_{i=1}^n [y_i - b_1 - b_2 e^{b_3 x_i}] e^{b_3 x_i} = 0,$$

$$\frac{\partial S(\mathbf{b})}{\partial b_3} = - \sum_{i=1}^n [y_i - b_1 - b_2 e^{b_3 x_i}] b_2 x_i e^{b_3 x_i} = 0.$$

Для этих уравнений не существует решения в явном виде.

Признавая некоторую возможную некорректность, пока мы будем определять модель нелинейной регрессии следующим образом.

Определение 7.1. Нелинейная модель регрессии. Модель называется **нелинейной**, если условия первого порядка для оценки параметров методом наименьших квадратов являются нелинейными функциями от параметров.

Таким образом, мы определяем нелинейность в терминах техники оценки параметров, а не вида регрессионной функции. В дальнейшем мы расширим это определение, чтобы включить процедуры оценки, отличные от метода наименьших квадратов.

7.2.3. Асимптотические свойства оценки нелинейным методом наименьших квадратов

Для нелинейной оценки наименьших квадратов можно получить некоторые аналитические результаты, такие как состоятельность и асимптотическая нормальность. Мы не можем гарантировать, что нелинейная оценка наименьших квадратов является наиболее эффективной, кроме как в случае нормально распределенных шоков. (К аналогичному заключению мы пришли для случая линейной модели.) Однако в полупараметрической постановке, рассматриваемой в этой главе, мы можем задаться вопросом, не является ли эта оценка в некотором смысле оптимальной, учитывая имеющуюся у нас информацию. Ответ на этот вопрос оказывается положительным. Несколько примеров, приведенных ниже, иллюстрируют этот момент.

Нам потребуется сделать некоторые предположения о регрессорах. Точные требования обсуждаются в работах Джаджа и др. (Judge et al. (1985)), Амеми (Amemiya (1985)) и Дэвидсона и Маккиннона (Davidson, MacKinnon (2004)). В линейной модели для вывода асимптотических результатов мы предполагали, что матрица выборочных моментов $(1/n)\mathbf{X}'\mathbf{X}$ сходится к положительно определенной матрице \mathbf{Q} . По аналогии мы потребуем выполнения этого условия для производных регрессионной функции, которые называются **псевдорегрессорами** в линеаризованной модели (определенной в (7-29)), *вычисленных при истинных значениях параметров*. Таким образом, в модели нелинейной регрессии аналогом для (4-20) будет

$$\text{plim} \frac{1}{n} \mathbf{X}'_0 \mathbf{X}_0 = \text{plim} \frac{1}{n} \sum_{i=1}^n \left(\frac{\partial h(x_i, \beta_0)}{\partial \beta_0} \right) \left(\frac{\partial h(x_i, \beta_0)}{\partial \beta'_0} \right) = \mathbf{Q}^0, \quad (7-12)$$

где \mathbf{Q}^0 — положительно определенная матрица. Для вывода состоятельности \mathbf{b} в линейной модели нам потребовалось условие $\text{plim}(1/n)\mathbf{X}'\boldsymbol{\varepsilon} = \mathbf{0}$. Мы будем использовать аналогичное условие для псевдорегрессоров:

$$\text{plim} \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i^0 \varepsilon_i = \mathbf{0}.$$

Как уже было отмечено в (4-24), это называется условием ортогональности. Заметим, в частности, что оно отличается от условия ортогональности шо-

ков и данных. Наконец, асимптотическая нормальность может быть выведена при общих условиях, если

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i^0 \varepsilon_i \xrightarrow{d} N[\mathbf{0}, \sigma^2 \mathbf{Q}^0].$$

Имея это в виду, мы можем получить асимптотические свойства нелинейной оценки наименьших квадратов. Фактически это те же условия, которые мы использовали в случае линейной модели, только вместо регрессоров теперь мы используем производные линеаризированной функции, вычисленные в точке β, \mathbf{X}^0 . [Amemiya (1985).]

Критерий нелинейного метода наименьших квадратов представляет собой

$$S(\mathbf{b}) = \frac{1}{2} \sum_{i=1}^n [y_i - h(x_i, \mathbf{b})]^2 = \frac{1}{2} \sum_{i=1}^n e_i^2, \quad (7-13)$$

где \mathbf{b} — искомое решение. Те значения параметров, при которых достигается минимум (половины) суммы квадратов остатков, и есть оценки нелинейным методом наименьших квадратов. Условия первого порядка для поиска минимума записываются как

$$g(\mathbf{b}) = - \sum_{i=1}^n [y_i - h(x_i, \mathbf{b})] \frac{\partial h(x_i, \mathbf{b})}{\partial \mathbf{b}} = 0. \quad (7-14)$$

В линейной модели, рассматриваемой в главе 3, это приводило нас к системе линейных уравнений (нормальным уравнениям) (3-4). Однако в общем случае (7-14) представляет собой систему нелинейных уравнений, которая не имеет решения в явном виде. Заметим, что σ^2 не влияет на решение [как и в (3-4)]. В искомой точке

$$\mathbf{g}(\mathbf{b}) = -\mathbf{X}^0' \mathbf{e} = \mathbf{0}$$

аналогично (3-12) в случае линейной модели.

При наших предположениях мы получаем следующий результат:

Теорема 7.1. *Состоятельность оценки нелинейного метода наименьших квадратов*

Если выполнены следующие предположения:

- a) пространство, содержащее β , является компактом (нет разрывов и невогнутых областей),*
- b) для каждого вектора β^0 в этом пространстве $\text{plim} (1/n)S(\beta^0) = q(\beta^0)$ является непрерывной и дифференцируемой функцией,*
- c) $q(\beta^0)$ имеет единственный минимум, который достигается при истинном значении параметра β , тогда оценка нелинейного метода наименьших*

квадратов, определенная в (7-13) и (7-14), является состоятельной. Мы кратко опишем доказательство, а затем рассмотрим отличия формулировки теоремы и ее доказательства от более простого случая линейной модели. Доказательство, несмотря на тонкость условий, проводится непосредственным образом. Оценка, скажем b^0 , минимизирует $(1/n)S(\beta^0)$. Если $(1/n)S(\beta^0)$ минимизируется для каждого n , то в точке b^0 достигается минимум и при неограниченном росте n . Мы предположили, что точка минимума $q(\beta^0)$ единственна и равна β . Если минимальное значение $\text{plim } S(\beta^0)$ равно пределу по вероятности минимального значения суммы квадратов, то теорема доказана. Это равенство следует из непрерывности в предположении b .

В линейной модели состоятельность оценки метода наименьших квадратов следовала из условий $\text{plim}(1/n)\mathbf{X}'\mathbf{X} = \mathbf{Q}$ и $\text{plim}(1/n)\mathbf{X}'\varepsilon = \mathbf{0}$. В рамках этого подхода мы могли бы использовать линеаризированную модель и получить в сущности тот же результат. Некоторая неточность этого подхода состоит в том, что линеаризированная модель не является истинной, она является приближением. Таким образом, для корректности этого метода требуется предположить или показать, что $\text{plim}(1/n)\mathbf{X}^0'\delta = \mathbf{0}$, где $\delta_i = h(x_i, \beta)$ за вычетом ее приближения рядом Тейлора. Обсуждение этого момента можно найти в работе Миттельхаммера и др. (Mittelhammer et al. (2000, с. 190–191)).

Заметим, что ни слова не было сказано о несмещенности оценки. Оценка линейного метода наименьших квадратов в модели линейной регрессии в сущности является единственной несмещенной, которая рассматривается в этой книге. Вообще говоря, невозможно установить несмещенность для любой другой оценки. Как мы уже видели, несмещенность оценки выступает незначительным преимуществом в большинстве случаев — так, например, это свойство может выполняться как для оценки, построенной для 10 наблюдений, так и для оценки, построенной для 10 000 наблюдений. За исключением линейного случая, главным требованием к оценке является ее состоятельность. После установления состоятельности мы переходим к вопросу об эффективности и в большинстве случаев полагаемся на асимптотическую нормальность как на основу всех статистических выводов.

Теорема 7.2. Асимптотическая нормальность оценки нелинейного метода наименьших квадратов

Если псевдорегрессоры, определенные в (7-12), являются «доброкачественными», то

$$b \stackrel{a}{\sim} N \left[\beta, \frac{\sigma^2}{n} (\mathbf{Q}^0)^{-1} \right],$$

где

$$\mathbf{Q}^0 = \text{plim} \frac{1}{n} \mathbf{X}^0' \mathbf{X}^0.$$

Выборочной оценкой асимптотической ковариационной матрицы является

$$\text{Est.Asy.Var}[b] = \sigma^2 \mathbf{X}^0' \mathbf{X}^0. \quad (7-15)$$

Без предположений о вероятностном распределении трудно установить асимптотическую эффективность оценки нелинейного метода наименьших квадратов. Одним из возможных косвенных подходов может стать следующий. Из предположения об ортогональности псевдорегрессоров и истинных шоков следует, что оценка нелинейного метода наименьших квадратов является GMM-оценкой. При предположениях о гомоскедастичности и об отсутствии автокорреляции оптимальной взвешивающей матрицей будет именно та, которая была нами задействована, и, таким образом, оценка нелинейного метода наименьших квадратов использует оптимальную взвешивающую матрицу в классе GMM-оценок. Значит, она является асимптотически эффективной в классе GMM-оценок.

Из требования о сходимости матрицы в (7-12) к положительно определенной матрице вытекает, что столбцы матрицы регрессоров X^0 должны быть линейно независимыми. Это **условие идентификации** аналогично требованию о линейной независимости независимых переменных в линейной модели. Модели нелинейной регрессии обычно насчитывают несколько независимых переменных, и на первый взгляд может показаться, что достаточно проверить данные на мультиколлинеарность. Однако это не так. Подобная ситуация рассмотрена в примере 7.4.

Состоятельная оценка величины σ^2 основана на остатках:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n [y_i - h(\mathbf{x}_i, \mathbf{b})]^2. \quad (7-16)$$

Поправка на число степеней свободы, $1/(n-K)$, где K — число элементов в β , здесь не так уж нужна, так как все результаты все равно верны лишь асимптотически. В работе Дэвидсона и Маккиннона (Davidson, MacKinnon (2004)) утверждается, что в среднем (7-16) будет недооценивать σ^2 и нужно использовать поправку на число степеней свободы. Большая часть программного обеспечения использует эту поправку, но рекомендуется уточнять этот момент при исследовании. Учитывая все сказанное, оценка асимптотической ковариационной матрицы для оценки нелинейного метода наименьших квадратов задается в (7-15).

После того как оценки нелинейного метода наименьших квадратов получены, статистические тесты и проверка гипотез могут быть проведены ровно так же, как это было описано в главе 5. Небольшая проблема может возникнуть с оценкой качества приближения данных моделью регрессии, так как известная нам мера качества

$$R^2 = 1 - \frac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (7-17)$$

более не обязана находиться между 0 и 1. Тем не менее она является удобной описательной мерой.

7.2.4. Проверка гипотез и ограничения на параметры

В большинстве случаев гипотезы, которые нам потребуется проверять, будут достаточно простыми линейными ограничениями. Тесты можно про-

вести с помощью известных формул, о которых мы говорили в главе 5, и асимптотической ковариационной матрицы, которая была выписана выше. Для более сложных гипотез и нелинейных ограничений процедура несколько менее очевидна. По сути, две процедуры тестирования были разобраны в разделе 5.4: тест Вальда, основанный на состоятельности и асимптотической нормальности оценки, и F -тест, который применим к конечным (любым) выборкам и основан на нормальном распределении шоков. В нелинейном случае мы полагаемся на свойства больших выборок, поэтому статистика Вальда является основным инструментом для статистических выводов. Аналог F -статистики, основанный на качестве приближения данных моделью регрессии, также будет выведен далее. Наконец, в общем случае могут быть построены **тесты множителей Лагранжа**. Поскольку мы (пока) не предположили нормальность распределения шоков, то отложим вывод статистики отношения правдоподобий до главы 14, в которой вернемся к этой модели.

Гипотеза, которую требуется проверить, выглядит как

$$H_0: \mathbf{r}(\boldsymbol{\beta}) = \mathbf{q}, \quad (7-18)$$

где $\mathbf{r}(\boldsymbol{\beta})$ — вектор-столбец из J непрерывных функций элементов $\boldsymbol{\beta}$. Эти ограничения могут быть как линейными, так и нелинейными. Однако необходимо, чтобы эти ограничения были **сверхидентифицирующими**. Это формально означает, что если вектор параметров насчитывает K свободных элементов, то гипотеза $\mathbf{r}(\boldsymbol{\beta}) = \mathbf{q}$ должна накладывать хотя бы одно функциональное соотношение на параметры. Если таких ограничений насчитывается более одного, то они должны быть функционально независимы. Из этих двух условий следует, что **Якобиан** размерности $J \times K$,

$$\mathbf{R}(\boldsymbol{\beta}) = \frac{\partial \mathbf{r}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}, \quad (7-19)$$

должен иметь полный ранг и что количество ограничений J должно быть строго меньше K . Эта ситуация аналогична линейной модели, в которой $\mathbf{R}(\boldsymbol{\beta})$ будет матрицей коэффициентов в ограничениях. (См. также раздел 5.4, где эти методы были рассмотрены применительно к линейной модели.)

Пусть \mathbf{b} есть оценка нелинейного метода наименьших квадратов без каких-либо ограничений и пусть \mathbf{b}_* — оценка, полученная при наложенных гипотезой ограничениях³. Какую тестовую статистику следует использовать, зависит от сложности вычислений. В отличие от случая линейной модели разные тестовые процедуры могут сильно различаться по сложности. Так, в нашем примере значительно проще строить статистику множителей Лагранжа. Из четырех методов, которые мы обсудим, только этот тест не требует вычисления нелинейной регрессии.

Нелинейный аналог знакомой нам F -статистики основан на качестве приближения (т.е. на сумме квадратов остатков) и имеет вид

³Такая вычислительная задача сама по себе может быть очень сложной, особенно если ограничения нелинейны. Мы предполагаем, что каким-либо образом мы все же получили эту оценку.

$$F[J, n - K] = \frac{[S(\mathbf{b}^*) - S(\mathbf{b})]/J}{S(\mathbf{b})/(n - K)}. \quad (7-20)$$

Это уравнение уже появлялось ранее в представлении F -статистики в виде (5-29). Однако в нелинейном случае ни числитель, ни знаменатель не имеют точного хи-квадрат-распределения, поэтому F -распределение получается лишь приближенно. Отметим, что эта F -статистика требует оценки как модели с ограничениями, так и модели без ограничений.

Тест Вальда основан на расстоянии между $\mathbf{r}(\mathbf{b})$ и \mathbf{q} . Если оценки, полученные без ограничений, не удовлетворяют ограничениям, то возникают сомнения в справедливости самих ограничений. Статистика представляет собой

$$\begin{aligned} W &= [\mathbf{r}(\mathbf{b}) - \mathbf{q}]' \{ \text{Est. Asy. Var}[\mathbf{r}(\mathbf{b}) - \mathbf{q}] \}^{-1} [\mathbf{r}(\mathbf{b}) - \mathbf{q}] \\ &= [\mathbf{r}(\mathbf{b}) - \mathbf{q}]' \{ \mathbf{R}(\mathbf{b}) \hat{\mathbf{V}} \mathbf{R}'(\mathbf{b}) \}^{-1} [\mathbf{r}(\mathbf{b}) - \mathbf{q}], \end{aligned} \quad (7-21)$$

где $\hat{\mathbf{V}} = \text{Est. Asy. Var}[\mathbf{b}]$, и $\mathbf{R}(\mathbf{b})$ вычисляется в точке \mathbf{b} , оценке β .

В условиях нулевой гипотезы эта статистика имеет предельное хи-квадрат-распределение с J степенями свободы. Если ограничения верны, то статистика Вальда и F -статистика, умноженная на J , асимптотически эквивалентны. Статистика Вальда может быть основана на оцененной ковариационной матрице, полученной с помощью оценок без ограничений, что может сильно облегчить вычисления, если ограничения нелинейны. Следует отметить, что поведение статистики W в малых выборках может быть непредсказуемым и более традиционная F -статистика может оказаться предпочтительней, если выборка невелика.

То же предупреждение относительно статистики Вальда, которое мы приводили для линейного случая, справедливо и здесь. Поскольку этот тест является чистым тестом значимости и не подвержен влиянию альтернативной гипотезы, то статистика Вальда не инвариантна к тому, как формулировать гипотезу. В том случае, если есть несколько эквивалентных способов сформулировать ограничения $\mathbf{r}(\beta) = \mathbf{q}$, W может приводить к разным результатам в зависимости от способа.

Тест множителей Лагранжа основан на уменьшении суммы квадратов остатков, которое произойдет, если убрать ограничения из модели с ограничениями. Строгое обоснование теста приведено в подразделе 14.6.3. Для модели нелинейной регрессии тест имеет особенно привлекательный вид⁴. Пусть \mathbf{e}_* есть вектор остатков $y_i - h(\mathbf{x}_i, \mathbf{b}_*)$, вычисленный с помощью оценок при наличии ограничений. Напомним, что мы определяли \mathbf{X}^0 как $n \times K$ матрицу производных, вычисленную в векторе истинных параметров (см. (7-29)). Пусть \mathbf{X}_*^0 — такая матрица, вычисленная в векторе оцененных параметров в модели с ограничениями. Тогда статистика множителей Лагранжа для модели нелинейной регрессии имеет вид

$$LM = \frac{\mathbf{e}'_* \mathbf{X}_*^0 [\mathbf{X}_*^{0'} \mathbf{X}_*^0]^{-1} \mathbf{X}_*^{0'} \mathbf{e}_*}{\mathbf{e}'_* \mathbf{e}_* / n}. \quad (7-22)$$

⁴Вывод этого теста можно найти в работе Джаджа и др. (Judge et al. (1985)). Подробное обсуждение содержится в работе Миттельхаммера и др. (Mittelhammer et al. (2000)).

В условиях гипотезы H_0 эта статистика имеет предельное хи-квадрат-распределение с J степенями свободы. Особенно привлекательным в этом подходе является то, что он требует лишь вычисления оценок с ограничениями. Это может дать некоторый выигрыш с точки зрения вычислений, если, например, в результате ограничений модель сводится к линейной. Заметим также, что статистика множителей Лагранжа представляется в виде n , умноженного на нецентрированный R^2 регрессии e_* на X_*^0 . Многие статистики множителей Лагранжа считаются именно в этом виде.

7.2.5. Примеры

В данном подразделе содержатся три примера оценки и статистических выводов в модели нелинейной регрессии. В примере 7.4 рассматривается нелинейная функция потребления в продолжение примеров 1.2 и 2.1. Эта модель является простым примером оценки и проверки гипотез в нелинейной модели. В примере 7.5 исследуется преобразование Бокса–Кокса (Box–Cox). Подобная спецификация полезна для обобщения случая линейной модели: она содержит линейную и логлинейную модели как частные случаи. Наконец, пример 7.6 представляет собой подробный разбор модели экспоненциальной регрессии. В нем мы изучим некоторые аспекты нелинейного моделирования, в частности «эффекты взаимодействия» (interaction effects). Нам уже встречались эффекты взаимодействия в подразделе 6.3.3 в модели вида

$$y = \beta_1 + \beta_2 x + \beta_3 z + \beta_4 xz + \varepsilon.$$

В этом случае эффект взаимодействия был равен $\partial^2 E[y|x, z]/\partial x \partial z = \beta_4$. Эффект отсутствовал, если β_4 равен нулю. Пример 7.6 рассматривает (возможно непреднамеренный) аспект нелинейной модели, который заключается в том, что если $E[y|x, z] = h(x, z, \beta)$, то существует эффект взаимодействия, даже если модель выписывается как

$$h(x, z, \beta) = h(\beta_1 + \beta_2 x + \beta_3 z).$$

Пример 7.4. Анализ нелинейной функции потребления

Линейная функция потребления, рассмотренная в начале главы 2, есть частный случай более общей модели потребления

$$C = \alpha + \beta Y^\gamma + \varepsilon,$$

где степень γ равна 1. При этом ограничении модель является линейной. Если γ может изменяться, то уравнение становится моделью нелинейной регрессии. Квартальные данные по потреблению, реальному располагаемому доходу и некоторым другим показателям экономики США в период с 1950 по 2000 г. приведены в табл. F5.2 приложения. Воспользуемся этими данными для оценки нелинейной функции потребления. (Подробности вычисления оценок приведены в подразделе 7.2.6, в примере 7.8.) Результаты оценки линейной регрессии с ограничениями и нелинейной

регрессий без ограничений с помощью метода наименьших квадратов приведены в табл. 7.1.

Таблица 7.1. Оцененные функции потребления

| Параметр | Линейная модель | | Нелинейная модель | |
|-----------|-----------------|--------------------|-------------------|--------------------|
| | Оценка | Стандартная ошибка | Оценка | Стандартная ошибка |
| α | -80,3547 | 14,3059 | 458,7990 | 22,5014 |
| β | 0,9217 | 0,003872 | 0,10085 | 0,01091 |
| γ | 1,0000 | — | 1,24483 | 0,01205 |
| $e'e$ | 1 536 321,881 | | 504 403,1725 | |
| σ | 87,20983 | | 0,0946 | |
| R^2 | 0,996448 | | 0,998834 | |
| Var[b] | — | | 0,00119037 | |
| Var[c] | — | | 0,00014532 | |
| Cov[b, c] | — | | -0,000131491 | |

С помощью ранее рассмотренных методов мы строим асимптотические стандартные ошибки и оценки σ^2 . (Для того чтобы иметь возможность сравнить ее с s^2 в линейной модели, мы использовали коррекцию на число степеней свободы.)

В предыдущем примере не мог не возникнуть вопрос о мультиколлинеарности в матрице данных $\mathbf{X} = [\mathbf{i}, \mathbf{y}]$; вариация в Y видна невооруженным глазом. Однако в новой модели R^2 регрессии равен 0,998834, а корреляция между двумя псевдорегрессорами $x_2^0 = Y^\gamma$ и $x_3^0 = \beta Y^\gamma \ln Y$ составляет 0,999752. Число обусловленности нормализованной матрицы суммы квадратов и перекрестных произведений равно 208,306. (Число обусловленности вычисляется как квадратный корень из отношения наибольшего собственного значения матрицы $\mathbf{D}^{-1}\mathbf{X}_0'\mathbf{X}_0\mathbf{D}^{-1}$ к наименьшему, где $x_1^0 = 1$ и \mathbf{D} — диагональная матрица, в которой на k -м месте на диагонали стоит квадратный корень из $x_k^0 x_k^0$.) Напомним, что значение 20 представляет собой порог для проблемных данных. Согласно правилам, обсуждаемым в подразделах 4.7.1 и А.6.6, проблема коллинеарности в наших «данных» не может быть проигнорирована. На самом деле оказывается, что здесь это вообще не является проблемой.

Для проверки гипотез и построения доверительных интервалов мы будем использовать знакомые процедуры, подразумевая, что все результаты имеют смысл лишь асимптотически. Таким образом, хи-квадрат-статистика более уместна, чем F -статистика. Например, для проверки гипотезы о том, что γ отлична от 1, асимптотический t -тест, основанный на стандартном нормальном распределении, дает нам

$$z = \frac{1,24483 - 1}{0,01205} = 20,3178.$$

Результат выходит за критическое значение 1,96 для 5%-го уровня значимости, и таким образом мы отвергаем линейную модель регрессии в пользу

нелинейной. Три других метода проверки гипотез приводят к тому же результату.

- F -статистика равна

$$F[1, 204 - 3] = \frac{(1\,536\,321,881 - 504\,403,17)/1}{504\,403,17/(204 - 3)} = 411,29.$$

Критическое значение по таблице равно 3,84, и гипотеза отвергается.

- Статистика Вальда, основанная на расстоянии от $\hat{\gamma}$ до 1, представляет собой просто квадрат асимптотической t -статистики, вычисленной ранее:

$$W = \frac{(1,24483 - 1)^2}{0,01205^2} = 412,805.$$

Критическое значение по таблице хи-квадрат равно 3,84.

- Для статистики множителей Лагранжа элементы в \mathbf{x}_i^* — это

$$\mathbf{x}_i^* = [1, Y^\gamma, \beta Y^\gamma \ln Y].$$

Для того чтобы вычислить оценки с ограничениями, мы используем обычные оценки, сделанные методом наименьших квадратов для α и β и 1 для γ и получаем

$$\mathbf{x}_i^* = [1, Y, \beta Y \ln Y].$$

Остатки здесь — это остатки, полученные из модели линейной регрессии при оценке методом наименьших квадратов. Подставляя значения, вычисленные ранее, мы получаем

$$LM = \frac{996\,103,9}{(1\,536\,321,881/204)} = 132,267.$$

Как и ожидалось, эта статистика также превышает критическое значение из таблицы значений хи-квадрат-распределения.

Нам также интересно найти предельную склонность к потреблению (MPC, Marginal Propensity to Consume). В расширенной модели гипотеза $H_0: \gamma = 1$ проверяет, что предельная склонность к потреблению постоянна, а не равна 1. (Для проверки гипотезы о равенстве 1 требуется провести совместный тест $\gamma = 1$ и $\beta = 1$.) Для данной модели предельная склонность к потреблению есть функция

$$MPC = dC/dY = \beta\gamma Y^{\gamma-1},$$

зависящая от Y . Для проверки гипотезы, что она равна 1, нам требуется конкретное значение Y . Выберем последнее значение $DPI_{2000,4} = 6634,9$. Для этого значения MPC оценивается как 1,08264. С помощью дельта-метода оценим ее стандартную ошибку как квадратный корень из

$$\begin{aligned} & [\partial MPC/\partial b \quad \partial MPC/\partial c] \begin{bmatrix} Var[b] & Cov[b,c] \\ Cov[b,c] & Var[c] \end{bmatrix} \begin{bmatrix} \partial MPC/\partial b \\ \partial MPC/\partial c \end{bmatrix} = \\ & = [cY^{c-1} \quad bY^{c-1}(1 + c \ln Y)] \begin{bmatrix} 0,000119037 & -0,000131491 \\ -0,000131491 & 0,00014532 \end{bmatrix} \times \\ & \times \begin{bmatrix} cY^{c-1} \\ bY^{c-1}(1 + c \ln Y) \end{bmatrix} = 0,00007469, \end{aligned}$$

что приводит к стандартному отклонению 0,0086423. Для проверки гипотезы, что МРС равна 1 в IV квартале 2000 г., обратимся к значению $z = (1,08264 - 1)/0,0086423 = 9,56299$ в таблице стандартного нормального распределения. Разница статистически значима, поэтому мы отвергаем гипотезу.

Пример 7.5. Преобразование Бокса–Кокса

Преобразование Бокса–Кокса (Box–Cox transformation) [Box, Cox (1964), Zarembka (1974)] используется как инструмент для обобщения линейной модели. Преобразование имеет вид

$$x^{(\lambda)} = (x^\lambda - 1)/\lambda.$$

Особый интерес представляет случай $\lambda = 1$, приводящий к линейному преобразованию $x^{(1)} = x - 1$, а также случай $\lambda = 0$. При λ , равном нулю, преобразование по правилу Лопиталя имеет вид

$$\lim_{\lambda \rightarrow 0} \frac{x^\lambda - 1}{\lambda} = \lim_{\lambda \rightarrow 0} \frac{d(x^\lambda - 1)/d\lambda}{1} = \lim_{\lambda \rightarrow 0} x^\lambda \times \ln x = \ln x.$$

Регрессионный анализ может быть проведен *условно* по λ . Для заданного значения λ модель

$$y = \alpha + \sum_{k=2}^K \beta_k x_k^{(\lambda)} + \varepsilon \quad (7-23)$$

является линейной регрессией и может быть оценена методом наименьших квадратов. Однако если λ в (7-23) является неизвестным параметром, то регрессия становится нелинейной по параметрам.

Вообще говоря, каждый регрессор можно преобразовывать с помощью своего значения λ , но в большинстве приложений это приводит к тому, что обобщение модели становится слишком громоздким и обычно λ предполагается единым для всех переменных в модели⁵. Для того чтобы быть определенным для любого значения λ , x должен быть строго положительным. В большинстве приложений некоторые регрессоры — например дамми-переменные — не преобразуются. Для таких переменных, скажем обозначенных как ν_k , $\nu_k^{(\lambda)} = \nu_k$, а соответствующие производные в (7-24) равны нулю. Также возможно преобразовать y , скажем $y^{(\theta)}$. Заметим, что преобразование зависимой переменной касается спецификации модели в целом, а не только функциональной формы условного среднего. Например, $\theta = 1$ соответствует линейному уравнению, в то время как $\theta = 0$ логарифмическому.

В некоторых приложениях такое преобразование возникает из необходимости рассматривать нулевые значения в логлинейной модели. Кейвс, Кристенсен и Трезевэй (Caves, Christensen, Trethaway (1980)) исследовали издержки производства для железнодорожных компаний, осуществляющих пассажирские и грузовые перевозки. Следуя мнению, изложенному в литературе по издержкам производства в естественных монополиях (regulated

⁵Seaks, Layson (1983).

industries), транслогарифмическая функция цены (см. подраздел 10.4.2) является естественной для моделирования выпуска неоднородной продукции. Однако некоторые из исследуемых фирм не предоставляли услуги пассажирских перевозок, что препятствует использованию транслогарифмической модели. (В этом случае потребуются брать логарифм нуля.) Выходом может стать преобразование Бокса–Кокса, которое может быть вычислено для нулевых значений выпуска продукции. В этой ситуации (равно как и в аналогичных) возникает вопрос, следует ли трактовать нулевой выпуск ровно так же, как ненулевой, или же нулевой выпуск представляет собой отдельное решение компании, отличающееся от остальных уровней выпуска. Кроме этого, как можно видеть из (7-24), это лишь частичное решение проблемы. Нулевые значения регрессоров препятствуют получению соответствующих стандартных отклонений.

Оценка методом наименьших квадратов получается непосредственно. В большинстве случаев мы ожидаем получить оценку λ методом наименьших квадратов между -2 и 2 . Поэтому обычно λ оценивается путем исследования этого интервала в поисках значения, которое минимизировало бы сумму квадратов остатков. Обсудим, что происходит, если в выборке встречаются нулевые значения x . В этом случае на λ в модели потребуются наложить ограничения, так как $0^{(\lambda)}$ определен лишь для строго положительных λ . Положительность значения λ , вообще говоря, не гарантируется в модели. После нахождения оптимального значения оценки λ методом наименьших квадратов оценка среднего квадрата остатков и найденное значение λ представляют собой нелинейные оценки параметров.

После определения оптимального значения λ ее иногда считают известной величиной для получения результатов оценивания методом наименьших квадратов. Однако $\hat{\lambda}$ является оценкой неизвестного параметра. Несложно показать, что стандартные отклонения, вычисленные по методу наименьших квадратов, всегда будут недооценивать истинные асимптотические стандартные отклонения⁶. Чтобы узнать истинные значения, нам потребуется вычислить производные правой части (7-23) по α , β и λ . Псевдорегрессоры равны

$$\begin{aligned}\frac{\partial h(\cdot)}{\partial \alpha} &= 1, \\ \frac{\partial h(\cdot)}{\partial \beta_k} &= x_k^{(\lambda)}, \\ \frac{\partial h(\cdot)}{\partial \lambda} &= \sum_{k=1}^K \beta_k \frac{\partial x_k^{(\lambda)}}{\partial \lambda} = \sum_{k=1}^K \beta_k \left[\frac{1}{\lambda} \left(x_k^\lambda \ln x_k - x_k^{(\lambda)} \right) \right].\end{aligned}\tag{7-24}$$

Теперь мы можем определить асимптотическую ковариационную матрицу оценок параметров с помощью (7-15) и (7-16). Заметим, что $\ln x_k$ присутствует в $\partial h(\cdot)/\partial \lambda$. Если $x_k = 0$, то эта матрица не может быть вычислена. Именно об этом моменте говорилось ранее.

Важно помнить, что коэффициенты в нелинейной модели не равны коэффициентам наклона (или эластичностям) переменных. Для конкретной

⁶Fomby, Hill, Johnson (1984, p. 426–431).

модели Бокса–Кокса $\ln Y = \alpha + \beta X^{(\lambda)} + \varepsilon$ получаем

$$\frac{\partial E[\ln y|\mathbf{x}]}{\partial \ln X} = x \frac{\partial E[\ln y|\mathbf{x}]}{\partial X} = \beta x^\lambda = \eta.$$

Стандартная ошибка этой оценки может быть получена с помощью **дельта-метода**. Производные равны соответственно $\partial\eta/\partial\beta = x^\lambda = \eta/\beta$ и $\partial\eta/\partial\lambda = \eta \ln X$. Приводя подобные члены, мы получаем

$$\begin{aligned} \text{Asy.Var}[\hat{\eta}] &= (\eta/\beta)^2 \left\{ \text{Asy.Var}[\hat{\beta}] + (\beta \ln x)^2 \text{Asy.Var}[\hat{\lambda}] + \right. \\ &\quad \left. + (2\beta \ln x) \text{Asy.Cov}[\hat{\beta}, \hat{\lambda}] \right\}. \end{aligned}$$

Случай, рассмотренный в примере 7.4, относится к моделям Бокса–Кокса такого вида. Мы можем переписать (7-23) в виде

$$\begin{aligned} y &= (\alpha - 1/\lambda) + (\beta/\lambda)X^\lambda + \varepsilon = \\ &= \alpha^* + \beta^*x^\gamma + \varepsilon. \end{aligned}$$

Мы видим, что одним из способов работы с моделью регрессии Бокса–Кокса может быть преобразование этой модели в модель нелинейной регрессии и использование регрессии Гаусса–Ньютона (см. подраздел 7.2.6) для оценки параметров. Искомые параметры восстанавливаются следующим образом: $\lambda = \gamma$, $\alpha = \alpha^* + 1/\gamma$ и $\beta = \gamma\beta^*$.

Пример 7.6. Эффекты взаимодействия в логлинейной модели дохода

Одним из недавних исследований в области экономики здравоохранения является работа «Incentive Effects in the Demand for Health Care: A Bivariate Panel Count Data Estimation» (Riphahn, Wambach, Million (2003)). Авторы интересуют число посещений врача и больниц, а также влияние наличия частной страховки на число посещений, т.е. свидетельствуют ли данные о существовании риска недобросовестного поведения. Выборка представляет собой несбалансированную панель по 7293 домохозяйствам из источника German Socioeconomic Panel (GSOEP)⁷. Среди переменных панели присутствуют такие показатели, как доход домохозяйства (Income), несколько социодемографических показателей, таких как возраст (Age), пол (Male или Female) и образование (Education). В данном примере мы будем моделировать распределение дохода, используя последние доступные данные (1988 г.), точнее, межобъектные данные, включающие 4483 наблюдения. Два индивида в этой выборке сообщили о нулевом доходе, что несовместимо с моделями, которые мы собираемся использовать ниже. Удаляя эти два наблюдения, мы получаем выборку из 4481 наблюдения. Рисунки 7.1 и 7.2 представляют собой гистограмму и

⁷Данные опубликованы на сайте *Journal of Applied Econometrics* по адресу <http://qed.econ.queensu.ca/jae/2003-v18.4/riphahn-wambach-million/>. Переменные, используемые в файле, перечислены в таблице приложения F7.1. Количество опросов одного домохозяйства отличается по годам и составляет от одного до семи раз в год. Общее число наблюдений равно 27 326. Мы будем использовать эти данные несколько раз по ходу изложения книги.

оценку ядерной плотности для дохода домохозяйств, построенные по этим данным.

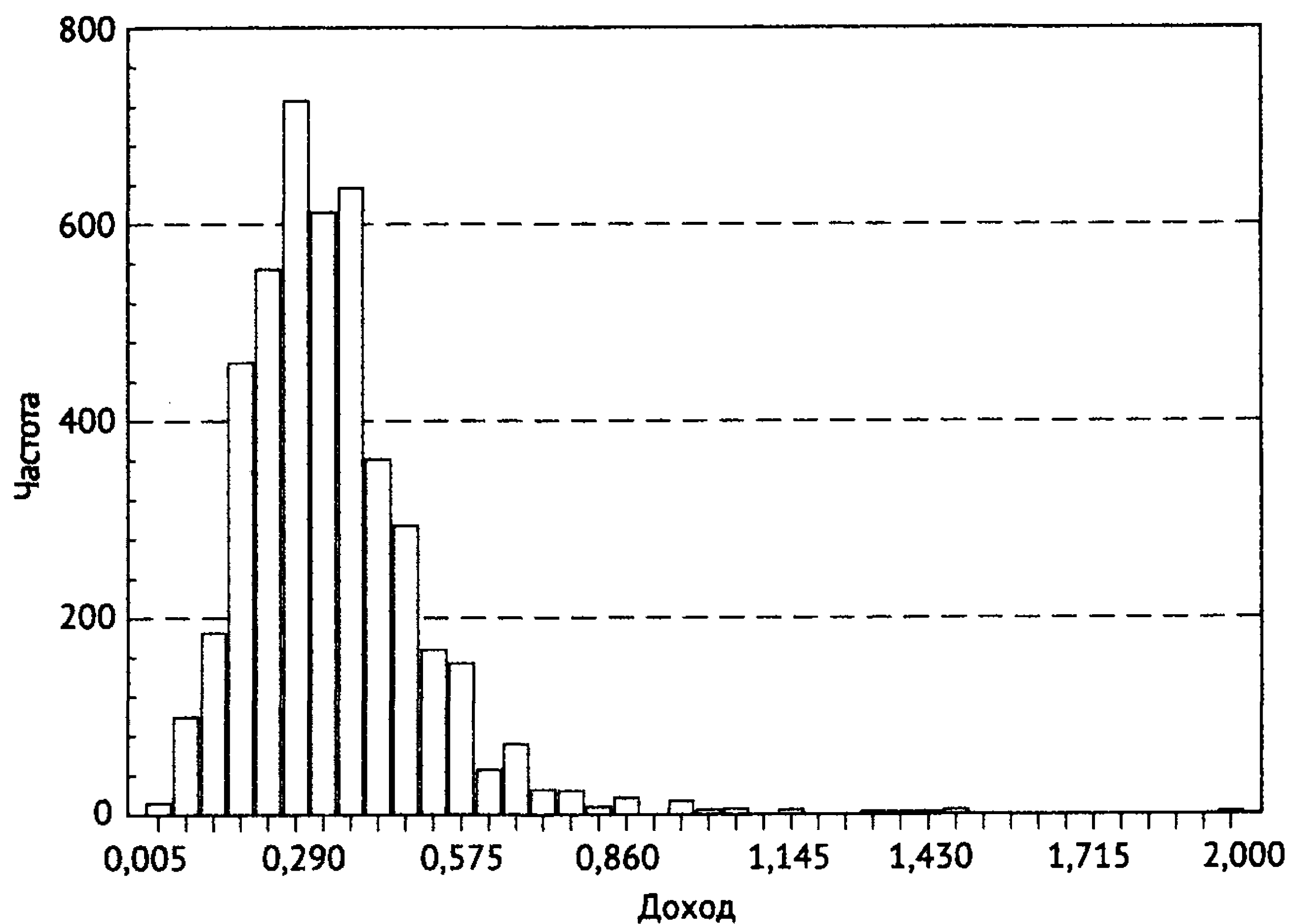


Рис. 7.1. Гистограмма дохода

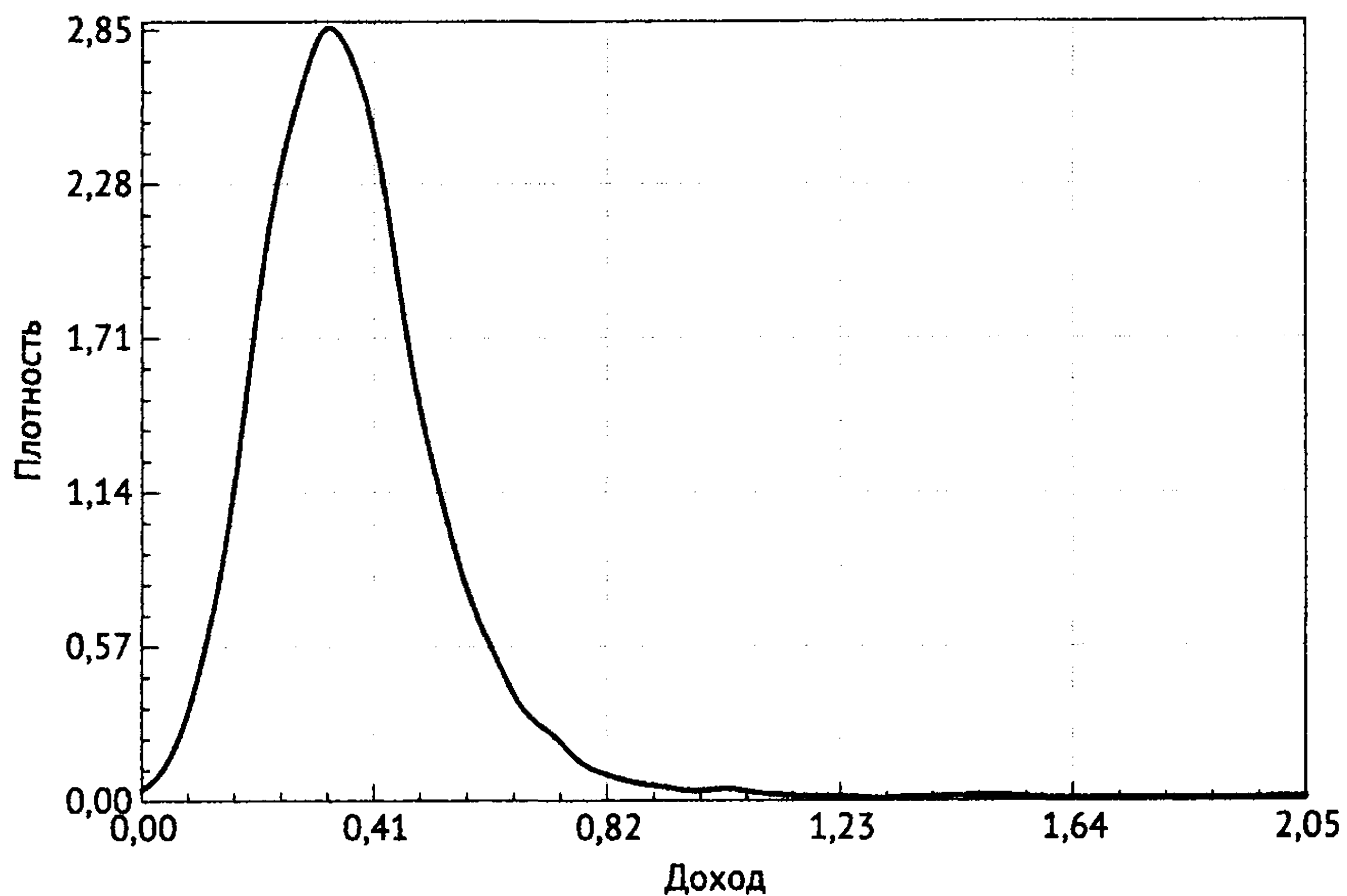


Рис. 7.2. Оценка ядерной плотности дохода

Для объяснения переменной дохода (Income) мы используем экспоненциальную модель регрессии вида

$$\begin{aligned} \text{Income} = & \exp(\beta_1 + \beta_2 \text{Age} + \beta_3 \text{Age}^2 + \beta_4 \text{Education} + \beta_5 \text{Female} + \\ & + \beta_6 \text{Female} \times \text{Education} + \beta_7 \text{Age} \times \text{Education}) + \varepsilon. \end{aligned}$$

В табл. 7.2 приведены описательные статистики для переменных, используемых в этом примере.

Таблица 7.2. Описательные статистики для переменных, используемых в нелинейной регрессии

| Величина | Среднее | Станд. откл. | Минимум | Максимум |
|----------|----------|--------------|---------|----------|
| INCOME | 0,348896 | 0,164054 | 0,0050 | 2 |
| AGE | 43,4452 | 11,2879 | 25 | 64 |
| EDUC | 11,4167 | 2,36615 | 7 | 18 |
| FEMALE | 0,484267 | 0,499808 | 0 | 1 |

Логлинейные модели играют заметную роль в статистике. Многие модели выводятся из функции плотности вида $f(y|\mathbf{x}) = p[y|\alpha^0 + \mathbf{x}'\beta, \theta]$, где α^0 есть постоянная, θ — некоторый параметр и

$$E[y|\mathbf{x}] = g(\theta) \exp(\alpha^0 + \mathbf{x}'\beta)$$

(откуда и происходит название «логлинейные модели»). Сюда относятся такие модели, как модели Вейбулла, гамма, логнормальная и экспоненциальная для случая непрерывных величин, а также пуассоновская и отрицательно биномиальная модели для целочисленных величин. Мы можем записать $E[y|\mathbf{x}]$ как $\exp[\ln g(\theta) + \alpha^0 + \mathbf{x}'\beta]$, затем включить $\ln g(\theta)$ в постоянную в $\ln E[y|\mathbf{x}] = \alpha + \mathbf{x}'\beta$. Логнормальное распределение (см. раздел В.4.4) часто используется при моделировании дохода. Для логнормальной случайной величины

$$p[y|\alpha^0 + \mathbf{x}'\beta, \theta] = \frac{\exp[-\frac{1}{2}(\ln y - \alpha^0 - \mathbf{x}'\beta)^2/\theta^2]}{\theta y \sqrt{2\pi}}, \quad y > 0,$$

$$E[y|\mathbf{x}] = \exp(\alpha^0 + \mathbf{x}'\beta + \theta^2/2) = \exp(\alpha + \mathbf{x}'\beta).$$

Экспоненциальная регрессионная модель также получается при использовании гамма-распределения. Плотность величины, распределенной по гамма-распределению, задается как

$$p[y|\alpha^0 + \mathbf{x}'\beta, \theta] = \frac{\lambda^\theta \exp(-\lambda y) y^{\theta-1}}{\Gamma(\theta)}, \quad y > 0, \theta > 0, \lambda = \exp(-\alpha^0 - \mathbf{x}'\beta),$$

$$E[y|\mathbf{x}] = \theta/\lambda = \theta \exp(\alpha^0 + \mathbf{x}'\beta) = \exp(\ln \theta + \alpha^0 + \mathbf{x}'\beta) = \exp(\alpha + \mathbf{x}'\beta).$$

Параметр θ определяет форму распределения. При $\theta > 2$ плотность гамма-распределения имеет форму хи-квадрат (которая является частным случаем). Наконец, модель Вейбулла имеет схожую форму:

$$p[y|\alpha^0 + \mathbf{x}'\beta, \theta] = \theta \lambda^\theta \exp[-(\lambda y)^\theta] y^{\theta-1}, \quad y \geq 0, \quad \theta > 0, \quad \lambda = \exp(-\alpha^0 - \mathbf{x}'\beta),$$

$$E[y|\mathbf{x}] = \Gamma(1 + 1/\theta) \exp(\alpha^0 + \mathbf{x}'\beta) = \exp[\ln \Gamma(1 + 1/\theta) + \alpha^0 + \mathbf{x}'\beta] = \exp(\alpha + \mathbf{x}'\beta).$$

Во всех случаях наиболее эффективной является оценка методом максимального правдоподобия. (Оценка параметров этой модели методом максимального правдоподобия рассмотрена в главе 14.) Тем не менее оценка нелинейным методом наименьших квадратов

$$E[y|\mathbf{x}] = \exp(\alpha + \mathbf{x}'\beta) + \varepsilon$$

имеет то преимущество, что она будет состоятельной, даже если предположение о виде распределения было ошибочным — она является *устойчивой (робастной)* к этому типу неправильной спецификации модели, так как не использует предположение о виде распределения в явной форме.

Таблица 7.3. Оцененные уравнения регрессии

| Переменная | Нелинейный метод наименьших квадратов | | | Линейный метод наименьших квадратов | | |
|------------------|--|------------------|--------|--|------------------|--------|
| | Оценка | Станд. ошибка | t | Оценка | Станд. ошибка | t |
| Constant | -2,58070 | 0,17455 | 14,78 | -0,13050 | 0,06261 | -2,08 |
| Age | 0,06020 | 0,00615 | 9,79 | 0,01791 | 0,00214 | 8,37 |
| Age ² | -0,00084 | 0,00006082 | -13,83 | -0,00027 | 0,00001985 | -13,51 |
| Education | -0,00616 | 0,01095 | -0,56 | -0,00281 | 0,00418 | -0,67 |
| Female | 0,17497 | 0,05986 | 2,92 | 0,07955 | 0,02339 | 3,40 |
| Female × Educ | -0,01476 | 0,00493 | -2,99 | -0,00685 | 0,00202 | -3,39 |
| Age × Educ | 0,00134 | 0,00024 | 5,59 | 0,00055 | 0,00009394 | 5,88 |
| $e'e$ | | 106,09825 | | | 106,24323 | |
| s | | 0,15387 | | | 0,15410 | |
| R^2 | | 0,12005 | | | 0,11880 | |

В табл. 7.3 представлены результаты оценки нелинейной регрессии методом наименьших квадратов. На первый взгляд знаки и значимость вполне ожидаемы, кроме дамми-переменной для женщин. Тем не менее два момента затрудняют интерпретацию коэффициентов этой модели. Во-первых, модель нелинейна, поэтому коэффициенты не дают нам величину интересующего нас влияния факторов в уравнении. В частности, для этой модели

$$\partial E[y|\mathbf{x}]/\partial x_k = \exp(\alpha + \mathbf{x}'\beta) \times \partial(\alpha + \mathbf{x}'\beta)/\partial x_k.$$

Во-вторых, по построению модели вторая часть производной не равна коэффициенту, потому что величины входят в уравнение не только линейно, но и в квадрате или в составе произведения с другой величиной. Более того, для дамми-переменной *Female* мы бы хотели посчитать частное влияние как

$$\Delta E[y|\mathbf{x}]/\Delta Female = E[y|\mathbf{x}, Female = 1] - E[y|\mathbf{x}, Female = 0].$$

Третий вопрос, который возникает, связан с тем, как считать частные влияния: по выборочным средним или по средним значениям переменных? Например,

$$\partial E[y|\mathbf{x}]/\partial Age = E[y|\mathbf{x}] \times (\beta_2 + 2\beta_3 Age + \beta_7 Educ).$$

Среднее значение для *Age* в выборке составляет 43,4452, а для *Education* — 11,4167. Частное влияние одного года образования оценивается как 0,000948, если его считать, усредняя частные влияния, вычисленные для каждого индивида. Оно же оценивается как 0,000925, если вычислять условное среднее и линейную составляющую в средних трех переменных. Без информации о размерности величины дохода частное влияние трудно интерпретировать. Так как средний доход по выборке составляет около 0,35, подобные частные влияния означают, что дополнительный год образования связан с изменением ожидаемого дохода примерно на 2,6% (т.е. 0,009/0,35).

Грубый подсчет частного влияния *Age* не обнаруживает связи в модели между возрастом и ожидаемым доходом. Так, заметим, что коэффициент при *Age* положительный, в то время как коэффициент при Age^2 отрицательный. Отсюда следует (если игнорировать компоненту взаимодействия в конце уравнения), что связь *Age*–*Income*, определяемая моделью, является параболической. Частное влияние положительно при низких значениях возраста (*Age*) и отрицательно при высоких. Чтобы подробнее изучить этот момент, мы посчитаем ожидаемый *Income* отдельно для мужчин и для женщин, предполагая, что они закончили колледж ($Educ = 16$), для выборки с 25 до 64 лет. Результат изображен на рис. 7.3. Верхняя кривая относится к мужчинам ($Female = 0$), а нижняя — к женщинам. Как и ожидалось, форма является параболической; рисунок демонстрирует относительно сильное влияние: при прочих равных предполагается, что доход возрастает до 80% между 25 и 64 годами. (Важным для этого вычисления остается вопрос о том, что автор модели хочет проанализировать. Остается невыясненным, является ли эта параболическая связь траекторией ожидаемого дохода для каждого индивида по мере того, как меняется его возраст, или это средние доходы различных возрастных групп в конкретный момент (1988). Последнее кажется более разумным, судя по данному рисунку, хотя можно придумать объяснение и первой теории.)

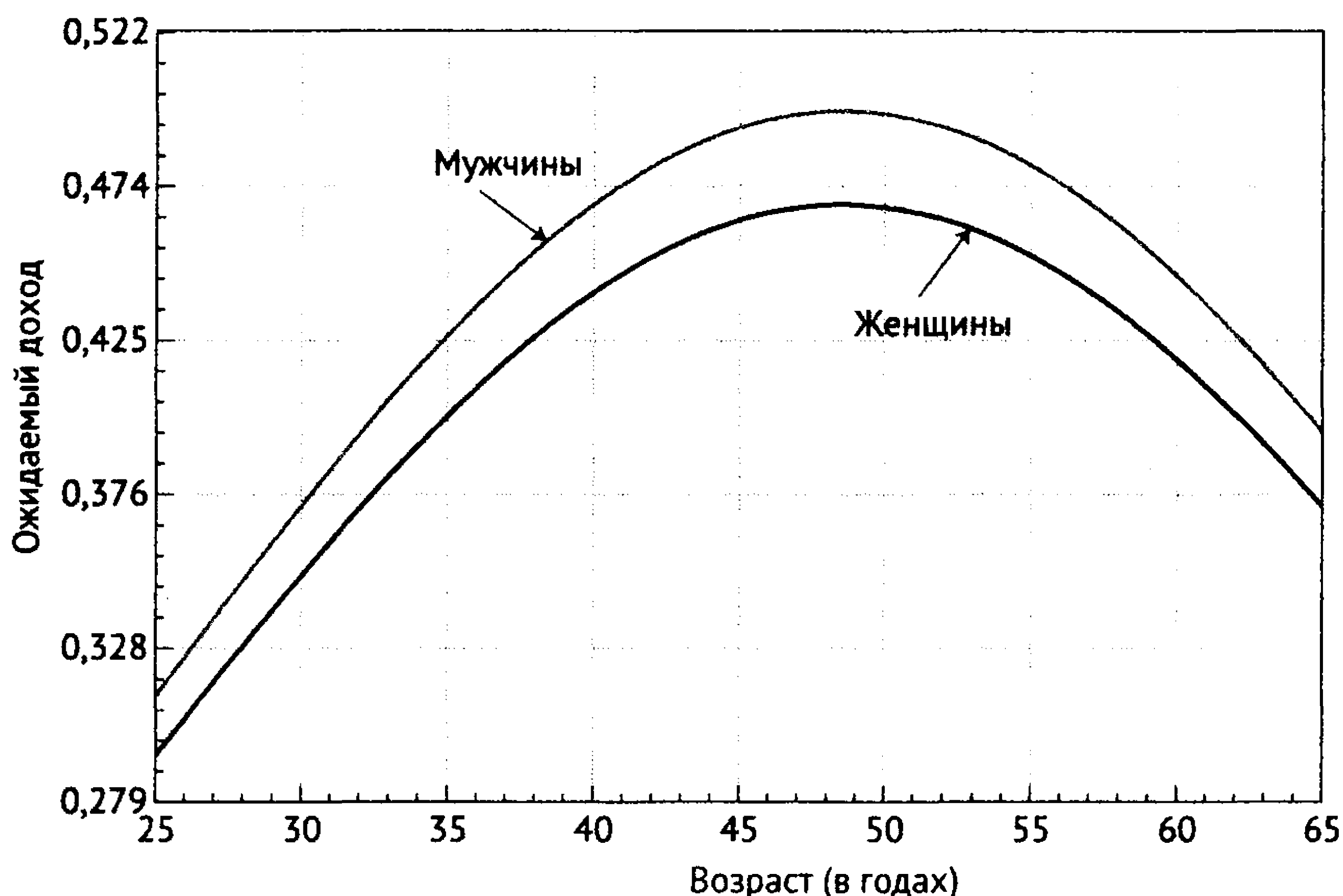


Рис. 7.3. Ожидаемые доходы

Рисунок демонстрирует также другой вывод из оцененной модели, который неочевиден из результатов регрессии. Коэффициент при дамми-переменной *Female* положителен, высокосignificant и является наибольшим по модулю в модели. Это может привести к выводу, что в среднем ожидаемый доход по этой выборке выше у женщин, чем у мужчин. Но рис. 7.3 явно демонстрирует обратное. Это различие объясняется компонентой взаимодействия, *Female* × *Education*. Отрицательный знак последнего члена порождает сомнения в положительности влияния. Однако знак совокупного влияния остается неясным без дополнительного анализа, предложенного ниже.

Наконец, кроме возраста (*Age*), входящего в квадрате, в модель включена компонента взаимодействия *Age* × *Education*. Коэффициент является положительным и высокосignificant. Однако его интерпретация совершенно неочевидна. В линейной модели

$$\begin{aligned} \text{Income} = & \beta_1 + \beta_2 \text{Age} + \beta_3 \text{Age}^2 + \beta_4 \text{Education} + \beta_5 \text{Female} \\ & + \beta_6 \text{Female} \times \text{Education} + \beta_7 \text{Age} \times \text{Education} + \varepsilon \end{aligned}$$

мы бы сказали, что $\beta_7 = \partial^2 E[\text{Income}|x] / \partial \text{Age} \partial \text{Education}$. То есть «эффект взаимодействия» есть изменение частного влияния *Age* при изменении *Education* (и наоборот). Конечно, если коэффициент β_7 равен нулю, т.е. это произведение отсутствует в модели, то нет никакого эффекта взаимодействия — вторая производная равна нулю. Тем не менее эта простая интерпретация часто неприменима в нелинейных моделях (т.е. в любой нелинейной модели). Рассмотрим нашу экспоненциальную регрессию и предположим, что на самом деле β_7 равен нулю. Для удобства пусть $\mu(x)$ обозначает функцию условного среднего. Тогда частное влияние *Age* равно

$$\partial \mu(x) / \partial \text{Age} = \mu(x) \times (\beta_2 + 2\beta_3 \text{Age})$$

и

$$\partial^2 \mu(x) / \partial \text{Age} \partial \text{Educ} = \mu(x) \times (\beta_2 + 2\beta_3 \text{Age})(\beta_4 + \beta_6 \text{Female}), \quad (7-25)$$

что не равно нулю, даже когда нет никакой **компоненты взаимодействия** в модели. Влияние взаимодействия в оцененной нами модели, которая содержит произведение, равно

$$\partial^2 E[y|x] / \partial \text{Age} \partial \text{Educ} = \mu(x) \times [\beta_7 + (\beta_2 + 2\beta_3 \text{Age} + \beta_7 \text{Educ})(\beta_4 + \beta_6 \text{Female} + \beta_7 \text{Age})]. \quad (7-26)$$

По крайней мере, часть того, что называется взаимодействием, происходит из-за того, что модель является нелинейной. Для того чтобы отделить «эффект функциональной формы» от настоящего «эффекта взаимодействия», мы можем вычесть (7-25) из (7-26) и перегруппировать члены:

$$\begin{aligned} \partial^2 \mu(x) / \partial \text{Age} \partial \text{Educ} = & \mu(x) [(\beta_2 + 2\beta_3 \text{Age})(\beta_4 + \beta_6 \text{Female})] + \mu(x) \beta_7 \times \\ & \times [1 + \text{Age}(\beta_2 + 2\beta_3) + \text{Educ}(\beta_4 + \beta_6 \text{Female}) + \text{Educ} \times \text{Age}(\beta_7)]. \quad (7-27) \end{aligned}$$

Ясно, что коэффициент при произведении не имеет никакого отношения к нужной нам величине (предполагая, что нас интересует изменение частного влияния). С другой стороны, второй член не равен нулю тогда, и только тогда, когда коэффициент β_7 не равен нулю. Поэтому можно связать вторую часть с «взаимодействием» в модели. Тем не менее вопрос о том, можно ли использовать такую неоднозначную интерпретацию, остается открытым. Более того, в этом случае остается без объяснения эффект функциональной формы. Целью этого упражнения было показать, что к интерпретации взаимодействия в нелинейной модели надо подходить крайне осторожно. Подобный анализ этой проблемы представлен в работе Аи и Нортон (Ai, Norton (2004)). Также существуют некоторые комментарии и обобщения этих результатов, включая работу Грина (Greene (2010)).

Сделаем одно заключительное замечание, касающееся нелинейной регрессии. В логлинейной модели, аналогичной той, которая была здесь рассмотрена, можно было бы для сравнения вычислить обычные оценки методом наименьших квадратов. Коэффициенты в правой части табл. 7.3 на первый взгляд заставляют думать, что нелинейный и линейный методы наименьших квадратов оценивают совершенно различные зависимости. Для того чтобы понять их сходство (если оно есть), полезно обратить внимание на частные влияния, а не на коэффициенты. Так, например, мы вычислили, что частное влияние образования в нелинейной модели при использовании усредненных переменных равно 0,000925. Хотя оценки коэффициента в линейной модели сильно отличаются, вычислив частное влияние образования в линейной модели, мы получаем $-0,00281 - 0,00685(0,5) + 0,00055(43,4452) = 0,01766$ при использовании значения 0,5 для переменной *Female*. Разделив полученный результат на 0,35, получаем 0,0504, что не так уж отличается от аналога в нелинейной модели. Общий результат состоит в том, что оценки коэффициентов линейным методом наименьших квадратов являются приближением.

7.2.6. Вычисление оценки нелинейным методом наименьших квадратов

Минимизация суммы квадратов остатков в нелинейной регрессии — это частая проблема, с которой приходится сталкиваться при нелинейной оптимизации. Она может решаться различными способами. (См. раздел Е.3.) Для этой цели часто используется метод Гаусса–Ньютона. Этот алгоритм основывается на линейном разложении в ряд Тейлора нелинейных регрессионных функций. Оценка получается итеративным способом после преобразования оптимизационной задачи в ряд регрессий линейного метода наименьших квадратов.

Нелинейная регрессионная модель записывается как $y = h(\mathbf{x}, \boldsymbol{\beta}) + \varepsilon$. (Для экономии места мы не пишем индекс номера наблюдения.) Метод основан на приближении $h(\mathbf{x}, \boldsymbol{\beta})$ рядом Тейлора в точке значений вектора параметров $\boldsymbol{\beta}^0$:

$$h(\mathbf{x}, \boldsymbol{\beta}) \approx h(\mathbf{x}, \boldsymbol{\beta}^0) + \sum_{k=1}^K \frac{\partial h(\mathbf{x}, \boldsymbol{\beta}^0)}{\partial \beta_k^0} (\beta_k - \beta_k^0). \quad (7-28)$$

Уравнение в этой форме называется **линеаризированной регрессионной моделью** (linearized regression model). Перегруппировав, получаем

$$h(\mathbf{x}, \boldsymbol{\beta}) \approx \left[h(\mathbf{x}, \boldsymbol{\beta}^0) - \sum_{k=1}^K \beta_k^0 \left(\frac{\partial h(\mathbf{x}, \boldsymbol{\beta}^0)}{\partial \beta_k^0} \right) \right] + \sum_{k=1}^K \beta_k \left(\frac{\partial h(\mathbf{x}, \boldsymbol{\beta}^0)}{\partial \beta_k^0} \right). \quad (7-29)$$

Пусть x_k^0 обозначает k -ю частную производную⁸ $\partial h(\mathbf{x}, \boldsymbol{\beta}^0) / \partial \beta_k^0$. Для данного значения $\boldsymbol{\beta}^0$ x_k^0 — это функция только от данных, но не от неизвестных параметров. В таком случае имеем

$$h(\mathbf{x}, \boldsymbol{\beta}) \approx \left[h^0 - \sum_{k=1}^K x_k^0 \beta_k^0 \right] + \sum_{k=1}^K x_k^0 \beta_k,$$

что можно переписать как

$$h(\mathbf{x}, \boldsymbol{\beta}) \approx h^0 - \mathbf{x}^{0'} \boldsymbol{\beta}^0 + \mathbf{x}^{0'} \boldsymbol{\beta},$$

откуда следует, что

$$y \approx h^0 - \mathbf{x}^{0'} \boldsymbol{\beta}^0 + \mathbf{x}^{0'} \boldsymbol{\beta} + \varepsilon.$$

Переносим линейные члены в левую часть уравнения, мы получаем линейное уравнение

$$y^0 = y - h^0 + \mathbf{x}^{0'} \boldsymbol{\beta}^0 = \mathbf{x}^{0'} \boldsymbol{\beta} + \varepsilon^0. \quad (7-30)$$

Заметим, что ε^0 включает как настоящий шок ε , так и ошибку в разложении Тейлора первого порядка настоящей регрессии, записанной в (7-29), т.е.

⁸Необходимо проверить, что для модели линейной регрессии эти производные являются независимыми величинами.

$$\varepsilon^0 = \varepsilon + \left[h(\mathbf{x}, \boldsymbol{\beta}) - \left\{ h^0 - \sum_{k=1}^K x_k^0 \beta_k^0 + \sum_{k=1}^K x_k^0 \beta_k \right\} \right]. \quad (7-31)$$

Поскольку мы учли все ошибки, то (7-30) представляет собой точное равенство, а не приближение. Имея значение $\boldsymbol{\beta}^0$, мы могли бы вычислить y^0 и \mathbf{x}^0 , а затем оценить параметры (7-30) линейным методом наименьших квадратов. Теперь нам требуется понять, является ли эта оценка состоятельной.

Пример 7.7. Линеаризованная регрессия

Для модели, рассмотренной в примере 7.3, регрессоры в линеаризованном уравнении будут иметь вид

$$x_1^0 = \frac{\partial h(\cdot)}{\partial \beta_1^0} = 1, \quad x_2^0 = \frac{\partial h(\cdot)}{\partial \beta_2^0} = e^{\beta_3^0 x}, \quad x_3^0 = \frac{\partial h(\cdot)}{\partial \beta_3^0} = \beta_2^0 e^{\beta_3^0 x}.$$

Имея набор параметров $\boldsymbol{\beta}^0$, можно было бы оценить линейную регрессию

$$y^0 = y - h(x, \beta_1^0, \beta_2^0, \beta_3^0) + \beta_1^0 x_1^0 + \beta_2^0 x_2^0 + \beta_3^0 x_3^0$$

на три псевдорегрессора и получить оценки коэффициентов β_1, β_2 и β_3 .

Линеаризованная модель регрессии, записанная в (7-30), может быть оценена линейным методом наименьших квадратов. Полученный вектор параметров может играть роль нового значения $\boldsymbol{\beta}^0$, и все вычисления могут быть произведены заново. **Итерации** можно проводить до тех пор, пока различие между векторами параметров не станет достаточно малым, чтобы можно было предположить сходимость. Одно из главных достоинств этого метода заключается в том, что при последней итерации оценка $(\mathbf{Q}^0)^{-1}$ независимо от масштабирующего множителя $\hat{\sigma}^2/n$ будет верной оценкой для асимптотической ковариационной матрицы вектора параметров.

Итерационное решение задачи минимизации есть

$$\begin{aligned} \mathbf{b}_{t+1} &= \left[\sum_{i=1}^n \mathbf{x}_i^0 \mathbf{x}_i^{0'} \right]^{-1} \left[\sum_{i=1}^n \mathbf{x}_i^0 (y_i - h_i^0 + \mathbf{x}_i^{0'} \mathbf{b}_t) \right] \\ &= \mathbf{b}_t + \left[\sum_{i=1}^n \mathbf{x}_i^0 \mathbf{x}_i^{0'} \right]^{-1} \left[\sum_{i=1}^n \mathbf{x}_i^0 (y_i - h_i^0) \right] \\ &= \mathbf{b}_t + (\mathbf{X}^{0'} \mathbf{X}^0)^{-1} \mathbf{X}^{0'} \mathbf{e}^0 \\ &= \mathbf{b}_t + \boldsymbol{\Delta}_t, \end{aligned} \quad (7-32)$$

где все члены в правой части уравнения вычисляются в точке \mathbf{b}_t и \mathbf{e}^0 — вектор остатков, полученных нелинейным методом наименьших квадратов. Этот алгоритм вполне понятен интуитивно. На каждой итерации мы обновляем предыдущий вектор оцененных параметров, регрессируя остатки, полученные нелинейным методом наименьших квадратов, на производные регрессионных функций. Процесс сходится (т.е. обновление будет равно 0), когда $\mathbf{X}^{0'} \mathbf{e}^0$ достаточно близко к 0. Эта производная имеет прямой аналог в нормальных уравнениях для линейной модели $\mathbf{X}' \mathbf{e} = 0$.

Обычно использование вычислительной техники не позволяет добиться полной сходимости, т.е. того, чтобы выражение $\mathbf{X}'\mathbf{e}^0$ было в точности равно нулю. Полезным, не зависящим от масштаба является критерий $\delta = \mathbf{e}^{0'}\mathbf{X}^0(\mathbf{X}^{0'}\mathbf{X}^0)^{-1}\mathbf{X}^{0'}\mathbf{e}^0$, который аналогичен критериям сходимости, рассмотренным в подразделе Е.3.6. [См. (7-22).] Наконец, заметим, что итерационный алгоритм линеаризированной регрессии хоть и является очень эффективным для многих задач, но работает не всегда. Как и метод Ньютона, этот алгоритм может «выпрыгнуть» к сильно разбросанным значениям на второй итерации, после чего может оказаться невозможным вычислить остатки для следующей итерации. Принципиальным здесь может стать выбор начальных значений. Вычисление оценок нелинейным методом наименьших квадратов представляет собой смесь науки и искусства. [McCullough, Vinod (1999).] Если мы не обладаем никакой информацией для выбора начальных значений, рабочей стратегией может стать предварительное использование алгоритма Гаусса–Ньютона. Если этот алгоритм не приводит к ответу, то можно вернуться к стартовым значениям и использовать более общие алгоритмы, такие как BFGS, рассматривая минимизацию суммы квадратов остатков как обычную оптимизационную задачу.

Пример 7.8. Нелинейный метод наименьших квадратов

В примере 7.4 была рассмотрена нелинейная функция потребления

$$C = \alpha + \beta Y^\gamma + \varepsilon.$$

Линеаризированная регрессионная модель записывается как

$$C - (\alpha^0 + \beta^0 Y^{\gamma^0}) + (\alpha^0 1 + \beta^0 Y^{\gamma^0} + \gamma^0 \beta^0 Y^{\gamma^0} \ln Y) = \alpha + \beta(Y^{\gamma^0}) + \gamma(\beta^0 Y^{\gamma^0} \ln Y) + \varepsilon^0.$$

Объединяя слагаемые, получаем, что нелинейный метод оценивания сводится к последовательности регрессий

$$C^0 = C + \gamma^0 \beta^0 Y^{\gamma^0} \ln Y$$

на

$$\mathbf{x}^0 = \left[\frac{\partial h(\cdot)}{\partial \alpha} \quad \frac{\partial h(\cdot)}{\partial \beta} \quad \frac{\partial h(\cdot)}{\partial \gamma} \right]' = \begin{bmatrix} 1 \\ Y^{\gamma^0} \\ \beta^0 Y^{\gamma^0} \ln Y \end{bmatrix}.$$

Нахождение **начальных значений** для нелинейной процедуры может оказаться трудной задачей. Простой выбор произвольных значений может оказаться непродуктивным. К сожалению, не существует хороших правил выбора начальных значений, кроме того, что они должны по возможности не сильно отличаться от конечных (такое правило не особенно нам помогает). Иногда возможно использовать какую-нибудь состоятельную оценку β в качестве начального значения. Однако во многих случаях единственная состоятельная оценка, которая возможна, и является оценкой методом наименьших квадратов, которую мы и пытаемся вычислить. В данном конкретном

случае естественный выбор начальных значений возникает из-за того, что линейная модель будет частным случаем нашей. Таким образом, мы можем вычислить α и β с помощью линейного метода наименьших квадратов, который является частным случаем при $\gamma = 1$, и взять 1 в качестве стартового значения для γ . Итак, **итерации** начинаются с оценок линейным методом наименьших квадратов для α и β и 1 для γ .

Решение достигается после восьми итераций; последующие итерации уже служат «тонкой настройкой» скрытых цифр (т.е. тех, которые исследователь не будет включать в отчет; «градиент» — это не зависящий от масштаба критерий сходимости δ , упомянутый выше). Заметим, что на второй итерации вектор коэффициентов резко «прыгает» — сумма квадратов становится очень большой, но после этого итерационный процесс успокаивается и процедура сходится.

Начинаем NLSQ-итерации. Линеаризированная регрессия.

Итерация = 1; сумма квадратов = 1536321,88; градиент = 996103,930

Итерация = 2; сумма квадратов = 0,184780956E+12;

градиент = 0,184780452E+12 ($\times 10^{12}$)

Итерация = 3; сумма квадратов = 20406917,6; градиент = 19902415,7

Итерация = 4; сумма квадратов = 581703,598; градиент = 77299,6342

Итерация = 5; сумма квадратов = 504403,969; градиент = 0,752189847

Итерация = 6; сумма квадратов = 504403,216; градиент = 0,526642396E-04

Итерация = 7; сумма квадратов = 504403,216; градиент = 0,511324981E-07

Итерация = 8; сумма квадратов = 504403,216; градиент = 0,606793426E-10

7.3. Медианная и квантильная регрессии

Мы придерживаемся основных предположений модели линейной регрессии

$$y = \mathbf{x}'\beta + \varepsilon,$$

где $E[\varepsilon|\mathbf{x}] = 0$ и $E[y|\mathbf{x}] = \mathbf{x}'\beta$. Если величины $\varepsilon|\mathbf{x}$ распределены нормально, то распределение $\varepsilon|\mathbf{x}$ симметрично и медиана $\text{Med}[\varepsilon|\mathbf{x}]$ тоже равна нулю, а $\text{Med}[y|\mathbf{x}] = \mathbf{x}'\beta$. При этих предположениях оценка методом наименьших квадратов остается естественным выбором для оценки β . Однако, как мы видели в примере 4.5, **оценка методом наименьших абсолютных отклонений** (least absolute deviations — LAD) представляет возможную альтернативу, которая может быть даже предпочтительнее. Предположим, что мы исходим непосредственно из второго предположения. В этом случае постановка модели выглядит как

$$\text{Med}[y|\mathbf{x}] = \mathbf{x}'\beta.$$

Здесь LAD-оценка, устойчивая к выбросам, возникает скорее сама по себе, чем как альтернатива методу наименьших квадратов⁹. Условная медиана $y_i|x_i$ вполне может быть функцией, представляющей самостоятельный интерес. В более широком смысле другие квантили распределения $y_i|x_i$ могут

⁹В примере 4.5 мы видели, что на малых выборках при распределениях с тяжелыми хвостами дисперсия LAD-оценки может быть меньше, чем оценки методом наименьших квадратов.

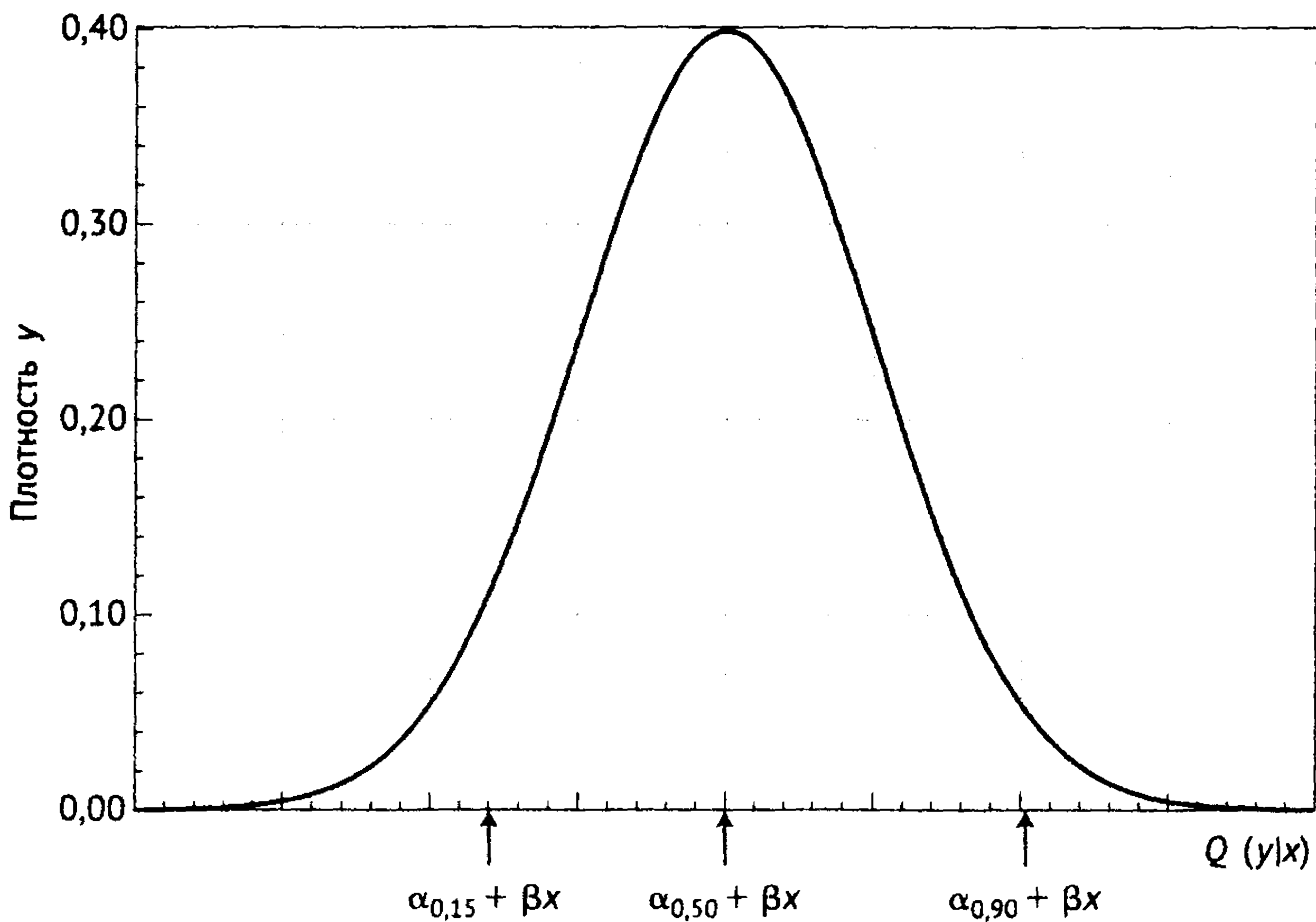


Рис. 7.4. Модель квантильной регрессии

также оказаться заслуживающими отдельного внимания. Например, нас может интересовать изучение различных квантилей распределения доходов (или расходов). В этом случае предпочтительнее использовать квантильную регрессию. (Линейная) модель квантильной регрессии может быть определена как

$$Q[y|x, q] = \mathbf{x}'\beta_q \text{ такая, что } \text{Prob}[y \leq \mathbf{x}'\beta_q|x] = q, 0 < q < 1. \quad (7-33)$$

Медианная регрессия получается при $q = \frac{1}{2}$. Другие интересные случаи — это нижняя и верхняя четверти $q = \frac{1}{4}$ и $q = \frac{3}{4}$ соответственно. Подробно медианная регрессия будет рассматриваться в подразделе 7.3.1 в основном как альтернатива модели линейной регрессии.

Модель квантильной регрессии относится к более широкому классу моделей, чем изученная нами модель линейной регрессии, так как коэффициенты в (7-33) имеют индекс q . Модель является непараметрической — для нее требуется куда менее подробно специфицировать распределение $y|x$. В простейшей модели линейной регрессии с фиксированными коэффициентами β квантили $y|x$ будут определяться разбросом постоянного члена. Этот вывод продемонстрирован на рис. 7.4. Для фиксированного β условно по x значение $\alpha_q + \beta x$ такое, что $\text{Prob}(y < \alpha_q + \beta x)$ вычислено для значений $q = 0,15, 0,5$ и $0,9$. Для любого квантиля существует соответствующее значение α_q . В подразделе 7.3.2 нами будет исследована более общая спецификация квантильной регрессии, в которой роль α_q на рис. 7.4 играет весь вектор коэффициентов.

7.3.1. Оценка методом наименьших абсолютных отклонений

Результаты оценки методом наименьших квадратов могут быть сильно искажены наблюдениями-выбросами. Так, современные исследования в микроэкономике и экономике финансов, включающие распределения с тяжелыми хвостами, достаточно часто содержат подобные наблюдения. (Разумеется, это не относится к исследованиям в финансовой сфере, в которых исследователь зачастую располагает сотнями тысяч наблюдений, что становится обычным для подобных исследований.) В таких исследованиях ставится вопрос о поиске «устойчивой» оценки, которая бы не подвергалась такому искажению из-за выбросов¹⁰. В этом подразделе мы изучим одну из таких оценок — оценку методом наименьших абсолютных отклонений, или LAD-оценку.

Из-за того что метод наименьших квадратов приписывает такой большой вес сильным отклонениям от линии регрессии, результаты регрессии очень чувствительны к наличию нетипичных наблюдений, если размер выборки мал или ограничен. Оценка методом наименьших абсолютных отклонений (LAD) была предложена как вариант, позволяющий снизить (хотя бы до определенного уровня) этот эффект. LAD-оценка определяется как решение оптимизационной задачи

$$\text{Min}_{\mathbf{b}_0} \sum_{i=1}^n |y_i - \mathbf{x}'_i \mathbf{b}_0|.$$

История LAD-оценки начинается раньше истории оценки методом наименьших квадратов (которая в свою очередь насчитывает более 200 лет). Редкость ее использования в эконометрике объясняется тем, что метод Гаусса (LS) вытеснил метод LAD; LS значительно проще для подсчета. Кроме того, в скором времени статистические свойства LS были изучены лучше, чем свойства оценки LAD, а выборки достаточно велики, и преимущества LAD-оценки на малых выборках оказались не востребуемыми.

Оценка LAD есть частный случай квантильной регрессии

$$\text{Prob}[y_i \leq \mathbf{x}'_i \boldsymbol{\beta}_q] = q.$$

Она оценивает *медианную регрессию*, т.е. является решением квантильной регрессии при $q = 0,5$. Анализ свойств этой регрессии посвящены работы (Koenker, Bassett (1978, 1982), Huber (1967) и Rogers (1993))¹¹. Их результаты представляют оценку асимптотической ковариационной матрицы для оценки квантильной регрессии вида

$$\text{Est. Asy. Var}[\mathbf{b}_q] = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{D}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1},$$

¹⁰Некоторые приложения описаны в работах: Taylor (1974), Amemiya (1985, p. 70–80), Andrews (1974), Koenker and Bassett (1978); простой для понимания обзор представлен в работе (Birkes, Dodge (1993)). Несколько более строгое изложение приведено в работе (Hardle (1990)).

¹¹В работе (Powell (1984)) оценка LAD была обобщена для получения свойства устойчивости в случае цензурированных выборок по зависимой переменной, т.е. в ситуации, когда отрицательные значения y_i записываются как нули. В качестве примера см. (Melenberg, van Soest (1996)). Похожие результаты, касающиеся других полупараметрических подходов, представлены в работах (Butler et al. (1990)) и (McDonald, White (1993)).

где \mathbf{D} — диагональная матрица, содержащая веса

$$d_i = \left[\frac{q}{f(0)} \right]^2, \text{ если } y_i - \mathbf{x}'_i \boldsymbol{\beta} \text{ больше нуля, и } \left[\frac{1-q}{f(0)} \right]^2, \text{ если иначе,}$$

а $f(0)$ — истинная плотность шоков, вычисленная в нуле¹². [Осталось лишь получить оценку величины $f(0)$.] В этом результате есть полезная симметрия. Предположим, что истинная плотность является нормальной с дисперсией σ^2 . Тогда предыдущее выражение сокращается до $\sigma^2(\pi/2)(\mathbf{X}'\mathbf{X})^{-1}$ и представляет собой результат, который мы использовали в примере 4.5. Для более общих случаев потребуются какие-то эмпирические оценки $f(0)$. Например, можно использовать непараметрические методы оценки плотности [см. раздел 12.4 и, например (Johnston, DiNardo (1997, p. 370–375))]. Однако в маленьких выборках, для которых подобные методы наиболее интересны (пример ниже насчитывает 25 наблюдений), непараметрическое ядерное оценивание единственного значения функции представляется слишком оптимистичным; в конечном счете подобные результаты являются асимптотическими. Однако асимптотически, как показывает пример 4.5, метод наименьших квадратов значительно предпочтительнее. Так или иначе удобной может стать **ядерная оценка плотности**, описанная в подразделе 12.4.1. Забегая вперед, заметим, что она будет вычисляться как

$$\hat{f}(0) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K \left[\frac{e_i}{h} \right],$$

где h — **ширина окна** (мы обсудим это чуть ниже), $K[\cdot]$ — взвешивающая, или ядерная, функция, а e_i , $i = 1, \dots, n$ — остатки. Не существует четких и ясных правил для выбора h ; один из популярных вариантов использует пакет Stata (2006), $h = 0,9s/n^{1/5}$. В выборе ядерной функции также наблюдается некоторый произвол, хотя редко возникают случаи, при которых важно, какую именно функцию выбрать; часто выбирают логистическое ядро (см. табл. 12.2).

Для подобных случаев хорошо подходит метод **бутстрэпа (самонастройки)**, используемый для получения статистических выводов. Поскольку эффективность бутстрэпа в подобных случаях хорошо известна, нам не требуется формула для стандартных ошибок LAD-оценки. Оценка асимптотической ковариационной матрицы методом бутстрэпа может быть вычислена следующим образом:

$$\text{Est. Var}[\mathbf{b}_{LAD}] = \frac{1}{R} \sum_{r=1}^R (\mathbf{b}_{LAD}(r) - \mathbf{b}_{LAD})(\mathbf{b}_{LAD}(r) - \mathbf{b}_{LAD})',$$

где \mathbf{b}_{LAD} — оценка LAD и $\mathbf{b}_{LAD}(r)$ — r -я оценка LAD-величины $\boldsymbol{\beta}$, основанная на n наблюдениях, полученных с помощью выборки с возвращением из начальных данных.

¹²Кознкер (Koenker) предположил, что в случае независимых одинаково распределенных наблюдений нужно заменить d_i на константу $a = q(1-q)/[f(F^{-1}(q))]^2 = [0,50/f(0)]^2$ для медианной (LAD) оценки. В этом случае выражение превращается в истинную асимптотическую ковариационную матрицу $a(\mathbf{X}'\mathbf{X})^{-1}$. Предложенная оценка будет вести себя аналогично и на больших выборках. (Из личной беседы с автором.)

Пример 7.9. Оценка LAD производственной функции Кобба–Дугласа

В работе Зельнера и Реванкара (Zellner, Revankar (1970)) было предложено обобщение производственной функции Кобба–Дугласа, которое позволяет отдаче от масштаба меняться в зависимости от выпуска. Используемые ими данные для разных штатов по показателям $Y = \text{value added (output)}$ (добавленная стоимость, выпуск), $K = \text{capital}$ (капитал), $L = \text{labor}$ (труд) и $N = \text{число сотрудников в транспортной отрасли}$ приведены в табл. F7.2 приложения.

$$\ln(Y_i/N_i) = \beta_1 + \beta_2 \ln(K_i/N_i) + \beta_3 \ln(L_i/N_i) + \varepsilon_i$$

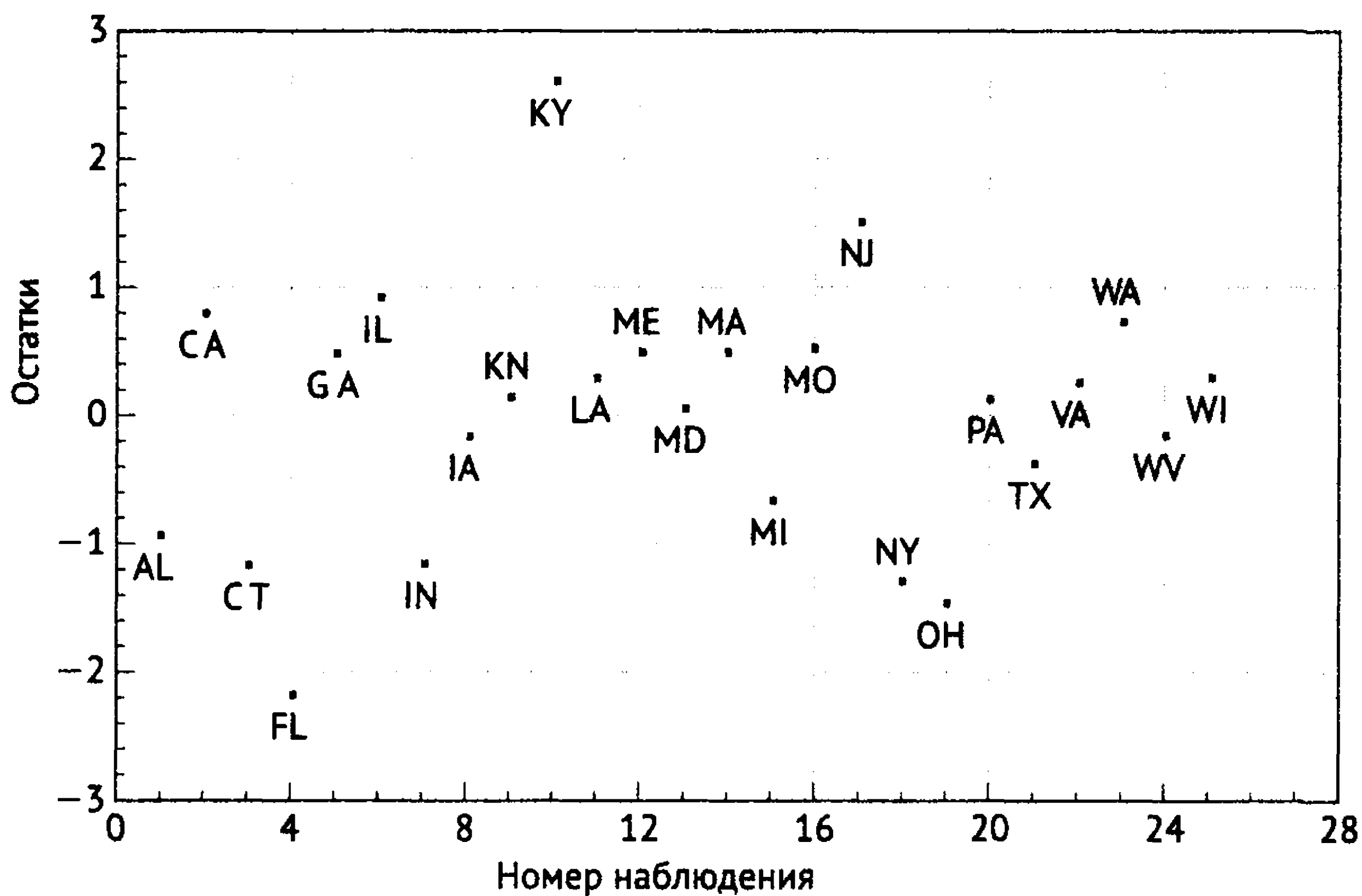


Рис. 7.5. Стандартизированные остатки для производственной функции

Таблица 7.4. LS- и LAD-оценки производственной функции

| Коэффициент | Метод наименьших квадратов | | | LAD | | | | |
|--------------|----------------------------|---------------|-----------------|--------|---------------|-----------------|-------------------|-----------------|
| | Оценка | Станд. ошибка | t -статистика | Оценка | Бутстрэп | | Ядерная плотность | |
| | | | | | Станд. ошибка | t -статистика | Станд. ошибка | t -статистика |
| Константа | 2,293 | 0,107 | 21,396 | 2,275 | 0,202 | 11,246 | 0,183 | 12,374 |
| β_k | 0,279 | 0,081 | 3,458 | 0,261 | 0,124 | 2,099 | 0,138 | 1,881 |
| β_l | 0,927 | 0,098 | 9,431 | 0,927 | 0,121 | 7,637 | 0,169 | 5,498 |
| Σe^2 | 0,7814 | | | 0,7984 | | | | |
| $\Sigma e $ | 3,3652 | | | 3,2541 | | | | |

Для этого случая были получены оценки производственной функции Кобба–Дугласа с помощью метода наименьших квадратов и LAD. Основываясь на графике стандартизированных остатков метода наименьших квадратов, приведенном на рис. 7.5, можно предположить, пользуясь стандартным подходом, что два наблюдения (Флорида (FL) и Кентукки (KY)) являются выбросами. Вектор коэффициентов, полученный методом наименьших квадратов с учетом и без учета этих двух наблюдений, есть $(2,293; 0,279; 0,927)$ и $(2,205; 0,261; 0,879)$ соответственно, откуда мы можем сделать вывод, что эти две точки оказывают сильное влияние. В табл. 7.4 представлены LAD-оценки тех же параметров со стандартными ошибками, построенными на 500 бутстрэповских повторениях. Оценки LAD с учетом и без учета этих двух наблюдений идентичны, поэтому представлены только первые. Если мы используем простое приближение, умножив соответствующую стандартную ошибку OLS на $(\pi/2)^{1/2} = 1,2533$, то получим значение, удивительно близкое к оценкам стандартных ошибок, полученных бутстрэпом: $(0,102; 0,123)$ против $(0,124; 0,121)$. Следующая группа оценок стандартных ошибок построена на оценке, предложенной Коэнкером: $0,25/\hat{f}^2(0) = 0,25/1,5467^2 = 0,104502$. Окно и ядерная функция выбраны так, как было предложено ранее. Для столь малой выборки результаты поразительно устойчивы.

7.3.2. Модели квантильных регрессий

Модель квантильной регрессии запишется как

$$Q[y|\mathbf{x}, q] = \mathbf{x}'\beta_q \text{ такая, что } Prob[y \leq \mathbf{x}'\beta_q|\mathbf{x}] = q, \quad 0 < q < 1.$$

Эта спецификация является совершенно непараметрической. Никакие предположения не были сделаны ни про распределение величины $y|\mathbf{x}$, ни про ее условную дисперсию. Из того, что q может непрерывно меняться (строго) между нулем и единицей, следует, что существует бесконечное число «векторов параметров». Кажется разумным смотреть на коэффициенты, которые мы можем записать как $\beta(q)$, не как на фиксированные «параметры», как мы делали в модели линейной регрессии, а, скорее, как на особенность распределения $y|\mathbf{x}$. Так, например, кажется неправильным рассматривать $\beta(0,49)$ совершенно отличающейся от $\beta(0,50)$ или точно вычислять разности вроде $\beta(0,5) - \beta(0,3)$. С другой стороны, качественная разница или, напротив, ее отсутствие между $\beta(0,3)$ и $\beta(0,5)$, как покажет пример ниже, может быть интересным свойством данных.

Оценка \mathbf{b}_q вектора коэффициентов β_q для конкретного квантиля вычисляется путем минимизации функции

$$\begin{aligned} F_n(\beta_q|\mathbf{y}, \mathbf{X}) &= \sum_{i: y_i \geq \mathbf{x}'_i \beta_q} q |y_i - \mathbf{x}'_i \beta_q| + \sum_{i: y_i < \mathbf{x}'_i \beta_q} (1 - q) |y_i - \mathbf{x}'_i \beta_q| \\ &= \sum_{i=1}^n g(y_i - \mathbf{x}'_i \beta_q | q), \end{aligned}$$

где

$$g(e_{i,q}|q) = \begin{cases} qe_{i,q}, & \text{если } e_{i,q} \geq 0 \\ (1-q)e_{i,q}, & \text{если } e_{i,q} < 0 \end{cases}, \quad e_{i,q} = y_i - \mathbf{x}'_i\beta_q.$$

При $q = 0,5$ она является оценкой методом наименьших абсолютных отклонений, которую мы изучали в примере 4.5 и подразделе 7.3.1. Для решения задачи минимизации требуется итеративная оценка. Задачу можно рассматривать как пример линейного программирования¹³. [См. (Keonker, D'Ogry (1987)).]

Мы не можем использовать методы, обсуждавшиеся в главе 4, чтобы определить асимптотическую ковариационную матрицу этой оценки. Однако то, что оценка получается путем минимизации суммы, влечет ряд результатов, схожих с выведенными нами для оценки методом наименьших квадратов в разделе 4.4. [См. (Buchinsky (1998)).] Предполагая, что регрессоры доброкачественные, оценка β_q в квантильной регрессии является состоятельной и асимптотически нормально распределенной с асимптотической ковариационной матрицей

$$\text{Asy.Var.}[b_q] = \frac{1}{n} \mathbf{H}^{-1} \mathbf{G} \mathbf{H}^{-1},$$

где

$$\mathbf{H} = \text{plim} \frac{1}{n} \sum_{i=1}^n f_q(0|\mathbf{x}_i) \mathbf{x}_i \mathbf{x}'_i$$

и

$$\mathbf{G} = \text{plim} \frac{q(1-q)}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}'_i.$$

Ранее этот результат был получен нами для LAD-оценки, а теперь и для квантиля q , не равного 0,5. Как и ранее, вычисление оценки усложняется из-за необходимости вычислять плотность ε_q в нуле. Для этого потребуется либо использовать приближение неизвестной точности, либо специфицировать конкретную плотность, чего мы стремились избежать. Обычный подход, как и ранее, — это использование бутстрэпа.

Пример 7.10. *Эластичность по кредитным картам в зависимости от дохода/расходов*

В работах Грина (Greene (1992, 2007)) были проанализированы невыплаты по кредитам и ежемесячные расходы по кредитным картам в большой выборке (13 444 наблюдения) пользователей кредитных карт. Среди целей работы была в том числе оценка эластичности ежемесячных расходов по доходу. Обычный регрессионный подход может быть основан на уравнении

$$Q[\ln \text{Spending}|\mathbf{x}, q] = \beta_{1,q} + \beta_{2,q} \ln \text{Income} + \beta_{3,q} \text{Age} + \beta_{4,q} \text{Dependents}.$$

Данные в табл. F7.3 приложения содержат информацию об этих и некоторых других переменных, которые могут объяснять расходы; эти

¹³Квантильная регрессия включается как процедура, встроенная в современное программное обеспечение, такое как Stata, SAS и NLOGIT.

три переменные были выбраны только для этого примера. 13 444 наблюдения были получены на основании заявлений на получение кредитных карт. Из всей выборки 10 499 заявлений были подтверждены, и для них проводились наблюдения по тратам и невыплатам по кредитам¹⁴. Расходы рассчитываются как среднемесячные в течение 12 месяцев после открытия счета. Среднемесячный доход и число иждивенцев есть среди демографических данных, указанных в заявлении. В табл. 7.5 представлены оценки коэффициентов функции условного среднего методом наименьших квадратов и результаты для нескольких квантилей¹⁵. Стандартные ошибки приводятся для результатов по методу наименьших квадратов и медианы ($q = 0,5$). Для остальных квантилей результаты в целом те же. Оценка методом наименьших квадратов, равная 1,08344, немного больше единицы, а оцененная стандартная ошибка равна 0,03212, соответственно t -статистика равна $(1,08344 - 1)/0,03212 = 2,60$. Такое поведение потребителя вполне ожидаемо. Однако большой разброс по различным квантилям несколько неожиданный. Мы можем объяснить это тем, что при высоком уровне расходов независимо от дохода существует (сравнительное) насыщение в реакции расходов на изменение дохода.

Таблица 7.5. Оцененные модели квантильных регрессий

| Квантиль | Оцененные параметры | | | |
|----------------------------|---------------------|-----------|-----------|------------|
| | Константа | ln Income | Age | Dependents |
| 1 | 2 | 3 | 4 | 5 |
| 0,1 | -6,73560 | 1,40306 | -0,03081 | -0,04297 |
| 0,2 | -4,31504 | 1,16919 | -0,02460 | -0,04630 |
| 0,3 | -3,62455 | 1,12240 | -0,02133 | -0,04788 |
| 0,4 | -2,98830 | 1,07109 | -0,01859 | -0,04731 |
| (Медиана) 0,5 | -2,80376 | 1,07493 | -0,01699 | -0,04995 |
| Станд. ошибка | (0,24564) | (0,03223) | (0,00157) | (0,01080) |
| t | -11,41 | 33,35 | -10,79 | -4,63 |
| Метод наименьших квадратов | -3,05581 | 1,08344 | -0,01736 | -0,04461 |
| Станд. ошибка | (0,23970) | (0,03212) | (0,00135) | (0,01092) |
| t | -12,75 | 33,73 | -12,88 | -4,08 |
| 0,6 | -2,05467 | 1,00302 | -0,01478 | -0,04609 |
| 0,7 | -1,63875 | 0,97101 | -0,01190 | -0,03803 |
| 0,8 | -0,94031 | 0,91377 | -0,01126 | -0,02245 |
| 0,9 | -0,05218 | 0,83936 | -0,00891 | -0,02009 |

На рис. 7.6 изображены оценки эластичности расходов по доходу для ряда квантилей — от 0,1 до 0,9 и оценка методом наименьших квадратов, относящаяся к фиксированному для всех указанных квантилей значению. Доверительный интервал, показанный на рисунке, основан

на асимптотической нормальности. Он вычислен как оцененный доход плюс-минус 1,96, умноженное на оцененную стандартную ошибку.

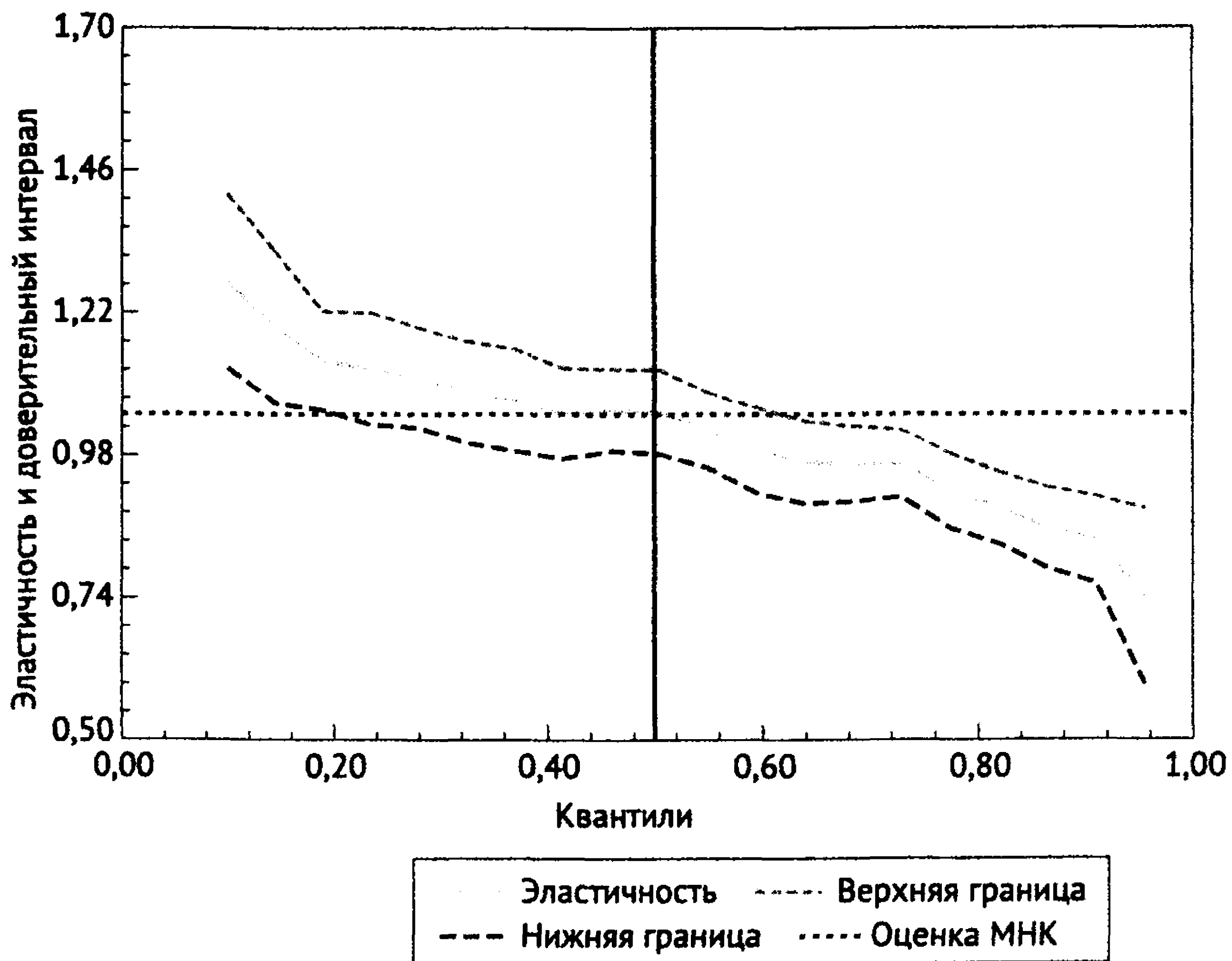


Рис. 7.6. Оценки эластичности расходов по доходу

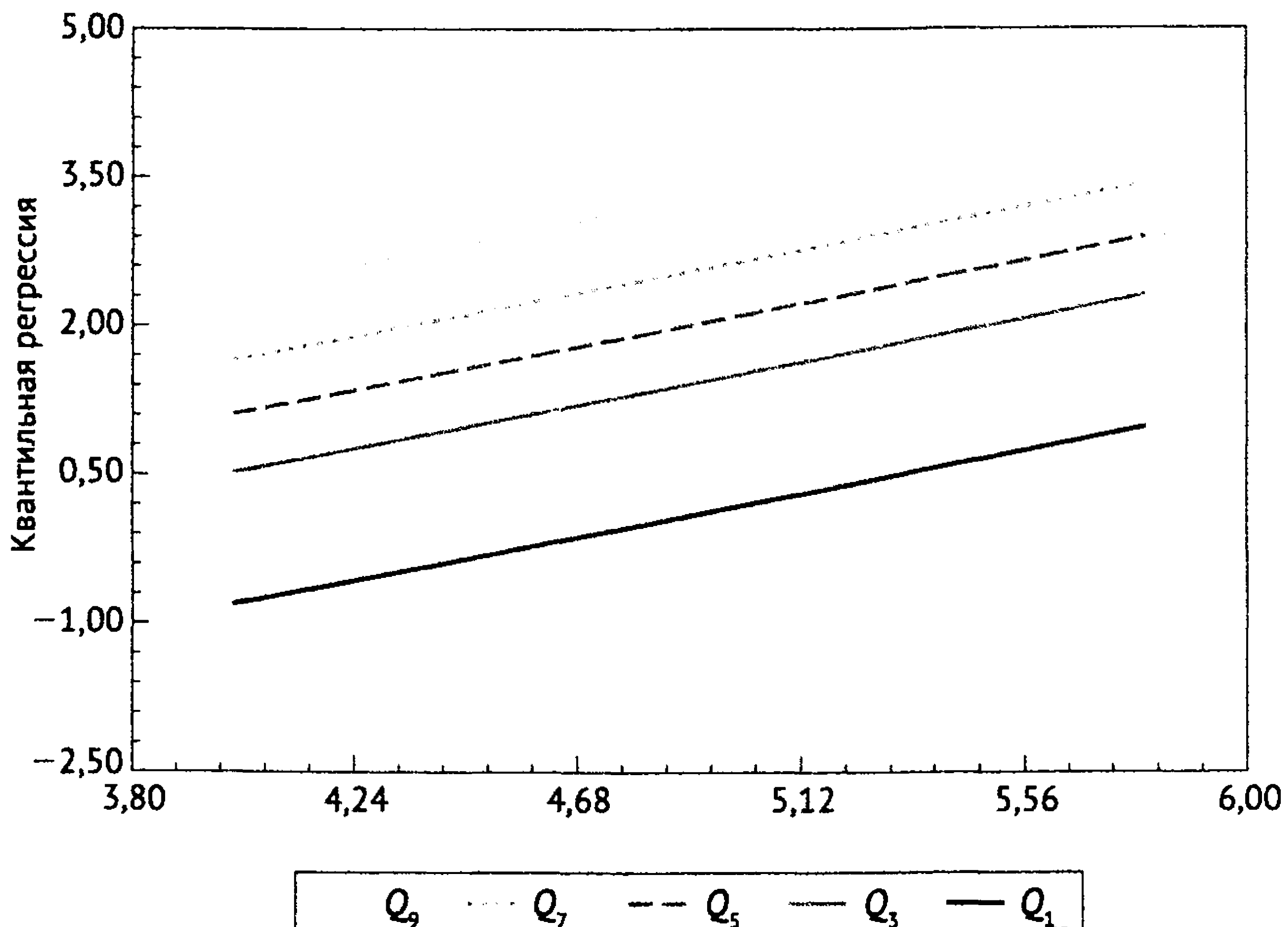


Рис. 7.7. Квантильные регрессии для логарифма расходов (Ln Spending)

Рисунок 7.7 показывает соответствующие квантильные регрессии для значений $q = 0,1; 0,3; 0,5; 0,7$ и $0,9$. Относительно большой рост — от

0,1 квантиля до 0,3 — заставляет предположить некоторую скошенность в распределении расходов. В широком смысле результаты, по-видимому, согласуются с полученными ранее, гласящими, что квантили в значительной степени различаются ввиду сдвигов в постоянном члене, несмотря на кажущиеся принципиальными различия в коэффициенте при $\ln \text{Income}$ в таблице результатов.

7.4. Частично линейная регрессия

Выбор точной функциональной формы — важный момент при спецификации линейной регрессии. Это было внимательно разобрано нами в главе 6. Некоторые подходы, в том числе использование дамми-переменных, логарифмов, квадратов и т.д., являются способами учета нелинейности. В частности, транслогарифмическая модель (пример 2.4) является хорошо известным способом приближения неизвестной нелинейной функции. Однако, даже используя такие подходы, исследователь может быть все еще заинтересован в ослаблении условия на функциональную форму модели. Частично линейная модель, [детально рассмотренная в работах (Yatchew (1998, 2000)) и (Härdle, Liang, Gao (2000))], может стать альтернативным подходом. Рассмотрим регрессионную модель, в которой одна из переменных, x , представляет особый интерес, но вид функциональной формы по отношению к x неясен. Запишем модель в виде

$$y_i = f(x_i) + \mathbf{z}'_i \boldsymbol{\beta} + \varepsilon_i,$$

где данные предполагаются доброкачественными и с точностью до функциональной формы все предположения классической модели выполнены. Функция $f(x_i)$ остается неспецифицированной. Как уже было сказано, нельзя получить оценки методом наименьших квадратов, пока $f(x_i)$ не будет специфицирована. Предположим, данные таковы, что состоят из пар наблюдений (y_{j1}, y_{j2}) , $j = 1, \dots, n/2$, для которых $x_{j1} = x_{j2}$ для каждой пары. В этом случае оценка $\boldsymbol{\beta}$ может быть получена при помощи простой преобразованной модели

$$y_{j2} - y_{j1} = (\mathbf{z}_{j2} - \mathbf{z}_{j1})' \boldsymbol{\beta} + (\varepsilon_{j2} - \varepsilon_{j1}), \quad j = 1, \dots, n/2.$$

¹⁴ Данные по расходам были взяты из базы по кредитным картам, а доход и демографические показатели были взяты из заявлений на получение карт. Может показаться, что стоит использовать цензурированную квантильную регрессию (как, например, сделано в работе (Powell (1986a,b))), чтобы учесть большое количество нулей зависимой переменной. Однако такой подход будет неверно специфицировать модель: нули относятся к несуществующим наблюдениям, а не к пропущенным. Более аккуратный подход, использованный в исследовании 1992 г., — это моделировать по расходам существование или отсутствие наблюдения, а затем моделировать расходы условно по тому факту, что заявление было принято. Мы вернемся к этому моменту в главе 19 при обсуждении модели выборочной селективности (sample selection model). Данные по доходу ограничены, 100 000 и 220 точек имеют в графе «расходы» 1 долл. или менее. Мы не чистили данные, чтобы убрать подобные наблюдения. Все 10499 наблюдений, которые присутствовали в исходных данных, были использованы.

¹⁵ Заметим, что если модель формулируется в виде (7-33), то отсюда не вытекает, что функция условного среднего является линейной регрессией. Это должно предполагаться отдельно.

Поскольку наблюдения независимы, построенные шоки v_i все еще имеют нулевое среднее, новую дисперсию $2\sigma^2$ и остаются попарно некоррелированными. В этом случае применима классическая модель и метод наименьших квадратов является оптимальным. В самом деле, имея оценку величины β , скажем $\hat{\beta}_d$, зашумленная оценка $f(x_i)$ может быть получена как $y_i - \mathbf{z}'_i \hat{\beta}$ (в оценке содержится как ошибка оценивания, так и ε_i)¹⁶.

Проблема, разумеется, заключается в том, что основное предположение малореалистично. Данные не будут вести себя таким образом, кроме случая, когда они были получены искусственно. Тем не менее общая логика оценивания частично линейной регрессии остается прежней. Предположим, что наблюдения переставлены таким образом, что $x_1 < x_2 < \dots < x_n$. Предположим также, что эта переменная является доброкачественной в том смысле, что при увеличении размера выборки вектор сортированных данных равномерно плотно заполняет пространство, в котором может меняться x_i . Тогда интуитивно разность почти равна нулю и уменьшается с ростом выборки. [В работах (Yatchew (1997, 1998)) соответствующая теория рассмотрена более аккуратно.] Также разработана теория деления на группы, состоящие из двух и более наблюдений. Преобразованные наблюдения в этом случае выглядят как $y_{d,i} = \sum_{m=0}^M d_m y_{i-m}$, где $\sum_{m=0}^M d_m = 0$ и $\sum_{m=0}^M d_m^2 = 1$. (Данные не разделены на непересекающиеся группы для такого преобразования — мы лишь используем этот способ для описания техники.) Пара весов для $M = 1$ очевидна: берется $\pm\sqrt{0,5}$, т.е. всего-навсего масштабированный вектор 1, -1 (обычная разность). В работе Ятчева [(Yatchew [1998, p. 697])] затабулированы оптимальные веса для дифференцирования при $M = 1, \dots, 10$. Для $M = 2$ значения составляют (0, 8090; -0, 500; -0, 3090), а для $M = 3$ (0, 8582; -0, 3832; -0, 2809; -0, 1942). Показывается, что эта оценка является состоятельной, асимптотически нормально распределена и имеет асимптотическую ковариационную матрицу¹⁷

$$\text{Asy. Var}[t\hat{\beta}_d] = \left(1 + \frac{1}{2M}\right) \frac{\sigma_v^2}{n} E_x[\text{Var}[\mathbf{z} | x]].$$

Матрица может быть оценена с помощью суммы квадратов и попарных произведений преобразованных данных. Дисперсия остатков вычисляется подобным обычному образом:

$$\hat{\sigma}_v^2 = \frac{\sum_{i=M+1}^n (y_{d,i} - \mathbf{z}'_{d,i} \hat{\beta}_d)^2}{n - M}.$$

Ятчев предлагает сгладить частичные остатки $y_{d,i} - \mathbf{z}'_{d,i} \hat{\beta}_d$ с помощью ядерной оценки плотности, чтобы получить улучшенную оценку $f(x_i)$. В работе (Manzan, Zeron (2010)) эта модель применяется для исследования рынка топлива в США.

¹⁶См. работу (Estes, Honoré (1995)), в которой был предложен этот подход (простое взятие попарной разности в данных).

¹⁷В работе (Yatchew (2000, p. 191)) ковариационная матрица определяется как $E[\text{Cov}[\mathbf{z} | x]]$.

Пример 7.11. Частично линейная транслогарифмическая функция издержек

В работах Ятчева (Yatchew (1998, 2000)) подобная схема была применена при исследовании эффекта масштаба для издержек потребления электричества. Следуя работам Нерлова (Nerlove (1963)) и Кристенсена и Грина (Christensen, Greene (1976)), функция издержек специфицируется как транслогарифмическая модель (см. пример 2.4 и подраздел 10.5.2), включающая цены на труд и капитал, различные характеристики коммунальных предприятий, а также интересующую нас величину — количество потребителей в системе C . Проведем похожий анализ, используя данные по потреблению электричества из работы Кристенсена и Грина (Christensen, Greene (1970)). Данные приведены в табл. F4.4 приложения. (См. подраздел 10.5.1, содержащий описание данных.) Данные содержат 158 наблюдений, но последние 35 из них представляют собой управляющие компании, т.е. включают комбинации остальных. К тому же несколько малых компаний в Новой Англии имеют совершенно нехарактерные для отрасли издержки. Мы будем строить наш анализ, используя наблюдения с 6-го по 123-й. Величины, которые присутствуют в данных, включают $Q = \text{output}$ (выпуск), $C = \text{total cost}$ (совокупные издержки) и $PK, PL, PF = \text{стоимость за условную единицу для капитала (capital), труда (labor) и топлива (fuel) соответственно}$. Специфицированная параметризованная модель является частным случаем модели, рассмотренной в работе Кристенсена и Грина (Christensen, Greene, 1976):

$$c = \beta_1 k + \beta_2 l + \beta_3 q + \beta_4 (q^2/2) + \beta_5 + \varepsilon,$$

где $c = \ln[C/(Q \times PF)]$, $k = \ln(PK/PF)$, $l = \ln(PL/PF)$ и $q = \ln Q$. В частично линейной модели последние три члена заменены на $f(q)$. Деление на PF требуется с точки зрения теории: средние издержки должны быть линейно однородны по ценам. Далее приведены оцененные уравнения с оценками стандартных ошибок:

(параметрическая)

$$c = \begin{matrix} -7,32 & + & 0,069k & + & 0,241 & - & 0,569q & + & 0,057q^2/2 & + & \varepsilon, \\ (0,333) & & (0,065) & & (0,069) & & (0,042) & & (0,006) & & s = 0,13949 \end{matrix}$$

(частично линейная)

$$c_d = \begin{matrix} 0,108k_d & + & 0,163l_d & + & f(q) & + & v \\ (0,076) & & (0,081) & & & & s = 0,16529 \end{matrix}$$

7.5. Непараметрическая регрессия

Функция регрессии переменной y на единственную переменную x определяется как

$$y = \mu(x) + \varepsilon.$$

Здесь мы не делаем никаких предположений о распределении, гомоскедастичности, автокорреляции и, главное, о функциональной форме; $\mu(x)$ может быть совершенно нелинейной. Поскольку мы рассматриваем функцию условного среднего, единственное ограничение состоит в том, что отклонения от этой функции не являются функциями от x (не коррелированы с x). Мы рассматривали различные варианты нелинейных функций, включающие сплайны, полиномы, логарифмы, дамми-переменные и т.д. Однако каждая из этих спецификаций является всеобъемлющей. Функциональная форма остается неизменной для любых значений x . Сейчас мы рассмотрим методы, которые не предполагают никакой конкретной функциональной формы.

Простейшим случаем для анализа был бы тот, в котором несколько (разных) наблюдений y_i соответствовали единственному конкретному значению x_i . Тогда функцию условного среднего можно было бы оценить естественным образом, используя средние по группам. Однако этот подход имеет два изъяна. Во-первых, простое соединение точек средних не приведет к получению гладкой функции. Для этого метода придется предполагать что-нибудь о поведении функции между точками, чего мы хотели бы избежать. Во-вторых, подобные данные вряд ли встретятся нам вне экспериментальной ситуации. Имея данные, не являющиеся сгруппированными, мы можем предложить следующий вариант. Около каждой интересующей нас точки определим соседние точки и проведем отдельно линейные или квадратичные регрессии в каждой окрестности. Это возвращает нас к проблеме непрерывности, о которой мы говорили ранее, но можно использовать метод сплайнов, который мы обсуждали в подразделе 6.3.1 и который, вообще говоря, специально для этой проблемы и создавался. Тем не менее, пока количество соседних точек не станет достаточно большим, функция будет выглядеть весьма грубой.

Сглаживающие методы приспособлены для построения оценок функции условного среднего без необходимости делать строгие предположения о поведении функции между точками. Они поддерживают принцип **ближайшего соседа**, но используют более тонкие схемы для построения гладких, доброкачественных функций. Общий класс подобных оценок может быть определен как функции оценивания условного среднего в виде

$$\hat{\mu}(x^*) = \sum_{i=1}^n w_i(x^* | x_1, x_2, \dots, x_n) y_i = \sum_{i=1}^n w_i(x^* | \mathbf{x}) y_i,$$

где сумма весов равна 1. Линейная регрессия методом наименьших квадратов является именно такой оценкой. Прогнозное значение выглядит как

$$\hat{\mu}(x^*) = a + bx^*,$$

где a и b есть постоянная и коэффициент наклона, оцененные при помощи метода наименьших квадратов. Для этой функции можно показать, что

$$w_i(x^* | \mathbf{x}) = \frac{1}{n} + \frac{x^*(x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Проблема с этой взвешивающей функцией, которую мы здесь не хотим затрагивать, заключается в том, что она позволяет каждому x_i быть в окрестности x^* , но не уменьшает вес x_i при удалении от x^* . Был предложен ряд **сглаживающих функций**, которые дают более качественные регрессионные функции. [См. Cleveland (1979) и Schimek (2000).] Рассмотрим две из них. Локально взвешенная сглаженная оценка регрессии (иногда называемая loess или lowess) основывается на явном определении окрестности точек, близких к x^* . Для этого требуется выбрать ширину окна h . **Окрестность** — это тот набор точек, для которого величина $|x^* - x_i|$ достаточно мала. Например, точки, которые попадают в интервал $x^* \pm h/2$, могут образовывать окрестность. Выбор окна существен и, как мы увидим в следующем примере, сложен. Простого, всегда пригодного правила не существует. Исследователи часто пользуются **эмпирическим правилом Сильвермана** (1986)

$$h_{Silverman} = \frac{0,9[\min(s, IQR)]}{1,349 n^{0,2}},$$

где s — это выборочное стандартное отклонение и IQR — межквантильный размах (0,75 квантиля минус 0,25 квантиля). Кроме этого требуется выбрать подходящие веса. В работе Кливлэнда (Cleveland (1979)) рекомендуются трикубические (tricube) веса

$$T_i(x^*|\mathbf{x}, h) = \left[1 - \left(\frac{|x_i - x^*|}{h} \right)^3 \right]^3.$$

Таким образом, вес для этой оценки равен

$$w_i(x^*|\mathbf{x}, h) = 1(x_i \text{ лежит в окрестности}) \times T_i(x^*|\mathbf{x}, h).$$

Выбор ширины окна принципиально влияет на результаты. Более широкое окно приведет к более гладкой функции, однако приближение данных будет хуже. Другой возможностью, похожей на подход метода наименьших квадратов, является рассмотрение окрестности, включающей все точки, но с выбором взвешивающей функции, плавно убывающей при увеличении расстояния между x^* и любым x_i . Для этой цели используется ряд **ядерных функций**. Частым выбором являются **логистическое ядро**

$$K(x^*|x_i, h) = \Lambda(v_i)[1 - \Lambda(v_i)], \text{ где } \Lambda(v_i) = \exp(v_i)/[1 + \exp(v_i)], v_i = (x_i - x^*)/h$$

и **ядро Епанечникова**

$$K(x^*|x_i, h) = 0,75(1 - 0,2v_i^2)/\sqrt{5}, \text{ если } |v_i| \leq 5, \text{ и } 0, \text{ если иначе.}$$

Это приводит к ядерной взвешенной оценке регрессии,

$$\hat{\mu}(x^*|\mathbf{x}, h) = \frac{\sum_{i=1}^n \frac{1}{h} K \left[\frac{x_i - x^*}{h} \right] y_i}{\sum_{i=1}^n \frac{1}{h} K \left[\frac{x_i - x^*}{h} \right]},$$

которая стала стандартным инструментом в непараметрическом анализе.

Пример 7.12. Непараметрическая функция средних издержек

В примере 7.11 мы применяли частично линейную регрессию для нахождения соотношения между средними издержками и выпуском на рынке электричества. Рисунки 7.8 и 7.9 показывают менее претенциозные непараметрические регрессии средних издержек на выпуск. Общая картина та же, что и в предыдущем примере. Ядерная функция в обоих случаях является логистической плотностью.

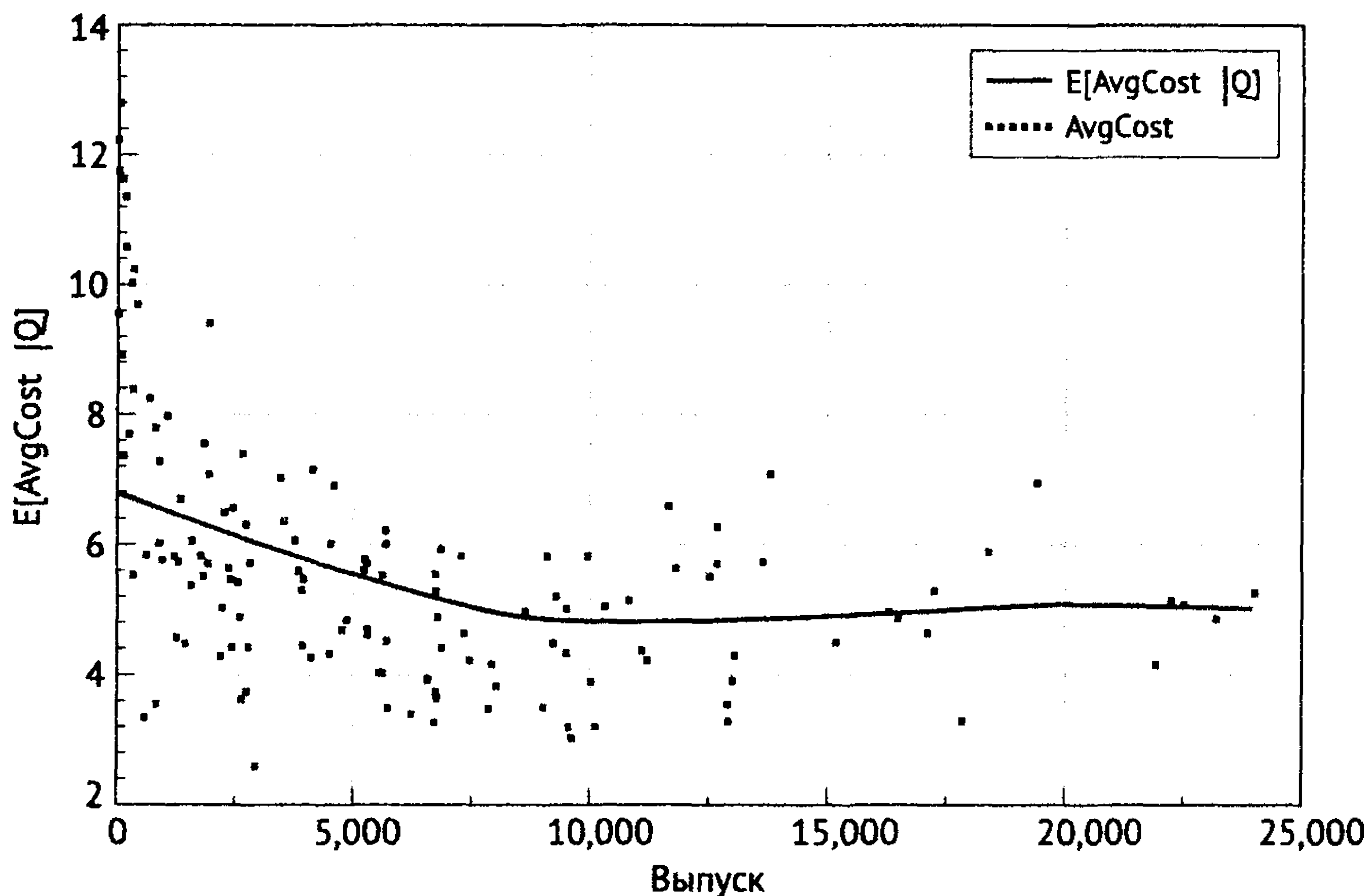


Рис. 7.8. Непараметрическая функция издержек

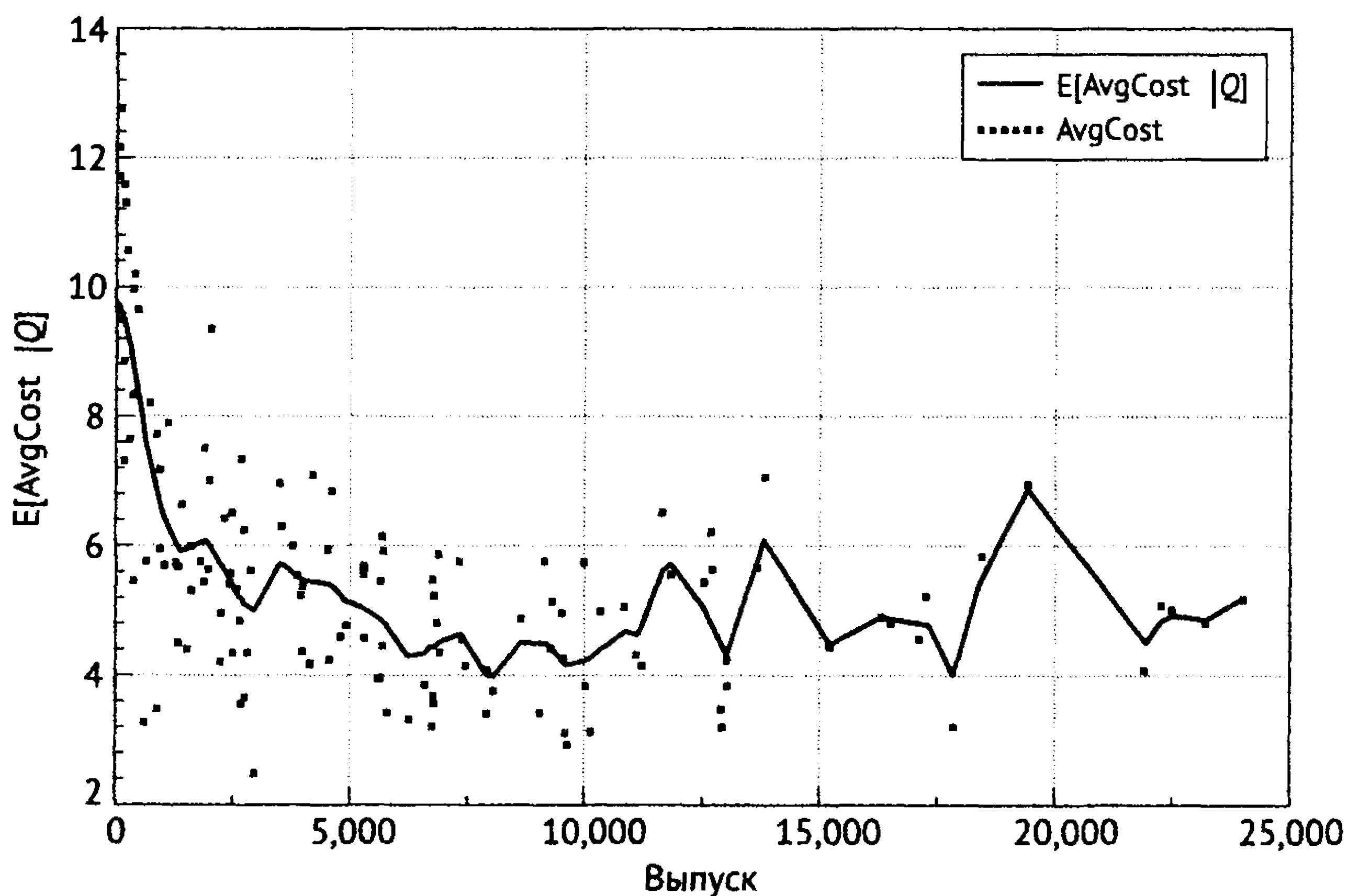


Рис. 7.9. Непараметрическая функция издержек

На рис. 7.8 мы используем ширину окна, равную 2000. Поскольку это достаточно большая часть от разброса выпуска, функция получается достаточно гладкой. Регрессия на рис. 7.9 использует ширину окна, равную 200. Функция приближает данные лучше, но мы поплатились за это гладкостью. Пример демонстрирует упомянутый выше факт: выбор ширины окна в случае с непараметрическими моделями оказывается весьма существенным.

Сглаживание данных в существенной мере определяется ими самими. Как и в случае с многими непараметрическими методами, статистические выводы не являются необходимой частью анализа и во многом носят описательный характер. Как мы увидели в предыдущем примере, непараметрическая регрессия может выявить интересные свойства данных. Тем не менее для исследователя этот подход имеет несколько существенных недостатков. С одной стороны, нет опасности неправильно специфицировать функцию условного среднего. Однако, с другой стороны, этот подход слишком общий, чтобы проверить конкретную спецификацию или лежащую в основе модели экономическую теорию. [См., например, работу Blundell, Browning, Crawford's (2003), представляющую собой всестороннее исследование британской структуры расходов.] Большая часть соотношений сложнее, чем простое условное среднее от одной переменной. В примере 7.12 некоторый разброс средних издержек определяется разницей в ценах факторов производства (в частности, топлива) и коэффициенте эксплуатационной мощности (load factor). Расширение полностью непараметрической регрессии на случай более одной переменной является возможным, но очень громоздким. [См. Härdle (1990), Li, Racine (2007).] Многообещающим подходом в данном случае будет частично линейная модель, рассмотренная ранее.

7.6. Заключение и выводы

В этой главе мы расширили модель линейной регрессии для случая, позволяющего учитывать нелинейность в параметрах функции регрессии. Интерпретация, оценивание и проверка гипотез в этой модели достаточно близки к линейному случаю. К двум наиболее существенным отличиям относятся, во-первых, более сложные процедуры оценивания, необходимые для оценки нелинейной модели, и, во-вторых, некоторая неясность в интерпретации коэффициентов модели (поскольку производные регрессионной функции не являются постоянными в отличие от случая линейной модели).

Основные термины и понятия

| | |
|---|---|
| Bootstrap — Бутстрэп | Median regression — Медианная регрессия |
| Box-Cox transformation — Преобразование Бокса-Кокса | Nearest neighbor — Метод ближайших соседей |
| Conditional mean function — Функция условного среднего | Neighborhood — Окрестность |
| Conditional median — Условная медиана | Nonlinear least squares — Нелинейный метод наименьших квадратов |
| Delta method — Дельта-метод | Nonlinear regression model — Нелинейная регрессионная модель |
| Epanechnikov kernel — Ядро Епанечникова | Nonparametric estimators — Непараметрические оценки |
| GMM estimator — GMM-оценка | Nonparametric regression — Непараметрическая регрессия |
| Identification condition — Условие идентификации | Normalization — Нормализация |
| Identification problem — Проблема идентификации | Orthogonality condition — Условие ортогональности |
| Index function model — Модель индексной функции | Overidentifying restrictions — Сверхидентифицирующие ограничения |
| Indirect utility function — Косвенная функция полезности | Partially linear model — Частично линейная модель |
| Interaction term — Член взаимодействия | Pseudoregressors — Псевдорегрессоры |
| Iteration — Итерация | Quantile regression — Квантильная регрессия |
| Jacobian — Якобиан | Roy's identity — Тождество Роя |
| Kernel density estimator — Ядерная оценка плотности | Semiparametric — Полупараметрический |
| Kernel functions — Ядерные функции | Semiparametric estimation — Полупараметрическая оценка |
| Least absolute deviations (LAD) — Метод наименьших абсолютных отклонений | Silverman's rule of thumb — Эмпирическое правило Сильвермана |
| Linear regression model — Линейная модель регрессии | Smoothing function — Сглаживающая функция |
| Linearized regression model — Линеаризованная модель регрессии | Starting values — Начальные значения |
| Lagrange multiplier test — Тест множителей Лагранжа | Two-step estimation — Двухшаговая оценка |
| Logistic kernel — Логистическое ядро | Wald test — Тест Вальда |
| Logit model — Логит-модель | |
| Loglinear model — Логлинейная модель | |

Упражнения

1. Объясните, как получить оценки нелинейным методом наименьших квадратов модели $y = \alpha x^\beta + \varepsilon$.
2. Проверьте следующее дифференциальное уравнение, связанное с преобразованием Бокса-Кокса:

$$\frac{d^i x^{(\lambda)}}{d\lambda^i} = \left(\frac{1}{\lambda}\right) \left[x^\lambda (\ln x)^i - \frac{id^{i-1} x^{(\lambda)}}{d\lambda^{i-1}} \right]. \quad (7-34)$$

Покажите, что предельной последовательностью в случае $\lambda = 0$ является

$$\lim_{\lambda \rightarrow 0} \frac{d^i x^{(\lambda)}}{d\lambda^i} = \frac{(\ln x)^{i+1}}{i+1}. \quad (7-35)$$

Этот результат может быть крайне полезен при вычислении второй производной функции логарифмического правдоподобия в модели Бокса–Кокса.

Эмпирические приложения

1. При помощи преобразования Бокса–Кокса мы можем определить следующую модель, альтернативную модели Кобба–Дугласа:

$$\ln Y = \alpha + \beta_k \frac{(K^\lambda - 1)}{\lambda} + \beta_l \frac{(L^\lambda - 1)}{\lambda} + \varepsilon.$$

Используя данные Зельнера и Реванкара (Zellner, Revankar) в табл. F7.2 приложения, оцените α , β_k , β_l и λ с помощью метода, рассмотренного в примере 7.5. (Не забудьте отнормировать Y , K и L в зависимости от числа служащих.) Используйте (7-24), (7-15) и (7-16) для вычисления нужных асимптотических стандартных ошибок для ваших оценок. Вычислите две эластичности выпуска, $\partial \ln Y / \partial \ln K$ и $\partial \ln Y / \partial \ln L$, в точках выборочных средних значений K и L . (Подсказка: $\partial \ln Y / \partial \ln K = K \partial \ln Y / \partial K$.)

2. Для модели из задачи 1 проверьте гипотезу, о том, что $\lambda = 0$, используя тест Вальда и тест множителей Лагранжа. Заметьте, что модель с ограничениями — это логлинейная модель Кобба–Дугласа. LM-статистика приведена в (7-22). Для того чтобы провести тест, вам необходимо будет вычислить элементы четвертого столбца X^0 , псевдорегрессор по отношению к λ имеет вид $\partial E[y | x] / \partial \lambda | \lambda = 0$. Результат (7-35) может быть полезен в этом случае.
3. Национальный институт стандартов и технологий (National Institute of Standards and Technology, NIST) создал веб-сайт, на котором можно найти большое число задач на оценивание с данными, предназначенными для проверки качества компьютерных программ (URL-сайт: <http://www.itl.nist.gov/div898/strd/>). Каждый из пяти наборов тестовых задач включает 27 задач на нелинейный метод наименьших квадратов, разделенных на три группы: простые, средние и сложные. Мы используем один из таких наборов для этой задачи. Вы можете попробовать какой-нибудь другой набор (например, чтобы посмотреть, может ли программный пакет, который вы используете, справиться с этими задачами). Это задача Misralc (<http://www.itl.nist.gov/div898/strd/nls/data/misralc.shtml>). Нелинейная регрессионная модель имеет вид

$$y_i = h(x, \beta) + \varepsilon = \beta_1 \left(1 - \frac{1}{\sqrt{1 + 2\beta_2 x_i}} \right) + \varepsilon_i.$$

Мы имеем дело со следующими данными:

| Y | X |
|-------|-------|
| 10,07 | 77,6 |
| 14,73 | 114,9 |
| 17,94 | 141,1 |
| 23,93 | 190,8 |
| 29,61 | 239,9 |
| 35,18 | 289,0 |
| 40,02 | 332,8 |
| 44,82 | 378,4 |
| 50,76 | 434,8 |
| 55,05 | 477,3 |
| 61,01 | 536,8 |
| 66,40 | 593,1 |
| 75,47 | 689,1 |
| 81,78 | 760,0 |

Для каждой задачи NIST также предоставляет «официальное решение» (т.е. правильный ответ). Для задачи Misralc решение следующее:

| | Оценка | Оцененная стандартная ошибка |
|---------------------|--------------------|------------------------------|
| β_1 | 6,3642725809E + 02 | 4,6638326572E + 00 |
| β_2 | 2,0813627256E - 04 | 1,7728423155E - 06 |
| $e'e$ | | 4,0966836971E - 02 |
| $s^2 = e'e/(n - K)$ | | 5,8428615257E - 02 |

Наконец, NIST предлагает два набора начальных значений для итерационного процесса: один далекий от решения, а другой близкий к нему. Для этой задачи предложенные начальные значения — это $\beta^1 = (5000, 0001)$ и $\beta^2 = (6000, 0002)$. Ваша задача повторить результаты NIST с помощью вашего программного пакета. [Подробный анализ основных моментов оценивания NIST нелинейным методом наименьших квадратов см. в (McCullough (1999)).]

4. В примере 7.1 функция CES используется для моделирования выпуска,

$$\ln y = \ln \gamma - \frac{\nu}{\rho} \ln [\delta K^{-\rho} + (1 - \delta)L^{-\rho}] + \varepsilon. \quad (7-36)$$

В примере 6.8 предложен косвенный метод оценки параметров этой модели. Функция линеаризуется около точки $\rho = 0$, что приводит, по сути, к линейной аппроксимации функции

$$\ln y = \beta_1 + \beta_2 \ln K + \beta_3 \ln L + \beta_4 [1/2(\ln K - \ln L)^2] + \varepsilon,$$

где $\beta_1 = \ln \gamma$, $\beta_2 = \nu \delta$, $\beta_3 = \nu(1 - \delta)$ и $\beta_4 = \rho \nu \delta(1 - \delta)$. Эта аппроксимация может быть оценена линейным методом наименьших квадратов. Оценки структурных параметров находятся решением четырех предыдущих уравнений. Оценка асимптотической ковариационной матрицы может быть получена с помощью дельта-метода. Параметры в (7-36) также могут быть оценены непосредственно с помощью нелинейного метода наименьших квадратов и результатов этой главы.

Данные Кристенсена и Грина (Christensen, Greene (1976)) по производству электроэнергии в США приведены в табл. F4.4 приложения. Данные содержат 158 наблюдений. Задействуя первые 123 из них, оцените CES — производственную функцию, используя капитал и топливо как факторы производства вместо капитала и труда. Сравните результаты, полученные при двух подходах, и прокомментируйте имеющиеся различия (которые являются существенными).

Для следующих заданий потребуется специальное программное обеспечение. Соответствующие процедуры есть в ряде программ, которые могут быть использованы, например SAS, Stata или LIMDEP. Задания предлагаются как отправные точки для исследований, которые можно проводить, используя некоторые рассмотренные в этой главе процедуры оценивания.

5. Используя данные по рынку бензина из табл. F2.2 приложения, с помощью метода частичной линейной регрессии, рассмотренного в разделе 7.4, оцените уравнение вида

$$\ln(G/Pop) = \beta_1 \ln(Income) + \beta_2 \ln P_{new\ cars} + \beta_3 \ln P_{used\ cars} + g(\ln P_{gasoline}) + \varepsilon.$$

6. В продолжение задачи 5 рассмотрите непараметрическую регрессию G/Pop на цену ($price$). Применяя метод непараметрического оценивания, предложенный в разделе 7.5, получите непараметрическую оценку, используя различную ширину окна, чтобы изучить влияние этого параметра на оценку.

ГЛАВА 8

Эндогенность и метод инструментальных переменных

8.1. Введение

До сих пор предположение о том, что в модели линейной регрессии

$$y_i = x_i' \beta + \varepsilon_i \quad (8-1)$$

x_i и ε_i не коррелированы, являлось определяющим. Однако часто встречаются ситуации, в которых это предположение оказывается неправдоподобным. В частности, это относится к моделям эффекта воздействия, как, например, в примере 6.5; к моделям, включающим переменные, измеренные с ошибкой; к динамическим моделям, включающим ожидания, а также к большому количеству различных ситуаций, в которых часть переменных ненаблюдаема или по каким-то причинам пропущена. Без предположения о некоррелированности шоков и регрессоров все утверждения о состоятельности или несмещенности оценки методом наименьших квадратов, рассмотренные в главе 4, становятся несправедливыми, а сама оценка соответственно теряет все свои достоинства. В данной главе описывается метод оценивания, пригодный в подобных ситуациях.

Удобно разделить вектор переменных x в (8-1) на два подвектора x_1 и x_2 , предполагая, что переменные в x_1 не коррелированы с ε , а переменные в x_2 коррелированы или могут быть коррелированы (т.е. это требует эмпирической проверки). Мы предполагаем, что вектор x_1 является **экзогенным** в модели (см. предположение А.3 в условиях линейной регрессионной модели в разделе 2.3). Таким образом, по этому определению вектор x_2 является **эндогенным** в модели. Откуда берется эндогенность? Некоторые часто встречающиеся случаи рассматриваются ниже, в примере 8.1.

Пример 8.1. *Модели с эндогенными переменными в правой части*

Следующие модели и ситуации будут встречаться в дальнейшем еще не раз.

Пропущенные переменные (omitted variables). В примере 4.2 мы изучали уравнение спроса на бензин вида

$$\ln G = \beta_1 + \beta_2 \ln Price + \beta_3 \ln Income + \varepsilon.$$

Если доход (Income) некорректно исключен из (любого) уравнения спроса, то оцениваемая «модель» приобретает вид

$$\ln G = \beta_1 + \beta_2 \ln Price + w,$$

где $w = \beta_3 \ln Income + \varepsilon$. Линейная регрессия $\ln G$ на константу $\ln Price$ не оценивает (β_1, β_2) состоятельно, если $\ln Price$ (логарифм цены) коррелирован с w . А корреляция обязательно будет присутствовать в агрегированных временных данных. Пропущенная переменная появляется в новом уравнении как часть шока, что приводит к **смещению, вызванному пропущенной переменной** (omitted variable bias), оценки методом наименьших квадратов в неверно специфицированном уравнении.

Эндогенный эффект воздействия (endogenous treatment effect.) В работе Крюгер и Дэйла (Kreuger, Dale (1999)) изучалось влияние посещения элитного колледжа на доходы в течение жизни. Регрессионная модель с «эффектом воздействия» — с дамми-переменной T , принимающей значение, равное единице, для посещающих элитный колледж и значение, равное нулю, для остальных — выглядит как

$$\ln y = \mathbf{x}'\beta + \delta T + \varepsilon.$$

Регрессия, оцененная при помощи метода наименьших квадратов, показателя сбережений ($\ln y$) на \mathbf{x} и T дает возможность получить оценку параметра δ , т.е. силы воздействия. Однако кажется совершенно ясным, что ненаблюдаемые факторы, влияющие на доходы в течение жизни, такие как амбиции, врожденные способности, упорство и т.д., также влияют на возможность индивида посещать элитный колледж. В этом случае оценка методом наименьших квадратов коэффициента δ будет некорректно приписывать эффект рассматриваемому воздействию вместо перечисленных факторов. В конечном счете метод наименьших квадратов не будет состоятельно оценивать δ именно из-за корреляции между T и ε .

Для того чтобы определить вклад посещения элитного колледжа в регрессии, исследователю надо было бы провести следующий невозможный эксперимент. Индивиды в выборке должны быть рассмотрены дважды: один раз с посещением колледжа, а второй раз (в следующей жизни) — без него. Вопрос, приводит ли сравнение индивидов, посещавших колледж, с индивидами, которые его не посещали, к адекватной оценке **эффекта воздействия на подвергшихся воздействию** (effect of treatment on treated individuals), является объектом исследования большого количества современных работ. Для обзора этой темы см., например, работу (Imbens, Wooldridge (2009)).

Одновременные уравнения (simultaneous equations.) В равновесной модели определения цен (price) и выпуска (output) на рынке рассматриваются вместе уравнения предложения (supply) и спроса (demand). Например, можно рассматривать модель определения цен и выпуска на рынке товаров следующего вида:

$$\text{(Demand)} \quad \text{Quantity}_D = \alpha_0 + \alpha_1 Price + \alpha_2 Income + \varepsilon_D,$$

$$\text{(Supply)} \quad \text{Quantity}_S = \beta_0 + \beta_1 Price + \beta_2 InputPrice + \varepsilon_S,$$

$$\text{(Equilibrium)} \quad \text{Quantity}_D = \text{Quantity}_S.$$

Допустим, мы пытаемся оценить параметры уравнения спроса, регрессируя временные данные равновесного выпуска на равновесные цены. Равновесная цена определяется уравнениями с двумя различными выпусками. Из условия равновесия получается, что $Price = (\alpha_0 - \beta_0 + \alpha_2 Income - \beta_2 InputPrice + \varepsilon_D - \varepsilon_S) / (\beta_1 - \alpha_1)$. Отсюда вытекает, что цена (*price*) коррелирована с ε_D : если внешнее воздействие меняет ε_D , то происходит сдвиг кривой спроса, что в конечном счете приводит к новой равновесной цене. Метод наименьших квадратов при оценке регрессии выпуска (*quantity*) на цену (*price*) и доход (*income*) несостоятельно оценивает параметры уравнения спроса. Эта «отдача» от ε_D на *Price* приводит к **смещению, вызванному одновременными уравнениями**, в оценке методом наименьших квадратов.

Динамические модели панельных данных (*dynamic panel data models*). В главе 11 мы изучим динамическую модель со **случайными эффектами** вида $y_{it} = x_{it}\beta + \gamma y_{i,t-1} + \varepsilon_{it} + u_i$, где u_i включает неизменные во времени (*time-invariant*) ненаблюдаемые характеристики индивида i . В этом случае ясно, что регрессор $y_{i,t-1}$ коррелирован с шоком ($\varepsilon_{it} + u_i$): ненаблюдаемая гетерогенность присутствует в y_{it} для каждого момента. В главе 13 мы рассмотрим модель расходов муниципальных органов власти вида $S_{it} = f(S_{i,t-1}, \dots) + \varepsilon_{it}$. Шоки предполагаются произвольно коррелированными во времени, поэтому и $S_{i,t-1}$, и ε_{it} коррелированы с $\varepsilon_{i,t-1}$. Таким образом, они коррелированы и между собой, что означает, что в этой модели даже без неизменных во времени эффектов не выполняются предположения линейной регрессионной модели. В ней коррелированы шоки и регрессоры.

Гетерогенность, вызванная пропущенным параметром (*omitted parameter heterogeneity*). Многие исследования экономического роста, проводимые по различным странам, имеют следующий вид (для нашего примера мы рассмотрим сильно упрощенный вариант):

$$\beta \Delta \ln Y_{it} = \alpha_i + \theta_{it} + \beta_i \Delta \ln Y_{i,t-1} + \varepsilon_{it},$$

где $\Delta \ln Y_{it}$ — темп роста страны i в году t . [См., например (Lee, Pesaran, Smith (1997)).] Заметим, что коэффициенты в модели предполагаются отдельными для каждой страны. Что произойдет, если оценить методом наименьших квадратов регрессию темпов роста дохода на временной тренд и лагированные (запаздывающие) темпы роста? Перепишем уравнение роста как

$$\begin{aligned} \beta \Delta \ln Y_{it} &= \alpha + \theta_t + \beta \Delta \ln Y_{i,t-1} + (\alpha_i - \alpha) + (\theta_i - \theta)_t + \\ &+ (\beta_i - \beta) \Delta \ln Y_{i,t-1} + \varepsilon_{it} = \alpha + \theta_t + \beta \Delta \ln Y_{i,t-1} + w_{it}. \end{aligned}$$

Мы предполагаем, что «средние» параметры α , θ и β есть фиксированные, имеющие некий экономический смысл величины, которые хотим оценить. Будет ли регрессия $\Delta \ln Y_{it}$ на константу, t и $\Delta \ln Y_{i,t-1}$, оцененная при помощи метода наименьших квадратов, оценивать эти параметры состоятельно? Мы можем предположить, что межстрановые различия

в константе являются чисто случайными и временной тренд θ_i также вызывается совершенно экзогенными факторами. Однако межстрановые различия в параметре сходимости β_i , по-видимому, коррелированы с ростом дохода в этих странах, который, в свою очередь, вызывает корреляцию между лагированным ростом дохода и членом $(\beta_i - \beta)$, включенным в w_{it} . Если же $(\beta_i - \beta)$ является случайным шумом, не коррелированным с $\Delta \ln Y_{i,t-1}$, то не коррелирован будет и член $(\beta_i - \beta)\Delta \ln Y_{i,t-1}$.

Ошибка измерения (measurement error). В работах (Ashenfelter, Krueger (1994)), (Ashenfelter, Zimmerman (1997)) и (Bonjour et al. (2003)) изучались ситуации, в которых уравнение заработной платы (Earnings) имело вид

$$y_{i,t} = f(\text{Education}_{i,t}, \dots) + \varepsilon_{i,t}$$

и было определено для пары $t = 1, 2$ для n семей. Образование (Education) представляет собой по сути неизмеримую величину. В качестве наилучшей возможной **прокси** (прокси-переменная, проху variable, проху) обычно используют число лет обучения (переменная Schooling). Рассмотрим крайне простую модель

$$y_{it} = \beta_1 + \beta_2 \text{Education}_{it} + \varepsilon_{it}$$

и попытаемся оценить ее параметры с помощью регрессии $Earnings_{it}$ на $Schooling_{it}$, причем

$$Schooling_{it} = Education_{it} + u_{it},$$

где u_{it} есть ошибка измерения. Подставляя, мы получаем, что

$$y_{it} = \beta_1 + \beta_2 Schooling_{it} + w_{it},$$

где $w_{it} = \varepsilon_{it} - \beta_2 u_{it}$. Переменная $Schooling$, очевидно, коррелирована с $w_{it} = (\varepsilon_{it} - \beta_2 u_{it})$. Интуитивно это связано с тем, что разброс переменной $Schooling$ хотя бы частично объясняется разбросом ошибки измерения u_{it} . Так как переменная $Schooling$ коррелирует с w_{it} , она является эндогенной и метод наименьших квадратов не подойдет для оценивания уравнения заработной платы. Как мы покажем ниже, в подобных случаях ошибка измерения в нужном регрессоре вызывает особый вид несостоятельности в оценке β_2 , **недооценку коэффициента** (attenuation bias).

Неслучайная выборка (nonrandom sampling). В модели исследования влияния обучающих программ, программ занятости или поведения предложения труда какой-нибудь части рынка наблюдения могут оказаться выбранными неслучайно. Job Training Partnership Act (JTPA) была программой повышения квалификации, направленной на увеличение занятости среди неблагополучной молодежи. В работе Андерсона и др. (Anderson et al. (1991)) было показано, что для исследуемой выборки программа была востребована большей частью среди наиболее квалифицированных кадров. Если для подобной выборки мы оценим уравнение заработной платы, то получим, что шоки не являются в полной мере случайными

величинами. В рассматриваемом примере, в частности, шоки слишком велики по сравнению с генеральной совокупностью в целом. Сами по себе слишком высокие шоки не считаются проблемой, пока не оказывается, что они являются объяснением неслучайности хотя бы части рассматриваемых переменных. Подобная неслучайность выборки объясняется **смещением из-за отбора наблюдений** (sample selection bias).

Истощение выборки (attrition). Возможны два близко связанных между собой важных случая неслучайной выборки. В исследованиях производительности фирм по панельным данным фирмы, оставшиеся в выборке к концу периода наблюдений, обычно представляют собой подвыборку тех, которые были вначале, — некоторые терпящие крах фирмы так или иначе выпадают из выборки. Оставшиеся фирмы не совсем обычны в том же смысле, что и в предыдущем примере об участниках программы JTRA. В подобных случаях оценки методом наименьших квадратов регрессии показателя производительности на регрессоры (какие бы то ни было) подвержены разновидности смещения из-за отбора наблюдений, называемой **смещением выживших** (survivorship bias). В этом случае распределение зависимой переменной, показателя производительности фирм, для оставшихся систематически выше, чем по выборке фирм в целом. Этот феномен также известен как **смещение из-за усечения выборки** (truncation bias). В клинических испытаниях и других статистических исследованиях по изучению влияния препаратов на здоровье пациентов исследуемые могут выпадать из выборки в связи с самим влиянием препарата — так, при исследовании жизненно важных препаратов, таких как лекарства от рака, пациент может выбыть как по причине выздоровления и нежелания принимать далее участие в исследовании, так и по причине смерти или потери дееспособности. В статистическом анализе часто говорят о **смещении из-за истощения выборки** (attrition bias). Те же соображения могут влиять и на анализ других исследований, связанных со здравоохранением. Так, например, в работе (Contoyannis, Jones, Rice (2004)) изучались самооценки здоровья по протяженному набору панельных данных из обследования домохозяйств Великобритании (British Household Panel Data). Каждый год существенное число наблюдений пропадает и в результате выборка истощается к концу периода наблюдений.

Во всех перечисленных в примере 8.1 случаях термин «смещение» обозначает тот факт, что оценка методом наименьших квадратов (или какой-нибудь его модификацией) является несостоятельной (постоянно смещенной) оценкой коэффициентов интересующей нас модели. Несмотря на то что причина этого меняется от задачи к задаче, в конечном счете все сводится к эндогенности ряда переменных в правой части, что, в свою очередь, означает корреляцию между регрессорами и шоками. В широком смысле мы можем выделить следующие эффекты:

- пропущенные переменные, как наблюдаемые, так и ненаблюдаемые,
- эффекты отдачи,
- динамические эффекты,

- эффекты, вызванные эндогенностью выборки, и т.д.

Можно выделить два общих решения для построения состоятельной оценки. В некоторых случаях можно построить более аккуратную, «структурную» **спецификацию** модели. Обычно для этого требуется привлечение дополнительных уравнений, объясняющих корреляцию между x_i и ε_i , после чего становится возможным оценить все интересующие нас параметры. Некоторые такие модели мы рассмотрим в следующих главах, включая, например, модель Хекмана, учитывающую проблему отбора наблюдений (см. Heckman (1979)), в главе 19. Другим подходом, становящимся все более популярным в современных исследованиях, является метод **инструментальных переменных**. Он основывается на следующих соображениях. Предположим, что модели (8-1) K переменных из вектора x_i могут быть коррелированы с ε_i . Предположим также, что существуют L переменных z_i такие, что z_i коррелируют с x_i , но не коррелируют с ε_i . Пользуясь привычным нам методом наименьших квадратов, мы не сможем получить состоятельную оценку параметра β . Однако из сделанных предположений о z_i , x_i и ε_i можно вывести ряд соотношений, которые позволят нам построить состоятельную оценку нашего параметра.

В этой главе мы подойдем к методу инструментальных переменных как к обобщению моделей и оценок, рассматривавшихся ранее, в главах 2–7. В разделе 8.2 мы строго сформулируем модель, которая будет оцениваться в дальнейшем. В разделе 8.3 изучим метод инструментальных переменных (instrumental variables, IV) и двухшаговый метод наименьших квадратов (two-stage least squares, 2SLS). В разделе 8.4 рассматриваются два теста на спецификацию модели. В разделе 8.5 мы в деталях изучим пример, в котором оценивание происходит при ошибке измерения. В разделе 8.6 рассмотрим нелинейные модели и начнем изучение обобщенного метода моментов (generalized method of moments, GMM). IV-оценивание — это мощный инструмент, который очень часто используется в современных эмпирических исследованиях. В разделе 8.7 рассмотрен недостаток этого метода — проблема слабых инструментов (problem of weak instruments). Наконец, раздел 8.8 посвящен некоторым замечаниям об инструментальных переменных и рассмотрению вопроса о наличии причинных эффектов.

В этой главе рассматриваются фундаментальные результаты IV-оценивания. Использование этого метода в различных задачах будет встречаться нам много раз в дальнейшем, в том числе в главе 10, посвященной системам уравнений, в главе 11, посвященной панельным данным, и в главе 13, посвященной обобщенному методу моментов.

8.2. Предположения расширенной модели

Предположения линейной регрессионной модели, введенные в главах 2 и 4, заключались в следующем.

А.1. Линейность. $y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \dots + x_{iK}\beta_K + \varepsilon_i$.

А.2. Полный ранг. $n \times K$ -матрица данных X имеет полный ранг по столбцам.

А.3. Экзогенность независимых переменных. $E[\varepsilon_i | x_{j1}, x_{j2}, \dots, x_{jk}] = 0$, $i, j = 1, \dots, n$. Независимые переменные не коррелированы с шоками.

А.4. Гомоскедастичность и отсутствие автокорреляции. Каждый шок ε_i имеет одинаковую конечную дисперсию σ^2 и не коррелирован с любым другим шоком ε_j условно относительно \mathbf{X} .

А.5. Случайные и неслучайные данные. $(x_{i1}, x_{i2}, \dots, x_{iK})$, $i = 1, \dots, n$.

А.6. Нормальное распределение. Шоки распределены нормально. Мы также продолжаем считать выполненным важный результат, что $\text{plim}(\mathbf{X}'\mathbf{X}/n) = \mathbf{Q}_{\mathbf{xx}}$. Тем не менее мы несколько изменим основные предположения модели. Во-первых, предположение А.3 (отсутствие автокорреляции между \mathbf{x} и ε) теперь формулируется как

$$\text{А.13. } E[\varepsilon_i | \mathbf{x}_i] = \eta_i.$$

Это предположение надо понимать в том смысле, что регрессоры теперь содержат некоторую информацию об ожидаемых шоках. Важным следствием А.13 является наличие корреляции между регрессорами и шоками. Из предположения А.13 следует, что

$$E[\mathbf{x}_i \varepsilon_i] = \gamma \quad (8-2)$$

для некоторого ненулевого γ . Если данные являются доброкачественными, то мы можем применить теорему D.5 (теорему Хинчина) и получить, что

$$\text{plim}(1/n)\mathbf{X}'\varepsilon = \gamma. \quad (8-3)$$

Заметим, что исходной модели соответствует случай $\eta_i = 0$. Следствием (8-3) является то, что регрессоры \mathbf{X} более не экзогенны.

Теперь предположим, что мы располагаем дополнительным набором переменных \mathbf{Z} , обладающих следующими двумя свойствами:

- 1) **экзогенностью:** они не коррелированы с шоками;
- 2) **релевантностью:** они коррелированы с независимыми переменными \mathbf{X} .

Более подробно эти предположения будут сформулированы по ходу изложения. В условиях представленной модели любые переменные, обладающие этими двумя свойствами, являются инструментальными. Предположим следующее:

А.17. $[\mathbf{x}_i, \mathbf{z}_i, \varepsilon_i]$, $i = 1, \dots, n$, являются i.i.d (н.о.р.) последовательностью случайных величин.

А.18а. $E[x_{ik}^2] = \mathbf{Q}_{\mathbf{xx},kk} < \infty$, конечная постоянная для $k = 1, \dots, K$.

А.18б. $E[z_{il}^2] = \mathbf{Q}_{\mathbf{zz},ll} < \infty$, конечная постоянная для $l = 1, \dots, L$.

А.18с. $E[z_{il}x_{ik}] = \mathbf{Q}_{\mathbf{zx},lk} < \infty$, конечная постоянная для $l = 1, \dots, L$, $k = 1, \dots, K$.

А.19. $E[\varepsilon_i | \mathbf{z}_i] = 0$.

В дальнейшем при рассмотрении моделей временных рядов нам потребуется ослабить предположение А.17. Конечность средних для \mathbf{z}_i следует из предположения А.18б. Аналогично разделу 4.4 мы можем получить

$\text{plim}(1/n)\mathbf{Z}'\mathbf{Z} = \mathbf{Q}_{\mathbf{zz}}$, конечную положительно определенную матрицу (доброкачественность данных),

$\text{plim } (1/n)\mathbf{Z}'\mathbf{X} = \mathbf{Q}_{\mathbf{z}\mathbf{x}}$, конечную $L \times K$ -матрицу ранга K (релевантность),
 $\text{plim } (1/n)\mathbf{Z}'\boldsymbol{\varepsilon} = \mathbf{0}$ (экзогенность).

До сих пор при формулировании модели регрессии мы рассматривали частный случай $\eta_i = 0$, откуда следовало, что $\gamma = 0$. Здесь нет необходимости отказываться от предположения А.17 (оно может продолжать выполняться), однако в этом частном случае оно становится несущественным.

На данный момент мы предположим, что $L = K$, т.е. количество инструментальных переменных совпадает с количеством переменных в правой части уравнения. Напомним, что во введении и в примере 8.1 мы разделяли вектор переменных \mathbf{x} в правой части уравнения (8.1) на два подвектора: \mathbf{x}_1 из K_1 экзогенных переменных и \mathbf{x}_2 из K_2 эндогенных переменных. На практике почти всегда «проблема эндогенности» возникает для одной или нескольких переменных из \mathbf{x} . В работе Крюгер и Дэйла (Kreuger, Dale (1999)), посвященной изучению эндогенных эффектов воздействия (см. пример 8.1), возникла единственная эндогенная переменная в уравнении — дамми-переменная воздействия T . При постановке задачи мы в таких случаях будем считать, что K_1 переменных из \mathbf{x}_1 входят в набор инструментальных переменных \mathbf{Z} , а оставшиеся K_2 переменных являются экзогенными, не совпадающими с переменными из \mathbf{x}_2 . Обычная интерпретация в этом случае состоит в том, что эти K_2 переменных, \mathbf{z}_2 , являются «инструментами \mathbf{x}_2 », в то время как \mathbf{x}_1 -переменные служат инструментами для самих себя. Продолжая рассматриваемый пример, матрица \mathbf{Z} для эндогенных эффектов воздействия в модели будет включать K_1 столбцов матрицы \mathbf{X} и дополнительную инструментальную переменную \mathbf{z} для дамми-переменной воздействия. В модели одновременных уравнений спроса и предложения эндогенной переменной в правой части уравнения является $x_2 = price$ (цена), а к экзогенным переменным относятся $(1, Income)$ (доход). Можно предположить (и вполне справедливо), что для этой модели в качестве набора инструментальных переменных можно рассмотреть $\mathbf{z} = (1, Income, InputPrice)$ (цены факторов производства). Из соотношений между переменными в модели следует, что этот интуитивный выбор приведет к разумному результату. Тем не менее по причинам, которые станут ясны чуть позже, нам нужно статистически рассматривать \mathbf{Z} как инструмент к полному набору переменных \mathbf{X} .

Есть еще один тонкий момент в использовании инструментальных переменных, который мы также рассмотрим чуть ниже. Условие релевантности должно представлять выражение в терминах условной корреляции. Вернемся к примеру с эффектами воздействия и предположим, что \mathbf{z} является инструментальной переменной для дамми-переменной воздействия T . Из сформулированного условия релевантности следует, что корреляция между \mathbf{z} и (\mathbf{x}, T) является ненулевой. Формально для нас требуется, чтобы условная корреляция между \mathbf{z} и $T|\mathbf{x}$ была ненулевой. На этот факт можно взглянуть с точки зрения проекций: инструментальная переменная \mathbf{z} является релевантной, если коэффициент при \mathbf{z} в регрессии T на (\mathbf{x}, \mathbf{z}) является ненулевым. На интуитивном уровне \mathbf{z} должна давать информацию об изменении T , которая является новой к той информации, которую дают переменные \mathbf{x} , уже присутствующие в модели.

8.3. Оценка

Для общей модели из раздела 8.2 у нас не используется большинство полезных свойств оценки методом наименьших квадратов. Сначала мы изучим свойства оценки методом наименьших квадратов в этой расширенной модели, а затем будем искать альтернативный способ построения оценки параметра β .

8.3.1. Метод наименьших квадратов

Оценка методом наименьших квадратов \mathbf{b} уже не является несмещенной:

$$E[\mathbf{b}|\mathbf{X}] = \beta + \mathbf{X}'\mathbf{X}^{-1}\mathbf{X}'\eta \neq \beta,$$

и соответственно теорема Гаусса–Маркова не выполняется. Оценка также перестает быть состоятельной:

$$\text{plim } \mathbf{b} = \beta + \text{plim} \left(\frac{\mathbf{X}'\mathbf{X}}{n} \right)^{-1} \text{plim} \left(\frac{\mathbf{X}'\boldsymbol{\varepsilon}}{n} \right) = \beta + \mathbf{Q}_{\mathbf{X}\mathbf{X}}^{-1}\boldsymbol{\gamma} \neq \beta. \quad (8-4)$$

(Асимптотическое распределение оценки изучается в примерах ниже.) Несостоятельность оценки наименьших квадратов не ограничивается коэффициентами при эндогенных переменных. Чтобы показать это, применим (8-4) к примеру с эффектами воздействия, рассмотренному ранее. В этом примере все переменные в \mathbf{X} , кроме последней, не коррелированы с $\boldsymbol{\varepsilon}$. Это означает, что

$$\text{plim} \left(\frac{\mathbf{X}'\boldsymbol{\varepsilon}}{n} \right) = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ \gamma_K \end{pmatrix} = \gamma_K \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

В этом частном случае результат (8-4) запишется как

$$\text{plim } \mathbf{b} = \beta + \gamma_K \times \text{последний столбец матрицы } \mathbf{Q}_{\mathbf{X}\mathbf{X}}^{-1}.$$

Нет никаких оснований полагать, что какие-то из элементов последнего столбца матрицы $\mathbf{Q}_{\mathbf{X}\mathbf{X}}^{-1}$ равны нулю. Отсюда следует, что, хотя лишь одна переменная из \mathbf{X} коррелирована с $\boldsymbol{\varepsilon}$, все элементы оценки \mathbf{b} будут несостоятельны, а не только оценка коэффициента при эндогенной величине. Такого рода ситуации называются **размыванием** (smearing): из-за эндогенности одной переменной несостоятельность размывается по всем оценкам, сделанным методом наименьших квадратов.

8.3.2. Оценка методом инструментальных переменных

Поскольку $E[\mathbf{z}_i \boldsymbol{\varepsilon}_i] = 0$ и дисперсия всех членов конечна,

$$\text{plim} \left(\frac{\mathbf{Z}'\boldsymbol{\varepsilon}}{n} \right) = \text{plim} \left(\frac{\mathbf{Z}'\mathbf{y}}{n} \right) - \text{plim} \left(\frac{\mathbf{Z}'\mathbf{X}\beta}{n} \right) = 0.$$

Таким образом,

$$\text{plim} \left(\frac{\mathbf{Z}'\mathbf{y}}{n} \right) = \left[\text{plim} \left(\frac{\mathbf{Z}'\mathbf{X}}{n} \right) \right] \beta + \text{plim} \left(\frac{\mathbf{Z}'\boldsymbol{\varepsilon}}{n} \right) = \left[\text{plim} \left(\frac{\mathbf{Z}'\mathbf{X}}{n} \right) \right] \beta. \quad (8-5)$$

Мы предположили, что \mathbf{Z} содержит столько же переменных, сколько \mathbf{X} . Так, можно рассмотреть функцию потребления, для которой $\mathbf{x}_t = [1, Y_t]$, а $\mathbf{z}_t = [1, Y_{t-1}]$. Мы предположили, что ранг матрицы $\mathbf{Z}'\mathbf{X}$ равен K , поэтому $\mathbf{Z}'\mathbf{X}$ является квадратной матрицей. Отсюда получаем, что

$$\left[\text{plim} \left(\frac{\mathbf{Z}'\mathbf{X}}{n} \right) \right]^{-1} \text{plim} \left(\frac{\mathbf{Z}'\mathbf{y}}{n} \right) = \beta, \quad (8-6)$$

что приводит нас к **оценке методом инструментальных переменных** (instrumental variable estimator)

$$\mathbf{b}_{IV} = (\mathbf{Z}'\mathbf{X})^{-1} \mathbf{Z}'\mathbf{y}.$$

Мы уже доказали, что оценка \mathbf{b}_{IV} состоятельна. Теперь мы заинтересованы в выводе ее **асимптотического распределения**. Мы будем действовать так же, как в подразделе 4.4.2. Рассмотрим величину

$$\sqrt{n}(\mathbf{b}_{IV} - \beta) = \left(\frac{\mathbf{Z}'\mathbf{X}}{n} \right)^{-1} \frac{1}{\sqrt{n}} \mathbf{Z}'\boldsymbol{\varepsilon},$$

которая имеет то же **предельное распределение**, что и $\mathbf{Q}_{zx}^{-1}[(1/\sqrt{n})\mathbf{Z}'\boldsymbol{\varepsilon}]$. Величина $(1/\sqrt{n})\mathbf{Z}'\boldsymbol{\varepsilon}$ может быть исследована так же, как $(1/\sqrt{n})\mathbf{X}'\boldsymbol{\varepsilon}$ в подразделе 4.4.3, откуда следует, что

$$\begin{aligned} \left(\frac{1}{\sqrt{n}} \mathbf{Z}'\boldsymbol{\varepsilon} \right) &\xrightarrow{d} N[0, \sigma^2 \mathbf{Q}_{zz}], \\ \text{и} \left(\frac{\mathbf{Z}'\mathbf{X}}{n} \right)^{-1} \left(\frac{1}{\sqrt{n}} \mathbf{Z}'\boldsymbol{\varepsilon} \right) &\xrightarrow{d} N[0, \sigma^2 \mathbf{Q}_{zx}^{-1} \mathbf{Q}_{zz} \mathbf{Q}_{xz}^{-1}]. \end{aligned}$$

Этот шаг завершает доказательство следующей теоремы.

Теорема 8.1. Асимптотическое распределение оценки метода инструментальных переменных

Если выполняются предположения А.1, А.2, А.13, А.4, А.5, А.17, А.18а-с и А.19 для $[y_i, \mathbf{x}_i, \mathbf{z}_i, \varepsilon_i]$, где \mathbf{z} является подходящим набором из $L = K$ инструментальных переменных, то асимптотическое распределение оценки методом инструментальных переменных $\mathbf{b}_{IV} = (\mathbf{Z}'\mathbf{X})^{-1} \mathbf{Z}'\mathbf{y}$ задается как

$$\mathbf{b}_{IV} \overset{a}{\sim} N \left[\beta, \frac{\sigma^2}{n} \mathbf{Q}_{zx}^{-1} \mathbf{Q}_{zz} \mathbf{Q}_{xz}^{-1} \right], \quad (8-7)$$

где $\mathbf{Q}_{zx} = \text{plim}(\mathbf{Z}'\mathbf{X}/n)$ и $\mathbf{Q}_{zz} = \text{plim}(\mathbf{Z}'\mathbf{Z}/n)$.

Для оценки **асимптотической ковариационной матрицы** нам требуется оценка параметра σ^2 . Естественной оценкой будет

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - x_i' \mathbf{b}_{IV})^2.$$

Коррекция на число степеней свободы будет излишней, поскольку все результаты здесь асимптотические и $\hat{\sigma}^2$ ни в каком случае нельзя считать несмещенной. (Тем не менее в большинстве компьютерных пакетов делается коррекция на число степеней свободы.) Запишем вектор остатков как

$$\mathbf{y} - \mathbf{X}\mathbf{b}_{IV} = \mathbf{y} - \mathbf{X}(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y}.$$

Подставим $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ и перегруппируем члены, чтобы получить $\hat{\boldsymbol{\varepsilon}} = [\mathbf{I} - \mathbf{X}(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}']\boldsymbol{\varepsilon}$. Тогда

$$\begin{aligned} \hat{\sigma}^2 = \frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{n} &= \frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{n} + \left(\frac{\boldsymbol{\varepsilon}'\mathbf{Z}}{n}\right) \left(\frac{\mathbf{X}'\mathbf{Z}}{n}\right)^{-1} \left(\frac{\mathbf{X}'\mathbf{X}}{n}\right) \left(\frac{\mathbf{Z}'\mathbf{X}}{n}\right)^{-1} \left(\frac{\mathbf{Z}'\boldsymbol{\varepsilon}}{n}\right) - \\ &\quad - 2\left(\frac{\boldsymbol{\varepsilon}'\mathbf{X}}{n}\right) \left(\frac{\mathbf{Z}'\mathbf{X}}{n}\right)^{-1} \left(\frac{\mathbf{Z}'\boldsymbol{\varepsilon}}{n}\right). \end{aligned}$$

Ранее мы показали, что можно (после некоторых преобразований) применить результат о произведении пределов по вероятности для того, чтобы получить предел по вероятности подобного выражения. Опуская формальный вывод, мы получаем, что $\hat{\sigma}^2$ является **состоятельной оценкой** параметра σ^2 из-за первого слагаемого. Второе и третье слагаемые сходятся к нулю. Для завершения нашего вывода нам осталось оценить $\text{Asy. Var}[\mathbf{b}_{IV}]$ как

$$\begin{aligned} \text{Est. Asy. Var}[\mathbf{b}_{IV}] &= \frac{1}{n} \left\{ \left(\frac{\hat{\boldsymbol{\varepsilon}}'\hat{\boldsymbol{\varepsilon}}}{n}\right) \left(\frac{\mathbf{Z}'\mathbf{X}}{n}\right)^{-1} \left(\frac{\mathbf{Z}'\mathbf{Z}}{n}\right) \left(\frac{\mathbf{X}'\mathbf{Z}}{n}\right)^{-1} \right\} = \\ &= \hat{\sigma}^2 (\mathbf{Z}'\mathbf{X})^{-1} (\mathbf{Z}'\mathbf{Z}) (\mathbf{X}'\mathbf{Z})^{-1}. \end{aligned} \quad (8-8)$$

8.3.3. Причины, приводящие к методу инструментальных переменных

При получении IV-оценки мы искали решения уравнений в (8.5):

$$\text{plim}(\mathbf{Z}'\mathbf{y}/n) = \text{plim}(\mathbf{Z}'\mathbf{X}/n)\boldsymbol{\beta}$$

или

$$\mathbf{Q}_{Zy} = \mathbf{Q}_{ZX}\boldsymbol{\beta}.$$

IV-оценка получалась как решение этой системы из K **моментных уравнений**. Поскольку здесь набор из K уравнений с K неизвестными, то если \mathbf{Q}_{ZX}^{-1} существует, то существует ровно одно решение для $\boldsymbol{\beta}$, задаваемое (8-6). Соответствующие моментные уравнения, если использовать только матрицу \mathbf{X} , приводят к

$$\text{plim}(\mathbf{X}'\mathbf{y}/n) = \text{plim}(\mathbf{X}'\mathbf{X}/n)\boldsymbol{\beta} + \text{plim}(\mathbf{X}'\boldsymbol{\varepsilon}/n) = \text{plim}(\mathbf{X}'\mathbf{X}/n)\boldsymbol{\beta} + \boldsymbol{\gamma}$$

или

$$\mathbf{Q}_{Xy} = \mathbf{Q}_{XX}\boldsymbol{\beta} + \boldsymbol{\gamma},$$

где, если не вводить дополнительные ограничения, K уравнений и $2K$ неизвестных. Чтобы решить эту систему для β или γ , числа имеющихся уравнений недостаточно. Дополнительные ограничения, которые позволят получить оценку для β , будут иметь вид $\gamma = 0$, что в точности является предположением А.3 об экзогенности. Как следствие вектор параметров β не **идентифицируется** при наличии лишь моментных уравнений для X и y — для него просто не существует единственного решения. Но в терминах моментных уравнений на Z , X и y из предположения о релевантности, или при наличии K ограничений, наложенных условием экзогенности, возможно вычислить оценку b_{IV} .

Рассмотрим полученные результаты применительно к простейшей модели

$$y = \beta x + \delta T + \varepsilon.$$

Для того чтобы метод наименьших квадратов состоятельно оценивал δ (и β), требуется предположить, что изменения в T экзогенны по отношению к модели и потому ковариация между y и T объясняется изменениями T , а не ε . Когда T и ε коррелированы и ε изменяется под влиянием некоторого фактора, не учтенного в уравнении, изменения в y оказываются связанными с изменениями в T , которые в свою очередь вызываются изменениями в ε , передаваемыми через T . Если T является экзогенной переменной, т.е. не коррелированной с ε , то изменения в ε не вызывают изменений в T (здесь мы используем термин «вызывают» в крайне неточном смысле) и не могут быть ошибочно приняты за экзогенные изменения в T . Именно это и утверждает условие экзогенности. Итак, для того чтобы верно идентифицировать параметр δ , необходимо предположить, что изменения в T никак не связаны с изменениями в ε . Если же это не так, то, как показывает формула (8-4), изменения в y связаны с дополнительным источником, с изменениями в ε , переданными через изменения в T . Это и есть влияние γ в (8-4). В этом случае, для того чтобы идентифицировать δ , нам требуется выявить изменение T , которое не вызвано изменением ε . Введем инструментальную переменную z . Если z является инструментальной переменной, $cov(z, T) \neq 0$ и $cov(z, \varepsilon) = 0$, то изменение в z вызывает то изменение, которое мы ищем. Если бы мы могли провести эксперимент для того, чтобы измерить «причинный эффект» изменения в T , нам бы пришлось изменить z , а затем посмотреть, насколько изменился y при изменении T , зная, что изменение в T было вызвано только изменением в z , но не в ε , т.е. $(\Delta y / \Delta z) / (\Delta T / \Delta z)$.

Пример 8.2. *Исследование с помощью метода инструментальных переменных*

В работах (Grootendorst (2007)) и (Deaton (1997)) подробно излагается исследование, в котором, судя по всему, впервые был применен метод инструментальных переменных.

Несмотря на то что IV-теория развивалась главным образом стараниями экономистов, сам метод произошел из эпидемиологии. Он использовался для исследования путей передачи холеры во время эпидемии

1853–1854 гг. в Лондоне. Один из ученых того времени, Джон Сноу (John Snow), предположил, что холера передается через воду. Для проверки своего предположения ему надо было понять, действительно ли те, кто пил более чистую воду, были менее подвержены заражению холерой. Другими словами, ему надо было изучить корреляцию между чистотой воды (x) и случаями заболевания холерой (y). Однако, как замечает Дитон (Deaton (1997)), это было бы не слишком убедительно: «Люди, пившие нечистую воду, скорее принадлежали к беднейшим слоям населения и жили во всевозможным образом загрязненной обстановке, причем не только «ядами миазмы», которые, как считали, и являлись причиной холеры». Сноу вместо этого ввел инструмент, который сильно коррелировал с чистотой воды, но не коррелировал с факторами, вызывающими холеру, как наблюдаемыми, так и ненаблюдаемыми. Этим инструментом была компания, снабжающая домохозяйства питьевой водой. В то время лондонцы использовали воду из реки Темзы в качестве питьевой. Одна из компаний, *Lambeth Water Company*, брала воду выше по течению реки, до главного слива сточных вод, другая же, *Southwark and Vauxhall Company*, брала воду ниже слива. Таким образом, инструмент z был строго коррелирован с чистотой воды x . Инструмент также не был коррелирован с ненаблюдаемыми факторами, вызывающими холеру. Согласно Сноу (Snow (1855, р. 74–75)), домохозяйства, обслуживаемые этими компаниями, были весьма похожи. Так, он пишет, что «смешение предложения было почти полным. Трубы каждой компании проходили по всем улицам, почти по всем дворам и аллеям... Эксперимент также был очень обширен. Не менее трехсот тысяч человек обоих полов, всех возрастов и видов занятий, всех рангов, занимавших разное общественное положение — от аристократии до беднейших слоев населения, были поделены на две группы без малейшего их участия и даже в большинстве случаев без их ведома. Одна группа, получающая воду, содержащую стоки Лондона, включая, быть может, использовавшуюся холерными больными, и другая группа, получающая воду более-менее свободную от подобных нечистот».

Пример 8.3. Реки как инструменты

В работе Хоксби (Hoxby (2000)) был исследован эффект влияния числа школ, которые можно «выбрать» на школьном «рынке», на успехи в учебе. Предполагалось, что исследуемые уравнения имеют следующий вид:

$$\frac{A_{ikm}}{\ln E_{km}} = \beta_1 C_m + \mathbf{x}'_{ikm} \beta_2 + \bar{\mathbf{x}}'_{.km} \beta_3 + \bar{\mathbf{x}}'_{..m} \beta_4 + \varepsilon_{ikm} + \varepsilon_{km} + \varepsilon_m,$$

где ikm — домохозяйство i в округе k на рынке m , A_{ikm} — показатель, характеризующий достижения в учебе, и E_{km} — подушевые расходы. Уравнение содержало данные по индивидам, средние по району и по рынку. Экзогенные переменные предназначались для того, чтобы учесть

различные источники гетерогенности на всех трех уровнях агрегирования. (Составные ошибки, которые мы будем рассматривать в моделях панельных данных в главе 10, также означают возможность рассмотрения случайных эффектов на всех трех уровнях.) Обосновав то, что возможность более широкого выбора, доступного ученикам, S_m , будет эндогенной переменной в этом уравнении, автор нашел надежную инструментальную переменную, которая «объясняла бы» (была коррелирована с) S_m , но не была бы коррелирована с ошибками в уравнении. На американском рынке границы школьных округов были определены в конце XVIII и на протяжении XIX столетий и до сих пор не изменены. Автор замечает, что изначально границы округов определялись в соответствии с естественными барьерами, возникающими при передвижении, например с реками. Из чего следует, как она замечает, что «число округов является возрастающей функцией от числа естественных барьеров»; следовательно, число рек на соответствующей территории и есть необходимая инструментальная переменная. [Дискуссионный вопрос исследования и нетрадиционный выбор инструментов привлекли внимание со стороны прессы, например <http://gsppi.berkeley.edu/faculty/jrothstein/hoxby/wsj.pdf>, и исследователей, включая Розштайна (Rothstein (2004)).] Это исследование представляет собой пример «естественного эксперимента», как описано в работе (Angrist, Pischke (2009)).

Пример 8.4. *Инструментальные переменные в регрессии*

Роль инструментальных переменных в задаче идентификации параметров регрессии была изучена в классическом исследовании (Working (1926)). Мы несколько упростили его здесь для нашего примера 8.1 о рыночном равновесии. На рис. 8.1a изображены «наблюдаемые данные» в равновесном состоянии рынка, на котором присутствуют случайные шоки (ε_S , ε_D), продавцы имеют дело с различными доходами покупателей и различными ценами факторов производства. Точки равновесия на рис. 8.1a получаются как совместный результат действия всех этих факторов. На рис. 8.1b изображены возможные условия спроса и предложения, которые могут привести к такому равновесию. Различные значения функции в уравнении предложения, соответствующие разным значениям цен факторов производства и разным значениям дохода, со стороны спроса приводят к девяти различным режимам, подверженным изменениям вследствие шоков. Имея в распоряжении лишь неопределенный набор точек, линейная регрессия выпуска (quantity) на цены (price) и доход (income) приведет примерно к результату, изображенному на рис. 8.1c частым пунктиром. Коэффициент наклона этой прямой очень слабо соотносится с коэффициентом наклона уравнений спроса. При столкновении с подобным результатом возникает вопрос: как вообще возможно узнать что-нибудь о кривой спроса? На рис. 8.2d изображен некоторый эксперимент, включающий два шага. (1) Мы считаем доход (income) постоянным и тогда можем сосредоточиться на конкретной

кривой спроса. Напомним, что рассматриваем функцию множественной регрессии и доход включен как условная переменная в уравнение. (2) Теперь мы сосредоточим наше внимание на одном из наборов данных по спросу (например, D_2) и будем двигать кривую предложения. В этом случае точки равновесия будут образовывать функцию спроса. Таким образом, мы получили функцию от изменяющихся цен факторов производства ($InputPrice$), которая и будет инструментом для идентификации функции (или функций) спроса в нашем эксперименте.

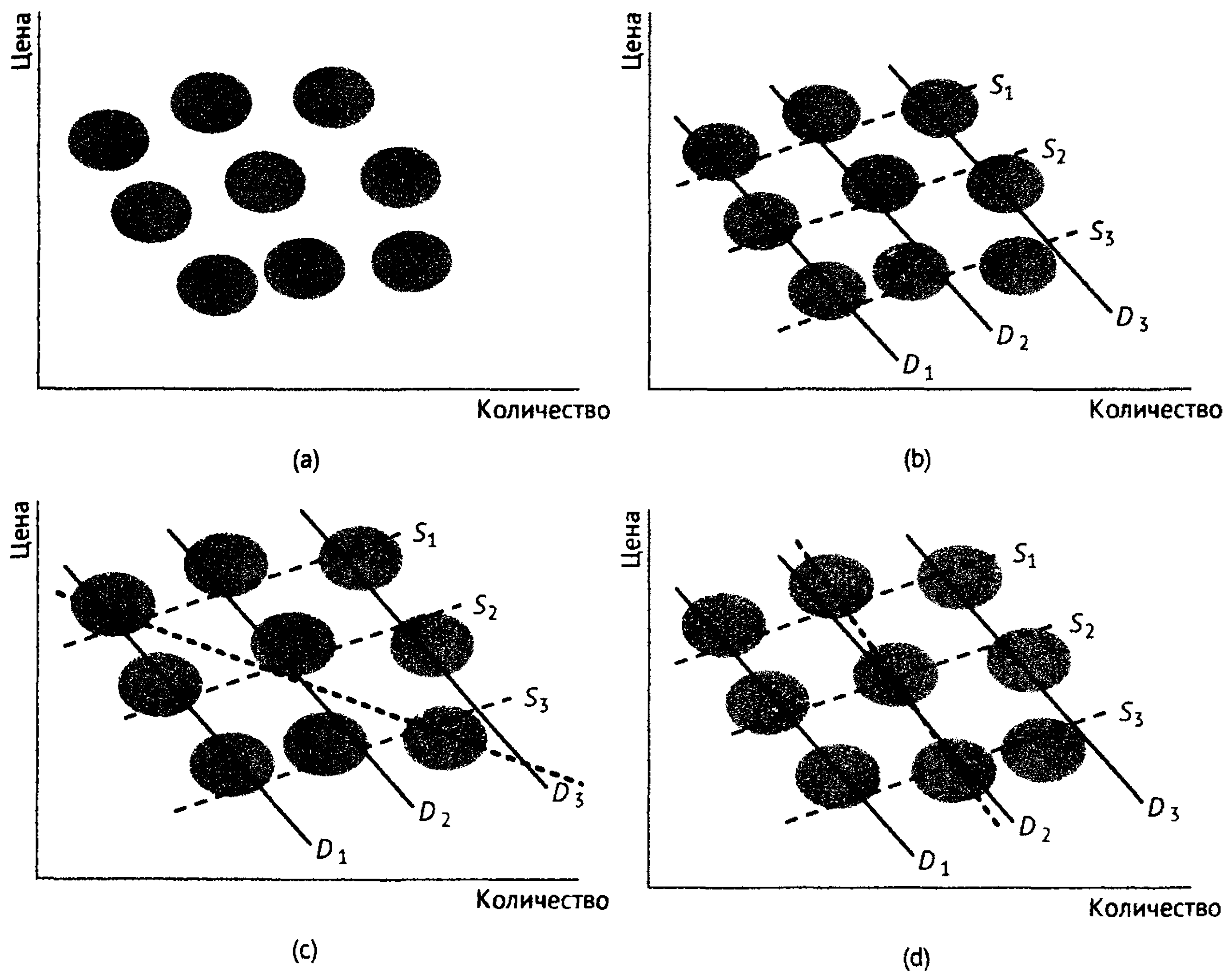


Рис. 8.1. Идентификация кривой спроса с помощью инструментальных переменных

8.3.4. Двухшаговый метод наименьших квадратов

До сих пор мы предполагали, что число инструментальных переменных в Z то же, что и число переменных (экзогенных и эндогенных) в X . (Обычно исследователю необходима одна инструментальная переменная для единственной эндогенной переменной в его уравнении. Тем не менее иногда бывает, что данные содержат дополнительные инструменты.) Вернемся в рыночному равновесию, рассмотренному в примерах 8.1 и 8.4. Предположим, что это был рынок сельскохозяйственных товаров с двумя экзогенными факторами условий для предложения: ценами факторов производства $InputPrice$ и количеством осадков $Rainfall$. Тогда уравнения модели имеют вид

(спрос) $Quantity_D = \alpha_0 + \alpha_1 Price + \alpha_2 Income + \varepsilon_D$,

(предложение) $Quantity_S = \beta_0 + \beta_1 Price + \beta_2 InputPrice + \beta_3 Rainfall + \varepsilon_S$,

(равновесие) $Quantity_D = Quantity_S$.

Ориентируясь на подход, предложенный в примере 8.4, кажется естественным выбрать одну из двух экзогенных переменных из уравнения предложения в качестве инструмента для идентификации уравнения спроса. (Мы вернемся к вопросу об идентификации уравнения предложения в подразделе 8.4.2.) Интуитивно ясно, что, ограничивая себя лишь частью множества всех доступных инструментов, мы теряем часть нужной информации. Таким образом, представляется совершенно ясным, что надо как-то использовать всю матрицу Z даже при $L > K$. Двухшаговый метод наименьших квадратов призван решить проблему использования всей информации, содержащейся в выборке, когда Z содержит более чем необходимое число переменных для построения IV-оценки.

Если Z содержит больше переменных, чем X , то предыдущие рассуждения не получится использовать, поскольку матрица $Z'X$ будет иметь размер $L \times K$ с рангом $K < L$ и не будет обратима. Существенным из полученных результатов остается условие $\text{plim}(Z'\varepsilon/n) = 0$. Таким образом, каждый столбец Z асимптотически не коррелирован с ε . Отсюда также следует, что любая линейная комбинация столбцов Z будет не коррелирована с ε , что приводит нас к новому подходу: составить K линейных комбинаций из столбцов Z . Однако какие именно комбинации выбрать? Одной из очевидных возможностей, минусы которой мы обсуждали выше, является простой выбор каких-нибудь K переменных из L , содержащихся в Z . Выкидывание информации, содержащейся в «дополнительных» $L - K$ столбцах, оборачивается неэффективностью оценки. Лучший способ представляет собой проекцию столбцов X в пространство, порожденное столбцами Z :

$$\hat{X} = Z(Z'Z)^{-1}Z'X.$$

Чуть ниже мы обсудим плюсы такого выбора. При таком выборе инструментальных переменных, \hat{X} вместо Z , получаем

$$b_{IV} = (\hat{X}'X)^{-1}\hat{X}'y = [X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}Z'y. \quad (8-9)$$

Оценкой асимптотической ковариационной матрицы будет $\hat{\sigma}^2$, умноженная на матрицу в скобках в формуле (8-9). Доказательство состоятельности и асимптотической нормальности этой оценки проводится так же, как и раньше, так как это доказательство остается верным для любого набора допустимых инструментов, к каковым относятся \hat{X} .

Есть две причины для использования этой оценки: теоретическая и практическая. Если какой-то из столбцов матрицы X встречается и в матрице Z , то этот же столбец матрицы X возникает и в матрице \hat{X} . Это несложно показать. Если в выражении для \hat{X} k -й столбец матрицы X является одним из столбцов матрицы Z (например, l -м), то k -й столбец матрицы $(Z'Z)^{-1}Z'X$ будет совпадать с l -м столбцом $L \times L$ идентифицирующей матрицы. Таким образом, k -й столбец матрицы $\hat{X} = Z(Z'Z)^{-1}Z'X$ будет l -м столбцом матрицы Z , который является k -м столбцом матрицы X . Это полезный и важный результат. Рассмотрим типичный пример. Пусть регрессия насчитывает K

переменных, из которых только одна, скажем k -я, коррелирована с шоками. Пусть у нас в наличии есть одна или несколько инструментальных переменных, как и $K - 1$ исходных переменных, которые являются инструментами сами по себе. Тогда в качестве инструментов мы будем использовать $\mathbf{Z} = [\mathbf{X}_{(k)}, \mathbf{z}_1, \mathbf{z}_2, \dots]$, где под индексом (k) понимаем удаление k -й переменной из матрицы. Другая полезная интерпретация матрицы $\hat{\mathbf{X}}$ состоит в том, что каждый ее столбец представляет собой набор подобранных значений, которые получаются при оценке регрессии соответствующего столбца матрицы \mathbf{X} на остальные столбцы \mathbf{Z} (эта интерпретация очевидно следует из определения). Отсюда также легко понять, почему каждый x_k , входящий в матрицу \mathbf{Z} , совершенно точно воспроизводится в матрице $\hat{\mathbf{X}}$. Каждый x_k представляет наилучший предиктор самого себя без какой-либо помощи остальных переменных в \mathbf{Z} . Таким образом, в нашем примере каждый столбец \mathbf{X} , за исключением последнего, который был исключен из матрицы $\mathbf{X}_{(k)}$, полностью повторяется, а исключенный столбец заменяется в матрице $\hat{\mathbf{X}}$ на свое подобранное значение из регрессии этой переменной на все переменные \mathbf{z} .

Из всех возможных линейных комбинаций из \mathbf{Z} $\hat{\mathbf{X}}$ является наиболее эффективной в том смысле, что асимптотическая ковариационная матрица IV-оценки, основанная на линейной комбинации $\mathbf{Z}\mathbf{F}$, меньше при $\mathbf{F} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}$, чем при любой другой \mathbf{F} , использующей все L столбцов матрицы \mathbf{Z} . Тем более этот результат исключает линейные комбинации, полученные без некоторых столбцов \mathbf{Z} . Этот факт был доказан в важной работе (Brundy, Jorgenson (1971)). [См. также Wooldridge (2002a, p. 96–97).]

Конец этого раздела мы посвятим некоторым практическим аспектам использования оценки инструментальных переменных. Простым перемножением матриц можно показать, что

$$\begin{aligned} b_{IV} &= (\hat{\mathbf{X}}'\mathbf{X})^{-1}\hat{\mathbf{X}}'\mathbf{y} \\ &= (\mathbf{X}'(\mathbf{I} - \mathbf{M}_Z)\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I} - \mathbf{M}_Z)\mathbf{y} \\ &= (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'\mathbf{y}, \end{aligned} \quad (8-10)$$

поскольку матрица $\mathbf{I} - \mathbf{M}_Z$ является идемпотентной. Таким образом, когда (и только когда) матрица $\hat{\mathbf{X}}$ представляет собой набор инструментальных переменных, IV-оценка представляет собой регрессию метода наименьших квадратов вектора \mathbf{y} на матрицу $\hat{\mathbf{X}}$. Отсюда предполагается (только для удобства, на самом деле нет никакой необходимости проводить два шага), что оценка b_{IV} может быть вычислена в два шага. Первым шагом будет вычисление матрицы $\hat{\mathbf{X}}$, а вторым — регрессия методом наименьших квадратов. По этой причине эту оценку часто называют *оценкой двухшагового метода наименьших квадратов* (two-stage least squares, 2SLS). Мы вернемся к этому представлению оценки еще несколько раз в дальнейшем, в частности при обсуждении моделей одновременных уравнений в разделе 10.5. Однако, применяя этот подход, следует соблюдать осторожность при вычислении асимптотической ковариационной матрицы: $\hat{\sigma}^2$ не должна строиться по матрице $\hat{\mathbf{X}}$. Оценка

$$s_{IV}^2 = \frac{(\mathbf{y} - b_{IV})'(\mathbf{y} - b_{IV})}{n}$$

является несостоятельной оценкой параметра σ^2 независимо от коррекции на число степеней свободы.

Очевидный вопрос состоит в том, где искать подходящие инструментальные переменные. Недавние работы по «естественным экспериментам» рассматривают изменения в политике, такие как «лодочный мост» из Мэриэля (пример 6.5), или природные показатели, такие как число рек (пример 8.3), или дни рождения [Angrist (1992, 1994)]. Во многих ситуациях, связанных с временными рядами, лагированные значения переменных в модели могут оказаться вполне естественным выбором. Во многих других случаях ответ на этот вопрос совершенно неочевиден. Асимптотическая ковариационная матрица может быть достаточно большой, если корреляция между Z и X слабая: элементы матрицы $(Z'X)^{-1}$ могут быть весьма велики. (См. разделы 8.7 и 10.6.6 для обсуждения «слабых» инструментов.) К сожалению, выбор инструментальных переменных обычно небогат. Часто выбор матрицы Z происходит *ad hoc*¹. Здесь присутствует некоторое затруднение. Кажется, что наилучшим выбором для инструмента будут переменные, которые сильно коррелируют с X . Но чем больше переменная коррелирует с проблемными столбцами матрицы X , тем более уязвима позиция, что эти же переменные не коррелируют с шоками.

Пример 8.5. Оценка уравнения предложения труда при помощи метода инструментальных переменных

Наиболее явным примером, в котором присутствует корреляция между регрессором и шоком, являются модели равновесного рынка. В работе Корнвелла и Руперта (Cornwell, Rupert (1988)) исследовалась отдача от образования в панели данных, состоящей из 595 наблюдений глав домохозяйств. Выборка для 1976–1982 гг. была взята из панельного исследования динамики доходов («Non-Survey of Economic Opportunity», the Panel Study of Income Dynamics). Оцениваемое уравнение имело вид

$$\ln Wage_{it} = \beta_1 + \beta_2 Exp_{it} + \beta_3 Exp_{it}^2 + \beta_4 Wks_{it} + \beta_5 Occ_{it} + \beta_6 Ind_{it} + \beta_7 South_{it} + \beta_8 SMSA_{it} + \beta_9 MS_{it} + \beta_{10} Union_{it} + \beta_{11} Ed_i + \beta_{12} Fem_i + \beta_{13} Blk_i + \varepsilon_{it},$$

где переменные обозначают

| | |
|-------|--|
| Exp | = число лет полной занятости (experience, опыт), |
| Wks | = количество отработанных недель (weeks), |
| Occ | = 1, если индивид является «синим воротничком» (occupation, род занятий), 0, если иначе, |
| Ind | = 1, если индивид принадлежит к производственной индустрии (industry), 0, если иначе, |
| South | = 1, если индивид проживает на юге, 0, если иначе, |
| SMSA | = 1, если индивид проживает в крупном городе (метрополия, SMSA), 0, если иначе, |
| MS | = 1, если индивид состоит в браке, 0, если иначе, |

¹ Результаты, касающиеся «оптимальных инструментов», встречаются в работах White (2001) и Hansen (1982). С другой стороны, существует большое количество работ, касающихся «слабых» инструментов, как, например, Staiger, Stock (1997), которые мы будем обсуждать в разделах 8.7 и 10.6.6.

| | |
|-------|---|
| Union | = 1, если зарплата индивида назначена по профсоюзному (union) контракту, 0, если иначе, |
| Ed | = число лет образования (education), |
| Fem | = 1 для женщин (female), 0, если иначе, |
| Blk | = 1 для черных (black), 0, если иначе. |

Данные приведены в табл. F8.1 приложения. Наибольший интерес в исследовании, помимо сравнения различных методов оценивания, представляет коэффициент β_{11} — отдача от образования. Это уравнение рассматривается в **усеченной (редуцированной) форме** (reduced form equation): оно включает все необходимые переменные, но не специфицирует структурные связи между ними. Напротив, модель, состоящая из трех уравнений, определенная в подразделе 8.3.4, носит название **структурная система уравнений** (structural equation system). Усеченная форма для этой модели будет состоять из отдельных регрессий *Price* (цены) и *Quantity* (выпуска) на $(1, Income, InputPrice, Rainfall)$. Мы вернемся к методам усеченных форм при рассмотрении моделей одновременных уравнений в главе 10. На настоящий момент из предложенной модели следует, что уравнение равновесия рынка показывает взаимодействие между спросом и предложением на рынке труда. Возможно рассмотреть рынок с точки зрения предложения труда. Уравнение предложения труда со стороны домохозяйств может иметь следующий вид:

$$Wks_{it} = \gamma_1 + \gamma_2 \ln Wage_{it} + \gamma_3 Ed_i + \gamma_4 Union_{it} + \gamma_5 Fem_i + u_{it}.$$

(В принципе можно рассмотреть и другой набор переменных в правой части уравнения.) Структурные уравнения более сложно специфицировать, чем их усеченные формы. Если количество отработанных недель и принятая заработная плата определяются совместно, то в этом уравнении $\ln Wage_{it}$ и u_{it} коррелированы. Мы рассмотрим две оценки инструментальных переменных, основанные на наборе переменных:

$$Z_1 = [1, Ind_{it}, Ed_i, Union_{it}, Fem_i]$$

и

$$Z_2 = [1, Ind_{it}, Ed_i, Union_{it}, Fem_i, SMSA_{it}].$$

На рис. 8.1 показаны три набора оценок. Оценка методом наименьших квадратов вычисляется при помощи стандартных методов, рассмотренных в главах 3 и 4. Одним из примечательных фактов является малость коэффициентов при переменной логарифма зарплаты. Второй набор оценок — это оценки методом инструментальных переменных, рассмотренным в подразделе 8.3.2. Заметим, что здесь мы используем единственный инструмент Ind_{it} . Как и ожидалось, коэффициент при логарифме заработной платы существенно выше. Оставшиеся коэффициенты, однако, несколько противоречат нашим представлениям. В принципе все три коэффициента отличаются от ожидаемых. Третий набор коэффициентов получен с помощью двухшагового метода наименьших квадратов, основанного на более широком наборе инструментальных переменных. В этом случае используются оба инструмента, *SMSA* и *Ind*.

Таблица 8.1. Оцененное уравнение спроса на труд

| Переменная | OLS | | IV с Z_1 | | IV с Z_2 | |
|------------|---------|---------------|------------|---------------|------------|---------------|
| | Оценка | Станд. ошибка | Оценка | Станд. ошибка | Оценка | Станд. ошибка |
| Константа | 4,7665 | 1,2153 | 8,8987 | 13,0590 | 0,7044 | 4,9997 |
| In Wage | 0,7326 | 0,1972 | 5,1828 | 2,2454 | 3,1518 | 0,8572 |
| Education | -0,1532 | 0,03206 | -0,4600 | 0,1578 | -0,3200 | 0,06607 |
| Union | -1,9960 | 0,1701 | -2,3602 | 0,2567 | -2,1940 | 0,1860 |
| Female | -1,3498 | 0,2642 | 0,6957 | 1,0650 | -0,2378 | 0,4679 |

8.4. Два теста на спецификацию модели

В нашей модели есть два момента, которые мы хотели бы скорее проверить, чем сразу постулировать их. Во-первых, как будет видно из вывода в подразделе 8.4.1, из двух рассмотренных оценок, методом наименьших квадратов и методом инструментальных переменных, оценка методом наименьших квадратов значительно более эффективна. IV-оценка, однако, является робастной: она состоятельна независимо от выполнения условия $\text{plim}(X'\epsilon/n) = 0$. Тем не менее если условие выполняется, т.е. если $\gamma = 0$, то оценка наименьших квадратов будет лучше, поскольку она имеет меньшую дисперсию². По этой причине, а к тому же желая проверить верность спецификации теоретической модели, мы заинтересованы в тесте, который бы давал некоторую информацию о смещении оценки наименьших квадратов. Во-вторых, при использовании оценки двухшаговым методом наименьших квадратов при $L > K$, т.е. с «дополнительными» инструментами, мы имеем $L - K$ ограничений на соотношения между переменными в модели. Как может быть видно из нашего вывода, если X насчитывает K переменных, некоторые из которых могут быть эндогенными, то должно насчитываться не менее K переменных в Z , чтобы параметры модели могли быть идентифицированы, т.е. чтобы с помощью информации в модели получить состоятельные оценки параметров. Если же у нас наблюдается избыток инструментов, то мы накладываем дополнительные, в общем-то излишние ограничения на процесс порождения данных. Вернемся еще раз к примеру рынка сельскохозяйственной продукции, рассмотренному в конце подраздела 8.3.3. В этой постановке вопроса, конечно, совершенно безопасно предположить, что переменная *Rainfall* (количество осадков) является экзогенным событием, которое не коррелирует с шоками уравнения спроса. Однако вполне возможно предположить, что вызванное взаимодействие рынков может быть таково, что *InputPrice* (цены факторов производства) могут быть коррелированы с шоками в уравнении спроса. На рынке

²Возможна, конечно, ситуация, что оценка наименьших квадратов, будучи несостоятельной, все равно будет более точной. Если LS лишь немного смещена, но имеет меньшую дисперсию, чем IV, то по критерию ожидаемого квадрата ошибки, дисперсия плюс квадрат смещения, оценка наименьших квадратов все равно предпочтительнее IV-оценки.

биотоплива зерно выступает как фактор производства в функции предложения рынка и как самостоятельный продукт на других рынках. Считать в этом примере *InputPrice* экзогенной величиной, т.е. считать выполненным предположение, что *InputPrice* не коррелирует с ε_D , в некотором смысле не нужно, поскольку параметры уравнения спроса могут быть оценены и без этого предположения. В данном разделе мы рассмотрим два теста на спецификацию модели, связанные с этими аспектами IV-оценивания.

8.4.1. Тесты Хаусмана и Ву на спецификацию модели

Вообще говоря, может быть совершенно неочевидно, что регрессоры модели коррелируют с шоками или что регрессоры измерены с ошибкой. Если таких ситуаций не происходит, то выгоднее использовать оценку методом наименьших квадратов (LS) вместо IV-оценки. Сравним две ковариационные матрицы в условиях нулевой гипотезы о том, что обе оценки являются состоятельными, т.е. при предположении $\text{plim}(1/n)\mathbf{X}'\varepsilon = \mathbf{0}$. Разность между асимптотическими ковариационными матрицами составляет

$$\begin{aligned} \text{Asy. Var}[\mathbf{b}_{IV}] - \text{Asy. Var}[\mathbf{b}_{LS}] &= \frac{\sigma^2}{n} \text{plim} \left(\frac{\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}}{n} \right)^{-1} - \\ &- \frac{\sigma^2}{n} \text{plim} \left(\frac{\mathbf{X}'\mathbf{X}}{n} \right)^{-1} = \frac{\sigma^2}{n} \text{plim} n [(\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X})^{-1} - (\mathbf{X}'\mathbf{X})^{-1}]. \end{aligned}$$

Для того чтобы сравнить две матрицы в скобках, сравним обратные к ним. Матрица, обратная к первой матрице, равна $\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X} = \mathbf{X}'(\mathbf{I} - \mathbf{M}_Z)\mathbf{X} = \mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{M}_Z\mathbf{X}$. Поскольку матрица \mathbf{M}_Z неотрицательно определена, то отсюда следует, что неотрицательно определена и матрица $\mathbf{X}'\mathbf{M}_Z\mathbf{X}$. Таким образом, $\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}$ равна $\mathbf{X}'\mathbf{X}$ минус неотрицательно определенная матрица. Поскольку матрица $\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}$ меньше с точки зрения сравнения матриц, чем $\mathbf{X}'\mathbf{X}$, то ее обратная матрица больше. При выполнении нулевой гипотезы асимптотическая ковариационная матрица LS-оценки никогда не больше, чем матрица IV-оценки. На самом деле она будет всегда меньше, за исключением случая, когда все столбцы \mathbf{X} полностью предсказываются регрессиями на \mathbf{Z} . Таким образом, мы показали, что если $\text{plim}(1/n)\mathbf{X}'\varepsilon = \mathbf{0}$, т.е. если LS-оценка состоятельна, то она является наилучшей оценкой. (Разумеется, это следует из уже известных нам результатов о плюсах оценки методом наименьших квадратов.)

Наш интерес к сравнению двух оценок вызван не только вопросом об эффективности. Интересующая нас гипотеза часто заключается в спецификации модели, а не в условии $\text{plim}(1/n)\mathbf{X}'\varepsilon = \mathbf{0}$. Искать ковариацию между \mathbf{X} и ε с помощью $(1/n)\mathbf{X}'\varepsilon$, разумеется, совершенно бесполезно, поскольку, как следует из нормальных уравнений, $(1/n)\mathbf{X}'\varepsilon = \mathbf{0}$. В важной работе Хаусманом (Hausman (1978)) был предложен альтернативный способ тестирования. [В более ранних работах Ву (Wu (1973)) и Дарбина (Durbin (1954)) был по сути предложен тот же тест.] Идея подхода Хаусмана состоит в следующем. В условиях нулевой гипотезы у нас есть две состоятельные оценки параметра β , \mathbf{b}_{LS} и \mathbf{b}_{IV} . При альтернативной гипотезе лишь одна из них, \mathbf{b}_{IV} , является состоятельной. Таким образом, предлагается исследовать разность

$\mathbf{d} = \mathbf{b}_{IV} - \mathbf{b}_{LS}$. В условиях нулевой гипотезы $\text{plim } \mathbf{d} = \mathbf{0}$, в то время как при альтернативе $\text{plim } \mathbf{d} \neq \mathbf{0}$. Используя тот же подход, что и в нескольких местах ранее, мы можем проверить эту гипотезу с помощью статистики Вальда:

$$H = \mathbf{d}' \{ \text{Est. Asy. Var}[\mathbf{d}] \}^{-1} \mathbf{d}.$$

Асимптотическая ковариационная матрица, требуемая для этого теста, равна

$$\begin{aligned} \text{Asy. Var}[\mathbf{b}_{IV} - \mathbf{b}_{LS}] = & \text{Asy. Var}[\mathbf{b}_{IV}] + \text{Asy. Var}[\mathbf{b}_{LS}] - \\ & - \text{Asy. Cov}[\mathbf{b}_{IV}, \mathbf{b}_{LS}] - \text{Asy. Cov}[\mathbf{b}_{LS}, \mathbf{b}_{IV}]. \end{aligned}$$

Теперь тест можно было бы провести непосредственно, но ситуация усложняется тем, что у нас нет выражения для члена ковариации. Хаусманом был получен фундаментальный результат, который позволяет нам применить этот подход. Слегка видоизмененный, он звучит так:

Ковариация между эффективной оценкой, \mathbf{b}_E , вектора параметров β и разницей между ней и неэффективной оценкой \mathbf{b}_I того же вектора параметров, $\mathbf{b}_E - \mathbf{b}_I$, равна нулю.

Определение 8.1. В нашем случае \mathbf{b}_E — это \mathbf{b}_{LS} , а \mathbf{b}_I — это \mathbf{b}_{IV} . Согласно результату Хаусмана мы получаем

$$\text{Cov}[\mathbf{b}_E, \mathbf{b}_E - \mathbf{b}_I] = \text{Var}[\mathbf{b}_E] - \text{Cov}[\mathbf{b}_E, \mathbf{b}_I] = 0$$

или

$$\text{Cov}[\mathbf{b}_E, \mathbf{b}_I] = \text{Var}[\mathbf{b}_E],$$

поэтому

$$\text{Asy. Var}[\mathbf{b}_{IV} - \mathbf{b}_{LS}] = \text{Asy. Var}[\mathbf{b}_{IV}] - \text{Asy. Var}[\mathbf{b}_{LS}].$$

Подставляя этот полезный результат в наше выражение для статистики Вальда и возвращаясь к эмпирическим оценкам этих соотношений, мы получаем

$$H = (\mathbf{b}_{IV} - \mathbf{b}_{LS})' \text{Est. Asy. Var}[\mathbf{b}_{IV}] - \text{Est. Asy. Var}[\mathbf{b}_{LS}] - 1(\mathbf{b}_{IV} - \mathbf{b}_{LS}).$$

В условиях нулевой гипотезы у нас есть две разные, но состоятельные оценки параметра σ^2 . Если мы будем использовать s^2 как более общепотребительную оценку, статистика превратится в

$$H = \frac{\mathbf{d}' \left[\left(\hat{\mathbf{X}}' \hat{\mathbf{X}} \right)^{-1} - (\mathbf{X}' \mathbf{X})^{-1} \right]^{-1} \mathbf{d}}{s^2}.$$

Возникает желание применить теорему о квадратичной форме полного ранга и нормальном векторе и сказать, что асимптотическое распределение нашей статистики есть хи-квадрат с K степенями свободы. Однако здесь этот метод не годится, и, более того, если только \mathbf{X} и \mathbf{Z} не имеют общих переменных, ранг матрицы в этой статистике меньше K , и обычная обратная матрица не существует. В большинстве случаев какие-то из переменных \mathbf{X} входят в \mathbf{Z} . (Почти в любой ситуации \mathbf{X} и \mathbf{Z} содержат константу.) Таким образом, известно, что некоторые из переменных \mathbf{X} не коррелированы с шоками. Например, часто бывает, что существует единственная переменная, являющаяся проблемной или измеренная с ошибкой. В этом случае наша гипотеза $\text{plim}(1/n)\mathbf{X}'\boldsymbol{\varepsilon} = \mathbf{0}$ не затрагивает все K переменных, потому что известно, что часть элементов этого вектора (скажем, K_0) равна нулю. Таким образом, квадратичная форма в тесте Вальда тестирует лишь $K^* = K - K_0$ гипотез. Легко (и достаточно полезно) показать, что на самом деле H является квадратичной формой ранга K^* . Поскольку матрица $\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ является идемпотентной, то $(\hat{\mathbf{X}}'\hat{\mathbf{X}}) = \hat{\mathbf{X}}'\mathbf{X}$. Используя это и расписывая \mathbf{d} , получаем

$$\begin{aligned} \mathbf{d} &= (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'\mathbf{y} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}[\hat{\mathbf{X}}'\mathbf{y} - (\hat{\mathbf{X}}'\hat{\mathbf{X}})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}] \\ &= (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'(\mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}) \\ &= (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'\mathbf{e}, \end{aligned}$$

где \mathbf{e} — вектор остатков метода наименьших квадратов. Напомним, что K_0 столбцов матрицы $\hat{\mathbf{X}}$ — это исходные переменные из \mathbf{X} . Предположим, что эти переменные составляют первые столбцы. Таким образом, первые K_0 строк матрицы $\hat{\mathbf{X}}'\mathbf{e}$ совпадают с первыми K_0 строками матрицы $\mathbf{X}'\mathbf{e}$, которые, разумеется, равны $\mathbf{0}$. (Это не значит, что первые K_0 элементов \mathbf{d} равны нулю.) Таким образом, мы можем записать \mathbf{d} как

$$\mathbf{d} = (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1} \begin{bmatrix} \mathbf{0} \\ \hat{\mathbf{X}}^{*'}\mathbf{e} \end{bmatrix} = (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1} \begin{bmatrix} \mathbf{0} \\ \mathbf{q}^* \end{bmatrix},$$

где \mathbf{X}^* представляет собой K^* переменные из \mathbf{x} , которые не входят в \mathbf{z} .

Наконец, обозначим всю матрицу в H как \mathbf{W} . (Поскольку обычной обратной матрицы может не существовать, эта матрица должна иметь обобщенную обратную матрицу, см. подраздел А.6.12.) Тогда, обозначая все произведение матриц за \mathbf{P} , мы получаем

$$H = [\mathbf{0}' \mathbf{q}^{*'}](\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\mathbf{W}(\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1} \begin{bmatrix} \mathbf{0} \\ \mathbf{q}^* \end{bmatrix} = [\mathbf{0}' \mathbf{q}^{*'}]\mathbf{P} \begin{bmatrix} \mathbf{0} \\ \mathbf{q}^* \end{bmatrix} = \mathbf{q}^{*'}\mathbf{P}_{**}\mathbf{q}^*,$$

где \mathbf{P}_{**} — нижний правый блок матрицы \mathbf{P} размером $K^* \times K^*$. Мы получили конечный результат. Алгебраически H представляет собой квадратичную форму вектора длины K^* , поэтому для теста мы будем использовать K^* степеней свободы.

Рассмотренный тест Вальда требует вычисления обобщенной обратной матрицы [Hausman, Taylor (1981)], что может быть несколько громоздко.

В принципе необязательно проводить этот тест именно в указанной форме, и к тому же он может быть проведен в любой регрессионной программе. Альтернативный подход *тест на включение переменной* (variable addition test), предложенный в работе Ву (Wu (1973)), оказывается проще. F -статистика с K^* и $n - K - K^*$ степенями свободы может быть использована для проверки совместной значимости элементов γ в расширенной регрессии

$$y = X\beta + \hat{X}^*\gamma + \epsilon^*, \quad (8-11)$$

где \hat{X}^* — подобранные значения в регрессиях переменных X^* на Z . Этот результат эквивалентен тесту Хаусмана для этой модели. [Алгебраические выкладки для этого результата можно найти в соответствующих статьях и в книге Дэвидсона и Маккиннона (Davidson, MacKinnon (2004, Section 8.7)).]

Пример 8.6. Модель предложения труда

Для уравнения предложения труда, оцененного в примере 8.5, мы используем тест Ву (дополнения переменной) для проверки эндогенности переменной $\ln Wage$ (логарифм заработной платы). На первом шаге оценивается регрессия $\ln Wage_{it}$ на $z_{1,it}$. Подбранное значение из этого уравнения далее добавляется в регрессию методом наименьших квадратов Wks_{it} на x_{it} . Мы получаем следующие результаты регрессии:

$$\begin{aligned} \widehat{Wks}_{it} = & 18,8987 + 0,6938 \ln Wage_{it} - 0,4600 Ed_i - 2,3602 Union_{it} + \\ & (12,3284) \quad (0,1980) \quad (0,1490) \quad (0,2423) \\ & + 0,6958 Fem_i + 4,4891 \ln \widehat{Wage}_{it} + u_{it}, \\ & (1,0054) \quad (2,1290) \end{aligned}$$

где в скобках указаны оцененные стандартные ошибки. t -статистика коэффициента при логарифме зарплаты равна 2,108, что превышает критическое значение нормального распределения, равное 1,96. Таким образом, гипотеза об экзогенности переменной $\ln Wage$ отвергается.

Несмотря на то что почти все предшествующие результаты были посвящены конкретному тесту о корреляции некоторых столбцов X и шоков ϵ , тест Хаусмана является общим. Возвратимся к ситуации, в которой у нас есть пара оценок $\hat{\theta}_E$ и $\hat{\theta}_I$, которые при гипотезе H_0 становятся состоятельными, причем $\hat{\theta}_E$ будет эффективной в отличие от $\hat{\theta}_I$, в то время как при альтернативе H_1 $\hat{\theta}_I$ остается состоятельной, в то время как $\hat{\theta}_E$ становится несостоятельной. Тогда мы можем построить тестовую статистику, называемую *статистикой Хаусмана*,

$$H = (\hat{\theta}_I - \hat{\theta}_E)' \text{Est. Var} [\hat{\theta}_I] - \text{Est. Var} [\hat{\theta}_E] - 1 (\hat{\theta}_I - \hat{\theta}_E) \xrightarrow{d} \chi^2[J],$$

и ориентироваться на подходящее критическое значение для таблиц хи-квадрат-распределения. Количество необходимых степеней свободы J зависит от конкретной ситуации. Более того, могут понадобиться обобщенные

обратные матрицы, хотя по крайней мере в одном частном случае, в модели со случайными эффектами (см. главу 11), для этого подхода требуется брать некоторые строки и столбцы матрицы. Ситуация, при которой возникает укороченный ранг матрицы, также не является общей. Во многих ситуациях этот метод применим с квадратичной формой полного ранга. Более того, можно применять подход Ву, как в примере выше. В других примерах, которые мы будем рассматривать, таких как тест независимости от неподходящих альтернатив для мультиномиальной логит-модели, тест не ограничивается регрессионным подходом и проводится с помощью статистики Вальда и квадратичной формы полного ранга. Наконец, заметим, что неполный ранг матрицы в статистике Вальда — это алгебраический результат. Иногда в конечных выборках можно столкнуться с тем, что матрица в статистике Вальда не является положительно определенной. Эта проблема не относится к структуре самой модели. В этом случае получение решения с помощью обобщенной обратной матрицы может приводить к неверным результатам. Хаусман предлагает в подобных случаях считать результат нулевым, т.е. не отвергать нулевую гипотезу.

Пример 8.7. Тест Хаусмана для функции потребления

Данные с I квартала 1950 г. по IV квартал 2000 г. по ряду макроэкономических переменных приведены в табл. F5.2 приложения. Функция потребления вида $C_t = \alpha + \beta Y_t + \varepsilon_t$ оценивается на 203 наблюдениях агрегированного реального потребления и реального располагаемого дохода, пропуская первый регрессор. Модель, по всей видимости, имеет смещение ввиду корреляции между Y_t и ε_t . Рассмотрим оценку методом инструментальных переменных, используя Y_{t-1} и C_{t-1} в качестве инструментов для Y_t , разумеется, константа используется как инструмент для самой себя. Одно наблюдение теряется, поскольку мы используем лагированные значения и результаты основываются на 203 наблюдениях. Статистика Хаусмана может быть вычислена одним из двух способов:

- с помощью статистики Вальда для H с использованием обобщенного обращения Мура–Пенроуза. Оценка s^2 считается по методу наименьших квадратов в условиях нулевой гипотезы об отсутствии корреляции. При таком подходе $H = 8,481$. Число степеней свободы $K^* = 1$. 95%-е критическое значение распределения хи-квадрат равно 3,84. Таким образом, мы отвергаем нулевую гипотезу об отсутствии корреляции между Y_t и ε_t ;
- используя статистику Ву, основанную на (8-11), мы оцениваем регрессию C_t на константу Y_t и подобранное значение регрессии Y_t на константу Y_{t-1} и C_{t-1} . t -статистика, соответствующая коэффициенту при подобранном значении, равна 2,968, поэтому F -статистика с 1 и 200 степенями свободы равна 8,809. Критическое значение этого F -распределения равно 3,888, и опять нулевая гипотеза отвергается.

8.4.2. Тест на сверхидентифицирующие ограничения

Идея, которая лежит в основе выбора в качестве метода оценивания метода инструментальных переменных, не преследует эффективность. Построенная оценка преследует лишь состоятельность, а вопрос об эффективности не рассматривается. В главе 13 мы вернемся к этому моменту в рамках получения эффективных оценок методом моментов. Тот факт, что 2SLS демонстрирует наиболее эффективное использование всех L -инструментов, говорит лишь о его эффективности в классе всех оценок, использующих K линейных комбинаций столбцов \mathbf{Z} . IV выводится из **условий ортогональности** (orthogonality conditions):

$$E[\mathbf{z}_i \varepsilon_i] = \mathbf{0}. \quad (8-12)$$

Выборочным аналогом этих условий будут **моментные условия** (моментные уравнения, moment equation):

$$\frac{1}{n} \sum_{i=1}^n \mathbf{z}_i \varepsilon_i = \mathbf{0}. \quad (8-13)$$

Решение при $L = K$ задается как $\mathbf{b}_{IV} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y}$, как мы уже убедились. Если $L > K$, решение не единственное, и мы избрали нашей стратегией метод 2SLS. Оценка все так же основывается на (8-13). Однако выборочный аналог уже представляет собой L уравнений с K неизвестными и (8-13) не имеет решения. Тем не менее при выполнении условий модели (8-12) остается справедливым. Мы можем рассматривать дополнительные ограничения как гипотезы, которые подтверждаются или не подтверждаются наблюдаемыми данными. Избыток моментных условий дает возможность построить тест на **сверхидентифицирующие ограничения** в модели. Тест будет основываться на уравнениях (8-13), которые, будучи вычисленными в \mathbf{b}_{IV} , не будут равны нулю при $L > K$, несмотря на то что гипотеза (8-12) может оставаться справедливой.

Тестовой будет статистика Вальда. (См. раздел 5.4.) Выборочная статистика, основывающаяся на (8-13) и оценке инструментальных переменных, будет равна

$$\bar{\mathbf{m}} = \frac{1}{n} \sum_{i=1}^n \mathbf{z}_i e_{IV,i} = \frac{1}{n} \sum_{i=1}^n \mathbf{z}_i (y_i - \mathbf{x}_i' \mathbf{b}_{IV}).$$

Статистика Вальда равна

$$\chi^2[L - K] = \bar{\mathbf{m}}' [\text{Var}(\bar{\mathbf{m}})]^{-1} \bar{\mathbf{m}}.$$

Для завершения нашего построения нам потребуется оценка дисперсии. Здесь есть два пути. При предположениях модели

$$\text{Var}[\bar{\mathbf{m}}] = \frac{\sigma^2}{n^2} \mathbf{Z}'\mathbf{Z},$$

что легко может быть оценено с помощью выборочной оценки параметра σ^2 . Другой путь — основывать нашу оценку на (8-12), откуда следует, что подходящей оценкой будет

$$\text{Est. Var}[\bar{m}] = \frac{1}{n^2} \sum_{i=1}^n (\mathbf{z}_i e_{IV,i})(\mathbf{z}_i e_{IV,i})' = \frac{1}{n^2} \sum_{i=1}^n e_{IV,i}^2 \mathbf{z}_i \mathbf{z}_i'$$

Эти две оценки будут численно различаться в конечных выборках, но при сделанных предположениях обе (умноженные на n) будут сходиться к одной матрице, поэтому выбор между ними не является существенным. В настоящее время на практике чаще используют второй вариант. Статистика Вальда в таком случае равна

$$\left(\frac{1}{n} \sum_{i=1}^n \mathbf{z}_i e_{IV,i} \right)' \left[\frac{1}{n^2} \sum_{i=1}^n e_{IV,i}^2 \mathbf{z}_i \mathbf{z}_i' \right]^{-1} \left(\frac{1}{n} \sum_{i=1}^n \mathbf{z}_i e_{IV,i} \right).$$

Единственное, что осталось, — это разобраться с числом степеней свободы. Тест может обнаружить невыполнение $L-K$ моментных условий, т.е. это и есть ранг квадратичной формы. Предельным распределением статистики будет хи-квадрат-распределение с $L-K$ степенями свободы.

Пример 8.8. Сверхидентификация в уравнении предложения труда

В примере 8.5 мы вычисляли оценки параметров уравнения для отработанных недель с помощью 2SLS-оценок:

$$x = [1, \ln Wage, Education, Union, Female]$$

и

$$z = [1, Ind, Education, Union, Female, SMSA].$$

Здесь есть одно сверхидентифицирующее ограничение. Выборочный момент, построенный по результатам 2SLS-оценки из табл. 8.1, имеет вид

$$(1/4165)Z'e_{2SLS} = [0; 0,03476; 0; 0; 0; -0,01543]'$$

Хи-квадрат-статистика с одной степенью свободы равна 1,09399. Если же использовать первый предложенный вариант оценки дисперсии, то статистика равна 1,05241. В обоих случаях они меньше 95%-го критического значения, равного 3,84, поэтому гипотеза о сверхидентифицирующих ограничениях не отвергается.

Сделаем еще заключительное замечание относительно этого теста. Можно решить, основываясь на соответствующей теории, что тест на сверхидентифицирующие ограничения относится только к одной инструментальной переменной, а не какой-нибудь другой. Так, в нашем примере о равновесном рынке с двумя инструментами для уравнения спроса — количеством осадков (*Rainfall*) и ценой факторов производства (*InputPrice*) — количество осадков, очевидно, является экзогенной переменной, поэтому отвержение гипотезы о сверхидентификации будет свидетельствовать о том, что *InputPrice* не является допустимым инструментом. Однако подобные выводы могут оказаться неверными: тест лишь указывает, что один или более

элементов в (8-12) являются ненулевыми. Он не указывает, о каких именно элементах идет речь.

8.5. Ошибка измерения

До сих пор мы предполагали (по крайней мере неявно), что данные, используемые нами для оценивания параметров, являются корректными измерениями своих теоретических аналогов. На практике такая ситуация возможна лишь при исключительно благоприятных обстоятельствах. Обычно наши данные содержат ошибки измерения всех возможных видов. Даже аккуратно собранные данные для исследования не всегда оказываются именно теми переменными, которые нужны нам для регрессий. Одни агрегированные показатели, такие как ВВП, есть лишь некоторые оценки истинных значений, а другие, такие как амортизация, капитал или «ставка процента», даже не существуют в согласованной теории. Наконец, и это худший случай, может вообще не быть физической меры, отражающей интересующий нас показатель: умственные способности, образование, постоянный доход — вот несколько примеров подобных величин. Тем не менее все эти величины включаются в точно специфицированные модели регрессий.

8.5.1. Недооценка метода наименьших квадратов

В этом подразделе мы обсудим некоторые известные результаты, касающиеся регрессий с плохо измеренными данными. Общие выводы здесь неутешительны. Смещение, вызываемое ошибкой измерения, может быть весьма значительным. Практически не существует результатов для модели с ошибкой измерения, касающихся ее свойств в конечных выборках: большинство полученных результатов носит асимптотический характер³. Рассмотрим несколько простых асимптотических результатов в рамках классической регрессионной модели.

Простейшим случаем для исследования будет модель парной линейной регрессии без константы. Несмотря на то что подобная модель, естественно, абсолютно нереалистична, на ее примере мы увидим некоторые существенные свойства, которые потом сможем распространить на более интересные модели. Пусть модель

$$y^* = \beta x^* + \varepsilon \quad (8-14)$$

удовлетворяет всем предположениям классической модели регрессии. Имея данные по y^* и x^* , мы можем оценить β методом наименьших квадратов. Предположим, однако, что наблюдаемые данные есть неидеально измеренные настоящие y^* и x^* . Так, пусть в нашем примере y^* — это $\ln(\text{output}/\text{labor})$, а x^* — $\ln(\text{capital}/\text{labor})$. Ни один фактор производства не может быть точно измерен, поэтому наблюдаемые y и x содержат ошибки измерения. Предположим, что

$$y = y^* + v, \quad \text{где } v \sim N[0, \sigma_v^2], \quad (8-15a)$$

$$x = x^* + u, \quad \text{где } u \sim N[0, \sigma_u^2]. \quad (8-15b)$$

³Imbens, Hyslop (2001).

Предположим также, что u и v независимы как между собой, так и от y^* и x^* . (Как мы увидим ниже, этих условий недостаточно для того, чтобы избежать проблем.)

Сначала подставим (8-15а) в (8-14), предполагая на данном этапе, что лишь y^* измерен с ошибкой:

$$y = \beta x^* + \varepsilon + v = \beta x^* + \varepsilon'.$$

Это уравнение удовлетворяет предположениям классической регрессионной модели. До тех пор пока нет ошибки измерения в регрессоре, ошибка измерения в зависимой переменной может быть включена в шоки регрессии и проигнорирована. В дальнейшем, чтобы слегка упростить изложение, мы будем предполагать, что проблемы ошибки измерения относятся лишь к регрессорам модели.

Рассмотрим теперь регрессию y на наблюдаемые x . Подставляя (8-15b) в (8-14), мы получаем

$$y = \beta x + [\varepsilon - \beta u] = \beta x + w. \quad (8-16)$$

Поскольку x равен $x^* + u$, регрессор в (8-16) коррелирован с шоком:

$$\text{Cov}[x, w] = \text{Cov}[x^* + u, \varepsilon - \beta u] = -\beta \sigma_u^2. \quad (8-17)$$

Таким образом, нарушается одно из главных предположений классической модели и вполне ожидаемо, что оценка методом наименьших квадратов

$$\mathbf{b} = \frac{(1/n) \sum_{i=1}^n x_i y_i}{(1/n) \sum_{i=1}^n x_i^2}$$

будет несостоятельна. Для нахождения ее предела по вероятности подставим (8-14) и (8-15b) в это выражение и воспользуемся теоремой Слуцкого:

$$\text{plim } \mathbf{b} = \frac{\text{plim}(1/n) \sum_{i=1}^n (x_i^* + u_i)(\beta x_i^* + \varepsilon_i)}{\text{plim}(1/n) \sum_{i=1}^n (x_i^* + u_i)^2}.$$

Так как x^* , ε и u независимы в совокупности, мы можем упростить это выражение до вида

$$\text{plim } \mathbf{b} = \frac{\beta Q^*}{Q^* + \sigma_u^2} = \frac{\beta}{1 + \sigma_u^2/Q^*}, \quad (8-18)$$

где $Q^* = \text{plim}(1/n) \sum_i x_i^{*2}$. Поскольку величина σ_u^2 положительна, оценка b является несостоятельной с постоянным смещением в сторону нуля. Очевидно, что чем выше разброс ошибки измерения, тем хуже смещение. Эффект смещения коэффициента в сторону нуля называется **недооценкой** (attenuation).

В модели множественной регрессии ситуация становится только хуже. Для начала допустим, что $y = \mathbf{X}^* \beta + \varepsilon$ и $\mathbf{X} = \mathbf{X}^* + \mathbf{U}$, предполагая, что наблюдение каждого регрессора может содержать ошибку. Обобщением ранее

полученного результата будет

$$\text{plim} \left(\frac{\mathbf{X}'\mathbf{X}}{n} \right) = \mathbf{Q}^* + \Sigma_{uu} \text{ и } \left(\frac{\mathbf{X}'\mathbf{y}}{n} \right) = \mathbf{Q}^*\beta.$$

Следовательно,

$$\text{plim } \mathbf{b} = [\mathbf{Q}^* + \Sigma_{uu}]^{-1} \mathbf{Q}^*\beta = \beta - [\mathbf{Q}^* + \Sigma_{uu}]^{-1} \Sigma_{uu}\beta. \quad (8-19)$$

Этот предел по вероятности представляет собой смесь различных параметров модели. В принципе, как и ранее, можно произвести **идентификацию** параметров модели, используя некоторую внешнюю информацию. Однако объем необходимой информации должен быть очень значительным, поэтому этот подход не представляется перспективным.

Часто возникает ситуация, когда с ошибкой измерена лишь одна переменная. Можно было бы предположить, что возникающие при этом проблемы должны быть схожи со случаем единственного коэффициента. Однако, к сожалению, это не так. В этом случае (предположим, что проблемной переменной является первая) матрица Σ_{uu} имеет вид

$$\Sigma_{uu} = \begin{bmatrix} \sigma_u^2 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ & & \vdots & \\ 0 & 0 & \dots & 0 \end{bmatrix}.$$

Можно показать, что в этом частном случае

$$\text{plim } b_1 = \frac{\beta_1}{1 + \sigma_u^2 q^{*11}} \quad (8-20a)$$

[заметим сходство этой формулы с (8-18)] и для $k \neq 1$

$$\text{plim } b_k = \beta_k - \beta_1 \left[\frac{\sigma_u^2 q^{*k1}}{1 + \sigma_u^2 q^{*11}} \right], \quad (8-20b)$$

где q^{*k1} есть $(k, 1)$ -й элемент матрицы $(\mathbf{Q}^*)^{-14}$. Эта формула включает несколько неизвестных параметров и не может быть оценена. Коэффициент при плохо измеренной переменной будет смещен, как и прежде, к нулю. Другие коэффициенты тоже будут смещены, однако направление смещения установить не удастся. Плохо измеренная величина портит все оценки метода наименьших квадратов⁵. Если же в модели более чем одна переменная содержит ошибку измерения, то известных результатов еще меньше⁶. Несмотря на то что для смещений в некоторых случаях можно получить выражения, обычно они зависят от ряда параметров, знаки и значения которых неизвестны и, судя по всему, не могут быть известны.

⁴Здесь используется (А-66) для обращения матрицы $[\mathbf{Q}^* + \Sigma_{uu}] = [\mathbf{Q}^* + (\sigma_u \mathbf{e}_1)(\sigma_u \mathbf{e}_1)']$, где \mathbf{e}_1 есть первый столбец единичной матрицы $K \times K$. Остальное получается непосредственно.

⁵Этот момент важно помнить, если есть подозрение в присутствии ошибки измерения в одной из переменных модели.

⁶Некоторые точные аналитические выражения были получены в работах Levi (1973), Theil (1961), Klepper and Leamer (1983), Garber and Klepper (1980), Griliches (1986) и Cragg (1997).

8.5.2. Оценка методом инструментальных переменных

Альтернативный способ оценивания этой модели (и некоторых других) основывается на методе инструментальных переменных. Вернемся к модели ошибок в данных, описанной уравнениями (8-14) и (8-15а, б). Параметры β , σ_ε^2 , q^* и σ_u^2 не могут быть идентифицированы в рамках x и y . Предположим, однако, что существует переменная z такая, что z коррелирована с x^* , но не коррелирована с u . Например, при исследовании домохозяйств о доходе обычно сообщают неохотно, частично это делается сознательно, частично из-за того, что респонденты игнорируют второстепенные источники дохода. Но предположим, что можно определить общую сумму оплаченных главой домохозяйства счетов. Очень похоже, что такая переменная z будет сильно коррелирована с доходом, но, вероятно, не будет существенно коррелировать с ошибками измерения. Если $\text{Cov}[x^*, z]$ не равна нулю, то становится возможным оценить параметры модели как

$$\text{plim} \frac{(1/n) \sum_i y_i z_i}{(1/n) \sum_i x_i z_i} = \frac{\beta \text{Cov}[x^*, z]}{\text{Cov}[x^*, z]} = \beta. \quad (8-21)$$

В общем случае $y = \mathbf{X}^* \beta + \varepsilon$, $\mathbf{X} = \mathbf{X}^* + \mathbf{U}$, предположим, что существует матрица переменных \mathbf{Z} , которая не коррелирована с шоками и ошибками измерения, но коррелирована с регрессорами \mathbf{X} . Тогда оценка методом инструментальных переменных, основанная на матрице \mathbf{Z} , $\mathbf{b}_{IV} = (\mathbf{Z}' \mathbf{X})^{-1} \mathbf{Z}' \mathbf{y}$, является состоятельной и асимптотически нормально распределенной с асимптотической ковариационной матрицей, которая может быть оценена как

$$\text{Est. Asy. Var}[\mathbf{b}_{IV}] = \hat{\sigma}^2 [\mathbf{Z}' \mathbf{X}]^{-1} [\mathbf{Z}' \mathbf{Z}] [\mathbf{X}' \mathbf{Z}]^{-1}. \quad (8-22)$$

Для более общих случаев применимы теорема 8.1 и результаты раздела 8.3.

8.5.3. Прокси-переменные

В некоторых случаях переменная в модели просто не имеет наблюдаемого аналога. Образование, способности, факторы предпочтений — вот, пожалуй, наиболее часто встречающиеся примеры. В этом случае, если не найден наблюдаемый индикатор для такой переменной, модель будет оцениваться в условиях пропущенной переменной. Обычно такой индикатор удается найти: для перечисленных переменных наиболее часто используют число лет учебы и оценки при написании тестов всевозможных видов. После этого с такими переменными имеют дело в рамках модели с ошибками измерения. Если, например,

$$\text{income} = \beta_1 + \beta_2 \text{education} + \varepsilon$$

и

$$\text{years of schooling} = \text{education} + u,$$

то мы можем использовать модель подраздела 8.5.1. Единственным отличием здесь будет то, что истинная переменная нашей модели является латентной (скрытой). Никакое улучшение качества измерения не приблизит прокси к переменной, которую она «представляет».

Преыдушие выводы могут показаться еще более удручающими, чем есть на самом деле. Рассмотрим *структурную модель*

$$Earnings = \beta_1 + \beta_2 Experience + \beta_3 Industry + \beta_4 Ability + \varepsilon.$$

Способности (*Ability*) ненаблюдаемы, но предположим, что у нас есть некоторый индикатор *IQ*. Если мы предположим, что *IQ* связан с переменной *Ability* соотношением

$$IQ = \alpha_1 + \alpha_2 Ability + v,$$

то можем выразить переменную *Ability* из этого уравнения и подставить в первое уравнение, что приводит нас к *уравнению в приведенной форме* (reduced form equation):

$$Earnings = (\beta_1 - \beta_4 \alpha_1 / \alpha_2) + \beta_2 Experience + \beta_3 Industry + (\beta_4 / \alpha_2) IQ + (\varepsilon - v \beta_4 / \alpha_2).$$

Это уравнение является внутренне линейным и может быть оценено методом наименьших квадратов. Тем самым мы не получим состоятельные оценки для β_1 и β_4 , но получим их для интересующих нас коэффициентов β_2 и β_3 . Такой подход решает нашу проблему. Отметим существенные части нашего решения: нам потребовалось, чтобы *индикатор IQ* не был связан с остальными переменными модели, а также чтобы v не коррелировал со всеми переменными. В этом случае некоторые параметры структурной модели идентифицировались в рамках имеющихся у нас наблюдений. Отметим, однако, что *IQ* является не прокси, а индикатором латентной переменной *Ability*. Такой подход к моделированию характерен преимущественно для исследований в области образования и психологии образования. Рассмотрим, как следовало поступать в предыдущей модели, если бы у нас был не единственный индикатор, а, скажем, целый ряд оценок по тестам, которые все выступают индикаторами одной и той же скрытой переменной способностей.

Важно обратить внимание на то, что прокси-переменная не является инструментом (и наоборот). Таким образом, в рамках метода инструментальных переменных мы не получаем оценки параметров с помощью регрессии y на Z . Рассмотрим вырожденный пример, предположив, что полная модель имеет вид

$$y = X^* \beta + \varepsilon,$$

$$X = X^* + U,$$

$$Z = X^* + W.$$

Таким образом, мы обладаем двумя плохо измеренными оценками X^* . Параметры этой модели легко оцениваются, если W не коррелирует с U и X^* , однако не с помощью регрессии y на Z . Для этого нам потребуется метод инструментальных переменных.

Если в модели есть переменная вроде образования или способностей, возникает естественный вопрос: нельзя ли, если мы заинтересованы в других коэффициентах модели, просто выбросить проблемную переменную из уравнения⁷? Такой способ приводит к знакомой нам проблеме пропущенной переменной, все равно приводящей к несостоятельности OLS-оценок. Так какая из этих двух оценок хуже? В работах Маккалума (McCallum (1972)) и Уикенса (Wickens (1972)) было показано, что асимптотическое смещение (по сути, степень несостоятельности) больше, если прокси-переменная не включается в уравнение даже в том случае, если она плохая (во многом определяется ошибкой измерения). Это утверждение, однако, не учитывает точности оценок. В работе Эйгнера (Aigner (1974)) исследовался этот момент и было показано, как, впрочем, и ожидалось, что такой вывод верен и для точности оценок. Автор заключает, что «это является свидетельством в пользу широкого использования прокси».

Пример 8.9. Доход и образование близнецов

Обычная модель, используемая в экономике труда для изучения влияния образования (education) на доход (income), описывается следующим уравнением:

$$y_i = \beta_1 + \beta_2 \text{age}_i + \beta_3 \text{age}_i^2 + \beta_4 \text{education}_i + \mathbf{x}'_i \beta_5 + \varepsilon_i,$$

где y_i обычно зарплата или годовой доход (возможно, в логарифмах), а \mathbf{x}_i содержит ряд переменных, таких как пол, регион проживания, отрасль, в которой индивид работает. В литературе встречается обсуждение различных проблем, которые возникают при оценке такого уравнения методом наименьших квадратов, с использованием собранных данных. Здесь мы рассмотрим две такие проблемы.

1. Несмотря на то что в уравнение включена переменная «образование», исследователь обычно располагает лишь информацией о числе лет обучения. Эта переменная является прокси-переменной для образования, поэтому в уравнении этого вида возможно наличие ошибки измерения. К тому же зачастую оказывается, что собранные данные по числу лет учебы сами по себе содержат ошибки, что является вторым источником ошибки измерения. Сейчас мы не будем рассматривать первую (более сложную) из этих двух проблем.

2. Другие переменные, такие как «способности» (обозначенные нами как μ_i), также влияют на доход, а кроме того, очевидно коррелированы с образованием. Если уравнение заработной платы оценивается в виде, указанном выше, то оценки будут смещены ввиду отсутствия «пропущенной переменной». По причинам, которые мы приведем в главе 19, подобное смещение носит, согласно недавним работам, название **эффект отбора** (selectivity effect).

⁷ В равной степени это относится к проблемам ошибки измерения и скрытых переменных.

⁸ Другие исследования по близнецам и родным братьям и сестрам включают работы (Bound, Chorkas, Haskel, Hawkes, Spector (2003)), (Ashenfelter, Rouse (1998)), (Ashenfelter, Zimmerman (1997)), (Behrman, Rosengweig (1999)), (Isacsson (1999)), (Miller, Mulvey, Martin (1995)), (Rouse (1999)) и (Taubman (1976)).

При исследовании по обычным межобъектным данным указанные сложности будут существенно мешать. Но в работе (Ashenfelter, Kreuger (1994)), посвященной исследованию близнецов, авторы использовали базу данных, позволяющую при нескольких простых предположениях обойти эти проблемы⁸. Каждый год во многих городах США проводятся «фестивали близнецов». Наиболее крупный проводится в городе Твинсберг (Twinsburg), штат Огайо (Ohio). Авторы работы проинтервьюировали около 500 человек старше 18 лет во время фестиваля в августе 1991 г. Имея в качестве объекта наблюдений пары близнецов, они смогли привести модель к следующему виду. Пусть (y_{ij}, A_{ij}) обозначают заработную плату и возраст соответственно для близнеца j , $j = 1, 2$, из пары i . В качестве переменной образования доступны лишь указанные респондентами «годы обучения» S_{ij} . Авторы подошли к проблеме ошибки измерения в этой переменной следующим образом. Они спрашивали каждого близнеца о годах его учебы и о годах учебы его брата/сестры. Будем обозначать данные, сообщенные близнецом m о близнеце j , как $S_{ij}(m)$. В этих обозначениях сообщенные данные о себе близнеца 1 — это $S_{i1}(1)$. Близнец 2 отвечает на этот же вопрос (о близнеце 1) как $S_{i1}(2)$. Модель ошибки измерения для переменной числа лет обучения представляет собой

$$S_{ij}(m) = S_{ij} + u_{ij}(m), \quad i, m = 1, 2,$$

где S_{ij} = «верная информация» о числе лет учебы близнеца j из пары i .

Мы предположим, что оба источника ошибки измерения $u_{ij}(m)$ не коррелированы между собой и что они, как и S_{ij} , имеют нулевые средние. Теперь рассмотрим простую двумерную модель, похожую на представленную в (8-14): $y_{ij} = \beta S_{ij} + \varepsilon_{ij}$. Как мы уже видели выше, оценка наименьших квадратов коэффициента β при использовании имеющихся данных будет недооценена:

$$\text{plim } \hat{\beta} = \frac{\beta \times \text{Var}[S_{ij}]}{\text{Var}[S_{ij}] + \text{Var}[u_{ij}(j)]} = \beta q.$$

(Поскольку нет естественной разницы между обозначениями близнеца 1 и близнеца 2, предположение о том, что дисперсии двух ошибок измерения одинаковы, является вполне безобидным.) Множитель q иногда называют отношением достоверности (reliability ratio). В этой простой модели мы могли бы состоятельно оценить β , если бы это отношение достоверности было нам известно. Оказывается, что модель устроена таким образом, что мы почти можем это сделать. Поскольку две ошибки измерения не коррелируют между собой, то

$$\begin{aligned} \text{Corr}[S_{i1}(1), S_{i1}(2)] &= \text{Corr}[S_{i2}(1), S_{i2}(2)] = \\ &= \frac{\text{Var}[S_{i1}]}{(\text{Var}[S_{i1}] + \text{Var}[u_{i1}(1)]) \times \{\text{Var}[S_{i1}] + \text{Var}[u_{i1}(2)]\}^{1/2}} = q, \end{aligned}$$

т.е. корреляция между двумя сообщенными столбцами данных по образованию измеряет отношение достоверности. Авторы получили значения 0,920 и 0,877 для 298 пар однояйцевых близнецов и 0,869 и 0,951 для 92 пар разнаяйцевых близнецов, что дает быструю оценку размера ошибки измерения в имеющихся данных. Уравнение заработной платы представляет собой множественную регрессию, поэтому этот результат хоть и полезен для проблемы в общем, но тем не менее не позволяет определить все смещения коэффициентов. Для этой задачи был использован метод инструментальных переменных. Уравнение для $y_{ij} = \ln Wage_{ij}$, оцененное при помощи методов наименьших квадратов (LS) и инструментальных переменных (IV), выглядит следующим образом:

$$\beta_6 \text{sex}_i y_{ij} = \beta_1 + \beta_2 \text{age}_i + \beta_3 \text{age}_i^2 + \beta_4 S_{ij}(j) + \beta_5 S_{im}(m) + \beta_6 \text{sex}_i + \beta_7 \text{race}_i + \varepsilon_{ij}$$

| | | | | | | |
|----|---------|----------|----------|---------|----------|-----------|
| LS | (0,088) | (-0,087) | (0,084) | (0,204) | (-0,410) | |
| IV | (0,206) | (0,088) | (-0,087) | (0,116) | (-0,037) | (-0,428). |

В этом уравнении $S_{ij}(j)$ представляет данные, сообщенные индивидом о числе лет своего обучения, а $S_{im}(m)$ — данные о числе лет обучения брата. Эта переменная (schooling) является проблемной. Для получения состоятельной оценки использовался метод инструментальных переменных, причем в качестве инструмента выступали чужие данные об образовании (данные близнеца). Оценки, полученные авторами, приведены под уравнением. (Оценка константы не приводится, и по некоторым причинам вторая переменная по годам обучения не была включена в оцениваемое уравнение в случае LS.) Эти данные приведены для того, чтобы сравнить их с другими известными исследованиями. Влияние возраста (age), числа лет обучения (schooling) и пола (sex) вполне согласуется с данными, полученными в других исследованиях, в то время как влияние цвета кожи (race) чрезвычайно отличается: -40% здесь против обычных +9% в других исследованиях. Особенный интерес представляет эффект влияния IV-метода на оценку коэффициента β_4 . Напомним, что отношение достоверности было оценено примерно как 0,9, откуда можно заключить, что IV-оценка будет, грубо говоря, на 11% выше (1/0,9). Поскольку мы имеем дело с множественной регрессией, эта оценка является очень грубым приближением. На самом деле, как указывалось ранее, она выше примерно на 38%. Авторы использовали также другой подход к оценке. Напомним, что источником смещения, вызванного отбором наблюдений, являются эффекты, которые не могут быть измерены. Авторы переписали модель в виде

$$y_{ij} = \beta_1 + \beta_2 \text{age}_i + \beta_3 \text{age}_i^2 + \beta_4 S_{ij}(j) + \beta_6 \text{sex}_i + \beta_7 \text{race}_i + \mu_i + \varepsilon_{ij}.$$

Латентные, не поддающиеся измерению эффекты, такие как «», содержатся в μ_i . Поскольку μ_i ненаблюдаема, но по предположению коррелирована с другими переменными модели, оценки регрессии методом наименьших квадратов y_{ij} на остальные переменные будут смещенными. [Это есть

модель с фиксированным эффектом — см. раздел 11.4. Предположение, что скрытый эффект «способностей» одинаков для близнецов и может быть полностью объяснен, является спорным предположением о том, что способности объясняются наследственностью, а не образованием. См., например, Behrman, Taubman (1989). Поиск в Интернете словосочетания «дебаты о том, что первично — наследственность или образование» («nature versus nurture debate») даст миллионы ссылок. Мы не будем останавливаться здесь на этом моменте.] Разность между двумя уравнениями заработной платы равна

$$y_{i1} - y_{i2} = \beta_4[S_{i1}(1) - S_{i2}(2)] + \varepsilon_{i1} - \varepsilon_{i2}.$$

В этом уравнении нет латентных эффектов, но, оказывается, ухудшается ситуация с проблемой ошибки измерения. Как и ранее, β_4 может быть оценен с помощью метода инструментальных переменных. Есть две доступные инструментальные переменные: $S_{i2}(1)$ и $S_{i1}(2)$. (Из статьи не ясно, используют ли авторы их отдельно или используют их разность.) Оценка методом наименьших квадратов равна 0,092, что сравнимо с оценкой, полученной ранее. Оценка инструментальных переменных равна 0,167, что выше примерно на 82%. Приведенные стандартные ошибки равны 0,024 и 0,043 соответственно. Учитывая все это, можно провести тест Хаусмана:

$$H = \frac{(0,167 - 0,092)^2}{0,043^2 - 0,024^2} = 4,418.$$

95%-е критическое значение для хи-квадрат-распределения с одной степенью свободы равно 3,84, поэтому гипотеза о том, что LS-оценка является состоятельной, отвергается. (Квадратный корень из H , равный 2,102, должен сравниваться со значением из стандартного нормального распределения, для которого критическое значение равно 1,96. Автор также приводит t -статистику для этой регрессии, равную 1,97. Причины этого несовпадения неясны.)

8.6. Оценка нелинейным методом инструментальных переменных

В разделе 8.2 мы расширили модель линейной регрессии, чтобы иметь возможность рассматривать случай регрессоров, которые коррелированы с шоками. Такая же проблема может возникнуть и в нелинейных моделях. Функция потребления, оцененная в подразделе 7.2.5, почти наверняка представляет собой подобный случай, и мы переоценили ее заново с помощью метода инструментальных переменных для линейных моделей в примере 8.7. В этом разделе мы расширим метод инструментальных переменных на случай нелинейных регрессионных моделей.

В нелинейной модели

$$y_i = h(x_i, \beta) + \varepsilon_i$$

ковариаты x_i могут быть коррелированы с шоками. Мы можем ожидать, что этот эффект проявится на псевдорегрессорах $x_i^0 = \partial h(x_i, \beta) / \partial \beta$. В этом случае все полученные нами результаты о линеаризированной регрессии более не являются справедливыми. Предположим, что есть набор переменных $[z_1, \dots, z_L]$ такой, что

$$\text{plim}(1/n)\mathbf{Z}'\varepsilon = 0 \quad (8-23)$$

и

$$\text{plim}(1/n)\mathbf{Z}'\mathbf{X}^0 = Qz x^0 \neq 0,$$

где \mathbf{X}^0 есть матрица псевдорегрессоров в линеаризированной регрессии, вычисленная при истинных значениях параметров. Если подход, который мы применяли к линейной модели в разделе 8.3, может быть применен к этому набору переменных, то мы можем построить состоятельную оценку параметра β с помощью метода инструментальных переменных. Итак, сначала мы попробуем воспроизвести тот же подход, который был нами использован для линейной модели. Линеаризированная модель регрессии задается (см. (7-30)) как

$$\mathbf{y} = \mathbf{h}(\mathbf{X}, \beta) + \varepsilon \approx \mathbf{h}^0 + \mathbf{X}^0(\beta - \beta^0) + \varepsilon$$

или

$$\mathbf{y}^0 \approx \mathbf{X}^0\beta + \varepsilon,$$

где $\mathbf{y}^0 = \mathbf{y} - \mathbf{h}^0 + \mathbf{X}^0\beta^0$.

На данном этапе мы опустим ошибку приближения, возникающую при линеаризации модели. В (8-23) мы предполагали, что

$$\text{plim}(1/n)\mathbf{Z}'\mathbf{y}^0 = \text{plim}(1/n)\mathbf{Z}'\mathbf{X}^0\beta. \quad (8-24)$$

Предположим, как мы уже делали, что число инструментальных переменных совпадает с числом параметров, т.е. с числом столбцов матрицы \mathbf{X}^0 . (Заметим, что это число не обязано совпадать с числом регрессоров.) Тогда «оценка», предложенная ранее, должна вычисляться как

$$\mathbf{b}_{IV} = (\mathbf{Z}'\mathbf{X}^0)^{-1}\mathbf{Z}'\mathbf{y}^0. \quad (8-25)$$

Логика построения этой оценки ясна, однако есть некоторая проблема. Незвестный вектор параметров β присутствует в обеих частях формулы (8-24). Мы могли бы предложить уже использованный нами в первый раз подход к решению нелинейной модели. Взяв некоторую начальную оценку, будем оценивать регрессию методом инструментальных переменных итеративно, пересчитывая псевдорегрессоры, пока процесс не сойдется к определенной точке, которая нам и нужна. Идея опять-таки ясна, и в принципе этот метод даст искомую оценку.

Если к предыдущим условиям мы добавим условие

$$\frac{1}{\sqrt{n}} \mathbf{Z}'\varepsilon \xrightarrow{d} N[0, \sigma^2 Q_{zz}],$$

то сможем использовать тот же вид асимптотического распределения, как и ранее в линейном случае. Перед этим, однако, требуется заполнить некоторые пробелы в предшествующем изложении. Во-первых, несмотря на интуитивную привлекательность, предложенная процедура нахождения оценки вряд ли будет хорошим алгоритмом. Во-вторых, нам бы не хотелось ограничиваться случаем, когда число инструментов совпадает с числом параметров. Поэтому мы рассмотрим задачу с более широкой точки зрения. Критерий оценки для нелинейного метода инструментальных переменных представляет собой квадратичную форму⁹

$$\begin{aligned} \text{Min}_{\beta} S(\beta) &= \frac{1}{2} \{ [\mathbf{y} - \mathbf{h}(\mathbf{X}, \beta)]' \mathbf{Z} \} (\mathbf{Z}'\mathbf{Z})^{-1} \{ \mathbf{Z}' [\mathbf{y} - \mathbf{h}(\mathbf{X}, \beta)] \} \quad (8-26) \\ &= \frac{1}{2} \varepsilon(\beta)' \mathbf{Z} (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}' \varepsilon(\beta). \end{aligned}$$

Условия первого порядка для минимизации этой взвешенной суммы квадратов представляют собой

$$\frac{\partial S(\beta)}{\partial \beta} = -\mathbf{X}^{0'} \mathbf{Z} (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}' \varepsilon(\beta) = 0. \quad (8-27)$$

Этот результат совершенно тот же, что и для случая линейной модели с заменой \mathbf{X} на \mathbf{X}^0 . Эта задача, однако, сильно нелинейна в большинстве случаев, и повторяющийся подход наименьших квадратов, по всей видимости, будет неэффективен. Но с точки зрения задач минимизации, рассматриваемых в приложении Е, это по сути простая минимизационная задача, и мы можем подходить к оцениванию как к задаче нелинейной оптимизации.

Мы подошли к оценке инструментальных переменных из более-менее общих соображений. Тем не менее возможен более структурированный подход. Условие ортогональности (orthogonality condition) $\text{plim}(1/n) \mathbf{Z}'\varepsilon = 0$ определяет GMM-оценку. С учетом предположений о гомоскедастичности и об отсутствии автокорреляции полученная **оценка минимального расстояния** (minimum distance estimator) приводит к предложенному выше критерию. Мы вернемся к этой оценке в главе 13.

При доброкачественных псевдорегрессорах и инструментальных переменных мы можем сформулировать общий результат оценки нелинейным методом инструментальных переменных. Более подробно об этом результате можно прочесть в книге Дэвидсона и Маккиннона (Davidson, MacKinnon (2004)).

⁹Пожалуй, наиболее естественной точкой для начала процедуры минимизации будет $S^0(\beta) = [\varepsilon(\beta)' \mathbf{Z}][\mathbf{Z}' \varepsilon(\beta)]$. Мы не останавливаемся подробно на этом моменте, поскольку из критерия (8-26) оценка, задаваемая условиями (8-27), оказывается (как будет следовать из дальнейших рассуждений и результатов главы 13) простой, но более эффективной GMM-оценкой.

Теорема 8.2. Асимптотическое распределение оценки нелинейным методом инструментальных переменных

При доброкачественных инструментальных переменных и псевдорегрессорах

$$b_{IV} \overset{a}{\sim} N[\beta, (\sigma^2/n)(Q_{xz}^0(Q_{zz})^{-1}Q_{zx}^0)^{-1}].$$

Оценка ковариационной матрицы имеет вид

$$\text{Est. Asy. Var}[b_{IV}] = \hat{\sigma}^2[\hat{X}^0'Z(Z'Z)^{-1}Z'\hat{X}^0]^{-1},$$

где \hat{X}^0 есть матрица X^0 , вычисленная с помощью b_{IV} .

Наконец, отметим, что интерпретация двухшагового метода наименьших квадратов для линейной модели применима и в этом случае по отношению к IV-оценке. Таким образом, в конечных оценках из условий первого порядка (нормальных уравнений) следует, что

$$X^0'Z(Z'Z)^{-1}Z'y = X^0'Z(Z'Z)^{-1}Z'X^0\beta,$$

откуда получается, что оценки удовлетворяют нормальным уравнениям линейной регрессии y (не y^0) на предсказанные значения, полученные при оценке регрессий столбцов X^0 на Z . Интерпретация все-таки немного отличается, поскольку для того, чтобы вычислить предсказанные значения X^0 , необходимо уже иметь оценку β . Таким образом, этот двухшаговый метод не помогает вычислять оценку b_{IV} , а просто характеризует ее.

Пример 8.10. Оценка функции потребления методом инструментальных переменных

В подразделе 7.2.5 функция потребления оценивалась нелинейным методом наименьших квадратов, при этом не учитывалось то, что наверняка имеется корреляция между X^0 и ϵ . Как мы уже делали, мы переоценим эту модель, используя подход инструментальных переменных. В этом примере будем использовать лагированное (на один период) значение функции потребления и лагированные (на один и на два периода) значения дохода в качестве инструментов. В табл. 8.2 приведены результаты оценок нелинейным методом наименьших квадратов и методом инструментальных переменных. Поскольку мы используем запаздывающее на два периода значение переменной, то теряем два наблюдения. Таким образом, оценки методом наименьших квадратов не совпадают с вычисленными ранее.

Оценки методом инструментальных переменных значительно отличаются от оценок методом наименьших квадратов. Однако это различие может вводить в заблуждение. Напомним, что MPC в модели равна $\beta\gamma Y^{\gamma-1}$. Значение в IV квартале 2000 г. для DPI , как мы уже видели, равно 6634,9. При этом значении оценки MPC метода инструментальных переменных и метода наименьших квадратов равны 1,1543 с оцененной

стандартной ошибкой, равной 0,01234, и 1,08406 с оцененной стандартной ошибкой, равной 0,008694, соответственно. Эти значения также немного разнятся, но уже значительно меньше, чем можно было бы предположить исходя из различий в оценках параметров. Отметим, что IV-оценка значительно больше оценки в линейной модели, равной 0,9217 (а также больше единицы, что может показаться несколько странным).

Таблица 8.2. Оценки нелинейным методом наименьших квадратов и нелинейным методом инструментальных переменных

| Параметр | Метод инструментальных переменных | | Метод наименьших квадратов | |
|----------|-----------------------------------|---------------|----------------------------|---------------|
| | Оценка | Станд. ошибка | Оценка | Станд. ошибка |
| α | 627,031 | 26,6063 | 468,215 | 22,788 |
| β | 0,040291 | 0,006050 | 0,0971598 | 0,01064 |
| γ | 1,34738 | 0,016816 | 1,24892 | 0,1220 |
| σ | 57,1681 | — | 49,87998 | — |
| $e'e$ | 650 369,805 | — | 495 114,490 | — |

8.7. Слабые инструменты

До сих пор наше исследование концентрировалось около условия «» для IV-оценки, т.е. условия «экзогенности» A.I9, которое гласит, что

$$\text{plim}(1/n)\mathbf{Z}'\varepsilon = 0. \quad (8-28)$$

Вместе с условием «релевантности»

$$\text{plim}(1/n)\mathbf{Z}'\mathbf{X} = \mathbf{Q}_{\mathbf{Z}\mathbf{X}} - \text{конечная, ненулевая матрица размером } L \times K, \text{ ранга } K, \quad (8-29)$$

что позволяет получить состоятельную IV-оценку. Точнее говоря, при выполнении условия (8-28) условие (8-29) является достаточным для свойства несмещенности. Учитывая это, исследователи сосредоточились на экзогенности как на определяющей проблеме для построения IV-оценки. Однако увеличивающееся количество работ подтверждает, что большее внимание следует уделять условию релевантности. Строго говоря, (8-29) и в самом деле достаточно для вывода рассмотренных нами асимптотических результатов, но часто возникает случай слабых инструментов, при котором условие (8-29) едва выполняется, это стало в последнее время предметом тщательного изучения. С прикладной точки зрения инструменты являются слабыми, если они лишь слегка коррелируют с переменными в правой части уравнения \mathbf{X} . Таким образом, величина $(1/n)\mathbf{Z}'\mathbf{X}$ близка к нулю. (Мы определим

этот момент теоретически, когда вернемся к нему в подразделе 10.6.6.) Исследователи начали внимательно изучать такие случаи, и иногда это приводило к объяснениям некоторых эмпирических результатов, не согласующихся с теорией¹⁰.

На первый взгляд проблема слабых инструментов заключается в асимптотической матрице ковариации IV-оценки:

$$\text{Asy. Var}[b_{IV}] = \frac{\sigma_{\varepsilon}^2}{n} \left[\left(\frac{X'Z}{n} \right) \left(\frac{Z'Z}{n} \right)^{-1} \left(\frac{Z'X}{n} \right) \right]^{-1},$$

которая будет большой при слабых инструментах и при прочих равных тем больше, чем слабее инструменты. Однако проблема лежит глубже. В работах Нельсона и Старца (Nelson, Startz (1990a,b)) и Хана и Хаусмана (Hahn, Hausman (2003)) перечислены два следствия слабых инструментов. Первое заключается в том, что оценка двухшаговым методом наименьших квадратов неверно смещена к оценке обычным методом наименьших квадратов, которая, как известно, является несостоятельной. Второе состоит в том, что стандартная асимптотика первого порядка (подобная той, которую мы использовали ранее) не предоставляет корректной основы для статистических выводов. Таким образом, ситуация значительно хуже, чем просто недостаток точности оценки. Кроме того, есть некоторые соображения, что этот аспект сильно проявляется и при рассмотрении «проблем малых выборок». [См. Bound, Jaeger, Baker (1995).]

Современный подход предлагает несколько рецептов для обнаружения слабости инструментов. В случае единственной эндогенной переменной x (коррелированной с ε) стандартный подход состоит в том, чтобы сделать первый шаг двухшаговым методом наименьших квадратов. Далее, используя обычную F -статистику, проверяющую гипотезу о том, что все коэффициенты регрессии

$$x_i = z_i' \pi + v_i$$

равны нулю, протестировать гипотезы о слабых инструментах. Значение F -статистики, меньшее 10, сигнализирует о проблеме. [См. работы Nelson, Startz (1990b), Staiger, Stock (1997), Stock, Watson (2007, глава 12) для объяснения этого специфического теста.] При наличии более одной эндогенной переменной в модели тестирования их по отдельности оказывается недостаточно. Слабость инструментов может быть вызвана коллинеарностью между переменными, и ни один из наших тестов этого не диагностирует. Для этих целей в работе Ши (Shea (1997)) была предложена многовариантная четырехшаговая процедура. В работе Годфри (Godfrey (1999)) был предложен на удивление простой с вычислительной точки зрения метод. Для эндогенной переменной k статистика Годфри представляет собой отношение оцененных дисперсий двух оценок, OLS и 2SLS:

$$R_k^2 = \frac{v_k(\text{OLS})/e'e(\text{OLS})}{v_k(\text{2SLS})/e'e(\text{SLS})},$$

¹⁰К важным ссылкам по этой теме относятся работы (Nelson, Startz (1990a,b)), (Staiger, Stock (1997)), (Stock, Wright, Yogo (2002)), (Hahn, Hausman (2002, 2003)), (Kleibergen (2002)), (Stock, Yogo (2005)) и (Hausman, Stock, Yogo (2005)).

где $v_k(\text{OLS})$ есть k -й диагональный элемент матрицы $[\mathbf{e}'\mathbf{e}(\text{OLS})/(n - K)] \times (\mathbf{X}'\mathbf{X})^{-1}$, аналогично определяется и $v_k(2\text{SLS})$. После масштабирования статистика принимает вид

$$R_k^2 = \frac{(\mathbf{X}'\mathbf{X})^{kk}}{(\hat{\mathbf{X}}'\hat{\mathbf{X}})^{kk}},$$

где верхний индекс обозначает элемент обратной матрицы. На основе этой статистики можно построить F -статистику в предположении, что \mathbf{Z} содержит константу: $F = [R_k^2/(L - 1)]/[(1 - R_k^2)/(n - L)]$.

Важно отметить, что тест на слабые инструменты не есть тест на спецификацию модели или тест, необходимый для построения модели. Он скорее представляет собой стратегию, помогающую исследователю избежать ситуации, в которой он будет полагаться при своих статистических выводах на ненадежные статистики, чьи свойства очень плохо описываются знакомыми асимптотическими результатами, такими как их распределения (при выполнении условий модели). Также существует несколько интересных обобщений. Некоторые статистические процедуры были предложены в работах (Hahn, Hausman (2002)) и (Kleibergen (2002)). Также нас интересуют ситуации, когда существует более одной эндогенной переменной. Мы вернемся к этому вопросу немного под другим углом зрения в подразделе 10.6.6, где будем рассматривать его в рамках работы с моделью одновременных уравнений.

Суровые выводы этого раздела ставят под сомнение IV-оценку. При некоторых ограниченных условиях альтернативой может быть «moment»-free LIML-оценка, обсуждаемая в подразделе 10.6.4. Несколько неожиданным подходом является возврат к методу наименьших квадратов. Оценка OLS не лишена некоторых достоинств. Ее асимптотическая дисперсия

$$\text{Asy. Var}[\mathbf{b}_{LS}] = (\sigma^2/n)\mathbf{Q}_{\mathbf{X}\mathbf{X}}^{-1}$$

однозначно меньше асимптотической дисперсии IV-оценки

$$\text{Asy. Var}[\mathbf{b}_{IV}] = (\sigma^2/n)(\mathbf{Q}_{\mathbf{X}\mathbf{Z}}\mathbf{Q}_{\mathbf{Z}\mathbf{Z}}^{-1}\mathbf{Q}_{\mathbf{Z}\mathbf{X}})^{-1}.$$

(Доказательство этого факта рассматривается в качестве упражнения в конце главы.) Имея в виду предыдущие выводы, она может быть много меньше. Однако OLS-оценка является несостоятельной:

$$\text{plim } \mathbf{b}_{LS} - \beta = \mathbf{Q}_{\mathbf{X}\mathbf{X}}^{-1}\gamma$$

[см. (8-4)]. По критерию среднеквадратичной ошибки совершенно неясно, какая из оценок является более точной: OLS, для которой

$$M(\mathbf{b}_{LS}|\beta) = (\sigma^2/n)\mathbf{Q}_{\mathbf{X}\mathbf{X}}^{-1} + \mathbf{Q}_{\mathbf{X}\mathbf{X}}^{-1}\gamma\gamma'\mathbf{Q}_{\mathbf{X}\mathbf{X}}^{-1},$$

или IV, для которой

$$M(\mathbf{b}_{IV}|\beta) = (\sigma^2/n)(\mathbf{Q}_{\mathbf{X}\mathbf{Z}}\mathbf{Q}_{\mathbf{Z}\mathbf{Z}}^{-1}\mathbf{Q}_{\mathbf{Z}\mathbf{X}})^{-1}.$$

Естественным выходом при слабых инструментах является исключение эндогенной переменной из модели или улучшение инструментов. И то, и другое относится к вопросам самой спецификации модели. Говоря только в категориях оценки модели, имея необходимые данные и спецификацию модели, вполне можно утверждать, что лучшим вариантом будет OLS.

8.8. Естественные эксперименты и поиск причинно-следственных связей

Эконометрику и статистику исторически преподавали, понимали и изучали с глубоким убеждением, что корреляция не есть причинно-следственная связь. Однако большое и все увеличивающееся количество работ по микроэконометрике и по темам, близким к рассмотренным в этой главе, может быть названо моделированием причинности¹¹. В современной литературе по эффектам влияния и оценке программных документов эконометрической задачей является нечто большее, чем статистические связи. Короче говоря, вопрос «*работает ли программа?*» требует более определенного эконометрического ответа, чем ответ «данные, по всей видимости, согласуются с этой гипотезой». Тем не менее аккуратный подход к эконометрическому моделированию «причинности» должен в существенной степени основываться на статистической базе¹².

Примером причинной модели, рассмотренной здесь, является уравнение, аналогичное модели Крюгер и Дэйла (Krueger, Dale (1999)) и описывающее влияние обучения в элитном колледже на последующие заработки:

$$\ln Earnings = \mathbf{x}'\boldsymbol{\beta} + \delta T + \varepsilon,$$

где δ обозначает причинное влияние факта посещения элитного колледжа. В этой модели T не может изменяться автономно, вне модели. Разброс T частично определяется какими-то скрытыми причинами, которые определяют и заработки в течение жизни. Несмотря на то что причинное влияние может быть приписано T , степень этого влияния δ не может быть определена в рамках множественной линейной регрессии. Линейный метод инструментальных переменных был введен как некое приспособление, позволяющее распутать причинные связи. Как и метод наименьших квадратов, метод инструментальных переменных должен быть защищен от риска, что лежащие в его основе статистические связи могут быть вызваны чем-то другим. Но если инструмент является результатом некоторого естественного эксперимента, то мы можем быть уверены в справедливости его экзогенности. Этот результат лежит в основе убеждений наиболее яростных сторонников такого направления исследования. Плюсы этого подхода лежат в, так сказать, неизбежности и стабильности полученных данных. Все это порождает го-

¹¹ См., например, главу 2 Cameron, Trivedi (2005), которая называется «Causal and Noncausal Models», и особенно работы Angrist, Imbens, Robin (1996), Angrist, Krueger (2001) и Angrist, Pischke (2009, 2010).

¹² Среди прочих работ по этой теме см. Heckman, Vytlacil (2007).

товность современных исследователей забывать об осторожности¹³. В примере 8.11 представлена другая точка зрения. Авторы на основе естественного эксперимента идентифицируют причинно-следственную связь, которая может быть рассмотрена вне области регрессионного анализа, в рамках более ранних концепций¹⁴.

Пример 8.11. Вызывает ли телевидение аутизм?

Текст ниже представляет собой аннотацию из работы (Waldman, Nicholson, Adilov (2008)), посвященной изучению аутизма¹⁵.

По современным оценкам, аутизму подвержен примерно один из 166 детей, хотя причины этой болезни до сих пор непонятны. Одна из современных теорий утверждает, что среди детей, предрасположенных к аутизму из-за наследственных особенностей, аутизм развивается, если ребенок подвергается некоторому (в настоящее время неизвестному) воздействию извне. В этой работе мы эмпирически проверяем гипотезу о том, что подобным спусковым крючком является просмотр телевизора в детстве. Опираясь на исследование Бюро статистики занятости (Bureau of Labor Statistics's – American Time Use Survey), мы сначала покажем, что количество времени, проведенного за просмотром телевизора в детстве, положительно связано с количеством осадков, выпадавших в это время в той местности, где жили дети. Отсюда можно предположить, что если телевидение является спусковым крючком для аутизма, то аутизм должен превалировать в округах, в которых выпадает больше осадков. Затем мы рассматриваем данные по аутизму по округам в трех штатах: Калифорнии (California), Орегоне (Oregon) и Вашингтоне (Washington), характеризующихся большими различиями в количестве выпадающих осадков. С помощью набора различных тестов мы покажем, что в каждом из трех штатов (и по трем штатам, рассматриваемым в совокупности) уровень аутизма в округе положительно связан с количеством выпадающих осадков. В конце исследования мы используем данные по детям из Калифорнии и Пенсильвании, родившимся между 1972 и 1989 гг., чтобы показать состоятельность гипотезы о просмотре телевизора в том смысле, что уровень аутизма в округе положительно связан с процентом домохозяйств, подключенных к кабельному телевидению. *Наши тесты о влиянии осадков на аутизм показывают, что только 40% диагностированного аутизма в трех штатах, которые мы изучили, являются следствием просмотра телевизора из-за выпадающих осадков, в то время как тесты о влиянии распространения кабельного телевидения показали, что около 17% роста уровня аутизма в Калифорнии и Пенсильвании в 1970–*

¹³См., например, Angrist, Pischke (2009, 2010). Keane (2010, p. 48), напротив, пишет: «Что всегда удивляло меня в научной школе “эксперименталистов”, так это их ложное чувство убежденности. Основная их идея состоит в том, что если у нас есть “по-настоящему хороший инструмент”, то мы можем прийти к “убедительным” оценкам “причинных влияний”, которые “не слишком чувствительны к предположениям”».

¹⁴См. сборник статей, вышедший весной 2010 г. в *Journal of Economic Perspectives* (Angrist, Pischke (2010), Leamer (2010), Sims (2010), Keane (2010), Stock (2010), Nevo, Whinston (2010)).

¹⁵Цит. по: <http://www.johnson.cornell.edu/faculty.profiles/waldman/autism-waldman-nicholson-adilov.pdf>.

1980-х гг. являются следствием распространения кабельного телевидения. Эти выводы согласуются с мнением о том, что просмотр телевизора в раннем детстве является важным механизмом при аутизме. (Выделено курсивом.) Мы также обсудим другие тесты, которые могут быть использованы, чтобы исследовать гипотезу более целенаправленно.

Авторы добавляют (на с. 3): «Несмотря на состоятельность гипотезы о том, что просмотр телевизора в раннем детском возрасте является важным механизмом развития аутизма, наш основной вывод согласуется с другой возможностью. Так как осадки, вероятно, коррелированы с тем, что маленькие дети проводят больше времени дома в принципе, а не только с тем, что они смотрят телевизор больше, наш основной вывод состоит в том, что первопричиной может быть чрезмерное нахождение дома. Следовательно, мы используем и вторую инструментальную переменную (или натуральный эксперимент), которая коррелирована с переменной, характеризующей просмотр телевизора в раннем возрасте, но скорее всего не сильно коррелирована с количеством времени, которое ребенок проводил дома». (Выделено курсивом.) Они делают вывод (на с. 39–40): «Используя результаты оценки сквозной регрессии уровня аутизма внутри округов по трем штатам (Калифорния, Орегон и Вашингтон), представленные в табл. 3, мы сделали вывод, что если просмотр телевизора в раннем детстве является единственным механизмом, приводящим к положительной корреляции между аутизмом и осадками, то 38% диагностированных случаев аутизма есть следствие увеличения просмотра телевизора из-за осадков».

Исследование (Waldman, Nicholson, Adilov (2008))¹⁶ спровоцировало мощную реакцию среди ученых, изучающих аутизм, и общественности. Уайтхаус (Whitehouse (2007)) изучил некоторые вопросы, касающиеся методологических аспектов поиска «причинного влияния» в эконометрических исследованиях:

Готовность проф. Уолдмена ставить на карту свое мнение в столь деликатном деле, как наука, по мнению критиков, отражает растущие амбиции экономистов, а также их высокомерие. Академические экономисты все чаще рискуют выходить за рамки традиционных принципов, что приводит к некоторым серьезным результатам и вызывает обеспокоенность тем, не заходят ли они слишком далеко.

Такие дискуссии, скорее всего, будут множиться, так как экономисты все больше вникают в вопросы образования, политики, истории и даже эпидемиологии. Проф. Уолдмен использует осадки как иллюстрацию одного из инструментов, который стимулировал его: инструментальной переменной, статистического метода, который при помощи введения некоторого случайного или естественного влияния помогает экономисту разобраться в вопросах причины и следствия. Используя эту технику, они могут создать «естественные эксперименты», которые стремятся приблизиться к строгим случайным экспериментам — традиционному золотому стандарту медицинских исследований.

¹⁶Опубликовано в препринтах NBER № 12632 в 2006 г.

Инструментальные переменные помогают известным ученым пролить свет на деликатные вопросы. Джошуа Энгрист из Массачусетского технологического института изучал издержки войны, Стивен Левит из Чикагского университета — влияние увеличения штата полиции на преступность, Кэролайн Хоксби из Гарварда — школьную успеваемость. Их работы сыграли важную роль в общественно-политических дискуссиях. Но из-за возрастающего относительно данного подхода энтузиазма растет и число проблем. Одна из таких проблем заключается в следующем. Когда экономисты используют одну переменную как прокси для другой, например количество осадков вместо просмотра телевизора, не всегда ясно, что в конечном счете измеряет результат. Кроме того, эксперименты сами по себе не очень помогают в понимании того, почему одна вещь влияет на другую. «Существует поговорка, что неведение — это благо, — говорит Джеймс Хекман, профессор экономики Университета Чикаго, который получил Нобелевскую премию в 2000 г. за работу по статистическим методам. — «Я думаю, что это характеризует большую часть восторгов относительно инструментов». Джерри Хаусман, экономист Массачусетского технологического института, говорит: «Если ваши инструменты не являются идеальными, то у вас может быть серьезная проблема».

Пример 8.12. *Является ли сезон рождения надежным инструментом?*

Баклс и Хангемэн (Buckles, Hungerman (ВН, 2008)) перечисляют более чем 20 исследований долгосрочных экономических показателей, в которых сезон рождения используется как инструментальная переменная. Одним из самых первых и наиболее известных исследований по «естественным экспериментам» является работа Энгриста и Крюгера (Angrist, Krueger (1991)). Утверждение о допустимости использования сезона рождения как подходящего инструмента исходит из того, что характеристики семьи не связаны с сезоном рождения, но они очевидно связаны с долгосрочными характеристиками, такими как доход и образование. Утверждение объясняет использование дамми-переменных сезона рождения в качестве инструментальных в уравнениях. С одной стороны, если сезон рождения коррелирован с характеристиками семьи, то будет «терять силу исключяющее ограничение в большинстве IV-условий, в которых он используется» (ВН, page 2). Следуя логике авторов, случайность квартала рождения в генеральной совокупности [см, например, Kleibergen (2002)] можно считать данностью без необходимости проведения научного исследования. Используя данные свидетельств о рождении и данные переписи населения, ВН обнаружили скромное в числовом выражении, но статистически значимое отношение между датой рождения и характеристиками семьи. Они нашли, что «женщины, рождающие зимой, отличаются от других женщин; они моложе, хуже образованы и с меньшей вероятностью замужем... Доля детей, рожденных женщинами, не имеющими аттестата о полном среднем образовании, приблизительно на 10% выше (на 2 п.п.) в январе, чем в мае. Мы также обнаружили, что доля детей, рожденных тинейджерами, уменьшается на 10% с января по май». Тем не менее вопрос, почему

наблюдается такая связь, остается без ответа. Исследователи (конечно) имеют разные точки зрения на многочисленные следствия из выводов ВН. [См. Lahart (2009).] Но методологические следствия их выводов вполне согласуются с наблюдением Хаусмана.

8.9. Заключение и выводы

Метод инструментальных переменных (IV) в различных видах является одним из наиболее фундаментальных в эконометрике. В широком смысле он включает большинство методов оценивания, которые мы рассматривали раньше. В этой главе нами были получены основные результаты для IV-оценки линейной модели. Мы начали с принципиального момента построения оценки: предположений экзогенности и релевантности, которые и определяют оценку инструментальных переменных. Затем мы рассмотрели IV-оценку как оценку двухшаговым методом наименьших квадратов. За некоторыми исключениями, относящимися к моделям одновременных уравнений с двумя переменными, для IV-оценки не удастся установить никаких конечновыборочных свойств. Затем мы изучили асимптотические свойства IV-оценок в линейных и нелинейных регрессионных моделях. Наконец, в разделе 8.7 осветили некоторые предостерегающие моменты, касающиеся использования IV-оценок в ситуации, когда инструменты являются слаборелевантными.

Основные термины и понятия

Asymptotic covariance matrix — Асимптотическая ковариационная матрица

Asymptotic distribution — Асимптотическое распределение

Attenuation — Недооценка

Attenuation bias — Недооценка коэффициента

Attrition — Истощение

Attrition bias — Смещение из-за истощения выборки

Consistent estimator — Состоятельная оценка

Effect of the treatment on the treated — Эффект воздействия на подвергшихся воздействию

Endogenous — Эндогенный

Endogenous treatment effect — Эндогенный эффект воздействия

Exogenous — Экзогенный

Hausman statistic — Статистика Хаусмана

Identification — Идентификация

Indicator — Индикатор

Instrumental variables — Инструментальные переменные

Instrumental variable estimator — Оценка методом инструментальных переменных

Limiting distribution — Предельное распределение

Measurement error — Ошибка измерения

Minimum distance estimator — Оценка методом наименьшего расстояния

Moment equations — Моментные условия

Natural experiment — Естественный эксперимент

Nonrandom sampling — Неслучайный отбор

Omitted parameter heterogeneity — Гетерогенность, вызванная пропущенным параметром

Omitted variables — Пропущенные переменные

| | |
|--|---|
| Omitted variable bias — Смещение, вызванное пропущенной переменной | нениями |
| Orthogonality conditions — Условия ортогональности | Smearing — Размывание |
| Overidentification — Сверхидентификация | Specification test — Тест на спецификацию модели |
| Panel data — Панельные данные | Strongly exogenous — Строго экзогенный |
| Proxy variable — Прокси, прокси-переменная | Structural equation system — Система структурных уравнений |
| Random effects — Случайные эффекты | Structural Model — Структурная модель |
| Reduced form equation — Приведенная форма уравнения | Structural specification — Структурная спецификация |
| Relevance — Релевантность | Survivorship bias — Смещение выживших |
| Reliability ratio — Отношение достоверности | Truncation bias — Смещение из-за усе- чения выборки |
| Sample selection bias — Смещение из-за отбора наблюдений | Two-stage least squares (2SLS) — Двух- шаговый метод наименьших квадратов |
| Selectivity effect — Эффект отбора на- блюдений | Variable addition test — Тест на включение переменной |
| Simultaneous equations — Одновремен- ные уравнения | Weak instruments — Слабые инструмен- ты |
| Simultaneous equations bias — Смеще- ние, вызванное одновременными урав- | Weakly exogenous — Слабоэкзогенный |
| | Wu test — Тест Ву |

Упражнения

1. При обсуждении оценки методом инструментальных переменных мы показали, что оценка методом наименьших квадратов \mathbf{b}_{LS} смещена и является несостоятельной. Тем не менее \mathbf{b}_{LS} кое-что оценивает — см. (8-4). Выведите формулу для асимптотической ковариационной матрицы \mathbf{b}_{LS} и покажите, что оценка \mathbf{b}_{LS} асимптотически нормально распределена.
2. Для модели с ошибкой измерения, описанной уравнениями (8-14) и (8-15b), докажите, что если лишь переменная x измерена с ошибкой, то квадрат коэффициента корреляции между y и x меньше, чем между y^* и x^* . (Заметим, что по предположению $y^* = y$.) Верно ли это соотношение, если переменная y^* тоже измерена с ошибкой?
3. Выведите формулы, аналогичные (8-20a) и (8-20b), для модели с ошибкой измерения. Используйте подсказку в сноске 4 в подразделе 8.5.1, которая предлагает использовать (A-66) при обращении матрицы

$$[\mathbf{Q}^* + \Sigma_{uu}] = [\mathbf{Q}^* + (\sigma_u \mathbf{e}_1)(\sigma_u \mathbf{e}_1)'].$$

4. В конце раздела 8.7 выдвигается предположение, что OLS-оценка может иметь меньшую среднеквадратичную ошибку, чем 2SLS-оценка. С помощью (8-4), результатов упражнения 1 и теоремы 8.1 покажите, что это верно в случае, если

$$\mathbf{Q}_{xx} - \mathbf{Q}_{xz} \mathbf{Q}_{zz}^{-1} \mathbf{Q}_{zx} \gg \frac{1}{(\sigma^2/n) + \gamma' \mathbf{Q}_{xx}^{-1} \gamma} \gamma \gamma'.$$

Как можно проверить, что это, по крайней мере, возможно? Правая часть

неравенства является неотрицательно определенной матрицей ранга 1. Что можно сказать о левой части неравенства?

5. Рассмотрим линейную модель $y_i = \alpha + \beta x_i + \varepsilon_i$, в которой $\text{Cov}[x_i, \varepsilon_i] = \gamma \neq 0$. Пусть z — экзогенный релевантный инструмент для этой модели. Предположим также, что z является бинарной величиной: принимает лишь значения 1 и 0. Выведите алгебраические выражения для LS- и IV-оценок коэффициентов α и β .
6. При обсуждении оценки методом инструментальных переменных мы показали, что оценка методом наименьших квадратов вектора \mathbf{b} смещена и несостоятельна. Тем не менее \mathbf{b} оценивает $\text{plim } \mathbf{b} = \boldsymbol{\theta} = \boldsymbol{\beta} + \mathbf{Q}^{-1}\boldsymbol{\gamma}$. Выведите асимптотическую ковариационную матрицу \mathbf{b} и покажите, что вектор \mathbf{b} асимптотически нормально распределен.

Эмпирические приложения

1. В примере 8.5 мы рассматривали модель рынка труда. Из «приведенной формы» первого уравнения можно получить весь набор переменных, входящих в модель (эндогенных переменных): $\ln Wage_{it}$, Wks_{it} , и экзогенных переменных. Предполагается, что уравнение предложения труда содержит эти две переменные и еще три экзогенные переменные. Исходя из этого можно предположить, какие переменные будут присутствовать в уравнении «спроса» на труд для $\ln Wage_{it}$. Предположим (в нашем примере), что $\ln Wage_{it}$ определяется переменной Wks_{it} и подходящими экзогенными переменными. (Нужно отметить, что этот пример является исключительно иллюстрацией вычислений — он не дает теоретически обоснованную модель равновесия на рынке труда.)
 - a. Каково предполагаемое уравнение спроса на труд?
 - b. Оцените параметры этого уравнения с помощью методов OLS и 2SLS и сравните полученные результаты. (Проигнорируйте панельную структуру данных. Действуйте так, как будто это пул (рассматривайте данные в совокупности).)
 - c. Являются ли использованные в этом уравнении инструменты релевантными? Почему?

ГЛАВА 9

Обобщенная регрессионная модель и гетероскедастичность

9.1. Введение

В этой и нескольких следующих главах мы расширим модель множественной регрессии до формы, в которой нарушается предположение А.4 об ошибках модели классической регрессии. **Обобщенная линейная регрессионная модель** имеет вид

$$\begin{aligned}y &= \mathbf{X}\beta + \varepsilon, \\E[\varepsilon | \mathbf{X}] &= \mathbf{0}, \\E[\varepsilon\varepsilon' | \mathbf{X}] &= \sigma^2\Omega = \Sigma,\end{aligned}\tag{9-1}$$

где Ω — положительно определенная матрица. (Ковариационная матрица записывается в форме $\sigma^2\Omega$, так что мы можем получить классическую модель $\sigma^2\mathbf{I}$ как удобный частный случай.)

Гетероскедастичность и автокорреляция — два основных вопроса, которые будут детально рассмотрены. Случайные ошибки регрессии являются гетероскедастичными, если они имеют разную дисперсию. Гетероскедастичность имеет место в данных по волатильным высокочастотным временным рядам, таких как дневные наблюдения на финансовых рынках, и межобъектных данных, у которых масштаб зависимых переменных и объясняющая сила модели имеют тенденцию меняться в зависимости от наблюдений. Микроэкономические данные, такие как опросы о расходах населения, являются типичными представителями подобных данных. Поскольку, как и прежде, предполагается некоррелированность ошибок внутри выборки, $\sigma^2\Omega$ имеет вид

$$\sigma^2\Omega = \sigma^2 \begin{bmatrix} \omega_1 & 0 & \cdots & 0 \\ 0 & \omega_2 & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & \omega_n \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}.$$

(Первая подобная ситуация, включающая финансовые данные, будет детально рассмотрена в главе 20 и будет несколько более сложной, чем эта.)

Автокорреляция обычно наблюдается во временных рядах. Экономические временные ряды часто демонстрируют «память» в том смысле, что отклонения ряда от регрессионной функции в разные моменты становятся за-

висимыми между собой. Примерами могут быть ряды сезонно скорректированных цен и различные количественные показатели, которые публикуются правительственными службами. Данные, имеющие структуру временных рядов, обычно гомоскедастичны, так что $\sigma^2\Omega$ может быть представлена в виде

$$\sigma^2\Omega = \sigma^2 \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{n-1} \\ \rho_1 & 1 & \cdots & \rho_{n-2} \\ & & \vdots & \\ \rho_{n-1} & \rho_{n-2} & \cdots & 1 \end{bmatrix}.$$

Значения матрицы вне диагонали зависят от модели, использованной для ошибок. В большинстве случаев из-за затухающей памяти значения уменьшаются по мере удаления от диагонали.

Панельные данные, соответствующие различным объектам, наблюдаемым в течение нескольких моментов, могут быть и автокоррелированными, и гетероскедастичными. Мы рассмотрим их в главе 11. В данной главе содержатся некоторые общие результаты для расширенной модели. Модель гетероскедастичности будет рассмотрена в этой главе, а также в главе 14. Общая модель автокорреляции представлена в главе 20. Главы 10 и 11 содержат детальный анализ отдельных типов обобщенной регрессионной модели.

Результаты, полученные ранее для классической модели, должны быть изменены. Мы будем действовать согласно следующему подходу для общих результатов и в специфических случаях гетероскедастичности и серийной корреляции:

1. Сначала мы рассмотрим изменения в оценках МНК из-за более общей формы регрессионной модели, а также влияние игнорирования структуры обобщенной модели и предложим процедуру оценки, основанную на методе наименьших квадратов.
2. Мы проанализируем альтернативные подходы к оценке, которые могут продемонстрировать лучшее использование характеристик модели. При этом сделаем минимальные предположения о Ω .
3. Затем мы сузим предположения и начнем искать методы для устранения недостатков классической модели, т.е. сформулируем процедуры для тестирования гипотезы о спецификации в виде классической модели против альтернативы в виде обобщенной регрессии.
4. Последним шагом в анализе станет формулировка **параметрической модели**, которая подразумевает специфические предположения о Ω . Оценки в этом случае — это некоторая форма обобщенного метода наименьших квадратов или метода максимального правдоподобия, который будет представлен в главе 14.

В этой главе модель рассматривается в общих чертах. Важные приложения к панельным данным и системам уравнений обсуждаются в главах 11 и 10 соответственно.

9.2. Неэффективность оценок методами наименьших квадратов и инструментальных переменных

Важные результаты для классической модели со **сферическими ошибками**

$$E[\varepsilon | \mathbf{X}] = \mathbf{0}$$

и

$$E[\varepsilon\varepsilon' | \mathbf{X}] = \sigma^2\mathbf{I} \quad (9-2)$$

представлены в главах 2–6. Напомним, мы установили, что **оценка обычным методом наименьших квадратов (МНК)**

$$\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon \quad (9-3)$$

является наилучшей линейной, несмещенной (BLU), состоятельной и асимптотически нормально распределенной (CAN), и если ошибки распределены нормально, то, как и другие оценки методом максимального правдоподобия, рассматриваемые в главе 14, она асимптотически эффективна среди всех CAN-оценок. Теперь мы проанализируем, какие из этих свойств выполняются в модели (9-1).

Таким образом, оценка методом наименьших квадратов сохраняет только некоторые из своих желаемых свойств в этой модели. Оценка МНК остается несмещенной, состоятельной и асимптотически нормально распределенной. Однако она больше не будет эффективной, это утверждение необходимо дополнительно проверить, и обычные процедуры вывода в этом случае не подойдут.

9.2.1. Свойства МНК в конечных выборках

Возьмем математическое ожидание обеих частей уравнения (9-3) и получим, что если $E[\varepsilon | \mathbf{X}] = \mathbf{0}$, то

$$E[\mathbf{b}] = E_{\mathbf{X}}[E[\mathbf{b} | \mathbf{X}]] = \beta. \quad (9-4)$$

Следовательно, мы получили следующую теорему.

Теорема 9.1. *Свойства оценок \mathbf{b} в конечных выборках в обобщенной регрессионной модели*

Если регрессоры и ошибки не коррелированы, то оценка МНК остается несмещенной, даже если нарушается предположение (9-2). Оценка МНК является несмещенной в обобщенной регрессионной модели. При условии наличия нестохастических регрессоров, или условно относительно \mathbf{X} , выборочная дисперсия оценки МНК будет равна

$$\begin{aligned}
\text{Var}[\mathbf{b} | \mathbf{X}] &= E[(\mathbf{b} - \beta)(\mathbf{b} - \beta)' | \mathbf{X}] = \\
&= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon\varepsilon'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} | \mathbf{X}] = \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\sigma^2\Omega)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \\
&= \frac{\sigma^2}{n} \left(\frac{1}{n}\mathbf{X}'\mathbf{X} \right)^{-1} \left(\frac{1}{n}\mathbf{X}'\Omega\mathbf{X} \right) \left(\frac{1}{n}\mathbf{X}'\mathbf{X} \right)^{-1}.
\end{aligned} \tag{9-5}$$

Если регрессоры стохастические, безусловная дисперсия определяется как $E_{\mathbf{X}}[\text{Var}[\mathbf{b} | \mathbf{X}]]$. В (9-3) \mathbf{b} является линейной функцией от ε . Следовательно, если вектор ε нормально распределен, то

$$\mathbf{b} | \mathbf{X} \sim N[\beta, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\Omega\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}].$$

Окончательный результат заключается в том, что \mathbf{b} имеет те же свойства, что и в классической регрессионной модели. Так как дисперсия оценки МНК не равна $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$, статистические выводы, основанные на $s^2(\mathbf{X}'\mathbf{X})^{-1}$, могут привести к неверному результату. Это проявляется не только в том, что используется неправильная матрица, но и в том, что s^2 может быть смещенной оценкой σ^2 . Часто бывает невозможно узнать, больше или меньше $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ по сравнению с истинной дисперсией \mathbf{b} , поэтому даже с хорошей оценкой σ^2 традиционная оценка $\text{Var}[\mathbf{b} | \mathbf{X}]$ может оказаться не особенно полезной. И наконец, из-за того что мы обошлись без одного из основных предположений, знакомая процедура вывода, основанная на распределениях F и t , больше не годится. В дальнейшем мы несколько раз рассмотрим вопрос о том, насколько сильно можно ошибиться, если не принимать во внимание результат (9-5) и использовать знакомые процедуры, основанные на $s^2(\mathbf{X}'\mathbf{X})^{-1}$.

9.2.2. Асимптотические свойства МНК

Если $\text{Var}[\mathbf{b} | \mathbf{X}]$ сходится к нулю, то \mathbf{b} сходится в среднеквадратичном. При условии наличия доброкачественных регрессоров $(\mathbf{X}'\mathbf{X}/n)^{-1}$ сходится к постоянной матрице. Но $(\sigma^2/n)(\mathbf{X}'\Omega\mathbf{X}/n)$ необязательно сходится к чему-либо. Запишем это произведение в виде

$$\frac{\sigma^2}{n} \left(\frac{\mathbf{X}'\Omega\mathbf{X}}{n} \right) = \left(\frac{\sigma^2}{n} \right) \left(\frac{\sum_{i=1}^n \sum_{j=1}^n \omega_{ij} \mathbf{x}_i \mathbf{x}_j'}{n} \right) \tag{9-6}$$

и увидим, что, хотя первый множитель будет сходиться к нулю, матрица является суммой членов n^2 , деленной на n . Поэтому произведение будет скаляром, который равен $O(1/n)$, умноженному на матрицу $O(n)$, т.е. $O(1)$. Таким образом, на первый взгляд, если произведение в (9-6) действительно сходится, оно может сходиться к матрице из ненулевых констант. В этом случае ковариационная матрица оценок МНК не сходится к нулю и будет сложно установить состоятельность. Мы рассмотрим в деталях условия, при которых матрица в (9-6) сходится к матрице констант¹. Если это так, то в

¹Для того чтобы произведение в (9-6) исчезло, достаточно, чтобы $(\mathbf{X}'\Omega\mathbf{X}/n)$ было $O(n^\delta)$, где $\delta < 1$.

силу того что σ^2/n сходится к нулю, оценка обычным МНК состоятельна и не смещена.

Теорема 9.2. *Состоятельность МНК в обобщенной регрессионной модели*
Если $\mathbf{Q} = \text{plim}(\mathbf{X}'\mathbf{X}/n)$ и $\text{plim}(\mathbf{X}'\mathbf{\Omega}\mathbf{X}/n)$ — конечные положительно определенные матрицы, то \mathbf{b} — состоятельная оценка для β . В этих предположениях

$$\text{plim } \mathbf{b} = \beta.$$

Условия из теоремы 9.2 зависят как от \mathbf{X} , так и от $\mathbf{\Omega}$. Альтернативный подход², который разделяет две компоненты, выглядит следующим образом. Оценка обычным методом наименьших квадратов состоятельна в обобщенной регрессии, если:

- 1) наименьший характеристический корень $\mathbf{X}'\mathbf{X}$ неограниченно возрастает при $n \rightarrow \infty$, что подразумевает, что $\text{plim}(\mathbf{X}'\mathbf{X})^{-1} = \mathbf{0}$. Если регрессоры удовлетворяют условиям Гренандера **G1–G3** из табл. 4.2 подраздела 4.4.1, то они удовлетворяют этому требованию;
- 2) наибольший характеристический корень $\mathbf{\Omega}$ конечен для всех n . В случае наличия гетероскедастичности в случайных ошибках дисперсии являются характеристическими корнями, что накладывает на них ограничение конечности. Для моделей с автокорреляцией требования состоят в том, что элементы $\mathbf{\Omega}$ должны быть конечными и недиагональные элементы не должны быть слишком велики по сравнению с диагональными элементами. Мы подробно рассмотрим это условие позднее.

Оценка методом наименьших квадратов асимптотически нормально распределена, если предельное распределение

$$\sqrt{n}(\mathbf{b} - \beta) = \left(\frac{\mathbf{X}'\mathbf{X}}{n} \right)^{-1} \frac{1}{\sqrt{n}} \mathbf{X}'\boldsymbol{\varepsilon} \quad (9-7)$$

является нормальным. Если $\text{plim}(\mathbf{X}'\mathbf{X}/n) = \mathbf{Q}$, то предельное распределение правой части уравнения такое же, как и у

$$\mathbf{v}_{n,LS} = \mathbf{Q}^{-1} \frac{1}{\sqrt{n}} \mathbf{X}'\boldsymbol{\varepsilon} = \mathbf{Q}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i \boldsymbol{\varepsilon}_i, \quad (9-8)$$

где \mathbf{x}'_i — это ряд \mathbf{X} (разумеется, в предположении, что предельное распределение существует). Вопрос состоит в том, применима ли центральная предельная теорема напрямую к \mathbf{v} . Если ошибки просто гетероскедастичные и некоррелированные, тогда ответ будет положительный. Фактически мы уже доказали этот результат в подразделе 4.4.2, когда использовали центральную предельную теорему в форме Линдберга–Феллера (D.19) или теорему Ляпунова (D.20). Теорема допускает различные дисперсии у слагаемых. Точная дисперсия суммы имеет вид

$$E_{\mathbf{x}} \left[\text{Var} \left[\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i \boldsymbol{\varepsilon}_i \right] \middle| \mathbf{X} \right] = \frac{\sigma^2}{n} \sum_{i=1}^n \omega_i \mathbf{Q}_i,$$

²Amemiya (1985, p. 184).

которая для наших целей должна сходиться к положительно определенной матрице. При анализе классической модели гетерогенность дисперсии возникла из-за регрессоров, но предельное распределение в (4-33) было нормальным. Единственным отличием от классической модели здесь является то, что дисперсия ε меняется *также* от наблюдения к наблюдению. Следовательно, *доказательство асимптотической нормальности в подразделе 4.4.2 является достаточно общим, чтобы было возможно включить эту модель без модификаций.* Поскольку \mathbf{X} доброкачественные, как и диагональные элементы Ω , которые к тому же конечны, оценка методом наименьших квадратов асимптотически нормально распределена с ковариационной матрицей, указанной в (9-5). Таким образом,

в случае гетероскедастичности, если дисперсия ε_i конечна и не доминируется любым отдельным членом так, что условия центральной предельной теоремы Линдберга–Феллера применимы к $\mathbf{v}_{n,LS}$ в (9-8), оценка методом наименьших квадратов асимптотически нормально распределена с ковариационной матрицей

$$\text{Asy. Var}[\mathbf{b}] = \frac{\sigma^2}{n} \mathbf{Q}^{-1} \text{plim} \left(\frac{1}{n} \mathbf{X}' \Omega \mathbf{X} \right) \mathbf{Q}^{-1}. \quad (9-9)$$

В самом общем случае асимптотическую нормальность гораздо сложнее установить, так как суммы в (9-8) необязательно являются суммами независимых или хотя бы некоррелированных случайных чисел. Тем не менее Амеция (Amemiya (1985, p. 187)) и Андерсон (Anderson (1971)) установили асимптотическую нормальность \mathbf{b} в модели автокоррелированных остатков достаточно общо, для того чтобы было можно включить в нее большинство ситуаций, с которыми мы сталкиваемся на практике. Мы вернемся к этому вопросу в главе 20, когда будем рассматривать модели временных рядов. В заключение приведем теорему, которая выполняется всегда, кроме некоторых особенно неудачных случаев.

Теорема 9.3. *Асимптотическое распределение \mathbf{b} в модели обобщенной регрессии*

Если регрессоры являются доброкачественными и недиагональные элементы матрицы Ω уменьшаются достаточно быстро, оценка методом наименьших квадратов асимптотически нормально распределена со средним β и ковариационной матрицей, определенной в (9-9).

9.2.3. Устойчивая оценка асимптотической ковариационной матрицы

Относительно предшествующих результатов остался следующий вопрос: должны ли мы отказаться от оценки методом наименьших квадратов ввиду (9-5)? Конечно, если матрица Ω известна, то, как будет показано в подразделе 9.6.1, существует простая и эффективная оценка, основанная на ковариационной матрице, и соответственно ответ положительный. Если же сама

матрица Ω неизвестна, но известна ее структура и мы можем оценить матрицу Ω , используя информацию о выборке, ответ не настолько однозначен. В некоторых случаях вместо оценки методом наименьших квадратов более предпочтительно использовать оценку β , основанную на некоторых альтернативных процедурах, которые применяют матрицу $\hat{\Omega}$. Этот вопрос освещается в главах 10 и 11. Третий вариант состоит в том, что матрица Ω совершенно неизвестна (неизвестны как структура, так и отдельные значения ее элементов). В этой ситуации оценка методами наименьших квадратов или инструментальных переменных может быть единственной доступной оценкой, а значит, единственной доступной стратегией является подбор оценки для соответствующей асимптотической ковариационной матрицы \mathbf{b} .

Если матрица $\sigma^2\Omega$ известна, оценка асимптотической ковариационной матрицы \mathbf{b} в (9-10) будет равна

$$\mathbf{V}_{OLS} = \frac{1}{n} \left(\frac{1}{n} \mathbf{X}'\mathbf{X} \right)^{-1} \left(\frac{1}{n} \mathbf{X}'[\sigma^2\Omega]\mathbf{X} \right) \left(\frac{1}{n} \mathbf{X}'\mathbf{X} \right)^{-1}.$$

Матрицы из сумм квадратов и перекрестных произведений в левой и правой частях являются данными из выборки, поэтому могут быть легко вычислены. Проблема состоит в центральной матрице, которая включает неизвестную матрицу $\sigma^2\Omega$. Необходимо обратить внимание на то, что σ^2 не является отдельным неизвестным параметром. Так как Ω — неизвестная матрица, она может быть отнормирована, скажем, на κ , что вместе с σ^2 , умноженным на $1/\kappa$, даст то же самое произведение. В нашем случае мы избавляемся от неопределенности, предполагая, что $\text{tr}(\Omega) = n$, что равнозначно предположению $\sigma^2\Omega = \sigma^2\mathbf{I}$ в классической модели. Давайте предположим, что $\Sigma = \sigma^2\Omega$. Может показаться, что для оценки $(1/n)\mathbf{X}'\Sigma\mathbf{X}$ потребуется оценка ковариационной матрицы Σ , которая содержит $n(n+1)/2$ неизвестных параметров. Но, к счастью (так как при наличии n наблюдений этот метод заведомо безнадежен), это наблюдение не совсем верно. То, что нам необходимо, — это оценка $K(K+1)/2$ неизвестных элементов в матрице

$$\text{plim } \mathbf{Q}_* = \text{plim} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij} \mathbf{x}_i \mathbf{x}_j'.$$

Суть в том, что \mathbf{Q}_* — матрица сумм квадратов и перекрестных произведений, которая включает σ_{ij} и строки \mathbf{X} . Оценка методом наименьших квадратов \mathbf{b} является состоятельной оценкой β , что подразумевает, что остатки метода наименьших квадратов e_i являются «точечно» состоятельными оценками их выборочных аналогов ε_i . Общая идея заключается в том, чтобы использовать \mathbf{X} и \mathbf{e} , чтобы построить оценку для \mathbf{Q}_* .

Этот (возможно, отчасти нелогичный) принцип очень полезен для современных исследований. Более важные применения, включая общие модели для гетероскедастичности, автокорреляции и большое разнообразие моделей для панельных данных, могут быть оценены этим методом. Большим преимуществом является то, что эта оценка освобождает аналитика от необходимости предположений относительно структуры Ω . С таким инструментом, как устойчивая оценка ковариационной матрицы, один из отдельных трендов в современных исследованиях состоит в том, чтобы отказаться от

узкого предположения в пользу более широких устойчивых моделей, таких как вышеназванные. Случаи гетероскедастичности и автокорреляции будут рассмотрены в разделе 9.4 и главе 20 соответственно, а некоторые модели для панельных данных — в главе 11.

9.2.4. Метод инструментальных переменных

В главе 8 были рассмотрены случаи, когда регрессоры \mathbf{X} были коррелированы с остатками ϵ . Оценка методом инструментальных переменных (IV), представленная здесь, обладает некоторой устойчивостью, которой лишена оценка методом наименьших квадратов, что позволяет ей достичь состоятельности независимо от того, коррелированы или нет \mathbf{X} и ϵ , тогда как \mathbf{b} не является ни несмещенной, ни состоятельной. Однако эффективность не была предметом обсуждения при построении оценок при помощи метода инструментальных переменных. Мы рассмотрим здесь оценку методом инструментальных переменных, но так как это не лучшее, с чего следовало бы начать, для начала скажем несколько слов о последствиях наличия несферических остатков для эффективности оценки, как было рассмотрено для \mathbf{b} в предыдущем разделе. По сути, в этой части важным для рассмотрения станет вопрос: «работает» ли метод инструментальных переменных в обобщенной регрессионной модели? Состоятельность и асимптотическая нормальность будут полезными свойствами.

Оценка методом инструментальных переменных имеет вид

$$\begin{aligned} \mathbf{b}_{IV} &= [\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} \\ &= \beta + [\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\epsilon, \end{aligned} \quad (9-10)$$

где \mathbf{X} — набор из K регрессоров и \mathbf{Z} — набор из $L \geq K$ инструментальных переменных. Теперь рассмотрим расширение теорем 9.2 и 9.3 на случай оценки методом инструментальных переменных при $E[\epsilon\epsilon'|\mathbf{X}] = \sigma^2\Omega$.

Предположим, что \mathbf{X} и \mathbf{Z} доброкачественные, как предполагается в разделе 8.2, т.е.

$$\begin{aligned} \text{plim}(1/n)\mathbf{Z}'\mathbf{Z} &= \mathbf{Q}_{ZZ} - \text{положительно определенная матрица,} \\ \text{plim}(1/n)\mathbf{Z}'\mathbf{X} &= \mathbf{Q}_{ZX} = \mathbf{Q}'_{XZ} - \text{ненулевая матрица,} \\ \text{plim}(1/n)\mathbf{X}'\mathbf{X} &= \mathbf{Q}_{XX} - \text{положительно определенная матрица.} \end{aligned}$$

Чтобы избежать большого числа матричных вычислений, которые могут не поместиться в одну строчку, для удобства положим

$$\begin{aligned} \mathbf{Q}_{XX.Z} &= [\mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1}\mathbf{Q}_{ZX}]^{-1}\mathbf{Q}_{XZ}\mathbf{Q}_{ZZ}^{-1} \\ &= \text{plim} \left[\left(\frac{1}{n}\mathbf{X}'\mathbf{Z} \right) \left(\frac{1}{n}\mathbf{Z}'\mathbf{Z} \right)^{-1} \left(\frac{1}{n}\mathbf{Z}'\mathbf{X} \right) \right]^{-1} \left(\frac{1}{n}\mathbf{X}'\mathbf{Z} \right) \left(\frac{1}{n}\mathbf{Z}'\mathbf{Z} \right)^{-1}. \end{aligned}$$

Если \mathbf{Z} — допустимый набор инструментальных переменных, т.е. если второй член в (9-10) асимптотически исчезает, то

$$\text{plim } \mathbf{b}_{IV} = \beta + \mathbf{Q}_{XX.Z} \text{plim} \left(\frac{1}{n}\mathbf{Z}'\epsilon \right) = \beta.$$

Результат в точности совпадает в том, который мы получили ранее. Мы должны отметить, что в нескольких пунктах, в которых установили несмещенность или состоятельность оценки методом наименьших квадратов или методом инструментальных переменных, свойства ковариационной матрицы вектора ошибок не играют никакой роли; несмещенность — свойство средних. Таким образом, этот результат не должен стать сюрпризом. Поведение \mathbf{b}_{IV} в больших выборках зависит от поведения

$$\mathbf{v}_{n,IV} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{z}_i \varepsilon_i.$$

Этот результат в точности повторяет тот, что мы получили в подразделе 4.4.2. Если выборочное распределение \mathbf{v}_n сходится к нормальному распределению, то мы сможем построить асимптотическое распределение для \mathbf{b}_{IV} . Этот набор условий соответствует тому, который был необходим для \mathbf{X} , когда мы рассматривали \mathbf{b} , но с \mathbf{Z} вместо \mathbf{X} . Мы снова полагаемся на результаты Андерсона (Anderson (1971)) или Амемии (Amemiya (1985)), что при общих условиях

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{z}_i \varepsilon_i \xrightarrow{d} \mathbf{N} \left[\mathbf{0}, \sigma^2 \text{plim} \left(\frac{1}{n} \mathbf{Z}' \Omega \mathbf{Z} \right) \right].$$

Учитывая уже имеющиеся результаты, получаем следующее.

Теорема 9.4. *Асимптотическое распределение оценок методом инструментальных переменных в обобщенной регрессионной модели*

Если регрессоры и инструментальные переменные являются доброкачественными, как было описано выше, то

$$\mathbf{b}_{IV} \stackrel{a}{\sim} N[\boldsymbol{\beta}, \mathbf{V}_{IV}],$$

$$\text{где } \mathbf{V}_{IV} = \frac{\sigma^2}{n} (\mathbf{Q}_{\mathbf{X}\mathbf{X}.\mathbf{Z}}) \text{plim} \left(\frac{1}{n} \mathbf{Z}' \Omega \mathbf{Z} \right) (\mathbf{Q}'_{\mathbf{X}\mathbf{X}.\mathbf{Z}}).$$

Теорема 9.4 представляет собой аналог для случая оценок методом инструментальных переменных теорем 9.2 и 9.3 для метода наименьших квадратов.

9.3. Эффективное оценивание обобщенным методом наименьших квадратов

Для того чтобы получить эффективную оценку $\boldsymbol{\beta}$ в обобщенной регрессионной модели, необходимо знать Ω . Для начала полезно рассмотреть случаи, в которых матрица Ω — известная, симметричная, положительно опре-

деленная матрица. Это предположение иногда выполняется, хотя в большинстве моделей матрица Ω будет содержать неизвестные параметры, которые также должны быть оценены. Мы рассмотрим эти случаи в подразделе 9.6.2.

9.3.1. Обобщенный метод наименьших квадратов (ОМНК, GLS)

Так как матрица Ω — положительно определенная симметричная матрица, она может быть представлена в виде

$$\Omega = \mathbf{C}\mathbf{\Lambda}\mathbf{C}',$$

где столбцы \mathbf{C} — это характеристические векторы матрицы Ω , а характеристические корни матрицы Ω упорядочены в диагональной матрице $\mathbf{\Lambda}$. Пусть $\mathbf{\Lambda}^{1/2}$ — диагональная матрица с i -м диагональным элементом $\sqrt{\lambda_i}$ и $\mathbf{T} = \mathbf{C}\mathbf{\Lambda}^{1/2}$. Тогда $\Omega = \mathbf{T}\mathbf{T}'$. Также пусть $\mathbf{P}' = \mathbf{C}\mathbf{\Lambda}^{-1/2}$, так что $\Omega^{-1} = \mathbf{P}'\mathbf{P}$. Умножим слева модель в (9-1) на \mathbf{P} и получим

$$\mathbf{P}\mathbf{y} = \mathbf{P}\mathbf{X}\boldsymbol{\beta} + \mathbf{P}\boldsymbol{\varepsilon}$$

или

$$\mathbf{y}_* = \mathbf{X}_*\boldsymbol{\beta} + \boldsymbol{\varepsilon}_*. \quad (9-11)$$

Условная дисперсия $\boldsymbol{\varepsilon}_*$ имеет вид

$$E[\boldsymbol{\varepsilon}_*\boldsymbol{\varepsilon}_*' | \mathbf{X}_*] = \mathbf{P}\sigma^2\boldsymbol{\Omega}\mathbf{P}' = \sigma^2\mathbf{I},$$

т.е. классическая регрессионная модель преобразуется к этой трансформированной модели. Так как Ω предполагается известной, \mathbf{y}_* и \mathbf{X}_* — наблюдаемые данные. В классической модели оценка методом наименьших квадратов эффективна; следовательно, оценка

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{X}_*\mathbf{X}_*)^{-1}\mathbf{X}_*\mathbf{y}_* \\ &= (\mathbf{X}'\mathbf{P}'\mathbf{P}\mathbf{X})^{-1}\mathbf{X}'\mathbf{P}'\mathbf{P}\mathbf{y} \\ &= (\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{y} \end{aligned}$$

является **эффективной оценкой** $\boldsymbol{\beta}$. Эта оценка называется оценкой **обобщенным методом наименьших квадратов (ОМНК)** или оценкой Айткена (Aitken (1935)) вектора коэффициентов $\boldsymbol{\beta}$. Эта оценка отличается от оценки методом наименьших квадратов (МНК), которая использует «взвешивающую матрицу» \mathbf{I} вместо $\boldsymbol{\Omega}^{-1}$. Применяя классическую регрессионную модель в (9-11), мы получаем следующую теорему, которая включает аналог обобщенной регрессионной модели к результатам главы 4:

Теорема 9.5. *Свойства оценки обобщенным методом наименьших квадратов*

Если $E[\boldsymbol{\varepsilon}_* | \mathbf{X}_*] = \mathbf{0}$, то

$$E[\hat{\boldsymbol{\beta}} | \mathbf{X}_*] = E[(\mathbf{X}_*\mathbf{X}_*)^{-1}\mathbf{X}_*\mathbf{y}_* | \mathbf{X}_*] = \boldsymbol{\beta} + E[(\mathbf{X}_*\mathbf{X}_*)^{-1}\mathbf{X}_*\boldsymbol{\varepsilon}_* | \mathbf{X}_*] = \boldsymbol{\beta}.$$

Оценка ОМНК вектора коэффициентов $\hat{\beta}$ является несмещенной. Этот результат эквивалентен тому, что $E[\mathbf{P}\varepsilon | \mathbf{P}\mathbf{X}] = \mathbf{0}$, но, так как \mathbf{P} — матрица известных констант, мы возвращаемся к уже известному условию: $E[\varepsilon | \mathbf{X}] = \mathbf{0}$. Требование о том, что регрессоры и ошибки не должны быть коррелированы, остается без изменений.

Оценка ОМНК состоятельна, если $\text{plim}(1/n)\mathbf{X}'_*\mathbf{X}_* = \mathbf{Q}_*$, где \mathbf{Q}_* — конечная, положительно определенная матрица. Подставив, получаем:

$$\text{plim}[(1/n)\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}]^{-1} = \mathbf{Q}_*^{-1}. \quad (9-12)$$

Мы требуем, чтобы $\mathbf{X}_* = \mathbf{P}\mathbf{X}$, а не исходные данные \mathbf{X} были доброкачественными³. При выполнении предположения (9-1) имеет место следующий результат:

оценка ОМНК будет асимптотически нормально распределена со средним β и дисперсией

$$\text{Var}[\hat{\beta} | \mathbf{X}_*] = \sigma^2(\mathbf{X}'_*\mathbf{X}_*)^{-1} = \sigma^2(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}. \quad (9-13)$$

Оценка ОМНК $\hat{\beta}$ — линейная несмещенная оценка с наименьшей дисперсией в обобщенной регрессионной модели. Это утверждение следует из применения теоремы Гаусса–Маркова к модели (9-11). Теорема 9.5 носит название **теорема Айткена (1935)**, а оценка $\hat{\beta}$ иногда называется оценкой Айткена. Этот общий результат включает теорему Гаусса–Маркова как частный случай при $\boldsymbol{\Omega} = \mathbf{I}$.

Для тестирования гипотез мы можем использовать полный набор результатов, полученных в главе 5, но применительно к трансформированной модели (9-11). Подходящей статистикой для тестирования J линейных ограничений $\mathbf{R}\beta = \mathbf{q}$ является

$$F[J, n - K] = \frac{(\mathbf{R}\hat{\beta} - \mathbf{q})'[\mathbf{R}\hat{\sigma}^2(\mathbf{X}'_*\mathbf{X}_*)^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\beta} - \mathbf{q})}{J} = \frac{(\hat{\varepsilon}'_c\hat{\varepsilon}_c - \hat{\varepsilon}'\hat{\varepsilon})/J}{\hat{\sigma}^2},$$

где вектор остатков

$$\hat{\varepsilon} = \mathbf{y}_* - \mathbf{X}_*\hat{\beta}$$

и

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n - K} = \frac{(\mathbf{y} - \mathbf{X}\hat{\beta})'\boldsymbol{\Omega}^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta})}{n - K}. \quad (9-14)$$

Остатки ОМНК регрессии с ограничениями $\hat{\varepsilon}_c = \mathbf{y}_* - \mathbf{X}_*\hat{\beta}_c$ вычисляются по формуле

$$\hat{\beta}_c = \hat{\beta} - [\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}]^{-1}\mathbf{R}'[\mathbf{R}(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\beta} - \mathbf{q})^4.$$

³И снова, чтобы включить временной тренд в модель, мы можем немного ослабить эту предпосылку.

⁴Заметим, что эта оценка является оценкой МНК с ограничениями для преобразованных данных. [См. (5-23).]

Подводя итог, скажем, что все результаты для классической модели, включая обычные процедуры вывода, применимы и к преобразованной модели (9-11).

В обобщенной регрессионной модели нет точного аналога для R^2 . В этом случае предлагаются некоторые альтернативы, но пользоваться ими нужно с осторожностью. Например, одна из них — R^2 в преобразованной регрессии (9-11). Но эта регрессия не обязательно должна иметь свободный член, так что R^2 не ограничен нулем и единицей. Даже если в регрессии имеется свободный член, преобразованная регрессия — инструмент для вычислений, а не модель, которая нас интересует. Нам неинтересно, хорошо или плохо такая «модель» (9-11) описывает данные; зависимая переменная в этой модели y_* отличается от той, что была в исходной модели. Обычный R^2 часто показывает, что качество подгонки модели улучшается при коррекции на гетероскедастичность и ухудшается при коррекции модели на автокорреляцию, но оба вида изменений могут чаще относиться к замене зависимой переменной на y_* . Более убедительная мера подгонки может быть основана на остатках исходной модели, оцененной при помощи ОМНК:

$$R_G^2 = 1 - \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{\sum_{i=1}^n (y_i - \bar{y})^2}.$$

Как и ранее, эта величина не ограничена единичным интервалом. Более того, этот параметр не может быть использован с высокой степенью надежности для сравнения моделей. Оценка обобщенным методом наименьших квадратов минимизирует **обобщенную сумму квадратов**

$$\epsilon_*' \epsilon_* = (y - X\beta)' \Omega^{-1} (y - X\beta),$$

а не $\epsilon' \epsilon$. Таким образом, нет никакой уверенности, что, например, пропуск переменных в модели приведет к уменьшению R_G^2 , что имеет место в случае с R^2 . Были также предложены другие индикаторы качества подгонки, представляющие собой в основном функции от суммы квадратов остатков (исходных или взвешенных на Ω^{-1}) и ограниченные нулем и единицей⁵. К сожалению, все они обладают как минимум одним недостатком, обозначенным выше. Индикаторы — аналоги R^2 — для обобщенной регрессионной модели играют исключительно описательную роль. В этом случае квадрат выборочной корреляции между наблюдаемыми и предсказанными по модели значениями $r_{y, \hat{y}}^2 = \text{corr}^2(y, \hat{y}) = \text{corr}^2(y, x' \hat{\beta})$ был бы полезным описательным показателем. Необходимо обратить внимание, что этот показатель не является долей объясненной дисперсии, как R^2 ; это показатель соответствия предсказанных моделью значений реальным данным.

9.3.2. Доступный обобщенный метод наименьших квадратов (ДОМНК, FGLS)

Чтобы использовать результаты подраздела 9.3.1, необходимо знать Ω . Если Ω содержит неизвестные параметры, которые должны быть оценены, то обобщенный метод наименьших квадратов нельзя реализовать. Но если

⁵Judge et al. (1985, p. 32) и Buse (1973).

на матрицу Ω не наложить дополнительные ограничения, то в $\sigma^2\Omega$ появятся $n(n+1)/2$ дополнительных параметров. Это число слишком велико для оценки модели, имея всего n наблюдений. Очевидно, на модель должна быть наложена некоторая структура, если мы собираемся ее оценить.

Обычно предполагается небольшое число параметров α таких, что $\Omega = \Omega(\alpha)$. Например, часто используемая формула в анализе временных рядов имеет вид

$$\Omega(\rho) = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \rho^2 & \dots & \rho^{n-2} \\ & & & & \vdots & \\ \rho^{n-1} & \rho^{n-2} & \dots & & & 1 \end{bmatrix},$$

она включает только один дополнительный неизвестный параметр. Модель, учитывающая гетероскедастичность, также имеет один новый параметр:

$$\sigma_i^2 = \sigma^2 z_i^\theta. \quad (9-15)$$

Предположим далее, что $\hat{\alpha}$ — состоятельная оценка α . (Позднее мы обсудим, как эта оценка может быть получена.) Чтобы сделать оценку ОМНК реализуемой, мы должны использовать $\hat{\Omega} = \Omega(\hat{\alpha})$ вместо истинной Ω . Вопрос, который будет рассмотрен здесь: требует ли использование $\Omega(\hat{\alpha})$ изменения результатов, полученных в подразделе 9.3.1.

Может показаться, что если $\text{plim } \hat{\alpha} = \alpha$, то использование $\hat{\Omega}$ асимптотически эквивалентно использованию реальной Ω^6 . Пусть оценкой доступным обобщенным методом наименьших квадратов (ДОМНК, FGLS) будет называться

$$\hat{\beta} = (\mathbf{X}'\hat{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\Omega}^{-1}\mathbf{y}.$$

Условия, которые подразумевают, что $\hat{\beta}$ асимптотически эквивалентна β , могут быть записаны как

$$\text{plim} \left[\left(\frac{1}{n} \mathbf{X}'\hat{\Omega}^{-1}\mathbf{X} \right) - \left(\frac{1}{n} \mathbf{X}'\Omega^{-1}\mathbf{X} \right) \right] = \mathbf{0} \quad (9-16)$$

и

$$\text{plim} \left[\left(\frac{1}{\sqrt{n}} \mathbf{X}'\hat{\Omega}^{-1}\boldsymbol{\varepsilon} \right) - \left(\frac{1}{\sqrt{n}} \mathbf{X}'\Omega^{-1}\boldsymbol{\varepsilon} \right) \right] = \mathbf{0}. \quad (9-17)$$

Первое уравнение говорит, что если матрица взвешенной суммы квадратов, основанная на истинной Ω , сходится к положительно определенной матрице, то матрица, основанная на $\hat{\Omega}$, сходится к той же самой матрице. Мы предполагаем, что это так. Во втором условии утверждается, что если преобразованные регрессоры являются доброкачественными, то правая часть суммы будет иметь предельное нормальное распределение. Это условие в точности такое же, какое мы использовали в главе 4 для вывода асимптотического распределения оценки МНК; здесь мы используем тот же результат

⁶Это уравнение иногда обозначают $\text{plim } \hat{\Omega} = \Omega$. Поскольку Ω имеет размер $n \times n$, она не может иметь предел по вероятности. Мы будем использовать этот термин для поэлементной сходимости.

для X_* и ϵ_* . Следовательно, (9-17) требует выполнения тех же условий при замене Ω на $\hat{\Omega}$ ⁷.

Эти условия в принципе могут различаться от случая к случаю. К счастью, в наиболее известных случаях они выполняются. Если мы предположим, что они выполнены, то оценка ДОМНК, основанная на $\hat{\alpha}$, имеет те же **асимптотические свойства**, что и оценка ОМНК. Этот результат чрезвычайно полезен. Особенно следует обратить внимание на следующую теорему.

Теорема 9.6. Эффективность оценки ДОМНК

Асимптотически эффективная оценка ДОМНК не подразумевает наличие эффективной оценки α ; для полной эффективности оценки ДОМНК достаточно только состоятельности оценки.

Как правило (за исключением простейших случаев), **свойства в конечных выборках** и точное распределение оценки ДОМНК неизвестны. Асимптотическая эффективность оценок ДОМНК может не распространяться на маленькие выборки вследствие изменчивости при оценке матрицы Ω . Анализ для случая гетероскедастичности представлен в работе Тейлора (Taylor (1977)). Модель автокорреляции проанализирована в статье Грилихеса и Рао (Griliches, Rao (1969)). В обоих случаях авторы установили, что при большом наборе параметров ДОМНК более эффективен, чем метод наименьших квадратов. Но если отклонение от классических предпосылок невелико, метод наименьших квадратов может быть более эффективен, чем ДОМНК, в маленькой выборке.

9.4. Гетероскедастичность и взвешенный метод наименьших квадратов

Случайные ошибки регрессии, дисперсии которых непостоянны, являются гетероскедастичными. **Гетероскедастичность** встречается как в межобъектных данных, так и во временных рядах. Например, даже после учета размера фирм мы ожидаем увидеть большую дисперсию в прибыли больших фирм по сравнению с маленькими. Разброс в прибыли может также зависеть от продуктовой диверсификации, расходов на исследования и особенностей индустрии, следовательно, может различаться даже среди фирм одинакового размера. При анализе расходов семей мы обнаружили, что существует большая дисперсия в расходах на одни и те же группы товаров среди семей с высоким доходом по сравнению с семьями с низким доходом вследствие большего выбора, доступного семьям с высоким доходом⁸.

⁷Точнее, требуется только, чтобы правая часть суммы имела *любое* предельное распределение, тогда и левая часть будет иметь то же распределение. Разумеется, это распределение может не быть нормальным, но это маловероятно, за исключением специально придуманного теоретического контрпримера.

⁸Prais, Houthakker (1955).

В регрессионной модели с гетероскедастичностью

$$\text{Var}[\varepsilon_i | \mathbf{X}] = \sigma_i^2, \quad i = 1, \dots, n.$$

Мы продолжаем предполагать, что ошибки попарно не коррелированы. Следовательно,

$$E[\varepsilon\varepsilon' | \mathbf{X}] = \sigma^2\mathbf{\Omega} = \sigma^2 \begin{bmatrix} \omega_1 & 0 & 0 & \dots & 0 \\ 0 & \omega_2 & 0 & \dots & \\ & & & \ddots & \\ 0 & 0 & 0 & \dots & \omega_n \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & 0 & \dots & \\ & & & \ddots & \\ 0 & 0 & 0 & \dots & \sigma_n^2 \end{bmatrix}.$$

Иногда может оказаться полезным записать $\sigma_i^2 = \sigma^2\omega_i$. Эта форма имеет произвольный масштаб, что позволяет нам использовать нормализацию

$$\text{tr}(\mathbf{\Omega}) = \sum_{i=1}^n \omega_i = n.$$

Это делает классическую регрессию с гомоскедастичными отклонениями простым частным случаем с $\omega_i = 1, i = 1, \dots, n$. Интуитивно можно думать о ω как о весах, которые отмасштабированы таким образом, чтобы отражать только различия в отклонениях ошибок. Масштабирующий фактор σ^2 обеспечивает полное масштабирование процесса ошибок.

Пример 9.1. Регрессия с гетероскедастичностью

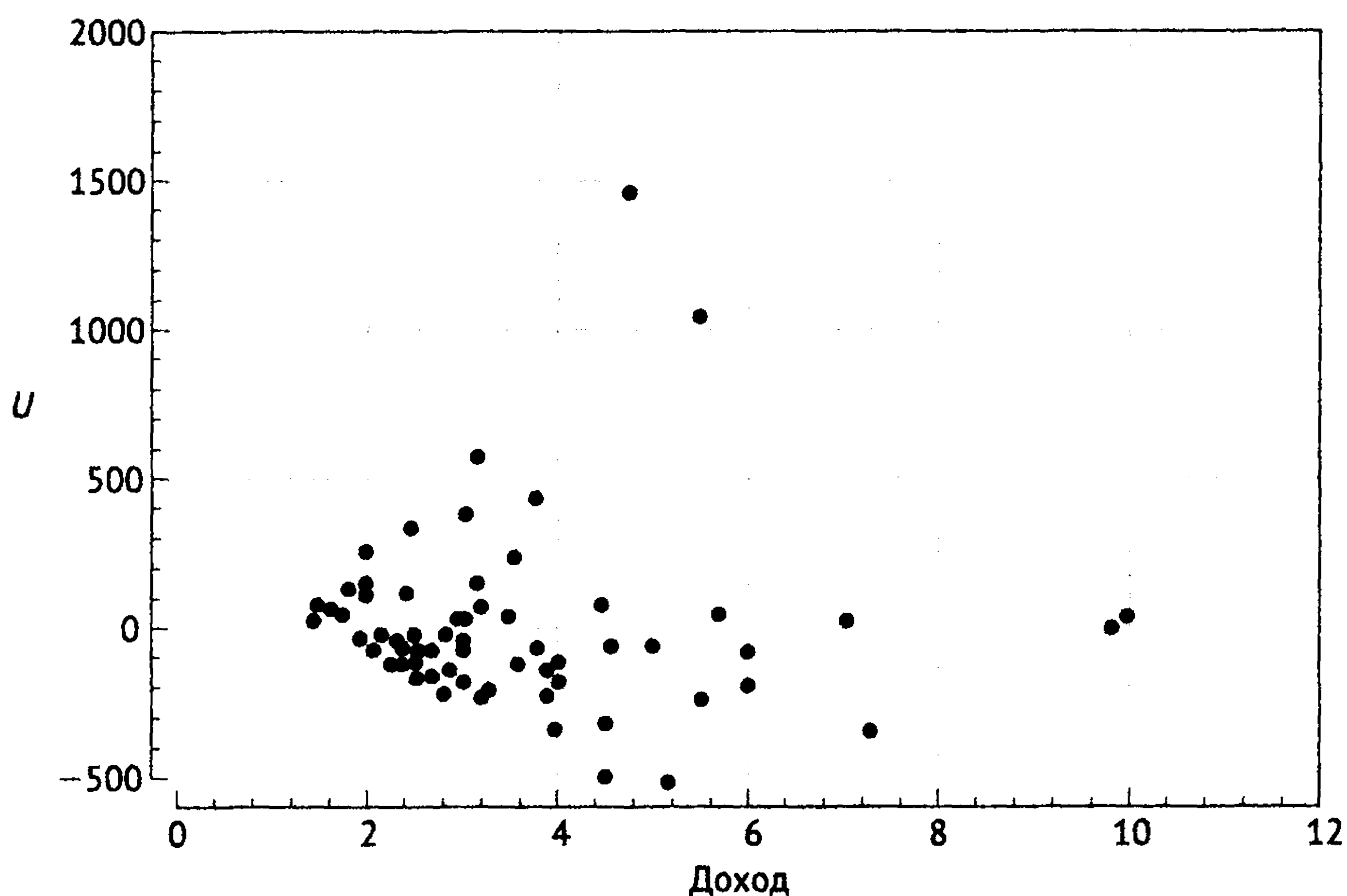


Рис. 9.1. График остатков по отношению к доходу

В табл. F7.3 приложения представлены месячные данные о расходах по кредитным картам 13 444 индивидов. Линейная регрессия месячных расходов на константу, возраст, доход и его квадрат, а также дамми-переменную для домохозяйств, оцененная по 72 наблюдениям, для которых расходы были не равны нулю, дает остатки, которые изображены на рис. 9.1. Вид остатков свидетельствует о наличии гетероскедастичности. (Подвыборка из 72 наблюдений представлена в табл. F9.1 приложения.)

Мы изучим регрессионную модель с гетероскедастичностью сначала в общем виде, а затем для некоторых отдельных форм ковариационной матрицы ошибок. Затем мы рассмотрим **устойчивые оценки**. В подразделе 9.4.4 представлены подходящие оценки для асимптотической ковариационной матрицы оценки методом наименьших квадратов. Тесты на спецификацию для случая гетероскедастичности рассмотрены в разделе 9.5. В разделе 9.6 описывается обобщенный (взвешенный) метод наименьших квадратов, который требует знания как минимум формы матрицы Ω . Наконец, в разделе 9.7 рассмотрены два общих приложения модели.

9.4.1. Обычный метод наименьших квадратов

Мы показали в разделе 9.2, что при наличии гетероскедастичности оценка \mathbf{b} , полученная при помощи метода наименьших квадратов, все еще является несмещенной, состоятельной и асимптотически нормально распределенной. Асимптотическая ковариационная матрица имеет вид

$$\text{Asy. Var}[\mathbf{b}] = \frac{\sigma^2}{n} \left(\text{plim} \frac{1}{n} \mathbf{X}'\mathbf{X} \right)^{-1} \left(\text{plim} \frac{1}{n} \mathbf{X}'\Omega\mathbf{X} \right) \left(\text{plim} \frac{1}{n} \mathbf{X}'\mathbf{X} \right)^{-1}. \quad (9-18)$$

Оценка асимптотической ковариационной матрицы должна выглядеть так:

$$\text{Var}[\mathbf{b} | \mathbf{X}] = (\mathbf{X}'\mathbf{X})^{-1} \left(\sigma^2 \sum_{i=1}^n \omega_i \mathbf{x}_i \mathbf{x}_i' \right) (\mathbf{X}'\mathbf{X})^{-1}.$$

[См. (9-5).] Предполагая, как всегда, что регрессоры являются доброкачественными, так что $(\mathbf{X}'\mathbf{X}/n)^{-1}$ сходится к положительно определенной матрице, мы обнаруживаем, что состоятельность \mathbf{b} в среднеквадратичном зависит от асимптотического поведения матрицы:

$$\mathbf{Q}_n^* = \frac{\mathbf{X}'\Omega\mathbf{X}}{n} = \frac{1}{n} \sum_{i=1}^n \omega_i \mathbf{x}_i \mathbf{x}_i'.$$

Если \mathbf{Q}_n^* сходится к положительно определенной матрице \mathbf{Q}^* , то при $n \rightarrow \infty$ \mathbf{b} будет сходиться к β в среднеквадратичном. В большинстве случаев если ω_i конечно для всех i , то мы ожидаем выполнения этого предположения. Обратите внимание на то, что \mathbf{Q}_n^* представляет собой взвешенную сумму квадратов и перекрестных произведений \mathbf{x} с весами ω_i/n , которые в сумме

дают 1. Мы уже предположили, что другая взвешенная сумма, $\mathbf{X}'\mathbf{X}/n$, в которой весами являются $1/n$, сходится к положительно определенной матрице \mathbf{Q} , поэтому будет неожиданно, если \mathbf{Q}_n^* не будет сходиться аналогичным образом. В общем случае мы ожидаем, что

$$\mathbf{b} \sim^a N \left[\boldsymbol{\beta}, \frac{\sigma^2}{n} \mathbf{Q}^{-1} \mathbf{Q}^* \mathbf{Q}^{-1} \right], \quad \text{с } \mathbf{Q}^* = \text{plim } \mathbf{Q}_n^*.$$

Формальное доказательство представлено в разделе 4.4 при $\mathbf{Q}_i = \omega_i \mathbf{x}_i \mathbf{x}_i'$.

9.4.2. Неэффективность обычного метода наименьших квадратов

Из предыдущих результатов следует, что оценка \mathbf{b} неэффективна по сравнению с оценкой обобщенным методом наименьших квадратов. Насколько сильно это различие, зависит от выборки, однако есть и некоторые общие моменты. Как можно было ожидать, чем больше разброс в ω_i по наблюдениям, тем больше эффективность оценки ОМНК по сравнению с МНК. Влияние на эффективность оценки будет сильно зависеть от природы дисперсий ошибок. В случаях когда ω_i зависит от переменных, которые также встречаются в модели, чем больше разброс в этих переменных, тем больше будет выигрыш от использования ОМНК. Важно заметить, однако, что оба этих сравнения основаны на знании матрицы $\boldsymbol{\Omega}$. На практике один из двух случаев, скорее всего, будет иметь место. Если у нас есть точная информация о $\boldsymbol{\Omega}$, наличие неэффективной оценки — спорный вопрос. В любом случае для оценивания мы будем использовать ОМНК или ДОМНК. В более общем случае у нас не будет детальной информации о матрице $\boldsymbol{\Omega}$, так что сравнение невозможно.

9.4.3. Оцененная ковариационная матрица \mathbf{b}

Если мы знаем наверняка тип гетероскедастичности в нашей модели, то оценка обычным методом наименьших квадратов нежелательна; мы должны использовать обобщенный метод наименьших квадратов. Однако точная форма гетероскедастичности чаще всего неизвестна. В этом случае обобщенный метод наименьших квадратов неприменим и нам необходимо улучшить то, что возможно, из результатов обычного метода наименьших квадратов. Ковариационная матрица, оцененная обычным способом, для оценки методом наименьших квадратов $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ не может быть использована в данном случае; необходимо использовать матрицу $\sigma^2(\mathbf{X}'\mathbf{X})^{-1} \times (\mathbf{X}'\boldsymbol{\Omega}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}$. Скорее всего, эти две оценки не совпадут, поэтому обычные оценки стандартных ошибок, скорее всего, некорректны. В этом разделе мы рассмотрим, насколько ошибочна оценка, полученная обычным способом.

Как обычно,

$$s^2 = \frac{\mathbf{e}'\mathbf{e}}{n-K} = \frac{\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon}}{n-K}, \quad (9-19)$$

где $\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Расписав уравнение, получим

$$s^2 = \frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{n-K} - \frac{\boldsymbol{\varepsilon}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}}{n-K}. \quad (9-20)$$

Рассмотрим условное математическое ожидание каждой из двух частей этой разности. Получим

$$E \left[\frac{\varepsilon' \varepsilon}{n - K} \mid \mathbf{X} \right] = \frac{\text{tr} E [\varepsilon \varepsilon' \mid \mathbf{X}]}{n - K} = \frac{n \sigma^2}{n - K}. \quad (9-21)$$

[Мы использовали условие $\text{tr}(\Omega) = n$.] В дополнение

$$\begin{aligned} E \left[\frac{\varepsilon' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \varepsilon}{n - K} \mid \mathbf{X} \right] &= \frac{\text{tr} \{ E [(\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \varepsilon \varepsilon' \mathbf{X} \mid \mathbf{X}] \}}{n - K} = \\ &= \frac{\text{tr} \left[\sigma^2 \left(\frac{\mathbf{X}' \mathbf{X}}{n} \right)^{-1} \left(\frac{\mathbf{X}' \Omega \mathbf{X}}{n} \right) \right]}{n - K} = \frac{\sigma^2}{n - K} \text{tr} \left[\left(\frac{\mathbf{X}' \mathbf{X}}{n} \right)^{-1} \mathbf{Q}_n^* \right], \end{aligned} \quad (9-22)$$

где \mathbf{Q}_n^* определяется после выражения (9-18). При $n \rightarrow \infty$ член в (9-21) сходится к σ^2 . Член в (9-22) сходится к нулю, если \mathbf{b} — состоятельная оценка, так как обе матрицы в произведении конечны. Следовательно,

$$\text{если } \mathbf{b} \text{ состоятельна, то } \lim_{n \rightarrow \infty} E [s^2] = \sigma^2.$$

Также можно показать — мы оставим это для упражнения, — что если четвертый момент каждой ошибки конечен и все наши предположения выполняются, то

$$\lim_{n \rightarrow \infty} \text{Var} \left[\frac{\mathbf{e}' \mathbf{e}}{n - K} \right] = \lim_{n \rightarrow \infty} \text{Var} \left[\frac{\mathbf{e}' \mathbf{e}}{n - K} \right] = 0.$$

Этот результат, следовательно, подразумевает, что

$$\text{если } \text{plim } \mathbf{b} = \boldsymbol{\beta}, \text{ то } \text{plim } s^2 = \sigma^2.$$

Прежде чем продолжить, полезно рассмотреть этот результат подробнее. Нормализация $\text{tr}(\Omega) = n$ подразумевает, что

$$\sigma^2 = \bar{\sigma}^2 = \frac{1}{n} \sum_i \sigma_i^2 \quad \text{и} \quad \omega_i = \frac{\sigma_i^2}{\bar{\sigma}^2}.$$

Следовательно, предыдущий результат о сходимости подразумевает, что оценка методом наименьших квадратов s^2 сходится к $\text{plim } \bar{\sigma}^2$, что есть предел по вероятности средней дисперсии ошибок в предположении, что он существует. Поэтому для получения результата необходимы некоторые дальнейшие предположения об этой дисперсии.

Разность между оценкой, полученной обычным способом, и подходящей (истинной) ковариационной матрицей для \mathbf{b} :

$$\text{Est. Var}[\mathbf{b} \mid \mathbf{X}] - \text{Var}[\mathbf{b} \mid \mathbf{X}] = s^2 (\mathbf{X}' \mathbf{X})^{-1} - \sigma^2 (\mathbf{X}' \mathbf{X})^{-1} (\mathbf{X}' \Omega \mathbf{X}) (\mathbf{X}' \mathbf{X})^{-1}. \quad (9-23)$$

В больших выборках (таких, что $s^2 \approx \sigma^2$) эта разность примерно равна

$$\mathbf{D} = \frac{\sigma^2}{n} \left(\frac{\mathbf{X}' \mathbf{X}}{n} \right)^{-1} \left[\frac{\mathbf{X}' \mathbf{X}}{n} - \frac{\mathbf{X}' \Omega \mathbf{X}}{n} \right] \left(\frac{\mathbf{X}' \mathbf{X}}{n} \right)^{-1}. \quad (9-24)$$

Разность между двумя матрицами:

$$\begin{aligned} \Delta &= \frac{\mathbf{X}'\mathbf{X}}{n} - \frac{\mathbf{X}'\Omega\mathbf{X}}{n} = \sum_{i=1}^n \left(\frac{1}{n}\right) \mathbf{x}_i \mathbf{x}_i' - \sum_{i=1}^n \left(\frac{\omega_i}{n}\right) \mathbf{x}_i \mathbf{x}_i' = \\ &= \frac{1}{n} \sum_{i=1}^n (1 - \omega_i) \mathbf{x}_i \mathbf{x}_i', \end{aligned} \quad (9-25)$$

где \mathbf{x}_i' — i -й ряд \mathbf{X} . Здесь наблюдаются две взвешенные средние матрицы $\mathbf{Q}_i = \mathbf{x}_i \mathbf{x}_i'$, где вес 1 используется для первого слагаемого и ω_i — для второго. Нормировка $\text{tr}(\Omega) = n$ подразумевает, что $\sum_i (\omega_i/n) = 1$. Тот факт, отличается ли взвешенное среднее, основанное на ω_i/n , от среднего, полученного с помощью $1/n$, зависит от весов. Если веса коррелированы со значениями в \mathbf{x}_i , то различие может быть значительным. Если веса не коррелированы с $\mathbf{x}_i \mathbf{x}_i'$, взвешенное среднее скорее будет равно невзвешенному среднему⁹.

Следовательно, сравнение основывается на том, зависит ли гетероскедастичность от x_k или $x_j \times x_k$. Общий вывод состоит в следующем: *если гетероскедастичность не коррелирована с переменными в модели, то, по крайней мере в больших выборках, расчеты в рамках метода наименьших квадратов не исказят конечный результат, хотя и не станут оптимальным использованием данных*. Например, в групповых моделях гетероскедастичности из подраздела 9.7.2, если наблюдения сгруппированы в подвыборки таким образом, что они не коррелированы с переменными в \mathbf{X} , оценка обычным МНК для $\text{Var}[\mathbf{b}]$ будет, по крайней мере в больших выборках, обеспечивать надежную оценку соответствующей ковариационной матрицы. Однако необходимо помнить, что оценка методом наименьших квадратов будет неэффективна настолько, насколько больше различия среди переменных в группах¹⁰.

Последний результат очень полезен, но мы не должны быть излишне оптимистичны. Во-первых, никто не отменял тот факт, что обычный метод наименьших квадратов неэффективен. Во-вторых, если первичная предпосылка анализа — что гетероскедастичность не связана с параметрами модели — неверна, то стандартные ошибки, оцененные обычным способом, могут быть достаточно далеки от необходимых значений.

9.4.4. Оценка подходящей ковариационной матрицы для метода наименьших квадратов

Из сказанного выше ясно, что гетероскедастичность имеет ряд потенциально серьезных последствий для тестирования гипотез, основанных на результатах метода наименьших квадратов. Однако применение более подходящих оценок требует детальной информации о Ω . Бывает, что форма

⁹Предположим, например, что \mathbf{X} содержит один столбец и что \mathbf{x}_i и ω_i независимые и одинаково распределенные случайные величины. Тогда $\mathbf{x}'\mathbf{x}/n$ сходится к $E[x_i^2]$, в то время как $\mathbf{x}'\Omega\mathbf{x}/n$ сходится к $\text{Cov}[\omega_i, x_i^2] + E[\omega_i]E[x_i^2]$. $E[\omega_i] = 1$, поэтому если ω и x^2 не коррелированы, то эти суммы имеют одинаковый предел по вероятности.

¹⁰Некоторые общие результаты, включая анализ свойств оценки, основанной на оцененных дисперсиях, представлены в работе Тейлора (Taylor (1977)).

гетероскедастичности неизвестна. Уайт (White (1980)) показал, что еще возможно получить подходящую оценку для дисперсии оценки методом наименьших квадратов, даже если гетероскедастичность не зависит от переменных в \mathbf{X} . Обращаясь к (9-18), мы ищем оценку

$$\mathbf{Q}_* = \frac{1}{n} \sum_{i=1}^n \sigma_i^2 \mathbf{x}_i \mathbf{x}_i'$$

Уайт (White (1980)) показал, что при общих условиях оценка¹¹

$$\mathbf{S}_0 = \frac{1}{n} \sum_{i=1}^n e_i^2 \mathbf{x}_i \mathbf{x}_i' \quad (9-26)$$

имеет

$$\text{plim } \mathbf{S}_0 = \text{plim } \mathbf{Q}_* \quad {}^{11}.$$

Мы можем набросать доказательство этого результата, используя результаты, которые получили в разделе 4.4¹². Сначала заметим, что \mathbf{Q}_* не является параметром самой матрицы. Это взвешенная сумма внешних произведений строк матрицы \mathbf{X} (или \mathbf{Z} для случая инструментальных переменных). Поэтому мы пытаемся не оценить \mathbf{Q}_* , а найти функцию от данных выборки, которая будет близка к функции наблюдаемых параметров, так как размер выборки растет. Это отличие важно. Мы не оцениваем среднюю матрицу в (9-9) или (9-18); мы пытаемся построить матрицу из выборочных значений, которая будет вести себя так же, как и матрица. По существу, если \mathbf{Q}_* сходится к конечной положительной матрице, то мы будем искать функцию от выборочных данных, которая сходится к той же самой матрице. Предположим, что правильные ошибки ε_i могут быть наблюдаемы. Тогда каждый член в \mathbf{Q}_* должен быть равен $E[\varepsilon_i^2 \mathbf{x}_i \mathbf{x}_i' | \mathbf{x}_i]$. При довольно слабых предпосылках относительно \mathbf{x}_i мы можем использовать закон больших чисел (см. теоремы D.4–D.9), чтобы получить, что если \mathbf{Q}_* имеет предел по вероятности, то

$$\text{plim } \frac{1}{n} \sum_{i=1}^n \sigma_i^2 \mathbf{x}_i \mathbf{x}_i' = \text{plim } \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2 \mathbf{x}_i \mathbf{x}_i'.$$

Последний штрих заключается в том, чтобы обосновать замену ε_i на e_i в \mathbf{S}_0 . Для этого будет достаточно состоятельности оценки \mathbf{b} для β . (Действительно, остатков, основанных на любой состоятельной оценке для β , будет достаточно для этой оценки, но на данный момент мы имеем только \mathbf{b} или \mathbf{b}_{IV} .) Конечный результат состоит в том, что **устойчивая к гетероскедастичности оценка Уайта**

$$\begin{aligned} \text{Est. Asy. Var}[\mathbf{b}] &= \frac{1}{n} \left(\frac{1}{n} \mathbf{X}' \mathbf{X} \right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n e_i^2 \mathbf{x}_i \mathbf{x}_i' \right) \left(\frac{1}{n} \mathbf{X}' \mathbf{X} \right)^{-1} = \\ &= n(\mathbf{X}' \mathbf{X})^{-1} \mathbf{S}_0 (\mathbf{X}' \mathbf{X})^{-1} \end{aligned} \quad (9-27)$$

¹¹ Eicker (1967), Horn, Horn, Duncan (1975), MacKinnon, White (1985).

¹² Мы рассмотрим только общую схему доказательства. Формальные результаты представлены в работах Уайта (White (1980, (2001))).

может быть использована для оценки асимптотической ковариационной матрицы для \mathbf{b} .

Этот результат чрезвычайно полезен и важен¹³. Он предполагает, что за неимением информации о типе гетероскедастичности мы все еще можем тестировать гипотезы на основе оценок методом наименьших квадратов. Этот вывод особенно полезен, если у нас нет уверенности относительно точной природы гетероскедастичности (что случается чаще всего). Мы рассмотрим некоторые примеры в разделе 9.7.

Авторы многих исследований пытались найти улучшение оценки Уайта для МНК¹⁴. Асимптотические свойства оценки не подлежат сомнению, но ее полезность в малых выборках остается открытым вопросом. Возможные проблемы возникают из общего результата, состоящего в том, что квадраты МНК остатков склонны недооценивать квадраты истинных ошибок. (Поэтому мы используем $1/(n - K)$, а не $1/n$ при вычислении s^2 .) Окончательный результат состоит в том, что в маленьких выборках, по крайней мере так предполагается некоторыми исследованиями с использованием метода Монте-Карло [например (MacKinnon, White (1985))], оценка Уайта представляется излишне оптимистичной; матрица оказывается слишком маленькой, поэтому асимптотические t -статистики завышены. Дэвидсон и Маккиннон (Davidson, MacKinnon (1993, p. 554)) предлагают следующие действия: (1) отнормировать результат на фактор $n/(n - K)$ и (2) использовать квадраты остатков, отнормированные на их истинную дисперсию e_i^2/m_{ii} вместо e_i^2 , где $m_{ii} = 1 - \mathbf{x}_i'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_i$ ¹⁵. [См. упражнение 9.6.b.] На основании этого исследования Дэвидсон и Маккиннон защищают ту или иную корректировку. Их предостережение «Никогда не используйте [оценку Уайта], так как [(2)] всегда работает лучше» звучит несколько сильно, но, по сути, правильно. Использование точных асимптотических результатов в маленьких выборках может быть проблематичным. Последние две строки табл. 9.1 показывают переоцененные стандартные ошибки с этими двумя модификациями.

Пример 9.2. Оценка Уайта

Результат использования оценки Уайта для регрессии в примере 9.1 представлен в колонке с названием «Стандартная ошибка Уайта» в табл. 9.1. Два коэффициента дохода совместно и индивидуально статистически значимы согласно t -статистикам и $F(2, 67) = [(0,244 - 0,064)/2]/[0,756/(72 - 5)] = 7,976$. Критическое значение на уровне значимости 1% равно 4,94.

¹³Дальнейшее обсуждение и некоторые детали можно найти в работе Крэга (Cragg (1982)). В ней показано, что наблюдение Уайта можно распространить и на оценку, которая улучшает эффективность МНК.

¹⁴MacKinnon, White (1985) и Messer, White (1984).

¹⁵Это стандартизированный остаток (4-61). Авторы также предлагают третью корректировку e_i^2/m_{ii} в качестве приближения к оценке, основанной на методе «складного ножа», но их аргументы в защиту этой оценки намного слабее, чем других двух.

Таблица 9.1. Результаты оценки регрессии методом наименьших квадратов

| | Константа | Age | OwnRent | Income | Income ² |
|----------------------|---------------------------------|---------|---------|--------|---------------------|
| Выборочное среднее | | 2,08 | 0,36 | 3,369 | |
| Коэффициент | -237,15 | -3,0818 | 27,941 | 234,35 | -14,997 |
| Стандартная ошибка | 199,35 | 5,5147 | 82,922 | 80,366 | 7,4693 |
| <i>t</i> -статистика | -1,19 | -0,5590 | 0,337 | 2,916 | -2,0080 |
| Стандартная ошибка | 212,99 | 3,3017 | 92,188 | 88,866 | 6,9446 |
| Уайта | | | | | |
| D. & M. (1) | 220,79 | 3,4227 | 95,566 | 92,122 | 7,1991 |
| D. & M. (2) | 221,09 | 3,4477 | 95,672 | 92,083 | 7,1995 |
| | $R^2 = 0,243578, s = 284,75080$ | | | | |

Средние расходы = 262,53 долл. Доход умножен на 10 000 долл. Тесты на гетероскедастичность: тест Уайта = 14,329, тест Бройша-Пэгана = 41,920, тест Коенкера-Бассета = 6,187.

Отличия в оцененных стандартных ошибках кажутся незначимыми в условиях сильной гетероскедастичности. Неожиданностью является уменьшение стандартной ошибки коэффициента при возрасте. *F*-тест уже неприменим для тестирования совместной значимости двух коэффициентов дохода из-за гетероскедастичности. Однако в этом случае может быть использован **тест Вальда (Wald)**. Хи-квадрат-тест основан на

$$W = (\mathbf{Rb})' [\mathbf{R}(\text{Est. Asy. Var}[\mathbf{b}])\mathbf{R}']^{-1} (\mathbf{Rb}), \quad \text{где } \mathbf{R} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

и оцененная асимптотическая ковариационная матрица представляет собой оценку Уайта. *F*-статистика, основанная на оценке методом наименьших квадратов, равна 7,976. Статистика Вальда, рассчитанная по оценке Уайта, составляет 20,604; критическое значение на уровне значимости 95% для распределения хи-квадрат с двумя степенями свободы составляет 5,99, т.е. результат не меняется.

9.5. Тесты на гетероскедастичность

Потенциально из-за гетероскедастичности возникает несколько проблем при тестировании гипотез с использованием оценок методом наименьших квадратов. В редких случаях мы можем быть точно уверены, что ошибки гетероскедастичные, и в случае, если это так, мы редко знаем, какую форму они имеют. В таком случае полезно уметь тестировать ошибки на гомоскедастичность и, если понадобится, модифицировать процедуру оценки

в соответствии с полученным результатом¹⁶. Для этих целей предлагаются несколько вариантов тестов. Они могут быть грубо поделены на группы в убывающем порядке для их общности и, как ожидается, в возрастающем порядке для их мощности¹⁷. Мы рассмотрим два наиболее часто используемых теста.

Тесты на гетероскедастичность основаны на следующей стратегии. Оценка методом наименьших квадратов является состоятельной для β даже при наличии гетероскедастичности. Если так, остатки метода наименьших квадратов будут повторяться, хотя и не полностью из-за случайности в конечной выборке, гетероскедастичность истинных ошибок. Следовательно, тесты, разработанные для обнаружения гетероскедастичности, будут применяться к остаткам метода наименьших квадратов.

9.5.1. Общий тест Уайта

Чтобы сформулировать большинство доступных тестов, необходимо специфицировать, по крайней мере грубо, природу гетероскедастичности. Желательно уметь проверять общую гипотезу в форме

$$H_0 : \sigma_i^2 = \sigma^2 \quad \text{для всех } i,$$

$$H_1 : H_0 \text{ неверна.}$$

В свете наших предыдущих результатов относительно сложности оценивания в модели с n неизвестными переменными эта задача выглядит довольно амбициозно. Тем не менее такой тест был предложен Уайтом (White (1980)). Правильная ковариационная матрица оценки методом наименьших квадратов выглядит следующим образом:

$$\text{Var}[\mathbf{b} | \mathbf{X}] = \sigma^2 [\mathbf{X}'\mathbf{X}]^{-1} [\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}] [\mathbf{X}'\mathbf{X}]^{-1},$$

она, как мы видели, может быть оценена с использованием (9-27). Обычная оценка имеет форму $\mathbf{V} = s^2 [\mathbf{X}'\mathbf{X}]^{-1}$. При отсутствии гетероскедастичности \mathbf{V} будет давать состоятельную оценку для $\text{Var}[\mathbf{b} | \mathbf{X}]$, в то время как при наличии гетероскедастичности оценка будет несостоятельна. Уайт разработал статистический тест, основанный на этом наблюдении. В простейшем случае тестовая статистика может быть получена как nR^2 в регрессии e_i^2 на константу, все переменные, содержащиеся в \mathbf{x} , и все квадраты, и перекрестные произведения переменных из \mathbf{x} . Статистика асимптотически распределена как хи-квадрат с $P - 1$ степенью свободы, где P — число регрессоров в уравнении, включая константу.

¹⁶ Существует вероятность того, что предварительный тест на гетероскедастичность ошибочно приведет нас к использованию взвешенного метода наименьших квадратов или не выявит гетероскедастичность и приведет нас к использованию обычного метода наименьших квадратов. Некоторые скромные данные о свойствах получаемых оценок приведены в работе (Ohtani, Toyoda (1980)). Их результаты говорят в пользу предварительного тестирования на гетероскедастичность, а не в пользу исходного предположения о ее присутствии.

¹⁷ Одно из исследований мощности нескольких тестов на гетероскедастичность было сделано в работе (Ali, Giaccotto (1984)).

Тест Уайта очень общий. Чтобы применять его, нам необходимо сделать некоторые специальные предположения относительно природы гетероскедастичности. Хотя эта характеристика виртуальная, в то же время она является потенциально серьезным недостатком. Тест может обнаружить гетероскедастичность, но вместо этого он может просто идентифицировать некоторые другие ошибки спецификации (такие как пропуск x^2 в простой регрессии)¹⁸. Если не учитывать контекст конкретной проблемы, мы не можем сказать много о мощности теста Уайта; его мощность может быть достаточно низкой относительно других альтернатив. В отличие от других тестов, которые мы обсудим, тест Уайта является **неконструктивным**. Если мы отвергаем нулевую гипотезу, тест не дает никаких рекомендаций относительно того, что делать дальше.

9.5.2. LM-тест Бройша–Пэгана/Годфри

Бройш и Пэган¹⁹ разработали **тест множителей Лагранжа** для проверки гипотезы $\sigma_i^2 = \sigma^2 f(\alpha_0 + \alpha' z_i)$, где z_i — вектор независимых переменных²⁰. Остатки в модели являются гомоскедастичными, если $\alpha = 0$. Тестовая статистика может быть получена из простой регрессии:

LM = $\frac{1}{2}$ объясненной суммы квадратов остатков регрессии $e_i^2/(e'e/n)$ на z_i .

Для вычислительных целей предположим, что Z есть $n \times P$ матрица наблюдений $(1, z_i)$, и пусть вектор g — вектор наблюдений $g_i = e_i^2/(e'e/n) - 1$. Тогда

$$LM = \frac{1}{2} [g'Z(Z'Z)^{-1}Z'g]. \quad (9-28)$$

В условиях нулевой гипотезы о гомоскедастичности остатков LM-статистика имеет хи-квадрат-распределение с числом степеней свободы, равным числу переменных в z_i . Этот тест может быть применен к большому количеству моделей, включая, например, модели, рассмотренные в примере 9.3 (2) и в подразделах 9.7.1 и 9.7.2²¹. Было доказано, что **тест множителей Лагранжа Бройша–Пэгана** чувствителен к предположению о нормальности остатков. Коэнкер (Koenker (1981)) и Коэнкер и Бассет (Koenker, Bassett (1982)) предложили метод расчета статистики множителей Лагранжа на основе более **устойчивой оценки** дисперсии ε_i^2 :

$$V = \frac{1}{n} \sum_{i=1}^n \left[e_i^2 - \frac{e'e}{n} \right]^2.$$

Дисперсия ε_i^2 необязательно равна дисперсии $2\sigma^4$, если ε_i не являются нормально распределенными. Пусть u равна $(e_1^2, e_2^2, \dots, e_n^2)$ и \mathbf{i} — столбец $n \times 1$,

¹⁸Thursby (1982) детально рассматривает этот случай.

¹⁹Breusch, Pagan (1979).

²⁰Тест множителей Лагранжа будет рассмотрен в подразделе 14.6.3.

²¹Модель $\sigma_i^2 = \sigma^2 \exp(\alpha' z_i)$ является одним из таких случаев. Анализируя эту модель, Харви (Harvey (1976)) вывел аналогичную тестовую статистику.

состоящий из единиц. Тогда $\bar{u} = e'e/n$. В этом случае мы получаем

$$LM = \left[\frac{1}{V} \right] (u - \bar{u}i)' Z(Z'Z)^{-1} Z'(u - \bar{u}i).$$

В условиях нормальности эта модифицированная статистика будет иметь то же предельное распределение, что и статистика Бройша–Пэгана, но есть некоторые свидетельства, что в отсутствие нормальности она дает более мощный тест. Уолдмэн (Waldman (1983)) показал, что если переменные в z_i те же самые, что были использованы для теста Уайта, описанного ранее, то с алгебраической точки зрения эти два теста одинаковы.

Пример 9.3. Тестирование гетероскедастичности

1. Тест Уайта: для данных, использованных в примере 9.1, мы имеем 15 переменных в $x \otimes x$, включая константу. Но так как $\text{Ownrent}^2 = \text{OwnRent}$ и $\text{Income} \times \text{Income} = \text{Income}^2$, только 13 из них являются уникальными. Регрессия квадратов остатков метода наименьших квадратов на эти 13 переменных дает $R^2 = 0,199013$. Следовательно, статистика хи-квадрат вычисляется как $72(0,199013) = 14,329$. 95%-е критическое значение распределения хи-квадрат с 12 степенями свободы равно 21,03, и, несмотря на очевидность, как может показаться из рис. 9.1, гипотеза о гомоскедастичности не отвергается.

2. Тест Бройша–Пэгана: для этого теста требуется специфицировать альтернативную гипотезу. Для этой цели мы определим тест на основе $z = [1, \text{Income}, \text{Income}^2]$. С использованием остатков метода наименьших квадратов мы рассчитываем $g_i = e_i^2 / (e'e/72) - 1$; тогда $LM = \frac{1}{2} g'Z(Z'Z)^{-1} Z'g$. Сумма квадратов остатков равна 5 432 562 033. После вычислений получим $LM = 41,920$. Критическое значение для хи-квадрат-распределения с двумя степенями свободы равно 5,99, так что гипотеза о гомоскедастичности отвергается. Вариант Коэнкера и Бассета этой статистики составляет 6,187, что все еще значительно, но гораздо меньше, чем LM-статистика. Большое различие между этими двумя статистиками наводит на мысль, что предположение о нормальности ошибочно. В отсутствие информации о гетероскедастичности мы можем использовать тесты на нормальность Харке и Бера (Bera, Jarque (1981, 1982)) и Кифера и Сэлмона (Kiefer, Salmon (1983))

$$\chi^2[2] = n[1/6(m_3/s^3)^2 + 1/25((m_4 - 3)/s^4)^2],$$

где $m_j = (1/n) \sum_i e_i^j$. В случае выполнения нулевой гипотезы о гомоскедастичности и нормального распределения ошибок эта статистика имеет хи-квадрат-распределение с двумя степенями свободы. Основанная на методе наименьших квадратов, она принимает значение 497,35, что определенно ведет к отвержению гипотезы. Однако здесь нужно быть осторожными. Непонятно, какая часть гипотезы должна быть отвергнута. У нас есть убедительное наблюдение, отраженное на рис. 9.1, что ошибки являются гетероскедастичными, так что предположение о гомоскедастичности в этом тесте ставится под сомнение. Это подтверждает

необходимость проверки данных перед использованием таких **тестов на спецификацию**, как этот.

9.6. Взвешенный метод наименьших квадратов

После проверки и обнаружения гетероскедастичности следующим логическим шагом является корректировка техники оценки регрессии. Оценка ОМНК имеет вид

$$\hat{\beta} = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}\mathbf{y}. \quad (9-29)$$

Рассмотрим наиболее общий случай $\text{Var}[\varepsilon_i | \mathbf{X}] = \sigma_i^2 = \sigma^2\omega_i$. Тогда Ω^{-1} — диагональная матрица, чей i -й элемент равен $1/\omega_i$. Оценка ОМНК получается из регрессии

$$\mathbf{Py} = \begin{bmatrix} y_1/\sqrt{\omega_1} \\ y_2/\sqrt{\omega_2} \\ \vdots \\ y_n/\sqrt{\omega_n} \end{bmatrix} \quad \text{на} \quad \mathbf{PX} = \begin{bmatrix} x'_1/\sqrt{\omega_1} \\ x'_2/\sqrt{\omega_2} \\ \vdots \\ x'_n/\sqrt{\omega_n} \end{bmatrix}.$$

Применяя метод наименьших квадратов к трансформированной модели, мы получаем оценку **взвешенного метода наименьших квадратов (ВМНК, WLS)**

$$\hat{\beta} = \left[\sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}'_i \right]^{-1} \left[\sum_{i=1}^n w_i \mathbf{x}_i y_i \right],$$

где $w_i = 1/\omega_i$ ²². Логика вычислений состоит в том, что наблюдения с меньшей дисперсией получают больший вес при вычислении сумм и, следовательно, имеют большее влияние в полученных оценках.

9.6.1. Взвешенный метод наименьших квадратов с известной матрицей Ω

Общий подход состоит в том, что дисперсия пропорциональна одному из регрессоров или его квадрату. Рассмотренный выше пример моделирования семейных расходов один из тех, в которых значимой переменной обычно является доход. Также в исследованиях прибыли фирм за основную переменную часто берется их размер. Если

$$\sigma_i^2 = \sigma^2 x_{ik}^2,$$

то трансформированной регрессионной моделью для ОМНК является

$$\frac{y}{x_k} = \beta_k + \beta_1 \left(\frac{x_1}{x_k} \right) + \beta_2 \left(\frac{x_2}{x_k} \right) + \dots + \frac{\varepsilon}{x_k}. \quad (9-30)$$

²²Веса обычно обозначаются $w_i = 1/\sigma_i^2$. Это выражение эквивалентно выражению $\hat{\beta} = [\mathbf{X}'(\sigma^2\Omega)^{-1}\mathbf{X}]^{-1}\mathbf{X}'(\sigma^2\Omega)^{-1}\mathbf{y}$. Все σ^2 сокращаются и дают предыдущее выражение.

Если переменная пропорциональна x_k вместо x_k^2 , то вес для каждого наблюдения должен быть $1/\sqrt{x_k}$ вместо $1/x_k$.

В (9-30) коэффициент при x_k становится константой. Но если дисперсия пропорциональна любой степени x_k , кроме двух, то преобразованная модель больше не будет содержать константу и мы сталкиваемся с проблемой интерпретации R^2 , упомянутой ранее. Например, мы не сможем сделать никаких выводов, если R^2 в регрессии y/z на $1/z$ и x/z выше, чем в регрессии y на константу и x для любого z , включая x . Хорошая подгонка взвешенной регрессии может получиться вследствие наличия $1/z$ в обеих частях уравнения.

Мы нечасто можем быть уверены относительно природы гетероскедастичности в регрессионной модели. В одном отношении проблема незначительна. Оценка взвешенным методом наименьших квадратов

$$\hat{\beta} = \left[\sum_{i=1}^n w_i \mathbf{x}_i \mathbf{x}_i' \right]^{-1} \left[\sum_{i=1}^n w_i \mathbf{x}_i y_i \right]$$

состоятельна относительно использованных весов, пока веса не коррелированы с ошибками.

Но использование неправильного набора весов имеет два других последствия, которые могут быть менее благоприятными. Во-первых, оценка неправильно взвешенным методом наименьших квадратов неэффективна. Этот вопрос может быть спорным, если правильные веса неизвестны, но стандартная ошибка ОМНК также будет неправильной. Асимптотическая ковариационная матрица оценки

$$\hat{\beta} = [\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \quad (9-31)$$

выглядит следующим образом:

$$\text{Asy. Var}[\hat{\beta}] = \sigma^2[\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{\Omega}\mathbf{V}^{-1}\mathbf{X}[\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1}. \quad (9-32)$$

Этот результат может как напоминать, так и не напоминать обычную оценку, которая будет матрицей в скобках и недооценит полезность оценки Уайта в (9-27).

Стандартный подход в литературе состоит в использовании МНК с оценкой Уайта или некоторого ее варианта для асимптотической ковариационной матрицы. Разумеется, можно оспорить как недостатки, так и сильные стороны этого подхода. В его пользу говорит **устойчивость к неизвестной гетероскедастичности**, что является сильным достоинством. При явном наличии гетероскедастичности, однако, метод наименьших квадратов может быть крайне неэффективен. Вопрос состоит в том, является ли использование неправильных весов лучшей альтернативой, чем отсутствие таковых. В этом вопросе есть несколько моментов. Если мы используем одну из моделей, упомянутых ранее, — например, модель Харви является универсальным кандидатом — то можем использовать неправильный набор весов, в дополнение оценка параметров дисперсии добавляет новый источник вариации в масштабе оценок модели. Устойчивая к гетероскедастичности

оценка взвешенным методом наименьших квадратов может быть сформирована посредством комбинирования (9-32) с оценкой Уайта. Оценка взвешенным методом наименьших квадратов в (9-31) является устойчивой с любым набором весов: $\mathbf{V} = \text{diag}[v_1, v_2, \dots, v_n]$. Ее асимптотическая ковариационная матрица может быть оценена как

$$\text{Est. Asy. Var}[\hat{\beta}] = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \left[\sum_{i=1}^n \left(\frac{e_i^2}{v_i^2} \right) \mathbf{x}_i \mathbf{x}_i' \right] (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}. \quad (9-33)$$

Любая устойчивая оценка может быть использована для расчета остатков. И оценка взвешенным методом наименьших квадратов является вполне естественным кандидатом.

9.6.2. Оценка при условии, что матрица Ω содержит неизвестные параметры

Общая форма регрессионной модели с гетероскедастичностью имеет слишком много параметров для оценивания стандартными методами. Обычно модель ограничена специфицированием $\sigma^2\Omega$ как функции от нескольких переменных, как в $\sigma_i^2 = \sigma^2 x_i^\alpha$ или $\sigma_i^2 = \sigma^2 (\mathbf{x}_i' \boldsymbol{\alpha})^2$. Запишем эти условия как $\Omega(\boldsymbol{\alpha})$. ДОМНК (FWLS), основанный на устойчивой оценке $\Omega(\boldsymbol{\alpha})$ (имеется в виду устойчивая оценка $\boldsymbol{\alpha}$), асимптотически эквивалентен полному ОМНК (WLS). Новая проблема состоит в том, что мы должны сначала найти устойчивые оценки неизвестных параметров в матрице $\Omega(\boldsymbol{\alpha})$. Чаще всего используются два метода: двухшаговый ОМНК и метод максимального правдоподобия. Мы рассмотрим двухшаговый ОМНК здесь, а метод максимального правдоподобия — в главе 14.

Для модели с гетероскедастичностью оценка ОМНК имеет вид

$$\hat{\beta} = \left[\sum_{i=1}^n \left(\frac{1}{\sigma_i^2} \right) \mathbf{x}_i \mathbf{x}_i' \right]^{-1} \left[\sum_{i=1}^n \left(\frac{1}{\sigma_i^2} \right) \mathbf{x}_i y_i \right]. \quad (9-34)$$

Двухшаговая оценка рассчитывается при помощи исходных оценок $\hat{\sigma}_i^2$, обычно получаемых с использованием некоторой функции остатков при помощи метода наименьших квадратов. Затем $\hat{\beta}$ использует (9-34) и $\hat{\sigma}_i^2$. Оценка методом наименьших квадратов β хоть и не является неэффективной, но все еще состоятельна. Если так, статистики, рассчитанные с использованием остатков метода наименьших квадратов, $e_i = (y_i - \mathbf{x}_i' \mathbf{b})$, будут иметь те же асимптотические свойства, что и остатки, рассчитанные с использованием истинных ошибок $\varepsilon_i = (y_i - \mathbf{x}_i' \boldsymbol{\beta})$. Этот результат предлагает регрессионный подход для истинных ошибок и переменных \mathbf{z}_i , которые могут как совпадать, так и не совпадать с \mathbf{x}_i . Вспомним, что $E[\varepsilon_i^2 | \mathbf{z}_i] = \sigma_i^2$, поэтому

$$\varepsilon_i^2 = \sigma_i^2 + v_i,$$

где v_i просто разность между ε_i^2 и его условным математическим ожиданием. Так как ε_i ненаблюдаемы, мы можем использовать остатки метода наименьших квадратов, для которых $e_i = \varepsilon_i - \mathbf{x}_i' (\mathbf{b} - \boldsymbol{\beta}) = \varepsilon_i + u_i$. Тогда

$e_i^2 = \varepsilon_i^2 + u_i^2 + 2\varepsilon_i u_i$. Но в больших выборках, так как $\mathbf{b} \xrightarrow{p} \beta$, компоненты u_i станут малы, так что, по крайней мере, приблизительно²³

$$e_i^2 = \sigma_i^2 + v_i^*.$$

Предложенная процедура состоит в рассмотрении функции дисперсии как регрессии и использовании квадратов или других функций остатков метода наименьших квадратов в качестве зависимых переменных²⁴. Например, если $\sigma_i^2 = \mathbf{z}_i' \alpha$, то состоятельная оценка для α представляет собой коэффициенты наклона метода наименьших квадратов \mathbf{a} в «модели»

$$e_i^2 = \mathbf{z}_i' \alpha + v_i^*.$$

В этой модели ошибки v_i^* являются и гетероскедастичными, и автокоррелированными, поэтому \mathbf{a} состоятельная, но неэффективная оценка. Но состоятельность — это все, что требуется для получения асимптотически эффективной оценки β с использованием матрицы $\Omega(\hat{\alpha})$. Остается установить, изменит ли улучшение оценки α в этой и других моделях, которые мы рассмотрим, свойства двухшаговой оценки β в маленьких выборках в лучшую сторону²⁵.

Двухшаговая оценка может быть получена пересчетом остатков после получения ДОМНК-оценок и повторного вычисления. Однако асимптотические свойства итерационной оценки совпадают с соответствующими свойствами двухшаговой оценки. В некоторых случаях такие итерации дают в пределе оценку максимального правдоподобия. Еще ни одна из оценок, основанных на регрессии квадратов ошибок на другие переменные, не удовлетворила требованию. Таким образом, в этом контексте итерационная процедура демонстрирует небольшое улучшение, если таковое вообще имеет место.

9.7. Приложения

В этом разделе рассматриваются два общих приложения регрессионной модели с гетероскедастичностью, модели Харви мультипликативной гетероскедастичности и модели групповой гетероскедастичности, которая добавляет к дисперсии некоторые члены, обычно связанные с изменениями в регрессионной функции.

9.7.1. Мультипликативная гетероскедастичность

Модель Харви (1976) мультипликативной гетероскедастичности — очень гибкая общая модель, которая включает большинство полезных моделей как частные случаи. Общий вид модели

$$\sigma_i^2 = \sigma^2 \exp(\mathbf{z}_i' \alpha).$$

²³Amemiya (1985) и Harvey (1976).

²⁴Jobson, Fuller (1980).

²⁵Fomby, Hill, Johnson (1984, p. 177–186) и Amemiya (1985, p. 203–207; 1977a) рассматривают такую модель.

Модель с гетероскедастичностью в форме

$$\sigma_i^2 = \sigma^2 \prod_{m=1}^M z_{im}^{\alpha_m}$$

имеет место, если логарифмы переменных содержатся в z_i . Модель групповой гетероскедастичности, описанная в примере 9.4, получена при помощи формирования z_i как набора групповых бинарных переменных (одна из них должна быть пропущена). В этом случае σ^2 является дисперсией ошибок для базовой группы, в то время как для других групп $\sigma_g^2 = \sigma^2 \exp(\alpha_g)$.

Пример 9.4. Мультипликативная гетероскедастичность

В примере 6.4 мы подобрали функцию затрат для авиаиндустрии США в виде

$$\ln C_{it} = \beta_1 + \beta_2 \ln Q_{it} + \beta_3 (\ln Q_{it})^2 + \beta_4 \ln P_{fuel,i,t} + \beta_5 Loadfactor_{i,t} + \varepsilon_{i,t},$$

где $C_{i,t}$ — общие затраты, $Q_{i,t}$ — выпуск, $P_{fuel,i,t}$ — цена топлива и 90 наблюдений в наборе данных соответствуют выборке по шести фирмам за 15 лет. (Модель также включает бинарные переменные для фирмы и года, которые мы опустили для простоты.) Теперь рассмотрим измененную модель, в которой фактор загруженности отражается в дисперсии $\varepsilon_{i,t}$, а не в регрессионной функции. Модель имеет вид

$$\sigma_{i,t}^2 = \sigma^2 \exp(\gamma Loadfactor_{i,t}) = \exp(\gamma_1 + \gamma_2 Loadfactor_{i,t}).$$

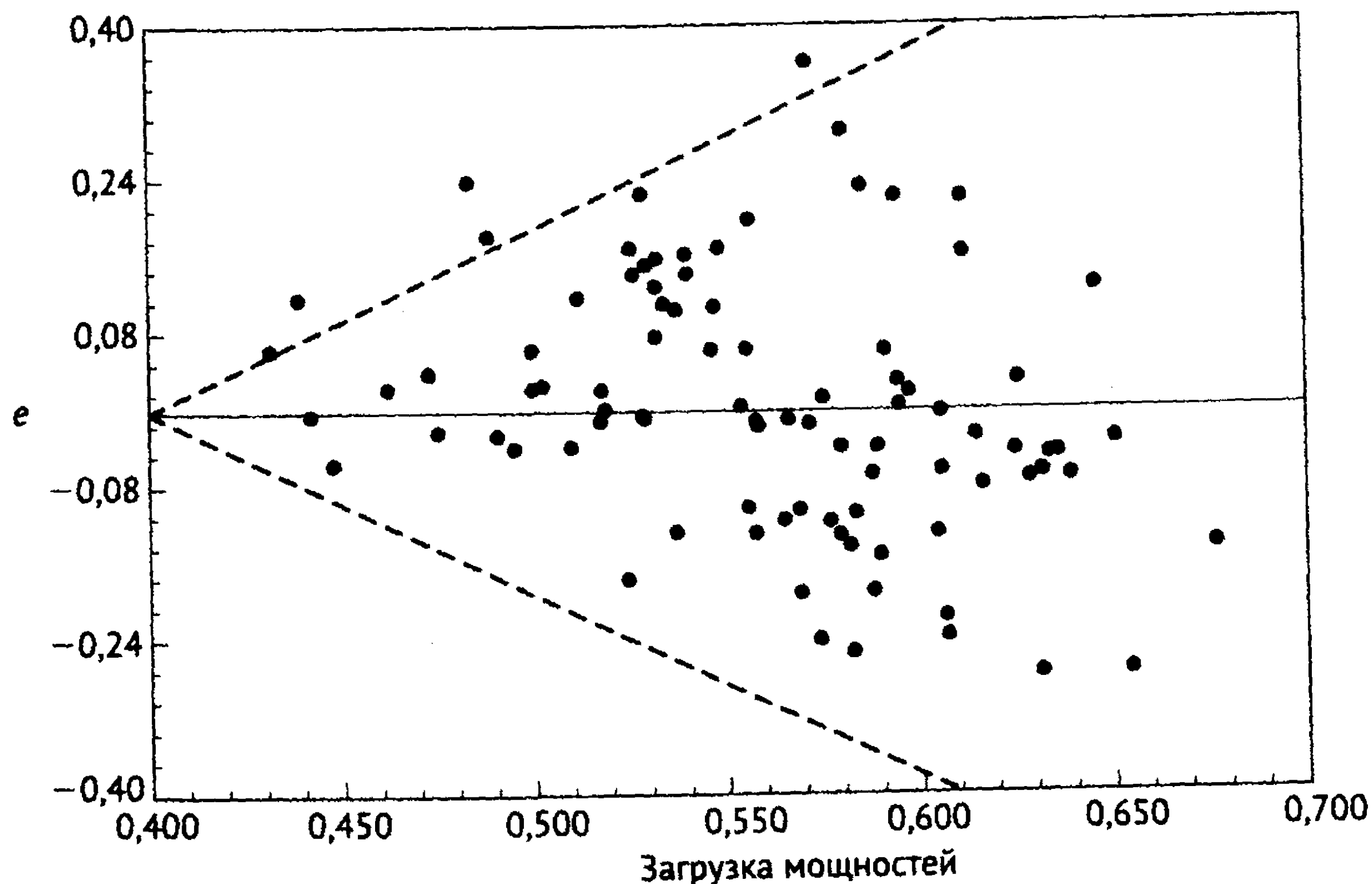


Рис. 9. 2. График зависимости остатков от входного фактора

Константой в предполагаемой регрессии является $\gamma_1 = \ln \sigma^2$. Рисунок 9.2 показывает зависимость остатков метода наименьших квадратов от переменной *Loadfactor* (загрузка мощностей) для 90 наблюдений. Рисунок

предполагает наличие гетероскедастичности. (Пунктирная линия подчеркивает наличие этого эффекта.) Мы рассчитали статистику множителей Лагранжа с использованием (9-28). Статистика хи-квадрат равна 2,959. Это меньше, чем критическое значение 3,84 с одной степенью свободы, на основании чего нулевая гипотеза о гомоскедастичности по отношению к фактору загрузки мощностей не отвергается.

Сначала мы используем МНК для оценки параметров функции затрат и набора остатков $e_{i,t}$. Регрессия $\log(e_{it}^2)$ на константу и входной фактор дает оценки для γ_1 и γ_2 , обозначенные c_1 и c_2 . Результаты приведены в табл. 9.2.

Таблица 9.2. Мультипликативная модель гетероскедастичности

| | Константа | Ln Q | Ln ² Q | Ln P _f | R ² | Сумма квад- ратов |
|----------------------------|----------------------|----------|-------------------|-------------------|------------------------|-------------------------|
| МНК | 9,1382 | 0,92615 | 0,029145 | 0,41006 | | |
| | 0,24507 ^a | 0,032306 | 0,012304 | 0,018807 | 0,9861674 ^c | 1,577479 ^d |
| | 0,22595 ^b | 0,030128 | 0,011346 | 0,017524 | | |
| Двух- шаговый метод | 9,2463 | 0,92136 | 0,024450 | 0,40352 | 0,986119 | 1,612938 |
| | 0,21896 | 0,033028 | 0,011412 | 0,016974 | | |
| Ите- рации ^e | 9,2774 | 0,91609 | 0,021643 | 0,40174 | 0,986071 | 1,645693 |
| | 0,20977 | 0,032993 | 0,011017 | 0,016332 | | |

^aОбычные стандартные ошибки МНК

^bУстойчивая оценка Уайта

^c Квадрат корреляции между действительными и модельными значениями

^dСумма квадратов ошибок

^eЗначения c_2 при разных итерациях: 8,254344, 11,622473, 11,705029, 11,710618, 11,711012, 11,711040, 11,11042

Как заметил Харви, $\exp(c_1)$ необязательно состоятельно оценивает σ^2 — для нормально распределенных ошибок этот показатель ниже в 1,2704 раза. Однако, как показано в (9-29), для оценки σ^2 (смещенной или нет) нужно вычислить оценку ДОМНК. Веса $w_{i,t} = \exp(-c_1 - c_2 \text{Loadfactor}_{i,t})$ рассчитываются с использованием этих оценок, затем взвешенный МНК из (9-20) используется для получения ДОМНК-оценок β . Результаты вычислений приведены в табл. 9.2.

Мы можем рассмотреть итерационную процедуру. Используя результаты ДОМНК на втором шаге, можно пересчитать c_1 и c_2 и веса, а затем повторить итерацию. Процесс сходится, когда оценка c_2 стабилизируется. Для этого потребуется семь итераций. Результаты приведены табл. 9.2. Как отмечено ранее, итерация здесь не дает никакого выигрыша. Оценка на втором шаге уже полностью эффективна. Более того, это также не дает оценки методом максимального правдоподобия. Она будет получена при оценке регрессии $[e_{i,t}^2 / \exp(c_1 + c_2 \text{Loadfactor}_{i,t}) - 1]$ на константу и загрузку мощностей на каждой итерации для получения новых оценок. Мы вернемся к этому примеру в главе 14.

9.7.2. Групповая гетероскедастичность

Регрессия с групповой гетероскедастичностью состоит из структурных уравнений

$$y_i = \mathbf{x}'_i \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \dots, n, \quad E[\varepsilon_i | \mathbf{x}_i] = 0, \quad i = 1, \dots, n.$$

n наблюдений сгруппированы в G групп, каждая из которых содержит n_g наблюдений. Вектор коэффициентов наклона одинаков во всех группах, но в группе g

$$\text{Var}[\varepsilon_{ig} | \mathbf{x}_{ig}] = \sigma_g^2, \quad i = 1, \dots, n_g.$$

Если дисперсии известны, оценка ОМНК имеет вид

$$\hat{\boldsymbol{\beta}} = \left[\sum_{g=1}^G \left(\frac{1}{\sigma_g^2} \right) \mathbf{X}'_g \mathbf{X}_g \right]^{-1} \left[\sum_{g=1}^G \left(\frac{1}{\sigma_g^2} \right) \mathbf{X}'_g \mathbf{y}_g \right]. \quad (9-35)$$

Так как $\mathbf{X}'_g \mathbf{y}_g = \mathbf{X}'_g \mathbf{X}_g \mathbf{b}_g$, где \mathbf{b}_g — оценка МНК, в g -й подгруппе наблюдений

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \left[\sum_{g=1}^G \left(\frac{1}{\sigma_g^2} \right) \mathbf{X}'_g \mathbf{X}_g \right]^{-1} \left[\sum_{g=1}^G \left(\frac{1}{\sigma_g^2} \right) \mathbf{X}'_g \mathbf{X}_g \mathbf{b}_g \right] = \left[\sum_{g=1}^G \mathbf{V}_g \right]^{-1} \left[\sum_{g=1}^G \mathbf{V}_g \mathbf{b}_g \right] = \\ &= \sum_{g=1}^G \mathbf{W}_g \mathbf{b}_g. \end{aligned}$$

Этот результат является матрицей взвешенных средних по G -оценкам МНК. Взвешенная матрица имеет вид

$$\mathbf{W}_g = \left[\sum_{g=1}^G (\text{Var}[\mathbf{b}_g])^{-1} \right]^{-1} (\text{Var}[\mathbf{b}_g])^{-1}.$$

Оценка с меньшей ковариационной матрицей, следовательно, получает больший вес. (Если \mathbf{X}_g одинакова в каждой группе, то матрица \mathbf{W}_g сводится к более простой, $w_g \mathbf{I} = (h_g / \sum_g h_g) \mathbf{I}$, где $h_g = 1/\sigma_g^2$.)

Рассмотренный выше метод представляет полезный способ вычисления оценки, но опирается на алгебраический результат, который может оказаться нереализуемым. Если количество наблюдений в любой группе меньше, чем количество регрессоров, то специфичная для группы оценка МНК не сможет быть рассчитана. Но, как можно увидеть из (9-35), это не то, что нам нужно; то, что нам нужно, — это веса. Как всегда, оценка МНК сквозной регрессии является состоятельной, это означает, что использование специфичных для группы подвекторов остатков МНК,

$$\hat{\sigma}_g^2 = \frac{\mathbf{e}'_g \mathbf{e}_g}{n_g}, \quad (9-36)$$

дает необходимую специфичную оценку для группы дисперсии ошибок. Соответственно (9-35) является оценкой и обратная матрица в этом выражении дает оценку асимптотической ковариационной матрицы.

Продолжая эту линию рассуждений, можно рассмотреть итерационную процедуру для оценки, вернувшись к (9-36) с двухшаговой оценкой ДОМНК, пересчитать веса, затем вернуться к (9-35), чтобы пересчитать вектор коэффициентов наклона. Эта процедура может быть продолжена до сходимости. Можно показать [см. Oberhofer, Kmenta (1974)], что пока (9-36) используется без коррекции на степени свободы, то, если выражение сходится, оно сходится и в оценке максимального правдоподобия (с нормально распределенными остатками).

Для тестирования предположения о гомоскедастичности можно довольно легко применить и тест Уайта, и тест множителей Лагранжа. Переменные, которые должны были войти в условную дисперсию, представляют собой простой набор из $G - 1$ групповой бинарной переменной, за исключением одной из них (чтобы исключить ловушку бинарных переменных), которую мы обозначим \mathbf{Z}^* . Так как столбцы \mathbf{Z}^* двумерные и ортогональные, для того чтобы применить тест Уайта, нам нужно всего лишь построить регрессию квадрата остатков метода наименьших квадратов на константу и \mathbf{Z}^* и посчитать NR^2 , где $N = \sum_g n_g$. Тест множителей Лагранжа также легкоосуществим. Для того чтобы его применить, заменим константу в \mathbf{Z} в (9-28) на оставшуюся групповую бинарную переменную. Так как пространство столбцов полного набора бинарных переменных такое же, как столбец констант и $G - 1$ бинарная переменная, все результаты останутся без изменений. В (9-28) \mathbf{g} теперь будет G -подвекторами, где каждый подвектор содержит n_g элементов $[(e_{ig}^2/\hat{\sigma}^2) - 1]$ и $\hat{\sigma}^2 = \mathbf{e}'\mathbf{e}/N$. Перемножив их, находим, что $\mathbf{g}'\mathbf{Z}$ — это G -вектор с элементами $n_g[(\hat{\sigma}_g^2/\hat{\sigma}^2) - 1]$, в то время как $(\mathbf{Z}'\mathbf{Z})^{-1} - G \times G$ матрица с диагональными элементами $1/n_g$. Отсюда следует, что

$$LM = \frac{1}{2} \mathbf{g}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{g} = \frac{1}{2} \sum_{g=1}^G n_g \left(\frac{\hat{\sigma}_g^2}{\hat{\sigma}^2} - 1 \right)^2. \quad (9-37)$$

В условиях нулевой гипотезы о гомоскедастичности обе статистики имеют предельное хи-квадрат-распределение с $G - 1$ степенью свободы. (Наблюдается только $G - 1$ степень свободы, так как гипотеза налагает $G - 1$ ограничений, что G переменных равны друг другу. Неявно одна из переменных является свободной, а другие $G - 1$ равны единице.)

Пример 9.5. Групповая гетероскедастичность

Балтаги и Гриффин (Baltagi, Griffin (1983)) исследовали использование бензина в 18 странах ОЭСР из 30. Модель, рассмотренная в этой работе, имеет вид

$$\begin{aligned} \ln(\text{Gasoline usage/car})_{i,t} &= \beta_1 + \beta_2 \ln(\text{Per capita income})_{i,t} + \\ &+ \beta_3 \ln \text{Price}_{i,t} + \beta_4 \ln(\text{Cars per capita})_{i,t} + \varepsilon_{i,t}, \end{aligned}$$

где i = страна и t = 1960, ..., 1978. Это сбалансированная панель (см. раздел 11.2) с $19 \times 18 = 342$ наблюдениями в сумме. Данные приведены в табл. F9.2 приложения.

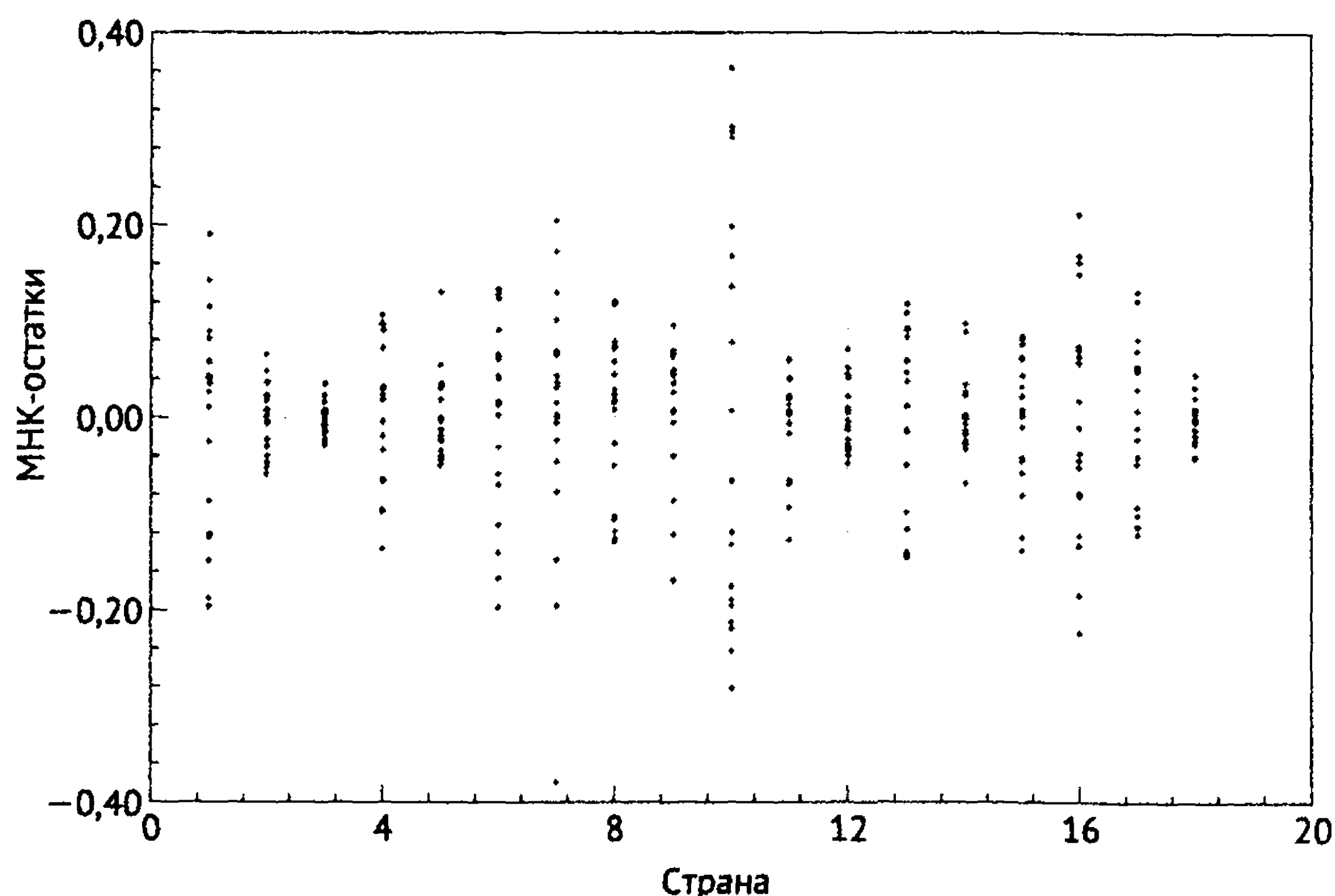


Рис. 9.3. График остатков МНК регрессии по странам

На рис. 9.3 изображены остатки регрессии, оцененной с использованием метода наименьших квадратов, рассмотренной выше в модели с дополнительными бинарными переменными по 18 странам (с 1-й по 18-ю) (и без константы). (Бинарные переменные по странам используются таким образом, что специфичные по странам остатки будут иметь в среднем ноль.) F -статистика для тестирования нулевой гипотезы о том, что все константы равны между собой, имеет вид

$$\begin{aligned}
 F[(G-1), (\sum_g n_g - K - G)] &= \frac{(e_0' e_0 - e_1' e_1) / (G-1)}{(e_1' e_1 / \sum_g n_g - K - G)} \\
 &= \frac{(14,90436 - 2,73649) / 17}{2,73649 / (342 - 3 - 18)} = 83,960798,
 \end{aligned}$$

где e_0 — вектор остатков в регрессии с одним свободным членом и e_1 — регрессия со специфичными для каждой страны свободными членами (т.е. со страновыми дамми-переменными). Критическое значение из таблицы для F -распределения с 17 и 321 степенью свободы равно 1,655. Результаты оценки регрессии представлены в табл. 9.3. Рисунок 9.3 говорит о наличии групповой гетероскедастичности. Статистики Уайта и метода множителей Лагранжа равны соответственно $342 \times (0,38365) = 131,21$ и 279588. Критическое значение для хи-квадрат-распределения с 17 степенями свободы равно 27,587, т.е. мы отвергаем гипотезу о гомоскедастичности и продолжаем оценивать модель при помощи ДОМНК. Двухшаговые оценки также приведены в табл. 9.3. Оценка ДОМНК рассчитана с использованием взвешенного метода наименьших квадратов с весами, равными $1/\hat{\sigma}_g^2$ для каждого наблюдения в стране g . Сравнивая стандартные ошибки Уайта и двухшаговых оценок, мы видим, что в этом случае появляется существенный выигрыш при использовании ДОМНК.

Таблица 9.3. Оценки модели потребления бензина

| | OLS | | | FGLS | |
|-----------------|-------------|---------------|------------------------|-------------|---------------|
| | Коэффициент | Станд. ошибка | Станд. ошибка по Уайту | Коэффициент | Станд. ошибка |
| Доход | 0,66225 | 0,07339 | 0,07277 | 0,57507 | 0,02927 |
| Цена | -0,32170 | 0,04410 | 0,05381 | -0,27967 | 0,03519 |
| Автомобиль/чел. | -0,64048 | 0,02968 | 0,03876 | -0,56540 | 0,01613 |
| Страна 1 | 2,28586 | 0,22832 | 0,22608 | 2,43707 | 0,11308 |
| Страна 2 | 2,16555 | 0,21290 | 0,20983 | 2,31699 | 0,10225 |
| Страна 3 | 3,04184 | 0,21864 | 0,22479 | 3,20652 | 0,11663 |
| Страна 4 | 2,38946 | 0,20809 | 0,20783 | 2,54707 | 0,10250 |
| Страна 5 | 2,20477 | 0,21647 | 0,21087 | 2,33862 | 0,10101 |
| Страна 6 | 2,14987 | 0,21788 | 0,21846 | 2,30066 | 0,10893 |
| Страна 7 | 2,33711 | 0,21488 | 0,21801 | 2,57209 | 0,11206 |
| Страна 8 | 2,59233 | 0,24369 | 0,23470 | 2,72376 | 0,11384 |
| Страна 9 | 2,23255 | 0,23954 | 0,22973 | 2,34805 | 0,10795 |
| Страна 10 | 2,37593 | 0,21184 | 0,22643 | 2,58988 | 0,11821 |
| Страна 11 | 2,23479 | 0,21417 | 0,21311 | 2,39619 | 0,10478 |
| Страна 12 | 2,21670 | 0,20304 | 0,20300 | 2,38486 | 0,09950 |
| Страна 13 | 1,68178 | 0,16246 | 0,17133 | 1,90306 | 0,08146 |
| Страна 14 | 3,02634 | 0,39451 | 0,39180 | 3,07825 | 0,20407 |
| Страна 15 | 2,40250 | 0,22909 | 0,23280 | 2,56490 | 0,11895 |
| Страна 16 | 2,50999 | 0,23566 | 0,26168 | 2,82345 | 0,13326 |
| Страна 17 | 2,34545 | 0,22728 | 0,22322 | 2,48214 | 0,10955 |
| Страна 18 | 3,05525 | 0,21960 | 0,22705 | 3,21519 | 0,11917 |

9.8. Заключение и выводы

В этой главе представлены основные расширения классической линейной модели. Добавив гетероскедастичность и автокорреляцию в ошибках, мы расширили модель на большое число случаев. Мы рассмотрим их в следующих нескольких главах. Формальные понятия, представленные в этой главе, включают ответы на следующие вопросы: как это расширение влияет на свойства оценки МНК? как нужная нам оценка асимптотической ковариационной матрицы оценки МНК может быть рассчитана в рамках расширенной модели? как использовать информацию о дисперсиях и ковариаци-

ях остатков, чтобы получить оценку, которая более эффективна, чем оценка МНК?

Мы детально проанализировали одну из форм обобщенной регрессионной модели — модель с гетероскедастичностью. Сначала мы рассмотрели оценку методом наименьших квадратов. Исходный результат для оценки методом наименьших квадратов состоял в том, что она сохраняет состоятельность и асимптотическую нормальность, однако для тестирования гипотез необходима некоторая коррекция оцененной асимптотической ковариационной матрицы. Оценка Уайта — это стандартный подход для таких вычислений. После знакомства с двумя общими тестами на гетероскедастичность мы сузили модель до некоторой специфичной параметрической формы и рассмотрели взвешенный (обобщенный) метод наименьших квадратов для эффективного оценивания и оценивания методом максимального правдоподобия. Если форма гетероскедастичности известна, но включает неизвестные параметры, все еще непонятно, является ли коррекция ДОМНК улучшением по сравнению с МНК. Асимптотически сравнение вполне понятно, но в малых или средних выборках дополнительная случайность из-за оцененных параметров дисперсии может свидетельствовать в пользу МНК.

Основные термины и понятия

Aitken's theorem — Теорема Айткена

Asymptotic properties — Асимптотические свойства

Autocorrelation — Автокорреляция

Breusch-Pagan Lagrange multiplier test — Тест множителей Лагранжа Бройша-Пэгана

Efficient estimator — Эффективная оценка

Feasible generalized least squares (FGLS) — Доступный обобщенный метод наименьших квадратов (ДОМНК)

Finite-sample properties — Свойства в конечных выборках

Generalized least squares (GLS) — Обобщенный метод наименьших квадратов (ОМНК)

Generalized linear regression model — Обобщенная линейная регрессионная модель

Generalized sum of squares — Обобщенная сумма квадратов

Groupwise heteroscedasticity — Групповая гетероскедастичность

Heteroscedasticity — Гетероскедастичность

Kruskal's theorem — Теорема Крускала

Lagrange multiplier test — Тест множителей Лагранжа

Multiplicative heteroscedasticity — Мультипликативная гетероскедастичность

Nonconstructive test — Неконструктивный тест

Ordinary least squares (OLS) — Метод наименьших квадратов (МНК)

Panel data — Панельные данные

Parametric model — Параметрическая модель

Robust estimation — Устойчивый метод оценивания

Robust estimator — Устойчивая оценка

Robustness to unknown heteroscedasticity — Устойчивость к неизвестной гетероскедастичности

Semiparametric model — Полупараметрическая модель

Specification test — Тест на спецификацию

Spherical disturbances — Сферические ошибки

Two-step estimator — Двухшаговая оценка

Wald test — Тест Вальда

Weighted least squares (WLS) — Взве-

шенный метод наименьших квадратов **estimator** — Ошибки в форме Уайта (ВМНК)

White heteroscedasticity consistent White test — Тест Уайта

Упражнения

1. Чему равна ковариационная матрица $\text{Cov}[\hat{\beta}, \hat{\beta} - \mathbf{b}]$ оценки ОМНК $\hat{\beta} = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}\mathbf{y}$, а также разность между последней оценкой и оценкой МНК $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$? Результат играет ключевую роль при выводе спецификационного теста Хаусмана (Hausman, 1978).
2. Это и два следующих упражнения основаны на тестовой статистике, обычно используемой для тестирования набора из J линейных ограничений в обобщенной регрессионной модели:

$$F[J, n - K] = \frac{(\mathbf{R}\hat{\beta} - \mathbf{q})'[\mathbf{R}(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\hat{\beta} - \mathbf{q})/J}{(\mathbf{y} - \mathbf{X}\hat{\beta})'\Omega^{-1}(\mathbf{y} - \mathbf{X}\hat{\beta})/(n - K)},$$

где $\hat{\beta}$ — оценка ОМНК. Покажите, что если Ω известна, ошибки нормально распределены и нулевая гипотеза $\mathbf{R}\beta = \mathbf{q}$ верна, то эта статистика распределена как F с J и $n - K$ степенями свободы. Какие предположения о регрессорах необходимы для получения такого результата? Должны ли они быть нестохастическими?

3. Теперь предположим, что ошибки распределены ненормально, хотя Ω все еще известна. Покажите, что предельное распределение предыдущей статистики равно $(1/J)$, умноженной на хи-квадрат-переменную с J степенями свободы. (Подсказка: знаменатель сходится к σ^2 .) Сделайте вывод, что в обобщенной регрессионной модели предельное распределение статистики Вальда

$$W = (\mathbf{R}\hat{\beta} - \mathbf{q})' \{ \mathbf{R}(\text{Est. Var}[\hat{\beta}])\mathbf{R}' \}^{-1} (\mathbf{R}\hat{\beta} - \mathbf{q})$$

есть хи-квадрат-распределение с J степенями свободы независимо от распределения остатков до тех пор, пока данные являются доброкачественными. Заметьте, что в конечных выборках истинное распределение может быть приближено $F[J, n - K]$ -распределением. Однако несколько сомнительно интерпретировать этот факт как утверждение о том, что статистика асимптотически распределена как F с J и $n - K$ степенями свободы, так как предельное распределение, которое использовалось для вывода нашего результата, было хи-квадрат-распределением, а не F . В этом случае $F[J, n - K]$ является случайной переменной, которая асимптотически сходится к случайной величине хи-квадрат.

4. В заключение предположим, что Ω должна быть оценена, но оценка удовлетворяет предположениям (9-16) и (9-17). Какие изменения необходимо внести в предыдущую задачу?
5. В обобщенной регрессионной модели, если K столбцов \mathbf{X} являются характеристическими векторами Ω , оценки МНК и ОМНК совпадают. (На самом деле результат немного шире. \mathbf{X} может быть любой линейной комбинацией K характеристических векторов. Этот результат называется **теоремой Крускала**.)

- a. Докажите этот результат, используя матричную алгебру.
 - b. Докажите, что если \mathbf{X} содержит свободный член и остальные столбцы рассматриваются в форме отклонений (т.е. сумма столбцов равна нулю), то модель из упражнения 8 является примером одного из таких случаев. (Модель кажущихся несвязанными регрессий с той же матрицей регрессоров, рассмотренная в главе 10, является другой.)
6. Предположим, что в обобщенной регрессионной модели Ω известна.
- a. Чему равна ковариационная матрица МНК и ОМНК-оценок β ?
 - b. Чему равна ковариационная матрица вектора остатков МНК $e = y - \mathbf{X}\hat{\beta}$?
 - c. Чему равна ковариационная матрица вектора остатков ОМНК $\hat{\varepsilon} = y - \mathbf{X}\hat{\beta}$?
 - d. Чему равна ковариационная матрица векторов остатков МНК и ОМНК?
7. Предположим, что y имеет плотность распределения

$$\text{pdf } f(y | \mathbf{x}) = (1/\mathbf{x}'\beta)e^{-y/(\mathbf{x}'\beta)}, y > 0.$$

- Тогда $E[y | \mathbf{x}] = \mathbf{x}'\beta$ и $\text{Var}[y | \mathbf{x}] = (\mathbf{x}'\beta)^2$. Для этой модели докажите, что оценки ОМНК и методом максимального правдоподобия совпадают, даже если это распределение содержит те же параметры в функции условного среднего и дисперсии остатков.
8. Предположим, что регрессионная модель представлена в виде $y = \mu + \varepsilon$, где ε имеет нулевое среднее, постоянную дисперсию и одинаковую корреляцию ρ по всем наблюдениям. Тогда $\text{Cov}[\varepsilon_i, \varepsilon_j] = \sigma^2\rho$, если $i \neq j$. Докажите, что оценка методом наименьших квадратов μ несостоятельна. Найдите характеристические корни Ω и покажите, что условие 2 после теоремы 9.2 нарушается.
9. Предположим, что регрессионная модель имеет вид $y_i = \mu + \varepsilon_i$, где $E[\varepsilon_i | x_i] = 0$, $\text{Cov}[\varepsilon_i, \varepsilon_j | x_i, x_j] = 0$ для $i \neq j$, но $\text{Var}[\varepsilon_i | x_i] = \sigma^2 x_i^2$, $x_i > 0$.
- a. Имея выборку наблюдений по y_i и x_i , найдите наиболее эффективную оценку μ . Чему равна ее дисперсия?
 - b. Чему равны оценка МНК для μ и дисперсия оценки МНК?
 - c. Докажите, что оценка из пункта а по крайней мере так же эффективна, как оценка из пункта б.
10. Рассмотрим модель из упражнения 9. Чему равен предел по вероятности для $s^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$? Заметьте, что s^2 — это оценка МНК остаточной дисперсии. Также это n , умноженное на обычную оценку дисперсии оценки МНК:

$$\text{Est. Var}[\bar{y}] = s^2(\mathbf{X}'\mathbf{X})^{-1} = \frac{s^2}{n}.$$

Как это уравнение соотносится с истинным значением, которое вы нашли в пункте б упражнения 9? Дает ли обычная оценка корректную оценку истинной асимптотической дисперсии оценки наименьших квадратов?

11. Для модели из упражнения 9 предположите, что ε нормально распределена со средним ноль и дисперсией $\sigma^2[1 + (\gamma x)^2]$. Покажите, что σ^2 и γ^2 могут быть состоятельно оценены регрессией МНК-остатков этой модели на константу и x^2 . Эффективна ли эта оценка?
12. Две выборки по 50 наблюдений каждая дают следующие матрицы моментов. (В каждом случае X состоит из константы и одной переменной.)

| | Выборка 1 | Выборка 2 |
|-------|--|--|
| $X'X$ | $\begin{bmatrix} 50 & 300 \\ 300 & 2100 \end{bmatrix}$ | $\begin{bmatrix} 50 & 300 \\ 300 & 2100 \end{bmatrix}$ |
| $y'X$ | $[300 \quad 2000]$ | $[300 \quad 2200]$ |
| $y'y$ | $[2100]$ | $[2800]$ |

- Вычислите МНК-оценки коэффициентов регрессии и остаточной дисперсии s^2 для каждой выборки. Вычислите R^2 для каждой регрессии.
- Вычислите МНК-оценки вектора коэффициентов, предполагая, что коэффициенты и дисперсия остатков одинаковы в обеих регрессиях. Также вычислите оценку асимптотической ковариационной матрицы оценки.
- Проверьте гипотезу о том, что дисперсии в двух регрессиях совпадают, не предполагая, что коэффициенты одинаковы.
- Вычислите двухшаговую ДОМНК-оценку коэффициентов регрессий, предполагая, что их свободный член и коэффициенты наклона одинаковы. Вычислите оценку ковариационной матрицы и сравните ее с результатом из пункта b.

Эмпирические приложения

1. В этом приложении используются следующие данные.

| 50 наблюдений y | | | | | | | | |
|---------------------|--------|-------|-------|-------|-------|-------|--------|-------|
| -1,42 | 2,75 | 2,10 | -5,08 | 1,49 | 1,00 | 0,16 | -1,11 | 1,66 |
| -0,26 | -4,87 | 5,94 | 2,21 | -6,87 | 0,90 | 1,61 | 2,11 | -3,82 |
| -0,62 | 7,01 | 26,14 | 7,39 | 0,79 | 1,93 | 1,97 | -23,17 | -2,52 |
| -1,26 | -0,15 | 3,41 | -5,45 | 1,31 | 1,52 | 2,04 | 3,00 | 6,31 |
| 5,51 | -15,22 | -1,47 | -1,48 | 6,66 | 1,78 | 2,62 | -5,16 | -4,71 |
| -0,35 | -0,48 | 1,24 | 0,69 | 1,91 | | | | |
| 50 наблюдений x_1 | | | | | | | | |
| -1,65 | 1,48 | 0,77 | 0,67 | 0,68 | 0,23 | -0,40 | -1,13 | 0,15 |
| -0,63 | 0,34 | 0,35 | 0,79 | 0,77 | -1,04 | 0,28 | 0,58 | -0,41 |
| -1,78 | 1,25 | 0,22 | 1,25 | -0,12 | 0,66 | 1,06 | -0,66 | -1,18 |
| -0,80 | -1,32 | 0,16 | 1,06 | -0,60 | 0,79 | 0,86 | 2,04 | -0,51 |
| 0,02 | 0,33 | -1,99 | 0,70 | -0,17 | 0,33 | 0,48 | 1,90 | -0,18 |
| -0,18 | -1,62 | 0,39 | 0,17 | 1,02 | | | | |

Окончание табл.

| 50 наблюдений x_2 | | | | | | | | |
|---------------------|--------|-------|-------|-------|-------|-------|--------|-------|
| -0,67 | 0,70 | 0,32 | 2,88 | -0,19 | -1,28 | -2,72 | -0,70 | -1,55 |
| -0,74 | -1,87 | 1,56 | 0,37 | -2,07 | 1,20 | 0,26 | -1,34 | -2,10 |
| 0,61 | 2,32 | 4,38 | 2,16 | 1,51 | 0,30 | -0,17 | 7,82 | -1,15 |
| 1,77 | 2,92 | -1,94 | 2,09 | 1,50 | -0,46 | 0,19 | -0,39 | 1,54 |
| 1,87 | -3,45 | -0,88 | -1,53 | 1,42 | -2,70 | 1,77 | -1,89 | -1,85 |
| 2,01 | 1,26 | -2,02 | 1,91 | -2,23 | | | | |
| 50 наблюдений y | | | | | | | | |
| -1,42 | 2,75 | 2,10 | -5,08 | 1,49 | 1,00 | 0,16 | -1,11 | 1,66 |
| -0,26 | -4,87 | 5,94 | 2,21 | -6,87 | 0,90 | 1,61 | 2,11 | -3,82 |
| -0,62 | 7,01 | 26,14 | 7,39 | 0,79 | 1,93 | 1,97 | -23,17 | -2,52 |
| -1,26 | -0,15 | 3,41 | -5,45 | 1,31 | 1,52 | 2,04 | 3,00 | 6,31 |
| 5,51 | -15,22 | -1,47 | -1,48 | 6,66 | 1,78 | 2,62 | -5,16 | -4,71 |
| -0,35 | -0,48 | 1,24 | 0,69 | 1,91 | | | | |
| 50 наблюдений x_1 | | | | | | | | |
| -1,65 | 1,48 | 0,77 | 0,67 | 0,68 | 0,23 | -0,40 | -1,13 | 0,15 |
| -0,63 | 0,34 | 0,35 | 0,79 | 0,77 | -1,04 | 0,28 | 0,58 | -0,41 |
| -1,78 | 1,25 | 0,22 | 1,25 | -0,12 | 0,66 | 1,06 | -0,66 | -1,18 |
| -0,80 | -1,32 | 0,16 | 1,06 | -0,60 | 0,79 | 0,86 | 2,04 | -0,51 |
| 0,02 | 0,33 | -1,99 | 0,70 | -0,17 | 0,33 | 0,48 | 1,90 | -0,18 |
| -0,18 | -1,62 | 0,39 | 0,17 | 1,02 | | | | |
| 50 наблюдений x_2 | | | | | | | | |
| -0,67 | 0,70 | 0,32 | 2,88 | -0,19 | -1,28 | -2,72 | -0,70 | -1,55 |
| -0,74 | -1,87 | 1,56 | 0,37 | -2,07 | 1,20 | 0,26 | -1,34 | -2,10 |
| 0,61 | 2,32 | 4,38 | 2,16 | 1,51 | 0,30 | -0,17 | 7,82 | -1,15 |
| 1,77 | 2,92 | -1,94 | 2,09 | 1,50 | -0,46 | 0,19 | -0,39 | 1,54 |
| 1,87 | -3,45 | -0,88 | -1,53 | 1,42 | -2,70 | 1,77 | -1,89 | -1,85 |
| 2,01 | 1,26 | -2,02 | 1,91 | -2,23 | | | | |

- Оцените МНК регрессию y на константу, x_1 и x_2 . Удостоверьтесь, что вы также рассчитали обычную оценку асимптотической ковариационной матрицы оценки МНК.
 - Вычислите оценку Уайта асимптотической ковариационной матрицы МНК-оценки.
 - Проверьте наличие гетероскедастичности, используя общий тест Уайта. Можете ли вы сделать вывод о природе гетероскедастичности на основе полученного результата?
 - Проверьте наличие гетероскедастичности, используя тест множителей Лагранжа Бройша–Пэгана/Годфри.
 - Переоцените параметры модели, используя двухшаговую оценку ДОМНК. Используйте представление Харви
$$\text{Var}[\varepsilon_i | x_{i1}, x_{i2}] = \sigma^2 \exp(\gamma_1 x_{i1} + \gamma_2 x_{i2}).$$
2. (Мы готовимся к использованию оценки методом максимального правдоподобия для оценки обсуждаемых в этой главе моделей в главе 14.) В примере 9.4 мы вычислили итерационную оценку ДОМНК с использованием данных по авиакомпаниям и модель $\text{Var}[\varepsilon_{it} | \text{Loadfactor}_{i,t}] = \exp(\gamma_1 +$

$+\gamma_2 \text{Loadfactor}_{i,t}$). Веса, рассчитанные для каждой итерации, были получены оценкой (γ_1, γ_2) при помощи МНК регрессии $\ln \hat{\epsilon}_{i,t}^2$ на константу и Loadfactor . Оценка методом максимального правдоподобия будет аналогична, однако теперь веса будут рассчитаны при помощи регрессии $[\hat{\epsilon}_{i,t}^2 / \hat{\sigma}_{i,t}^2 - 1]$ на константу и $\text{Loadfactor}_{i,t}$. Используйте эту альтернативную процедуру для оценки модели. Получили ли вы другой результат?

ГЛАВА 10

Системы уравнений

10.1. Введение

Существует множество условий, при которых однофакторные модели, рассмотренные в предыдущих главах, применяются для групп взаимосвязанных переменных. В этом контексте имеет смысл рассматривать несколько моделей вместе. Ниже мы приводим некоторые примеры.

1. Модель Манэлла (Munnell, 1990) для валового продукта по 48 американским штатам выглядит так:

$$\ln GSP_{it} = \beta_{1i} + \beta_{2i} \ln pc_{it} + \beta_{3i} \ln hwy_{it} + \beta_{4i} \ln water_{it} + \beta_{5i} \ln util_{it} + \\ + \beta_{6i} \ln emp_{it} + \beta_{7i} unemp_{it} + \varepsilon_{it}.$$

В модели рассматривается отдельно каждый штат и, таким образом, получена система, состоящая из 48 линейных регрессионных моделей. В данной модели наблюдения коррелированы во времени в пределах каждого штата. Главным здесь и в приложении из следующего примера является вопрос о том, насколько обоснованно допущение, что вектор коэффициентов одинаков для всех штатов (индивидуальных наблюдений) в выборке.

2. Модель ценообразования финансовых активов (CAPM) определяет, что для данной ценной бумаги

$$r_{it} - r_{ft} = \alpha_i + \beta_i(r_{mt} - r_{ft}) + \varepsilon_{it},$$

где r_{it} — доходность за период t по бумаге i , r_{ft} — доходность по безрисковой бумаге, r_{mt} — рыночная доходность, а β_i — бета-коэффициент ценной бумаги. Очевидно, что остатки по различным ценным бумагам коррелированы. Знание того, что прибыль по ценной бумаге i превышает показатель безрисковой бумаги на определенную величину, дает некоторую информацию относительно чрезмерного дохода от бумаги j , по крайней мере для некоторых j . Это может оказаться полезным для того, чтобы оценить уравнения вместе, не игнорируя эту связь.

3. Песаран и Смит (Pesaran, Smith (1995)) предложили динамическую модель заработной платы в 38 отраслях производства в Великобритании. Центральное уравнение имеет следующий вид:

$$y_{it} = \alpha_i + \mathbf{x}'_{it}\beta_i + \gamma_i y_{i,t-1} + \varepsilon_{it}.$$

В работе (Nair-Reichert, Weinhold (2001)) аналогичная модель используется для анализа роста развивающихся стран. В обоих случаях каждая группа (будь то отрасль или страна) могла быть изучена отдельно. Однако анализ связи между этими группами и занимательный вопрос «возможности объединения в пул», т.е. обоснованности допущения о равных коэффициентах,

составили основную часть анализа. Наличие в правой части регрессии запаздывающего значения зависимой переменной создает существенное затруднение.

4. В модели производства условия оптимизации, которые подразумевает экономическая теория, предполагают, что если у фирмы есть набор цен факторов производства \mathbf{p} , то набор факторов, минимизирующих издержки и необходимых для производства выпуска Q , будет определяться системой уравнений вида $x_m = f_m(Q, \mathbf{p})$. На практике модель выглядит следующим образом:

$$\begin{aligned}x_1 &= f_1(Q, \mathbf{p}|\boldsymbol{\theta}) + \varepsilon_1, \\x_2 &= f_2(Q, \mathbf{p}|\boldsymbol{\theta}) + \varepsilon_2, \dots \\x_M &= f_M(Q, \mathbf{p}|\boldsymbol{\theta}) + \varepsilon_M,\end{aligned}$$

где $\boldsymbol{\theta}$ — вектор параметров, составляющих часть технологии, а ε_m представляют ошибки оптимизации. Опять же остатки взаимосвязаны. Кроме того, одинаковые параметры производственной технологии включены во все уравнения спроса, чтобы система уравнений имела перекрестные ограничения. Оценка уравнений по отдельности приведет к потере информации об одинаковом множестве параметров, имеющемся во всех уравнениях.

5. Ключевой формой модели равновесия на рынке является

$$\begin{aligned}Q_{Demand} &= \alpha_1 + \alpha_2 Price + \alpha_3 Income + \mathbf{d}'\boldsymbol{\alpha} + \varepsilon_{Demand}, \\Q_{Supply} &= \beta_1 + \beta_2 Price + \mathbf{s}'\boldsymbol{\beta} + \varepsilon_{Supply}, \\Q_{Equilibrium} &= Q_{Demand} = Q_{Supply},\end{aligned}$$

где \mathbf{d} и \mathbf{s} — некоторые переменные, влияющие на равновесие посредством их воздействия на кривые спроса и предложения соответственно. Эта модель отличается от предложенных выше, так как включение третьего уравнения делает *Price* (цену) неэкзогенной в данной системе уравнений. Уравнения этой модели больше вписываются в структуру метода инструментальных переменных, рассмотренную в главе 8, нежели в регрессионные модели из глав 1–7. Структура системы уравнений, приведенная в этой главе, позволяет получить дополнительные результаты по оцениванию таких «моделей одновременных уравнений», как эта.

Модель регрессии, состоящая из нескольких уравнений, рассматриваемая в этой главе, задает определенные рамки, которые можно использовать при различных условиях. Модели производства и затрат, приведенные в разделе 10.5, создают базу, используемую в академической литературе об эмпирическом анализе поведения фирм. На макроэкономическом уровне «модели векторной авторегрессии», рассматриваемые в главе 21, являются характерными формами модели квазинезависимых регрессий из раздела 10.5. Модель одновременных уравнений из раздела 10.6 выходит за пределы спецификации различных моделей главы 8.

В этой главе рассматриваются основные теоретические вопросы, касающиеся систем взаимосвязанных регрессионных уравнений. В разделе 10.2 анализируются общая модель, в которой каждое уравнение имеет собственное фиксированное множество параметров, а также эффективные методы

оценки таких систем. В подразделе 10.2.6 показана «объединенная» модель с равными коэффициентами во всех уравнениях. Модели производства и потребительского спроса представляют собой особый случай общей модели, в котором уравнения подчиняются дополнительному ограничению, имеющему важные последствия для спецификации и оценки. В разделе 10.3 предлагается распространение систем внешне не связанных уравнений на обобщенные регрессионные модели с гетероскедастичностью и автокорреляцией, которые рассматриваются в главах 9 и 20. В разделе 10.4 модель внешне не связанных регрессий распространяется на нелинейные системы уравнений. В разделе 10.5 мы рассматриваем классическое приложение модели внешне не связанных регрессий, которое наглядно показывает интересные особенности этого вида исследования спроса, изучаемого в приведенных источниках. Далее модель внешне не связанных регрессий распространяется на транслогарифмическую спецификацию, которая является основой для многих современных исследований производства и издержек в области микроэкономики. В заключение в разделе 10.6 мы объединяем выводы из главы 8, касающиеся моделей с эндогенными переменными, и результаты, полученные для систем нескольких уравнений в этой главе. В разделе 10.6 рассмотрим **модели одновременных уравнений**. Это системы уравнений, которые строятся на системе внешне не связанных уравнений и предполагают наличие взаимосвязи между зависимыми переменными. Примером практического применения модели одновременных уравнений является модель спроса и предложения, приведенная в примере 5 выше: это равновесие, где цена и количество на рынке определяются вместе.

10.2. Системы внешне не связанных уравнений

Все примеры, приведенные во введении к данной главе, имеют общую структуру, состоящую из нескольких уравнений, которую можно записать как

$$\begin{aligned} y_1 &= X_1\beta_1 + \epsilon_1, \\ y_2 &= X_2\beta_2 + \epsilon_2, \\ &\dots \\ y_M &= X_M\beta_M + \epsilon_M. \end{aligned} \tag{10-1}$$

Здесь имеется M уравнений и T наблюдений из выборки данных, используемой для их оценки¹. Второй и третий примеры демонстрируют разные виды ограничений для уравнений и разные структуры остатков. Однако ко всем из них будет применен некоторый исходный набор принципов². Система внешне не связанных уравнений в (10-1) имеет вид

$$y_i = X_i\beta_i + \epsilon_i, i = 1, \dots, M. \tag{10-2}$$

Определим вектор случайных возмущений $MT \times 1$:

¹Использование T необязательно подразумевает привязку к временным рядам. Так, в примере 4, рассмотренном выше, данные могли быть межобъектными.

²См. обзоры (Srivastava, Dwivedi (1979)), (Srivastava, Giles (1987)) и (Fiebig (2001)).

$$\boldsymbol{\varepsilon} = [\boldsymbol{\varepsilon}'_1, \boldsymbol{\varepsilon}'_2, \dots, \boldsymbol{\varepsilon}'_M]'$$

Мы предполагаем строгую экзогенность \mathbf{X}_i

$$E[\boldsymbol{\varepsilon} | \mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_M] = \mathbf{0},$$

а также гомоскедастичность

$$E[\boldsymbol{\varepsilon}_m \boldsymbol{\varepsilon}'_m | \mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_M] = \sigma_{mm} \mathbf{I}_T.$$

Мы также допускаем, что все T наблюдений используются при оценке параметров в M уравнениях³. Каждое уравнение содержит K_i регрессоров, в сумме $K = \sum_{i=1}^M K_i$. Необходимо, чтобы выполнялось неравенство $T > K$. Предполагается также, что рассматриваемые данные являются доброкачественными, как описано в подразделе 4.4.1, и мы не будем рассматривать уравнения системы по отдельности. Пока мы также предполагаем, что шоки не взаимосвязаны по наблюдениям, но при этом взаимосвязаны между уравнениями. Таким образом, $E[\boldsymbol{\varepsilon}_{js} | \mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_M] = \sigma_{ij}$, если $t = s$, и 0 в противном случае.

Исходя из этого шоки выражаются как

$$E[\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}'_j | \mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_M] = \sigma_{ij} \mathbf{I}_T$$

или

$$E[\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}' | \mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_M] = \boldsymbol{\Omega} = \begin{bmatrix} \sigma_{11} \mathbf{I} & \sigma_{12} \mathbf{I} & \cdots & \sigma_{1M} \mathbf{I} \\ \sigma_{21} \mathbf{I} & \sigma_{22} \mathbf{I} & \cdots & \sigma_{2M} \mathbf{I} \\ & \vdots & & \\ \sigma_{M1} \mathbf{I} & \sigma_{M2} \mathbf{I} & \cdots & \sigma_{MM} \mathbf{I} \end{bmatrix}. \quad (10-3)$$

В дальнейшем будет удобнее использовать общее название для моделей, в которых матрицы данных представляют собой специфические наборы, сгруппированные по определенному признаку. Модель Манэлла, рассмотренная во введении, именно такой случай. Этот особый вид систем внешне не связанных уравнений является **многомерной регрессионной моделью**. В отличие от этого примера модель функции затрат, представленная в подразделе 10.4.1, не является моделью такого вида, она состоит из функции затрат, включающей выпуск и цены, а также уравнения распределения затрат, в которые входит лишь набор констант. Важно отметить, что это всего лишь удобное название для особого вида систем внешне не связанных уравнений, а не видоизмененная модель.

10.2.1. Обобщенный метод наименьших квадратов

Каждое уравнение само по себе является классической регрессией. Поэтому параметры могли быть оценены последовательно, если не эффективно, в каждом уравнении по отдельности по обычному методу наименьших

³ Существует не много результатов для неравного числа наблюдений в уравнениях, например, Schmidt (1977), Baltagi, Garvin, Kerman (1989), Conniffe (1985), Hwang (1990) и Im (1994). Но случай фиксированного T является нормой на практике.

квадратов. **Обобщенная регрессионная модель** имеет вид

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ & & \vdots & \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_M \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_M \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_M \end{bmatrix} = \mathbf{X}\beta + \varepsilon. \quad (10-4)$$

Поэтому для эффективной оценки используется обобщенный метод наименьших квадратов (GLS)⁴. Модель принимает особенно удобную форму. Для t -го наблюдения ковариационная матрица шоков $M \times M$ выглядит как

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1M} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2M} \\ & \vdots & & \\ \sigma_{M1} & \sigma_{M2} & \cdots & \sigma_{MM} \end{bmatrix}, \quad (10-5)$$

так что в (10-3)

$$\Omega = \Sigma \otimes \mathbf{I} \quad (10-6)$$

$$\text{и } \Omega^{-1} = \Sigma^{-1} \otimes \mathbf{I}^5.$$

Обозначая ij -й элемент матрицы Σ^{-1} через σ^{ij} , мы находим ОМНК (GLS)-оценку:

$$\hat{\beta} = [\mathbf{X}'\Omega^{-1}\mathbf{X}]^{-1}\mathbf{X}'\Omega^{-1}\mathbf{y} = [\mathbf{X}'(\Sigma^{-1} \otimes \mathbf{I})\mathbf{X}]^{-1}\mathbf{X}'(\Sigma^{-1} \otimes \mathbf{I})\mathbf{y}. \quad (10-7)$$

Раскрывая **кронекеровы произведения**, получим:

$$3\hat{\beta} = \begin{bmatrix} \sigma^{11}\mathbf{X}'_1\mathbf{X}_1 & \sigma^{12}\mathbf{X}'_1\mathbf{X}_2 & \cdots & \sigma^{1M}\mathbf{X}'_1\mathbf{X}_M \\ \sigma^{21}\mathbf{X}'_2\mathbf{X}_1 & \sigma^{22}\mathbf{X}'_2\mathbf{X}_2 & \cdots & \sigma^{2M}\mathbf{X}'_2\mathbf{X}_M \\ & \vdots & & \\ \sigma^{M1}\mathbf{X}'_M\mathbf{X}_1 & \sigma^{M2}\mathbf{X}'_M\mathbf{X}_2 & \cdots & \sigma^{MM}\mathbf{X}'_M\mathbf{X}_M \end{bmatrix}^{-1} \begin{bmatrix} \sum_{j=1}^M \sigma^{1j}\mathbf{X}'_1\mathbf{y}_j \\ \sum_{j=1}^M \sigma^{2j}\mathbf{X}'_2\mathbf{y}_j \\ \vdots \\ \sum_{j=1}^M \sigma^{Mj}\mathbf{X}'_M\mathbf{y}_j \end{bmatrix}.$$

Асимптотическая ковариационная матрица для GLS-оценки представляет собой обратную матрицу в (10-7), взятую в скобки. Все выводы главы 9 для обобщенной модели регрессии применимы к этой модели (которая обладает и **гетероскедастичностью**, и **автокорреляцией**).

Очевидно, что этот метод отличается от обычного метода наименьших квадратов. Однако в этом случае уравнения связаны только своими шоками – отсюда и название «система внешне не связанных уравнений», – поэтому интересным представляется узнать, насколько эффективнее использовать обобщенный метод наименьших квадратов вместо обычного. Зельнер (Zellner (1962)) и Двиведи и Сривастава (Dwivedi, Srivastava (1978)) детально рассмотрели несколько случаев.

⁴Zellner (1962) и Telser (1964).

⁵См. приложение А.5.5.

1. Если уравнения *действительно* несвязанные, т.е. $\sigma_{ij} = 0$ для $i \neq j$, тогда мы не получаем выигрыша при использовании ОМНК-оценки полной системы уравнений. Действительно, ОМНК-оценка системы уравнений в данном случае представляет оценку каждого отдельного уравнения методом наименьших квадратов⁶.
2. Если уравнения имеют **одинаковые объясняющие переменные**, т.е. $X_i = X_j$, тогда МНК и ОМНК тождественны. Мы рассмотрим этот случай в подразделе 10.2.2⁷.
3. Если регрессоры в одном блоке уравнений принадлежат подмножеству другого, тогда ОМНК не приносит большей эффективности по сравнению с МНК при оценке меньшего набора уравнений; таким образом, ОМНК и МНК снова тождественны⁸.

В более общем случае, с неограниченной корреляцией шоков и разными регрессорами в уравнениях, выводы усложняются и зависят от данных. Два следующих утверждения имеют место в общем случае:

1. Чем больше корреляция шоков, тем больший выигрыш в эффективности мы получаем при использовании ОМНК.
2. Чем меньше корреляция между матрицами X , тем больший выигрыш в эффективности мы получаем при использовании ОМНК⁹.

10.2.2. Системы внешне не связанных уравнений с одинаковыми регрессорами

Случай **одинаковых регрессоров** встречается довольно часто, особенно в модели ценообразования финансовых активов (САРМ) в практической области финансов — что было указано во введении к главе. Именно в этом случае оценка обобщенным методом наименьших квадратов соответствует оценке обычным методом наименьших квадратов каждого уравнения в отдельности. Зафиксируем допущение о том, что $X_i = X_j = X$, так что $X'_i X_j = X' X$ для всех i и j в (10-7). Отсюда обратная матрица в правой части приобретает вид $[\Sigma^{-1} \otimes X' X]^{-1}$, что равно $[\Sigma \otimes (X' X)^{-1}]$, если использовать (A-76). Также в правой части каждый член $X'_i y_j$ равен $X' y_j$, что, в свою очередь, равно $X' X b_j$. Исходя из полученных результатов после извлечения

⁶См. также другие случаи, где МНК = ОМНК, у Baltagi (1989) и Bartels, Fiebig (1992).

⁷Интересный результат, хотя и не имеющий большого практического значения, состоит в том, что этот результат также применим, если X представляет собой неединичные и необязательно одинаковые линейные комбинации одного набора переменных. Формальный результат, выводимый из теоремы Крускала [Davidson, MacKinnon (1993, p. 294)], таков, что оценки МНК и ОМНК совпадут, если K столбцов в X будут линейными комбинациями конкретных K собственных векторов матрицы Ω . Показывая здесь равенство МНК и ОМНК, мы проверили условия нашего вывода. Общий результат выводится в упражнениях.

⁸Такой результат был проанализирован Goldberger (1970) и позже Revankar (1974) и Conniffe (1982a, b).

⁹Binkley (1982) и Binkley, Nelson (1988).

$\mathbf{X}'\mathbf{X}$ из суммирования в правой части получаем

$$\hat{\beta} = \begin{bmatrix} \sigma_{11}(\mathbf{X}'\mathbf{X})^{-1} & \sigma_{12}(\mathbf{X}'\mathbf{X})^{-1} & \cdots & \sigma_{1M}(\mathbf{X}'\mathbf{X})^{-1} \\ \sigma_{21}(\mathbf{X}'\mathbf{X})^{-1} & \sigma_{22}(\mathbf{X}'\mathbf{X})^{-1} & \cdots & \sigma_{2M}(\mathbf{X}'\mathbf{X})^{-1} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{M1}(\mathbf{X}'\mathbf{X})^{-1} & \sigma_{M2}(\mathbf{X}'\mathbf{X})^{-1} & \cdots & \sigma_{MM}(\mathbf{X}'\mathbf{X})^{-1} \end{bmatrix} \begin{bmatrix} (\mathbf{X}'\mathbf{X}) \sum_{l=1}^M \sigma^{1l} \mathbf{b}_l \\ (\mathbf{X}'\mathbf{X}) \sum_{l=1}^M \sigma^{2l} \mathbf{b}_l \\ \vdots \\ (\mathbf{X}'\mathbf{X}) \sum_{l=1}^M \sigma^{Ml} \mathbf{b}_l \end{bmatrix}. \quad (10-8)$$

Теперь выделим один из подвекторов, скажем первый, из $\hat{\beta}$. После умножения матрицы моментов сокращаются и остается следующее:

$$\hat{\beta}_1 = \sum_{j=1}^M \sigma_{1j} \sum_{l=1}^M \sigma^{jl} \mathbf{b}_l = \mathbf{b}_1 \left(\sum_{j=1}^M \sigma_{1j} \sigma^{j1} \right) + \mathbf{b}_2 \left(\sum_{j=1}^M \sigma_{1j} \sigma^{j2} \right) + \cdots + \mathbf{b}_M \left(\sum_{j=1}^M \sigma_{1j} \sigma^{jM} \right).$$

Члены в скобках являются элементами первой строки $\Sigma \Sigma^{-1} = \mathbf{I}$, поэтому конечный результат равен $\hat{\beta}_1 = \mathbf{b}_1$. Для оставшихся подвекторов, получаемых таким же образом, $\hat{\beta}_i = \mathbf{b}_i$, это и есть искомый результат¹⁰.

Опять же важный вывод, который мы получили здесь, заключается в том, что в системе внешне не связанных уравнений, если все уравнения включают одинаковые регрессоры, эффективной будет оценка простым методом наименьших квадратов, примененным к каждому уравнению; МНК-оценка будет такой же, как и ОМНК. Кроме того, асимптотическая ковариационная матрица $\hat{\beta}$ для этого случая дана в виде большой обратной матрицы в скобках в (10-8) и будет оценена так:

$$\text{Est. Asy. Cov}[\hat{\beta}_i, \hat{\beta}_j] = \hat{\sigma}_{ij} (\mathbf{X}'\mathbf{X})^{-1}, \quad i, j = 1, \dots, M, \quad \text{где } \hat{\Sigma}_{ij} = \hat{\sigma}_{ij} = \frac{1}{T} \mathbf{e}'_i \mathbf{e}_j.$$

За исключением некоторых отдельных случаев, этот общий результат утратит силу, если в модели есть любые ограничения на β в каком-либо уравнении или в системе.

10.2.3. Допустимый обобщенный метод наименьших квадратов (ДОМНК, FGLS)

Выше предполагалось, что Σ неизвестна, что, как обычно, маловероятно. Тем не менее были разработаны так называемые FGLS-оценки¹¹. Остатки наименьших квадратов могут быть использованы для получения состоятельных оценок элементов матрицы Σ с

$$\hat{\sigma}_{ij} = s_{ij} = \frac{\mathbf{e}'_i \mathbf{e}_j}{T}. \quad (10-9)$$

¹⁰Hashimoto, Ohtani (1990) для обсуждения вопроса тестирования гипотез в этом случае.

¹¹Zellner (1962) и Zellner, Huang (1962). FGLS-оценка для данной модели также имеет название **эффективная оценка Zellнера (ZEF)** по имени автора работы (Zellner (1962)), в которой она была введена.

Состоятельность s_{ij} следует из b_i и b_j ¹².

Иногда допустима поправка на степени свободы в делителе. Два случая, в которых оценки не смещены при $i = j$, имеют вид¹³

$$s_{ij}^* = \frac{e_i' e_j}{[(T - K_i)(T - K_j)]^{1/2}} \quad \text{и} \quad s_{ij}^{**} = \frac{e_i' e_j}{T - \max(K_i, K_j)}. \quad (10-10)$$

Является ли в данном случае несмещенность оценки Σ , используемой для ДОМНК, хорошим свойством — неясно. Асимптотические свойства оценки допустимого ОМНК $\hat{\beta}$ не зависят от несмещенной оценки Σ ; требуется только состоятельность. Все результаты из глав 8 и 9 для оценок ДОМНК распространяются на данную модель без изменений. Далее будем использовать (10-9). Имея

$$S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1M} \\ s_{21} & s_{22} & \cdots & s_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ s_{M1} & s_{M2} & \cdots & s_{MM} \end{bmatrix}, \quad (10-11)$$

мы можем получить FGLS-оценку как обычно.

10.2.4. Проверка гипотез

Для проверки гипотез о β существует статистический аналог F -критерия для систем уравнений, который выглядит так:

$$F[J, MT - K] = \frac{(\mathbf{R}\hat{\beta} - \mathbf{q})' [\mathbf{R}(\mathbf{X}'\hat{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{R}']^{-1} (\mathbf{R}\hat{\beta} - \mathbf{q}) / J}{\hat{\epsilon}'\hat{\Omega}^{-1}\hat{\epsilon} / (MT - K)}. \quad (10-12)$$

Для того чтобы вычислить эту статистику, необходимо знать неизвестную матрицу Ω . Если мы подставим FGLS-оценку $\hat{\Omega}$, основанную на (10-9), и будем использовать тот факт, что знаменатель стремится к единице, тогда для больших выборок статистика будет вести себя как

$$\hat{F} = \frac{1}{J} (\mathbf{R}\hat{\beta} - \mathbf{q})' [\mathbf{R} \widehat{\text{Var}}[\hat{\beta}] \mathbf{R}']^{-1} (\mathbf{R}\hat{\beta} - \mathbf{q}). \quad (10-13)$$

Это можно свести к стандартной таблице значений F -критерия. Так как в ней используется оцененная Σ , даже при нормально распределенных шоках F -распределение обосновано лишь приблизительно. В общем случае если $n \rightarrow \infty$, $F[J, n]$ -статистика стремится к распределению хи-квадрат с J степенями свободы, умноженному на $1/J$. Исходя из этого альтернативная тестовая статистика для проверки того, что нулевая гипотеза верна, имеющая предельное распределение хи-квадрат с J степенями свободы, выглядит как

$$J\hat{F} = (\mathbf{R}\hat{\beta} - \mathbf{q})' [\mathbf{R} \widehat{\text{Var}}[\hat{\beta}] \mathbf{R}']^{-1} (\mathbf{R}\hat{\beta} - \mathbf{q}). \quad (10-14)$$

¹²Может показаться удивительным, что в предположении симметричности плотности ε , как это было бы при нормальном распределении, оценки b_i также не смещены (Kakwani (1967)).

¹³Judge et al. (1985), Theil (1971), Srivastava и Giles (1987).

Эта статистика также носит название **статистика Вальда**, которая измеряет расстояние между $\mathbf{R}\hat{\beta}$ и \mathbf{q} . Обе статистики обоснованы лишь асимптотически, однако (10-13) лучше работает в небольших или средних выборках¹⁴. Опять же делитель, используемый для вычисления $\hat{\sigma}_{ij}$, может меняться, но общего правила здесь не существует.

Гипотеза, представляющая особый интерес, — это **ограничение однородности** о равенстве векторов коэффициентов в многомерной регрессионной модели. Этот случай обычно записывается в следующем виде. Ограничение однородности заключается в том, что $\beta_i = \beta_M, i = 1, \dots, M - 1$. В соответствии с (10-13)–(10-14) сформулируем гипотезу так:

$$\mathbf{R}\beta = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} & -\mathbf{I} \\ \mathbf{0} & \mathbf{I} & \dots & \mathbf{0} & -\mathbf{I} \\ & & \dots & & \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{I} & -\mathbf{I} \end{bmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_M \end{pmatrix} = \begin{pmatrix} \beta_1 - \beta_M \\ \beta_2 - \beta_M \\ \dots \\ \beta_{M-1} - \beta_M \end{pmatrix} = \mathbf{0}. \quad (10-15)$$

Эта запись определяет сумму $(M - 1)K$ ограничений для вектора параметров размером $KM \times 1$. Обозначим оценку асимптотической ковариации для $(\hat{\beta}_i, \hat{\beta}_j)$ через $\hat{\mathbf{V}}_{ij}$. Матрица в (10-13), взятая в скобки, будет иметь типичный вид

$$[\mathbf{R} \widehat{\text{Var}}[\hat{\beta}] \mathbf{R}']_{ij} = \hat{\mathbf{V}}_{ij} - \hat{\mathbf{V}}_{iM} - \hat{\mathbf{V}}_{Mj} + \hat{\mathbf{V}}_{MM}.$$

Ее вычисление может быть довольно громоздким. Тест будет проще, если модель подождена согласно максимуму правдоподобия, который мы рассмотрим в подразделе 14.9.3. Песаран и Ямагата (Pesaran, Yamagata (2008)) предложили альтернативный тест, который может использоваться при больших M и относительно небольших T .

10.2.5. Тест на спецификацию для систем внешне не связанных уравнений

Одна из интересных задач — определение того, равны ли внедиагональные элементы матрицы Σ нулю. Если это так, то для полного вектора параметров в отсутствие гетероскедастичности и автокорреляции эффективной является оценка, полученная для каждого уравнения системы простым методом наименьших квадратов. Вообще говоря, не существует стандартного теста для общего случая системы внешне не связанных уравнений, за исключением случая, когда в (10-2) и (10-3) вводится дополнительное допущение о нормальности шоков. При условии нормальности распределения шоков можно использовать три классических теста — тест Вальда, **критерий отношения правдоподобия** и **тест множителей Лагранжа**. По определенным причинам, к которым мы вскоре вернемся, тест Вальда в большинстве случаев будет слишком громоздким для применения. При нормально распределенных шоках статистика теста отношения правдоподобия для

¹⁴Judge et al. (1985, p. 476). Статистика Вальда часто дает плохие результаты для выборок небольшого размера, типичных в данной области. Fiebig (2001, p. 108–110) исследовал современную литературу по методам улучшения мощности тестовых процедур в системах внешне не связанных уравнений.

проверки нулевой гипотезы о том, что матрица Σ в (10-5) является диагональной матрицей при наличии альтернативной гипотезы о том, что Σ просто неограниченная положительно определенная матрица, будет иметь вид

$$\lambda_{LR} = T[\ln |\mathbf{S}_0| - \ln |\mathbf{S}_1|], \quad (10-16)$$

где \mathbf{S}_1 — ковариационная матрица остатков, определенная в (10-9) (без поправки на степени свободы). Остатки вычисляют, используя оценки параметров методом максимального правдоподобия (MLE), а не FGLS¹⁵. В условиях нулевой гипотезы модель может быть оценена эффективно, если оценивать отдельно каждое уравнение МНК, так что

$$\ln |\mathbf{S}_0| = \sum_{i=1}^M \ln(\mathbf{e}_i' \mathbf{e}_i / T),$$

где $\mathbf{e}_i = \mathbf{y}_i - \mathbf{X}_i \mathbf{b}_i$. Предельное распределение критерия отношения правдоподобия в условиях нулевой гипотезы будет распределено по хи-квадрат-распределению с $M(M-1)/2$ степенями свободы.

Критерий отношения правдоподобия требует наличия неограниченной MLE для вычисления ковариационной матрицы остатков в условиях альтернативной гипотезы, так что вычисления могут оказаться громоздкими. Более простая альтернатива — критерий множителей Лагранжа, предложенный Бройшем и Пэганом (Breusch, Pagan (1980)), который имеет вид

$$\begin{aligned} \lambda_{LM} &= T \sum_{i=2}^M \sum_{j=1}^{i-1} r_{ij}^2 = \\ &= (T/2)[\text{trace}(\mathbf{R}'\mathbf{R}) - M], \end{aligned} \quad (10-17)$$

где \mathbf{R} — матрица выборочных корреляций M множеств T МНК-остатков. Он имеет такое же выборочное распределение в условиях нулевой гипотезы, как и критерий отношения правдоподобия, но вычисления при этом упрощаются, так как этот критерий требует лишь наличия МНК-остатков.

Третий статистический тест — статистика Вальда. В принципе статистику Вальда для систем внешне не связанных уравнений можно вычислить, используя формулу

$$W = \hat{\sigma}' [\text{Asy. Var}(\hat{\sigma})]^{-1} \hat{\sigma},$$

¹⁵В системе внешне не связанных уравнений (SUR) в этой главе оценка методом максимального правдоподобия при нормально распределенных шоках может быть вычислена с помощью итеративной процедуры ДОМНК при повторении шагов (10-7) — (10-9) до тех пор, пока оценки не перестанут изменяться. Отметим, что такая процедура дает оценки максимального правдоподобия, когда она сходится, но нет гарантии, что она сойдется, и при этом не утверждается, что такая оценка будет единственной. Для нашей региональной базы данных итеративная FGLS-процедура не сходится даже после 1000 итераций. Результат работы Оберхофера и Кменты (Oberhofer, Kmenta (1974)) предполагает, что если процесс сходится, то он сходится к MLE. Но гарантии, что процесс будет сходящимся, нет. Возможно, что в рассматриваемом приложении проблема заключается в наличии очень маленькой выборки — всего 17 наблюдений. Обычно для выборок такого небольшого размера метод максимального правдоподобия не используется.

где $\hat{\sigma}$ — вектор длины $M(M-1)/2$, содержащий оценки внедиагональных (нижнетреугольных) элементов Σ , и в скобках — асимптотическая ковариационная матрица оценки. В условиях нормальности асимптотическая ковариационная матрица включает соответствующие элементы $2\Sigma \otimes \Sigma/T$. Можно оценить ковариации, используя оценки, основанные на моментах. Так как

$$\hat{\sigma}_{ij} = \frac{1}{T} \sum_{t=1}^T e_{it}e_{jt}$$

является средним значением T наблюдений, можно использовать обычные оценки дисперсии и ковариации с $\hat{\sigma}_{lm}$, откуда

$$f_{ij,lm} = \frac{1}{T} \frac{1}{T-1} \sum_{t=1}^T (e_{it}e_{jt} - \hat{\sigma}_{ij})(e_{lt}e_{mt} - \hat{\sigma}_{lm}). \quad (10-18)$$

Модифицированная статистика Вальда будет тогда иметь вид

$$W' = \hat{\sigma}'[\mathbf{F}]^{-1}\hat{\sigma},$$

где элементы \mathbf{F} — это соответствующие значения из (10-18). В этом случае вычисления, очевидно, более сложные, нежели в двух других случаях. Однако здесь не требуется предположение о нормальности шоков в модели. Требуется же (а) состоятельность оценок β_i , чтобы мы могли утверждать, что (б) оценки σ_{ij} состоятельны, и, наконец, (с) асимптотическая нормальность оценок в (б), необходимая для возможности использования теоремы 4.4. Все три условия должны выполняться в системе внешне не связанных уравнений (SUR) с доброкачественными регрессорами.

Предложенные альтернативные подходы [см., например, Johnson, Wichern (2005, p. 424)] строятся на следующей общей стратегии. В условиях альтернативной гипотезы о неограниченности Σ выборочные оценки Σ будут равны $\hat{\Sigma} = [\hat{\sigma}_{ij}]$, как определено в (10-9). В условиях любых ограничений в нулевой гипотезе оценка Σ будет равна $\hat{\Sigma}_0$, т.е. матрице, которая по структуре будет больше, чем $\hat{\Sigma}_1$, в смысле, определенном в приложении А. Статистические критерии, основанные на «избыточной вариации», такие как $T(\hat{\Sigma}_0 - \hat{\Sigma}_1)$, предлагаются в качестве тестовых процедур. Одна из них — критерий отношения правдоподобия (10-16).

10.2.6. Объединенная модель

Если все переменные в X_i одинаковы, а все векторы коэффициентов в (10-2) по предположению равны, то **объединенная модель** имеет вид

$$y_{it} = x'_{it}\beta + \varepsilon_{it}.$$

Эта модель отличается от модели панельных данных из главы 11, в которой предполагается, что корреляция между наблюдениями возникает в период t , но не внутри группы i . (Безусловно, даже при незначительном изменении данных получится такая же модель. Однако ее интерпретация другая, поэтому мы будем считать их разными.) Собрав T наблюдений для группы i , мы получим

$$y_i = X_i\beta + \epsilon_i$$

или для всех n групп

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \beta + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} = \mathbf{X}\beta + \epsilon, \quad (10-19)$$

где

$$E[\epsilon_i | \mathbf{X}] = 0,$$

$$E[\epsilon_i \epsilon_j' | \mathbf{X}] = \sigma_{ij} \Omega_{ij}. \quad (10-20)$$

Если $\Omega_{ij} = \mathbf{I}$, тогда это будет равнозначно SUR-модели (10-2) с одинаковыми векторами коэффициентов. Оценка обобщенным методом наименьших квадратов при таком предположении о **структуре ковариационной матрицы** будет иметь вид

$$\begin{aligned} \hat{\beta} &= [\mathbf{X}'(\Sigma \otimes \mathbf{I})^{-1}\mathbf{X}]^{-1}[\mathbf{X}'(\Sigma \otimes \mathbf{I})^{-1}\mathbf{y}] = \\ &= \left[\sum_{i=1}^n \sum_{j=1}^n \sigma^{ij} \mathbf{X}'_i \mathbf{X}_j \right]^{-1} \left[\sum_{i=1}^n \sum_{j=1}^n \sigma^{ij} \mathbf{X}'_i y_j \right], \end{aligned} \quad (10-21)$$

где σ^{ij} обозначает ij -й элемент из Σ^{-1} . Оценку FGLS можно вычислить с помощью (10-9), где e_i может быть либо рассчитано при помощи характерных для группы МНК-остатков, либо представлено как подвектор вектора остатков объединенной МНК-регрессии, оцененной с использованием всех nT наблюдений.

Сделаем важное замечание для допустимого обобщенного метода наименьших квадратов в этой модели. Вычисления требуют обращения матрицы $\hat{\Sigma}$, в которой ij -й элемент имеет вид (10-9). Эта матрица имеет размерность $n \times n$. Она рассчитывается по остаткам наименьших квадратов с использованием

$$\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^T e_t e_t' = \frac{1}{T} E' E, \quad (10-22)$$

где e_t' — вектор размерностью $1 \times n$, включающий все n остатков для n групп в период t , расположенных в t -й строке матрицы остатков E размерностью $T \times n$. Ранг этой матрицы не может превосходить T . Заметим, что происходит, если $n > T$. В этом случае матрица размерностью $n \times n$ имеет ранг T , что меньше n , т.е. она должна быть вырожденной, а оценка FGLS не может быть вычислена. Рассмотрим пример 10.1. Мы объединили 48 штатов в $n = 9$ регионов. Составить полную модель для $n = 48$ штатов лишь по $T = 17$ наблюдениям не представляется возможным. Это следует из ограниченности набора данных, а не самой модели. Матрица Σ генеральной совокупности

положительно определена. Но в случае если наблюдений недостаточно, набор данных будет слишком мал, чтобы получить положительно определенную оценку матрицы.

Пример 10.1. Региональная модель производства для стоимости фондов в общественном секторе

Маннэлл (1990) предложил модель производительности общественного сектора на уровне штатов. Центральное уравнение, которое мы расширим здесь, имеет вид производственной функции Кобба–Дугласа

$$\ln gsp_{it} = \alpha_i + \beta_{1i} \ln pc_{it} + \beta_{2i} \ln hwy_{it} + \beta_{3i} \ln water_{it} + \beta_{4i} \ln util_{it} + \beta_{5i} \ln emp_{it} + \beta_{6i} \ln unemp_{it} + \varepsilon_{it},$$

где переменные модели, измеренные для 48 штатов США в период 1970–1986 гг., таковы:

gsp = валовой продукт штата (gross state product),

pc = частный капитал (private capital),

hwy = стоимость основных фондов в дорожном хозяйстве (highway capital),

$water$ = стоимость основных фондов в водоснабжении (water utility capital),

$util$ = стоимость основных фондов в сфере коммунальных услуг (utility capital),

emp = занятость (employment (labor)),

$unemp$ = уровень безработицы (unemployment rate).

Данные приведены в табл. F10.1 приложения. Мы обозначили девять регионов, включающих группы из 48 штатов:

| | | | | |
|-------|---|---------------------|---|---|
| 1. GF | = | <i>Gulf</i> | = | Алабама, Флорида, Луизиана, Миссисипи |
| 2. MF | = | <i>Midwest</i> | = | Иллинойс, Индиана, Кентукки, Мичиган, Миннесота, Огайо, Висконсин |
| 3. MA | = | <i>Mid Atlantic</i> | = | Делавэр, Мэриленд, Нью-Джерси, Нью-Йорк, Пенсильвания, Вирджиния |
| 4. MT | = | <i>Mountain</i> | = | Колорадо, Айдахо, Монтана, Северная Дакота, Южная Дакота, Вайоминг |
| 5. NE | = | <i>New England</i> | = | Коннектикут, Мэн, Массачусетс, Нью-Хэмпшир, Род-Айленд, Вермонт |
| 6. SO | = | <i>South</i> | = | Джорджия, Северная Каролина, Южная Каролина, Теннесси, Западная Вирджиния, Арканзас |
| 7. SW | = | <i>Southwest</i> | = | Аризона, Невада, Нью-Мексико, Техас, Юта |
| 8. SN | = | <i>Central</i> | = | Аляска, Айова, Канзас, Миссури, Небраска, Оклахома |
| 9. WC | = | <i>West Coast</i> | = | Калифорния, Орегон, Вашингтон |

В данном примере мы будем использовать объединенные данные для анализа модели уравнений по девяти регионам. Данные по выпуску,

основным фондам и занятости получены простым суммированием значений по отдельным штатам (до того, как мы взяли логарифмы). Уровень безработицы для каждого региона m за период t определяется как средневзвешенная величина уровней безработицы для штатов в регионе, где веса составляют

$$w_{it} = emp_{it} / \sum_j emp_{jt}.$$

Затем уровень безработицы по региону m за период t считается как среднее значение уровней безработицы штатов (j) в регионе (m) за период t :

$$unemp_{mt} = \sum_j w_{jt}(m) unemp_{jt}(m).$$

Изначально мы оценивали отдельно каждое из девяти уравнений модели производительности регионов с помощью МНК. МНК-оценки приведены в табл. 10.1. Корреляционная матрица для МНК-остатков имеет вид:

| | GF | MW | MA | MT | NE | SO | SW | CN | WC | |
|-----|----|---------|---------|---------|---------|---------|--------|---------|--------|--------|
| R = | GF | 1,0000 | | | | | | | | |
| | MW | 0,1036 | 1,0000 | | | | | | | |
| | MA | 0,3421 | 0,0634 | 1,0000 | | | | | | |
| | MT | 0,4243 | 0,6970 | -0,0158 | 1,0000 | | | | | |
| | NE | -0,5127 | -0,2896 | 0,1915 | -0,5372 | 1,0000 | | | | |
| | SO | 0,5897 | 0,4893 | 0,2329 | 0,3434 | -0,2411 | 1,0000 | | | |
| | SW | 0,3115 | 0,1320 | 0,6514 | 0,1301 | -0,3220 | 0,2594 | 1,0000 | | |
| | CN | 0,7958 | 0,3370 | 0,3904 | 0,4957 | -0,2980 | 0,8050 | 0,3465 | 1,0000 | |
| | WC | 0,2340 | 0,5654 | 0,2116 | 0,5736 | -0,0576 | 0,2693 | -0,0375 | 0,3818 | 1,0000 |

Значения в R достаточно велики, чтобы предположить, что существует значимая корреляция шоков по регионам.

В табл. 10.1 также представлены оценки ДОМНК (FGLS) параметров модели внешне не связанных регрессий для регионального выпуска. Они вычисляются в два этапа. Результаты МНК-регрессии на первом этапе позволяют получить оценку Σ для FGLS. (Результаты оценки объединенной регрессии, которые также приведены, рассматриваются в подразделе 10.2.8.) Корреляции, перечисленные выше, показывают, что с точки зрения эффективности оценки, по всей видимости, существует значительное преимущество в использовании FGLS. МНК-оценки по отдельным уравнениям состоятельны, однако при этом не учитывается перекрестная корреляция между уравнениями. Существенно меньшие оценки стандартных ошибок при использовании FGLS для каждого уравнения, похоже, подтверждают это.

Мы использовали (10-14) при построении статистики для двух гипотез. Сначала проверили гипотезу о постоянной отдаче от масштаба внутри всей системы. Для постоянной отдачи от масштаба потребуется, чтобы коэффициенты $\beta_1 - \beta_5$ (четыре характеристики основных фондов и показатель занятости) в сумме были равны 1,0. Матрица R размерностью $9 \times 9(7)$ для (10-14) будет содержать строки, равные

$$\begin{aligned} \mathbf{R}_1 &= (0, 1, 1, 1, 1, 1, 0)0'0'0'0'0'0'0'0', \\ \mathbf{R}_2 &= 0'(0, 1, 1, 1, 1, 1, 0)0'0'0'0'0'0'0'0' \end{aligned}$$

и т.д. В (10-14) мы имеем $q' = (1, 1, 1, 1, 1, 1, 1, 1, 1)$. Эта гипотеза накладывает девять ограничений. Рассчитанная статистика хи-квадрат равна 102,305. Критическое значение составило 16,919, поэтому гипотеза отвергается. Вектор разности для этих результатов равен

$$\begin{aligned} (\mathbf{R}\beta - q') &= (-0,64674; -0,12883; 0,96435; 0,03930; 0,06710; 1,79472; \\ &2,30283; 0,12907; 1,10534). \end{aligned}$$

Расстояние по некоторым регионам достаточно велико, поэтому гипотеза о постоянной отдаче от масштаба (в той мере, в которой она значима на уровне объединения), по всей видимости, не соответствует полученным результатам.

Ограничение «объединения» для многофакторной регрессии (для тех же переменных, но необязательно для тех же данных, как в нашем примере) формулируется как

$$H_0: \beta_1 = \beta_2 = \dots = \beta_M, \quad H_1: H_0 \text{ неверна.}$$

Для этой гипотезы матрица \mathbf{R} приведена в (10-15), проверочная статистика — в (10-14). Для нашей модели с девятью уравнениями и семью параметрами в каждом уравнении нулевая гипотеза налагает $8(7) = 56$ ограничений. Рассчитанная статистика составила 10 554,77, что намного превосходит критическое значение из таблицы, равное 74,468. Поэтому гипотеза об однородности отвергается.

Как отмечалось в подразделе 10.2.7, не существует стандартного теста на спецификацию для систем внешне не связанных уравнений против альтернативной гипотезы об отсутствии корреляции между шоками для общей модели внешне не связанных регрессий без предположения о нормальности. Тем не менее тест множителей Лагранжа, предложенный Бройшем и Пэганом (Breusch, Pagan (1980)) и основанный на корреляционной матрице, имеет некую интуитивную привлекательность. Мы использовали (10-17) для вычисления критерия множителей Лагранжа для модели с девятью уравнениями, приведенной в табл. 10.1. Для корреляционной матрицы, приведенной ранее, статистика хи-квадрат с $8(9)/2 = 36$ степенями свободы равна 102,305. Критическое значение по таблице распределения хи-квадрат равно 50,998, поэтому нулевая гипотеза о том, что внешне не связанные регрессии действительно не связаны, отвергается. Мы заключаем, что шоки в региональной модели на самом деле не являются несвязанными. Нулевая гипотеза о том, что $\sigma_{ij} = 0$ для всех $i \neq j$, отвергается. Для более глубокого анализа мы повторили тест с полностью дезагрегированными (в масштабе штата) данными. Соответствующая статистика хи-квадрат с $48 \times (47)/2 = 1128$ степенями свободы равна 8399,41. Критическое значение равно 1207,25, поэтому нулевая гипотеза отвергается и на уровне штатов.

10.3. Обобщенные модели внешне не связанных регрессий

В принципе случайные ошибки в SUR-модели могут быть гетероскедастичными и автокоррелированными. Бартелс, Фибиг (Bartels, Fiebig (1992)) предложили использовать обобщенную модель внешне не связанных регрессий $\Omega = A[\Sigma \otimes I]A'$, где A — блочно-диагональная матрица. В идеале A становится функцией измеренных характеристик индивидуального и отдельного вектора параметров θ , поэтому модель может быть оценена поэтапно. На первом шаге могут быть использованы МНК-остатки, чтобы сформировать предварительную оценку θ , а далее данные могут быть пересчитаны к гомоскедастичному виду, при этом Σ и β остаются для оценки на последующих этапах с использованием этих новых данных. Одно из практических применений в этом направлении — модель со случайными параметрами, предложенная в работе Фибиг, Бартелс, Эйгнер (Fiebig, Bartels, Aigner (1991)); (9-50) показывает, как модель со случайными параметрами приводит к гетероскедастичности. Другое приложение предложили Мэнди, Мартинс-Фило (Mandy, Martins-Filho (1993)), которые определили $\sigma_{ij}(t) = z_{ij}(t)' \alpha_{ij}$. (Линейная спецификация дисперсии приводит к ряду проблем, так как отрицательное значение не устраняется.) В работе (Kumbhakar, Heshmati (1996)) предложена система издержек и спроса, которая сочетает транслогарифмическую модель из подраздела 10.4.2 с системой полных уравнений из 10.4.1. В этом практическом применении лишь уравнение издержек было специфицировано с включением гетероскедастичных шоков.

Автокорреляция шоков в регрессионных моделях обычно возникает как характерная черта моделей временных рядов, т.е. это одно из неотъемлемых свойств временных рядов. (Мы подробно изучим этот вопрос спецификации моделей в главе 20.) В системах уравнений, рассматриваемых в данной главе, свойства временных рядов для данных, как правило, не являются главным предметом изучения. Основное преимущество спецификации систем внешне не связанных уравнений — возможность анализа корреляции между наблюдениями в определенный момент. Зачастую спецификации панельных данных, как в примерах 3 и 4 из вводной части главы, могут быть также изучены в рамках модели внешне не связанных регрессий из этой главы. В этих случаях могут наблюдаться устойчивые эффекты в шоках, но здесь опять же такие эффекты часто рассматриваются в качестве последствий присутствия скрытой, не зависящей от времени гетерогенности. Тем не менее, так как системы уравнений, изучаемые в данной главе, часто содержат умеренно длинные временные ряды, уместным будет несколько более формально подойти к рассмотрению автокорреляции. Мнения о целесообразности «корректировки» автокорреляции расходятся. На одном полюсе находится Мизон (Mizon (1995)), который считает, что автокорреляция возникает как следствие поправимой неспособности включить в модель динамические эффекты. Однако в системах уравнений анализ, приводящий к такому выводу, по всей видимости, будет намного более сложным, нежели в модели с одним уравнением¹⁶.

¹⁶Динамические модели внешне не связанных регрессий в духе наставления Мизона были

Достаточно лишь сказать, что этот вопрос остается открытым до сих пор.

10.4. Нелинейные системы уравнений

Теперь мы рассмотрим методы оценки нелинейных систем уравнений. Лежащая в их основе теория, по существу, такая же, как и для линейных систем. Исходя из этого большая часть раздела будет описывать практические аспекты оценивания. Рассмотрим оценивание параметров следующей системы уравнений:

$$\begin{aligned} y_1 &= \mathbf{h}_1(\boldsymbol{\beta}, \mathbf{X}) + \varepsilon_1, \\ y_2 &= \mathbf{h}_2(\boldsymbol{\beta}, \mathbf{X}) + \varepsilon_2, \\ &\vdots \\ y_M &= \mathbf{h}_M(\boldsymbol{\beta}, \mathbf{X}) + \varepsilon_M. \end{aligned} \quad (10-23)$$

[Обратите внимание на аналогию с (10-19).]

Здесь всего M уравнений, которые следует оценить по $t = 1, \dots, T$ наблюдениям. В модели K параметров. Предположение о том, что каждое уравнение имеет «собственный» вектор параметров, не делается; мы просто используем некоторые из них или все K элементов в $\boldsymbol{\beta}$ в каждом уравнении. Также есть набор T наблюдений по каждой из P независимых переменных $x_p, p = 1, \dots, P$, некоторые из них или все присутствуют в каждом уравнении. Для удобства уравнения записаны в общем виде с полными $\boldsymbol{\beta}$ и \mathbf{X} . Предполагается, что шоки имеют нулевые средние значения и одновременную ковариационную матрицу $\boldsymbol{\Sigma}$. Мы не будем рассматривать наличие автокорреляции, оставив этот случай для более продвинутых исследований.

В многомерной регрессионной модели если $\boldsymbol{\Sigma}$ известна, то оценка вектора $\boldsymbol{\beta}$, полученная при помощи обобщенного метода наименьших квадратов, будет вектором, который минимизирует обобщенную сумму квадратов остатков:

$$\boldsymbol{\varepsilon}(\boldsymbol{\beta})' \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon}(\boldsymbol{\beta}) = \sum_{i=1}^M \sum_{j=1}^M \sigma^{ij} [y_i - \mathbf{h}_i(\boldsymbol{\beta}, \mathbf{X})]' [y_j - \mathbf{h}_j(\boldsymbol{\beta}, \mathbf{X})], \quad (10-24)$$

где $\boldsymbol{\varepsilon}(\boldsymbol{\beta})$ — вектор шоков $MT \times 1$, полученный объединением уравнений, $\boldsymbol{\Omega} = \boldsymbol{\Sigma} \otimes \mathbf{I}$ и σ^{ij} — ij -й элемент $\boldsymbol{\Sigma}^{-1}$. [См. (10-7).] Как и в подразделе 7.2.3,

предложены Андерсоном и Бланделлом (Anderson, Blundell (1982)). Несколько последних их применений представили Кивье, Филлипс, Шип (Kiviet, Phillips, Schipp (1995)) и Дешамп (DesChamps (1998)). Однако относительно немного работ было посвящено именно динамическим моделям внешне не связанных регрессий. VAR-модели составляют важную группу, но они исходят из других аналитических предпосылок. Подобно этому приложения к анализу панельных данных, отмеченные во введении и подразделе 11.8.3, будут вписываться в рамки моделей, рассматриваемых здесь. Однако в этих приложениях регрессии будут действительно независимыми — авторы не моделировали корреляцию между элементами, что является центральным моментом в этой главе. Схожие результаты можно найти у Guilkey, Schmidt (1973), Guilkey (1974), Berndt, Savin (1977), Moschino, Moro (1994), McLaren (1996) и Holt (1998).

Таблица 10.1. Оценки системы внешне не связанных уравнений для регионального выпуска (в скобках приведены стандартные ошибки)

| Регион | Метод оценки | α | β_1 | β_2 | β_3 | β_4 | β_5 | β_6 | σ_m | R^2 |
|--------|-----------------|----------------------|-------------------------|----------------------|----------------------|---------------------|---------------------|------------------------|------------|--------|
| GF | OLS | 12,1458 (3,3154) | -0,007117 (0,01114) | -2,1352 (0,8677) | 0,1161 0,06278 | 1,4247 (0,5944) | 0,7851 (0,1493) | -0,00742 0,00316 | 0,01075 | 0,9971 |
| | FGLS | 10,4792 (1,5912) | -0,003160 (0,005391) | -1,5448 (0,3888) | 0,1139 (0,03651) | 0,8987 (0,2516) | 0,8886 (0,07715) | -0,005299 0,00182 | 0,008745 | 0,9967 |
| MW | OLS | 3,0282 (1,7835) | 0,1635 (0,1660) | -0,07471 (0,2205) | -0,1689 (0,09896) | 0,6372 (0,2078) | 0,3622 (0,1650) | -0,01736 0,004741 | 0,009942 | 0,9984 |
| | FGLS | 4,1206 (1,0091) | 0,06370 (0,08739) | -0,1275 (0,1284) | -0,1292 (0,06152) | 0,5144 (0,1118) | 0,5497 (0,08597) | -0,01545 0,00252 | 0,008608 | 0,9980 |
| MA | OLS | -11,2110 (3,5867) | 0,4120 (0,2281) | 2,1355 (0,5571) | 0,5122 (0,1192) | -0,4740 (0,2519) | -0,4620 (0,3529) | -0,03022 0,00853 | 0,01040 | 0,9950 |
| | FGLS | -9,1438 (2,2025) | 0,3511 (0,1077) | 1,7972 (0,3410) | 0,5168 (0,06405) | -0,3616 (0,1294) | -0,3391 (0,1997) | -0,02954 0,00474 | 0,008625 | 0,9946 |
| MT | OLS | 3,5902 (6,9490) | 0,2949 (0,2054) | 0,1740 (0,2082) | -0,2257 (0,3840) | -0,2144 (0,9712) | 0,9166 (0,3772) | -0,008143 0,00839 | 0,01688 | 0,9940 |
| | FGLS | 2,8150 (3,4428) | 0,1843 (0,09220) | 0,1164 (0,1165) | -0,3811 (0,1774) | 0,01648 (0,4654) | 1,1032 (0,1718) | -0,005507 (0,00422) | 0,01321 | 0,9938 |
| | OLS | 6,3784 (2,3823) | -0,1526 (0,08403) | -0,1233 (0,2850) | 0,3065 (0,08917) | -0,5326 (0,2375) | 1,3437 (0,1876) | 0,005098 (0,00517) | 0,008601 | 0,9986 |

Окончание табл. 10.1

| Регион | Метод оценки | α | β_1 | β_2 | β_3 | β_4 | β_5 | β_6 | σ_m | R^2 |
|--------|-----------------|-----------------------|-----------------------|------------------------|----------------------|----------------------|---------------------|------------------------|------------|---------------------|
| NE | FGLS | 3,5331 (1,3388) | -0,1097 (0,04570) | 0,1637 (0,1676) | 0,2459 (0,04974) | -0,3155 (0,1194) | 1,0828 (0,09248) | -0,000664 (0,00263) | 0,007249 | 0,9983 |
| | OLS | -13,7297 (18,0199) | -0,02041 (0,2856) | 0,6621 (1,8111) | -0,9693 (0,2843) | -0,1074 (0,5634) | 3,3803 (1,1643) | 0,03378 (0,02150) | 0,02241 | 0,9852 |
| SO | FGLS | -13,1186 (7,6009) | 0,1007 (0,1280) | 0,9923 (0,7827) | -0,5851 (0,1373) | -0,3029 (0,2412) | 2,5897 (0,4665) | 0,02143 (0,00809) | 0,01908 | 0,9817 |
| | OLS | -22,8553 (4,8739) | -0,3776 (0,1673) | 3,3478 (1,8584) | -0,2637 (0,4317) | -1,7783 (1,1757) | 2,6732 (1,0325) | 0,02592 (0,01727) | 0,01293 | 0,9864 |
| SW | FGLS | -19,9917 (2,8649) | -0,3386 (0,08943) | 3,2821 (0,8894) | -0,1105 (0,1993) | -1,7812 (0,5609) | 2,2510 (0,4802) | 0,01846 (0,00793) | 0,01055 | 0,9846 |
| | OLS | 3,4425 (1,2574) | 0,05040 (0,2662) | -0,5938 (0,3219) | 0,06351 (0,3333) | -0,01294 (0,3787) | 1,5731 (0,4125) | 0,006125 (0,00892) | 0,01753 | 0,9937 |
| CN | FGLS | 2,8172 (0,8434) | 0,01412 (0,8833) | -0,5086 (0,1869) | -0,02685 (0,1405) | 0,1165 (0,1774) | 1,5339 (0,1762) | 0,006499 (0,00421) | 0,01416 | 0,9930 |
| | OLS | -9,1108 (3,9704) | 0,2334 (0,2062) | 1,6043 (0,7449) | 0,7174 (0,1613) | -0,3563 (0,3153) | -0,2592 (0,3029) | -0,03416 (0,00629) | 0,01085 | 0,9895 |
| WC | FGLS | -10,2989 (2,4189) | 0,03734 (0,1107) | 1,8176 (0,4503) | 0,6572 (0,1011) | -0,4358 (0,1912) | 0,02904 (0,1828) | -0,02867 (0,00373) | 0,008837 | 0,9881 |
| | OLS | 3,1567 (0,1377) | 0,08692 (0,01058) | -0,02956 (0,03405) | 0,4922 (0,04167) | 0,06092 (0,03833) | 0,3676 (0,04018) | -0,01746 (0,00304) | 0,05558 | 0,9927 |
| Pooled | FGLS | 3,1089 (0,0208) | 0,08076 (0,005148) | -0,01797 (0,006186) | 0,3728 (0,01311) | 0,1221 (0,00557) | 0,4206 (0,01442) | -0,01506 (0,00101) | NA | 0,9882 ^a |
| | OLS | 3,0977 (0,1233) | 0,08646 (0,01144) | -0,02141 (0,02830) | 0,03874 (0,03529) | 0,1215 (0,02805) | 0,4032 (0,03410) | -0,01529 (0,00256) | NA | 0,9875 ^a |

^a R^2 модели, оцененной методом FGLS, вычислен с использованием $1 - 9/\text{tr}(\mathbf{S}^{-1}\mathbf{S}_{yy})$.

найдем псевдорегрессоры как производные функций $h(\beta, X)$ по β , т.е. приведем каждое уравнение к линейному виду. Тогда условие первого порядка для минимизации этой суммы квадратов имеет вид

$$\frac{\partial \epsilon(\beta)' \Omega^{-1} \epsilon(\beta)}{\partial \beta} = \sum_{i=1}^M \sum_{j=1}^M \sigma^{ij} [2X_i^0(\beta)' \epsilon_j(\beta)] = 0, \quad (10-25)$$

где $X_i^0(\beta)$ — матрица псевдорегрессоров размерностью $T \times K$, полученная линеаризацией i -го уравнения. (См. подраздел 7.2.6.) Если какой-либо из параметров β отсутствует в i -м уравнении, то соответствующий столбец $X_i^0(\beta)$ будет столбцом, состоящим из нулей.

Проблема оценивания нелинейной системы сложна вдвойне. Во-первых, почти при любых обстоятельствах ее решение потребует использования итеративной процедуры и одного из методов, описанных в приложении Е. Во-вторых, Σ неизвестна и должна быть оценена. Важно помнить, что эффективное оценивание многомерной регрессионной модели не требует эффективности оценки Σ , а только ее состоятельности. Исходя из этого один подход будет заключаться в оценивании параметров каждого уравнения по отдельности с помощью метода нелинейных наименьших квадратов. Этот метод становится неэффективным, если какие-либо из уравнений включают одинаковые параметры, так как в этом случае такая информация будет игнорироваться. Но на данном этапе важна состоятельность, а не эффективность. Полученные остатки могут использоваться для вычисления

$$S = \frac{1}{T} E'E. \quad (10-26)$$

Второй шаг в методе FGLS — решение (10-25), для чего снова потребуется итерационная процедура, и это может строиться на S вместо Σ . При доброкачественных псевдорегрессорах оценка на втором шаге становится полностью эффективной. И опять теория, использованная для FGLS в линейном случае с одним уравнением, применяется и здесь¹⁷. Когда мы получаем FGLS-оценку, соответствующая асимптотическая ковариационная матрица оценивается с помощью

$$\text{Est. Asy. Var}[\hat{\beta}] = \left[\sum_{i=1}^M \sum_{j=1}^M s^{ij} X_i^0(\beta)' X_j^0(\beta) \right]^{-1}. \quad (10-27)$$

В описанной стратегии присутствует возможный изъян. Может быть, будет невозможно оценить отдельно все уравнения при помощи нелинейного метода наименьших квадратов. Возможно, что идентификация некоторых параметров требует совместной оценки нескольких уравнений. Но до тех пор пока полная система идентифицирует все параметры, это представляет

¹⁷ Ни нелинейность, ни наличие нескольких уравнений в этой модели не ставят новые статистические вопросы. При объединении уравнений мы видим, что эта модель представляет собой лишь вариант нелинейной регрессионной модели с усложнением в виде не скалярной ковариационной матрицы шоков, которую мы рассматривали в главе 9. Новые усложнения носят в первую очередь практический характер.

собой простой способ решения проблемы. Напомним, что все, что нам нужно на первом шаге, — это состоятельность оценок элементов β . Легко показать, что все это определяет GMM-оценку (См. главу 13.) Мы можем использовать этот результат для разработки альтернативной, более простой стратегии. Взвешивание сумм квадратов и скрещенных произведений в (10-24) по σ^{ij} позволяет получить эффективную оценку для β . Любое другое взвешивание, основанное на некоторой положительно определенной матрице A , дало бы состоятельные, хотя и неэффективные, оценки. На данном этапе эффективность отходит на второй план, поэтому удобно выбрать $A = I$. Таким образом, на нашем первом шаге мы можем найти β , чтобы минимизировать

$$\varepsilon(\beta)' \varepsilon(\beta) = \sum_{i=1}^M [y_i - h_i(\beta, X)]' [y_i - h_i(\beta, X)] = \sum_{i=1}^M [y_{it} - h_i(\beta, x_{it})]^2.$$

(Эта оценка представляет собой просто оценку нелинейным методом наименьших квадратов обобщенной (сквозной) регрессии, где функция регрессии изменяется по набору наблюдений.) Этот шаг позволит получить оценку вектора $\hat{\beta}$, которая необходима нам для вычисления S .

10.5. Системы уравнений спроса: вырожденные системы

Большая часть упомянутых примеров приложений многомерной регрессионной модели¹⁸ относится к **системам уравнений спроса**: спроса либо на продукты потребления, либо на факторы производства в контексте производства.

Пример 10.2. Система расходов Стоуна

Система расходов Стоуна¹⁹, основанная на наборе логарифмических уравнений спроса на товар, доходе Y и ценах на продукты потребления p_n , имеет следующий вид:

$$\log q_i = \alpha_i + \beta_i \log \left(\frac{Y}{P} \right) + \sum_{j=1}^M \eta_{ij}^* \log \left(\frac{p_j}{P} \right),$$

где P — обобщенный (взвешенный по долям) индекс цен, η_i — эластичность дохода и η_{ij}^* — компенсированная эластичность цен. Мы

¹⁸Обратите внимание на различие между многомерной моделью, или моделью с несколькими уравнениями, рассматриваемой здесь, и моделью множественной регрессии.

¹⁹Очень легко читаемый обзор методов оценки систем спроса на товары представлен в работе Дитона и Мюэльбауэра (Deaton, Muellbauer (1980)). Пример, упомянутый здесь, взят из главы 3 и ссылок на работы Стоуна (Stone (1954a,b)), приведенных в ней. Дитон (Deaton (1986)) предлагает другой полезный обзор. Приложение к построению моделей производственной функции можно найти у (Chambers (1988)). Другие исследования по теме спецификации систем уравнений спроса включают работы (Chavez, Segerson (1987)), (Brown, Walker (1995)) и (Fry, Fry, McLaren (1996)).

можем интерпретировать эту систему как уравнение спроса по реальным расходам и реальным ценам. Полученный набор уравнений образует эконометрическую модель в виде системы нескольких внешне не связанных регрессий. При оценке мы должны учитывать ряд ограничений, включая однородность первой степени в доходах, $\sum_i S_i \eta_i = 1$ и симметричность матрицы компенсированных эластичностей цен, $\eta_{ij}^* = \eta_{ji}^*$, где S_i — доля товара i в бюджете.

Другие примеры включают систему спроса на факторы и доли факторных затрат в производстве, чего мы снова коснемся далее. По сути, каждый из них лишь частный случай приложения модели из раздела 10.2. Но при таких условиях возникает ряд особых проблем. Во-первых, параметры этих систем ограничены между уравнениями, т.е. неограниченная модель не согласуется с центральной теорией²⁰. Примером могут служить многочисленные ограничения в системе уравнений спроса, приведенной выше. Вторая существенная черта многих таких моделей состоит в том, что ковариационная матрица шоков Σ является вырожденной²¹.

10.5.1. Функция издержек Кобба–Дугласа

Рассмотрим производственную функцию Кобба–Дугласа

$$Q = \alpha_0 \prod_{i=1}^M x_i^{\alpha_i}.$$

Максимизация прибыли с экзогенно определенной ценой на выпуск требует от фирмы максимизировать выпуск для данного уровня издержек C (или минимизировать затраты на данный выпуск Q). Лагранжиан задачи максимизации имеет вид

$$\Lambda = \alpha_0 \prod_{i=1}^M x_i^{\alpha_i} + \lambda(C - p'x),$$

где p — вектор, состоящий из M цен на факторы. Обязательными для максимизации этой функции являются условия

$$\frac{\partial \Lambda}{\partial x_i} = \frac{\alpha_i Q}{x_i} - \lambda p_i = 0 \quad \text{и} \quad \frac{\partial \Lambda}{\partial \lambda} = C - p'x = 0.$$

²⁰Такая непостоянность не предполагает, что теоретические ограничения непроверяемы или что неограниченная модель не может быть оценена. В некоторых случаях, однако, смысл модели становится неясен без ограничений. Статистическое отрицание ограничений, которые предполагает теория и которые использовались для получения в первую очередь эконометрических моделей, может поставить нас в не очень удобную ситуацию. К примеру, при исследованиях функций полезности (Christensen, Jorgenson, Lau (1975)) после отвержения гипотезы о наличии симметрии показателей спроса на товары по уравнениям заключили: «С таким выводом мы можем завершить проверку, так как эти результаты признают несостоятельность теории спроса» (с. 380). См. Silver и Ali (1989) для изучения методов тестирования симметричных ограничений. Модель может противоречить теории и по другим направлениям. Так, линейная система расходов Стоуна в примере 10.7 не согласуется с какой-либо теоретически верной функцией полезности. См. Goldberger (1987).

²¹В работе (Denton (1978)) проверяется несколько таких случаев.

Совместное решение дает $x_i(Q, p)$ и $\lambda(Q, p)$. Совокупные издержки составляют

$$\sum_{i=1}^M p_i x_i = \sum_{i=1}^M \frac{\alpha_i Q}{\lambda}.$$

Доля издержек i -го фактора равна

$$\frac{p_i x_i}{\sum_{i=1}^M p_i x_i} = \frac{\alpha_i}{\sum_{i=1}^M \alpha_i} = \beta_i. \quad (10-28)$$

Полная модель имеет вид²²

$$\begin{aligned} \ln C &= \beta_0 + \beta_q \ln Q + \sum_{i=1}^M \beta_i \ln p_i + \varepsilon_c, \\ s_i &= \beta_i + \varepsilon_i, \quad i = 1, \dots, M. \end{aligned} \quad (10-29)$$

По построению $\sum_{i=1}^M \beta_i = 1$ и $\sum_{i=1}^M s_i = 1$. (Анализ такой функции издержек мы начали в примере 6.6. Мы вернемся к этому примеру ниже.) Доли издержек также будут в сумме равны единице в этом случае. Из этого следует, что $\sum_{i=1}^M \varepsilon_i = 0$ для каждой точки, т.е. система вырождена. В данный момент будем пренебрегать функцией издержек. Пусть вектор шоков размерностью $M \times 1$ из наших долей равен $\varepsilon = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_M]'$. Так как $\varepsilon' \mathbf{i} = 0$, где \mathbf{i} — столбец из единиц, получаем, что $E[\varepsilon \varepsilon' \mathbf{i}] = \Sigma \mathbf{i} = \mathbf{0}$, а это подразумевает вырожденность Σ . Таким образом, методы из предыдущих разделов не могут быть использованы в этом случае. (Следует проверить, что выборочная ковариационная матрица МНК-остатков также будет вырождена.)

Как представляется, решение проблемы вырожденности — выкинуть из системы одно из уравнений, оценить оставшуюся часть и рассчитать последний параметр из других $M - 1$. Ограничение $\sum_{i=1}^M \beta_i = 1$ предполагает, что функция затрат должна быть гомогенной первой степени по ценам исходя из теоретической необходимости. Если мы введем ограничение

$$\beta_M = 1 - \beta_1 - \beta_2 - \dots - \beta_{M-1}, \quad (10-30)$$

то система сведется к невырожденной:

$$\begin{aligned} \ln \left(\frac{C}{p_M} \right) &= \beta_0 + \beta_q \ln Q + \sum_{i=1}^{M-1} \beta_i \ln \left(\frac{p_i}{p_M} \right) + \varepsilon_c, \\ s_i &= \beta_i + \varepsilon_i, \quad i = 1, \dots, M - 1. \end{aligned}$$

Эта система дает оценки β_0, β_q и $\beta_1, \dots, \beta_{M-1}$. Последний параметр оценивается с использованием (10-30). Не имеет значения, какой фактор выбран в качестве эталонного. Обе оценки, FGLS и **максимального правдоподобия**, которые можно получить как итеративный FGLS или прямо оценивая функцию максимального правдоподобия, **неизменны** по отношению к тому, какой фактор выбран как эталонный²³.

²²Мы оставляем в качестве упражнения вывод β_0 , который является смесью всех параметров, и β_q , который равен $1/\sum_m \alpha_m$.

²³Неизменность результата доказана в работе (Barten (1969)). Некоторые дополнительные результаты по этому методу даны в работах (Revankar (1976)), Deaton (1986), Powell (1969) и McGuire et al. (1968).

Таблица 10.2. Оценки регрессий (в скобках даны стандартные ошибки)

| | Метод наименьших квадратов | | | | Многомерная регрессия | | | |
|--------------|----------------------------|----------|--------|-----------|-----------------------|-----------|--------|-----------|
| β_0 | -4,686 | (0,885) | -3,764 | (0,702) | -7,069 | (0,107) | -5,707 | (0,165) |
| β_q | 0,721 | (0,0174) | 0,153 | (0,0618) | 0,766 | (0,0154) | 0,238 | (0,0587) |
| β_{qq} | — | | 0,0505 | (0,00536) | — | | 0,0451 | (0,00508) |
| β_k | -0,00847 | (0,191) | 0,0739 | (0,150) | 0,424 | (0,00946) | 0,424 | (0,00944) |
| β_l | 0,594 | (0,205) | 0,481 | (0,161) | 0,106 | (0,00386) | 0,106 | (0,00382) |
| β_f | 0,414 | (0,0989) | 0,445 | (0,0777) | 0,470 | (0,0101) | 0,470 | (0,0100) |
| R^2 | 0,9316 | — | 0,9581 | — | — | — | — | — |

Исследование отрасли электроэнергетики, проведенное Нерловым (Nerlove (1963)), которое мы рассматривали в примере 6.6, является примером практического применения модели функции издержек Кобба–Дугласа. Его оценки параметров по методу наименьших квадратов были приведены в примере 6.6. В этих результатах присутствует (к сожалению) отрицательный коэффициент капитала в трех из шести регрессий. Нерлов также обнаружил, что простая модель Кобба–Дугласа не объяснила в полной мере отношение между выпуском и средней стоимостью. Кристенсен и Грин (Christensen, Greene (1976)) глубже проанализировали данные Нерлова и добавили в набор данные о долях издержек, чтобы оценить полную **систему спроса**. В табл. F6.2 приложения приведены 145 наблюдений Нерлова вместе с данными Кристенсена и Грина о долях издержек. Издержки — это общая стоимость сгенерированной энергии в миллионах долларов, данные по выпуску приведены в миллионах киловатт-часов, стоимость капитала — это индекс расходов на строительство, ставка заработной платы дана в долларах в час в производстве и выработке электроэнергии, цена на топливо — это индекс затрат за БТЕ (британские тепловые единицы) топлива, купленного фирмами, и данные отражают стоимость продукции в 1955 г. Оценки регрессии приводятся в табл. 10.2.

Оценки функции издержек Кобба–Дугласа при помощи метода наименьших квадратов приведены в первом столбце²⁴. Коэффициент при стоимости капитала отрицательный. Так как $\beta_i = \beta_q \partial \ln Q / \partial \ln x_i$ — иными словами, это положительное число, кратное эластичности выпуска i -го фактора, — то этот факт вызывает беспокойство. В третьем столбце представлены FGLS-оценки с ограничениями. Для получения оценки с ограничениями мы придаем модели форму оценки объединенной системы внешне не связанных регрессий в (10-19):

$$y = \begin{bmatrix} \ln(C/P_f) \\ s_k \\ s_l \end{bmatrix} = \begin{bmatrix} \mathbf{i} & \ln Q & \ln(P_k/P_f) & \ln(P_l/P_f) \\ 0 & 0 & \mathbf{i} & 0 \\ 0 & 0 & 0 & \mathbf{i} \end{bmatrix} \begin{pmatrix} \beta_0 \\ \beta_q \\ \beta_k \\ \beta_l \end{pmatrix} + \begin{bmatrix} \varepsilon_c \\ \varepsilon_k \\ \varepsilon_l \end{bmatrix}.$$

[Здесь $3 \cdot (145) = 435$ наблюдений в матрицах данных.] Тогда оценка становится FGLS, как показано в (10-21). Дополнительный столбец добавлен для

²⁴Результаты, основанные на полных данных Нерлова, приведены в примере 6.6.

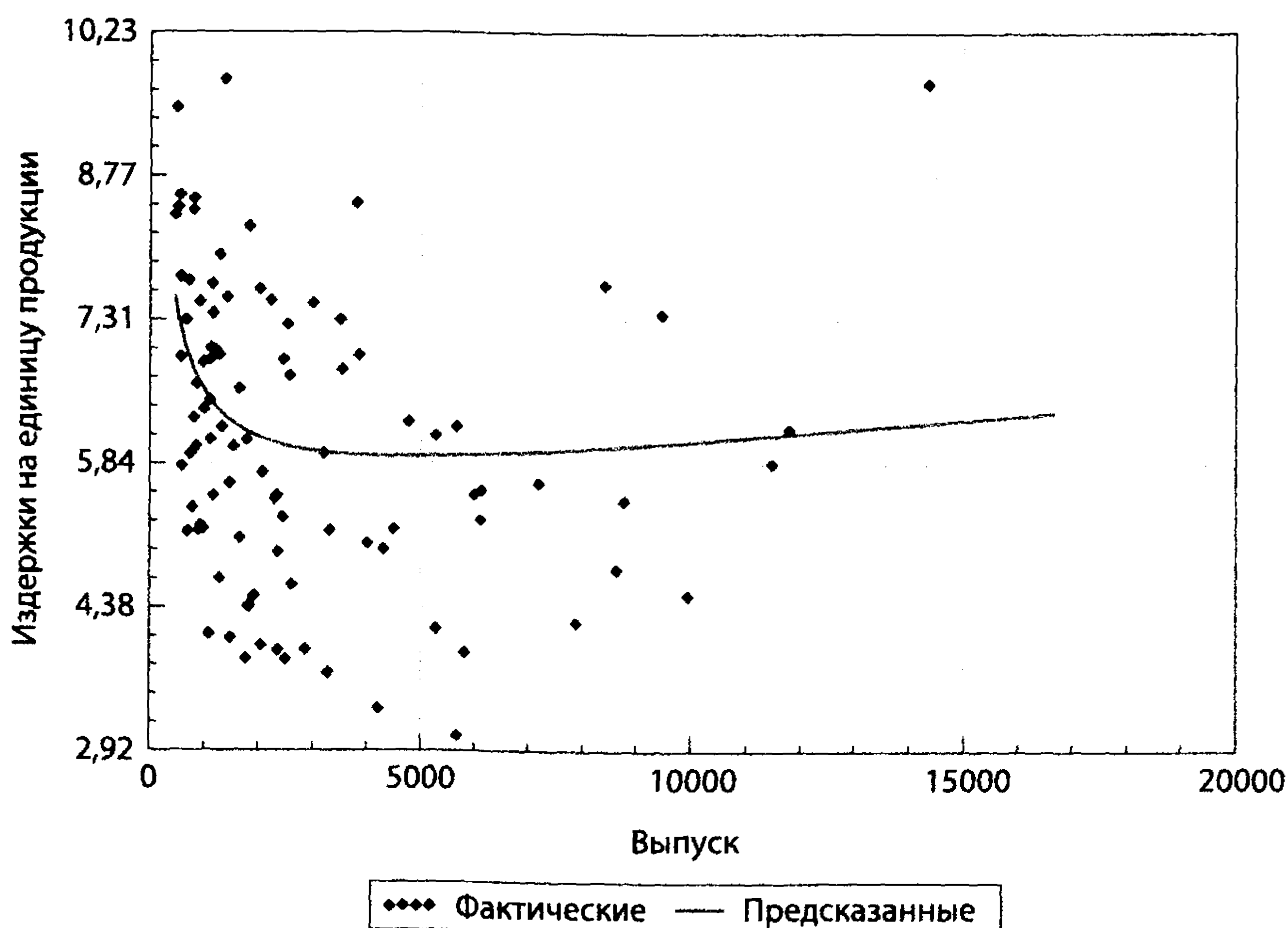


Рис. 10.1. Предсказанные и фактические средние издержки

логарифмической квадратичной модели. Важно отметить две вещи: значительно меньшие стандартные ошибки и теперь положительную (и приемлемую) оценку коэффициента при стоимости капитала. Оценки экономии от масштаба производства в базовой модели Кобба–Дугласа составляют $1/\beta_q = 1,39$ (столбец 1) и $1,31$ (столбец 3), что предполагает некоторую возрастающую отдачу от масштаба. Нерлов, однако, обнаружил доказательство того, что при чрезвычайно большом масштабе фирмы экономия от масштаба уменьшается и в конечном счете исчезает. Для объяснения этого (по сути, классической U-образной кривой средних затрат) он добавил квадрат логарифма выпуска в функцию издержек. Оценки одного уравнения и многомерной модели приведены во втором и четвертом множествах результатов.

Квадрат выпуска придает функции издержек ожидаемую U-форму. Мы можем определить точку, где средние затраты достигают минимума, приравняв $\partial \ln C / \partial \ln Q$ к 1. Это точка $Q^* = \exp[(1 - \beta_q) / (2\beta_{qq})]$. Для многомерной регрессии это значение равно $Q^* = 4665$. Примерно 85% фирм в выборке имели выпуск меньший, чем этот, поэтому по этим оценкам большее число фирм в выборке еще не использовали доступную им экономию от масштаба. На рис. 10.1 показаны предсказанные моделью и фактические средние издержки для нашей выборки. (Чтобы получить приемлемый масштаб, треть наиболее мелких фирм не изображена на графике.) Предсказанные средние издержки вычисляются по выборочным средним значениям цен факторов. График показывает, что за пределами достаточно малых фирм экономия от масштаба хоть статистически и значима, но достаточно мала с экономической точки зрения.

10.5.2. Гибкие функциональные формы: транслогарифмическая функция издержек

Академическая литература о производстве и затратах и о полезности и спросе развивалась по нескольким направлениям. В области моделей поведения производителя классическая работа Эрроу и др. (Arrow et al. (1961)) поставила под вопрос неотъемлемое ограничение модели Кобба–Дугласа о том, что все эластичности замещения факторов производства равны единице. С этого времени ученые разработали многочисленные **гибкие функции**, позволяющие эластичности замещения быть неограниченной (т.е. даже непостоянной)²⁵. Схожие направления в литературе появились и в анализе спроса на товары²⁶. В этом подразделе мы подробно рассматриваем модель производства.

Предположим, что производство описывается производственной функцией $Q = f(x)$. Решение проблемы минимизации издержек производства указанного уровня выпуска при данном наборе цен факторов дает набор показателей спроса на факторы, минимизирующие издержки, $x_i = x_i(Q, p)$. Общие издержки производства задаются функцией издержек

$$C = \sum_{i=1}^M p_i x_i(Q, p) = C(Q, p). \quad (10-31)$$

Если существует **постоянная отдача от масштаба**, то выходит, что $C = Qc(p)$, или $C/Q = c(p)$, где $c(p)$ — это некая единица, или функция, средних затрат²⁷. Спрос на факторы, минимизирующие издержки, получаем, применяя **лемму Шепарда** (Shephard, 1970), которая гласит, что если $C(Q, p)$ дает минимальную общую стоимость производства, то минимизирующий набор показателей спроса на факторы следует из

$$x_i^* = \frac{\partial C(Q, p)}{\partial p_i} = \frac{Q \partial c(p)}{\partial p_i}. \quad (10-32)$$

В качестве альтернативы, дифференцируя логарифмы, мы получаем минимизирующие издержки факторов производства:

$$s_i = \frac{\partial \ln C(Q, p)}{\partial \ln p_i} = \frac{p_i x_i}{C}. \quad (10-33)$$

При постоянной отдаче от масштаба $\ln C(Q, p) = \ln Q + \ln c(p)$ получаем

$$s_i = \frac{\partial \ln c(p)}{\partial \ln p_i}. \quad (10-34)$$

²⁵Berndt, Christensen (1973). Два полезных обзора по этой теме были предложены Jorgenson (1983) и Diewert (1974).

²⁶Christensen, Jorgenson, Lau (1975) и два обзора, Deaton, Muellbauer (1980) и Deaton (1983). В работе Berndt (1990) содержится много интересных результатов.

²⁷Функция Кобба–Дугласа из предыдущего подраздела иллюстрирует это. Ограничением постоянной отдачи от масштаба является $\beta_q = 1$, что тождественно $C = Qc(p)$. Более общая версия функции издержек у Нерлова дает непостоянную отдачу от масштаба. См. у Christensen, Greene (1976) и Diewert (1974) о некоторых свойствах функции издержек и ее связи со структурой производства.

Во многих практических исследованиях объектами оценки являются эластичности замещения факторов и собственно эластичность цен спроса, что задается через

$$\theta_{ij} = \frac{c(\partial^2 c / \partial p_i \partial p_j)}{(\partial c / \partial p_i)(\partial c / \partial p_j)}$$

и

$$\eta_{ii} = s_i \theta_{ii}.$$

С помощью соответствующей параметризации функции затрат (10-31) и долей затрат (10-34) мы получим эконометрическую модель с M или $M + 1$ уравнениями, которую можно использовать для оценивания этих величин²⁸.

Трансцендентная логарифмическая или **транслогарифмическая функция** — наиболее часто используемая гибкая функция в эмпирических исследованиях²⁹. Раскладывая $\ln c(p)$ в **ряд Тейлора** второго порядка в окрестности точки $\ln p = 0$, мы получаем

$$\ln c \approx \beta_0 + \sum_{i=1}^M \left(\frac{\partial \ln c}{\partial \ln p_i} \right) \log p_i + \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^M \left(\frac{\partial^2 \ln c}{\partial \ln p_i \partial \ln p_j} \right) \ln p_i \ln p_j, \quad (10-35)$$

где все производные вычисляются в точке разложения. Если мы рассмотрим эти производные как коэффициенты, то функция затрат примет вид

$$\begin{aligned} \ln c = & \beta_0 + \beta_1 \ln p_1 + \dots + \beta_M \ln p_M + \delta_{11} \left(\frac{1}{2} \ln^2 p_1 \right) + \delta_{12} \ln p_1 \ln p_2 + \\ & + \delta_{22} \left(\frac{1}{2} \ln^2 p_2 \right) + \dots + \delta_{MM} \left(\frac{1}{2} \ln^2 p_M \right). \end{aligned} \quad (10-36)$$

Это транслогарифмическая функция издержек. Если δ_{ij} равен нулю, то она сводится к функции Кобба–Дугласа, рассмотренной нами ранее. Доли издержек задаются как

$$\begin{aligned} s_1 &= \frac{\partial \ln c}{\partial \ln p_1} = \beta_1 + \delta_{11} \ln p_1 + \delta_{12} \ln p_2 + \dots + \delta_{1M} \ln p_M, \\ s_2 &= \frac{\partial \ln c}{\partial \ln p_2} = \beta_2 + \delta_{21} \ln p_1 + \delta_{22} \ln p_2 + \dots + \delta_{2M} \ln p_M, \\ &\vdots \\ s_M &= \frac{\partial \ln c}{\partial \ln p_M} = \beta_M + \delta_{M1} \ln p_1 + \delta_{M2} \ln p_2 + \dots + \delta_{MM} \ln p_M. \end{aligned} \quad (10-37)$$

Доли издержек в сумме должны быть равны 1, что требует, чтобы

$$\begin{aligned} \beta_1 + \beta_2 + \dots + \beta_M &= 1, \\ \sum_{i=1}^M \delta_{ij} &= 0 \quad (\text{сумма столбцов равна нулю}), \\ \sum_{j=1}^M \delta_{ij} &= 0 \quad (\text{сумма строк равна нулю}). \end{aligned} \quad (10-38)$$

²⁸Функция затрат является лишь одним из нескольких подходов для данного исследования. См. Jorgenson (1983) по этой теме.

²⁹См. пример 2.4. Функция была разработана Кментой (Kmenta (1967)) как средство аппроксимации производственной функции CES и официально представлена в серии работ Berndt, Christensen, Jorgenson и Lau, включая Berndt и Christensen (1973) и Christensen et al. (1975). Литература стала чем-то вроде арены соревнования по разработке экзотических функциональных форм. Однако транслогарифмическая функция осталась самой популярной и, по мнению Guilkey, Lovell и Sickles (1983), наиболее достоверной из нескольких доступных альтернатив. См. также пример 5.4.

Мы также накладываем (теоретическое) ограничение симметричности $\delta_{ij} = \delta_{ji}$.

Система **уравнений долей** позволяет получить модель внешне не связанных регрессий, которую можно использовать для оценки параметров модели³⁰. Чтобы сделать модель работающей, следует наложить ограничения в (10-38) и решить проблему **вырожденности ковариационной матрицы шоков** в уравнениях долей. Первое выполняется с помощью деления первых $M - 1$ цен на M , таким образом мы исключаем последний член в каждой строке и столбце матрицы параметров. Как и в модели Кобба–Дугласа, мы получаем невырожденную систему, опуская уравнение M -й доли. Мы вычисляем оценки параметров при помощи метода максимального правдоподобия, чтобы обеспечить **инвариантность** относительно выбора исключаемого уравнения доли. Что касается транслогарифмической функции издержек, то эластичности замещения довольно легко вычислить, если параметры уже оценены:

$$\theta_{ij} = \frac{\delta_{ij} + s_i s_j}{s_i s_j}, \quad \theta_{ii} = \frac{\delta_{ii} + s_i(s_i - 1)}{s_i^2}. \quad (10-39)$$

Эти эластичности будут различаться для каждой точки из выборки данных. Легче вычислить их для некой центральной точки, например для среднего³¹.

Пример 10.3. Функция издержек для промышленного производства США.

В ряде недавних исследований с использованием транслогарифмической методологии была применена четырехфакторная модель с капиталом K , трудом L , энергией E и сырьем M — факторами производства. Одно из первых исследований с использованием такой методологии включало оценку транслогарифмической функции издержек для промышленного сектора США у Берндта и Вуда (Berndt, Wood (1975)). Доли трех факторов, использованных для оценки модели, имеют вид

$$s_K = \beta_K + \delta_{KK} \ln \left(\frac{p_K}{p_M} \right) + \delta_{KL} \ln \left(\frac{p_L}{p_M} \right) + \delta_{KE} \ln \left(\frac{p_E}{p_M} \right),$$

$$s_L = \beta_L + \delta_{KL} \ln \left(\frac{p_K}{p_M} \right) + \delta_{LL} \ln \left(\frac{p_L}{p_M} \right) + \delta_{LE} \ln \left(\frac{p_E}{p_M} \right),$$

$$s_E = \beta_E + \delta_{KE} \ln \left(\frac{p_K}{p_M} \right) + \delta_{LE} \ln \left(\frac{p_L}{p_M} \right) + \delta_{EE} \ln \left(\frac{p_E}{p_M} \right).$$

³⁰При желании можно включать функцию затрат, что позволит получить оценку β_0 , но в ином случае она несущественна. Однако при отсутствии предположения о постоянной отдаче от масштаба функция затрат будет содержать интересующие параметры, которые не возникают в уравнениях долей. Поэтому кому-то может захотеться включить его в модель. См. у Christensen и Greene (1976) практическое применение.

³¹Они также будут являться в высшей степени нелинейными функциями параметров и данных. Метод для вычисления асимптотических стандартных ошибок для оцененных эластичностей представлен у Anderson и Thursby (1986). Krinsky и Robb (1986, 1990) (см. раздел 15.3) предложили свой метод в качестве альтернативного подхода к расчету.

Данные Берндта и Вуда приведены в табл. F10.2 приложения. FGLS-оценки параметров в ограничениями, представленные в табл. 10.3, были получены при помощи построения «объединенной регрессии» в (10-19) по матрицам данных

$$y = \begin{bmatrix} s_K \\ s_L \\ s_E \end{bmatrix}, \quad (10-40)$$

$$X = \begin{bmatrix} i & 0 & 0 & \ln P_K/P_M & \ln P_L/P_M & \ln P_E/P_M & 0 & 0 & 0 \\ 0 & i & 0 & 0 & \ln P_K/P_M & 0 & \ln P_L/P_M & \ln P_E/P_M & 0 \\ 0 & 0 & i & 0 & 0 & \ln P_K/P_M & 0 & \ln P_L/P_M & \ln P_E/P_M \end{bmatrix},$$

$$\beta' = (\beta_K, \beta_L, \beta_E, \delta_{KK}, \delta_{KL}, \delta_{KE}, \delta_{LL}, \delta_{LE}, \delta_{EE}).$$

Таблица 10.3. Оценки параметров (в скобках даны стандартные ошибки)

| | | | | | |
|---------------|-----------|------------|---------------|-----------|-----------|
| β_K | 0,05682 | (0,00131) | δ_{KM} | -0,02169* | (0,00963) |
| β_L | 0,25355 | (0,001987) | δ_{LL} | 0,07488 | (0,00639) |
| β_E | 0,04383 | (0,00105) | δ_{LE} | -0,00321 | (0,00275) |
| β_M | 0,64580* | (0,00299) | δ_{LM} | -0,07169* | (0,00941) |
| δ_{KK} | 0,02987 | (0,00575) | δ_{EE} | 0,02938 | (0,00741) |
| δ_{KL} | 0,0000221 | (0,00367) | δ_{EM} | -0,01797* | (0,01075) |
| δ_{KE} | -0,00820 | (0,00406) | δ_{MM} | 0,11134* | (0,02239) |

* Оценены с использованием (10-38).

Таблица 10.4. Оцененные эластичности

| | Капитал | Труд | Энергия | Материалы |
|---|----------|----------|----------|-----------|
| Доли издержек, 1959 г. | | | | |
| Установленные доли | 0,05646 | 0,27454 | 0,04424 | 0,62476 |
| Фактические доли | 0,06185 | 0,27303 | 0,04563 | 0,61948 |
| Оцененные эластичности замещения, 1959 г. | | | | |
| Капитал | -7,34124 | | | |
| Труд | 1,0014 | -1,64902 | | |
| Энергия | -2,28422 | 0,73556 | -6,59124 | |
| Материалы | 0,38512 | 0,58205 | 0,34994 | -0,31536 |
| Оцененные ценовые эластичности | | | | |
| | -0,41448 | -0,45274 | -0,29161 | -0,19702 |

Оценки получены при помощи использования двухшаговой процедуры (10-7) и (10-9)³². Все полученные оценки приведены в табл. 10.4. Параметры, которые не были оценены напрямую в (10-36), вычислены с помощью (10-38).

Предполагаемые оценки эластичностей замещения и спроса на 1959 г. (центральный год в данных) приведены в табл. 10.4 посредством установленных долей издержек и оцененных параметров в (10-39). Отступление от модели Кобба-Дугласа с единичными эластичностями существенно. К примеру, результаты почти не предполагают замещения между энергией и трудом и говорят о некоторой взаимодополняемости капитала и энергии³³.

10.6. Модели одновременных уравнений

Существует качественная разница между моделью рыночного равновесия, представленной во вводной части главы,

$$\begin{aligned} Q_{Demand} &= \alpha_1 + \alpha_2 Price + \alpha_3 Income + \mathbf{d}'\boldsymbol{\alpha} + \varepsilon_{Demand}, \\ Q_{Supply} &= \beta_1 + \beta_2 Price + \mathbf{s}'\boldsymbol{\beta} + \varepsilon_{Supply}, \\ Q_{Equilibrium} &= Q_{Demand} = Q_{Supply}, \end{aligned}$$

и другими примерами, рассмотренными далее. Модель внешне не связанных регрессий

$$y_{im} = \mathbf{x}'_{im}\boldsymbol{\beta}_m + \varepsilon_{im}$$

выводится из набора регрессионных уравнений, связанных между собой через шоки. Регрессоры \mathbf{x}_{im} экзогенны и могут меняться автономно ввиду того, что они не объясняются в рамках модели. Таким образом, коэффициенты могут быть непосредственно интерпретированы как частные эффекты и могут быть оценены при помощи метода наименьших квадратов или другими методами, основанными на функциях условных средних $E[y_{im}|\mathbf{x}_{im}] = \mathbf{x}'_{im}\boldsymbol{\beta}$. В таких моделях, как рассмотренная выше модель равновесия, отношения между уравнениями являются явными и ни одно из двух рыночных уравнений не будет регрессионной моделью. Как следствие частичный эксперимент, заключающийся в изменении цены и приводящий к изменению равновесного выпуска, чтобы получить оценку ценовой эластичности спроса, α_2 (или эластичности предложения β_2), не имеет смысла. Модель предполагает совместное определение выпуска и цены. Цена изменяется, когда меняется рыночное равновесие, однако это же стимулируется изменениями других факторов, например изменениями в доходах или других переменных, которые влияют на функцию предложения. (См. графический анализ на рис. 8.1.)

Как мы видели в примере 8.4, с помощью оценки регрессии для наблюдаемых равновесных величин цен и других факторов методом наименьших квадратов мы получим неясное смешение из функций спроса и предложения. Такой вывод следует из эндогенности цены (Price) в каждом уравнении. «Модели одновременных уравнений» появляются в подобных условиях, когда есть набор взаимозависимых уравнений. Модели одновременных уравнений соответствуют теории, представленной в главе 8, в ней мы рассматривали уравнения, где некоторые переменные с правой стороны эндогенные и таким образом коррелируют с шоками. Принципиальное разли-

³²Эти оценки не совпадают с изложенными у Берндта и Вуда. Они использовали итерационную оценку, в то время как наша представляет собой двухшаговый FGLS. Чтобы очистить эти данные от возможной корреляции с шоками, они сначала вычислили коэффициент регрессии для цен 10 экзогенных макроэкономических переменных, таких как население США, государственные закупки трудовых услуг, реальный экспорт товаров длительного спроса и фонд материальных активов США, а затем использовали в анализе подобранные значения. Оценки, приведенные здесь, являются в общем смысле достаточно близкими к тем, что приводят Берндт и Вуд. К примеру, их оценки первых пяти параметров равны 0,0564, 0,2539, 0,0442, 0,6455 и 0,0254.

³³Оценка Берндта и Вуда для на 1959 г. равна 0,64.

чие заключается в источнике эндогенности. В ситуации из главы 8 эндогенность возникает, к примеру, в моделях с пропущенными переменными, из-за ошибок измерения или является эффектом эндогенного воздействия и, по сути, ненамеренным отклонением от условий модели линейной регрессии. В анализе с использованием одновременных уравнений эндогенность выступает центральной частью спецификации. В этом разделе мы рассматриваем проблемы спецификации и оценки систем одновременных уравнений. Подраздел 10.6.1 мы начнем с изложения общих теоретических рамок для анализа и формулировки некоторых фундаментальных идей. В подразделе 10.6.2 представлена модель одновременных уравнений как продолжение модели внешне не связанных регрессий из раздела 10.2. Конечная цель анализа — выводы о коэффициентах модели. Вопрос о том, является ли это возможным, рассматривается в подразделе 10.6.3, где мы изучим проблему идентификации. После решения проблемы идентификации в подразделах 10.6.4 и 10.6.5 рассмотрим методы оценки и основные выводы.

10.6.1. Системы уравнений

Рассмотрим упрощенный вариант указанной выше модели равновесия:

$$\text{уравнение спроса: } q_{d,t} = \alpha_1 p_t + \alpha_2 x_t + \varepsilon_{d,t},$$

$$\text{уравнение предложения: } q_{s,t} = \beta_1 p_t + \varepsilon_{s,t},$$

$$\text{условие равновесия: } q_{d,t} = q_{s,t} = q_t.$$

Эти уравнения являются **структурными**, так как выводятся из теории и каждое из них имеет целью описать отдельную сторону экономики³⁴. Поскольку эта модель является одной из частей совместного определения цены и выпуска, они называются **совместно зависимыми** или **эндогенными** переменными. Предполагается, что доход x определен вне модели, что делает его **экзогенным**. Шоки добавляют в классическое описание в учебниках, чтобы получить **эконометрическую модель**. Все три уравнения необходимы для определения равновесных значений цены и выпуска, т.е. система **взаимозависима**. Наконец, так как равновесное решение для значений цены и выпуска находится при некоторых доходах и шоках в предположении, что α_1 не равно β_1 , то система называется **полной системой уравнений**. *Полнота системы требует, чтобы число уравнений было равно числу эндогенных переменных*. Как правило, оценить все параметры неполных систем не представляется возможным (хотя иногда существует возможность оценить некоторые из них).

Предположим, что наша главная задача — оценить эластичность спроса α_1 . Для простоты предположим, что ε_d и ε_s являются доброкачественными классическими шоками с

$$\begin{aligned} E[\varepsilon_{d,t} | x_t] &= E[\varepsilon_{s,t} | x_t] = 0, \\ E[\varepsilon_{d,t}^2 | x_t] &= \sigma_d^2, \\ E[\varepsilon_{s,t}^2 | x_t] &= \sigma_s^2, \\ E[\varepsilon_{d,t}\varepsilon_{s,t} | x_t] &= 0. \end{aligned}$$

³⁴Различие между структурными и неструктурными моделями иногда проводится на этой основе. См., к примеру, Cooley, LeRoy (1985).

Все переменные не коррелированы друг с другом по наблюдениям в различные моменты. Цена, выпуск и доход измеряются в логарифмах отклонений от своих выборочных средних. Решение уравнений для p и q в переводе на x , ε_d и ε_s дает **приведенную форму модели**

$$\begin{aligned} p &= \frac{\alpha_2 x}{\beta_1 - \alpha_1} + \frac{\varepsilon_d - \varepsilon_s}{\beta_1 - \alpha_1} = \pi_1 x + v_1, \\ q &= \frac{\beta_1 \alpha_2 x}{\beta_1 - \alpha_1} + \frac{\beta_1 \varepsilon_d - \alpha_1 \varepsilon_s}{\beta_1 - \alpha_1} = \pi_2 x + v_2. \end{aligned} \quad (10-41)$$

(Обратите внимание на роль требования «полноты» в том, что α_1 не равно β_1 .)

Из этого следует, что $\text{Cov}[p, \varepsilon_d] = \sigma_d^2 / (\beta_1 - \alpha_1)$ и $\text{Cov}[p, \varepsilon_s] = -\sigma_s^2 / (\beta_1 - \alpha_1)$, т.е. оба уравнения (и для спроса, и для предложения) не удовлетворяют условиям классической регрессионной модели. Эластичность спроса по цене не может быть состоятельно оценена регрессией МНК q на x и p . Этот вывод характерен для систем одновременных уравнений. Поскольку все эндогенные переменные коррелированы с шоками, МНК-оценки параметров уравнений с эндогенными переменными в правой части несостоятельны³⁵.

Представим, что у нас есть выборка из T наблюдений по p , q и x такая, что

$$\text{plim}(1/T)\mathbf{x}'\mathbf{x} = \sigma_x^2.$$

Так как оценки МНК несостоятельны, мы могли бы вместо этого использовать **метод инструментальных переменных (IV-оценку)**³⁶. Единственной переменной, которая не коррелирована с шоками, в этой системе будет x . Рассмотрим далее IV-оценку $\hat{\beta}_1 = \mathbf{q}'\mathbf{x}/\mathbf{p}'\mathbf{x}$. Для этой оценки

$$\text{plim}\hat{\beta}_1 = \text{plim} \frac{\mathbf{q}'\mathbf{x}/T}{\mathbf{p}'\mathbf{x}/T} = \frac{2\beta_1\alpha_2/(\beta_1 - \alpha_1)}{2\alpha_2/(\beta_1 - \alpha_1)} = \beta_1.$$

Очевидно, что параметр кривой предложения может быть оценен с помощью метода инструментальных переменных. В МНК-регрессии p на x предсказанные моделью значения равны $\hat{p} = (\mathbf{p}'\mathbf{x} / \mathbf{x}'\mathbf{x})x$. Следовательно, в регрессии с инструментальными переменными инструментом будет \hat{p} , т.е.

$$\hat{\beta}_1 = \frac{\mathbf{p}'\mathbf{q}}{\mathbf{p}'\hat{\mathbf{p}}}.$$

Так как $\hat{p}'\mathbf{p} = \hat{p}'\hat{p}$, $\hat{\beta}_1$ также является коэффициентом наклона в регрессии для q на эти предсказанные значения. Такая трактовка определяет оценку **двухшаговым методом наименьших квадратов**.

Желательно использовать эту же технику для оценки параметров уравнения спроса, однако, к сожалению, информация в выборке исчерпана. Проблема заключается не только в том, что с помощью метода наименьших

³⁵Этот недостаток метода наименьших квадратов иногда называют смещением в системах одновременных уравнений.

³⁶См. раздел 8.3.

квадратов не удастся оценить уравнение спроса, но и в том, что без некоторых дополнительных предположений в выборке не хватает информации, с которой можно было бы работать. В этом примере возникает **проблема идентификации**, упомянутая во введении к данному разделу.

Различие между «экзогенными» и «эндогенными» переменными в модели — тонкое, а иногда и спорное дополнительное усложнение. Оно является предметом исследования в многочисленных работах. Мы обрисовали различие в пригодном для использования экономическом виде с точки зрения того, может ли переменная в модели изменяться «автономно», независимо от других переменных. Так, в модели спроса и предложения переменная погоды в уравнении предложения, скорее всего, будет экзогенной в чистом виде по отношению к определению цены и выпуска, в то время как текущая цена, очевидно, «эндогенна» при любом разумном построении. К сожалению, такое четкое разделение имеет достаточно ограниченное применение в макроэкономике, где почти ни одну переменную нельзя назвать по-настоящему экзогенной в том виде, в каком этот термин понимает большинство исследователей. Если взять типичный пример — оценку функций потребления при помощи обычного метода наименьших квадратов, как мы делали в некоторых предыдущих примерах, такой подход обычно рассматривается как допустимый шаг, даже несмотря на то что большое число макроэкономических моделей (включая приведенные здесь примеры) отклоняется от функции потребления, в которой доход экзогенен. Такое отклонение заставило исследователей, хорошо это или плохо, проводить различие скорее на статистических основаниях. Методологическое развитие данного вопроса в литературе привело к некоторому консенсусу. Как мы увидим далее, определения формализуют экономические характеристики, описанные нами ранее. Мы перечислим здесь некоторые выводы, необходимые нам в дальнейшем изложении. Заинтересованный читатель может обратиться к соответствующей литературе (предупреждаем о некоторых сложностях чтения).

Энгл, Хендри и Ричард (Engle, Hendry, Richard (1983)) определяют переменные x_t в параметризованной модели как **слабоэкзогенные**, если полная модель может быть записана с точки зрения безусловного распределения вероятности для x_t и условного распределения для $y_t|x_t$ такого, что оценка параметров условного распределения не менее эффективна, чем оценка полного набора параметров совместного распределения. Такой случай будет иметь место, если ни один из параметров условного распределения не будет присутствовать в безусловном распределении x_t . В данном контексте нам понадобится такой вид построения, чтобы получить приведенные формы, как мы делали до этого. Со ссылкой на примеры использования временных рядов (хотя эта идея распространяется и на межобъектные данные) переменные x_t считаются **предопределенными** в модели, если x_t независим от всех *последующих* структурных шоков ε_{t+s} для $s \geq 0$. Переменные, которые предопределены в модели, могут рассматриваться, по крайней мере асимптотически, как если бы они были экзогенными в том смысле, что постоянные оценки могут быть получены, если они появляются в качестве регрессоров. Мы используем этот вывод в главе 21, когда будем выводить свойства регрессий, содержащих запаздывающие значения зави-

симой переменной. С этим связано понятие **причинности по Грейнджеру–Симсу** (Granger (1969), Sims (1977)). Причинность по Грейнджеру (вид статистической обратной связи) отсутствует, когда $f(x_t|x_{t-1}, y_{t-1})$ равно $f(x_t|x_{t-1})$. Определение указывает, что условное распределение и запаздывающие значения y_t не добавляют информации к объяснению x_t за пределы, предоставляемые запаздывающими значениями самого x_t . Эта идея полезна в построении прогнозных моделей. Наконец, если переменные x_t слабоэкзогенны и если y_{t-1} не является причиной по Грейнджеру для x_t , тогда x_t **строго экзогенны**.

10.6.2. Общие замечания для моделей линейных одновременных уравнений³⁷

Структурная форма модели имеет вид³⁸

$$\begin{aligned} \gamma_{11}y_{t1} + \gamma_{21}y_{t2} + \cdots + \gamma_{M1}y_{tM} + \beta_{11}x_{t1} + \cdots + \beta_{K1}x_{tK} &= \varepsilon_{t1}, \\ \gamma_{12}y_{t1} + \gamma_{22}y_{t2} + \cdots + \gamma_{M2}y_{tM} + \beta_{12}x_{t1} + \cdots + \beta_{K2}x_{tK} &= \varepsilon_{t2}, \\ &\vdots \\ \gamma_{1M}y_{t1} + \gamma_{2M}y_{t2} + \cdots + \gamma_{MM}y_{tM} + \beta_{1M}x_{t1} + \cdots + \beta_{KM}x_{tK} &= \varepsilon_{tM}. \end{aligned} \quad (10-42)$$

У нас есть M уравнений и M эндогенных переменных, обозначаемых y_1, \dots, y_M , а также K экзогенных переменных x_1, \dots, x_K , которые могут включать предопределенные значения y_1, \dots, y_M . Первый элемент x_t , как обычно, будет константой 1. И наконец, $\varepsilon_{t1}, \dots, \varepsilon_{tM}$ — **структурные шоки**. Индекс t будет использоваться для обозначения наблюдений, $t = 1, \dots, T$.

В матричном виде система может быть записана как

$$\begin{aligned} [y_1 \quad y_2 \quad \cdots \quad y_M]_t &\begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1M} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2M} \\ & & \vdots & \\ \gamma_{M1} & \gamma_{M2} & \cdots & \gamma_{MM} \end{bmatrix} + \\ &+ [x_1 \quad x_2 \quad \cdots \quad x_K]_t \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1M} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2M} \\ & & \vdots & \\ \beta_{K1} & \beta_{K2} & \cdots & \beta_{KM} \end{bmatrix} = [\varepsilon_1 \varepsilon_2 \cdots \varepsilon_M]_t \end{aligned}$$

или

$$y'_t \Gamma + x'_t \mathbf{B} = \varepsilon'_t.$$

Каждый столбец матрицы параметров — это вектор коэффициентов в отдельном уравнении, а каждая строка относится к конкретной эндогенной переменной.

³⁷Мы ограничимся рассмотрением линейных моделей. **Нелинейные системы** являются другим направлением в этой области. В нелинейных системах возникает множество сложностей помимо рассмотренных здесь, что выходит за рамки данного текста. Gallant (1987), Gallant, Holly (1980), Gallant, White (1988), Davidson, MacKinnon (2004) и Wooldridge (2002a) рассматривают этот вопрос более подробно.

³⁸Сейчас будет удобнее пренебречь особой природой запаздывающих эндогенных переменных и рассматривать их так же, как строго экзогенные переменные.

Лежащая в основе таких моделей теория будет предполагать некоторое число ограничений для Γ и \mathbf{B} . Одна из переменных в каждом уравнении имеет название *зависимая переменная*, и ее коэффициент в модели будет равен 1. Поэтому здесь будет как минимум одна «1» в каждом столбце Γ . Подобная **нормировка** не является основным ограничением. Соотношения, определенные для данного уравнения, останутся неизменными, если каждый коэффициент в уравнении умножить на одну и ту же константу. Выбор «зависимой переменной» просто снимает эту неопределенность. Если существуют какие-либо тождества, то соответствующие столбцы Γ и \mathbf{B} будут полностью известны и для этого уравнения не будет существовать шока. Поскольку не все переменные возникают во всех уравнениях, некоторые параметры будут равны нулю. Теория по данной теме может также налагать ограничения другого вида на параметрические матрицы.

Если Γ — верхнетреугольная матрица, то система называется **треугольной**. В этом случае модель принимает вид

$$\begin{aligned} y_{t1} &= f_1(\mathbf{x}_t) + \varepsilon_{t1}, \\ y_{t2} &= f_2(y_{t1}, \mathbf{x}_t) + \varepsilon_{t2}, \\ &\vdots \\ y_{tM} &= f_M(y_{t1}, y_{t2}, \dots, y_{t,M-1}, \mathbf{x}_t) + \varepsilon_{tM}. \end{aligned}$$

Совместное определение переменных в этой матрице имеет **рекурсивный** характер. Первая из них полностью определяется экзогенными факторами. Затем, имея первую переменную, определяем вторую аналогичным образом и т.д.

Решение системы уравнений путем определения y_t в терминах \mathbf{x}_t и ε_t представляет собой **приведенную форму** модели

$$\begin{aligned} \mathbf{y}'_t &= [x_1 \ x_2 \ \dots \ x_K]_t \begin{bmatrix} \pi_{11} & \pi_{12} & \dots & \pi_{1M} \\ \pi_{21} & \pi_{22} & \dots & \pi_{2M} \\ & & \vdots & \\ \pi_{K1} & \pi_{K2} & \dots & \pi_{KM} \end{bmatrix} + [\nu_1 \ \dots \ \nu_M]_t \\ &= -\mathbf{x}'_t \mathbf{B} \Gamma^{-1} + \varepsilon'_t \Gamma^{-1} \\ &= \mathbf{x}'_t \mathbf{\Pi} + \mathbf{v}'_t. \end{aligned}$$

Чтобы существовала возможность такого решения, модель должна удовлетворять **условию полноты** для систем одновременных уравнений: Γ должна быть невырожденной.

Пример 10.4. Структурная и приведенная формы в малой макроэкономической модели

Рассмотрим следующую модель:

$$\text{потребление: } c_t = \alpha_0 + \alpha_1 y_t + \alpha_2 c_{t-1} + \varepsilon_{t1},$$

инвестиции: $i_t = \beta_0 + \beta_1 r_t + \beta_2 (y_t - y_{t-1}) + \varepsilon_{t2}$,

спрос: $y_t = c_t + i_t + g_t$.

Модель включает авторегрессионную функцию потребления, зависящую от выпуска y_t и первого запаздывания потребления, уравнение инвестиций как зависимости от процента r_t и прироста выпуска, а также условие равновесия. Модель определяет значения трех эндогенных переменных c_t , i_t и y_t . Это динамическая модель. Помимо экзогенных переменных r_t и государственных расходов g_t она содержит две **предопределенные переменные** c_{t-1} и y_{t-1} . Очевидно, что они не являются экзогенными, однако с учетом текущих значений эндогенных переменных могут рассматриваться как уже определенные. Решающим фактором будет то, являются ли они некоррелированными с текущими шоками, что мы могли бы предположить. Приведенная форма модели выглядит так:

$$Ac_t = \alpha_0(1 - \beta_2) + \beta_0\alpha_1 + \alpha_1\beta_1 r_t + \alpha_1 g_t + \alpha_2(1 - \beta_2)c_{t-1} - \alpha_1\beta_2 y_{t-1} + (1 - \beta_2)\varepsilon_{t1} + \alpha_1\varepsilon_{t2},$$

$$Ai_t = \alpha_0\beta_2 + \beta_0(1 - \alpha_1) + \beta_1(1 - \alpha_1)r_t + \beta_2 g_t + \alpha_2\beta_2 c_{t-1} - \beta_2(1 - \alpha_1)y_{t-1} + \beta_2\varepsilon_{t1} + (1 - \alpha_1)\varepsilon_{t2},$$

$$Ay_t = \alpha_0 + \beta_0 + \beta_1 r_t + g_t + \alpha_2 c_{t-1} - \beta_2 y_{t-1} + \varepsilon_{t1} + \varepsilon_{t2},$$

где $A = 1 - \alpha_1 - \beta_2$. Заметим, что в приведенной форме сохраняется условие равновесия.

Обозначим, что $y' = [c, i, y]$, $x' = [1, r, g, c_{-1}, y_{-1}]$, тогда

$$\Gamma = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ -\alpha_1 & -\beta_2 & 1 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} -\alpha_0 & -\beta_0 & 0 \\ 0 & -\beta_1 & 0 \\ 0 & 0 & -1 \\ -\alpha_2 & 0 & 0 \\ 0 & \beta_2 & 0 \end{bmatrix},$$

$$\Gamma^{-1} = \frac{1}{\Delta} \begin{bmatrix} 1 - \beta_2 & \beta_2 & 1 \\ \alpha_1 & 1 - \alpha_1 & 1 \\ \alpha_1 & \beta_2 & 1 \end{bmatrix},$$

$$\Pi' = \frac{1}{\Delta} \begin{bmatrix} \alpha_0(1 - \beta_2) + \beta_0\alpha_1 & \alpha_1\beta_1 & \alpha_1 & \alpha_2(1 - \beta_2) & -\beta_2\alpha_1 \\ \alpha_0\beta_2 + \beta_0(1 - \alpha_1) & \beta_1(1 - \alpha_1) & \beta_2 & \alpha_2\beta_2 & -\beta_2(1 - \alpha_1) \\ \alpha_0 + \beta_0 & \beta_1 & 1 & \alpha_2 & -\beta_2 \end{bmatrix},$$

где $\Delta = 1 - \alpha_1 - \beta_2$. Условие полноты предполагает, что сумма α_1 и β_2 не равна единице.

В интерпретации коэффициентов системы одновременных уравнений существует некоторая неоднозначность. Эффекты в структурной форме модели были бы названы «причинными» из-за того, что они вытекают непосредственно из теории. Однако чтобы отследить эффекты автономных изменений в переменных модели, необходимо рассматривать приведенную форму. К примеру, в функции потребления нет процентной ставки. Но из этого не следует, что изменения в r_t не станут «причиной» изменений в потреблении, поскольку изменения в r_t меняют инвестиции,

а это оказывает воздействие на спрос, который, в свою очередь, возникает в функции потребления. Так, мы видим из приведенной формы, что $\Delta c_t / \Delta r_t = \alpha_1 \beta_1 / A$. Точно так же «эксперимент» $\Delta c_t / \Delta y_t$ не имеет смысла без предварительного определения того, что вызвало изменение y_t . Если это изменение было спровоцировано изменением процентной ставки, мы бы обнаружили, что $(\Delta c_t / \Delta r_t) / (\Delta y_t / \Delta r_t) = (\alpha_1 \beta_1 / A) / (\beta_1 / A) = \alpha_1$.

Предполагается, что структурные шоки случайным образом выбраны из M -мерного распределения, где

$$E[\epsilon_t | x_t] = 0 \text{ и } E[\epsilon_t \epsilon_t' | x_t] = \Sigma.$$

Пока предположим, что

$$E[\epsilon_t \epsilon_s' | x_t, x_s] = 0, \quad \forall t, s.$$

Позже мы опустим это предположение, чтобы сделать возможным наличие гетероскедастичности и автокорреляции. Иногда может быть удобно предположить, что ϵ_t имеет многомерное нормальное распределение, однако мы не будем вводить это предположение до тех пор, пока оно не станет необходимым. Может быть удобно оставить тождества без шоков как отдельные уравнения. В этом случае для спецификации стохастических шоков нужно проставить строки и столбцы из нулей в соответствующие места в Σ . Из этого следует, что **матрица ошибок в приведенной форме** $v_t' = \epsilon_t' \Gamma^{-1}$ будет иметь вид

$$\begin{aligned} E[v_t | x_t] &= (\Gamma^{-1})' 0 = 0, \\ E[v_t v_t' | x_t] &= (\Gamma^{-1})' \Sigma \Gamma^{-1} = \Omega. \end{aligned}$$

Это предполагает, что

$$\Sigma = \Gamma' \Omega \Gamma.$$

Предшествующая формула описывает то, как модель применяется к тому или иному наблюдению $[y', x', \epsilon']_t$ в определенный момент или для определенного объекта. В выборке каждое совместное наблюдение соответствует одной строке в матрице данных

$$[\mathbf{Y} \ \mathbf{X} \ \mathbf{E}] = \begin{bmatrix} y_1' & x_1' & \epsilon_1' \\ y_2' & x_2' & \epsilon_2' \\ \vdots & \vdots & \vdots \\ y_T' & x_T' & \epsilon_T' \end{bmatrix}.$$

В терминах всего набора из T наблюдений модель имеет вид

$$\mathbf{Y}\Gamma + \mathbf{X}\mathbf{B} = \mathbf{E},$$

где

$$E[\mathbf{E} | \mathbf{X}] = 0 \text{ и } E[(1/T)\mathbf{E}'\mathbf{E} | \mathbf{X}] = \Sigma.$$

При выполнении ряда общих предположений мы можем свести эту структуру к виду

$$\text{plim}[(1/T)\mathbf{E}'\mathbf{E}] = \Sigma.$$

Важное предположение, сравнимое со сделанным в главе 4 для классической регрессионной модели, состоит в том, что

$$\text{plim}(1/T)\mathbf{X}'\mathbf{X} = \mathbf{Q} \text{ — конечная положительно определенная матрица.} \quad (10-43)$$

Мы также предполагаем, что

$$\text{plim}(1/T)\mathbf{X}'\mathbf{E} = \mathbf{0}. \quad (10-44)$$

Именно это предположение и отличает предопределенные переменные от эндогенных переменных. Приведенная форма имеет вид

$$\mathbf{Y} = \mathbf{X}\Pi + \mathbf{V}, \text{ где } \mathbf{V} = \mathbf{E}\Gamma^{-1}. \quad (10-45)$$

Объединяя полученные ранее результаты, имеем

$$\text{plim} \frac{1}{T} \begin{bmatrix} \mathbf{Y}' \\ \mathbf{X}' \\ \mathbf{V}' \end{bmatrix} [\mathbf{Y} \ \mathbf{X} \ \mathbf{V}] = \begin{bmatrix} \Pi' \mathbf{Q} \Pi + \Omega \Pi' \mathbf{Q} \Omega \\ \mathbf{Q} \Pi & \mathbf{Q} \ \mathbf{0}' \\ \Omega & \mathbf{0} \ \Omega \end{bmatrix}.$$

10.6.3. Проблема идентификации

Решение **проблемы идентификации** логически предшествует оцениванию. Это ключевая часть этапа спецификации модели. Главная проблема заключается в том, существует ли *какой-либо* способ получить оценки параметров для специфицированной модели. У нас есть определенная информация, необходимая для того, чтобы сделать выводы о структуре модели. Если с одними и теми же «данными» согласуется более чем одна теория, то эти теории называют **эквивалентными с точки зрения наблюдений** и не существует способа их различить. Мы уже сталкивались с этой проблемой в главе 4, где рассматривалась проблема *мультиколлинеарности*. «Модель»

$$\begin{aligned} \text{consumption} = & \beta_1 + \beta_2 \text{WageIncome} + \beta_3 \text{NonWageIncome} + \\ & + \beta_4 \text{TotalIncome} + \varepsilon \end{aligned} \quad (10-46)$$

нельзя отличить от альтернативной модели

$$\begin{aligned} \text{consumption} = & \gamma_1 + \gamma_2 \text{WageIncome} + \gamma_3 \text{NonWageIncome} + \\ & + \gamma_4 \text{TotalIncome} + \omega, \end{aligned} \quad (10-47)$$

где $\gamma_1 = \beta_1$, $\gamma_2 = \beta_2 + a$, $\gamma_3 = \beta_3 + a$, $\gamma_4 = \beta_4 - a$ для некоторого ненулевого a , если наши данные состоят только из потребления и двух значений дохода (и их суммы). Однако если мы знаем, что β_4 равно нулю, то, как мы видели

в главе 4, γ_2 должно быть равно β_2 и γ_3 должно быть равно β_3 . Дополнительная информация служит для того, чтобы исключить альтернативную модель. Представление об эквивалентности по наблюдениям относится к тому, что может быть извлечено из имеющейся информации, которая состоит из выборки данных и ограничений, налагаемых теорией на уравнения данной модели. В главе 8, где мы рассматривали IV-оценку, мы определяли идентификацию с точки зрения достаточных моментных уравнений. Действительно, рис. 8.1 является точным отражением принципа эквивалентности по наблюдениям. Случай ошибки измерения, рассмотренный нами в разделе 8.5, также имеет отношение к проблеме идентификации. Выборочный коэффициент регрессии b сходится к функции двух параметров β и σ_u^2 ; $\text{plim } b = \beta/[1 + \sigma_u^2/Q^{**}]$, где $Q^{**} = \text{plim}(x^{**'}x^*/n)$. Без дополнительной информации о σ_u^2 мы не можем вывести β из информации о выборке b и Q^{**} , ведь есть различные пары β и σ_u^2 , которые дадут тот же $\text{plim } b$.

Математически эта идея может быть изложена в терминах функции правдоподобия, которая включает информацию о выборке. В этом случае можно провести статистическое различие между «у» и «х» и в общем виде изучить совместное распределение вероятности для наблюдаемых данных $p(\mathbf{Y}, \mathbf{X}|\theta)$ при данных параметрах модели. Две структуры модели эквивалентны с точки зрения наблюдений, если

$$p(\mathbf{Y}, \mathbf{X}|\theta_1) = p(\mathbf{Y}, \mathbf{X}|\theta_2) \quad \text{при } \theta_1 \neq \theta_2 \text{ для всех возможных } (\mathbf{Y}, \mathbf{X}).$$

Структура называется *неидентифицируемой*, если она эквивалентна с точки зрения наблюдений другой структуре³⁹. (В предыдущем примере о потреблении, как и во всех случаях, когда модель не идентифицирована, существует бесконечное число структур, каждая из которых будет эквивалентна (10-46), по одной на каждое ненулевое значение a в (10-47).

Общая модель одновременных уравнений, которую мы специфицировали в (10-42), неидентифицируема. Неявно мы предполагали, что безусловное распределение \mathbf{X} может быть отделено от условного распределения $\mathbf{Y}|\mathbf{X}$. Мы можем записать эту модель как

$$p(\mathbf{Y}, \mathbf{X}|\Gamma, \mathbf{B}, \Sigma, \Theta) = p(\mathbf{Y}|\mathbf{X}, \Pi, \Omega)p(\mathbf{X}|\Theta) \text{ с } \Pi = -\mathbf{B}\Gamma^{-1} \\ \text{и } \Omega = (\Gamma')^{-1}\Sigma(\Gamma)^{-1}.$$

Мы предполагаем, что Θ и $(\Gamma, \mathbf{B}, \Sigma)$ не имеют общих элементов. Пусть \mathbf{F} будет любой невырожденной матрицей $M \times M$, и определим $\mathbf{B}_2 = \mathbf{F}\mathbf{B}$ и $\Gamma_2 = \mathbf{F}\Gamma$ и $\Sigma_2 = \mathbf{F}'\Sigma\mathbf{F}$ (т.е. мы просто умножили всю модель на \mathbf{F}). Если \mathbf{F} не равна единичной матрице, то \mathbf{B}_2 , Γ_2 и Σ_2 – это другие \mathbf{B} , Γ и Σ , соответствующие тем же данным, а значит, и тем же (\mathbf{Y}, \mathbf{X}) , которые подразумевают $(\Pi$ и $\Omega)$. Это происходит, поскольку $\Pi_2 = -\mathbf{B}_2\Gamma_2^{-1} = -\mathbf{B}\Gamma^{-1} = \Pi$, и точно так же для Ω_2 . Чтобы понять, что произойдет далее, обратим внимание на то, что в каждом уравнении есть «зависимая переменная», т.е. переменная, коэффициент при которой равен единице. Следовательно, один конкретный элемент из Γ в каждом уравнении (столбце) равен единице. Это исключает любую матрицу \mathbf{F} , в которой не остается единицы на этом месте

³⁹См. обзор по этой проблеме у Hsiao (1983).

в Γ_2 . Подобно этому в примере рыночного равновесия в подразделе 10.6.1 коэффициент при x в уравнении предложения равен нулю. Это значит, что в одном из столбцов матрицы \mathbf{B} есть элемент, равный нулю. Любая матрица \mathbf{F} , в которой нет такого нулевого ограничения, не подходит. Таким образом, определенные ограничения в модели, наложенные теорией, исключают некоторые альтернативные модели. Со всеми возможными ограничениями единственной подходящей матрицей \mathbf{F} будет $\mathbf{F} = \mathbf{I}$, и модель при этом становится идентифицируемой.

Структурная модель состоит из системы уравнений

$$y'\Gamma = -x'\mathbf{B} + \epsilon'$$

Каждый столбец в Γ и \mathbf{B} соответствует конкретному уравнению системы. Информация в выборке состоит в первую очередь из данных (\mathbf{Y}, \mathbf{X}) и другой информации не из выборки в виде ограничений для матриц параметров, таких как нормировки, приведенные в предыдущем примере. Данные выборки позволяют получить выборочные моменты $\mathbf{X}'\mathbf{X}/n$, $\mathbf{X}'\mathbf{Y}/n$ и $\mathbf{Y}'\mathbf{Y}/n$. Для осуществления идентификации, которая не зависит от объема выборки, представим, что мы могли бы рассматривать настолько большую выборку, насколько нам бы хотелось. Тогда мы могли бы рассмотреть [с (10-45)]

$$\text{plim}(1/n)\mathbf{X}'\mathbf{X} = \mathbf{Q},$$

$$\text{plim}(1/n)\mathbf{X}'\mathbf{Y} = \text{plim}(1/n)\mathbf{X}'(\mathbf{X}\Pi + \mathbf{V}) = \mathbf{Q}\Pi,$$

$$\text{plim}(1/n)\mathbf{Y}'\mathbf{Y} = \text{plim}(1/n)(\mathbf{X}\Pi + \mathbf{V})'(\mathbf{X}\Pi + \mathbf{V}) = \Pi'\mathbf{Q}\Pi + \mathbf{\Omega}.$$

Следовательно, матрица Π коэффициентов приведенной формы является наблюдаемой:

$$\Pi = [\text{plim}(1/n)\mathbf{X}'\mathbf{X}]^{-1}[\text{plim}(1/n)\mathbf{X}'\mathbf{Y}].$$

Эта оценка получена при помощи регрессии \mathbf{Y} на \mathbf{X} с использованием метода наименьших квадратов для каждого уравнения системы. Поскольку теперь Π является наблюдаемой, $\mathbf{\Omega}$ также равна

$$\mathbf{\Omega} = [\text{plim}(1/n)\mathbf{Y}'\mathbf{Y}] - [\text{plim}(1/n)\mathbf{Y}'\mathbf{X}][\text{plim}(1/n)\mathbf{X}'\mathbf{X}]^{-1}[\text{plim}(1/n)\mathbf{X}'\mathbf{Y}].$$

В этом результате следует признать дисперсионно-ковариационную матрицу остатков МНК-регрессии. Следовательно, Π и $\mathbf{\Omega}$ могут быть оценены соответственно при помощи МНК-регрессии \mathbf{Y} на \mathbf{X} . Имеющаяся информация, таким образом, состоит из Π , $\mathbf{\Omega}$ и любой другой вневыборочной информации о структуре⁴⁰.

Таким образом, Π и $\mathbf{\Omega}$ стали «наблюдаемыми». Последний вопрос: можем ли мы вывести Γ , \mathbf{B} , Σ из Π , $\mathbf{\Omega}$? Простое вычислительное упражнение сразу же показывает, что ответ отрицательный — нужно вывести M^2 параметров в Γ , $M(M+1)/2$ в Σ и KM в \mathbf{B} . Данные в выборке содержат KM элементов в Π и $M(M+1)/2$ элементов в $\mathbf{\Omega}$. Если просто посчитать уравнения и неизвестные, мы обнаружим, что наших данных недостаточно на M^2

⁴⁰Вовсе не обязательно, что это вся информация в выборке. В общем случае мы наблюдаем условное распределение $f(y_i|x_i)$, которое устанавливает правдоподобие для приведенной формы. Если шоки распределены нормально, такое распределение представляет собой функцию только из Π и $\mathbf{\Omega}$. В случае других распределений другие или более высокие моменты переменных могли бы давать дополнительную информацию. См., например, Goldberger (1964, p. 311), Hausman (1983, p. 402-403) и особенно Reiersol (1950).

элементов информации. Мы уже использовали информацию в выборке, поэтому эти M^2 дополнительных ограничений будут получены из теории, касающейся нашей модели. Небольшой пример поможет закрепить эти идеи.

Пример 10.5. Идентификация

Рассмотрим рынок, на котором q — выпуск товара Q , p — цена, а z — цена родственного товара Z . Мы предполагаем, что z входит в оба уравнения: и спроса, и предложения. К примеру, Z может быть сельскохозяйственной культурой, которую покупают потребители и которую будут выращивать фермеры вместо Q , если ее цена по отношению к p достаточно вырастет. То есть мы бы ожидали, что $\alpha_2 > 0$ и $\beta_2 < 0$. Таким образом,

$$\begin{aligned} q_d &= \alpha_0 + \alpha_1 p + \alpha_2 z + \varepsilon_d && \text{(спрос),} \\ q_s &= \beta_0 + \beta_1 p + \beta_2 z + \varepsilon_s && \text{(предложение),} \\ q_d &= q_s = q && \text{(равновесие).} \end{aligned}$$

Приведенная форма имеет вид

$$\begin{aligned} q &= \frac{\alpha_1 \beta_0 - \alpha_0 \beta_1}{\alpha_1 - \beta_1} + \frac{\alpha_1 \beta_2 - \alpha_2 \beta_1}{\alpha_1 - \beta_1} z + \frac{\alpha_1 \varepsilon_s - \beta_1 \varepsilon_d}{\alpha_1 - \beta_1} = \pi_{11} + \pi_{21} z + \nu_q, \\ p &= \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{\beta_2 - \alpha_2}{\alpha_1 - \beta_1} z + \frac{\varepsilon_s - \varepsilon_d}{\alpha_1 - \beta_1} = \pi_{12} + \pi_{22} z + \nu_p. \end{aligned}$$

При наличии только четырех коэффициентов приведенной формы и шести структурных параметров очевидно, что не будет полного решения для всех шести структурных параметров в терминах четырех параметров в приведенной форме. Но представим, что известно, что $\beta_2 = 0$ (фермеры не заменяют альтернативную культуру на эту). Тогда решением для β_1 будет π_{21} / π_{22} . После некоторых операций мы также получим $\beta_0 = \pi_{11} - \pi_{12} \pi_{21} / \pi_{22}$. Ограничение позволяет идентифицировать параметры предложения, но дальше этого шага мы пойти не сможем.

Теперь предположим, что вместо z в уравнении спроса появляется доход x . Новая модель будет иметь вид

$$\begin{aligned} q &= \alpha_0 + \alpha_1 p + \alpha_2 x + \varepsilon_1, \\ q &= \beta_0 + \beta_1 p + \beta_2 z + \varepsilon_2. \end{aligned}$$

Теперь структура станет следующей:

$$[q \ p] \begin{bmatrix} 1 & 1 \\ -\alpha_1 & -\beta_1 \end{bmatrix} + [1 \ x \ z] \begin{bmatrix} -\alpha_0 & -\beta_0 \\ -\alpha_2 & 0 \\ 0 & -\beta_2 \end{bmatrix} = [\varepsilon_1 \ \varepsilon_2],$$

а приведенная форма примет вид

$$[q \ p] = [1 \ x \ z] \begin{bmatrix} (\alpha_1 \beta_0 - \alpha_0 \beta_1) / \Delta & (\beta_0 - \alpha_0) / \Delta \\ -\alpha_2 \beta_1 / \Delta & -\alpha_2 / \Delta \\ \alpha_1 \beta_2 / \Delta & \beta_2 / \Delta \end{bmatrix} + [\nu_1 \ * \ \nu_2],$$

где $\Delta = (\alpha_1 - \beta_1)$. Любая ложная структура имеет такую же приведенную форму. Но в матрице коэффициентов

$$\tilde{\mathbf{B}} = \mathbf{BF} = \begin{bmatrix} \alpha_0 f_{11} + \beta_0 f_{21} & \alpha_0 f_{12} + \beta_0 f_{22} \\ \alpha_2 f_{11} & \alpha_2 f_{12} \\ \beta_2 f_{21} & \beta_2 f_{22} \end{bmatrix}.$$

Если f_{12} не равно нулю, то в уравнении предложения в ложной структуре возникнет доход, хотя мы его исключили из теоретических предположений модели. Подобно этому если f_{21} не нуль, то z возникнет в уравнении спроса, что тоже исключается нашей теорией. Таким образом, несмотря на то что все ложные структуры имеют одинаковый приведенный вид, что и истинная, единственная структура, соответствующая нашей теории (т.е. **приемлемая**) и имеющая коэффициенты 1 для q в обоих уравнениях (проверьте \mathbf{GF}), — это $\mathbf{F} = \mathbf{I}$. Это преобразование лишь воспроизводит первоначальную структуру.

Однозначные решения для структурных параметров, выраженные через параметры приведенной формы, теперь имеют вид

$$\begin{aligned} \alpha_0 &= \pi_{11} - \pi_{12} \left(\frac{\pi_{31}}{\pi_{32}} \right), & \beta_0 &= \pi_{11} - \pi_{12} \left(\frac{\pi_{21}}{\pi_{22}} \right), \\ \alpha_1 &= \frac{\pi_{31}}{\pi_{32}}, & \beta_1 &= \frac{\pi_{21}}{\pi_{22}}, \\ \alpha_2 &= \pi_{22} \left(\frac{\pi_{21}}{\pi_{22}} - \frac{\pi_{31}}{\pi_{32}} \right), & \beta_2 &= \pi_{32} \left(\frac{\pi_{31}}{\pi_{32}} - \frac{\pi_{21}}{\pi_{22}} \right). \end{aligned}$$

Можем заключить, что одни системы уравнений могут быть идентифицируемы, а другие нет. Формальные математические условия, при которых система уравнений идентифицируется, включают некоторые осложнения, известные как **условия ранга и порядка**.

Условие порядка — простое правило подсчета. В контексте системы уравнений условие порядка состоит в том, что число экзогенных переменных, возникающих где-либо в системе уравнений, должно быть, по крайней мере, таким же, как и число эндогенных переменных в уравнении. Мы использовали это правило при построении IV-оценки в главе 8. В том случае нам было необходимо, чтобы наша модель была «идентифицируема» и число инструментальных переменных, не содержащихся в \mathbf{X} , было таким же, как число эндогенных переменных. В этом примере приложение к случаю единственного уравнения с определенным здесь условием состоит в том, что оставшаяся часть системы уравнений, по сути, весь остальной мир (т.е. источник инструментальных переменных)⁴¹. Простое достаточное условие порядка для системы уравнений таково, что каждое уравнение должно содержать «свою собственную» экзогенную переменную, которая не возникает больше нигде в системе.

⁴¹ Это рождает извечный вопрос (неоднократно упоминаемый в приложениях главы 8): откуда берутся инструментальные переменные? См. раздел 8.8 по этой теме.

Условие порядка является необходимым для идентификации, а **условие ранга** — достаточным. Структурная форма системы уравнений в (10-42) имеет вид $y'\Gamma = -x'V + \epsilon'$. Приведенная форма выглядит так: $y' = x'(-V\Gamma^{-1}) + \epsilon'\Gamma^{-1} = x'\Pi + v'$. То, как мы собираемся вывести параметры в (Γ, V, Σ) , следует из параметров в приведенной форме (Π, Ω) . Для конкретного уравнения, скажем j -го, решение содержится в $\Pi\Gamma = -V$, т.е. в $\Pi\Gamma_j = -V_j$ для конкретного уравнения, где Γ_j включает все коэффициенты в j -м уравнении, на которые умножаются эндогенные переменные. Один из этих коэффициентов будет равен единице, некоторые из них, как правило, будут равны нулю, а остальные — ненулевые коэффициенты эндогенных переменных в уравнении Y_j [далее в (10-48) они обозначаются как γ_j]. Подобно этому V_j содержит коэффициенты в уравнении j для всех экзогенных переменных в модели — некоторые из них будут нулевыми, а оставшиеся будут умножаться на переменные в X_j , экзогенные переменные, возникающие в этом уравнении [далее в (10-48) они обозначаются как β_j]. Эмпирический аналог имеет вид:

$$[\text{plim}(1/n)X'X]^{-1}[\text{plim}(1/n)X'Y]\Gamma_j - V_j = 0.$$

Условие ранга обеспечивает наличие единственного решения для этого набора уравнений. На практике условие ранга сложно установить для больших систем уравнений. Обычно это принимается как данность. В небольших системах с двумя или тремя уравнениями, доминирующих в современных исследованиях, это несложно. Мы уже использовали условие ранга в главе 8, где оно выступало в роли условия «релевантности» для оценки по инструментальным переменным. Так, обратите внимание, что после обозначения допущений для оценки по инструментальным переменным мы предположили, что $\text{plim}(1/n)Z'X$ — матрица с рангом K . (В современных исследованиях его часто называют условием ранга. Оно неидентично, но достаточно для выполнения условия, рассматриваемого здесь.)

В добавление ко всему сказанному важно вернуться к условию порядка. Мы пытаемся решить систему моментных уравнений, связывающих параметры структурной и приведенной форм. Информация в выборке дает $KM + M(M + 1)/2$ элементов в Π и Ω . Нам требуется M^2 дополнительных **ограничений**, следующих из теории для модели. Ограничения представлены в виде нормировок, как правило **исключающих ограничений**, а также других связей между параметрами, к примеру линейной связи или особых значений коэффициентов. Вопрос идентификации — это теоретическое упражнение. Он возникает во всех экономических задачах, где параметры модели должны быть выведены из сочетания информации в выборке и вне выборки (теоретической информации). Главный вопрос в таких случаях — наша способность (или ее отсутствие) выводить значения параметров структурной формы однозначно из информации, присутствующей в выборке, в терминах выборочных моментов в сочетании с **информацией вне выборки**, включающей в основном ограничения на значения параметров. Проблема идентификации — предмет очень большого числа работ, включая работы Working (1927) (эта работа была использована для рис. 8.1), Gabrielsen (1978), Amemiya (1985), Bekker и Wansbeek (2001), что имело продолжение в современных естественных экспериментах (раздел 8.8 и Angrist, Pischke (2010) с комментариями).

10.6.4. Оценка одного уравнения и проверка гипотез

Для того чтобы оценить и сделать выводы, запишем систему одновременных уравнений в том виде, в котором обычно она записывается в исследованиях:

$$y_j = X_j\beta_j + Y_j\gamma_j + \epsilon_j = Z_j\delta_j + \epsilon_j, \quad (10-48)$$

где y_j — «зависимая переменная» в уравнении, X_j — набор объясняющих переменных в j -м уравнении — обратите внимание, что это не все переменные в модели, — и $Z_j = (X_j, Y_j)$. Полный набор экзогенных переменных в модели, включая X_j и переменные, присутствующие где-либо в модели (в том числе постоянный член, если его содержит какое-либо уравнение), обозначается как X . К примеру, в модели спроса/предложения в примере 10.5 полный набор экзогенных переменных имеет вид $X = (1, x, z)$, в то время как для уравнения предложения $X_{Demand} = (1, x)$ и $X_{Supply} = (1, z)$. Наконец, Y_j — эндогенные переменные, присутствующие в правой части j -го уравнения. Опять же это, вероятно, будет подмножеством эндогенных переменных в полной модели. В примере 10.5 это $Y_j = (\text{price})$ в обоих случаях.

Существует два подхода к оценке систем одновременных уравнений. **Оценки при ограниченной информации** строятся для каждого уравнения отдельно. Этот подход аналогичен оценке системы внешне не связанных уравнений из раздела 10.2 методом наименьших квадратов отдельно каждого уравнения системы. **Оценки при полной информации** используются для оценки всех уравнений одновременно. Аналогом для системы внешне не связанных уравнений является оценка при помощи допустимого обобщенного метода наименьших квадратов, рассмотренного в подразделе 10.2.3. Главное отличие, важное здесь, — это эндогенность Y_j в (10-48).

Система уравнений в (10-48) — это в точности модель, рассмотренная в главе 8. Метод наименьших квадратов в большинстве случаев будет давать несостоятельные оценки из-за корреляции между Y_j и ϵ_j . Обычно применяется двухшаговый метод наименьших квадратов, как показано в подразделах 8.3.2–8.3.4. Единственное различие между случаем, рассматриваемым здесь, и случаем из главы 8 — это источник инструментальных переменных. В общей модели из главы 8 их источником оставалось нечто неопределенное; общее правило было «вне модели». В случае же, рассматриваемом здесь, инструментальные переменные берутся откуда-то в модели, т.е. «не в j -м уравнении». Таким образом, наиболее распространенная оценка для модели одновременных уравнений имеет вид

$$\begin{aligned} \hat{\delta}_{j,2SLS} &= [\hat{Z}'_j \hat{Z}_j]^{-1} \hat{Z}'_j y_j \\ &= [(Z'_j X)(X'X)^{-1}(X'Z_j)]^{-1} (Z'_j X)(X'X)^{-1} X'y_j, \end{aligned} \quad (10-49)$$

где все столбцы \hat{Z}'_j получены как прогнозы в регрессии соответствующего столбца Z_j на X . Это уравнение также дает удобные упрощения оцененной асимптотической ковариационной матрицы

$$\text{Est.Asy. Var}[1\hat{\delta}_{j,2SLS}] = \hat{\sigma}_{jj} [\hat{Z}'_j \hat{Z}_j]^{-1}.$$

Важно отметить, что σ_{jj} оценивается так:

$$\hat{\sigma}_{jj} = \frac{(y_j - Z_{j1}\hat{\delta}_j)'(y_j - Z_{j1}\hat{\delta}_j)}{T} \quad (10-50)$$

с использованием исходных данных, а не \hat{Z}_j .

Здесь необходимо обратить внимание на роль условия порядка для идентификации в оценке, полученной при помощи двухшагового метода наименьших квадратов. Формально условие порядка требует, чтобы число экзогенных переменных, появляющихся где-либо в модели (не в этом уравнении), было, по крайней мере, таким же, как число эндогенных переменных, присутствующих в этом уравнении. Подразумевается, что мы будем предсказывать $Z_j = (X_j, Y_j)$, используя $X = (X_j, X_j^*)$. Чтобы эти предсказания были линейно независимыми, используемых для их вычисления переменных должно быть столько же, сколько и переменных, которые требуется предсказать. Из сравнения (X_j, Y_j) и (X_j, X_j^*) мы видим, что переменных в X_j^* должно быть, по крайней мере, столько же, сколько и в Y_j , что является условием порядка. Практическое правило о том, что каждое уравнение имеет как минимум одну переменную, которая не присутствует в любом другом уравнении, будет гарантировать выполнение этого условия. Двухшаговый метод наименьших квадратов почти всегда используется при оценке моделей одновременных уравнений по причинам, отмеченным в главе 8. Однако в некоторых приложениях (и теоретических работах) указывается, что оценка **методом максимального правдоподобия при ограниченной информации (LIML-оценка)**, основанная на нормальном распределении, может обладать более предпочтительными свойствами. Такая техника уже нашла применение в современных работах о слабых инструментах переменных, рассмотренных в подразделе 10.6.5. Вывод логарифмической функции правдоподобия представлен в книге Davidson, MacKinnon (2004). Мы же сосредоточимся на практической стороне этой оценки и направим читателя к этому источнику для ознакомления с особенностями техники. Результатом является вывод о том, что LIML-оценка имеет такое же асимптотическое распределение, как и 2МНК-оценка, и последняя не опирается на предположение о нормальности. Тогда возникает вопрос, зачем использовать технику LIML, если доступна более надежная (и более простая в вычислении) альтернатива. Результаты для малых выборок немногочисленны, но и здесь предпочтительнее 2МНК. [См. Phillips (1983).] Одним важным достоинством LIML-оценки является ее независимость от нормировки уравнения. Рассмотрим пример в системе уравнений

$$y_1 = y_2\gamma_2 + y_3\gamma_3 + x_1\beta_1 + x_2\beta_2 + \varepsilon_1.$$

Эквивалентное уравнение имеет вид

$$\begin{aligned} y_2 &= y_1(1/\gamma_2) + y_3(-\gamma_3/\gamma_2) + x_1(-\beta_1/\gamma_2) + x_2(-\beta_2/\gamma_2) + \varepsilon_1(-1/\gamma_2) \\ &= y_1\tilde{\gamma}_1 + y_3\tilde{\gamma}_3 + x_1\tilde{\beta}_1 + x_2\tilde{\beta}_2 + \tilde{\varepsilon}_1. \end{aligned}$$

Чтобы получить параметры первого уравнения, можно сначала после некоторых преобразований получить параметры второго. Но, как можно легко

проверить, 2МНК-оценка не неизменна по отношению к нормировке уравнения – 2МНК даст численно другие результаты. LIML же даст одни и те же числовые решения для обеих задач оценивания, рассмотренных выше. Вторым преимуществом является лучшее поведение LIML-оценки при наличии слабых инструментов.

LIML-оценка, или оценка **отношения наименьших дисперсий**, может быть вычислена следующим образом⁴². Пусть

$$\mathbf{W}_j^0 = \mathbf{E}_j^{0'} \mathbf{E}_j^0, \quad (10-51)$$

где

$$\mathbf{Y}_j^0 = [y_j, \mathbf{Y}_j]$$

и

$$\mathbf{E}_j^0 = \mathbf{M}_j \mathbf{Y}_j^0 = [\mathbf{I} - \mathbf{X}_j (\mathbf{X}_j' \mathbf{X}_j)^{-1} \mathbf{X}_j'] \mathbf{Y}_j^0. \quad (10-52)$$

Каждый столбец \mathbf{E}_j^0 – это остатки метода наименьших квадратов в регрессии соответствующего столбца в \mathbf{Y}_j^0 на \mathbf{X}_j , т.е. это экзогенные переменные, появляющиеся в j -м уравнении. Итак, \mathbf{W}_j^0 – это матрица сумм квадратов остатков и их перекрестных произведений. Определим

$$\mathbf{W}_j^1 = \mathbf{E}_j^{1'} \mathbf{E}_j^1 = \mathbf{Y}_j^{0'} [\mathbf{I} - \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}'] \mathbf{Y}_j^0. \quad (10-53)$$

То есть \mathbf{W}_j^1 находится подобно \mathbf{W}_j^0 , но с одним отличием: регрессии оцениваются на все x в модели, а не только на те, которые есть в j -м уравнении. Пусть

$$\lambda_1 = \text{наименьший характеристический корень } (\mathbf{W}_j^1)^{-1} \mathbf{W}_j^0. \quad (10-54)$$

Эта матрица асимптотическая, но все ее корни реально существующие и больше единицы или равны единице. В зависимости от доступной программы может быть удобнее получить тождественный наименьший корень симметричной матрицы $\mathbf{D} = (\mathbf{W}_j^1)^{-1/2} \mathbf{W}_j^0 (\mathbf{W}_j^1)^{-1/2}$. Теперь разобьем \mathbf{W}_j^0 на

$\mathbf{W}_j^0 = \begin{bmatrix} w_{jj}^0 & w_j^{0'} \\ w_j^0 & w_{jj}^0 \end{bmatrix}$ в соответствии с $[y_j, \mathbf{Y}_j]$ и так же разобьем \mathbf{W}_j^1 . Далее, имея эти части, получим

$$\hat{\gamma}_{j,\text{LIML}} = [\mathbf{W}_{jj}^0 - \lambda_1 \mathbf{W}_{jj}^1]^{-1} (w_j^0 - \lambda_1 w_j^1) \quad (10-55)$$

и

$$\hat{\beta}_{j,\text{LIML}} = [\mathbf{X}_j' \mathbf{X}_j]^{-1} \mathbf{X}_j' (y_j - \mathbf{Y}_j \hat{\gamma}_{j,\text{LIML}}).$$

Обратите внимание, что β_j оценено при помощи простой регрессии методом наименьших квадратов. [См. (3-18).] Асимптотическая ковариационная матрица для LIML-оценки тождественна матрице для 2МНК-оценки⁴³.

⁴²Оценка по критерию наименьших дисперсий выведена, к примеру, у Johnston (1984). LIML-оценка была получена у Anderson, Rubin (1949, 1950). С момента ее получения Anderson, Rubin в 1949 и 1950 гг. LIML-оценка была предметом лишь теоретического интереса. Гораздо более простая и настолько же эффективная оценка по двухступенчатому методу наименьших квадратов использовалась в качестве оценки выбора. Но LIML- и AR-тесты на спецификацию были вновь открыты и активизированы с момента их использования в анализе слабых инструментальных переменных. См. Hahn, Hausman (2002, 2003), а также разделы 8.7 и 10.6.6.

⁴³Чтобы доказать это, нужно показать, что обе оценки принадлежат к оценкам « k -класса», которые имеют одинаковые асимптотические ковариационные матрицы. Подробности даны в работах Theil (1971) и Schmidt (1976).

Результатом является то, что при нормально распределенных шоках 2МНК-оценка полностью эффективна.

Оценки **k-класса** определяются по следующей формуле:

$$\hat{\delta}_{j,k} = \begin{pmatrix} \hat{\gamma}_{j,k} \hat{\beta}_{j,k} \end{pmatrix} = \begin{bmatrix} \mathbf{Y}'_j \mathbf{Y}_j - k \mathbf{V}'_j \mathbf{V}_j & \mathbf{Y}'_j \mathbf{X}_j \\ \mathbf{X}'_j \mathbf{Y}_j & \mathbf{X}'_j \mathbf{X}_j \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{Y}'_j \mathbf{y}_j - k \mathbf{V}'_j \mathbf{v}_j \\ \mathbf{X}'_j \mathbf{y}_j \end{bmatrix}, \quad (10-56)$$

где \mathbf{V}_j и \mathbf{v}_j — шоки приведенной формы из (10-45). Допустимая оценка вычисляется с использованием остатков из МНК-регрессий \mathbf{Y}_j и \mathbf{y}_j на \mathbf{X} (не \mathbf{X}_j). Мы уже рассматривали три члена этого класса оценок — МНК с $k = 0$, 2МНК с $k = 1$ и, что можно показать, LIML с $k = \lambda_1$. [Последний результат следует из (10-55).] Помимо таких оценок k -класса были получены и многие другие; этот вопрос рассматривается в книге Davidson, MacKinnon (2004, р. 537–538 и 548–549) и работе Mariano (2001). Было показано, что все члены класса k , для которых k сходится к 1 значительно быстрее, чем $1/\sqrt{n}$, имеют одинаковое асимптотическое распределение, подобное распределению 2МНК-оценки, рассмотренной ранее. Такие оценки вызывают больший теоретический интерес, учитывая повсеместное использование 2МНК или МНК, если бы не одно замечание. Свойства всех оценок k -класса при больших выборках одинаковы, однако свойства при конечных выборках, возможно, очень различаются. Davidson, MacKinnon (2004, р. 537–538 и 548–549) и Mariano (1982, 2001) утверждают, что существуют некоторые основания отдать предпочтение LIML, если размер выборки мал или средний, а число сверхидентифицирующих ограничений относительно велико.

10.6.5. Методы оценки систем уравнений

Можно выразить полную систему уравнений в виде

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} Z_1 0 \cdots 0 & Z_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & Z_M \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_M \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_M \end{bmatrix} \quad (10-57)$$

$$\text{или } \mathbf{y} = \mathbf{Z}\boldsymbol{\delta} + \boldsymbol{\epsilon},$$

где

$$E[\boldsymbol{\epsilon} | \mathbf{X}] = \mathbf{0} \quad \text{и} \quad E[\boldsymbol{\epsilon}\boldsymbol{\epsilon}' | \mathbf{X}] = \bar{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} \otimes \mathbf{I}. \quad (10-58)$$

[См. (10-6).] МНК-оценка

$$\mathbf{d} = [\mathbf{Z}'\mathbf{Z}]^{-1} \mathbf{Z}'\mathbf{y}$$

находится при помощи оценки простым методом наименьших квадратов отдельно каждого уравнения и будет несостоятельной. Но даже если бы оценка простым методом наименьших квадратов была состоятельна, мы знаем из наших выводов для систем внешне не связанных уравнений, что она будет неэффективна по сравнению с оценкой, использующей перекрестные корреляции шоков в уравнениях. По первому вопросу снова обратимся к IV-оценке. По второму — и это следует из подраздела 10.2.1 — мы используем

обобщенный метод наименьших квадратов. Итак, допуская, что матрица инструментальных переменных $\bar{\mathbf{W}}$ удовлетворяет требованиям IV-оценки, состоятельная (хоть и неэффективная) оценка будет иметь вид

$$\hat{\delta}_{IV} = [\bar{\mathbf{W}}' \mathbf{Z}]^{-1} \bar{\mathbf{W}}' \mathbf{y}. \quad (10-59)$$

По аналогии с моделью внешне не связанных регрессий более эффективной была бы оценка, построенная на основе обобщенного метода наименьших квадратов

$$\hat{\delta}_{IV, GLS} = [\bar{\mathbf{W}}' (\Sigma^{-1} \otimes \mathbf{I}) \mathbf{Z}]^{-1} \bar{\mathbf{W}}' (\Sigma^{-1} \otimes \mathbf{I}) \mathbf{y} \quad (10-60)$$

или где \mathbf{W}_j — множество инструментальных переменных для j -го уравнения:

$$\hat{\delta}_{IV, GLS} = \begin{bmatrix} \sigma^{11} \mathbf{W}'_1 \mathbf{Z}_1 & \sigma^{12} \mathbf{W}'_1 \mathbf{Z}_2 & \cdots & \sigma^{1M} \mathbf{W}'_1 \mathbf{Z}_M \\ \sigma^{21} \mathbf{W}'_2 \mathbf{Z}_1 & \sigma^{22} \mathbf{W}'_2 \mathbf{Z}_2 & \cdots & \sigma^{2M} \mathbf{W}'_2 \mathbf{Z}_M \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^{M1} \mathbf{W}'_M \mathbf{Z}_1 & \sigma^{M2} \mathbf{W}'_M \mathbf{Z}_2 & \cdots & \sigma^{MM} \mathbf{W}'_M \mathbf{Z}_M \end{bmatrix}^{-1} \times$$

$$\times \begin{bmatrix} \sum_{j=1}^M \sigma^{1j} \mathbf{W}'_1 \mathbf{y}_j \\ \sum_{j=1}^M \sigma^{2j} \mathbf{W}'_2 \mathbf{y}_j \\ \vdots \\ \sum_{j=1}^M \sigma^{Mj} \mathbf{W}'_M \mathbf{y}_j \end{bmatrix}.$$

Для совместной оценки всей системы уравнений обычно используются три метода IV: трехшаговый метод наименьших квадратов, GMM и критерий максимального правдоподобия при полной информации (FIML). Здесь мы рассмотрим трехшаговый метод наименьших квадратов. GMM и FIML будут рассмотрены в главах 13 и 14 соответственно.

Рассмотрим IV-оценку, построенную из

$$\bar{\mathbf{W}} = \hat{\mathbf{Z}} = \text{diag}[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}_1, \dots, \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}_M] = \begin{bmatrix} \hat{\mathbf{Z}}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{Z}}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \hat{\mathbf{Z}}_M \end{bmatrix}.$$

IV-оценка

$$\hat{\delta}_{IV} = [\hat{\mathbf{Z}}' \mathbf{Z}]^{-1} \hat{\mathbf{Z}}' \mathbf{y}$$

представляет собой простую 2МНК каждого уравнения в отдельности. Мы уже обозначили состоятельность для 2МНК. По аналогии с моделью внешне не связанных регрессий из раздела 10.2 мы, однако, ожидаем, что эта оценка будет менее эффективной, нежели GLS-оценка. Самым подходящим вариантом будет

$$\hat{\delta}_{3SLS} = [\hat{\mathbf{Z}}' (\Sigma^{-1} \otimes \mathbf{I}) \mathbf{Z}]^{-1} \hat{\mathbf{Z}}' (\Sigma^{-1} \otimes \mathbf{I}) \mathbf{y}.$$

Для того чтобы эта оценка была обоснованной IV-оценкой, мы должны показать, что

$$\text{plim} \frac{1}{T} \hat{\mathbf{Z}}' (\Sigma^{-1} \otimes \mathbf{I}) \boldsymbol{\varepsilon} = \mathbf{0},$$

что представляет собой M наборов уравнений, каждое из которых имеет форму

$$\text{plim} \frac{1}{T} \sum_{j=1}^M \sigma^{ij} \hat{\mathbf{Z}}_i' \boldsymbol{\varepsilon}_j = 0.$$

Каждое уравнение — это сумма векторов, каждый из которых сходится к нулю, как мы видели в ходе изучения 2МНК-оценки. Второе требование, что

$$\text{plim} \frac{1}{T} \hat{\mathbf{Z}}' (\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}) \mathbf{Z} \neq 0$$

и что матрица невырожденная, может быть доказано аналогично случаю 2МНК. Идентификация каждого уравнения по условию ранга является достаточной. [Однако см. Mariano (2001) по вопросу «слабых инструментов.»]

Опять же используя идемпотентность $\mathbf{I} - \mathbf{M}$, мы также можем представить эту оценку как GLS-оценку вида

$$\hat{\boldsymbol{\delta}}_{3\text{SLS}} = [\hat{\mathbf{Z}}' (\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}) \hat{\mathbf{Z}}]^{-1} \hat{\mathbf{Z}}' (\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}) \mathbf{y}. \quad (10-61)$$

Соответствующая асимптотическая ковариационная матрица для этой оценки такова:

$$\text{Asy. Var}[\hat{\boldsymbol{\delta}}_{3\text{SLS}}] = [\bar{\mathbf{Z}}' (\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}) \bar{\mathbf{Z}}]^{-1}, \quad (10-62)$$

где $\bar{\mathbf{Z}} = \text{diag}[\mathbf{X}\boldsymbol{\Pi}_j, \mathbf{X}_j]$. Эта матрица будет оцениваться как обратная матрица в скобках из (10-61).

Используя данные выборки, находим, что $\bar{\mathbf{Z}}$ может быть оценена как $\hat{\mathbf{Z}}$. Оставшаяся трудность — получить оценку для $\boldsymbol{\Sigma}$. При оценке многомерной регрессионной модели для получения эффективной оценки подойдет любая состоятельная оценка для $\boldsymbol{\Sigma}$. Разработчики 3МНК-метода Зельнер и Тейл (Zellner, Theil (1962)) предложили естественный выход, следующий из оценки двухшаговым методом наименьших квадратов. Тогда оценка **трехшаговым методом наименьших квадратов (3МНК)** определяется следующим образом:

1. Оцените $\boldsymbol{\Pi}$ по простому методу наименьших квадратов и вычислите $\hat{\mathbf{Y}}_j$ для каждого уравнения.
2. Вычислите $\hat{\boldsymbol{\delta}}_{j,2\text{SLS}}$ для каждого уравнения, затем

$$\hat{\sigma}_{ij} = \frac{(\mathbf{y}_i - \mathbf{Z}_i \hat{\boldsymbol{\delta}}_i)' (\mathbf{y}_j - \mathbf{Z}_j \hat{\boldsymbol{\delta}}_j)}{T}. \quad (10-63)$$

3. Вычислите GLS-оценку по (10-61) и оценку асимптотической ковариационной матрицы по (10-62), используя $\hat{\mathbf{Z}}$ и $\hat{\boldsymbol{\Sigma}}$.

Возможно также повторить вычисление 3МНК. Однако в отличие от случая оценки систем внешне не связанных уравнений этот метод не позволяет получить оценку по критерию максимального правдоподобия, а также не повышает асимптотическую эффективность⁴⁴.

⁴⁴Член якобиана, необходимый для максимизации логарифмической функции правдоподобия, нельзя получить с помощью 3МНК-оценки. Dhrymes (1973).

Показав, что 3МНК-оценка удовлетворяет требованиям для IV-оценки, мы установили ее постоянство. Вопрос асимптотической эффективности остается в силе. Можно показать, что среди всех IV-оценок, которые используют только информацию выборки, содержащуюся в системе, 3МНК асимптотически эффективна⁴⁵. При нормально распределенных шоках также можно показать, что 3МНК имеет такое же асимптотическое распределение, как оценка по критерию максимального правдоподобия при полной информации.

Пример 10.6. Модель I Клейна

Модель I Клейна (1950) — широко распространенный пример системы одновременных уравнений в экономике. Модель можно записать так:

$$C_t = \alpha_0 + \alpha_1 P_t + \alpha_2 P_{t-1} + \alpha_3 (W_t^p + W_t^g) + \varepsilon_{1t} \quad (\text{потребление}),$$

$$I_t = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + \beta_3 K_{t-1} + \varepsilon_{2t} \quad (\text{инвестиции}),$$

$$W_t^p = \gamma_0 + \gamma_1 X_t + \gamma_2 X_{t-1} + \gamma_3 A_t + \varepsilon_{3t} \quad (\text{частные зарплаты}),$$

$$X_t = C_t + I_t + G_t \quad (\text{равновесный спрос}),$$

$$P_t = X_t - T_t - W_t^p \quad (\text{частные прибыли}),$$

$$K_t = K_{t-1} + I_t \quad (\text{акционерный капитал}).$$

Все эндогенные переменные расположены в левой части уравнений, а их названия указаны справа. Экзогенные переменные — это G_t = государственные расходы без заработной платы, T_t = косвенные налоги на бизнес плюс чистый экспорт, W_t^g = государственные счета по заработной плате, A_t = временной тренд, измеренный в годах с 1931 г., и постоянный член. Также есть три предопределенные переменные: запаздывающие значения основного капитала, частных прибылей и общего спроса. Модель содержит три **поведенческих уравнения**, **условие равновесия** и два расчетных тождества. Модель является отличным примером небольшой динамической модели в экономике. Она всегда широко применялась в качестве тестовой модели для оценок систем одновременных уравнений. Клейн оценивал параметры, используя годовые данные с 1921 по 1941 г. Данные приведены в табл. F10.3 приложения. В табл. 10.5 представлены оценки при ограниченной и полной информации для модели I Клейна, основанные на исходных данных за 1920–1941 гг.⁴⁶

⁴⁵См. доказательство эффективности относительно 2МНК в работе Schmidt (1976).

⁴⁶Асимптотическая ковариационная матрица для LIML-оценки будет отличаться от аналогичной матрицы для оценки 2МНК при ограниченной выборке, поскольку оценка σ_{jj} , на которую умножается обратная матрица, будет другой, а также поскольку в вычислении обращаемой матрицы значение « k » [см. уравнение после (10-55)] равно единице для 2МНК и наименьшему корню из (10-54) для LIML. Асимптотически k равно единице и оценки σ_{jj} для эквиваленты.

Таблица 10.5. Оценки модели I Клейна (в скобках приведены оцененные асимптотические стандартные ошибки)

| | Оценки при ограниченной информации | | | | Оценки при полной информации | | | |
|----------------------|------------------------------------|-------------------|------------------|-------------------|------------------------------|-------------------|------------------|-------------------|
| | 2МНК | | | | 3МНК | | | |
| <i>C</i> | 16,6 (1,32) | 0,017 (0,118) | 0,216 (0,107) | 0,810 (0,040) | 16,4 (1,30) | 0,125 (0,108) | 0,163 (0,100) | 0,790 (0,038) |
| <i>I</i> | 20,3 (7,54) | 0,150 (0,173) | 0,616 (0,162) | -0,158 (0,036) | 28,2 (6,79) | -0,013 (0,162) | 0,756 (0,153) | -0,195 (0,033) |
| <i>W^p</i> | 1,50 (1,15) | 0,439 (0,036) | 0,147 (0,039) | 0,130 (0,029) | 1,80 (1,12) | 0,400 (0,032) | 0,181 (0,034) | 0,150 (0,028) |
| | LIML | | | | FIML | | | |
| <i>C</i> | 17,1 (1,84) | -0,222 (0,202) | 0,396 (0,174) | 0,823 (0,055) | 18,3 (2,49) | -0,232 (0,312) | 0,388 (0,217) | 0,802 (0,036) |
| <i>I</i> | 22,6 (9,24) | 0,075 (0,219) | 0,680 (0,203) | -0,168 (0,044) | 27,3 (7,94) | -0,801 (0,491) | 1,052 (0,353) | -0,146 (0,30) |
| <i>W^p</i> | 1,53 (2,40) | 0,434 (0,137) | 0,151 (0,135) | 0,132 (0,065) | 5,79 (1,80) | 0,234 (0,049) | 0,285 (0,045) | 0,235 (0,035) |
| | МНК | | | | I3МНК | | | |
| <i>C</i> | 16,2 (1,30) | 0,193 (0,091) | 0,090 (0,091) | 0,796 (0,040) | 16,6 (1,22) | 0,165 (0,096) | 0,177 (0,090) | 0,766 (0,035) |
| <i>I</i> | 10,1 (5,47) | 0,480 (0,097) | 0,333 (0,101) | -0,112 (0,027) | 42,9 (10,6) | -0,356 (0,260) | 1,01 (0,249) | -0,260 (0,051) |
| <i>W^p</i> | 1,50 (1,27) | 0,439 (0,032) | 0,146 (0,037) | 0,130 (0,032) | 2,62 (1,20) | 0,375 (0,031) | 0,194 (0,032) | 0,168 (0,029) |

В свете всей дискуссии может показаться, что одна из структурных оценок, описанных ранее, должна всегда быть предпочтительнее оценки простым методом наименьших квадратов, которая, единственная из рассмотренных здесь оценок, является несостоятельной. Но, к сожалению, этот вопрос не так уж очевиден. Во-первых, часто оказывается, что МНК-оценка удивительно близка к структурной оценке. Можно показать, что МНК имеет меньшую дисперсию вокруг своего среднего, чем 2МНК вокруг своего, по крайней мере, в нескольких случаях это ведет к тому, что МНК мог бы быть более точным в отношении среднеквадратических ошибок⁴⁷. Но такое заключение может омрачиться тем, что стандартные ошибки МНК, по всей вероятности, нельзя использовать для получения выводов, основанных на этой оценке⁴⁸. Тем не менее МНК — это широко используемая оценка. Очевидно, что эта дискуссия имеет отношение лишь к конечным выборкам. Асимптотически 2МНК должна подавлять МНК, и в корректно специфицированной модели любая оценка при полной информации должна доминировать над любой оценкой при ограниченной информации. Свойства

⁴⁷Goldberger (1964, p. 359–360).

⁴⁸Cragg (1967).

в конечных выборках представляют особую важность. Почти все, что мы знаем, — асимптотические свойства, но наибольшее число примеров основано на довольно малых выборках или же выборках среднего размера.

Большие различия между несостоятельной МНК-оценкой и другими оценками предполагают смещение, о котором говорилось ранее. С другой стороны, неправильные знаки у LIML- и FIML-оценок коэффициента при P и еще большее отличие коэффициента при P_{-1} в уравнении C поразительны. При допущении, что уравнение правильно специфицировано, такие аномалии, вероятнее всего, будут связаны со свойствами конечной выборки, поскольку LIML и 2МНК асимптотически эквивалентны.

Интуитивно кажется, что методы оценки систем 3МНК и FIML должны быть более предпочтительными по сравнению с методами оценки отдельного уравнения 2МНК и LIML. Действительно, если преимущество настолько очевидно, зачем тогда вообще использовать оценки отдельных уравнений? Верная аналогия — использование МНК для отдельного уравнения против GLS в модели внешне не связанных регрессий из раздела 10.2. Очевидное практическое соображение — простота в вычислении методов оценки отдельного уравнения. Однако возможности современного программного обеспечения сняли это преимущество.

Несмотря на то что методы оценки систем уравнений предпочтительнее асимптотически, в них заключены две проблемы. Во-первых, любая ошибка спецификации в структуре модели будет распространяться по всей системе через 3МНК или FIML. Оценки при ограниченной информации будут в общем и целом сводить эту проблему к конкретному уравнению, в котором она возникнет. Во-вторых, так же как и в модели внешне не связанных регрессий, вариация конечной выборки оцененной ковариационной матрицы передается через всю систему. Так, вариация конечной выборки у 3МНК вполне может оказаться такой же или даже большей, чем аналогичная вариация у 2МНК. Результаты для модели I Клейна, хотя и являются оценками отдельных уравнений, дают поразительный пример. Результат вполне мог оказаться таким, что преимущество оценок систем в конечных выборках могло бы быть более скромным, чем следует из асимптотических результатов. Исследование этой проблемы методом Монте-Карло приводит к аналогичному заключению⁴⁹.

10.6.6. Тестирование при наличии слабых инструментальных переменных

В разделе 8.7 мы обозначили проблемы оценки и получения выводов при оценке методом инструментальных переменных при наличии **слабых инструментов**. Метод оценки регрессии на первом шаге Стэйджера и Стока (Staiger, Stock (1997)) часто используется для обнаружения условий. Предлагались и другие тесты, особенно тест Хана–Хаусмана (Hahn, Hausman (2002, 2003)). Рассмотрим уравнение с одной эндогенной переменной в правой части

$$y_1 = \gamma y_2 + x_1' \beta_1 + \varepsilon_1.$$

⁴⁹См. Cragg (1967) и многие другие работы, приведенные у Judge et al. (1985, p. 646–653).

Учитывая вид модели, расположение y_1 в левой части этого уравнения и y_2 в правой отражает не более чем нормировку коэффициента матрицы Γ в (10-42). На время назовем это уравнение прямым. Если мы перенормируем модель относительно y_2 , то получим полностью равносильное уравнение

$$\begin{aligned} y_2 &= (1/\gamma)y_1 + \mathbf{x}'_1(-\beta_1/\gamma) + (-\varepsilon_1/\gamma) \\ &= \theta y_1 + \mathbf{x}'_1 \lambda_1 + v_1, \end{aligned}$$

которое мы [а точнее, Хан и Хаусман (Hahn, Hausman (2002))] назовем обращенным уравнением. В принципе для оценки γ не должно иметь никакого значения, в какой форме оценивать уравнение; мы можем оценить γ напрямую в первом уравнении или косвенно посредством $1/\theta$ во втором. Однако на практике из всех оценок k -класса, приведенных в подразделе 10.6.4, включающих все оценки, которые мы рассматривали, только LIML-оценка инвариантна к перенормировке, а 2МНК-оценка, конечно, нет. Если мы рассмотрим 2МНК-оценку прямого уравнения $\hat{\gamma}$ и оценку обращенного уравнения $1/\hat{\theta}$, то полученные нами оценки в принципе должны быть схожими. Однако в 2МНК-оценке есть смещение, которое становится более выраженным, когда инструменты становятся слабее. Статистический тест Хана–Хаусмана основан на разности между двумя оценками (с поправкой на известное смещение 2МНК-оценки в этом случае). [Исследования этого и других тестов продолжаются. Hausman, Stock, Yogo (2005) представляют довольно разочаровывающие результаты о мощности этого теста при наличии нерелевантных инструментов.]

Проблема получения статистических выводов из оценок остается. Заключение на данный момент таково, что обычный статистический тест, вероятнее всего, ненадежен. Некоторые полезные результаты были получены при разработке процедур тестирования, которые были бы более надежными, нежели стандартные асимптотики первого порядка, которые мы использовали (например, в теореме 8.1 и разделе 10.4). Клейберген (Kleibergen (2002)) разработал класс статистических тестов, основанных на выводах из работ Андерсона и Рубина (Anderson, Rubin (1949, 1950)), предлагающих развитие этого направления. Интересная сторона этого направления исследований заключается в том, что тест Андерсона–Рубина разрабатывался в их работах за 1949 и 1950 гг. и опередил на несколько лет разработку двухшагового метода наименьших квадратов у Тэйла (Theil (1953)) и Басмана (Basman (1957)). [См. (Stock, Trebbi (2003)) обзор ранних работ по методу инструментальных переменных.] Более детальное описание метода Клейбергена (Kleibergen, 2002) и некоторых продолжающих его работ представлено в обзоре Дюфо (Dufour (2003)), который мы используем здесь для беглого знакомства со статистикой Андерсона–Рубина.

Система одновременных уравнений в терминах уравнения 1 записывается так:

$$\begin{aligned} y_1 &= \mathbf{X}_1 \beta_1 + \mathbf{Y}_1 \gamma_1 + \varepsilon_1, \\ \mathbf{Y}_1 &= \mathbf{X}_1 \Pi_1 + \mathbf{X}_1^* \Pi_1^* + \mathbf{V}_1, \end{aligned} \tag{10-64}$$

где y_1 — переменная, состоящая из n наблюдений, в левой части интересующего нас уравнения, \mathbf{Y}_1 — n наблюдений для каждой из M_1 эндогенных

переменных в данном уравнении, γ_1 — структурный вектор параметров в этом уравнении, а X_1 — K_1 экзогенных переменных, включенных в уравнение 1. Второе уравнение — это набор M_1 уравнений приведенной формы для включенных эндогенных переменных, возникающих в уравнении 1. (Обратите внимание, что M_1^* эндогенных переменных Y_1^* исключены из уравнения 1.) Полный набор экзогенных переменных в модели имеет вид

$$X = [X_1, X_1^*],$$

где X_1^* — K_1^* экзогенных переменных, исключенных из уравнения 1. (Мы несколько изменили обозначения из работы Дюфо, чтобы соответствовать допущениям, использованным в нашей модели.) Заметим, что второе уравнение представляет собой первый этап двухшагового метода наименьших квадратов.

Нас интересует проверка гипотез о γ_1 . В первую очередь мы должны предположить, что модель идентифицируема. Как и прежде, мы применяем условия ранга и порядка. Условие порядка состоит в том, что исключенных переменных должно быть, по крайней мере, столько же, сколько включенных эндогенных переменных, т.е. $K_1^* \geq M_1$. Чтобы соблюсти условие ранга, нам необходимо выполнение равенства

$$\pi_1^* - \Pi_1^* \gamma_1 = 0,$$

где π_1^* — вторая часть вектора коэффициентов в уравнении приведенной формы для y_1 , т.е.

$$y_1 = X_1 \pi_1 + X_1^* \pi_1^* + v_1.$$

Чтобы получить этот результат, матрица Π_1^* должна иметь полный ранг по столбцам K_1^* . Проблема слабых инструментов включена в Π_1^* . Если у этой матрицы невысокий ранг, вектор параметров не идентифицирован. Проблема слабых инструментов возникает, когда γ_1 имеет почти низкий ранг. Важным аспектом этого наблюдения является то, что слабые инструменты могут быть охарактеризованы как проблема идентификации.

Андерсон и Рубин (Anderson, Rubin (1949, 1950)) (AR) предложили метод проверки гипотезы $H_0: \gamma_1 = \gamma_1^0$. AR-статистика строится следующим образом: совмещая два уравнения из (10-64), имеем

$$y_1 = X_1 \beta_1 + X_1 \Pi_1 \gamma_1 + X_1^* \Pi_1^* \gamma_1 + \varepsilon_1 + V_1 \gamma_1.$$

Используя (10-64) снова, вычитаем $Y_1 \gamma_1^0$ из обеих частей этого уравнения, чтобы получить

$$\begin{aligned} y_1 - Y_1 \gamma_1^0 &= X_1 \beta_1 + X_1 \Pi_1 \gamma_1 + X_1^* \Pi_1^* \gamma_1 + \varepsilon_1 + V_1 \gamma_1 - \\ &\quad - X_1 \Pi_1 \gamma_1^0 - X_1^* \Pi_1^* \gamma_1^0 - V_1 \gamma_1^0 \\ &= X_1 [\beta_1 + \Pi_1 (\gamma_1 - \gamma_1^0)] + X_1^* [\Pi_1^* (\gamma_1 - \gamma_1^0)] + \varepsilon_1 + V_1 (\gamma_1 - \gamma_1^0) \\ &= X_1 \theta_1 + X_1^* \theta_1^* + w_1. \end{aligned}$$

В условиях нулевой гипотезы это уравнение сокращается до

$$y_1 - Y_1 \gamma_1^0 = X_1 \theta_1 + w_1,$$

поэтому проверку нулевой гипотезы можно осуществить посредством проверки гипотезы о том, что вектор θ_1^* равен нулю в предшествующем неполном уравнении приведенной формы. Андерсон и Рубин предложили простой F-тест

$$AR(\gamma_1^0) = \frac{[(y_1 - Y_1\gamma_1^0)'M_1(y_1 - Y_1\gamma_1^0) - (y_1 - Y_1\gamma_1^0)'M(y_1 - Y_1\gamma_1^0)]/K_1^*}{(y_1 - Y_1\gamma_1^0)'M(y_1 - Y_1\gamma_1^0)/(n - K)}$$

$$\sim F[K_1^*, n - K],$$

где $M_1 = [I - X_1(X_1'X_1)^{-1}X_1']$ и $M = [I - X(X'X)^{-1}X']$. Это стандартная F-статистика для проверки гипотезы о том, что множество коэффициентов равно нулю для классической линейной регрессии. [См. (5-29).] [Дюфо (Dufour (2003)) показал, как можно расширить статистику для получения более общих ограничений, также включающих β_1 .]

У такого подхода есть некоторые удивительные свойства помимо того факта, что он был доступен с 1949 г.: (1) его распределение свободно от параметров модели в конечных выборках (в предположении нормальности шоков); (2) *он устойчив (робастен) относительно проблемы слабых инструментов*; (3) он устойчив по отношению к исключению других инструментов и (4) он устойчив по отношению к ошибкам спецификации в структурных уравнениях для Y_1 , других переменных в уравнении. При этом существуют и недостатки, а именно (1) тесты, основанные на этом методе, применимы только к полному вектору параметров; (2) мощность этого теста может ослабевать, если добавить больше (или намного больше) инструментальных переменных; (3) он опирается на предположение о нормальности шоков и (4) отсутствует его вариант для нелинейных систем уравнений.

10.7. Заключение и выводы

В этой главе мы рассмотрели методы оценки систем уравнений. Системы внешне не связанных регрессий являются приложением обобщенной регрессионной модели, представленной в главе 9. Преимуществом концепции внешне не связанных регрессий является разнообразие поведенческих моделей, вписывающихся в эти рамки. Мы начали главу со знакомства с методами оценки и проверки гипотез для модели внешне не связанных уравнений, анализируя ее, по сути, как обобщенную регрессию. Основное различие между полученными здесь результатами и моделью для одного уравнения из главы 9 – практическая сторона вопроса. Так как система внешне не связанных уравнений является в принципе моделью обобщенной регрессии с одним уравнением и со сложной ковариационной структурой, особые проблемы возникают, когда мы открыто признаем ее как набор уравнений, связанных своими шоками. Главным результатом оценивания на данном этапе является оценка допустимым обобщенным методом наименьших квадратов. Несмотря на его кажущуюся сложность, мы можем оценить модель внешне не связанных регрессий прямо, при помощи двушагового обобщенного метода наименьших квадратов, схожего с тем, который мы использовали для моделей с гетероскедастичностью в главе 9. Мы также расширили

модель внешне не связанных регрессий до случая наличия автокорреляции и гетероскедастичности. Опять же природа множественных уравнений модели усложняет эти примеры. В разделе 10.5 представлено общее приложение систем внешне не связанных регрессий к оценке систем спроса. Одной из важных черт литературы по данной теме является плавный переход от теоретических моделей оптимизации потребителей и производителей к набору эмпирических уравнений спроса, полученных из тождества Роя для потребителей и из леммы Шепарда для производителей.

Системы уравнений, рассмотренные в этой главе, включают большую часть проблем, возникающих в эконометрике при анализе линейных уравнений. Перед тем как начать процедуру оценивания, необходимо убедиться, что данные в выборке действительно содержат достаточно информации для получения требуемых оценок параметров. В этом состоит вопрос идентификации. Идентификация затрагивает и статистические свойства оценок, и роль теории в спецификации модели. Когда идентификация осуществлена, появляются многочисленные методы оценивания. Мы рассматривали некоторые методы оценки уравнений в отдельности, включая методы наименьших квадратов, инструментальных переменных и максимального правдоподобия. Полностью эффективное использование данных в выборке требует совместного оценивания всех уравнений системы. Опять же существует несколько методов — это расширенные варианты методов для одного уравнения, включая трехшаговый метод наименьших квадратов и критерий максимального правдоподобия при полной информации. Обе концепции создают одну из тех благоприятных ситуаций, в которых наиболее простая с вычислительной точки зрения оценка, как правило, оказывается самой эффективной.

Основные термины и понятия

| | |
|---|--|
| Admissible — Допустимый | Dynamic model — Динамическая модель |
| Autocorrelation — Автокорреляция | Econometric model — Эконометрическая модель |
| Balanced panel — Сбалансированные панельные данные | Endogenous — Эндогенный |
| Behavioral equation — Поведенческое уравнение | Equilibrium condition — Условие равновесия |
| Causality — Причинность | Exactly identified model — Точно идентифицированная модель |
| Cobb-Douglas model — Модель Кобба-Дугласа | Exclusion restrictions — Исключающие ограничения |
| Complete system of equations — Полная система уравнений | Exogenous — Экзогенный |
| Completeness condition — Условие полноты | Feasible GLS — Допустимый GLS |
| Consistent estimators — Состоятельные оценки | Flexible functional form — Гибкая функциональная форма |
| Constant returns to scale — Постоянная отдача от масштаба | Flexible functions — Гибкие функции |
| Covariance structures model — Структура ковариационной матрицы | Full information estimator — Оценка при полной информации |
| Demand system — Система спроса | Full information maximum likelihood (FIML) — Критерий максимального правдоподобия при полной информации |

- Generalized regression model** — Обобщенная регрессионная модель
- Granger causality** — Причинность по Грейнджеру
- Heteroscedasticity** — Гетероскедастичность
- Homogeneity restriction** — Ограничение однородности
- Identical explanatory variables** — Тожественные объясняющие переменные
- Identical regressors** — Тожественные регрессоры
- Identification** — Идентификация
- Instrumental variable estimator** — Оценка методом инструментальных переменных
- Interdependent** — Взаимозависимый
- Invariance** — Инвариантность
- Invariant** — Инвариантный
- Jointly dependent** — Совместно зависимый
- k class** — k-класс
- Kronecker product** — Произведение Кронекера
- Lagrange multiplier test** — Критерий множителей Лагранжа
- Least variance ratio** — Отношение наименьших дисперсий
- Likelihood ratio test** — Тест отношения правдоподобий
- Limited information estimator** — Оценка при ограниченной информации
- Limited information maximum likelihood (LIML) estimator** — Оценка методом максимального правдоподобия при ограниченной информации (LIML)
- Maximum likelihood** — Максимальное правдоподобие
- Multivariate regression model** — Многофакторная регрессионная модель
- Nonlinear systems** — Нелинейные системы
- Nonsample information** — Вневыборочная информация
- Nonstructural** — Неструктурный
- Normalization** — Нормировка
- Observationally equivalent** — Эквивалентный с точки зрения наблюдений
- Order condition** — Условие порядка
- Overidentification** — Сверхидентификация
- Pooled model** — Объединенная модель
- Predetermined variable** — Предопределенные переменные
- Problem of identification** — Проблема идентификации
- Projection** — Проекция
- Rank condition** — Условие ранга
- Recursive model** — Рекурсивная модель
- Reduced form** — Приведенная форма
- Reduced form disturbance** — Шок приведенной формы
- Restrictions** — Ограничения
- Seemingly unrelated regressions** — Внешне не связанные регрессии
- Share equations** — Уравнения долей
- Shephard's lemma** — Лемма Шепарда
- Simultaneous equations models** — Смещение в системах одновременных уравнений
- Singularity of the disturbance covariance matrix** — Вырожденность ковариационной матрицы шоков
- Simultaneous equations bias** — Смещение в системах одновременных уравнений
- Specification test** — Тест на спецификацию
- Strongly exogenous** — Строго экзогенный
- Structural disturbance** — Структурный шок
- Structural equation** — Структурное уравнение
- Structural form** — Структурная форма
- System methods of estimation** — Методы оценки систем
- Systems of demand equations** — Системы уравнений спроса
- Taylor series** — Ряды Тейлора
- Three-stage least squares (3SLS) estimator** — Оценка трехшаговым методом наименьших квадратов (3МНК)
- Translog function** — Транслогарифмическая функция
- Triangular system** — Треугольная система
- Two-stage least squares (2SLS) estimator** — Оценка двухшаговым методом наименьших квадратов (2МНК)
- Underidentified** — Неидентифицируемый
- Weak instruments** — Слабые инструменты
- Weakly exogenous** — Слабоэкзогенный

Упражнения

1. Для выборки из 100 наблюдений получены следующие выборочные характеристики:

$$\begin{aligned}\bar{y}_1 &= 1, & \bar{y}_2 &= 2, \\ y_1' y_1 &= 150, \\ y_2' y_2 &= 550, \\ y_1' y_2 &= 260.\end{aligned}$$

Основная двумерная регрессионная модель имеет вид

$$\begin{aligned}y_1 &= \mu + \varepsilon_1, \\ y_2 &= \mu + \varepsilon_2.\end{aligned}$$

- Вычислите МНК-оценку μ и оцените выборочную дисперсию этой оценки.
 - Вычислите FGLS-оценку μ и выборочную дисперсию оценки.
2. Рассмотрите оценки следующей модели из двух уравнений:

$$\begin{aligned}y_1 &= \beta_1 + \varepsilon_1, \\ y_2 &= \beta_2 x + \varepsilon_2.\end{aligned}$$

Выборка из 50 наблюдений дает следующую матрицу моментов:

$$\begin{array}{c} 1 \\ y_1 \\ y_2 \\ x \end{array} \begin{bmatrix} 1 & y_1 & y_2 & x \\ 50 & & & \\ 150 & 500 & & \\ 50 & 40 & 90 & \\ 100 & 60 & 50 & 100 \end{bmatrix} .$$

- Запишите точную формулу для GLS-оценки $[\beta_1, \beta_2]$. Чему равна асимптотическая ковариационная матрица оценки?
 - Найдите МНК-оценку и ее выборочную дисперсию для данной модели.
 - Найдите МНК-оценки для β_1 и β_2 и оцените выборочную ковариационную матрицу двух оценок. Используйте n вместо $(n-1)$ в качестве делителя для вычисления оценок дисперсий шоков.
 - Вычислите FGLS-оценки β_1 и β_2 и оцененную выборочную ковариационную матрицу.
 - Проверьте гипотезу о том, что $\beta_2 = 1$.
3. Модель

$$\begin{aligned}y_1 &= \beta_1 x_1 + \varepsilon_1, \\ y_2 &= \beta_2 x_2 + \varepsilon_2\end{aligned}$$

удовлетворяет всем предположениям модели классической многофакторной регрессии. Все переменные имеют нулевые средние значения. Следующая матрица моментов второго порядка получена для выборки из 20 наблюдений:

$$\begin{array}{c} y_1 \\ y_2 \\ x_1 \\ x_2 \end{array} \begin{bmatrix} 20 & 6 & 4 & 3 \\ 6 & 10 & 3 & 6 \\ 4 & 3 & 5 & 2 \\ 3 & 6 & 2 & 10 \end{bmatrix}.$$

- Вычислите FGLS-оценки для β_1 и β_2 .
 - Проверьте гипотезу о том, что $\beta_1 = \beta_2$.
 - Вычислите оценки параметров модели при помощи метода максимального правдоподобия.
 - Используйте тест отношения правдоподобий для проверки гипотезы из пункта b.
4. Докажите, что в модели

$$\begin{aligned} y_1 &= \mathbf{X}_1 \beta_1 + \varepsilon_1, \\ y_2 &= \mathbf{X}_2 \beta_2 + \varepsilon_2 \end{aligned}$$

результаты оценки обобщенным методом наименьших квадратов равносильны результатам оценки обычным методом наименьших квадратов, примененным к каждому отдельному уравнению, если $\mathbf{X}_1 = \mathbf{X}_2$. Сохранится ли ваш вывод, если будет также известно, что $\beta_1 = \beta_2$?

5. Рассмотрите систему из двух уравнений

$$\begin{aligned} y_1 &= \beta_1 x_1 + \varepsilon_1, \\ y_2 &= \beta_2 x_2 + \beta_3 x_3 + \varepsilon_2. \end{aligned}$$

Предположим, что дисперсии и ковариация шоков известны. Теперь представьте, что исследователь этой модели применяет GLS, но ошибочно опускает x_3 из второго уравнения. Какой эффект эта ошибка спецификации окажет на состоятельность оценки β_1 ?

6. Рассмотрите систему

$$\begin{aligned} y_1 &= \alpha_1 + \beta x + \varepsilon_1, \\ y_2 &= \alpha_2 + \varepsilon_2. \end{aligned}$$

Шоки коррелированы между собой. Докажите, что в результате использования метода GLS для оценки данной системы получатся МНК-оценки α_1 и α_2 , но со смещением оценок наименьших квадратов коэффициентов наклонов в регрессиях y_1 и y_2 на x как оценки β . Что это за смещение? Для простоты подсчета предположим (без потери общности), что $\bar{x} = 0$.

7. Для модели

$$\begin{aligned} y_1 &= \alpha_1 + \beta x + \varepsilon_1, \\ y_2 &= \alpha_2 + \varepsilon_2, \\ y_3 &= \alpha_3 + \varepsilon_3 \end{aligned}$$

предположим, что $y_{i2} + y_{i3} = 1$ для любого наблюдения. Докажите, что выборочная ковариационная матрица остатков наименьших квадратов из этих трех уравнений будет вырожденной, что препятствует вычислению FGLS-оценки. Что можно сделать в этом случае?

8. Рассмотрите следующую систему из двух уравнений:

$$\begin{aligned} y_1 &= \gamma_1 y_2 + \beta_{11} x_1 + \beta_{21} x_2 + \beta_{31} x_3 + \varepsilon_1, \\ y_2 &= \gamma_2 y_1 + \beta_{12} x_1 + \beta_{22} x_2 + \beta_{32} x_3 + \varepsilon_2. \end{aligned}$$

a. Убедитесь, что, как уже указывалось, ни одно из уравнений не идентифицируемо.

b. Выясните, являются ли следующие ограничения достаточными для идентификации (или частичной идентификации) модели:

(1) $\beta_{21} = \beta_{32} = 0$,

(2) $\beta_{12} = \beta_{22} = 0$,

(3) $\gamma_1 = 0$,

(4) $\gamma_1 = \gamma_2$ и $\beta_{32} = 0$,

(5) $\sigma_{12} = 0$ и $\beta_{31} = 0$,

(6) $\gamma_1 = 0$ и $\sigma_{12} = 0$,

(7) $\beta_{21} + \beta_{22} = 1$,

(8) $\sigma_{12} = 0$, $\beta_{21} = \beta_{22} = \beta_{31} = \beta_{32} = 0$,

(9) $\sigma_{12} = 0$, $\beta_{11} = \beta_{21} = \beta_{22} = \beta_{31} = \beta_{32} = 0$.

9. Найдите приведенную форму для модели из упражнения 8 при допущениях, сделанных в пунктах a, b(1) и b(9).

10. Следующая модель имеет вид

$$\begin{aligned} y_1 &= \gamma_1 y_2 + \beta_{11} x_1 + \varepsilon_1, \\ y_2 &= \gamma_2 y_1 + \beta_{22} x_2 + \beta_{32} x_3 + \varepsilon_2. \end{aligned}$$

Все переменные измерены как отклонения от своего среднего. Следующая матрица сумм квадратов и перекрестных произведений получена для выборки из 25 наблюдений:

$$\begin{array}{c} \begin{array}{ccccc} & y_1 & y_2 & x_1 & x_2 & x_3 \\ y_1 & 20 & 6 & 4 & 3 & 5 \\ y_2 & 6 & 10 & 3 & 6 & 7 \\ x_1 & 4 & 3 & 5 & 2 & 3 \\ x_2 & 3 & 6 & 2 & 10 & 8 \\ x_3 & 5 & 7 & 3 & 8 & 15 \end{array} \end{array} .$$

a. Оцените оба уравнения по МНК.

b. Оцените параметры двух уравнений по 2МНК. Также оцените асимптотическую ковариационную матрицу 2МНК-оценок.

c. Найдите LIML-оценки параметров первого уравнения.

d. Оцените эти два уравнения по 3МНК.

e. Оцените матрицу коэффициентов приведенной формы по МНК и косвенно, используя ваши структурные оценки из пункта b.

11. Для модели

$$\begin{aligned} y_1 &= \gamma_1 y_2 + \beta_{11} x_1 + \beta_{21} x_2 + \varepsilon_1, \\ y_2 &= \gamma_2 y_1 + \beta_{32} x_3 + \beta_{42} x_4 + \varepsilon_2 \end{aligned}$$

покажите, что существует два ограничения для коэффициентов приведенной формы. Опишите процедуру оценивания модели по мере включения ограничений.

12. Докажите, что

$$\text{plim} \frac{\mathbf{Y}'_j \varepsilon_j}{T} = \omega_j - \Omega_{jj} \gamma_j.$$

13. Докажите, что неидентифицируемое уравнение не может быть оценено по 2МНК.

Эмпирические приложения

1. Продолжая анализ из подраздела 10.5.2, мы обнаруживаем, что транслогарифмическая функция затрат для случая выпуска одного продукта и трех факторов производства, для которых не выполняется постоянная отдача от масштаба, имеет вид

$$\begin{aligned} \ln C &= \alpha + \beta_1 \ln p_1 + \beta_2 \ln p_2 + \beta_3 \ln p_3 + \delta_{11} \frac{1}{2} \ln^2 p_1 + \delta_{12} \ln p_1 \ln p_2 + \\ &+ \delta_{13} \ln p_1 \ln p_3 + \delta_{22} \frac{1}{2} \ln^2 p_2 + \delta_{23} \ln p_2 \ln p_3 + \delta_{33} \frac{1}{2} \ln^2 p_3 + \\ &+ \gamma_{q1} \ln Q \ln p_1 + \gamma_{q2} \ln Q \ln p_2 + \gamma_{q3} \ln Q \ln p_3 + \\ &+ \beta_q \ln Q + \beta_{qq} \frac{1}{2} \ln^2 Q + \varepsilon_c. \end{aligned}$$

Уравнения для долей факторов имеют вид

$$\begin{aligned} S_1 &= \beta_1 + \delta_{11} \ln p_1 + \delta_{12} \ln p_2 + \delta_{13} \ln p_3 + \gamma_{q1} \ln Q + \varepsilon_1, \\ S_2 &= \beta_2 + \delta_{12} \ln p_1 + \delta_{22} \ln p_2 + \delta_{23} \ln p_3 + \gamma_{q2} \ln Q + \varepsilon_2, \\ S_3 &= \beta_3 + \delta_{13} \ln p_1 + \delta_{23} \ln p_2 + \delta_{33} \ln p_3 + \gamma_{q3} \ln Q + \varepsilon_3. \end{aligned}$$

[См. анализ этой модели в работе Кристенсена и Грина (Christensen, Greene (1976)).]

- Сумма долей трех факторов должна быть равна 1. Какие ограничения это требование накладывает на параметры модели?
- Покажите, что дополняющее условие из (10-38) может быть наложено прямо на модель посредством спецификации транслогарифмической модели в (C/p_3) , (p_1/p_3) и (p_2/p_3) и пропуска третьего уравнения доли. (См. пример 10.3.) Заметьте, что это сокращает число свободных параметров в модели до 10.
- В продолжение пункта b модель, специфицированная с ограничениями симметрии и перекрестного равенства коэффициентов в уравнениях, имеет 15 параметров. Налагая ограничения, вы сокращаете это число до 10 в оценочных уравнениях. Как вы получите оценки параметров, не оцененных напрямую?

Для выполнения остальных пунктов этого упражнения потребуется специализированное программное обеспечение. Можно использовать четыре программы, отмеченные в предисловии: **E-Views**, **TSP**, **Stata** или **LIMDEP**. Все вычисления должны проводиться по данным, использованным в подразделе 10.5.1.

- a. Оцените каждое из трех уравнений, полученных в пункте b, простым методом наименьших квадратов. Удовлетворяют ли эти оценки ограничениям перекрестного равенства коэффициентов в уравнениях и симметрии, которые подразумевает теория?
- b. Используя данные из подраздела 10.5.1, оцените полную систему из трех уравнений (затрат и двух независимых долей), налагая ограничения симметрии и перекрестного равенства коэффициентов в уравнениях.
- c. Используя ваши оценки параметров, вычислите оценки эластичностей в (10-39) при средних значениях переменных.
- d. Используйте статистику отношения правдоподобий для проверки совместной гипотезы о том, что $\gamma_{qi} = 0, i = 1, 2, 3$. [Подсказка: просто исключите нужные переменные из модели.]

2. Данные по инвестициям Грюнфельда из табл. F10.4 приложения — классическое множество данных, применяемое десятилетиями для получения и демонстрации оценок для внешне не связанных регрессий⁵⁰. Несмотря на то что они немного устарели на данном этапе, они остаются идеальным примером применения методов оценки, представленных в этой главе⁵¹. Данные состоят из временных рядов наблюдений за 20 лет по 10 фирмам. Рассматриваются следующие три переменные:

- I_{it} = валовые инвестиции,
 F_{it} = рыночная стоимость фирмы на конец предшествующего года,
 C_{it} = стоимость акций на машины и оборудование на конец предшествующего года.

Главное уравнение в этом исследовании имеет вид $I_{it} = \beta_1 + \beta_2 F_{it} + \beta_3 C_{it} + \varepsilon_{it}$ ⁵².

1. Оцените 10 уравнений отдельно методом наименьших квадратов и покажите ваши результаты.
 - a. Используйте тест Вальда (Чоу) для проверки ограничения «агрегирования» о том, что 10 векторов коэффициентов одинаковы.
 - b. Используйте оценку внешне не связанных регрессий (FGLS), чтобы переоценить параметры модели, опять же допуская отличия коэффициентов по 10 уравнениям. Теперь используйте объединенную мо-

⁵⁰Grunfeld (1958), Grunfeld, Griliches (1960), Boot, de Witt (1960) и Kleiber, Zeileis (2010).

⁵¹Zellner (1962, 1963) и Zellner, Huang (1962).

⁵²Обратите внимание, что модель определяет инвестиции, поток как функцию от двух капиталов. Это могло быть теоретической ошибкой спецификации. И предпочтительнее было бы специфицировать модель в пересчете на плановые инвестиции. Но спустя 50 лет после ее появления будем использовать модель в том виде, как она есть.

- дель и снова FGLS, чтобы оценить уравнение с ограничениями с равными векторами параметров, а также проверьте гипотезу об агрегировании.
- c. Используя МНК-остатки из отдельных регрессий, примените статистику множителей Лагранжа из (10-17) для проверки наличия корреляции между уравнениями.
 - d. Альтернативный вариант спецификации модели в пункте c, сосредоточенный больше на дисперсиях, чем на средних значениях, — это модель групповой гетероскедастичности. Для данного приложения вы можете получить эту модель, используя (10-19), (10-20) и (10-21) и устанавливая более простую модель, где $\sigma_{ij} = 0$ при $i \neq j$. Различаются ли результаты по объединенной модели в рассмотренных трех случаях: при простом МНК, групповой гетероскедастичности и полных неограниченных ковариациях [из чего получим (10-20)] при $\Omega_{ij} = I$?
2. Данные из табл. F5.2 приложения можно использовать для оценки малых макроэкономических моделей. Используйте эти данные для оценки модели из примера 10.4. Оцените параметры этих двух уравнений по двух- или трехшаговому методу наименьших квадратов.

ГЛАВА 11

Модели панельных данных

11.1. Введение

Наборы данных, сочетающие временные ряды и межобъектные данные, широко распространены в экономике. Публикуемая статистика ОЭСР содержит многочисленные ряды сводных экономических показателей, ежегодно наблюдаемых во многих странах. Penn World Tables [CIC (2010)] — база данных по национальному доходу для 188 стран более чем за 50 лет. Составленные недавно **лонгитюдные данные** содержат наблюдения по тысячам человек или семей, каждое из которых относится к определенным моментам. Другие эмпирические исследования включают временные ряды одновременно по фирмам, штатам, странам и отраслям. Такие данные являются ценными источниками информации об экономике. Анализ панельных данных позволяет исследователю, строящему модель, получить информацию об экономических процессах, изучая и неоднородность по агентам, фирмам, странам и т.д., и динамические эффекты, которые невидимы в межобъектных данных. Построение моделей в данном контексте часто требует сложных стохастических спецификаций. В этой главе мы проведем обзор наиболее часто используемых методик для смешанных данных, имеющих одновременно структуру временных рядов и межобъектных данных (например, по странам), и панельных (например, лонгитюдных) данных. Методы, рассматриваемые здесь, можно распространить на большую часть моделей, о которых шла речь в предыдущих главах. В разделе 11.2 описаны специфические черты панельных данных. Значительная часть анализа в этом разделе сосредоточена на индивидуальных данных, нежели на сводных данных по странам. Мы рассмотрим некоторые аспекты построения моделей по сводным данным в разделе 11.11. В разделах 11.3, 11.4 и 11.5 мы по очереди изучим три главных подхода к регрессионному анализу панельных данных, объединенных регрессией, моделями с фиксированными и случайными эффектами. В разделе 11.6 будут рассмотрены методы надежной оценки ковариационных матриц оценок по панельным данным, включая общий анализ «кластерных» эффектов. В разделах 11.7–11.11 изучаются некоторые специфические примеры применения и распространения методов работы с панельными данными. Пространственная автокорреляция обсуждается в разделе 11.7. В разделе 11.8 мы рассмотрим источники эндогенности в моделях со случайными эффектами, включая модель такого вида, как в главе 8, с эндогенной переменной в правой части, и далее два подхода к анализу динамических моделей. В разделе 11.9 модели с фиксированными и случайными эффектами трансформируются в нелинейные регрессионные модели. В разделе 11.10 модель со случайными эффектами расширяется до систем нескольких уравнений, рассмотренных в главе 10. Наконец, в разделе 11.11

рассматриваются модели со случайными параметрами. Подход с использованием случайных параметров является расширением моделей с фиксированными и случайными эффектами, в которых гетероскедастичность учитывается включением в постоянные члены, а также распространяется на другие параметры. Панельные данные используются во всей оставшейся части этого учебника. Мы рассмотрим некоторые расширенные варианты моделей с фиксированными и случайными эффектами в главе 14 для методов максимального правдоподобия и в главе 15, где продолжим тему моделей со случайными параметрами, начатую в подразделе 11.11. В главе 14 будут также представлены методы обработки дискретных распределений случайных параметров под заглавием о моделях латентных классов. В главе 21 мы вернемся к моделям нестационарных панельных данных, предложенным в подразделе 11.8.4. Теория фиксированных и случайных эффектов будет применяться в примерах моделей дискретных и ограниченных зависимых переменных в микроэконометрике в главах 17, 18 и 19.

11.2. Модели панельных данных

Во многих современных исследованиях используются **панельные**, или лонгитюдные, данные. Два наиболее известных массива таких данных — Национальное лонгитюдное обследование рынка труда (National Longitudinal Survey of Labor Market Experience, <http://www.bls.gov/nls/nlsdoc.htm>) и Мичиганское панельное обследование динамики дохода (Michigan Panel Study of Income Dynamics, PSID, <http://psidonline.isr.umich.edu/>). Здесь очень многие пространственные данные, состоящие из тысяч микроэлементов, отслеживаются во времени, хотя число периодов зачастую невелико. Так, PSID — это обследование примерно 6000 семей и 15 000 человек, опрашиваемых периодически с 1968 г. до настоящего времени. Осуществляемое в настоящий момент обследование в Соединенном Королевстве — Панельное обследование британских домохозяйств (British Household Panel Survey, <http://www.iser.essex.ac.uk/ulsc/bhps/>) — началось в 1991 г. и сейчас находится на 18-й волне. Оно охватывает несколько тысяч домохозяйств (сейчас более 5000) за несколько лет. Много ценных панельных данных недавно было получено в области здравоохранения и его экономики, включая Социоэкономическую панель Германии (German Socioeconomic Panel, http://dpls.dacc.wisc.edu/apdu/GSOEP/gsoep_cd_data.html) и Панельное обследование расходов на медицину (Medical Expenditure Panel Survey, <http://www.meps.ahrq.gov/>). Построение длинных, равномерно распределенных временных рядов в таких ситуациях было бы слишком дорого, однако для тех целей, для которых такие данные обычно используются, это необязательно. Временные эффекты часто рассматриваются в качестве «переходов» или дискретных изменений состояния. Текущее обследование населения (Current Population Survey, <http://www.census.gov/cps/>), к примеру, представляет собой ежемесячный обзор 50 000 домохозяйств, в ходе которого домохозяйства опрашиваются в течение четырех месяцев, затем наступает перерыв, после чего опрос осуществляется снова. Такого рода двухволновый формат **обновляемой панели** позволяет анализировать краткосрочные изменения наряду

с получением более общего обзора национального рынка труда в США. Как правило, данные моделируются как характерные для того периода, в котором они возникают, и не распространяются на периоды внутри объекта¹. Панельные данные больше ориентированы на пространственные виды анализа; они широкие, но, как правило, не длинные. Неоднородность объектов является неотъемлемой их частью, а на самом деле центральным аспектом анализа.

Анализ панельных, или лонгитюдных, данных — предмет одной из наиболее динамичных и инновационных областей исследований в эконометрике² отчасти из-за того, что панельные данные дают обширную почву для развития технических аспектов оценивания и теоретических результатов. С практической стороны, однако, исследователи способны использовать пространственные данные временных рядов для изучения проблем, которые нельзя было рассматривать отдельно при пространственных данных или временных рядах. Приведем два примера.

1. В широко цитируемом исследовании предложения труда Бен-Пораз (Ben-Porath (1973)) наблюдал, что в определенный момент в некоторой группе женщин 50% могут оказаться трудящимися. Неясно, предполагает ли это открытие, что в данной группе половина женщин будет работать в среднем или что та же половина будет работать в каждом периоде. Это может иметь очень разные последствия для политики и интерпретации любых статистических результатов. Взятые отдельно межобъектные данные не прольют свет на этот вопрос.
2. Давней проблемой в анализе производственных функций была неспособность разграничить эффект масштаба и технологические изменения³. Межобъектные данные дают информацию только о первом, в то время как временные ряды спутывают эти два эффекта без какой-либо возможности разделить их. Например, часто предполагается постоянная отдача от масштаба, с тем чтобы выявить технические изменения⁴. Конечно,

¹Формальная процедура моделирования временных рядов для панельных данных кратко рассматривается в разделе 21.5.

²Работы по панельным данным конкурируют с общепринятым исследованием единичных корней и коинтеграции в эконометрике по темпам роста. Обзор ранних работ представлен у Maddala (1993). Обзоры по эконометрике панельных данных есть у Hsiao (2003), Dielman (1989), Matyas, Sevestre (1996), Raj, Baltagi (1992), Nerlove (2002), Arellano (2003) и Baltagi (2001, 2008). Существуют и масштабные обзоры, посвященные определенным темам, например моделям с ограниченными зависимыми переменными [Hsiao, Lahiri, Lee, и Pesaran (1999)] или полупараметрическим методам [Lee (1998)]. Обширная библиография приведена у Baltagi (2008).

³Различие между этими двумя эффектами фигурировало в политическом вопросе о том, нужно ли было расформировать корпорацию AT&T в 1980-х гг. и в итоге допустить конкуренцию в предоставлении междугородней телефонной связи.

⁴В классическом исследовании этой проблемы Солоу (Solow (1957)) утверждает: «Из временных рядов для $\Delta Q/Q$, w_K , $\Delta K/K$, w_L и $\Delta L/L$ или их дискретных аналогов по отношению к предыдущему году мы могли бы оценить $\Delta A/A$ и из этого сам $A(t)$. На самом деле здесь происходит нечто удивительное. До сих пор ничего не было сказано об отдаче от масштаба. Но если все факторные издержки классифицируются как K или L , тогда имеющиеся данные всегда показывают w_K и w_L , в сумме равные единице. Поскольку мы допускаем, что факторы оплачиваются их предельными продуктами, это ведет к принятию гипотезы теоремы Эйлера. Оставив исчисления такими, какие они есть, мы могли бы лишь также допустить вывод, что F является гомогенной первой степени».

такая процедура не принимает во внимание проблему. Панельные данные по затратам или выпуску для определенного числа фирм, каждая из которых наблюдаема в течение нескольких лет, позволяют получить оценки и уровня технологических изменений (с течением времени), и эффекта от масштаба (для выборки из фирм разных размеров в каждый момент).

Современные разработки позволили исследователям изучать влияние изменений политики здравоохранения [к примеру, анализ реформ всеобщего медицинского страхования в Германии у Riphahn et al.'s (2003)] и более общую динамику поведения рынка труда. В принципе методы из глав 6 и 21 могут быть применены к лонгитюдным данным. В классических панельных данных, однако, существует большое число пространственных элементов и всего несколько периодов. Таким образом, методы временных рядов, рассмотренные здесь, могут быть несколько проблематичными. Последние работы в целом сосредоточены на моделях, более подходящих для этих коротких и широких данных. Данные методы сосредоточены на межобъектной изменчивости или неоднородности. В этой главе мы подробно изучим наиболее широко используемые модели и кратко рассмотрим некоторые расширенные варианты.

11.2.1. Общие подходы к моделированию панельных данных

Фундаментальным преимуществом панельных данных по сравнению с межобъектными является то, что они дают исследователю значительную гибкость в моделировании различий в поведении агентов. Базовой конструкцией здесь является регрессионная модель следующего вида:

$$\begin{aligned} y_{it} &= \mathbf{x}'_{it}\beta + \mathbf{z}'_i\alpha + \varepsilon_{it} \\ &= \mathbf{x}'_{it}\beta + c_i + \varepsilon_{it}. \end{aligned} \quad (11-1)$$

В \mathbf{x}_{it} — это K регрессоров без учета постоянного члена. **Неоднородность (гетерогенность)** или **индивидуальный эффект** — это $\mathbf{z}'_i\alpha$, где \mathbf{z}_i содержит постоянный член и набор индивидуальных или специфических групповых переменных, которые могут быть наблюдаемы, как раса, пол, местоположение и т.д., или ненаблюдаемы, как специфические характеристики семьи, индивидуальная неоднородность в навыках или предпочтениях и т.д., которые считаются постоянными в период t . Можно показать, что это модель классической регрессии. Если \mathbf{z}_i рассматривается для всех людей, тогда всю модель можно анализировать как обычную линейную модель и использовать метод наименьших квадратов. Осложнения возникают, когда c_i ненаблюдаемы, что будет случаться в большинстве примеров. Рассмотрите, например, исследования влияния образования и опыта работы на зарплату, в которых «способность» всегда будет отсутствующей и ненаблюдаемой переменной. При изучении здравоохранения, например использования систем здравоохранения, «здоровье» и «здоровоохранение» будут ненаблюдаемыми факторами в анализе.

Главной целью анализа будет получение постоянной и эффективной оценки **частных эффектов**

$$\beta = \partial E[y_{it} | \mathbf{x}_{it}] / \partial \mathbf{x}_{it}.$$

Возможно ли это, зависит от допущений о ненаблюдаемых эффектах. Мы начнем с предположения о **строгой экзогенности** для независимых переменных

$$E[\varepsilon_{it} | \mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots] = 0,$$

т.е. действующий шок не коррелирован с независимыми переменными в каждом периоде, прошлом, настоящем и будущем. Ключевой аспект модели касается неоднородности. Особенно удобным допущением была бы **независимость средних значений**

$$E[c_i | \mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots] = \alpha.$$

Если отсутствующая переменная не коррелирована с включенными переменными, то, как мы увидим, они могут быть включены в шок модели. Это предположение лежит в основе модели со случайными эффектами, как мы обнаружим далее. Однако это частично строгое предположение — оно маловероятно в примерах о рынке труда и здравоохранении, упомянутых ранее. Альтернативой была бы модель

$$\begin{aligned} E[c_i | \mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots] &= h(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots) \\ &= h(\mathbf{X}_i). \end{aligned}$$

Такая формулировка более общая, но в то же время гораздо более сложная, тем более что она может потребовать дальнейших допущений о природе этой функции.

11.2.2. Структуры модели

Мы изучим ряд различных моделей для панельных данных. В широком смысле они могут быть упорядочены следующим образом.

1. **Объединенная регрессия:** если \mathbf{z}_i содержит только постоянный член, то метод наименьших квадратов позволяет получить постоянные и эффективные оценки обычного α и вектора наклона β .
2. **Модель с фиксированными эффектами:** если \mathbf{z}_i ненаблюдаем, но коррелирован с \mathbf{x}_{it} , то МНК-оценка β смещена и несостоятельна как следствие пропуска переменной. Однако в этом случае модель

$$y_{it} = \mathbf{x}'_{it}\beta + \alpha_i + \varepsilon_{it},$$

где $\alpha_i = \mathbf{z}'_i\alpha$, включает в себе все наблюдаемые эффекты и определяет оцениваемое условное среднее значение. В подходе с **фиксированными эффектами** α_i принимается за специфический для группы постоянный член в регрессионной модели. Следует отметить, что использованный здесь термин «фиксированные» означает наличие корреляции между c_i и \mathbf{x}_{it} , а не то, что c_i нестохастическая.

3. **Модель со случайными эффектами:** если можно предположить, что ненаблюдаемая индивидуальная неоднородность, все же формулируемая, не коррелирована с включенными переменными, тогда модель может быть записана как

$$\begin{aligned} y_{it} &= \mathbf{x}'_{it}\beta + E[\mathbf{z}'_i\alpha] + \{\mathbf{z}'_i\alpha - E[\mathbf{z}'_i\alpha]\} + \varepsilon_{it} \\ &= \mathbf{x}'_{it}\beta + \alpha + u_i + \varepsilon_{it}, \end{aligned}$$

т.е. как модель линейной регрессии с составным шоком, который может быть состоятельно, хоть и неэффективно, оценен методом наименьших квадратов. Такой подход со **случайными эффектами** определяет, что u_i является характерным для группы случайным элементом, схожим с ε_{it} , но для каждой группы существует единственный исход, возникающий в регрессии одинаково в каждый период. Опять же ключевое различие между фиксированными и случайными эффектами заключается в том, отражает ли ненаблюдаемый индивидуальный эффект элементы, коррелирующие с регрессорами в модели, а не в том, стохастические эти эффекты или нет. Мы рассмотрим эту основную формулировку, а затем изучим расширение на случай динамической модели.

4. Модель со случайными параметрами: модель со случайными эффектами может рассматриваться как регрессионная модель со случайным постоянным членом. При достаточно обширных данных мы можем распространить эту идею на модель, в которой другие коэффициенты случайным образом также различаются по индивидуальным наблюдениям. Расширенная модель могла быть представлена как

$$y_{it} = \mathbf{x}'_{it}(\boldsymbol{\beta} + \mathbf{h}_i) + (\alpha + u_i) + \varepsilon_{it},$$

где \mathbf{h}_i — случайный вектор, характеризующий изменчивость параметров по индивидуальным наблюдениям. Такая модель со случайными параметрами была предложена в литературе достаточно рано, но получила широкое применение в нескольких областях весьма недавно. Она представляет собой естественное расширение модели, в котором исследователи расширяют неоднородность среди индивидуальных наблюдений, при этом оставляя некоторые общности — векторы параметров также имеют общее среднее. Некоторые современные примеры расширили ее еще на один шаг, позволив среднему значению распределения параметров быть персонифицированным, как в

$$y_{it} = \mathbf{x}'_{it}(\boldsymbol{\beta} + \Delta \mathbf{z}_i + \mathbf{h}_i) + (\alpha + u_i) + \varepsilon_{it},$$

где \mathbf{z}_i — множество наблюдаемых персонифицированных переменных и Δ — матрица параметров, которые необходимо оценить. Как мы увидим в главе 17, эта **иерархическая модель** чрезвычайно разносторонняя.

11.2.3. Расширения

Краткий список типов моделей, представленный выше, лишь предваряет разнообразие приложений методов панельных данных в эконометрике. В этой главе мы начнем изучение некоторых формулировок и примеров использования линейных моделей. Модели с фиксированными и случайными эффектами и модели со случайными параметрами также широко используются в моделях с цензурированными переменными, в моделях бинарного и другого дискретного выбора и моделях со счетными данными. Мы рассмотрим их в последующих главах. В одних случаях, как в моделях со счетными данными из главы 18, расширение моделей с фиксированными и случайными эффектами применяется непосредственно, хоть и влечет несколько более сложные вычисления. В других же случаях, как в моделях бинарного выбора из главы 17 и в моделях с цензурированными данными из главы 19,

модели панельных данных применяются, но не раньше преодоления некоторых значительных методологических и вычислительных трудностей.

11.2.4. Сбалансированные и несбалансированные панели

В качестве предисловия к дальнейшему анализу отметим важный аспект анализа панельных данных. Как было указано в предшествующем обсуждении, панельные данные будут состоять из n наборов индивидуальных наблюдений, обозначаемых как $i = 1, \dots, n$. Если каждый агент в данных наблюдается одинаковое число раз, обычно обозначаемых как T , то данные являются **сбалансированной панелью**. Данные **несбалансированной панели** — это те, в которых агенты могут наблюдаться различное число раз. Мы обозначим это через T_i . **Фиксированная панель** та, в которой одно и то же множество агентов наблюдается на протяжении всего исследования. Данные, которые мы рассмотрим в этой главе, являются фиксированными, но не всегда сбалансированными. Обновляемая панель та, в которой состав агентов изменяется от одного периода к другому. К примеру, Гонзалес и Мэлони (Gonzalez, Maloney (1999)) рассматривали решения о самозанятости в Мексике, используя Национальное обследование занятости в городах (National Urban Employment Survey). Это ежеквартальные данные за 1987–1993 гг., когда агентов опрашивали пять раз. Каждый квартал пятая часть агентов в данных обновляется. Мы не будем анализировать обновляемые панели здесь. Некоторые обсуждения и множество ссылок можно найти у Балтаги (Baltagi (2008)).

Последующее развитие темы построено таким образом, что различие между сбалансированными и несбалансированными панелями будет вызывать не более чем незначительное изменение в названии, где для удобства мы пишем T , предполагая сбалансированную панель, только изменение T на T_i обобщает выводы. Мы будем отмечать, когда этого не происходит, как, например, в статистике множителей Лагранжа у Бройша и Пэгана (Breusch, Pagan (1980)).

11.2.5. Доброкачественные панельные данные

Асимптотические свойства оценок в классической регрессионной модели были сформулированы в разделе 4.4 при следующих допущениях:

А.1. Линейность: $y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \dots + x_{iK}\beta_K + \varepsilon_i$.

А.2. Полный ранг: матрица данных $n \times K$, \mathbf{X} имеет полный ранг по столбцам.

А.3. Экзогенность независимых переменных: $E[\varepsilon_i | x_{j1}, x_{j2}, \dots, x_{jK}] = 0$, $i, j = 1, \dots, n$.

А.4. Гомоскедастичность и отсутствие автокорреляции.

А.5. Механизм порождения данных дает независимые наблюдения.

Далее следуют ключевые необходимые нам выводы: для состоятельности \mathbf{b} необходимо, чтобы

$$\begin{aligned} \text{plim}(1/n)\mathbf{X}'\mathbf{X} &= \text{plim} \bar{\mathbf{Q}}_n = \mathbf{Q}, & \text{положительно определенная матрица,} \\ \text{plim}(1/n)\mathbf{X}'\boldsymbol{\varepsilon} &= \text{plim} \bar{\mathbf{w}}_n = E[\bar{\mathbf{w}}_n] = \mathbf{0}. \end{aligned}$$

(Для состоятельности s^2 мы добавили довольно слабое допущение о моментах шоков.) Чтобы установить асимптотическую нормальность, нам потребовались состоятельность и

$$\sqrt{n} \bar{w}_n \xrightarrow{d} N[0, \sigma^2 \mathbf{Q}].$$

При утверждении этого необходимые характеристики далее выводятся при помощи методов из подразделов 4.4.1 и 4.4.2.

Исключения в допущениях, вероятнее всего, возникнут в **панельных данных**. Выборка будет включать множественные наблюдения по каждому из многочисленных наблюдаемых элементов. Например, обследование может состоять из множества наблюдений, сделанных в различные моменты в большом количестве семей. В этом случае x будут обязательно коррелированы по наблюдениям по крайней мере внутри наблюдаемых объектов. Они могут даже быть одинаковыми для всех наблюдений одной семьи.

Панельные данные могут анализироваться следующим образом. Сейчас допустим, что данные состоят из фиксированного числа наблюдений, скажем T , на некоторое количество n семей, так что общее число строк в X равно $N = nT$. Матрица

$$\bar{Q}_n = \frac{1}{n} \sum_{i=1}^n Q_i,$$

в которой N — все наблюдения в выборке, может рассматриваться как

$$\bar{Q}_n = \frac{1}{n} \sum_i \frac{1}{T} \sum_{\substack{\text{наблюдения} \\ \text{для семьи } i}} Q_{it} = \frac{1}{n} \sum_{i=1}^n \bar{Q}_i,$$

где \bar{Q}_i — среднее Q_{it} для семьи i . Далее мы могли бы рассмотреть множество наблюдений по i -му объекту, как если бы они были одним наблюдением, и применить наши доводы о сходимости к количеству семей, увеличивающемуся без ограничений. Дело в том, что условия, необходимые для установления сходимости, будут применяться по отношению к числу наблюдаемых элементов. Число наблюдений, принятое для каждого наблюдаемого элемента, может быть фиксированным и достаточно небольшим.

Эта глава содержит относительно небольшое расширение свойств оценок, как это было сделано в главе 4. Мы будем опираться на выводы из глав 4, 8 и 9 и сосредоточимся на разнообразии моделей и спецификаций.

11.3. Объединенная регрессионная модель

Начнем анализ с самой простой версии — **объединенной модели**:

$$\begin{aligned} y_{it} &= \alpha + \mathbf{x}'_{it} \boldsymbol{\beta} + \varepsilon_{it}, \quad i = 1, \dots, n, t = 1, \dots, T_i, \\ E[\varepsilon_{it} \mid \mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iT_i}] &= 0, \\ \text{Var}[\varepsilon_{it} \mid \mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iT_i}] &= \sigma_\varepsilon^2, \\ \text{Cov}[\varepsilon_{it}, \varepsilon_{js} \mid \mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iT_i}] &= 0, \text{ если } i \neq j \\ &\text{или } t \neq s. \end{aligned} \tag{11-2}$$

(В контексте панельных данных эта модель также называется **усредненной по генеральной совокупности моделью** при допущении, что любая скрытая неоднородность была исключена усреднением.) В этом виде, если оставшиеся предположения классической модели соблюдены (нулевое условное среднее значение ε_{it} , гомоскедастичность, независимость по наблюдениям i и строгая экзогенность \mathbf{x}_{it}), дальнейший анализ, выходящий за пределы выводов из главы 4, не нужен. Простой метод наименьших квадратов дает эффективную оценку, и статистические выводы надежны в соответствии с подходом, развитым в главе 5.

11.3.1. Оценка объединенной модели методом наименьших квадратов

Главная проблема при анализе панельных данных в этой главе состоит в том, что допущения, предполагающие оценку объединенной модели при помощи обычного метода наименьших квадратов, вряд ли могут быть соблюдены. Из этого следует вопрос: что можно ожидать от оценки, когда неоднородность различается по объектам? Случай с фиксированными эффектами понятен. Как мы увидим позже, пропуск (или игнорирование) неоднородности, когда модель с фиксированными эффектами уместна, превращает оценку методом наименьших квадратов в несостоятельную — иногда даже так. В примере со случайными эффектами, когда правильная модель имеет вид

$$y_{it} = c_i + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}$$

при $E[c_i | \mathbf{X}_i] = \alpha$, мы можем записать модель как

$$\begin{aligned} y_{it} &= \alpha + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it} + (c_i - E[c_i | \mathbf{X}_i]) \\ &= \alpha + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it} + u_i \\ &= \alpha + \mathbf{x}'_{it}\boldsymbol{\beta} + w_{it}. \end{aligned}$$

Из этой записи видно, что ненаблюдаемая неоднородность вызывает **автокорреляцию**; $E[w_{it}w_{is}] = \sigma_u^2$ при $t \neq s$. Как мы обнаружили в главе 9, — мы вернемся к этому в главе 20 — оценка методом наименьших квадратов в обобщенной модели регрессии может быть состоятельной, однако стандартная оценка асимптотической дисперсии, скорее всего, будет недооценивать истинную дисперсию оценки.

11.3.2. Устойчивая оценка ковариационной матрицы

Предположим, что мы рассматриваем более общую модель, нежели здесь. Соберем все T_i наблюдений по объекту i в одно уравнение

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{w}_i,$$

где $\boldsymbol{\beta}$ теперь включает постоянный член. В таких обстоятельствах может появиться гетероскедастичность по наблюдениям. Однако в панельных данных более значительной проблемой является корреляция между наблюдениями, или автокорреляция. В лонгитюдных данных вся группа наблюдений может относиться к одному объекту и любые скрытые эффекты, остав-

ленные вне модели, будут сохраняться во всех периодах. Теперь предположим, что вектор шоков состоит из ε_{it} плюс эти пропущенные компоненты. Тогда

$$\begin{aligned}\text{Var}[\mathbf{w}_i | \mathbf{X}_i] &= \sigma_\varepsilon^2 \mathbf{I}_{T_i} + \Sigma_i \\ &= \Omega_i.\end{aligned}$$

Оценка β обычным методом наименьших квадратов имеет вид

$$\begin{aligned}\mathbf{b} &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} \\ &= \left[\sum_{i=1}^n \mathbf{X}'_i \mathbf{X}_i \right]^{-1} \sum_{i=1}^n \mathbf{X}'_i \mathbf{y}_i \\ &= \left[\sum_{i=1}^n \mathbf{X}'_i \mathbf{X}_i \right]^{-1} \sum_{i=1}^n \mathbf{X}'_i (\mathbf{X}_i \beta + \mathbf{w}_i) \\ &= \beta + \left[\sum_{i=1}^n \mathbf{X}'_i \mathbf{X}_i \right]^{-1} \sum_{i=1}^n \mathbf{X}'_i \mathbf{w}_i.\end{aligned}$$

Можно вывести состоятельность так, как это было сделано в главе 4. Истинная асимптотическая ковариационная матрица примет форму, знакомую нам по обобщенной регрессионной модели (9-9):

$$\begin{aligned}\text{Asy. Var}[\mathbf{b}] &= \frac{1}{n} \left\{ \text{plim} \left[\frac{1}{n} \sum_{i=1}^n \mathbf{X}'_i \mathbf{X}_i \right]^{-1} \text{plim} \left[\frac{1}{n} \sum_{i=1}^n \mathbf{X}'_i \mathbf{w}_i \mathbf{w}'_i \mathbf{X}_i \right] \text{plim} \left[\frac{1}{n} \sum_{i=1}^n \mathbf{X}'_i \mathbf{X}_i \right]^{-1} \right. \\ &= \left. \frac{1}{n} \left\{ \text{plim} \left[\frac{1}{n} \sum_{i=1}^n \mathbf{X}'_i \mathbf{X}_i \right]^{-1} \text{plim} \left[\frac{1}{n} \sum_{i=1}^n \mathbf{X}'_i \Omega_i \mathbf{X}_i \right] \text{plim} \left[\frac{1}{n} \sum_{i=1}^n \mathbf{X}'_i \mathbf{X}_i \right]^{-1} \right. \right.\end{aligned}$$

Такой результат дает аналог для (9-18). Как и раньше, должна быть оценена центральная матрица. Аналогично оценке Уайта мы можем оценить матрицу с помощью

$$\text{Est. Asy. Var}[\mathbf{b}] = \frac{1}{n} \left[\frac{1}{n} \sum_{i=1}^n \mathbf{X}'_i \mathbf{X}_i \right]^{-1} \left[\frac{1}{n} \sum_{i=1}^n \mathbf{X}'_i \hat{\mathbf{w}}_i \hat{\mathbf{w}}'_i \mathbf{X}_i \right] \left[\frac{1}{n} \sum_{i=1}^n \mathbf{X}'_i \mathbf{X}_i \right]^{-1}, \quad (11-3)$$

где $\hat{\mathbf{w}}'_i$ — вектор T_i остатков для агента i . Логика оценки Уайта действительно сохраняется и для этой оценки. Однако обратите внимание, что это не совсем одно и то же с (9-27). Вполне вероятно, что более важной проблемой для подходящей оценки является корреляция между наблюдениями, а не гетероскедастичность. Таким образом, вполне вероятно, что оценка Уайта в (9-27) не будет решением проблемы получения статистических выводов здесь. Пример 11.1 показывает этот эффект на практике.

Пример 11.1. Уравнение заработной платы

Корнвелл и Руперт (Cornwell, Rupert (1988)) проанализировали отдачу от обучения в сбалансированной панели из 595 наблюдений глав домохозяйств. Выборочные данные получены за 1976–1982 гг. из «Неопросного

обследования экономических возможностей» — панельного обследования динамики доходов («Non-Survey of Economic Opportunity», Panel Study of Income Dynamics). Оцениваемое уравнение имеет вид

$$\begin{aligned} \ln Wage_{it} = & \beta_1 + \beta_2 Exp_{it} + \beta_3 Exp_{it}^2 + \beta_4 Wks_{it} + \beta_5 Occ_{it} \\ & + \beta_6 Ind_{it} + \beta_7 South_{it} + \beta_8 SMSA_{it} + \beta_9 MS_{it} \\ & + \beta_{10} Union_{it} + \beta_{11} Ed_i + \beta_{12} Fem_i + \beta_{13} Blk_i + \varepsilon_{it}, \end{aligned}$$

где используются следующие переменные:

Exp = число лет работы на полную ставку,

Wks = отработанные недели,

Occ = 1, если индивид «синий воротничок», 0, если иначе,

Ind = 1, если индивид работает в обрабатывающей промышленности, 0, если иначе,

South = 1, если индивид проживает на юге, 0, если иначе,

SMSA = 1, если индивид проживает в SMSA, 0, если иначе,

MS = 1, если индивид состоит в браке, 0, если иначе,

Union = 1, если заработная плата индивида устанавливается по соглашению с профсоюзом, 0, если иначе,

Ed = годы обучения,

Fem = 1, если индивид женского пола, 0, если иначе,

Blk = 1, если индивид темнокожий, 0, если иначе.

Обратите внимание, что *Ed*, *Fem* и *Blk* являются **не изменяющимися во времени**. См. табл. 8.1 приложения со ссылками на источники данных. Помимо сравнения различных методов оценки нас интересует коэффициент β_{11} — отдача от обучения. В табл. 11.1 представлены оценки методом наименьших квадратов, основанные на полной выборке из 4165 наблюдений. [Авторы не представили МНК-оценки. Однако они представили линейные оценки методом наименьших квадратов модели с фиксированными эффектами, которые являются оценками обычным методом наименьших квадратов с использованием отклонений от индивидуальных средних. (См. раздел 11.4.) Не было возможности сопоставить представленные авторами результаты с этими или любыми другими результатами. Так как наша задача — сравнить различные оценки друг с другом, мы не пробовали разрешить это несоответствие.] Стандартные ошибки МНК-оценок приведены во втором столбце таблицы. В третьем столбце даны устойчивые стандартные ошибки, вычисленные с помощью (11-3). Для этих данных вычисления имеют вид

$$\text{Est.Asy.Var}[\mathbf{b}] = \left[\sum_{i=1}^{595} \mathbf{X}'_i \mathbf{X}_i \right]^{-1} \left[\sum_{i=1}^{595} \left(\sum_{t=1}^7 \mathbf{x}_{it} e_{it} \right) \left(\sum_{t=1}^7 \mathbf{x}_{it} e_{it} \right)' \right] \left[\sum_{i=1}^{595} \mathbf{X}'_i \mathbf{X}_i \right]^{-1}.$$

Устойчивые стандартные ошибки, как правило, вдвое больше нескорректированных. В отличие от этого устойчивые стандартные ошибки по Уайту почти такие же, как нескорректированные. Это показывает, что для данной модели игнорирование внутригрупповых корреляций на самом деле значительно влияет на получаемые выводы.

Таблица 11.1. Уравнение заработной платы, оцененное МНК

| Коэффициент | Оцененный коэффициент | Стандартная ошибка МНК | Устойчивая панельная стандартная ошибка | Устойчивая к гетероскедастичности состоятельная стандартная ошибка Уайта |
|------------------------------|-----------------------|------------------------|---|--|
| β_1 : Constant | 5,2511 | 0,07129 | 0,1233 | 0,07435 |
| β_2 : Exp | 0,04010 | 0,002159 | 0,004067 | 0,002158 |
| β_3 : Exp ² | -0,0006734 | 0,00004744 | 0,00009111 | 0,00004789 |
| β_4 : Wks | 0,004216 | 0,001081 | 0,001538 | 0,001143 |
| β_5 : Occ | -0,1400 | 0,01466 | 0,02718 | 0,01494 |
| β_6 : Ind | 0,04679 | 0,01179 | 0,02361 | 0,01199 |
| β_7 : South | -0,05564 | 0,01253 | 0,02610 | 0,01274 |
| β_8 : SMSA | 0,1517 | 0,01207 | 0,02405 | 0,01208 |
| β_9 : MS | 0,04845 | 0,02057 | 0,04085 | 0,02049 |
| β_{10} : Union | 0,09263 | 0,01280 | 0,02362 | 0,01233 |
| β_{11} : Ed | 0,05670 | 0,002613 | 0,005552 | 0,002726 |
| β_{12} : Fem | -0,3678 | 0,02510 | 0,04547 | 0,02310 |
| β_{13} : Blk | -0,1669 | 0,02204 | 0,04423 | 0,02075 |

11.3.3. Кластеризация и стратификация

Во многих современных обследованиях, например в текущем обследовании населения (Current Population Survey), анализируются данные опросов. Их иногда объединяют в кластеры частично ради сокращения затрат. К примеру, интервьюеры могут посетить все семьи в определенном жилищном массиве. В других случаях эффекты, напоминающие обычные случайные эффекты в исследованиях панельных данных, могут естественно возникать в выборочной совокупности. Рассмотрим, к примеру, исследование результатов экзаменов школьников в нескольких штатах. Обычные эффекты могут возникнуть в таких данных на разных уровнях. Образовательная программа или политика финансирования в штате могут стать причиной «эффекта штата»; могут возникать эффекты школьного округа, эффекты школы внутри округа и даже эффекты учителей внутри конкретной школы. Каждый из них с большой вероятностью может вызвать корреляцию между наблюдениями, похожую на случайные (или фиксированные) эффекты, которые мы определили. Можно неохотно допустить, что имеет место узкоструктурированная модель, как то спецификация простых случайных эффектов. Однако, как мы видели в примере 11.1, игнорирование обычных эффектов может привести к серьезным ошибкам при получении статистических выводов. Устойчивая оценка, предложенная в подразделе 11.3.2, дает полезный метод.

Для двухуровневой модели, которая может возникнуть для выборки фирм, сгруппированных по отрасли, или школьников, обучающихся у одних преподавателей в определенной школе, естественным методом такой

«кластеризации» был бы подход с устойчивыми общими эффектами, показанный ранее. Сходство стандартной **кластерной оценки** для одноуровневой модели с панельной моделью с общими эффектами, рассмотренной ранее, более чем случайно. Однако существует различие механизма обобщения данных в том, что при этих условиях объекты в группе наблюдаются, как правило, один раз, а их объединение, что и является обычным эффектом, вероятно, будет менее четко определено, нежели в панелях, как мы поняли в примере 11.1. Усовершенствование (11-3) часто применяется для объяснения эффектов в малых выборках, когда число кластеров с высокой вероятностью является значительной долей конечного общего, как то число школьных районов в штате. Корректировка на степени свободы, как показано в (11-4), часто используется для этой цели. Устойчивая оценка ковариационной матрицы будет иметь вид

$$\begin{aligned} \text{Est.Asy.Var}[\mathbf{b}] &= \\ &= \left[\sum_{g=1}^G \mathbf{X}'_g \mathbf{X}_g \right]^{-1} \left[\frac{G}{G-1} \sum_{g=1}^G \left(\sum_{i=1}^{n_g} \mathbf{x}_{ig} \hat{\mathbf{w}}_{ig} \right) \left(\sum_{i=1}^{n_g} \mathbf{x}_{ig} \hat{\mathbf{w}}_{ig} \right)' \right] \left[\sum_{g=1}^G \mathbf{X}'_g \mathbf{X}_g \right]^{-1} = \\ &= \left[\sum_{g=1}^G \mathbf{X}'_g \mathbf{X}_g \right]^{-1} \left[\frac{G}{G-1} \sum_{g=1}^G (\mathbf{X}'_g \hat{\mathbf{w}}_g) (\hat{\mathbf{w}}_g' \mathbf{X}_g) \right] \left[\sum_{g=1}^G \mathbf{X}'_g \mathbf{X}_g \right]^{-1}, \quad (11-4) \end{aligned}$$

где G — количество кластеров в выборке, каждый кластер состоит из n_g , $g = 1, \dots, G$ наблюдений. [Обратите внимание, что эта матрица просто в $G/(G-1)$ раз увеличенная матрица из (11-3).] Иногда применяемая дальнейшая корректировка (без явной формальной причины) является «корректировкой на степени свободы» $\sum_g n_g / [(\sum_g n_g) - K]$.

Многие другие усовершенствования для более сложных выборок — см. пример по результатам экзаменов — также предлагались. Более подробный анализ см. в книге Кэмерона и Триведи (Cameron, Trivedi (2005, глава 24)). Некоторые вычислительные моменты обсуждаются также в книге Вулдриджа (Wooldridge (2003)). Возникает важный вопрос относительно использования результатов об асимптотических распределениях в случаях, когда число кластеров может быть достаточно мало. Энгрист, Лэви (Angrist, Lavy (2002)) обнаружили, что кластерная корректировка после объединенного МНК, как мы делали в примере 11.1, не настолько полезна, как казалось (хотя наша корректировка с 595 кластерами, каждый из которых имел размер 7, была бы «безопасной» по этим стандартам). Однако могла бы возникнуть сложность, по крайней мере частично, при использовании МНК при наличии обычных эффектов. Кезд (Kezde (2001)) и Бертран, Дюфло, Маллэтан (Bertrand, Dufflo, Mullainathan (2002)) отметили более позитивные результаты, если корректировку применять после оценивания регрессии с фиксированными эффектами. Еще одна сложность возникает, когда группы очень большие, а число групп относительно мало, к примеру, если панель состоит из большого числа крупных выборок из подмножества (или даже из всех) штатов США. Поскольку асимптотическая теория, которую мы использовали, в этом случае предполагает прямо противоположное, то результаты здесь будут менее надежными. Дональд, Лэнг (Donald, Lang (2007)) обнаружили, что этот случай больше тяготеет к анализу групповых средних значе-

ний, нежели индивидуальных данных. Вулдридж (Wooldridge (2003)) представил выводы, помогающие объяснить это открытие. Наконец, существует естественный вопрос о том, необходима ли вообще корректировка, если мы используем обобщенный метод наименьших квадратов для оценки модели со случайными эффектами (см. раздел 11.5) к оценке на первом этапе. Если бы механизм обобщения данных был строго состоятелен в модели со случайными эффектами, то ответ был бы, очевидно, отрицательный. Если принять точку зрения, что модель со случайными эффектами лишь приближение к корреляции по наблюдениям в кластере, тогда осталась бы «корреляция остатков», полученная из корректировки в (11-4) (или неким аналогом GLS). (Однако это поставило бы под вопрос особую корректировку случайных эффектов в разделе 11.5.) Похожий довод также мотивировал бы корректировку после построения модели с фиксированными эффектами. Мы проследим эти возможности в подразделе 11.6.4, после того как подробно рассмотрим оценки фиксированных и случайных эффектов.

Пример 11.2. Повторная продажа картин Моне

В примерах 3.4, 4.5, 4.10 и 6.2 мы рассматривали взаимосвязь между ценой продажи и площадью картины в выборке из 430 торговых сделок по продаже картин Моне. Фактически это была не продажа 430 картин. Многие торговые сделки были повторной продажей тех же картин в разные моменты. На самом деле выборка включает 376 картин. Один раз были проданы 333 картины, два раза — 34, три раза — 7 и четыре раза — 2. Если цена продажи картины как минимум частично обусловлена внутренними ее свойствами, тогда это обусловило бы корректировку стандартных ошибок методом наименьших квадратов, как было указано в (11-4). В табл. 11.2 приведены результаты оценки МНК-регрессии со стандартными и скорректированными стандартными ошибками. Даже при достаточно скромном числе сгруппированных данных влияние корректировки в предполагаемой области больших стандартных ошибок очевидно.

Таблица 11.2. Уравнение цены продажи

| Переменная | Оцененный коэффициент | Станд. ошибка МНК | Скорректированная станд. ошибка |
|------------------------|-----------------------|-------------------|---------------------------------|
| <i>Constant</i> | -9,7068 | 0,5661 | 0,6791 |
| <i>ln Area</i> | 1,3473 | 0,0822 | 0,1030 |
| <i>Signature</i> | 1,2614 | 0,1250 | 0,1281 |
| <i>ln Aspect Ratio</i> | 0,0225 | 0,1479 | 0,1661 |

11.3.4. Устойчивое оценивание с использованием группового среднего значения

Объединенную регрессионную модель можно оценить при помощи выборочных средних значений данных. Предполагаемая регрессионная модель получается из умножения в обратном порядке каждой группы на $(1/T)i'$, где

\mathbf{i}' — строка из единиц:

$$(1/T)\mathbf{i}'\mathbf{y}_i = (1/T)\mathbf{i}'\mathbf{X}_i\boldsymbol{\beta} + (1/T)\mathbf{i}'\mathbf{w}_i$$

или

$$\bar{y}_i = \bar{\mathbf{x}}_i'\boldsymbol{\beta} + \bar{w}_i.$$

В преобразованной линейной регрессии у шоков сохраняются нулевые условные средние значения, но дисперсии гетероскедастичны и равны $\sigma_i^2 = (1/T^2)\mathbf{i}'\boldsymbol{\Omega}_i\mathbf{i}$. При неспецифицированном $\boldsymbol{\Omega}_i$ это будет гетероскедастичной регрессией, для которой мы использовали бы оценку Уайта при получении надежных статистических выводов. Почему мы используем эту оценку, когда доступны полные данные? Если соблюдены стандартные допущения, тогда можно непосредственно показать, что асимптотическая ковариационная матрица для оценки по групповым средним значениям однозначно больше и ответом было бы, что никакой выгоды в этом нет. Однако невыполнение стандартных допущений приводит нас к тому, что вопрос становится уже менее ясным. При наличии неструктурированных кластерных эффектов действенность метода наименьших квадратов может быть значительно ослаблена, как мы видели в предшествующем примере. Потеря информации, возникающая при усреднении, может быть относительно невелика, хотя в принципе неагрегированные данные все же будут лучше.

Мы подчеркиваем, что использование **групповых средних значений** не решает проблему, обусловленную оценкой модели с фиксированными эффектами. Рассмотрим общую модель

$$y_i = \mathbf{X}_i\boldsymbol{\beta} + c_i\mathbf{i} + \mathbf{w}_i,$$

где, как и прежде, c_i — это скрытый эффект. Если предположение о независимости средних значений, $E[c_i | \mathbf{X}_i] = \alpha$, не соблюдается, то эффект будет распространяться и на групповые средние значения. В таком случае $E[c_i | \mathbf{X}_i] = h(\mathbf{X}_i)$. Обычной спецификацией будет модель Мундлака (Mundlak, 1978)

$$E[c_i | \mathbf{X}_i] = \bar{\mathbf{x}}_i'\boldsymbol{\gamma}.$$

(Мы вернемся к этой спецификации в подразделе 11.5.6.) Тогда

$$\begin{aligned} y_{it} &= \mathbf{x}'_{it}\boldsymbol{\beta} + c_i + \varepsilon_{it} \\ &= \mathbf{x}'_{it}\boldsymbol{\beta} + \bar{\mathbf{x}}_i'\boldsymbol{\gamma} + [\varepsilon_{it} + c_i - E[c_i | \mathbf{X}_i]] \\ &= \mathbf{x}'_{it}\boldsymbol{\beta} + \bar{\mathbf{x}}_i'\boldsymbol{\gamma} + u_{it}, \end{aligned}$$

где по построению $E[u_{it} | \mathbf{X}_i] = 0$. Усредняя, как выше, получаем

$$\begin{aligned} \bar{y}_i &= \bar{\mathbf{x}}_i'\boldsymbol{\beta} + \bar{\mathbf{x}}_i'\boldsymbol{\gamma} + \bar{u}_i \\ &= \bar{\mathbf{x}}_i'(\boldsymbol{\beta} + \boldsymbol{\gamma}) + \bar{u}_i. \end{aligned}$$

Подразумевается, что при оценке групповых средних значений оценивается не $\boldsymbol{\beta}$, а $\boldsymbol{\beta} + \boldsymbol{\gamma}$. Усреднение наблюдений в группе дает полный набор эффектов, наблюдаемых и скрытых, в групповых средних значениях.

Остается одно соображение, которое, к сожалению, мы не можем решить аналитически, — это возможность **ошибки измерения**. Если регрессоры измерены с ошибкой, тогда, как мы видели в разделе 8.5, оценка методом наименьших квадратов будет несостоятельна и как следствие эффективность станет вопросом спорным. В условиях панельных данных, если ошибка техники измерения случайна, использование групповых средних значений работало бы в направлении ее усреднения; действительно, в этом случае, предполагая наличие эталонного стандарта $x_{itk} = x_{itk}^* + u_{itk}$, можно было бы показать, что оценка по групповым средним значениям была бы состоятельна, так как $T \rightarrow \infty$, в то время как МНК-оценки не было бы.

Пример 11.3. Устойчивые оценки уравнения заработной платы

В табл. 11.3 приведены оценки уравнения заработной платы из примера 11.1 по групповым средним значениям и исходные оценки методом наименьших квадратов. Как видно, схожие результаты получаются при оценке с помощью средних значений.

Таблица 11.3. Уравнение заработной платы, оцененное МНК

| Коэффициент | Оцененный МНК коэффициент | Устойчивая панельная станд. ошибка | Оценки по групповым средним значениям | Устойчивая станд. ошибка Уайта |
|---------------------|---------------------------|------------------------------------|---------------------------------------|--------------------------------|
| $\beta_1: Constant$ | 5,2511 | 0,1233 | 5,1214 | 0,2078 |
| $\beta_2: Exp$ | 0,04010 | 0,004067 | 0,03190 | 0,004597 |
| $\beta_3: Exp^2$ | -0,0006734 | 0,00009111 | -0,0005656 | 0,0001020 |
| $\beta_4: Wks$ | 0,004216 | 0,001538 | 0,009189 | 0,003578 |
| $\beta_5: Occ$ | -0,1400 | 0,02718 | -0,1676 | 0,03338 |
| $\beta_6: Ind$ | 0,04679 | 0,02361 | 0,05792 | 0,02636 |
| $\beta_7: South$ | -0,05564 | 0,02610 | -0,05705 | 0,02660 |
| $\beta_8: SMSA$ | 0,1517 | 0,02405 | 0,1758 | 0,02541 |
| $\beta_9: MS$ | 0,04845 | 0,04085 | 0,1148 | 0,04989 |
| $\beta_{10}: Union$ | 0,09263 | 0,02362 | 0,1091 | 0,02830 |
| $\beta_{11}: Ed$ | 0,05670 | 0,005552 | 0,05144 | 0,005862 |
| $\beta_{12}: Fem$ | -0,3678 | 0,04547 | -0,3171 | 0,05105 |
| $\beta_{13}: Blk$ | -0,1669 | 0,04423 | -0,1578 | 0,04352 |

11.3.5. Оценка в первых разностях

Взятие первых разностей — еще один метод оценки. Здесь прямой целью будет преобразование скрытой неоднородности в модели. Основной случай имеет вид

$$y_{it} = c_i + \mathbf{x}'_{it}\beta + \varepsilon_{it},$$

что подразумевает уравнение первых разностей в виде

$$\Delta y_{it} = \Delta c_i + (\Delta \mathbf{x}_{it})'\beta + \Delta \varepsilon_{it},$$

или

$$\begin{aligned}\Delta y_{it} &= (\Delta \mathbf{x}_{it})' \boldsymbol{\beta} + \varepsilon_{it} - \varepsilon_{i,t-1} \\ &= (\Delta \mathbf{x}_{it})' \boldsymbol{\beta} + u_{it}.\end{aligned}$$

Преимуществом метода **оценки в первых разностях** является то, что он устраняет скрытую неоднородность из модели в зависимости от того, соответствует ли ей модель с фиксированными или случайными эффектами. Недостаток заключается в том, что взятие разностей также устраняет любые не изменяющиеся во времени переменные из модели. В нашем примере их было три, Ed , Fem и Blk . Если не изменяющиеся во времени переменные в модели не представляют интереса, тогда это надежный подход, который может последовательно оценить параметры изменяющихся во времени переменных. Конечно, это бесполезно в нашем примере, так как влияние Ed на $\ln Wage$ было первоочередной целью анализа. Также отметим, что процедура взятия разностей заменяет корреляцию между наблюдениями в c_i на шок скользящего среднего (МА), $u_{i,t} = \varepsilon_{i,t} - \varepsilon_{i,t-1}$. У нового шока $u_{i,t}$ есть автокорреляция, хотя всего в одном периоде. Процедуры доступны для использования двухшагового допустимого GLS с ошибками в виде МА (см. главу 20). Кроме того, эта модель является естественным вариантом для МНК с устойчивой оценкой ковариации Ньюи–Веста, поскольку истинное число лагов (один) известно. (См. подраздел 20.5.2.)

Общее замечание: при многообразии доступных подходов оценка в первых разностях не является настолько хорошей, чтобы ее рекомендовать, за исключением одного очень важного приложения. Во многих исследованиях используются «панели» с двумя периодами — до и после воздействия. В таких случаях интересующим нас явлением вполне определенно может быть изменение в результирующей переменной — «эффект воздействия». Рассмотрим модель

$$y_{it} = c_i + \mathbf{x}'_{it} \boldsymbol{\beta} + \theta S_{it} + \varepsilon_{it},$$

где $t = 1, 2$ и $S_{it} = 0$ в периоде 1 и 1 в периоде 2; S_{it} показывает «воздействии», которое имело место между двумя наблюдениями. «Эффект воздействия» имеет вид

$$E[\Delta y_i | (\Delta \mathbf{x}_i = 0)] = \theta,$$

что как раз является постоянным членом в регрессии в первых разностях:

$$\Delta y_i = \theta + (\Delta \mathbf{x}_i)' \boldsymbol{\beta} + u_i.$$

Мы подробно рассмотрим подобного рода случаи в разделе 19.6.

11.3.6. Внутри- и межгрупповые оценки

Объединенную регрессионную модель можно записать тремя способами. Первая, исходная, формулировка имеет вид

$$y_{it} = \alpha + \mathbf{x}'_{it} \boldsymbol{\beta} + \varepsilon_{it}. \quad (11-5a)$$

В терминах групповых средних она имеет вид

$$\bar{y}_i = \alpha + \bar{\mathbf{x}}'_i \boldsymbol{\beta} + \bar{\varepsilon}_i, \quad (11-5b)$$

а в терминах отклонений от групповых средних значений —

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' \beta + \varepsilon_{it} - \bar{\varepsilon}_i. \quad (11-5c)$$

[Мы предполагаем, что в \mathbf{x}_{it} нет не изменяющихся во времени переменных, как Ed в примере 11.1. Все они становятся нулевыми в (11-5c).] Все три записи являются классическими регрессионными моделями, и в принципе все они могут быть оценены, по крайней мере, состоятельно, если не эффективно, при помощи метода наименьших квадратов. [Обратите внимание, что (11-5b) определяет лишь n наблюдений групповых средних значений.] Теперь рассмотрим матрицы сумм квадратов остатков и перекрестных произведений, которые использовались бы в каждом случае, сосредотачиваясь только на оценке β . В (11-5a) моменты будут накапливать изменчивость относительно совокупных средних значений, \bar{y} и $\bar{\mathbf{x}}$, и мы будем использовать совокупные суммы квадратов и перекрестных произведений

$$\mathbf{S}_{xx}^{total} = \sum_{i=1}^n \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}})(\mathbf{x}_{it} - \bar{\mathbf{x}})' \quad \text{и} \quad \mathbf{S}_{xy}^{total} = \sum_{i=1}^n \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}})(y_{it} - \bar{y}). \quad (11-6)$$

Для (11-5c), поскольку данные уже рассматриваются в отклонениях, средние значения $(y_{it} - \bar{y}_i)$ и $(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$ равны нулю. Матрицы моментов представляют собой **внутригрупповые** (т.е. с изменчивостью относительно групповых средних) суммы квадратов и перекрестных произведений

$$\mathbf{S}_{xx}^{within} = \sum_{i=1}^n \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' \quad \text{и} \quad \mathbf{S}_{xy}^{within} = \sum_{i=1}^n \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(y_{it} - \bar{y}_i).$$

Наконец, для (11-5b) среднее значение групповых средних есть совокупное среднее. Матрицы моментов являются **межгрупповыми** суммами квадратов и перекрестных произведений, т.е. изменчивостью групповых средних относительно совокупных средних:

$$\mathbf{S}_{xx}^{between} = \sum_{i=1}^n T(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})' \quad \text{и} \quad \mathbf{S}_{xy}^{between} = \sum_{i=1}^n T(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{y}_i - \bar{y}).$$

Легко убедиться в том, что

$$\mathbf{S}_{xx}^{total} = \mathbf{S}_{xx}^{within} + \mathbf{S}_{xx}^{between} \quad \text{и} \quad \mathbf{S}_{xy}^{total} = \mathbf{S}_{xy}^{within} + \mathbf{S}_{xy}^{between}.$$

Таким образом, существует три возможные оценки β методом наименьших квадратов в соответствии с разложением. Оценка методом наименьших квадратов имеет вид

$$\mathbf{b}^{total} = [\mathbf{S}_{xx}^{total}]^{-1} \mathbf{S}_{xy}^{total} = [\mathbf{S}_{xx}^{within} + \mathbf{S}_{xx}^{between}]^{-1} [\mathbf{S}_{xy}^{within} + \mathbf{S}_{xy}^{between}]. \quad (11-7)$$

Внутригрупповая оценка имеет вид

$$\mathbf{b}^{within} = [\mathbf{S}_{xx}^{within}]^{-1} \mathbf{S}_{xy}^{within}. \quad (11-8)$$

Это оценка с фиктивной переменной, рассмотренная в разделе 11.4. Альтернативной была бы межгрупповая оценка

$$\mathbf{b}^{between} = [\mathbf{S}_{xx}^{between}]^{-1} \mathbf{S}_{xy}^{between}. \quad (11-9)$$

Это оценка по групповым средним значениям. Эта оценка (11-5b) методом наименьших квадратов основана на n наборах групповых средних значений. (Обратите внимание, мы предполагаем, что n как минимум такое же по размеру, как K .) Из предшествующих выражений (и известных прежних результатов) получаем

$$\mathbf{S}_{xy}^{within} = \mathbf{S}_{xx}^{within} \mathbf{b}^{within} \quad \text{и} \quad \mathbf{S}_{xy}^{between} = \mathbf{S}_{xx}^{between} \mathbf{b}^{between}.$$

Подставив это в (11-7), мы видим, что оценка методом наименьших квадратов будет равна средневзвешенному по матрице внутри- и межгрупповых оценок:

$$\mathbf{b}^{total} = \mathbf{F}^{within} \mathbf{b}^{within} + \mathbf{F}^{between} \mathbf{b}^{between}, \quad (11-10)$$

где

$$\mathbf{F}^{within} = [\mathbf{S}_{xx}^{within} + \mathbf{S}_{xx}^{between}]^{-1} \mathbf{S}_{xx}^{within} = \mathbf{I} - \mathbf{F}^{between}.$$

Вид полученного результата напоминает байесовскую оценку в классической модели, рассматриваемую в главе 16. Но это сходство более чем мимолетно; можно показать [см., например, Judge et al. (1985)], что

$$\mathbf{F}^{within} = \{ [\text{Asy. Var}(\mathbf{b}^{within})]^{-1} + [\text{Asy. Var}(\mathbf{b}^{between})]^{-1} \}^{-1} [\text{Asy. Var}(\mathbf{b}^{within})]^{-1},$$

что является, по сути, тем же смещенным результатом, что и для байесовской оценки. Во взвешенном среднем оценка с меньшей дисперсией получает больший вес.

Пример 11.4. Анализ ковариации и данные Всемирной организации здравоохранения

Декомпозиция совокупной дисперсии в подразделе 11.3.6 распространяет на линейную модель регрессии уже знакомый нам «дисперсионный анализ», или ANOVA, который часто используется для декомпозиции дисперсии в переменной в кластеризованной или стратифицированной выборке или в панельных данных. Одним из полезных свойств анализа панельных данных, который мы выполняем здесь, является способность рассматривать межгрупповую изменчивость (неоднородность) для изучения основных регрессионных взаимосвязей, внутригрупповой изменчивости, а также динамических эффектов.

Данные Всемирной организации здравоохранения, использованные в примере 6.10, представляют собой несбалансированные панельные данные – в примере 6.10 мы использовали лишь данные за один год. Из 191 страны в выборке 140 наблюдались в течение полных пяти лет, одна наблюдалась четыре раза, а 50 только один раз. В исходных исследованиях ВОЗ (2000a, 2000b) эти данные рассматривались с использованием модели с фиксированными эффектами, которую мы обсудим в следующем

разделе. Оценка такая же, как и в (11-8). Легко заметить, что группы с одним наблюдением выпадают из вычислений, поскольку если $T_i = 1$, то это наблюдение равно групповому среднему. Эти данные использовали многие исследователи в похожих исследованиях панельных данных. [См., например, Greene (2004c) и некоторые ссылки.] Gravelle et al. (2002a) сильно раскритиковали эти исследования, утверждая, что данные ВОЗ больше похожи на межобъектные данные, а не на панельные.

В примере 6.10 модель, которую применяли исследователи в ВОЗ, имела вид

$$\ln DALE_{it} = \alpha_i + \beta_1 \ln Health\ Expenditure_{it} + \beta_2 \ln Education_{it} + \beta_3 \ln^2 Education_{it} + \varepsilon_{it}.$$

Дополнительные модели оценивались с использованием сводного показателя ВОЗ обеспечения охраны здоровья *COMP*. Анализ дисперсии для переменной x_{it} основан на декомпозиции

$$\sum_{i=1}^n \sum_{t=1}^{T_i} (x_{it} - \bar{x})^2 = \sum_{i=1}^n \sum_{t=1}^{T_i} (x_{it} - \bar{x}_{i.})^2 + \sum_{t=1}^n T_i (\bar{x}_{i.} - \bar{x})^2.$$

Деление обеих частей уравнения на левую часть дает декомпозицию

$$1 = Within\text{-}groups\ proportion + Between\text{-}groups\ proportion.$$

Первый член в правой части представляет собой внутригрупповую изменчивость, которая дифференцирует панельные данные из межобъектных (или просто домножим наблюдения на одну и ту же переменную).

В табл. 11.4 представлена декомпозиция дисперсии в переменных, использованных в исследованиях ВОЗ.

Таблица 11.4. Анализ дисперсии по данным обеспечения охраны здоровья ВОЗ

| Переменная | Внутригрупповая изменчивость, % | Межгрупповая изменчивость, % |
|--------------------|---------------------------------|------------------------------|
| <i>COMP</i> | 5,635 | 94,635 |
| <i>DALE</i> | 0,150 | 99,850 |
| <i>Expenditure</i> | 0,635 | 99,365 |
| <i>Education</i> | 0,177 | 99,823 |

Полученные результаты дают повод авторам для такого отношения к данным. Для всех, кроме *COMP*, фактически вся изменчивость в данных сосредоточена между группами, т.е. это пространственная изменчивость. Как утверждают авторы, эти данные лишь ненамного отличаются от пространственных.

11.4. Модель с фиксированными эффектами

Модель с фиксированными эффектами возникает из предположения о том, что исключенные эффекты c_i в общей модели

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + c_i + \varepsilon_{it}$$

коррелированы с включенными переменными. В общем виде

$$E[c_i | \mathbf{X}_i] = h(\mathbf{X}_i). \quad (11-11)$$

Поскольку условное среднее одинаково для каждого периода, мы можем записать модель как

$$\begin{aligned} y_{it} &= \mathbf{x}'_{it}\boldsymbol{\beta} + h(\mathbf{X}_i) + \varepsilon_{it} + [c_i - h(\mathbf{X}_i)] \\ &= \mathbf{x}'_{it}\boldsymbol{\beta} + \alpha_i + \varepsilon_{it} + [c_i - h(\mathbf{X}_i)]. \end{aligned}$$

По построению взятый в скобки член не коррелирован с \mathbf{X}_i , поэтому мы можем включить его в шок и записать модель как

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \alpha_i + \varepsilon_{it}. \quad (11-12)$$

Дальнейшее допущение (обычно неустановленное) заключается в том, что $\text{Var}[c_i | \mathbf{X}_i]$ постоянна. При таком допущении (11-12) становится классической линейной регрессионной моделью. (Скоро мы пересмотрим допущение о гомоскедастичности.) Подчеркнем, что именно (11-11) отличает модель с фиксированными эффектами, а не то, что любая переменная «фиксирована» в этой ситуации и случайна в любой другой. Формулировка о фиксированных эффектах подразумевает, что различия по группам можно уловить в различиях постоянных членов⁵. Каждый α_i принимается за неизвестный оцениваемый параметр.

Перед тем как пойти дальше, снова отметим главный недостаток модели с фиксированными эффектами. Любые **не изменяющиеся во времени** переменные в \mathbf{x}_{it} будут имитировать конкретный постоянный член. Рассмотрите примеры 11.1 и 11.3. Мы могли бы записать модель с фиксированными эффектами как

$$\ln Wage_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + [\beta_{10}Ed_i + \beta_{11}Fem_i + \beta_{12}Blk_i + c_i] + \varepsilon_{it}.$$

Модель с фиксированными эффектами в таком виде будет убирать последние четыре переменные и включать их в α_i . Коэффициенты не изменяющихся во времени переменных нельзя оценить. Такое отсутствие идентификации является своеобразной ценой за устойчивость спецификации к неизмеренной корреляции между обычными эффектами и экзогенными переменными.

⁵Также можно позволить коэффициентам наклона различаться по i , однако такой метод влечет новые методологические проблемы и значительное усложнение в вычислениях. Исследование по этой теме представлено у Cornwell, Schmidt (1984).

11.4.1. Оценки методом наименьших квадратов

Пусть y_i и X_i будут T наблюдениями для i -го элемента, i — столбцом из единиц размером $T \times 1$, а ϵ_i — объединенным $T \times 1$ вектором шоков⁶. Тогда

$$y_i = X_i\beta + i\alpha_i + \epsilon_i.$$

Соединение этих членов дает

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \beta + \begin{bmatrix} i & 0 & \cdots & 0 \\ 0 & i & \cdots & 0 \\ & & \vdots & \\ 0 & 0 & \cdots & i \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

или

$$y = [X \quad d_1 \quad d_2, \dots, d_n] \begin{bmatrix} \beta \\ \alpha \end{bmatrix} + \epsilon, \quad (11-13)$$

где d_i — фиктивная переменная, обозначающая i -й элемент. Пусть матрица $nT \times n$ $D = [d_1, d_2, \dots, d_n]$. Тогда собирание всех nT строк даст

$$y = X\beta + D\alpha + \epsilon.$$

На эту модель, как правило, ссылаются как на **метод наименьших квадратов с фиктивной переменной (LSDV)** (хотя часть «наименьших квадратов» этого названия относится к технике, которая обычно используется для оценки модели, а не к ней самой).

Эта модель является классической регрессионной моделью, поэтому никаких новых результатов для ее анализа не требуется. Если n достаточно мало, то модель можно оценить при помощи обычного метода наименьших квадратов с K регрессорами в X и n столбцами в D как множественную регрессию с $K + n$ параметрами. Конечно, если n исчисляется тысячами, что довольно типично, тогда модель, вероятнее всего, превысит емкость любого компьютера. Однако при использовании уже известных нам результатов для частичной регрессии мы можем сократить число вычислений⁷. Запишем оценку методом наименьших квадратов для β как

$$b = [X'M_D X]^{-1} [X'M_D y] = b^{within}, \quad (11-14)$$

где

$$M_D = I - D(D'D)^{-1}D'.$$

При использовании преобразованных данных $X_* = M_D X$ и $y_* = M_D y$ получим оценку регрессии методом наименьших квадратов. Структура D очень удобна; ее столбцы ортогональны, поэтому

$$M_D = \begin{bmatrix} M^0 & 0 & 0 & \cdots & 0 \\ 0 & M^0 & 0 & \cdots & 0 \\ & & \cdots & & \\ 0 & 0 & 0 & & M^0 \end{bmatrix}.$$

⁶Предположение о фиксированном размере группы T здесь сделано только для удобства. Как отмечается в подразделе 11.2.4, несбалансированный случай является незначительным изменением.

⁷См. теорему 3.3.

Каждая матрица на диагонали равна

$$\mathbf{M}^0 = \mathbf{I}_T - \frac{1}{T} \mathbf{i} \mathbf{i}' \quad (11-15)$$

Умножение слева любого $T \times 1$ вектора \mathbf{z}_i на \mathbf{M}^0 дает $\mathbf{M}^0 \mathbf{z}_i = \mathbf{z}_i - \bar{z}_i \mathbf{i}$. (Обратите внимание, что среднее значение взято только по T наблюдениям для элемента i .) Таким образом, оценка регрессии $\mathbf{M}_D \mathbf{y}$ на $\mathbf{M}_D \mathbf{X}$ методом наименьших квадратов тождественна оценке регрессии $[y_{it} - \bar{y}_i]$ на $[\mathbf{x}_{it} - \bar{\mathbf{x}}_i]$, где \bar{y}_i и $\bar{\mathbf{x}}_i$ скалярные и есть вектор размером $K \times 1$ средних значений y_{it} и \mathbf{x}_{it} по T наблюдениям для группы i ⁸. Коэффициенты фиктивной переменной можно восстановить из другого нормального уравнения частичной регрессии:

$$\mathbf{D}' \mathbf{D} \mathbf{a} + \mathbf{D}' \mathbf{X} \mathbf{b} = \mathbf{D}' \mathbf{y}$$

или

$$\mathbf{a} = [\mathbf{D}' \mathbf{D}]^{-1} \mathbf{D}' (\mathbf{y} - \mathbf{X} \mathbf{b}).$$

Это подразумевает, что для каждого i

$$a_i = \bar{y}_i - \bar{\mathbf{x}}_i' \mathbf{b}. \quad (11-16)$$

Подходящая оценка асимптотической ковариационной матрицы для \mathbf{b} имеет вид

$$\text{Est. Asy. Var}[\mathbf{b}] = s^2 [\mathbf{X}' \mathbf{M}_D \mathbf{X}]^{-1} = s^2 [\mathbf{S}_{xx}^{\text{within}}]^{-1}, \quad (11-17)$$

где используется матрица вторых моментов с иксами, которые сейчас выражены как отклонения от их соответствующих групповых средних. Оценка дисперсии шоков такова:

$$s^2 = \frac{\sum_{i=1}^n \sum_{t=1}^T (y_{it} - \mathbf{x}_{it}' \mathbf{b} - a_i)^2}{nT - n - K} = \frac{(\mathbf{M}_D \mathbf{y} - \mathbf{M}_D \mathbf{X} \mathbf{b})' (\mathbf{M}_D \mathbf{y} - \mathbf{M}_D \mathbf{X} \mathbf{b})}{nT - n - K}. \quad (11-18)$$

it -й остаток, использованный в вычислении, имеет вид

$$e_{it} = y_{it} - \mathbf{x}_{it}' \mathbf{b} - a_i = y_{it} - \mathbf{x}_{it}' \mathbf{b} - (\bar{y}_i - \bar{\mathbf{x}}_i' \mathbf{b}) = (y_{it} - \bar{y}_i) - (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' \mathbf{b}.$$

Итак, числитель в s^2 в точности представляет собой сумму квадратов остатков метода наименьших квадратов с использованием коэффициентов наклонов метода наименьших квадратов и данные в форме отклонений от групповых средних значений. Но можно было бы далее использовать $nT - K$ вместо $nT - n - K$ в качестве делителя в вычислении s^2 , и тогда была бы необходима корректировка. Для индивидуальных эффектов

$$\text{Asy. Var}[a_i] = \frac{\sigma_\varepsilon^2}{T} + \bar{\mathbf{x}}_i' \{ \text{Asy. Var}[\mathbf{b}] \} \bar{\mathbf{x}}_i, \quad (11-19)$$

и так можно вычислить простую оценку, основанную на s^2 .

⁸Интересная ситуация возникает, если $T = 2$. В случае с двумя периодами вы можете показать — мы оставляем это в качестве упражнения, — что эта регрессия, оцененная методом наименьших квадратов, построена с использованием $nT/2$ наблюдений первых разностей через построение регрессии наблюдения $(y_{i2} - y_{i1})$ (и его отрицательного) на $(\mathbf{x}_{i2} - \mathbf{x}_{i1})$ (и его отрицательного).

11.4.2. Асимптотики при малых T

Из (11-17) найдем

$$\begin{aligned}
 \text{Asy. Var}[\mathbf{b}] &= \sigma_\varepsilon^2 [\mathbf{X}' \mathbf{M}_D \mathbf{X}]^{-1} \\
 &= \frac{\sigma_\varepsilon^2}{n} \left[\frac{1}{n} \sum_{i=1}^n \mathbf{X}_i' \mathbf{M}^0 \mathbf{X}_i \right]^{-1} \\
 &= \frac{\sigma_\varepsilon^2}{n} \left[\frac{1}{n} \sum_{i=1}^n \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_{i.})(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i.})' \right]^{-1} \quad (11-20) \\
 &= \frac{\sigma_\varepsilon^2}{n} \left[T \frac{1}{n} \sum_{i=1}^n \frac{1}{T} \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_{i.})(\mathbf{x}_{it} - \bar{\mathbf{x}}_{i.})' \right]^{-1} \\
 &= \frac{\sigma_\varepsilon^2}{n} [T \bar{S}_{xx,i}]^{-1}.
 \end{aligned}$$

Так как оценки методом наименьших квадратов в этой модели не смещены, вопрос о (среднеквадратической) состоятельности сводится к рассмотрению ковариационной матрицы. Сходится ли матрица, рассмотренная выше, к нулю? Нужно быть конкретными в том, что мы подразумеваем под сходимостью. В этой ситуации увеличение размера выборки имеет отношение к увеличению n , т.е. числа групп. Предполагается, что размер группы T является фиксированным. С увеличением n главный скаляр, очевидно, исчезает. Матрица в квадратных скобках в T раз превышает среднее по n группам во внутригрупповых ковариационных матрицах переменных в \mathbf{X}_i . Поскольку данные являются доброкачественными, мы можем допустить, что матрица в скобках не сходится к нулевой матрице (или матрице с нулями по диагонали). На этом основании мы можем ожидать состоятельности оценки методом наименьших квадратов. С практической точки зрения для этого необходима внутригрупповая изменчивость данных. Заметим, что этот результат рассыпается, если в \mathbf{X}_i есть не изменяющиеся во времени переменные, поскольку тогда по диагоналям матрицы в скобках будут нули. Этот результат также раскрывает характер проблемы данных ВОЗ в примере 11.4, что было проанализировано в работе Gravelle et al. (2002).

Теперь рассмотрим результат в (11-19) для асимптотической дисперсии a_i . Предположим, что \mathbf{b} состоятельна, как было показано ранее. Тогда с увеличением n асимптотическая дисперсия a_i снижается до более низкой границы σ_ε^2/T , которая не сходится к нулю. Оценки постоянных членов в модели с фиксированными эффектами не являются состоятельными оценками α_i . Они несостоятельны, поскольку сходятся к неверному параметру. Они таковы, так как их асимптотические дисперсии не сходятся к нулю даже по мере увеличения размера выборки. Легко понять, почему так происходит. Из (11-16) мы видим, что каждый a_i оценивается с использованием только T наблюдений, предположим, что n было бесконечно, чтобы β был известен. Поскольку не предполагается, что T будет увеличиваться, мы получаем удивительный результат. Постоянные члены несостоятельны, если только нет $T \rightarrow \infty$, что не является частью модели.

11.4.3. Проверка значимости групповых эффектов

t -критерий для α_i может использоваться для проверки гипотезы о том, что α_i равен нулю. Однако эта гипотеза об одной конкретной группе, как правило, непригодна для проверки в контексте данной регрессии. Если нас интересуют различия по группам, то мы можем проверить гипотезу о том, что все постоянные члены равны, по F -тесту. В условиях нулевой гипотезы о равенстве фиксированных эффектов эффективной будет оценка методом наименьших квадратов объединенной регрессии. F -критерий, используемый для этого теста, имеет вид

$$F(n-1, nT-n-K) = \frac{(R_{LSDV}^2 - R_{Pooled}^2)/(n-1)}{(1 - R_{LSDV}^2)/(nT-n-K)}, \quad (11-21)$$

где $LSDV$ — модель с фиктивной переменной, а $Pooled$ (объединенные) — объединенная или приведенная модель с одним общим постоянным членом. Кроме того, модель можно было бы вместо этого оценить, используя общую константу и $n-1$ фиктивную переменную. Все другие результаты (например, коэффициенты наклонов s^2 , R^2) останутся неизменными, но вместо оценки α_i каждый коэффициент фиктивной переменной теперь будет оценкой $\alpha_i - \alpha_1$, где группа «1» — пропущенная группа. F -тест о том, что коэффициенты этих $n-1$ фиктивных переменных равны нулю, тождествен тесту, описанному выше. Однако важно помнить, что, хотя статистические результаты одни и те же, интерпретация коэффициентов при фиктивной переменной в двух формулировках разная⁹.

11.4.4. Фиксированные временные и групповые эффекты

Метод наименьших квадратов с фиктивной переменной можно расширить, включив в него также временной эффект. Один из способов записать расширенную модель — это просто включить в нее временной эффект:

$$y_{it} = \mathbf{x}'_{it}\beta + \alpha_i + \delta_t + \varepsilon_{it}. \quad (11-22)$$

Эта модель получена из предшествующей включением дополнительной $T-1$ фиктивной переменной. (Один из временных эффектов должен быть пропущен, чтобы избежать совершенной коллинеарности, — групповые эффекты и временные эффекты в сумме равны единице.) Если число переменных слишком велико, чтобы оценивать обычную регрессию, то эту модель также можно оценить с использованием частичной регрессии¹⁰. Однако в этой формулировке существует асимметрия, поскольку каждый из групповых эффектов представляет собой пересечение в рамках конкретной группы, в то время как временные эффекты являются **противопоставлениями**,

⁹Обсуждение различий см. в работе Suits (1984).

¹⁰Матричная алгебра и теоретическое рассмотрение двусторонних эффектов в моделях панельных регрессий очень трудны. См., например, Baltagi (2008). К счастью, практическое применение гораздо проще. Число периодов, анализируемых в большей части панельных данных, редко превышает небольшое число. Так как современные компьютерные программы, как правило, позволяют работать с дюжинами (и даже сотнями) регрессоров, почти любой пример, включающий второй фиксированный эффект, можно решить через буквальное включение второго эффекта в качестве набора действующих фиктивных переменных.

т.е. сравнениями с базовым периодом (с тем, который исключен). Симметричная форма модели имеет вид

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \mu + \alpha_i + \delta_t + \varepsilon_{it}, \quad (11-23)$$

где эффекты полных n и T включены, однако накладываются ограничения:

$$\sum_i \alpha_i = \sum_t \delta_t = 0.$$

Оценки коэффициента наклона методом наименьших квадратов в этой модели получены оценкой регрессии

$$y_{*it} = y_{it} - \bar{y}_{i.} - \bar{y}_{.t} + \bar{\bar{y}} \quad (11-24)$$

на

$$\mathbf{x}_{*it} = \mathbf{x}_{it} - \bar{\mathbf{x}}_{i.} - \bar{\mathbf{x}}_{.t} + \bar{\bar{\mathbf{x}}},$$

где средние значения для конкретного периода и общие имеют вид

$$\bar{y}_{.t} = \frac{1}{n} \sum_{i=1}^n y_{it} \quad \text{и} \quad \bar{\bar{y}} = \frac{1}{nT} \sum_{i=1}^n \sum_{t=1}^T y_{it},$$

так же и для $\bar{\mathbf{x}}_{.t}$ и $\bar{\bar{\mathbf{x}}}$. Общая константа и коэффициенты при фиктивных переменных теперь можно извлечь из нормальных уравнений вида

$$\begin{aligned} \hat{\mu} &= m = \bar{\bar{y}} - \bar{\bar{\mathbf{x}}}'\mathbf{b}, \\ \hat{\alpha}_i &= a_i = (\bar{y}_{i.} - \bar{\bar{y}}) - (\bar{\mathbf{x}}_{i.} - \bar{\bar{\mathbf{x}}})'\mathbf{b}, \\ \hat{\delta}_t &= d_t = (\bar{y}_{.t} - \bar{\bar{y}}) - (\bar{\mathbf{x}}_{.t} - \bar{\bar{\mathbf{x}}})'\mathbf{b}. \end{aligned} \quad (11-25)$$

Оцененная асимптотическая ковариационная матрица для \mathbf{b} посчитана с использованием сумм квадратов и перекрестных произведений \mathbf{x}_{*it} , вычисленных в (11-24), и

$$s^2 = \frac{\sum_{i=1}^n \sum_{t=1}^T (y_{it} - \mathbf{x}'_{it}\mathbf{b} - m - a_i - d_t)^2}{nT - (n-1) - (T-1) - K - 1}. \quad (11-26)$$

Если одно из чисел, n или T , маленькое, а другое большое, то может оказаться проще рассмотреть меньшее множество как обычный набор переменных и применить полученные ранее результаты к модели с односторонними фиксированными эффектами, определенными по большему множеству. Несмотря на то что эта модель общепринятая, она редко используется на практике. Для этого есть две причины. Во-первых, издержки в виде числа степеней свободы часто неоправданны. Во-вторых, в тех ситуациях, в которых необходима модель временной эволюции шока, обычно используют более общую модель, нежели эту простую формулировку с фиктивной переменной.

11.4.5. Постоянные во времени переменные и декомпозиция вектора фиксированных эффектов

Наличие не изменяющихся во времени переменных (TIV) в регрессии общих эффектов представляет весьма неприятную проблему для исследователя, строящего модель. Значительной проблемой для модели с фиксированными эффектами (FEM) является то, что оценка согласуется с TIV. Так, в уравнении заработной платы из примера 11.5 мы опустили три переменные, представляющие значительный интерес в модели с фиксированными эффектами, — Ed , Fem и Blk . Если мы запишем FEM с набором не изменяющихся во времени переменных в ней как

$$y = X\beta + Z\gamma + D\alpha + \varepsilon$$

с Z , который был бы матрицей M не изменяющихся во времени переменных, тогда это становится проблемой мультиколлинеарности. Поскольку столбцы матрицы D являются полным набором n фиктивных переменных, любая не изменяющаяся во времени переменная в Z может быть записана как линейная комбинация из столбцов D . Пусть m -й столбец Z будет TIV, $Z(m) = (z_{m1}, z_{m1}, \dots, z_{m2}, z_{m2}, \dots, z_{mn}, z_{mn}, \dots)'$; каждое конкретное значение z_{mi} повторяется T_i раз. Тогда $Z(m)$ равно Dz_m , где z_m — $n \times 1$ вектор $(z_{m1}, z_{m2}, \dots, z_{mn})'$. Собрав все M столбцов, мы получим $Z = DZ_n$, где Z_n — матрица $n \times m$ (z_1, z_2, \dots, z_m) . Если мы сделаем попытку вычислить LSDV-оценку для $(\beta', \gamma)'$ из (11-14), используя преобразованные переменные $M_D[X, Z]$, столбцы Z будут преобразованы в отклонения от групповых средних, которые являются столбцами из нулей, поскольку Z уже представляет собой средние значения периода, отклонения от групповых средних равны нулю. LSDV-регрессию нельзя вычислить при наличии TIV. С теоретической точки зрения проблема состоит в том, что γ не идентифицирована. Никакое количество данных не может отделить γ от α . Модель имела бы такой вид:

$$y = X\beta + D(Z_n\gamma) + D\alpha + \varepsilon = X\beta + D[Z_n\gamma + \alpha] + \varepsilon.$$

В случае с фиксированными эффектами идентифицирующее ограничение равно $\gamma = 0$, т.е. в модели с фиксированными эффектами коэффициенты для постоянных во времени переменных неидентифицируемы с точки зрения моментов данных, поэтому их коэффициенты зафиксированы в нуле, чтобы идентифицировать α .

Плампер, Трёгер (Plümper, Troeger (2007)) предложили трехшаговую процедуру, которую назвали **декомпозицией вектора фиксированных эффектов** (FEVD), она предлагает решение проблемы оценки коэффициентов не изменяющихся во времени переменных в модели с фиксированными эффектами и одновременно значительно увеличивает эффективность оценки параметров. Выполняются следующие три шага:

Шаг 1: линейная регрессия y на X и D для оценки α , т.е. нужно вычислить LSDV-оценку для β из (11-14) и использовать (11-16), чтобы вычислить оценки индивидуальных постоянных членов.

Шаг 2: линейная регрессия n оцененных постоянных членов $a_i, i = 1, \dots, n$ на постоянный член и Z_n . С помощью этой регрессии можно посчитать n

остатков, \mathbf{h}_n . Затем мы распространим этот вектор на всю длину выборки, используя $\mathbf{h} = \mathbf{D}\mathbf{h}_n$.

Шаг 3: линейная регрессия y на $[\mathbf{X}, (\mathbf{i}, \mathbf{Z}), \mathbf{h}]$, где \mathbf{i} — общая константа, чтобы оценить $(\beta, \alpha^0, \gamma, \delta)$ в $y = \mathbf{X}\beta + \alpha^0 + \mathbf{Z}\gamma + \mathbf{h}\delta + \varepsilon$.

Такое решение дает интересные алгебраические результаты, которые будут полезны для анализа в этой главе. Удивительный результат, который обсуждался в нескольких недавних комментариях, включая работы (Breusch, Ward, Nguyen, Kompas (2010)), (Chatelain, Ralf (2010)) и (Greene (2010)), состоит в том, что шаг 3 просто воспроизводит результаты шагов 1 и 2, однако ковариационная матрица, вычисленная для оценки β на шаге 3, нетождественна и однозначно меньше, чем матрица в (11-17). Весьма полезно сделать подробный вывод результатов.

Мы докажем следующие результаты:

FEVD 1. Оцененные коэффициенты для \mathbf{X} на шаге 3 тождественны полученным на шаге 1.

FEVD 2. Оцененные коэффициенты для (\mathbf{i}, \mathbf{Z}) на шаге 3 тождественны полученным на шаге 2.

FEVD 3. Оцененный коэффициент для \mathbf{h} на шаге 3 равен 1.

FEVD 4. Сумма квадратов остатков в регрессии на шаге 3 тождественна сумме, полученной на шаге 1.

FEVD 5. s^2 , вычисленная на шаге 3, меньше, чем полученная на шаге 1.

FEVD 6. Асимптотическая ковариационная матрица, вычисленная для оценки β на шаге 3, меньше, чем полученная на шаге 1 (даже несмотря на то, что оценки тождественны алгебраически), из-за FEVD 5 и поскольку использованная матрица меньше.

(Заметьте, что существует намного больше сжатых доказательств этих результатов. Следующие подходы используются, чтобы продемонстрировать инструментарий, рассмотренный нами в этой и предшествующих главах.)

Доказательства результатов: запишите результаты оценки трех регрессий методом наименьших квадратов как

$$\text{(шаг 1)} \quad \mathbf{y} = \mathbf{X}\mathbf{b}_{\text{LSDV}} + \mathbf{D}\mathbf{a}_{\text{LSDV}} + \mathbf{e}_{\text{LSDV}},$$

$$\text{(шаг 2)} \quad \mathbf{a}_{\text{LSDV}} = \mathbf{W}_n \mathbf{c}_{\text{LSDV}} + \mathbf{h}_n, \text{ где } \mathbf{W}_n = (\mathbf{i}_n, \mathbf{Z}_n),$$

$$\text{(шаг 3)} \quad \mathbf{y} = \mathbf{X}\mathbf{b}_{\text{FEVD}} + \mathbf{W}\mathbf{c}_{\text{FEVD}} + \mathbf{h}d_{\text{FEVD}} + \mathbf{e}_{\text{FEVD}}, \text{ где } \mathbf{W} = (\mathbf{i}, \mathbf{Z}).$$

Итак, \mathbf{W} на шаге 3 включает M не изменяющихся во времени переменных и общую константу. Начнем с того, что докажем равенство $\mathbf{e}_{\text{LSDV}} = \mathbf{e}_{\text{FEVD}}$. Вспомним, что $\mathbf{Z} = \mathbf{D}\mathbf{Z}_n$ и $\mathbf{i} = \mathbf{D}\mathbf{i}_n$, где \mathbf{i}_n — это вектор-столбец размерностью $n \times 1$ из единиц. Остатки, полученные на втором шаге, равны $\mathbf{h}_n = \mathbf{a}_{\text{LSDV}} - \mathbf{W}_n \mathbf{c}_{\text{LSDV}}$ и $\mathbf{h} = \mathbf{D}\mathbf{h}_n$. Отсюда результат на шаге 3 эквивалентен

$$\mathbf{y} = \mathbf{X}\mathbf{b}_{\text{FEVD}} + \mathbf{D}\mathbf{W}_n \mathbf{c}_{\text{FEVD}} + \mathbf{D}(\mathbf{a}_{\text{LSDV}} - \mathbf{W}_n \mathbf{c}_{\text{LSDV}})d_{\text{FEVD}} + \mathbf{e}_{\text{FEVD}}.$$

После небольших преобразований

$$\mathbf{y} = \mathbf{X}\mathbf{b}_{\text{FEVD}} + \mathbf{D}\mathbf{a}_{\text{LSDV}} + \mathbf{D}\mathbf{W}_n \mathbf{c}_{\text{FEVD}} - \mathbf{D}\mathbf{W}_n \mathbf{c}_{\text{LSDV}}(d_{\text{FEVD}}) + \mathbf{e}_{\text{FEVD}}. \quad (11-27)$$

Первые два члена — предсказанные значения из линейной регрессии y на \mathbf{X} и \mathbf{D} , а третий и четвертый члены просто добавляют больше линейных ком-

бинаций столбцов \mathbf{D} . Поскольку (\mathbf{X}, \mathbf{D}) имеет (как мы предположили) полный ранг по столбцам, регрессия, оцененная методом наименьших квадратов, должна давать тот же результат, что и на шаге 1. Остатки должны быть тождественными, т.е. $\mathbf{e}_{\text{FEVD}} = \mathbf{e}_{\text{LSDV}}$. Теперь умножим (11-27) в обратном порядке на $\mathbf{X}'\mathbf{M}_D$. Так как $\mathbf{M}_D\mathbf{D} = \mathbf{0}$ и $\mathbf{M}_D\mathbf{e}_{\text{LSDV}} = \mathbf{e}_{\text{LSDV}}$, найдем

$$\mathbf{X}'\mathbf{M}_D\mathbf{y} = \mathbf{X}'\mathbf{M}_D\mathbf{X}\mathbf{b}_{\text{FEVD}} + \mathbf{X}'\mathbf{e}_{\text{LSDV}}.$$

Поскольку $\mathbf{X}'\mathbf{e}_{\text{LSDV}} = \mathbf{0}$ (из шага 1), имеем $\mathbf{b}_{\text{FEVD}} = (\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1}(\mathbf{X}'\mathbf{M}_D\mathbf{y}) = \mathbf{b}_{\text{LSDV}}$, что доказывает FEVD 1.

Чтобы вычислить \mathbf{c}_{FEVD} , на шаге 3 мы имеем следующее решение (используя $\mathbf{b}_{\text{FEVD}} = \mathbf{b}_{\text{LSDV}}$ и $\mathbf{e}_{\text{FEVD}} = \mathbf{e}_{\text{LSDV}}$):

$$\mathbf{y} - \mathbf{X}\mathbf{b}_{\text{LSDV}} = \mathbf{W}\mathbf{c}_{\text{FEVD}} + \mathbf{h}d_{\text{FEVD}} + \mathbf{e}_{\text{LSDV}}.$$

Умножим это выражение слева на \mathbf{W}' . На шаге 2 получим $\mathbf{W}'\mathbf{h} = \mathbf{W}_n'\mathbf{D}'\mathbf{D}\mathbf{h}_n = \mathbf{0}$. Это верно, поскольку $\mathbf{D}'\mathbf{D}$ является диагональной матрицей с T_i по диагоналям. Таким образом, каждый элемент в $\mathbf{W}'\mathbf{h}$ имеет вид $T_i\mathbf{W}(m)'\mathbf{h}_n = 0$, где $\mathbf{W}(m)$ — m -й столбец \mathbf{W}_n . Из шага 3 $\mathbf{W}'\mathbf{e}_{\text{FEVD}} = \mathbf{W}'\mathbf{e}_{\text{LSDV}} = \mathbf{0}$. Таким образом,

$$\mathbf{W}'(\mathbf{y} - \mathbf{X}\mathbf{b}_{\text{LSDV}}) = \mathbf{W}'\mathbf{W}\mathbf{c}_{\text{FEVD}},$$

поэтому

$$\mathbf{c}_{\text{FEVD}} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'(\mathbf{y} - \mathbf{X}\mathbf{b}_{\text{LSDV}}).$$

Из шага 1 $\mathbf{y} - \mathbf{X}\mathbf{b}_{\text{LSDV}} = \mathbf{D}\mathbf{a}_{\text{LSDV}} + \mathbf{e}_{\text{LSDV}}$. Поскольку $\mathbf{W}'\mathbf{e}_{\text{FEVD}} = \mathbf{W}'\mathbf{e}_{\text{LSDV}} = \mathbf{0}$, из шага 3 имеем

$$\mathbf{c}_{\text{FEVD}} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{D}\mathbf{a}_{\text{LSDV}}.$$

Однако при умножении слева выражения из шага 2 на \mathbf{D} найдем $\mathbf{D}\mathbf{a}_{\text{LSDV}} = \mathbf{D}\mathbf{W}_n\mathbf{c}_{\text{LSDV}} + \mathbf{D}\mathbf{h}_n$. Из этого следует, что решение имеет вид

$$\mathbf{c}_{\text{LSDV}} = (\mathbf{W}_n'\mathbf{D}'\mathbf{D}\mathbf{W}_n)^{-1}\mathbf{W}_n'\mathbf{D}'\mathbf{D}\mathbf{a}_{\text{LSDV}} + (\mathbf{W}_n'\mathbf{D}'\mathbf{D}\mathbf{W}_n)^{-1}\mathbf{W}_n'\mathbf{D}'\mathbf{D}\mathbf{h}_n.$$

Второй член равен нулю, как было показано ранее. Итоговый результат принимает вид $\mathbf{c}_{\text{LSDV}} = \mathbf{c}_{\text{FEVD}}$, что представляет FEVD 2.

Снова применив шаг 3, найдем теперь d_{FEVD} , используя уже полученные результаты. Решение имеет следующий вид:

$$\mathbf{y} - \mathbf{X}\mathbf{b}_{\text{LSDV}} - \mathbf{W}\mathbf{c}_{\text{LSDV}} = \mathbf{h}d_{\text{FEVD}} + \mathbf{e}_{\text{LSDV}}.$$

Однако $\mathbf{y} - \mathbf{X}\mathbf{b}_{\text{LSDV}} = \mathbf{a} + \mathbf{e}_{\text{LSDV}} = \mathbf{D}\mathbf{a}_{\text{LSDV}} + \mathbf{e}_{\text{LSDV}}$ и $\mathbf{W}\mathbf{c}_{\text{LSDV}} = \mathbf{a} - \mathbf{h} = \mathbf{D}\mathbf{a}_{\text{LSDV}} - \mathbf{h}$. Подставляя, получаем

$$\mathbf{D}\mathbf{a}_{\text{LSDV}} + \mathbf{e}_{\text{LSDV}} - \mathbf{D}\mathbf{a}_{\text{LSDV}} + \mathbf{h} = \mathbf{h}d_{\text{FEVD}} + \mathbf{e}_{\text{LSDV}}$$

или

$$\mathbf{h} + \mathbf{e}_{\text{LSDV}} = \mathbf{h}d_{\text{FEVD}} + \mathbf{e}_{\text{LSDV}},$$

из чего следует, что $d_{\text{FEVD}} = 1$. Это доказывает FEVD 3.

Выражение FEVD 4 уже было доказано, поскольку $\mathbf{e}_{\text{FEVD}} = \mathbf{e}_{\text{LSDV}}$. R^2 в двух регрессиях также равны, так как $R_{\text{FEVD}}^2 = 1 - (\mathbf{e}_{\text{FEVD}}' \mathbf{e}_{\text{FEVD}} / \mathbf{y}' \mathbf{M}^0 \mathbf{y}) = R_{\text{LSDV}}^2$, поскольку равны векторы остатков. [См. (3-26).] Но

$$\begin{aligned} s_{\text{FEVD}}^2 &= \mathbf{e}_{\text{FEVD}}' \mathbf{e}_{\text{FEVD}} / (\sum_i T_i - K - M - 1 - 1) < s_{\text{LSDV}}^2 = \\ &= \mathbf{e}_{\text{LSDV}}' \mathbf{e}_{\text{LSDV}} / (\sum_i T_i - K - n). \end{aligned}$$

Различие состоит в корректировке на степени свободы, которая может быть значительной. В нашем примере, рассмотренном далее, $DF_{\text{FEVD}} = 4165 - 9 - 3 - 1 - 1 = 4151$, в то время как $DF_{\text{LSDV}} = 4165 - 9 - 595 = 3561$. Тогда для этого примера $s_{\text{FEVD}}^2 / s_{\text{LSDV}}^2 = 0,85787$. Это устанавливает FEVD 5. Для доказательства FEVD 6, основанного на (11-17), сравним

$$\text{Est.Asy.Var}[\mathbf{b}_{\text{FEVD}}] = s_{\text{FEVD}}^2 (\mathbf{X}' \mathbf{M}_{\mathbf{W}, \mathbf{h}} \mathbf{X})^{-1}$$

с

$$\text{Est.Asy.Var}[\mathbf{b}_{\text{LSDV}}] = s_{\text{LSDV}}^2 (\mathbf{X}' \mathbf{M}_{\mathbf{D}} \mathbf{X})^{-1}.$$

Мы уже показали, что $s_{\text{LSDV}}^2 > s_{\text{FEVD}}^2$. Чтобы сравнить матрицы, сравним обратные к ним матрицы и покажем, что матрица разностей

$$\mathbf{A} = \mathbf{X}' \mathbf{M}_{\mathbf{W}, \mathbf{h}} \mathbf{X} - \mathbf{X}' \mathbf{M}_{\mathbf{D}} \mathbf{X}$$

положительно определена. Это подразумевает, что обратная матрица в $\text{Est.Asy.Var}[\mathbf{b}_{\text{FEVD}}]$ меньше, нежели в $\text{Est.Asy.Var}[\mathbf{b}_{\text{LSDV}}]$. Чтобы это показать, отметим, что $\mathbf{R} = (\mathbf{W}, \mathbf{h}) = \mathbf{D}(\mathbf{W}_n, \mathbf{h}_n)$ представляет собой $M + 2$ линейных комбинаций столбцов \mathbf{D} , в то время как \mathbf{D} состоит из всех n столбцов из \mathbf{D} . Остатки, определенные через $\mathbf{M}_{\mathbf{D}} \mathbf{X}$ [см. (3-15)], получены оценкой регрессий \mathbf{X} на все n столбцов из \mathbf{D} . Они будут тождественны остаткам, полученным оценкой регрессии \mathbf{X} на любые n линейно независимых комбинаций столбцов из \mathbf{D} . Для них мы будем использовать $[\mathbf{R}, \mathbf{Q}]$, где \mathbf{Q} ортогональна \mathbf{R} . Отсюда $\mathbf{X}' \mathbf{M}_{\mathbf{D}} \mathbf{X} = \mathbf{X}' \mathbf{M}_{\mathbf{R}, \mathbf{Q}} \mathbf{X}$. Развивая это, получим

$$\mathbf{A} = \mathbf{X}' \mathbf{X} - \mathbf{X}' \mathbf{R} (\mathbf{R}' \mathbf{R})^{-1} \mathbf{R}' \mathbf{X} - \mathbf{X}' \mathbf{X} + \mathbf{X}' (\mathbf{RQ}) \left[\left(\begin{array}{c} \mathbf{R}' \\ \mathbf{Q}' \end{array} \right) (\mathbf{RQ}) \right]^{-1} \left(\begin{array}{c} \mathbf{R}' \\ \mathbf{Q}' \end{array} \right) \mathbf{X}.$$

Обратная матрица упрощается при помощи $\mathbf{R}' \mathbf{Q} = \mathbf{0}$, поэтому матрица в скобках и обратная ей блочно-диагональны. Перемножая, получаем

$$\mathbf{A} = \mathbf{X}' \mathbf{Q} (\mathbf{Q}' \mathbf{Q})^{-1} \mathbf{Q}' \mathbf{X} = \mathbf{X}' (\mathbf{I} - \mathbf{M}_{\mathbf{Q}}) \mathbf{X}.$$

Поскольку матрица $\mathbf{I} - \mathbf{M}_{\mathbf{Q}}$ идемпотентна, матрица $\mathbf{A} = \mathbf{X}' (\mathbf{I} - \mathbf{M}_{\mathbf{Q}})' (\mathbf{I} - \mathbf{M}_{\mathbf{Q}}) \mathbf{X} = \mathbf{X}'^* \mathbf{X}^*$ положительно определена. Это доказывает, что вычисленная ковариационная матрица для \mathbf{b}_{FEVD} всегда будет строго меньше, чем ковариационная матрица для \mathbf{b}_{LSDV} , что представляет собой FEVD 6.

После этого остается то, что можно назвать неопределенным окончанием анализа. Как было возможно оценить γ (на шаге 2 или шаге 3), учитывая, что она не идентифицирована в исходной модели? Ответом является решающее допущение, уже отмеченное несколько раз. По исходной спецификации \mathbf{Z} не коррелирована с ϵ . Однако для регрессии (на шаге 2), которая нужна

для оценки ненулевого γ , далее должно предполагаться, что переменная z_i не коррелирована с u_i . Это предположение ограничивает исходную модель с фиксированными эффектами, это гибрид, в котором допускается, что изменяющиеся во времени переменные коррелированы с u_i , однако это не допускается для не изменяющихся во времени переменных. Авторы отмечают это на странице 6 и в сноске 7, где утверждают: «Если предполагается, что не изменяющиеся во времени переменные ортогональны по отношению к ненаблюдаемым элементам, т.е. если предположение, лежащее в основе нашей оценки, верно, то оценка состоятельна. Если это предположение нарушается, оцененные коэффициенты для не изменяющихся во времени переменных смещены... Заметьте, что оцененные коэффициенты изменяющихся во времени переменных остаются несмещенными даже при наличии корреляции. Однако предположение, лежащее в основе модели с фиксированными эффектами, должно выполняться (не допускается корреляция изменяющихся во времени переменных)». (Выделение добавлено, вероятно, «изменяющиеся» должны быть «инвариантными».) Есть и другие оценки, которые состоятельно оценивали бы β и γ в этой пересмотренной модели, среди которых оценка Хаусмана–Тейлора, рассматриваемая в подразделе 11.8.1, и оценки методом инструментальных переменных, предложенные Breusch et al. (2010) и Chatelain, Ralf (2010).

Главной интересующей авторов рассматриваемой работы проблемой был промежуточный случай, несколько отличный от того, который мы рассматриваем здесь. Существует два направления работы. Если лишь несколько элементов Z , а не все из них коррелированы с u_i , тогда мы получим ситуацию, проанализированную Хаусманом и Тейлором, которая рассматривается в подразделе 11.8.1. FEVD-оценка, по (Plümper, Troeger), будет в таком случае несостоятельной, у нее может быть меньшая дисперсия, чем у IV-оценки, предложенной Хаусманом и Тейлором. Второй вопрос, который интересует авторов, — это когда Z является не строго не изменяющейся во времени, а «медленно меняющейся». Когда существует крайне небольшая внутригрупповая изменчивость, к примеру, как показано для данных Всемирной организации здравоохранения в примере 11.4, тогда опять же оценка, предложенная здесь, может иметь ряд преимуществ по сравнению с инструментальными переменными и другими методами. В том случае, когда в модели нет строго не изменяющихся во времени переменных, анализ следует проводить, используя модель случайных эффектов, рассмотренную в следующем разделе.

Пример 11.5. Уравнение заработной платы с фиксированными эффектами

В табл. 11.5 представлены оценки уравнения заработной платы с индивидуальными эффектами для данных Корнуэлла и Руперта из примеров 11.1 и 11.3. Модель включает три не изменяющиеся во времени переменные: Ed , Fem , Blk , которые необходимо исключить из уравнения. Как следствие оценки модели с фиксированными эффектами, вычисленные здесь, нельзя сравнить с уже известными результатами по объединенной модели. Для сравнения в таблице также представлены оценки методом наименьших квадратов с устойчивыми панельными стандартными ошибками. Мы

также добавили в модель набор временных фиктивных переменных. F -статистика для проверки значимости индивидуальных эффектов, основанная на R^2 для этого уравнения, равна

$$F[594, 3561] = \frac{(0,9072422 - 0,3154548)/594}{(1 - 0,9072422)/(4165 - 9 - 595)} = 38,247.$$

Критическое значение по F -таблице с 594 и 3561 степенями свободы равно 1,106, поэтому есть явные доказательства в пользу специфического индивидуального эффекта. Как это часто происходит, качество подгонки модели значительно улучшается, когда включаются индивидуальные эффекты. Мы также добавили в модель временные эффекты. Результаты по модели с временными эффектами без индивидуальных эффектов приведены во втором столбце результатов. F -статистика для проверки значимости временных эффектов (при отсутствии индивидуальных) равна

$$F[6, 4149] = \frac{(0,4636788 - 0,3154548)/6}{(1 - 0,4636788)/(4165 - 10 - 6)} = 191,11.$$

Критическое значение из F -таблицы равно 2,101, поэтому гипотеза о том, что временные эффекты нулевые, также отвергается. В последнем столбце приведены оценки модели и с временными, и с индивидуальными эффектами. Для этой модели необходимо исключить второй временной эффект, поскольку переменная опыта работы Exp является индивидуальным специфическим временным трендом. Переменную Exp можно выразить как

$$Exp_{i,t} = E_{i,0} + (t - 1), \quad t = 1, \dots, 7,$$

что также можно выразить через линейную комбинацию индивидуальной фиктивной переменной и шести временных переменных. В последней модели мы можем исключить первый и последний временные эффекты. В этой модели F -статистика для проверки значимости временных эффектов равна

$$F[5, 3556] = \frac{(0,9080847 - 0,9072422)/5}{(1 - 0,9080847)/(4165 - 9 - 5 - 5595)} = 6,519.$$

Временные эффекты остаются значимыми — критическое значение равно 2,217, — однако тестовая статистика значительно уменьшается. Временные эффекты выявляют важную тенденцию. В уравнении без индивидуальных эффектов мы обнаруживаем устойчивое увеличение зарплат на 7–9% в год. Однако при добавлении индивидуальных эффектов в модель этот рост исчезает.

Могло бы показаться удобным вычислить устойчивые стандартные ошибки как для оценки с фиксированными эффектами, так и для объединенной оценки. Однако в принципе это необязательно. Если модель корректна и полностью специфицирована, тогда индивидуальные эффекты должны захватывать пропущенную неоднородность и то, что остается, — это классический гомоскедастичный шок без автокорреляции. Это предполагает грубый тест корректности спецификации модели.

Если стандартная асимптотическая ковариационная матрица в (11-17) и устойчивая оценка в (11-3) при замене X_i на данные в форме отклонений от группового среднего дают очень разные оценки, то можно было бы поставить под сомнение спецификацию модели. [Эта логика лежит в основе теста информационной матрицы Уайта (1982а) (и его развития у Ньюи (Newey (1985а)) и Точена (Tauchen (1985))).] Устойчивые стандартные ошибки представлены в скобках под стандартными ошибками для оценок фиксированных эффектов в шестом столбце табл. 11.5. Они значительно больше, чем некорректированные стандартные ошибки, — 50% к 100%, — это может предполагать, что спецификацию с фиксированными эффектами следует пересмотреть.

Вычисления FEVD также приведены в табл. 11.5. Третий набор результатов, отмеченный как индивидуальные эффекты, показывает результаты оценок на шагах 1 и 2. Обратите внимание, что они посчитаны в двух регрессиях методом наименьших квадратов. Результаты оценок шага 3 показаны в двух последних столбцах таблицы. Как и ожидалось, оцененные коэффициенты совпадают с регрессиями, оцененными на первом и втором шагах. Для b_{LSDV} стандартные ошибки снизились в 2 раза — до 4. Для c_{LSDV} , оценок γ , они снизились в 7 раз — до 10. В свете предыдущих результатов оценки в последнем столбце табл. 11.5 рассматривались бы как достаточно оптимистичные.

11.5. Случайные эффекты

Модель с фиксированными эффектами допускает наличие корреляции ненаблюдаемых индивидуальных эффектов с включенными в модель переменными. Затем мы смоделировали различия между элементами строго как параметрические сдвиги регрессионной функции. Эту модель можно рассматривать как применимую только к пространственным элементам в исследовании, а не к дополнительным вне выборки. Например, межстрановое сравнение может включать полное множество стран, для которых разумно предположить, что модель постоянна. Если индивидуальные эффекты строго не коррелированы с регрессорами, то может быть удобно смоделировать индивидуальные специфические постоянные члены как случайно распределенные по межобъектным элементам. Этот подход будет необходим, если мы допускаем, что выборочные межобъектные данные были получены из большой генеральной совокупности. Так, несомненно, было бы для лонгитюдных данных, примеры которых приводятся во вводной части этой главы¹¹. Преимущество такой формы в том, что она значительно сокращает число параметров для оценки. Недостатком же является вероятность получения несостоятельных оценок, если предположение оказывается необоснованным.

¹¹Это различие не является жестким и прочным, оно носит скорее эвристический характер. Мы вернемся к этому вопросу позже. См. у Мундлака (Mundlak (1978)) методологическое рассмотрение различий между фиксированными и случайными эффектами.

Таблица 11.5. Оценки фиксированных эффектов для уравнения заработной платы Корнуэлла–Руперта

| Пере- менная | Объединенные | | Временные эффекты | | Индивидуальные эффекты | | Врем. и инд. эффекты | | FEVD шаг 3 | |
|------------------------|--------------|-------------------|-------------------|-------------------|------------------------|----------------------------------|----------------------|------------------|------------|------------------|
| | Оценка | Станд. ошибка* | Оценка | Станд. ошибка* | Оценка | Станд. ошибка (устойчивая) | Оценка | Станд. ошибка | Оценка | Станд. ошибка |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| <i>Constant</i> | 5,8802 | 0,09654 | 5,6963 | 0,09425 | | | | | | |
| <i>Exp</i> | 0,03611 | 0,0045241 | 0,02738 | 0,004556 | 0,1132 | 0,002471 (0,00437) | 0,1114 | 0,002618 | 0,1132 | 0,00100 |
| <i>Exp²</i> | -0,00066 | 0,0001013 | -0,00053 | 0,000101 | -0,00042 | 0,000055 (0,000089) | -0,00004 | 0,000054 | -0,00042 | 0,0000192 |
| <i>Wks</i> | 0,00446 | 0,001725 | 0,00409 | 0,001694 | 0,00084 | 0,000600 (0,00094) | 0,00068 | 0,0005991 | 0,00084 | 0,00044 |
| <i>Occ</i> | -0,3176 | 0,02721 | -0,3045 | 0,02684 | -0,02148 | 0,01378 (0,02052) | -0,01916 | 0,01275 | -0,02148 | 0,00596 |
| <i>Ind</i> | 0,03213 | 0,02521 | 0,04010 | 0,02489 | 0,01921 | 0,01545 (0,02450) | 0,02076 | 0,1540 | 0,01921 | 0,00476 |
| <i>South</i> | -0,1137 | 0,028626 | -0,1157 | 0,02834 | -0,00186 | 0,03430 (0,09646) | 0,00309 | 0,03419 | -0,00186 | 0,00506 |
| <i>SMSA</i> | 0,1586 | 0,025967 | 0,1722 | 0,02566 | -0,04247 | 0,01942 (0,03185) | -0,04188 | 0,01937 | -0,04247 | 0,00504 |
| <i>MS</i> | 0,3203 | 0,03487 | 0,3425 | 0,03459 | -0,02973 | 0,01898 (0,02902) | -0,02856 | 0,018918 | -0,02973 | 0,00831 |
| <i>Union</i> | 0,06975 | 0,026618 | 0,06272 | 0,02578 | 0,03278 | 0,01492 (0,02708) | 0,02952 | 0,01488 | 0,03278 | 0,00517 |
| <i>Constant</i> | | | | | 2,8286 | 0,18599 | | | 2,8286 | 0,03315 |
| <i>Fem</i> | | | | | -0,13003 | 0,12557 | | | -0,13003 | 0,01024 |
| <i>Ed</i> | | | | | 0,14438 | 0,01403 | | | 0,14438 | 0,00121 |
| <i>Blk</i> | | | | | -0,27507 | 0,15440,15 | | | -0,14438 | 0,00891 |

Окончание табл. 11.5

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|----------|-----------|---|-----------|----------|-----------|---|-----------|----------|-----------|---------|
| h_i | | | | | | | | | 1,00000 | 0,00683 |
| Год 1 | | | 0,0000 | 0,0000 | | | 0,0000 | 0,0000 | | |
| Год 2 | | | 0,07812 | 0,006860 | | | -0,00775 | 0,008167 | | |
| Год 3 | | | 0,2050 | 0,01072 | | | 0,02557 | 0,007769 | | |
| Год 4 | | | 0,2926 | 0,01125 | | | 0,02845 | 0,007639 | | |
| Год 5 | | | 0,3724 | 0,01095 | | | 0,02418 | 0,007772 | | |
| Год 6 | | | 0,4498 | 0,01245 | | | 0,00737 | 0,008161 | | |
| Год 7 | | | 0,5422 | 0,013015 | | | 0,0000 | 0,0000 | | |
| $e'e$ | 607,1265 | | 475,6659 | | 82,26732 | | 81,52012 | | 82,26732 | |
| Deg,Free | 4155 | | 4149 | | 3561 | | 3557 | | 4151 | |
| s | 0,3822588 | | 0,3385940 | | 0,1519944 | | 0,1514089 | | 0,1407788 | |
| R^2 | 0,3154548 | | 0,4636788 | | 0,9072422 | | 0,9080847 | | 0,9072422 | |

* Устойчивые стандартные ошибки, использующие (11-3), включая коррекцию на конечную генеральную совокупность

 $[(\sum_i T_i) - 1] / [(\sum_i T_i) - K - m] \times n / (n - 1)$.

Теперь рассмотрим переформулированную модель

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + (\alpha + u_i) + \varepsilon_{it}, \quad (11-28)$$

где есть K регрессоров, включая константу, и теперь единственный постоянный член — это среднее значение ненаблюдаемой неоднородности $E[z'_i\alpha]$. Компонента u_i — случайная неоднородность (гетерогенность), характерная для i -го наблюдения и постоянная во времени; из подраздела 11.2.1 вспомним, что $u_i = \{z'_i\alpha - E[z'_i\alpha]\}$. Например, при анализе семей мы можем рассматривать u_i как набор факторов $z'_i\alpha$ не из регрессии, характерной для данной семьи. Мы продолжаем допускать строгую экзогенность:

$$\begin{aligned} E[\varepsilon_{it} | \mathbf{X}] &= E[u_i | \mathbf{X}] = 0, \\ E[\varepsilon_{it}^2 | \mathbf{X}] &= \sigma_\varepsilon^2, \\ E[u_i^2 | \mathbf{X}] &= \sigma_u^2, \\ E[\varepsilon_{it}u_j | \mathbf{X}] &= 0 \quad \text{для всех } i, t \text{ и } j, \\ E[\varepsilon_{it}\varepsilon_{js} | \mathbf{X}] &= 0, \quad \text{если } t \neq s \text{ или } i \neq j, \\ E[u_i u_j | \mathbf{X}] &= 0, \quad \text{если } i \neq j. \end{aligned} \quad (11-29)$$

Как и раньше, полезно было бы рассмотреть запись модели по блокам из T наблюдений для группы i , y_i , \mathbf{X}_i , $u_i \mathbf{i}$ и ε_i . Для этих T наблюдений пусть

$$\eta_{it} = \varepsilon_{it} + u_i$$

и

$$\boldsymbol{\eta}_i = [\eta_{i1}, \eta_{i2}, \dots, \eta_{iT}]'$$

При такой форме η_{it} мы получим то, что часто называют **моделью компонент ошибки**. Для данной модели

$$\begin{aligned} E[\eta_{it}^2 | \mathbf{X}] &= \sigma_\varepsilon^2 + \sigma_u^2, \\ E[\eta_{it}\eta_{is} | \mathbf{X}] &= \sigma_u^2, \quad t \neq s \\ E[\eta_{it}\eta_{js} | \mathbf{X}] &= 0 \quad \text{для всех } t \text{ и } s, \text{ если } i \neq j. \end{aligned} \quad (11-30)$$

Для T наблюдений элемента i пусть $\boldsymbol{\Sigma} = E[\boldsymbol{\eta}_i \boldsymbol{\eta}'_i | \mathbf{X}]$. Тогда

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_\varepsilon^2 + \sigma_u^2 & \sigma_u^2 & \sigma_u^2 & \dots & \sigma_u^2 \\ \sigma_u^2 & \sigma_\varepsilon^2 + \sigma_u^2 & \sigma_u^2 & \dots & \sigma_u^2 \\ & \dots & & & \\ \sigma_u^2 & \sigma_u^2 & \sigma_u^2 & \dots & \sigma_\varepsilon^2 + \sigma_u^2 \end{bmatrix} = \sigma_\varepsilon^2 \mathbf{I}_T + \sigma_u^2 \mathbf{i}_T \mathbf{i}'_T, \quad (11-31)$$

где \mathbf{i}_T — вектор-столбец из единиц размерностью $T \times 1$. Поскольку наблюдения i и j независимы, то ковариационная матрица шоков для всех nT наблюдений равна

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} & \mathbf{0} & \dots & \mathbf{0} \\ & & & \vdots & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\Sigma} \end{bmatrix} = \mathbf{I}_n \otimes \boldsymbol{\Sigma}. \quad (11-32)$$

11.5.1. Оценка методом наименьших квадратов

Модель, определенная при помощи (11-28),

$$y_{it} = \alpha + \mathbf{x}_{it}'\beta + u_i + \varepsilon_{it},$$

в предположениях о строгой экзогенности в (11-29) и ковариационной матрице, подробно описанной в (11-31) и (11-32), представляет собой обобщенную регрессионную модель, которая входит в рамки, описанные нами в главе 9. Шоки автокоррелированы, так как наблюдения коррелированы во времени в пределах группы, хотя и не между группами. Здесь применимы все следствия из подраздела 9.2.1. В частности, параметры модели со случайными эффектами могут быть оценены состоятельно, хотя и неэффективно, простым методом наименьших квадратов (МНК). Соответствующая устойчивая асимптотическая ковариационная матрица для МНК-оценки может быть найдена с использованием (11-3).

Есть и другие доступные состоятельные оценки. Используя отклонения от группового среднего, получим

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' \beta + \varepsilon_{it} - \bar{\varepsilon}_i.$$

Это предполагает, что (допуская, что в \mathbf{x}_{it} нет не изменяющихся во времени регрессоров) LSDV-оценка из (11-14) является состоятельной для β . (Заметим, что LSDV-оценка, единственная из четырех оценок, рассмотренных здесь, является устойчивой к тому, будет ли корректная спецификация на самом деле моделью со случайными или фиксированными эффектами.) Как и МНК, LSDV неэффективна, так как существует эффективная GLS-оценка, не равная \mathbf{b}_{LSDV} , что будет показано в подразделе 11.5.2. Регрессионная модель групповых средних (между группами)

$$\bar{y}_i = \alpha + \bar{\mathbf{x}}_i' \beta + u_i + \bar{\varepsilon}_i, i = 1, \dots, n$$

дает третий метод состоятельного оценивания коэффициентов β . В данной ситуации ни одна из них не является более предпочтительной оценкой, поскольку GLS-оценка будет эффективнее, чем любая из них. Однако, как мы видели в главах 9 и 10, многие обобщенные регрессионные модели оцениваются в два шага, причем первый шаг представляет собой устойчивую оценку регрессии методом наименьших квадратов, которая используется для получения первых оценок дисперсий параметров модели. Здесь это также будет иметь место. Чтобы предположить, к чему приведет эта логика в подразделе 11.5.3, заметим, что для трех отмеченных случаев среднеквадратичные остатки давали бы следующие состоятельные оценки функций дисперсий:

$$\begin{aligned} \text{(объединенные)} \quad & \text{plim} [\mathbf{e}_{\text{pooled}}' \mathbf{e}_{\text{pooled}} / (nT)] = \sigma_u^2 + \sigma_\varepsilon^2, \\ \text{(LSDV)} \quad & \text{plim} [\mathbf{e}_{\text{LSDV}}' \mathbf{e}_{\text{LSDV}} / (nT)] = \sigma_\varepsilon^2 [1 - 1/T], \\ \text{(средние)} \quad & \text{plim} [\mathbf{e}_{\text{means}}' \mathbf{e}_{\text{means}} / (nT)] = \sigma_u^2 + \sigma_\varepsilon^2 / T. \end{aligned}$$

Любая пара этих оценок дает метод оценки моментов $(\sigma_u^2, \sigma_\varepsilon^2)$ с двумя уравнениями. Зная это, мы теперь рассмотрим эффективную оценку обобщенным методом наименьших квадратов.

11.5.2. Обобщенный метод наименьших квадратов

Оценка обобщенным методом наименьших квадратов для коэффициентов наклона имеет вид

$$\hat{\beta} = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}\mathbf{y} = \left(\sum_{i=1}^n \mathbf{X}'_i \Sigma^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{X}'_i \Sigma^{-1} \mathbf{y}_i \right).$$

Для вычисления этой оценки, как мы делали в главе 9 при помощи преобразования данных и использования обычного метода наименьших квадратов для преобразованных данных, нам необходимо, чтобы $\Omega^{-1/2} = [\mathbf{I}_n \otimes \Sigma]^{-1/2}$. Нужно лишь найти $\Sigma^{-1/2}$, которая будет равна

$$\Sigma^{-1/2} = \frac{1}{\sigma_\varepsilon} \left[\mathbf{I} - \frac{\theta}{T} \mathbf{i}_T \mathbf{i}'_T \right],$$

где

$$\theta = 1 - \frac{\sigma_\varepsilon}{\sqrt{\sigma_\varepsilon^2 + T\sigma_u^2}}.$$

Отсюда преобразование \mathbf{y}_i и \mathbf{X}_i для GLS будет

$$\Sigma^{-1/2} \mathbf{y}_i = \frac{1}{\sigma_\varepsilon} \begin{bmatrix} y_{i1} - \theta \bar{y}_i \\ y_{i2} - \theta \bar{y}_i \\ \vdots \\ y_{iT} - \theta \bar{y}_i \end{bmatrix}, \quad (11-33)$$

так же и для строк \mathbf{X}_i ¹². Тогда для данных в совокупности оценка обобщенным методом наименьших квадратов вычисляется оценкой регрессии этих частных отклонений y_{it} на те же преобразования x_{it} . Обратите внимание на схожесть этой процедуры с вычислениями в LSDV-модели, в которых используется $\theta = 1$ в (11-15). (Можно было бы трактовать θ как эффект, который останется в случае, если σ_ε будет нулевым, поскольку единственным эффектом в такой ситуации была бы u_i . Тогда модели с фиксированными и случайными эффектами были бы неразличимы, так что этот результат имеет значение.)

Можно показать, что GLS-оценка, как объединенная МНК-оценка, представляет собой средневзвешенное матрицы оценок внутри- и межгрупповых элементов

$$\hat{\beta} = \hat{\mathbf{F}}^{within} \mathbf{b}^{within} + (\mathbf{I} - \hat{\mathbf{F}}^{within}) \mathbf{b}^{between} \text{ }^{13}, \quad (11-34)$$

где теперь

$$\hat{\mathbf{F}}^{within} = [\mathbf{S}_{xx}^{within} + \lambda \mathbf{S}_{xx}^{between}]^{-1} \mathbf{S}_{xx}^{within},$$

$$\lambda = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_u^2} = (1 - \theta)^2.$$

¹²Такое преобразование представляет собой частный случай более общего анализа у Нерлова (Nerlove (1971b)).

¹³Альтернативная форма этого выражения, в которой взвешиваемые матрицы пропорциональны ковариационным матрицам двух оценок, дана у Judge et al. (1985).

В той степени, в какой λ отличается от единицы, мы видим, что неэффективность обычного метода наименьших квадратов будет следовать из неэффективного взвешивания двух оценок. По сравнению с обобщенным методом наименьших квадратов обычный метод придает слишком большой вес межобъектной изменчивости. Он включает всю ее в изменчивость в X , а не соразмерно распределяет некоторые ее части между случайной изменчивостью по группам, относимой к изменчивости в u_i по элементам.

Несбалансированные панели добавляют некоторый налет сложности в модель случайных эффектов. Первую проблему можно увидеть в (11-32). Матрица Ω больше не является $I_n \otimes \Sigma$, поскольку диагональные блоки в Ω разных размеров. Кроме того, существует групповая гетероскедастичность в (11-33), так как i -й диагональный блок в $\Omega^{-1/2}$ имеет вид

$$\Sigma_i^{-1/2} = \frac{1}{\sigma_\varepsilon} \left[I_{T_i} - \frac{\theta_i}{T_i} \mathbf{i}_{T_i} \mathbf{i}'_{T_i} \right], \quad \theta_i = 1 - \frac{\sigma_\varepsilon}{\sqrt{\sigma_\varepsilon^2 + T_i \sigma_u^2}}.$$

В принципе оценка остается прямой, потому что источником групповой гетероскедастичности являются только неравные размеры групп. Таким образом, для GLS или FGLS с оцененными компонентами дисперсии необходимо лишь применить характерный для группы θ_i в преобразовании (11-33).

11.5.3. Допустимый обобщенный метод наименьших квадратов при неизвестной матрице Σ

Если компоненты дисперсии известны, то оценку обобщенным методом наименьших квадратов можно вычислить, как было показано ранее. Конечно, это вряд ли получится, поэтому, как обычно, мы должны в первую очередь оценить дисперсии шоков и затем применить FGLS-процедуру. Эвристическим подходом к оценке компонент дисперсии будет

$$y_{it} = \mathbf{x}'_{it} \beta + \alpha + \varepsilon_{it} + u_i \quad (11-35)$$

и

$$\bar{y}_i = \bar{\mathbf{x}}'_i \beta + \alpha + \bar{\varepsilon}_i + u_i.$$

Следовательно, использование отклонения от групповых средних убирает неоднородность:

$$y_{it} - \bar{y}_i = [\mathbf{x}_{it} - \bar{\mathbf{x}}_i]' \beta + [\varepsilon_{it} - \bar{\varepsilon}_i]. \quad (11-36)$$

Поскольку

$$E \left[\sum_{t=1}^T (\varepsilon_{it} - \bar{\varepsilon}_i)^2 \right] = (T-1) \sigma_\varepsilon^2,$$

если β был бы наблюдаемым, тогда несмещенная оценка для σ_ε^2 , основанная на T наблюдениях в группе i , была бы равна

$$\hat{\sigma}_\varepsilon^2(i) = \frac{\sum_{t=1}^T (\varepsilon_{it} - \bar{\varepsilon}_i)^2}{T-1}. \quad (11-37)$$

Так как β должен быть оценен — (11-33) предполагает, что LSDV-оценка состоятельная, а на самом деле вообще несмещенная, — мы корректируем на

степени свободы и используем остатки LSDV в

$$s_e^2(i) = \frac{\sum_{t=1}^T (e_{it} - \bar{e}_i)^2}{T - K - 1}. \quad (11-38)$$

(Обратите внимание, что основанное на LSDV-оценках среднее \bar{e}_i на самом деле равно нулю. Мы доведем выкладки до конца, чтобы все же сохранить аналогию с (11-34), где \bar{e}_i является не нулем, но все равно остается оценкой $E[\varepsilon_{it}] = 0$.) Мы имеем n таких оценок, так что усредняем их, чтобы получить

$$\bar{s}_e^2 = \frac{1}{n} \sum_{i=1}^n s_e^2(i) = \frac{1}{n} \sum_{i=1}^n \left[\frac{\sum_{t=1}^T (e_{it} - \bar{e}_i)^2}{T - K - 1} \right] = \frac{\sum_{i=1}^n \sum_{t=1}^T (e_{it} - \bar{e}_i)^2}{nT - nK - n}. \quad (11-39)$$

Корректировка на степени свободы в \bar{s}_e^2 избыточна, так как она предполагает, что α и β переоцениваются для каждого i . Оцененные параметры представляют собой n средних значений \bar{y}_i и K коэффициентов наклона. Таким образом, мы предлагаем несмещенную оценку¹⁴

$$\hat{\sigma}_\varepsilon^2 = s_{LSDV}^2 = \frac{\sum_{i=1}^n \sum_{t=1}^T (e_{it} - \bar{e}_i)^2}{nT - n - K}. \quad (11-40)$$

Это есть оценка дисперсии в модели с фиксированными эффектами в (11-18), надлежащим образом скорректированная на степени свободы. Остается оценить σ_u^2 . Вернемся к спецификации исходной модели в (11-35). Несмотря на корректировку по наблюдениям, это классическая регрессионная модель, в которой и коэффициенты наклона, и оценки дисперсии, оцененные обычным методом наименьших квадратов, состоятельны и в большинстве случаев не смещены. Таким образом, используя остатки обычного метода наименьших квадратов из модели с единственной общей константой, имеем

$$\text{plim} s_{Pooled}^2 = \text{plim} \frac{\mathbf{e}'\mathbf{e}}{nT - K - 1} = \sigma_\varepsilon^2 + \sigma_u^2. \quad (11-41)$$

Это дает две оценки, необходимые для компонент дисперсии; второй была бы $\hat{\sigma}_u^2 = s_{Pooled}^2 - s_{LSDV}^2$. Возможным осложнением является то, что эта вторая оценка может быть отрицательной. Но вспомним, что для допустимого обобщенного метода наименьших квадратов нам не требуется несмещенная оценка дисперсии, а только состоятельная. И поэтому мы можем опустить корректировку на степени свободы в (11-40) и (11-41). Если так, то обе оценки дисперсии должны быть неотрицательными, так как сумма квадратов в модели LSDV не может быть больше, чем в простой регрессии только с одним постоянным членом. Предлагались и альтернативные оценки, все они были основаны на этом принципе использования двух различных сумм квадратов остатков¹⁵. Это тот момент, в котором современные компьютерные программы сильно различаются. Как правило, оценку компонент дисперсии программы начинают с выражений (11-40) и (11-41). Если

¹⁴Формальное доказательство этого предложения можно найти у Maddala (1971) или Judge et al. (1985, p. 551).

¹⁵Wallace, Hussain (1969), Maddala (1971), Fuller, Battese (1974) и Amemiya (1971).

оценка σ_u^2 неположительная, их дальнейшие действия весьма разнообразны. Исключение корректировки на число степеней свободы — часто используемая стратегия, однако как минимум одна из широко распространенных программ просто считает, что σ_u^2 равна нулю, а другие прибегают к различным стратегиям, основанным, например, на оценке по групповым средним. Неблагоприятным следствием для доверчивого исследователя является то, что разные программы могут систематически давать различные результаты при использовании одной и той же модели и одних и тех же данных. Поэтому исследователю-практику настоятельно рекомендуется изучить программные документы для решения этой проблемы.

Остается еще одна сложность. Если есть какие-либо регрессоры, не изменяющиеся внутри группы, LSDV-оценка не может быть вычислена. Например, в модели семейного дохода или предложения труда один из регрессоров мог бы быть фиктивной переменной местоположения, состава семьи или жилищных условий. Любая из них может быть полностью коллинеарна с фиксированным эффектом для этой семьи, что препятствует вычислению LSDV-оценки. В этом случае все же возможно оценить компоненты дисперсии случайных эффектов. Пусть $[b, a]$ будет любой состоятельной оценкой $[\beta, \alpha]$ в (11-35), такой как оценка обычным методом наименьших квадратов. Тогда (11-41) позволяет получить состоятельную оценку $m_{ee} = \sigma_\varepsilon^2 + \sigma_u^2$. Среднее квадратов остатков регрессии, основанной только на n групповых средних значениях в (11-35), дает состоятельную оценку $m_{**} = \sigma_u^2 + (\sigma_\varepsilon^2/T)$, так что мы можем использовать

$$\hat{\sigma}_\varepsilon^2 = \frac{T}{T-1}(m_{ee} - m_{**}),$$

$$\hat{\sigma}_u^2 = \frac{T}{T-1}m_{**} - \frac{1}{T-1}m_{ee} = \omega m_{**} + (1 - \omega)m_{ee},$$

где $\omega > 1$. Как и раньше, этот способ оценивания может давать отрицательную оценку σ_u^2 , что опять же ставит под вопрос спецификацию модели. [Обратите внимание, что остатки в (11-40) и (11-41) могут быть основаны на одном и том же векторе коэффициентов.]

Как это ни удивительно, существует более простой выход из дилеммы, поставленной не изменяющимися во времени переменными. В (11-36) мы обнаружили, что оценка по отклонениям от групповых средних еще дает состоятельную оценку σ_ε^2 . Не изменяющиеся во времени переменные выпадают из модели, поэтому оценить полный вектор коэффициентов β невозможно. Однако вспомним, что оценка β не является целью в данный момент, мы ищем оценку σ_ε^2 . Отсюда следует, что остатки оценки по отклонениям групповых средних (LSDV) могут также использоваться для оценки σ_ε^2 . По той же логике можно применить и первые разности. (См. подраздел 11.3.5.) При помощи дисперсии остатков в регрессии в первых разностях можно было бы оценить $2\sigma_\varepsilon^2$. Эти результаты не зависят от того, есть ли в данной модели не изменяющиеся во времени регрессоры.

11.5.4. Проверка гипотез для случайных эффектов

Бройш и Пэган (Breusch, Pagan (1980)) разработали **тест множителей Лагранжа** для модели со случайными эффектами, основанной на МНК-ос-

татках¹⁶. Для

$$\begin{aligned} H_0: \sigma_u^2 &= 0 \quad (\text{или } \text{Corr}\{\eta_{it}, \eta_{is}\} = 0), \\ H_1: \sigma_u^2 &\neq 0 \end{aligned}$$

тестовая статистика равна

$$\text{LM} = \frac{nT}{2(T-1)} \left[\frac{\sum_{i=1}^n \left[\sum_{t=1}^T e_{it} \right]^2}{\sum_{i=1}^n \sum_{t=1}^T e_{it}^2} - 1 \right]^2 = \frac{nT}{2(T-1)} \left[\frac{\sum_{i=1}^n (T \bar{e}_i)^2}{\sum_{i=1}^n \sum_{t=1}^T e_{it}^2} - 1 \right]^2. \quad (11-42)$$

В условиях нулевой гипотезы предельным распределением множителей Лагранжа является хи-квадрат с одной степенью свободы.

Пример 11.6. Проверка гипотез о случайных эффектах

Нас интересует сравнение оценок с фиксированными и случайными эффектами в уравнении заработной платы Корнуэлла–Руперта. Как мы видели ранее, в уравнении есть три не изменяющиеся во времени переменные: Ed , Fem и Blk . Поэтому мы не можем непосредственно сравнить две оценки. Модель со случайными эффектами может давать отдельные оценки параметров по не изменяющимся во времени переменным, в то время как оценка по фиксированным эффектам — нет. Чтобы это продемонстрировать, мы на время сосредоточим внимание на ограниченной модели общих эффектов

$$\ln Wage_{it} = \beta_1 Exp_{it} + \beta_2 Exp_{it}^2 + \beta_3 Wks_{it} + \beta_4 Occ_{it} + \beta_5 Ind_{it} + \beta_6 South_{it} + \beta_7 SMSA_{it} + \beta_8 MS_{it} + \beta_9 Union_{it} + c_i + \varepsilon_{it}.$$

Модели с фиксированными и случайными эффектами различаются в трактовке c_i .

Оценки параметров методом наименьших квадратов, включая постоянный член, представлены в табл. 11.6. Затем мы вычислили остатки групповых средних для семи наблюдений для каждого агента. Сумма квадратов средних значений равна 53,824384. Общая сумма квадратов остатков для регрессии составила 607,1265. При T и n , равных 7 и 595 соответственно, (11-42) дает статистику хи-квадрат, равную 3881,34. Это намного превышает 95%-е критическое значение для распределения хи-квадрат с одной степенью свободы, 3,84. На данном этапе мы заключаем, что классическая регрессионная модель с единственным постоянным членом не подходит для этих данных. Результатом теста является отклонение нулевой гипотезы в пользу модели случайных эффектов. Однако лучше не спешить с таким суждением, поскольку существует другая, конкурирующая спецификация, которая могла бы дать те же результаты,

¹⁶До сих пор мы сосредотачивались на обобщенном методе наименьших квадратов и моментах, основанных на состоятельных оценках компонент дисперсии. В отличие от них тест множителей Лагранжа основан на оценке методом максимального правдоподобия. См. об этом подходе к оцениванию у Maddala (1971) и Balestra, Nerlove (1966, 2003).

что и модель со случайными эффектами. Мы рассмотрим эту возможность в последующих примерах.

Имея оценки дисперсии, можно использовать FGLS для оценки параметров модели. Все полученные ранее результаты по FGLS-оценке применимы и здесь. В частности, все, что нужно для эффективного оценивания параметров модели, — это состоятельные оценки компонент дисперсии, и здесь их несколько. [См. Hsiao (2003), Baltagi (2005), Nerlove (2002), Berzeg (1979) и Maddala, Mount (1973).]

Пример 11.7. Оценки модели со случайными эффектами

В предыдущем примере мы нашли, что общая сумма квадратов остатков оценки методом наименьших квадратов равна 607,1265. Оценки модели с фиксированными эффектами (LSDV) для этой модели представлены в табл. 11.5 (и 11.6), где сумма квадратов равна 82,26732. Отсюда моментные оценки параметров дисперсии составят

$$\hat{\sigma}_\varepsilon^2 + \hat{\sigma}_u^2 = \frac{607,1265}{4165 - 10} = 0,1461195$$

и

$$\hat{\sigma}_\varepsilon^2 = \frac{82,26732}{4165 - 595 - 9} = 0,0231023.$$

Вытекающая из этого оценка σ_u^2 равна 0,12301719. (Проблемы отрицательных компонент дисперсии не возникло.) Оценка θ для FGLS равна

$$\hat{\theta} = 1 - \sqrt{\frac{0,0231023}{0,0231023 + 7(0,12301719)}} = 0,8383608.$$

FGLS-оценки посчитаны через построение регрессии для частичных разностей $\ln Wage_{it}$ на частичные разности константы и девять регрессоров с использованием этой оценки θ в (11-33). Оценки параметров при помощи МНК, оценки с фиксированными и случайными эффектами приведены в табл. 11.6.

Ни одно из необходимых свойств оценок в модели со случайными эффектами не основывается на T , стремящемся к бесконечности¹⁷. На самом деле вероятнее всего, что T будет достаточно мало. Оценка σ_ε^2 равна среднему по n оценкам, каждая из которых основана на T наблюдениях для элемента i . [См. (11-39).] Каждая компонента в этом среднем значении является в принципе состоятельной. Так что ее дисперсия порядка $1/T$ или меньше. Поскольку T мало, эта дисперсия может быть относительно большой. Однако каждый член дает некоторую информацию о параметре. Среднее значение по n пространственным элементам имеет дисперсию порядка

¹⁷Nickell (1981).

Таблица 11.6. Оценки уравнения заработной платы

| Переменная | Объединенная МНК-регрессия | | LSDV с фиксированными эффектами | | FGLS со случайными эффектами | |
|-------------------------|----------------------------|----------------------------|---------------------------------|---------------|------------------------------|---------------|
| | Оценка | Станд. ошибка ^a | Оценка | Станд. ошибка | Оценка | Станд. ошибка |
| <i>Exp</i> | 0,0361 | 0,004533 | 0,1132 | 0,002471 | 0,08906 | 0,002280 |
| <i>Exp</i> ² | -0,0006550 | 0,0001016 | -0,0004184 | 0,0000546 | -0,0007577 | 0,00005036 |
| <i>Wks</i> | 0,004461 | 0,001728 | 0,0008359 | 0,0005997 | 0,001066 | 0,0005939 |
| <i>Occ</i> | -0,3176 | 0,02726 | -0,02148 | 0,01378 | -0,1067 | 0,01269 |
| <i>Ind</i> | 0,03213 | 0,02526 | 0,01921 | 0,01545 | -0,01637 | 0,01391 |
| <i>South</i> | -0,1137 | 0,02868 | 0,001861 | 0,03430 | -0,06899 | 0,02354 |
| <i>SMSA</i> | 0,1586 | 0,02602 | -0,04247 | 0,01943 | -0,01530 | 0,01649 |
| <i>MS</i> | 0,3203 | 0,03494 | -0,02973 | 0,01898 | -0,02398 | 0,01711 |
| <i>Union</i> | 0,06975 | 0,02667 | 0,03278 | 0,01492 | 0,03597 | 0,01367 |
| <i>Constant</i> | 5,8802 | 0,09673 | | | 5,3455 | 0,04361 |

| | Мундлак: групповые средние | | Мундлак: изменяющиеся во времени | |
|-------------------------|----------------------------|---------------|----------------------------------|---------------|
| | Оценка | Станд. ошибка | Оценка | Станд. ошибка |
| <i>Exp</i> | -0,08574 | 0,005821 | 0,1132 | 0,002474 |
| <i>Exp</i> ² | -0,0001168 | 0,0001281 | -0,0004184 | 0,00005467 |
| <i>Wks</i> | 0,008020 | 0,004006 | 0,0008359 | 0,0006004 |
| <i>Occ</i> | -0,3321 | 0,03363 | -0,02148 | 0,01380 |
| <i>Ind</i> | 0,02677 | 0,03203 | 0,01921 | 0,01547 |
| <i>South</i> | -0,1064 | 0,04444 | -0,001861 | 0,03434 |
| <i>SMSA</i> | 0,2239 | 0,03421 | 0,04247 | 0,01945 |
| <i>MS</i> | 0,4134 | 0,03984 | -0,02972 | 0,01901 |
| <i>Union</i> | 0,05637 | 0,03549 | 0,03278 | 0,01494 |
| <i>Constant</i> | | | 5,7222 | 0,1906 |

^a Стандартные робастные оценки.

$1/(nT)$, что будет стремиться к нулю, если n возрастет, даже если мы будем считать T фиксированным. Вывод состоит в том, что ничего в этом анализе не будет зависеть от увеличивающегося T . Хотя можно показать, что некоторые результаты о состоятельности являются следствием возрастающего T , стандартная панель данных основывается на данных, для которых не имеет смысла допускать, что T неограниченно возрастает или, в некоторых случаях, что оно вообще возрастает¹⁸. Общее предложение состоит в том, что необходимо проявлять осторожность в разработке оценок, свойства которых зависят от того, большое T или нет. Наиболее широко распространенные оценки, которые обсуждались здесь, не зависят от этого, но мы не исключили все возможности.

Модель со случайными эффектами была представлена в работе (Balestra, Nerlove (1966)). Их формулировка включала как временной, κ_t , так и индивидуальный эффекты:

$$y_{it} = \alpha + \beta'x_{it} + \varepsilon_{it} + u_i + \kappa_t.$$

Расширенная формулировка является аналитически более сложной. В исследовании (Balestra, Nerlove (1966)) она была еще больше осложнена из-за включения в модель запаздывающей зависимой переменной. Было рассмотрено полное множество результатов для этой расширенной модели, включая метод работы с запаздывающей зависимой переменной¹⁹. Мы вернемся к этому в разделе 11.8.

11.5.5. Тест на спецификацию Хаусмана для модели со случайными эффектами

По разным причинам мы различали модели с фиксированными и случайными эффектами. Неизбежен вопрос: какую из моделей следует использовать? С чисто практической точки зрения метод фиктивной переменной является «дорогостоящим» из-за потери числа степеней свободы. Вместе с тем метод с фиксированными эффектами имеет одно существенное преимущество. Существует мало оснований для анализа индивидуальных эффектов как не коррелированных с остальными регрессорами, что предполагается в модели со случайными эффектами. Модель со случайными эффектами, таким образом, может пострадать от несостоятельности по причине этой корреляции между включенными переменными и случайным эффектом²⁰.

Тест на спецификацию, разработанный Хаусманом (Hausman, 1978)²¹, используется для проверки на ортогональность обычных эффектов и регрессоров. Тест основан на идее о том, что при гипотезе об отсутствии корреля-

¹⁸В связи с этим Chamberlain (1984) предлагает некоторые передовые методы исследования панельных данных, которые фактически берут T , как указано в модели, и которые основывают результаты о состоятельности исключительно на возрастании n . Некоторые дополнительные результаты для динамических моделей приведены у Bhargava, Sargan (1983).

¹⁹Balestra, Nerlove (1966), Fomby, Hill, Johnson (1984), Judge et al. (1985), Hsiao (1986), Anderson, Hsiao (1982), Nerlove (1971a, 2002) и Baltagi (2008).

²⁰Hausman, Taylor (1981) и Chamberlain (1978).

²¹Связанные с этим результаты приведены у Baltagi (1986).

ции LSDV- и FGLS-оценки состоятельны, но оценка по МНК неэффективна²², в то время как при альтернативной гипотезе LSDV состоятельна, а FGLS нет. Таким образом, в условиях нулевой гипотезы эти две оценки не должны систематически различаться, а тест может быть основан на разности. Другим важным элементом для теста является ковариационная матрица вектора разности $[\mathbf{b} - \hat{\beta}]$:

$$\text{Var}[\mathbf{b} - \hat{\beta}] = \text{Var}[\mathbf{b}] + \text{Var}[\hat{\beta}] - \text{Cov}[\mathbf{b}, \hat{\beta}] - \text{Cov}[\hat{\beta}, \mathbf{b}]. \quad (11-43)$$

Важный вывод Хаусмана состоит в том, что ковариация эффективной оценки и разность между эффективной и неэффективной оценками равна нулю, что предполагает

$$\text{Cov}[(\mathbf{b} - \hat{\beta}), \hat{\beta}] = \text{Cov}[\mathbf{b}, \hat{\beta}] - \text{Var}[\hat{\beta}] = \mathbf{0}$$

или

$$\text{Cov}[\mathbf{b}, \hat{\beta}] = \text{Var}[\hat{\beta}].$$

Применение этого результата в (11-43) дает искомую ковариационную матрицу

$$\text{Var}[\mathbf{b} - \hat{\beta}] = \text{Var}[\mathbf{b}] - \{\text{Var}[\hat{\beta}]\} = \Psi.$$

Хи-квадрат-тест основан на критерии Вальда:

$$W = \chi^2[K - 1] = [\mathbf{b} - \hat{\beta}]' \hat{\Psi}^{-1} [\mathbf{b} - \hat{\beta}]. \quad (11-44)$$

Для $\hat{\Psi}$ мы используем оцененные ковариационные матрицы оценок коэффициентов наклона в LSDV-модели и оцененную ковариационную матрицу в модели случайных эффектов без учета постоянного члена. В условиях нулевой гипотезы W имеет предельное распределение хи-квадрат с $K - 1$ степенями свободы.

Тест Хаусмана — полезный инструмент для определения предпочтительной спецификации модели общих эффектов. Как отмечалось, у него есть один практический недостаток. Построение в (11-43) соответствует теории. Однако нет гарантии, что разность двух ковариационных матриц будет положительно определенной в конечной выборке. Вследствие этого ничто не мешает статистике быть отрицательной, если вычислять ее в соответствии с (11-44). В связи с этим можно было бы заключить, что модель со случайными эффектами не отвергается, поскольку причиной проблемы является схожесть ковариационных матриц, а при альтернативной (с фиксированными эффектами) гипотезе они были бы очень разными. Однако существует несколько альтернативных методов вычисления статистики для теста Хаусмана, одни из которых эквивалентны асимптотически, а другие тождественны по числовым показателям. Обширный анализ представлен в книге Балтаги (Baltagi (2005, p. 65–73)). Одна крайне удобная форма этого теста позволяет обойти отмеченную здесь практическую проблему. Асимптотически эквивалентная тестовая статистика задается через

$$H' = (\hat{\beta}_{LSDV} - \hat{\beta}_{MEANS})' \left[\text{Asy. Var}[\hat{\beta}_{LSDV}] + \text{Asy. Var}[\hat{\beta}_{MEANS}] \right]^{-1} \times \\ \times (\hat{\beta}_{LSDV} - \hat{\beta}_{MEANS}), \quad (11-45)$$

²²При обращении к матричному взвешенному среднему FGLS, рассмотренному ранее, мы видим, что для эффективной меры веса используется θ , в то время как LSDV устанавливает $\theta = 1$.

где $\hat{\beta}_{MEANS}$ — оценка группового среднего, рассмотренная в подразделе 11.3.4. Как уже отмечалось, это одна из нескольких эквивалентных форм теста. Преимущество этой формы в том, что ковариационная матрица всегда будет неотрицательно определенной.

Пример 11.8. *Тест Хаусмана на спецификацию: фиксированные эффекты против случайных*

Используя результаты из предшествующего примера, мы имеем вектор коэффициентов и оцененную асимптотическую ковариационную матрицу \mathbf{b}_{FE} и \mathbf{V}_{FE} из результатов оценки модели с фиксированными эффектами и первых девяти элементов из $\hat{\beta}_{RE}$ и \mathbf{V}_{RE} (без учета постоянного члена). Тестовая статистика равна

$$H = (\mathbf{b}_{FE} - \hat{\beta}_{RE})' [\mathbf{V}_{FE} - \mathbf{V}_{RE}]^{-1} (\mathbf{b}_{FE} - \hat{\beta}_{RE}).$$

Значение тестовой статистики равно 2636,08. Критическое значение по таблице распределения хи-квадрат равно 16,919, поэтому нулевая гипотеза о модели со случайными эффектами отвергается. Мы заключаем, что модель с фиксированными эффектами является более предпочтительной спецификацией для рассматриваемых данных. Это неудачный оборот событий, поскольку главным объектом исследования здесь является влияние образования, что в данной выборке представляет собой не изменяющуюся во времени переменную. Используя вместо этого (11-42), мы получим тестовую статистику, равную 3177,58. Конечно, это не меняет наш вывод.

Имбенс и Вулдридж (Imbens, Wooldridge (2007)) утверждают, что, несмотря на практические соображения о тесте Хаусмана в (11-44) и (11-45), тест должен основываться на устойчивых оценках ковариационных матриц, которые не зависят от предположения нулевой гипотезы (модели случайных эффектов). (То есть «нет смысла представлять полностью устойчивую дисперсионную матрицу для фиксированных и случайных эффектов, а затем вычислять тест Хаусмана, который содержит полный набор допущений для случайных эффектов».) Предложенным ими подходом является тест сложения переменных с устойчивой ковариационной матрицей, описанный в следующем разделе.

11.5.6. Расширение модели с ненаблюдаемыми эффектами: подход Мундлака

Даже при наличии доступного теста Хаусмана выбор между спецификациями с фиксированными и случайными эффектами представляет собой дилемму. У обеих спецификаций есть недостатки. Модель с фиксированными эффектами устойчива к корреляции между пропущенной неоднородностью и регрессорами, но быстро разрастается по параметрам и не может быть приспособлена к не изменяющимся во времени регрессорам. Модель со случайными эффектами зависит от неправдоподобного допущения о том, что пропущенная гетерогенность не коррелирована с регрессорами.

Некоторыми авторами были предложены модификации модели со случайными эффектами, которые как минимум частично преодолевали бы этот недочет. Несостоятельность модели со случайными эффектами в том, что предположение о независимости средних значений, $E[c_i | \mathbf{X}_i] = 0$, необоснованно. В рамках подхода Мундлака (Mundlak, 1978) предлагается следующая спецификация²³:

$$E[c_i | \mathbf{X}_i] = \bar{\mathbf{x}}_i' \boldsymbol{\gamma}.$$

После подстановки этого выражения в модель со случайными эффектами получим

$$\begin{aligned} y_{it} &= \mathbf{x}_{it}' \boldsymbol{\beta} + c_i + \varepsilon_{it} \\ &= \mathbf{x}_{it}' \boldsymbol{\beta} + \bar{\mathbf{x}}_i' \boldsymbol{\gamma} + \varepsilon_{it} + (c_i - E[c_i | \mathbf{X}_i]) \\ &= \mathbf{x}_{it}' \boldsymbol{\beta} + \bar{\mathbf{x}}_i' \boldsymbol{\gamma} + \varepsilon_{it} + u_i. \end{aligned} \quad (11-46)$$

Это сохраняет спецификацию модели со случайными эффектами, однако (как можно было бы надеяться) напрямую касается проблемы корреляции между эффектами и регрессорами. Заметим, что дополнительные члены в $\bar{\mathbf{x}}_i' \boldsymbol{\gamma}$ будут включать лишь изменяющиеся во времени переменные — не изменяющиеся во времени переменные уже являются групповыми средними. Этот дополнительный набор оценок приведен в нижней части табл. 11.6 примера 11.6.

Подход Мундлака часто используется как компромисс между моделями с фиксированными и случайными эффектами. Преимуществом такой спецификации является то, что она дает другой, более удобный подход к тесту Хаусмана. Как следует из записи модели выше, разница между моделью с фиксированными эффектами и моделью со случайными эффектами заключается в ненулевом $\boldsymbol{\gamma}$. Поэтому статистическая проверка нулевой гипотезы о том, что $\boldsymbol{\gamma}$ равен нулю, должна обеспечивать альтернативный подход к двум методам, предложенным ранее.

Пример 11.9. Включение дополнительной переменной для тестирования фиксированных случайных эффектов

Используя результаты из примера 11.6, мы восстановили подвектор оценок в нижней половине табл. 11.6 в соответствии с $\boldsymbol{\gamma}$ и подматрицу полной ковариационной матрицы. (Нижняя панель в табл. 11.6 оценивалась отдельно.) Тестовая статистика равна

$$H' = \hat{\boldsymbol{\gamma}}' [\text{Est. Asy. Var}(\hat{\boldsymbol{\gamma}})]^{-1} \hat{\boldsymbol{\gamma}}.$$

²³ Другие исследования, например Chamberlain (1982) и Wooldridge (2002a), трактуют линейную функцию как проекцию c_i на групповые средние, а не как условное среднее. Разница состоит в том, что нам не нужно делать какие-либо особые предположения о функции условного среднего, так как всегда существует линейная проекция. Трактровка условного среднего налагает на модель необходимость дополнительного допущения, но при этом вносит и значительное упрощение. Несколько авторов проанализировали расширение модели в сторону проекции на полное множество индивидуальных наблюдений, а не на средние. Дополнительное обобщение закладывает основы для некоторых других оценок, включая оценку, основанную на минимизации расстояния [Chamberlain (1982)], GMM [Arellano, Bover (1995)], оценку ограниченных внешне не связанных уравнений и оценку трехшаговым методом наименьших квадратов [Wooldridge (2002a)].

Значение тестовой статистики равно 3193,69. Критическое значение по таблице распределения хи-квадрат с девятью степенями свободы равно 16,919, поэтому нулевая гипотеза о модели со случайными эффектами отвергается. Как и раньше, мы заключаем, что оценка с фиксированными эффектами является более предпочтительной спецификацией для данной модели.

11.5.7. Расширение моделей с фиксированными и случайными эффектами: подход Чемберлена

Модель ненаблюдаемых линейных эффектов имеет вид

$$y_{it} = c_i + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}. \quad (11-47)$$

Модель **со случайными эффектами** предполагает, что $E[c_i | \mathbf{X}_i] = \alpha$, где $T - \mathbf{X}_i$ строк из \mathbf{x}'_{it} . Как мы видели в подразделе 11.5.1, эту модель можно состоятельно оценить обычным методом наименьших квадратов. Независимо от того, как смоделированы ε_{it} , существует автокорреляция, вызванная обычным ненаблюдаемым c_i , поэтому используется обобщенная регрессионная модель. Формулировка, включающая случайные эффекты, основана на допущении $E[\mathbf{w}_i \mathbf{w}'_i | \mathbf{X}_i] = \sigma_\varepsilon^2 \mathbf{I}_T + \sigma_u^2 \mathbf{i} \mathbf{i}'$, где $w_{it} = (\varepsilon_{it} + u_i)$. Мы рассматривали GLS- и FGLS-оценки в такой же формулировке, как и стратегию устойчивого оценивания ковариационной матрицы МНК. Одно из следствий анализа в разделе 11.5 состоит в том, что такая запись ковариационной матрицы шоков больше ограничительная, чем необходимая, учитывая информацию, содержащуюся в данных. Предположение о том, что $E[\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}'_i | \mathbf{X}_i] = \sigma_\varepsilon^2 \mathbf{I}_T$, подразумевает, что корректировка по периодам равна для всех пар наблюдений и возникает исключительно за счет неизменного c_i . В подразделе 10.2.6 мы оценивали эквивалентную модель с неограниченной ковариационной матрицей $E[\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}'_i | \mathbf{X}_i] = \boldsymbol{\Sigma}$. Из этого следует, что анализ с использованием случайных эффектов включает два ограничительных допущения: независимость средних значений $E[c_i | \mathbf{X}_i] = \alpha$ и гомоскедастичность $E[\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}'_i | \mathbf{X}_i] = \sigma_\varepsilon^2 \mathbf{I}_T$. [Заметим, что исключение второго допущения будет стоить нам возможности идентифицировать σ_u^2 как оцениваемый параметр. Это имеет значение, если корреляция по периодам t и s может возникать или из их общего u_i , или из корреляции $(\varepsilon_{it}, \varepsilon_{is})$, тогда не существует способа отдельно оценить дисперсию для u_i , не говоря уже о ковариациях ε_{it} и ε_{is} .] Однако важно отметить, что модель панельных данных можно рассматривать и формулировать как модель внешне не связанных регрессий с общими коэффициентами, в которой каждый период образует отдельное уравнение. На самом деле возможно, хоть и необязательно, наложить ограничение $E[\mathbf{w}_i \mathbf{w}'_i | \mathbf{X}_i] = \sigma_\varepsilon^2 \mathbf{I}_T + \sigma_u^2 \mathbf{i} \mathbf{i}'$.

Предположение о независимости средних значений — главный недостаток модели со случайными эффектами. Центральным свойством модели с фиксированными эффектами из раздела 11.4 является вероятность того, что $E[c_i | \mathbf{X}_i]$ будет непостоянным $g(\mathbf{X}_i)$. Поэтому МНК-регрессия y_{it} на \mathbf{x}_{it} дает несостоятельную оценку $\boldsymbol{\beta}$. Модель с фиктивными переменными, рассмотренная в разделе 11.4, представляет собой естественную альтернати-

ву. Модель с **фиксированными эффектами** имеет преимущество, которое заключается в возможности обходиться без неправдоподобного предположения о том, что c_i и x_{it} не коррелированы. Однако ее недостаток — в необходимости оценки n «параметров» α_i .

Чемберлен (Chamberlain, 1982, 1984) и Мундлак предложили альтернативные подходы, лежащие между этими двумя. Их модификации модели с фиксированными эффектами усиливают ее через **проекции** c_i на все строки X_i (Чемберлен) или на групповые средние (Мундлак). (См. подраздел 11.5.5.) Рассмотрим первый метод и предположим (как требуется) наличие сбалансированной панели из T наблюдений на группу. Чтобы развить модель подобным образом, допустим, что $T = 3$. Обобщение станет очевидно в заключении. Тогда проекция, предложенная Чемберленом, имеет вид

$$c_i = \alpha + x'_{i1}\gamma_1 + x'_{i2}\gamma_2 + x'_{i3}\gamma_3 + r_i, \quad (11-48)$$

где теперь r_i ортогонально x_{it} по построению²⁴. Подставим (11-48) в (11-47) и получим

$$y_{it} = \alpha + x'_{i1}\gamma_1 + x'_{i2}\gamma_2 + x'_{i3}\gamma_3 + x'_{it}\beta + \varepsilon_{it} + r_i.$$

Оценка $1 + 3K + K$ параметров этой модели рождает ряд сложностей. [Заметим, что у такого подхода есть вероятность быстро увеличивать (очень сильно) число параметров. Для нашей достаточно небольшой модели региональной производительности в примере 11.19 исходная модель с шестью главными коэффициентами плюс константы становится моделью с $1 + 6 + 17(6) = 109$ параметрами для оценки.]

Если используются только n наблюдений за период 1, тогда вектор параметров

$$\theta_1 = \alpha, (\beta + \gamma_1), \gamma_2, \gamma_3 = \alpha, \pi_1, \gamma_2, \gamma_3 \quad (11-49)$$

можно оценить состоятельно, хотя и неэффективно, обычным методом наименьших квадратов. «Модель» имеет вид

$$y_{i1} = z'_{i1}\theta_1 + w_{i1}, i = 1, \dots, n.$$

Собрав n наблюдений, мы имеем

$$y_1 = Z_1\theta_1 + w_1.$$

Если вместо этого используются только n наблюдений за период 2 или период 3, то МНК-оценки, в свою очередь,

$$\theta_2 = \alpha, \gamma_1, (\beta + \gamma_2), \gamma_3 = \alpha, \gamma_1, \pi_2, \gamma_3$$

²⁴Здесь есть некоторые тонкости, разрешить которые можно только теоретически. Если проекция в (11-48) не является условным средним, тогда получим $E[r_i \times x_{it}] = 0$, $t = 1, \dots, T$, а не $E[r_i | X_i] = 0$. Это не влияет на асимптотические свойства FGLS-оценки, рассматриваемой здесь, хотя и имеет следствия, например, в отношении несмещенности. Состоятельность сохранится независимо от этого. Предположения за пределами (11-48) не включают, что $\text{Var}[r_i | X_i]$ гомоскедастична. Она может и не быть таковой. Это могло бы изучаться эмпирически. Это следствие затрагивает эффективность, но не состоятельность. FGLS-оценка, рассматриваемая здесь, осталась бы постоянной, но GMM-оценка была бы более эффективна — см. главу 13. Кроме того, без гомоскедастичности неочевидно, что FGLS-оценка, предложенная здесь, более эффективна, нежели МНК (с устойчивой оценкой ковариационной матрицы). Наша задача — начать исследование с этого. Дальнейшие детали можно найти у Chamberlain (1984) и, например, у Im, Ahn, Schmidt, Wooldridge (1999).

или

$$\theta_3 = \alpha, \gamma_1, \gamma_2, (\beta + \gamma_3) = \alpha, \gamma_1, \gamma_2, \pi_3.$$

Остается лишь согласовать многочисленные оценки одних и тех же векторов параметров. В терминах изложенных выше схем мы получим следующее:

| | | | |
|--------------|---|---|---|
| МНК-оценки : | $a_1, p_1, c_{2,1}, c_{3,1},$ | $a_2, c_{1,2}, p_2, c_{3,2},$ | $a_3, c_{1,3}, c_{2,3}, p_3,$ |
| оцененные | | | |
| параметры : | $\alpha, (\beta + \gamma_1), \gamma_2, \gamma_3,$ | $\alpha, \gamma_1, (\beta + \gamma_2), \gamma_3,$ | $\alpha, \gamma_1, \gamma_2, (\beta + \gamma_3),$ |
| структурные | | | |
| параметры: | $\alpha, \beta, \gamma_1, \gamma_2, \gamma_3.$ | | |

(11-50)

Чемберлен предложил оценку, основанную на минимизации расстояния (MDE). С точки зрения данной проблемы MDE является, по сути, средневзвешенным для нескольких оценок каждой части вектора параметров. В главе 13 мы более подробно рассмотрим MDE для этого примера. (Другое, более простое применение оценки по минимальному расстоянию, показывающее процедуру «взвешивания» на практике, см. в теме о согласовании четырех конкурирующих оценок одного параметра в конце примера 11.20.) Существует альтернативный способ задать немного более очевидную оценку. Для первого периода

$$\begin{aligned} y_1 &= \begin{pmatrix} y_{1,1} \\ y_{2,1} \\ \vdots \\ y_{n,1} \end{pmatrix} = \begin{bmatrix} 1 & x_{1,1} & x_{1,1} & x_{1,2} & x_{1,3} \\ 1 & x_{2,1} & x_{2,1} & x_{2,2} & x_{2,3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n,1} & x_{n,1} & x_{n,2} & x_{n,3} \end{bmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{pmatrix} + \begin{pmatrix} r_{1,1} \\ r_{2,1} \\ \vdots \\ r_{n,1} \end{pmatrix} = \\ &= \tilde{X}_1 \theta + r_1. \end{aligned} \quad (11-51)$$

Мы рассматриваем это в качестве первого уравнения в модели внешне не связанных регрессий с T уравнениями. Второе уравнение для периода 2 такое же (с теми же коэффициентами) с данными из второго периода, возникающего в блоках, а затем также для периода 3 (и периодов 4, ..., T в общем случае). Объединяя данные по T уравнениям (периодам), мы получим

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} = \begin{pmatrix} \tilde{X}_1 \\ \tilde{X}_2 \\ \vdots \\ \tilde{X}_T \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma_1 \\ \vdots \\ \gamma_T \end{pmatrix} + \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_T \end{pmatrix} = \tilde{X} \theta + r, \quad (11-52)$$

где $E[\tilde{X}'r] = 0$ и (по допущению) $E[rr' | \tilde{X}] = \Sigma \otimes I_n$. При предположении о гомоскедастичности $r_{i,t}$ это как раз и будет приложением из подраздела 10.2.6. Параметры можно оценить при помощи FGLS, как показано в подразделе 10.2.6.

Пример 11.10. Затраты больницы

Кэри (Carey (1997)) изучил затраты больницы для выборки из 1733 госпиталей, наблюдаемых в течение пяти лет, за 1987–1991 гг. Оцененная модель имела вид

$$\begin{aligned} \ln (TC/P)_{it} = & \alpha_i + \beta_D DIS_{it} + \beta_O OPV_{it} + \beta_3 ALS_{it} + \beta_4 CM_{it} \\ & + \beta_5 DIS_{it}^2 + \beta_6 DIS_{it}^3 + \beta_7 OPV_{it}^2 + \beta_8 OPV_{it}^3 \\ & + \beta_9 ALS_{it}^2 + \beta_{10} ALS_{it}^3 + \beta_{11} DIS_{it} \times OPV_{it} \\ & + \beta_{12} FA_{it} + \beta_{13} HI_{it} + \beta_{14} HT_i + \beta_{15} LT_i + \beta_{16} Large_i \\ & + \beta_{17} Small_i + \beta_{18} NonProfit_i + \beta_{19} Profit_i \\ & + \varepsilon_{it}, \end{aligned}$$

где

TC = общие затраты,

P = индекс цен,

DIS = погашение долгов,

OPV = амбулаторные посещения,

ALS = средняя продолжительность пребывания,

CM = смешанный индекс,

FA = основные фонды,

HI = индекс рыночной концентрации Герфиндаля на уровне страны,

HT = фиктивная переменная для госпиталей с большой учебной нагрузкой,

LT = фиктивная переменная для госпиталей с низкой учебной нагрузкой,

$Large$ = фиктивная переменная для обширной городской местности,

$Small$ = фиктивная переменная для небольшой городской местности,

$NonProfit$ = фиктивная переменная для некоммерческого госпиталя,

$Profit$ = фиктивная переменная для некоммерческого госпиталя.

Мы использовали индексы D и O для коэффициентов DIS и OPV, это обособление будет произведено в последующей дискуссии. Модель, используемая в этом исследовании, аналогична (11-47) и (11-48). Начальные МНК-оценки получены для полной функции затрат в каждом году. Оценки внешне не связанных регрессий получены далее с использованием ограниченной версии системы Чемберлена. Этот второй шаг предполагал гибридную модель, которая преобразовала бы (11-49) так, чтобы в каждом периоде вектор коэффициентов имел вид

$$\theta_t = [\alpha_t, \beta_{Dt}(\gamma), \beta_{Ot}(\gamma), \beta_{3t}(\gamma), \beta_{4t}(\gamma), \beta_{5t}, \dots, \beta_{19t}],$$

где $\beta_{Dt}(\gamma)$ означает, что все пять лет переменной (DIS_{it}) включены в уравнение, и то же происходит для $\beta_{Ot}(\gamma)(OPV)$, $\beta_{3t}(\gamma)(ALS)$ и $\beta_{4t}(\gamma)(CM)$. Все это равносильно использованию

$$c_i = \alpha + \sum_{t=1987}^{1991} (DIS, OPV, ALS, CM)'_{it} \gamma_t + r_i$$

в (11-48).

Система внешне не связанных уравнений без ограничений, оцениваемая на втором шаге, позволяет получить многочисленные оценки различных параметров модели. Например, каждое из пяти уравнений дает оценку $(\beta_5, \dots, \beta_{19})$. Автор добавил еще одну модификацию модели, позволив коэффициентам по DIS_{it} и OPV_{it} изменяться во времени. Поэтому интересующими нас структурными параметрами являются $(\beta_{D1}, \dots, \beta_{D5}), (\gamma_{D1}, \dots, \gamma_{D5})$ (коэффициенты по DIS) и $(\beta_{O1}, \dots, \beta_{O5}), (\gamma_{O1}, \dots, \gamma_{O5})$ (коэффициенты по OPV). Всего нас интересуют 20 параметров. Оценки внешне не связанных регрессий в каждом году (уравнении) позволяют получить параметры по DIS за пять лет и по OPV за пять лет, т.е. всего 50 оценок. Согласование их будет означать наложение 30 ограничений. В табл. 11.7 показаны соотношения для изменяющегося во времени параметра по DIS_{it} в модели с пятью уравнениями. Числовые значения, приведенные автором, следуют за теоретическими выводами. Схожая таблица будет получена для коэффициентов по OPV, ALS и CM. (В последних двух не предполагалось, что коэффициент β будет изменяющимся во времени.) Из таблицы видно, например, что существуют непосредственно четыре разные оценки $\gamma_{D,87}$ во втором и пятом уравнениях и так же для всех остальных параметров. Объединяя записи в табл. 11.7 с аналогом для коэффициентов по OPV, мы видим 50 SUR/FGLS-оценок, которые будут использоваться для оценивания 20 преимущественных параметров. Чтобы согласовать разные оценки, автор использовал метод минимального расстояния. Мы вернемся к этому в примере 13.6, где рассмотрим MDE более подробно.

Таблица 11.7. Оценки коэффициентов в модели внешне не связанных регрессий для затрат больниц

| Уравнение | Коэффициент переменной в уравнении | | | | |
|-----------|------------------------------------|--------------------------------|--------------------------------|--------------------------------|--------------------------------|
| | DIS87 | DIS88 | DIS89 | DIS90 | DIS91 |
| | $\beta_{D,87} + \gamma_{D,87}$ | $\gamma_{D,88}$ | $\gamma_{D,89}$ | $\gamma_{D,90}$ | $\gamma_{D,91}$ |
| SUR87 | 1,76 | 0,116 | -0,0881 | 0,0570 | -0,0617 |
| | $\gamma_{D,87}$ | $\beta_{D,88} + \gamma_{D,88}$ | $\gamma_{D,89}$ | $\gamma_{D,90}$ | $\gamma_{D,91}$ |
| SUR88 | 0,254 | 1,61 | -0,0934 | 0,0610 | -0,0514 |
| | $\gamma_{D,87}$ | $\gamma_{D,88}$ | $\beta_{D,89} + \gamma_{D,89}$ | $\gamma_{D,90}$ | $\gamma_{D,91}$ |
| SUR89 | 0,217 | 0,0846 | 1,51 | 0,0454 | -0,0253 |
| | $\gamma_{D,87}$ | $\gamma_{D,88}$ | $\gamma_{D,89}$ | $\beta_{D,90} + \gamma_{D,90}$ | $\gamma_{D,91}$ |
| SUR90 | 0,179 | 0,0822 ^a | 0,0295 | 1,57 | 0,0244 |
| | $\gamma_{D,87}$ | $\gamma_{D,88}$ | $\gamma_{D,89}$ | $\gamma_{D,90}$ | $\beta_{D,91} + \gamma_{D,91}$ |
| SUR91 | 0,153 | 0,0363 | -0,0422 | 0,0813 | 1,70 |

^a Значение, представленное в опубликованной работе, равно 8,22. Верное значение составляет 0,0822. (Примеч. авт.)

11.6. Несферические шоки и устойчивые оценки ковариационной матрицы

Поскольку рассматриваемые здесь модели являются расширением классической модели регрессии, мы можем анализировать гетероскедастичность так же, как делали в главе 9, т.е. вычислить оценки обычным или допустимым обобщенным методом наименьших квадратов и получить подходящую устойчивую оценку ковариационной матрицы или же придать некоторую структуру дисперсиям шоков и использовать обобщенный метод наименьших квадратов. В случае с панельными данными существует больше гибкости для применения второго способа без строгих допущений о природе гетероскедастичности.

11.6.1. Устойчивые оценки модели с фиксированными эффектами

Как отмечалось в подразделе 11.3.2, в случае с панельными данными корреляция по наблюдениям внутри группы, вероятнее всего, будет иметь более значимое влияние на оцененную ковариационную матрицу оценки методом наименьших квадратов, чем гетероскедастичность. Это ясно видно по оценкам в табл. 11.1. В модели с фиксированными (или случайными) эффектами цель прямого включения общего эффекта в модель состоит в объяснении источника корреляции. Однако объяснение общего эффекта в модели не устраняет гетероскедастичность — оно должным образом центрирует условное среднее. Здесь мы рассмотрим прямое расширение оценки Уайта на модели с фиксированными и случайными эффектами.

В модели с фиксированными эффектами полная матрица регрессии равна $\mathbf{Z} = [\mathbf{X}, \mathbf{D}]$. Состоятельная при наличии гетероскедастичности оценка Уайта ковариационной матрицы для МНК, т.е. для оценки фиксированных эффектов, представляет собой правый нижний блок разделенной матрицы

$$\text{Est. Asy. Var}[\mathbf{b}, \mathbf{a}] = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{E}^2\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1},$$

где \mathbf{E} — диагональная матрица остатков наименьших квадратов (оценка фиксированных эффектов). Эти подсчеты обещают быть громоздкими, но, к счастью, эта проблема решается просто. Оценка Уайта для коэффициентов наклона получена при помощи использования данных в форме отклонений от группового среднего [см. (11-15) и (11-18)] при уже знакомом нам вычислении \mathbf{S}_0 [см. (9-26) и (9-27)]. Помимо этого оценка дисперсии шоков из (11-18) является аналогом оценки из (9-20), когда мы показывали, что соответствующее масштабирование Ω дало состоятельную оценку $\sigma^2 = \text{plim}[1/(nT)] \times \sum_{i=1}^n \sum_{t=1}^T \sigma_{it}^2$. Следствием этого является тот факт, что мы все еще можем использовать (11-18) для оценки дисперсий фиксированных эффектов.

Немного менее общее, но полезное упрощение этого результата можно получить, если мы допустим, что дисперсия шоков постоянна внутри i -й группы. Если $E[\varepsilon_{it}^2 | \mathbf{Z}_i] = \sigma_i^2$, тогда при панельных данных σ_i^2 можно оценить как $\mathbf{e}'_i \mathbf{e}_i / T$ с использованием остатков наименьших квадратов. Центральную матрицу в $\text{Est. Asy. Var}[\mathbf{b}, \mathbf{a}]$ можно заменить на $\sum_i (\mathbf{e}'_i \mathbf{e}_i / T) \mathbf{Z}'_i \mathbf{Z}_i$. Является ли такая оценка более предпочтительной — неясно. Если групповая модель корректна, то она и оценка Уайта будут оценивать одну и ту же матрицу.

С другой стороны, если дисперсии шоков все же изменяются внутри групп, то это пересмотренное вычисление может оказаться неподходящим.

Ареллано (Arellano (1987)) и Ареллано и Бовер (Arellano, Bover (1995)) продвинулись в анализе этой проблемы на шаг дальше. Если рассматривать i -ю группу в целом, то мы можем оценивать наблюдения в

$$y_i = \mathbf{X}_i \boldsymbol{\beta} + \alpha_i i_T + \boldsymbol{\varepsilon}_i$$

как обобщенную модель регрессии с ковариационной матрицей шоков $\boldsymbol{\Omega}_i$. В подразделе 11.3.2 мы видели, что эта общая модель без структуры у $\boldsymbol{\Omega}_i$ дает мало надежды на устойчивое или иное получение оценок. Но эта проблема более управляема при наличии панельных данных, где можно допустить, что корреляция по элементам равна нулю. Как и раньше, пусть \mathbf{X}_{i*} обозначает данные в форме отклонений от группового среднего. Здесь аналогом $\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}$ является

$$\mathbf{X}'_* \boldsymbol{\Omega} \mathbf{X}_* = \sum_{i=1}^n (\mathbf{X}'_{i*} \boldsymbol{\Omega}_i \mathbf{X}_{i*}).$$

По тем же соображениям, которыми мы руководствовались при построении оценки Уайта в главе 9, мы можем рассматривать оценку $\boldsymbol{\Omega}_i$ с выборкой из одного, $\mathbf{e}_i \mathbf{e}'_i$. Как и раньше, это является не состоятельной оценкой индивидуальных $\boldsymbol{\Omega}_i$, в получении которых мы заинтересованы, а оценкой суммы. Если n достаточно велико, тогда мы могли бы утверждать, что

$$\begin{aligned} \text{plim} \frac{1}{nT} \mathbf{X}'_* \boldsymbol{\Omega} \mathbf{X}_* &= \text{plim} \frac{1}{nT} \sum_{i=1}^n \mathbf{X}'_{i*} \boldsymbol{\Omega}_i \mathbf{X}_{i*} \\ &= \text{plim} \frac{1}{n} \sum_{i=1}^n \frac{1}{T} \mathbf{X}'_{*i} \mathbf{e}_i \mathbf{e}'_i \mathbf{X}_{*i} \\ &= \text{plim} \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T e_{it} e_{is} \mathbf{x}_{*it} \mathbf{x}'_{*is} \right). \end{aligned} \quad (11-53)$$

Это выражение представляет собой расширение (11-3) на случай с фиксированными эффектами.

11.6.2. Гетероскедастичность в модели со случайными эффектами

Поскольку модель со случайными эффектами представляет собой обобщенную регрессионную модель с известной структурой, то МНК с устойчивой оценкой асимптотической ковариационной матрицы — самый лучший способ оценки данных. GLS-оценка будет эффективна, в то время как МНК-оценка нет. Если предполагается абсолютно общая ковариационная структура, то можно было бы просто применить оценку Ареллано, описанную в предшествующем разделе, с одним общим постоянным членом, а не набором фиксированных эффектов. Но для модели со случайными эффектами $\eta_{it} = \varepsilon_{it} + u_i$ допущение о возможности дисперсии шоков изменяться по группам выглядело бы полезным расширением модели.

В ряде работ, в частности в работах (Mazodier, Trognon (1978)), (Baltagi, Griffin (1988)) и в недавней монографии (Baltagi (2005, p. 77–79)), описано,

как можно позволить характерной для группы компоненте u_i быть гетероскедастичной. Однако с эмпирической точки зрения у этого подхода есть непреодолимая проблема. В итоговом анализе все оценки компонент дисперсии должны основываться на суммах квадратов остатков и, например, оценка $\sigma_{u_i}^2$ получалась бы с использованием множества остатков из распределения u_i . Однако данные содержат только одно наблюдение по u_i , повторяющееся в каждом наблюдении в группе i . Поэтому представленные оценки, например, у Baltagi (2001) используют, в сущности, один остаток в каждом случае, чтобы оценить $\sigma_{u_i}^2$. То, что кажется среднеквадратичным остатком, является лишь $(1/T) \sum_{t=1}^T \hat{u}_i^2 = \hat{u}_i^2$. Свойства такой оценки неопределенные, а эффективность кажется маловероятной. С увеличением размера выборки и даже T оценки не сходятся к какому-либо числу из генеральной совокупности. [См. Hsiao (2003, p. 56).] Гетероскедастичность в компоненте ε_{it} представляет возможность моделирования, которая легче поддается обработке.

В подразделе 11.5.2 мы ввели гетероскедастичность в оценки модели со случайными эффектами тем, что позволили размерам групп меняться. Но оценка здесь (и ее допустимый аналог из следующего раздела) будет той же, если вместо $\theta_i = 1 - \sigma_\varepsilon / (T_i \sigma_u^2 + \sigma_\varepsilon^2)^{1/2}$ мы столкнулись бы с

$$\theta_i = 1 - \frac{\sigma_{\varepsilon i}}{\sqrt{\sigma_{\varepsilon i}^2 + T_i \sigma_u^2}}.$$

Таким образом, для вычисления необходимой оценки при помощи допустимого обобщенного метода наименьших квадратов нам снова потребуется лишь разработать состоятельные оценки для компонент дисперсии и затем применить GLS-преобразование, показанное ранее. Один из способов сделать это состоит в следующем: поскольку оценки объединенным МНК все же последовательны, МНК дает удобное множество остатков. Используя МНК-остатки для каждой конкретной группы, получим:

$$\widehat{\sigma_{\varepsilon i}^2 + u_i^2} = \frac{\mathbf{e}_i' \mathbf{e}_i}{T}.$$

Остатки по модели с фиктивными переменными очищены от индивидуального специфического эффекта u_i , поэтому $\sigma_{\varepsilon i}^2$ может быть состоятельно (по T) оценена при помощи

$$\widehat{\sigma_{\varepsilon i}^2} = \frac{\mathbf{e}_i^{lsdv'} \mathbf{e}_i^{lsdv}}{T},$$

где $e_{it}^{lsdv} = y_{it} - \mathbf{x}_{it}' \mathbf{b}^{lsdv} - a_i$. Объединяя слагаемые, получим

$$\hat{\sigma}_u^2 = \frac{1}{n} \sum_{i=1}^n \left[\left(\frac{\mathbf{e}_i^{ols'} \mathbf{e}_i^{ols}}{T} \right) - \left(\frac{\mathbf{e}_i^{lsdv'} \mathbf{e}_i^{lsdv}}{T} \right) \right] = \frac{1}{n} \sum_{i=1}^n (\widehat{u_i^2}).$$

Теперь мы можем вычислить FGLS-оценку, как и раньше.

11.6.3. Автокорреляция в моделях панельных данных

Как мы видели в подразделе 11.3.2 и примере 11.1, автокорреляция, т.е. корреляция по наблюдениям в группах панели, вероятно, будет основной

характеристикой модели. Однако рассмотрение данного эффекта здесь должно было вмещать автокорреляцию в широком смысле, т.е. ненулевые ковариации по наблюдениям в группе. Полученные результаты равно применимы к кластеризованным наблюдениям, что отмечалось в подразделе 11.3.3. Важным аспектом той спецификации было то, что при кластеризованных данных очевидной структуры для автокорреляции могло и не быть. Если множество панельных данных состоит точно из групп временных рядов и особенно если эти временные ряды относительно длинные, как в примере 11.11, возникает желание начать применять более детализированные и структурированные модели временных рядов, рассматриваемые в главе 20.

11.6.4. Кластерные (и панельные) устойчивые ковариационные матрицы для оценок с фиксированными и случайными эффектами

Как было предложено ранее, в тех ситуациях, когда подходят кластерные корректировки, внутри групп может возникать корреляция остатков, которая не объясняется полностью оценкой по обобщенному методу наименьших квадратов или моделью с фиксированными эффектами. Альтернативой (11-4) для оценок с фиксированными и случайными эффектами является прямое построение, основанное на результатах, которые мы уже получили.

Для оценки с фиксированными эффектами, основанной на (11-14) и (11-20), получим

$$\mathbf{b}_{LSDV} = \left[\sum_{g=1}^G \sum_{i=1}^{n_g} (\Delta(1)\mathbf{x}_{ig})(\Delta(1)\mathbf{x}_{ig})' \right]^{-1} \left[\sum_{g=1}^G \sum_{i=1}^{n_g} (\Delta(1)\mathbf{x}_{ig})(\Delta(1)y_{ig}) \right], \quad (11-54)$$

где $\Delta(1)\mathbf{x}_{it} = \mathbf{x}_{it} - (1)\bar{\mathbf{x}}_i$ — отклонение \mathbf{x}_{it} от вектора группового среднего в одном периоде. Чем обусловлено (1), вскоре станет ясно. Так же как и в (11-3), мы построим оценку устойчивой ковариационной матрицы, используя

$$\text{Est.Asy.Var}[\mathbf{b}_{LSDV}] = \left[\sum_{g=1}^G \left\{ \sum_{i=1}^{n_g} (\Delta(1)\mathbf{x}_{ig}) e_{ig} \right\} \left\{ \sum_{i=1}^{n_g} (\Delta(1)\mathbf{x}_{ig}) e_{ig} \right\}' \right]^{-1} \times (11-55)$$

$$\left[\sum_{g=1}^G \sum_{i=1}^{n_g} (\Delta(1)\mathbf{x}_{ig})(\Delta(1)\mathbf{x}_{ig})' \right]^{-1}.$$

Эта оценка эквивалентна (11-3), основанной на данных отклонений от их кластерных средних. (После небольших изменений в обозначениях она становится устойчивой оценкой для ковариационной матрицы оценки с фиксированными эффектами.) Из (11-32) и (11-33) GLS-оценка β для модели со

случайными эффектами имеет вид

$$\begin{aligned} \hat{\beta}_{GLS} &= \left[\sum_{g=1}^G \mathbf{X}'_g \Sigma_g^{-1} \mathbf{X}_g \right]^{-1} \left[\sum_{g=1}^G \mathbf{X}'_g \Sigma_g^{-1} y_g \right] = \\ &= \left[\sum_{g=1}^G \sum_{i=1}^{n_g} (\Delta(\theta_g) \mathbf{x}_{ig}) (\Delta(\theta_g) \mathbf{x}_{ig})' \right]^{-1} \times \\ &\times \left[\sum_{g=1}^G \sum_{i=1}^{n_g} (\Delta(\theta_g) \mathbf{x}_{ig}) (\Delta(\theta_g) y_{ig}) \right], \end{aligned} \quad (11-56)$$

где $\theta_g = 1 - (\sigma_\varepsilon / \sqrt{\sigma_\varepsilon^2 + n_g \sigma_u^2})$. Из этого следует, что оценка асимптотической ковариационной матрицы будет

$$\begin{aligned} \text{Est. Asy. Var}[\hat{\beta}_{GLS}] &= \left[\sum_{g=1}^G \sum_{i=1}^{n_g} (\Delta(\theta_g) \mathbf{x}_{ig}) (\Delta(\theta_g) \mathbf{x}_{ig})' \right]^{-1} \times \\ &\times \left[\sum_{g=1}^G \left\{ \sum_{i=1}^{n_g} (\Delta(\theta_g) \mathbf{x}_{ig}) e_{ig} \right\} \left\{ \sum_{i=1}^{n_g} (\Delta(\theta_g) \mathbf{x}_{ig}) e_{ig} \right\}' \right] \times \\ &\times \left[\sum_{g=1}^G \sum_{i=1}^{n_g} (\Delta(\theta_g) \mathbf{x}_{ig}) (\Delta(\theta_g) \mathbf{x}_{ig})' \right]^{-1}. \end{aligned} \quad (11-57)$$

Также см. (Cameron, Trivedi (2005, p. 838–839)).

Пример 11.11. Устойчивые стандартные ошибки оценок с фиксированными и случайными эффектами

В табл. 11.8 представлены оценки моделей с фиксированными и случайными эффектами, уже рассмотренные в табл. 11.5 и 11.6. Корректировка полученных стандартных ошибок является весьма существенным изменением в оценках. Это следствие особенно видно в примере со случайными эффектами, где оцененные стандартные ошибки возрастают в пять или более раз.

Таблица 11.8. Кластерные корректировки для оценок с фиксированными и случайными эффектами

| Переменная | Фиксированные эффекты | | | Случайные эффекты | | |
|------------------|-----------------------|---------------|------------|-------------------|---------------|------------|
| | Оценка | Станд. ошибка | Устойчивая | Оценка | Станд. ошибка | Устойчивая |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Constant | | | | 5,3455 | 0,04361 | 0,19866 |
| Exp | 0,1132 | 0,002471 | 0,00437 | 0,08906 | 0,002280 | 0,01276 |
| Exp ² | -0,00042 | 0,000055 | 0,000089 | -0,0007577 | 0,00005036 | 0,00031 |

Окончание табл. 11.8

| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|-------|----------|----------|---------|----------|-----------|----------|
| Wks | 0,00084 | 0,000600 | 0,00094 | 0,001066 | 0,0005939 | 0,00331 |
| Occ | -0,02148 | 0,01378 | 0,02052 | -0,1067 | 0,01269 | 0,05424 |
| Ind | 0,01921 | 0,01545 | 0,02450 | -0,01637 | 0,01391 | 0,053003 |
| South | -0,00186 | 0,03430 | 0,09646 | -0,06899 | 0,02354 | 0,05984 |
| SMSA | -0,04247 | 0,01942 | 0,03185 | -0,01530 | 0,01649 | 0,05421 |
| MS | -0,02973 | 0,01898 | 0,02902 | -0,02398 | 0,01711 | 0,06984 |
| Union | 0,03278 | 0,01492 | 0,02708 | 0,03597 | 0,01367 | 0,05653 |

11.7. Пространственная автокорреляция

Эффекты кластеризации, отмеченные в подразделе 11.3.3, обусловлены вероятностью того, что влияние соседних объектов друг на друга создаст своего рода корреляцию в пространстве, а не во времени, на которой мы сосредотачивались до сих пор. Такой эффект часто происходит в межрегиональных исследованиях, например в сельском хозяйстве, городской и региональной экономике. Современные исследования этого феномена включают изучение структуры расходов в работе Case (1991), реальных цен на собственность в работе Bell и Bockstael (2000), а также анализ побочных результатов R&D в работе Baltagi и Li (2001). Для формализации этих идей были построены модели **пространственной автокорреляции** [см. классическую справку у Anselin (1988, 2001) и современный обзор у Le Sage, Pace (2009)].

Модель с пространственной автокорреляцией может быть сформулирована следующим образом: регрессионная модель принимает уже известную панельную структуру

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it} + u_i, i = 1, \dots, n; t = 1, \dots, T.$$

Единое u_i является обычным эффектом влияния объекта (к примеру, страны). Корреляция в пространстве проявляется в структуре пространственной автокорреляции

$$\varepsilon_{it} = \lambda \sum_{j=1}^n W_{ij} \varepsilon_{jt} + v_t.$$

Скаляр λ служит **коэффициентом пространственной авторегрессии**. Элементы W_{ij} являются пространственными (или **смежными**) весами, которые считаются известными. Элементы в верхней сумме — это ряд пространственного веса или **матрицы смежности** \mathbf{W} , поэтому для n элементов мы получим

$$\boldsymbol{\varepsilon}_t = \lambda \mathbf{W} \boldsymbol{\varepsilon}_t + \mathbf{v}_t, \mathbf{v}_t = v_t \mathbf{i}.$$

Структура модели находит отражение в матрице симметричных весов \mathbf{W} . Для примера рассмотрите округа или страны, географически заключенные в сетку координат или некоторую линейную шкалу, как то линия от одного

берега страны к другому. Как правило, W_{ij} будет равно единице для i, j соседних пар и нулю иначе. В свою очередь W_{ij} может отражать расстояния в пространстве так, что W_{ij} уменьшается при увеличении $|i - j|$. Это похоже на матрицу автокорреляции во времени. Предполагая, что $|\lambda|$ меньше единицы, а элементы \mathbf{W} таковы, что матрица $(\mathbf{I} - \lambda\mathbf{W})$ невырождена, мы можем записать следующее:

$$\boldsymbol{\varepsilon}_t = (\mathbf{I}_n - \lambda\mathbf{W})^{-1}\mathbf{v}_t,$$

отсюда для n наблюдений в период t

$$\mathbf{y}_t = \mathbf{X}_t\boldsymbol{\beta} + (\mathbf{I}_n - \lambda\mathbf{W})^{-1}\mathbf{v}_t + \mathbf{u}.$$

Далее мы допускаем, что u_i и v_i имеют нулевые средние значения, дисперсии σ_u^2 и σ_v^2 и независимы по странам и друг с другом. Из этого следует, что обобщенная модель регрессии применяется к n наблюдениям в период t :

$$\begin{aligned} E[\mathbf{y}_t | \mathbf{X}_t] &= \mathbf{X}_t\boldsymbol{\beta}, \\ \text{Var}[\mathbf{y}_t | \mathbf{X}_t] &= (\mathbf{I}_n - \lambda\mathbf{W})^{-1}[\sigma_v^2\mathbf{I}\mathbf{I}'](\mathbf{I}_n - \lambda\mathbf{W})^{-1} + \sigma_u^2\mathbf{I}_n. \end{aligned}$$

С этого момента оценивание могло бы осуществляться, как в главе 9, за исключением необходимости оценить λ . В этом случае не существует естественной оценки λ , основанной на остатках. Недавние разработки в этом направлении включают допущение о нормальности и применяют методы максимального правдоподобия. [Логарифмическую функцию правдоподобия для данной модели и множество ссылок можно найти у Baltagi (2005, р. 196). Глубокий анализ проблемы оценки сделан в работе Bell, Bockstael (2000).]

Очевидно, первый шаг в анализе — это тестирование пространственных эффектов. Классической процедурой для межобъектных данных является I -статистика Морана (1950), которая вычисляется для каждого множества остатков e_t при помощи

$$I_t = \frac{n \sum_{i=1}^n \sum_{j=1}^n W_{ij}(e_{it} - \bar{e}_t)(e_{jt} - \bar{e}_t)}{\left(\sum_{i=1}^n \sum_{j=1}^n W_{i,j}\right) \sum_{i=1}^n (e_{it} - \bar{e}_t)^2}. \quad (11-58)$$

Для панели из T независимых наборов наблюдений в $\bar{I} = \frac{1}{T} \sum_{t=1}^T I_t$ будет использоваться вся имеющаяся информация. Аппроксимация в больших выборках дисперсии статистики в условиях нулевой гипотезы об отсутствии пространственной автокорреляции имеет вид

$$V^2 = \frac{1}{T} \frac{n^2 \sum_{i=1}^n \sum_{j=1}^n W_{ij}^2 + 3 \left(\sum_{i=1}^n \sum_{j=1}^n W_{ij}\right)^2 - n \sum_{i=1}^n \left(\sum_{j=1}^n W_{ij}\right)^2}{(n^2 - 1) \left(\sum_{i=1}^n \sum_{j=1}^n W_{ij}\right)^2}. \quad (11-59)$$

Статистика \bar{I}/V сходится к стандартной нормальной в условиях нулевой гипотезы и может создать основу теста. (Предположение о независимости во времени, вероятнее всего, будет в лучшем случае сомнительно.) Baltagi, Song, Koh (2003) определяют множество тестов множителей Лагранжа, основанных на допущении о нормальности. Два из них, которые применяются в

анализе межобъектных данных [см. Bell и Bockstael (2000, p. 78)], имеют следующий вид:

$$LM(1) = \frac{(e'W'e/s^2)^2}{tr(W'W + W^2)}$$

для пространственной автокорреляции и

$$LM(2) = \frac{(e'W'y/s^2)^2}{b'X'WMW'Xb/s^2 + tr(W'W + W^2)}$$

для пространственно запаздывающих зависимых переменных, где e — вектор МНК остатков, $s^2 = e'e/n$ и $M = I - X(X'X)^{-1}X'$. [См. Anselin, Hudak (1992).]

Анселин (Anselin (1988)) определяет некоторые возможные направления распространения пространственной модели на динамические регрессии. «Чистая пространственно-рекурсивная модель» определяет, что автокорреляция относится к соседним значениям в предыдущем периоде:

$$y_{it} = \gamma[Wy_{t-1}]_i + x'_{it}\beta + \varepsilon_{it}.$$

«Пространственно-временная рекурсивная модель» определяет чисто авторегрессионную в отношении соседних значений зависимость в предыдущем периоде:

$$y_{it} = \rho y_{i,t-1} + \gamma[Wy_{t-1}]_i + x'_{it}\beta + \varepsilon_{it}.$$

«Пространственно-временная одновременная» модель определяет, что пространственная зависимость в отношении соседних значений существует в текущем периоде:

$$y_{it} = \rho y_{i,t-1} + \lambda[Wy_t]_i + x'_{it}\beta + \varepsilon_{it}.$$

Наконец, «пространственно-временная динамическая модель» определяет, что авторегрессия зависит от соседних значений и в текущем, и в предыдущем периоде:

$$y_{it} = \rho y_{i,t-1} + \lambda[Wy_t]_i + \gamma[Wy_{t-1}]_i + x'_{it}\beta + \varepsilon_{it}.$$

Пример 11.12. Пространственная автокорреляция в продажах недвижимости

Белл и Боксталь (Bell и Bockstael) анализировали проблему моделирования пространственной автокорреляции в больших выборках. Это с большой долей вероятности становится обычной проблемой при данных ГИС (геоинформационной системы). Центральной проблемой является максимизация функции правдоподобия, включающей разреженную матрицу, $(I - \lambda W)$. Прямые подходы к решению этой проблемы могут повлечь серьезные неточности в оценке обратной матрицы и определителя. Келехан и Пруша (Kelejian и Prucha (1999)) разработали оценку для λ , основанную на моментах, которая помогает смягчить эту проблему. Как только мы имеем оценку λ в своем распоряжении, оценка модели пространственной

автокорреляции осуществляется по FGLS. Авторы применили этот метод к анализу межобъектных данных по 1000 продаж жилья в округе Энн-Эрандел, штат Мэриленд, с 1993 по 1996 г. Все проданные участки включали дома, построенные в течение одного года перед продажей. Для измерения интересующих авторов свойств использовалась программа GIS. Модель имеет следующий вид:

$$\begin{aligned} \ln Price = & \alpha + \beta_1 \text{ оценочная стоимость (LIV)} \\ & + \beta_2 \text{ площадь участка (LLT)} \\ & + \beta_3 \text{ расстояние в км до Вашингтона (LDC)} \\ & + \beta_4 \text{ расстояние в км до Балтимора (LBA)} \\ & + \beta_5 \% \text{ земли вокруг участка} \\ & \text{в собственности государства (POPN)} \\ & + \beta_6 \% \text{ земли вокруг участка в природном частном} \\ & \text{пространстве (PNAT)} \\ & + \beta_7 \text{ земли вокруг участка в интенсивно развивающемся} \\ & \text{использовании (PDEV)} \\ & + \beta_8 \% \text{ земли вокруг участка в использовании с низкой} \\ & \text{плотностью населения (PLOW)} \\ & + \beta_9 \text{ государственная канализационная служба} \\ & (1, \text{ если существует или запланирована, } 0, \text{ если нет) (PSEW)} \\ & + \varepsilon. \end{aligned}$$

(Земля вокруг участка — это все участки в данных GIS, центроиды которых расположены в пределах 500 метров от участка, по которому осуществляется сделка.) Для всей модели спецификация имела вид

$$\begin{aligned} y &= X\beta + \varepsilon, \\ \varepsilon &= \lambda W\varepsilon + v. \end{aligned}$$

Авторы обозначили четыре матрицы смежности:

$$\begin{aligned} W1:W_{ij} &= 1, \text{ если расстояние между } i \text{ и } j < 600 \text{ м,} \\ & 0, \text{ если иначе,} \\ W2:W_{ij} &= 1, \text{ если расстояние между } i \text{ и } j < 200 \text{ м, } 0, \text{ если иначе,} \\ W3:W_{ij} &= 1, \text{ если расстояние между } i \text{ и } j < 400 \text{ м, } 0, \text{ если иначе,} \\ W4:W_{ij} &= 1, \text{ если расстояние между } i \text{ и } j < 600 \text{ м, } 0, \text{ если иначе.} \end{aligned}$$

Все матрицы смежности были стандартизированы по строкам, т.е. элементы в каждой строке взвешивались так, чтобы строка в сумме давала единицу. Одной из целей этого исследования было изучение влияния стандартизации строк на оценки. Она применяется для повышения числовой стабильности процесса оптимизации. Поскольку оценки численно зависят от нормировки, то она может оказаться не полностью безвредной.

Тестовые статистики для пространственной автокорреляции, основанные на МНК-остатках, приведены в табл. 11.9. (Они взяты из табл. 3 у авторов.) Статистика Морана имеет стандартное нормальное распределение,

в то время как статистика множителей Лагранжа (LM) распределена по хи-квадрат с одной степенью свободы. Все статистики, кроме статистики множителей Лагранжа (LM(2)) для W3, превосходят 99%-е критическое значение из соответствующей таблицы, поэтому мы можем заключить, что здесь есть свидетельство о наличии пространственной автокорреляции. Оценки некоторых регрессий приведены в табл. 11.10. В оставшейся части исследования авторы обнаружили, что результаты в определенном смысле чувствительны к спецификации пространственной матрицы весов и не очень чувствительны к методу оценивания λ .

Таблица 11.9. Статистические тесты на пространственную автокорреляцию

| | W1 | W2 | W3 | W4 |
|----------|-------|-------|--------|-------|
| I Морана | 7,89 | 9,67 | 13,66 | 6,88 |
| LM(1) | 49,95 | 84,93 | 156,48 | 36,46 |
| LM(2) | 7,40 | 17,22 | 2,33 | 7,42 |

Таблица 11.10. Оцененные модели пространственной регрессии

| Пара- метр | МНК | | FGLS ^a | | Пространственные по максимальному правдоподобию для WL | | Пространст- венные по обобщ. моментам для W1 | |
|---------------|----------|---------------|-------------------|---------------|---|---------------|---|---------------|
| | Оценка | Станд. ош. | Оценка | Станд. ош. | Оценка | Станд. ош. | Оценка | Станд. ош. |
| | α | 4,7332 | 0,2047 | 4,7380 | 0,2048 | 5,1277 | 0,2204 | 5,0648 |
| β_1 | 0,6926 | 0,0124 | 0,6924 | 0,0214 | 0,6537 | 0,0135 | 0,6638 | 0,0132 |
| β_2 | 0,0079 | 0,0052 | 0,0078 | 0,0052 | 0,0002 | 0,0052 | 0,0020 | 0,0053 |
| β_3 | -0,1494 | 0,0195 | -0,1501 | 0,0195 | -0,1774 | 0,0245 | -0,1691 | 0,0230 |
| β_4 | -0,0453 | 0,0114 | -0,0455 | 0,0114 | -0,0169 | 0,0156 | -0,0278 | 0,0143 |
| β_5 | -0,0493 | 0,0408 | -0,0484 | 0,0408 | -0,0149 | 0,0414 | -0,0269 | 0,0413 |
| β_6 | 0,0799 | 0,0177 | 0,0800 | 0,0177 | 0,0586 | 0,0213 | 0,0644 | 0,0204 |
| β_7 | 0,0677 | 0,0180 | 0,0680 | 0,0180 | 0,0253 | 0,0221 | 0,0394 | 0,0211 |
| β_8 | -0,0166 | 0,0194 | -0,0168 | 0,0194 | -0,0374 | 0,0224 | -0,0313 | 0,0215 |
| β_9 | -0,1187 | 0,0173 | -0,1192 | 0,0174 | -0,0828 | 0,0180 | -0,0939 | 0,0179 |
| λ | — | — | — | — | 0,4582 | 0,0454 | 0,3517 | — |

^a Автор представил результаты с использованием модели гетероскедастичности $\sigma_i^2 \times f(LIV_i, LIV_i^2)$. Функция $f(\cdot)$ не идентифицирована.

Пример 11.13. *Пространственные лаги в расходах на здравоохранение*
 Москон, Кнап, Тосетти (Moscone, Knapp, Tosetti (2007)) исследовали детерминанты расходов на психическое здоровье за шесть лет в 148 британских органах местной власти с использованием двух видов модели пространственной корреляции, чтобы выявить возможное взаимодействие между властями, а также ненаблюдаемую пространственную гетерогенность. В дополнение к объединенной регрессии и модели со случайными эффектами были получены следующие оценки. Первая из них — модель с **пространственными лагами**

$$y_t = \gamma_t \mathbf{i} + \rho \mathbf{W} y_t + \{ \mathbf{X}_t \beta + \mathbf{u} + \varepsilon_t,$$

где \mathbf{u} — вектор случайных эффектов размером 148×1 , а \mathbf{i} — столбец из единиц размером 148×1 . Для каждого органа местной власти

$$y_{it} = \gamma_t + \rho (w'_i y_t) + x'_{it} \beta + u_i + \varepsilon_{it},$$

где w'_i — i -я строка матрицы смежности \mathbf{W} . Показатели смежности в \mathbf{W} обозначались единицей, если район имел общую границу или высшую точку, и нулем, если иначе. (Авторы также экспериментировали с другими матрицами смежности, в основе которых лежали социально-демографические различия.) Вторая оцененная модель — модель **пространственной корреляции ошибок**

$$\begin{aligned} y_t &= \gamma_t \mathbf{i} + \{ \mathbf{X}_t \beta + \mathbf{u} + \varepsilon_t, \\ \varepsilon_t &= \lambda \mathbf{W} \varepsilon_t + \mathbf{v}_t. \end{aligned}$$

Для каждого органа местной власти эта модель предполагает

$$y_{it} = \gamma_t + x'_{it} \beta + u_i + \lambda \sum_j w_{ij} \varepsilon_{jt} + v_{it}.$$

Авторы использовали метод максимального правдоподобия для оценки параметров модели. Чтобы упростить вычисления, они отмечают, что максимизацию можно осуществить при помощи двухшаговой процедуры. Как мы видели в других примерах, когда матрица Ω в обобщенной модели регрессии известна, подходящей оценкой является GLS. Для обеих рассматриваемых моделей при известном параметре пространственной автокорреляции GLS-трансформация данных позволяет получить классическую регрессионную модель. [См. (9-11).] Используемый метод нужен для проведения итераций к простому оцениванию и от него γ_t , β и σ_ε^2 при помощи МНК и максимизации функции «концентрированного логарифмического правдоподобия», которая при других оценках является функцией параметра пространственной автокорреляции ρ или λ и дисперсии гетерогенности σ_u^2 .

Зависимая переменная в моделях — логарифм расходов на психическое здоровье на душу населения. Ковариаты — процентная доля мужчин и людей младше 20 лет в данной области, средние ставки по ипотечным кредитам, число заявлений о безработице, занятость,

средняя цена дома, медианная недельная зарплата, процент семей с одним родителем, фиктивные переменные на представительство Лейбористской или Либерально-демократической партии в органах власти и плотности населения («для контроля факторов со стороны предложения»). Оцененные коэффициенты пространственной автокорреляции в двух моделях равны 0,1579 и 0,1220, оба значения более чем в два раза превышают оцененную стандартную ошибку. Гипотеза об отсутствии пространственной корреляции, в основе которой простые тесты Вальда, была бы отвергнута. Значения логарифмического правдоподобия для обеих пространственных моделей равны +206,3 и +202,8 в сравнении с -211,1 для модели без пространственных эффектов или эффектов региона, поэтому получаемый результат, по всей видимости, отдает предпочтение пространственным моделям, основанным на тестовой статистике хи-квадрат (с одной степенью свободы) с разницей в два раза. Однако в этом исходе заключается некоторая неоднозначность, поскольку улучшенная «подгонка» могла с большей вероятностью произойти из-за эффектов региона, нежели из-за пространственных. Простая модель случайных эффектов дает значение логарифмического правдоподобия, равное +202,3, что подтверждает это. По сравнению с этим значением модель с пространственными лагами выглядит более предпочтительной спецификацией, в то время как модель пространственной автокорреляции не придает ничего значительного функции логарифмического правдоподобия в сравнении с базовой моделью случайных эффектов.

11.8. Эндогенность

Последние приложения **панельных данных** в значительной степени полагаются на методы инструментальных переменных. Мы подробно обсудим эту методологию в главе 13, где рассмотрим оценки обобщенного метода моментов (GMM). На данном этапе мы можем рассмотреть два главных основообразующих блока в этом наборе методов: оценку Хаусмана–Тейлора (1981) для модели случайных эффектов и предложения Бхаргава–Сарган (1983) для оценивания модели динамических панельных данных. Оба этих инструмента играют значительную роль в GMM-оценке моделей динамических панелей в главе 13.

11.8.1. Оценка инструментальных переменных Хаусмана–Тейлора

Вспомним исходную спецификацию линейной модели панельных данных (11-1):

$$y_{it} = \mathbf{x}'_{it}\beta + \mathbf{z}'_i\alpha + \varepsilon_{it}. \quad (11-60)$$

Модель со случайными эффектами основана на предположении о том, что ненаблюдаемые индивидуальные (специфические для объекта) эффекты \mathbf{z}_i не коррелированы с включенными переменными \mathbf{x}_{it} . Это допущение является главным недостатком данной модели. Однако введение случайных эффектов позволяет модели содержать наблюдаемые не изменяющиеся во

времени характеристики, к примеру демографические, в то время как модель с фиксированными эффектами не дает такой возможности — если они и есть, то их просто поглощают фиксированные эффекты. **Оценка Хаусмана–Тейлора** (Hausman, Taylor (1981)) для модели со случайными эффектами предлагает способ преодолеть первое из затруднений и одновременно избежать второго.

Модель имеет следующую форму:

$$y_{it} = \mathbf{x}'_{1it}\beta_1 + \mathbf{x}'_{2it}\beta_2 + \mathbf{z}'_{1i}\alpha_1 + \mathbf{z}'_{2i}\alpha_2 + \varepsilon_{it} + u_i,$$

где $\beta = (\beta'_1, \beta'_2)'$ и $\alpha = (\alpha'_1, \alpha'_2)'$. В такой формулировке все индивидуальные эффекты, обозначаемые как z_i , являются наблюдаемыми. Как и раньше, ненаблюдаемые индивидуальные эффекты, содержащиеся в $\mathbf{z}'_i\alpha$ в (11-60), включены в индивидуальный случайный член u_i . Хаусман и Тейлор выделяют четыре множества *наблюдаемых* переменных в данной модели:

\mathbf{x}_{1it} — K_1 переменных, которые изменяются во времени и не коррелированы с u_i ,

\mathbf{z}_{1i} — L_1 переменных, которые не изменяются во времени и не коррелированы с u_i ,

\mathbf{x}_{2it} — K_2 переменных, которые изменяются во времени и коррелированы с u_i ,

\mathbf{z}_{2i} — L_2 переменных, которые не изменяются во времени и коррелированы с u_i .

Допущения о случайных членах в модели имеют вид

$$E[u_i | \mathbf{x}_{1it}, \mathbf{z}_{1i}] = 0, \text{ хотя } E[u_i | \mathbf{x}_{2it}, \mathbf{z}_{2i}] \neq 0,$$

$$\text{Var}[u_i | \mathbf{x}_{1it}, \mathbf{z}_{1i}, \mathbf{x}_{2it}, \mathbf{z}_{2i}] = \sigma_u^2,$$

$$\text{Cov}[\varepsilon_{it}, u_i | \mathbf{x}_{1it}, \mathbf{z}_{1i}, \mathbf{x}_{2it}, \mathbf{z}_{2i}] = 0,$$

$$\text{Var}[\varepsilon_{it} + u_i | \mathbf{x}_{1it}, \mathbf{z}_{1i}, \mathbf{x}_{2it}, \mathbf{z}_{2i}] = \sigma^2 = \sigma_\varepsilon^2 + \sigma_u^2,$$

$$\text{Corr}[\varepsilon_{it} + u_i, \varepsilon_{is} + u_i | \mathbf{x}_{1it}, \mathbf{z}_{1i}, \mathbf{x}_{2it}, \mathbf{z}_{2i}] = \rho = \sigma_u^2 / \sigma^2.$$

Обратите внимание на ключевое допущение о том, что можно различать множества переменных \mathbf{x}_1 и \mathbf{z}_1 , не коррелированных с u_i , от \mathbf{x}_2 и \mathbf{z}_2 , которые коррелированы с ошибками. Возможное наличие \mathbf{x}_2 и \mathbf{z}_2 усложняет спецификацию и оценку модели со случайными эффектами в первую очередь.

Попутно заметим, что мы можем противопоставить эти четыре допущения предположениям, сделанным Плампером, Трёгером (Plümper, Troeger (2007)) при рассмотрении FEVD в подразделе 11.4.5 и состоящим (в данном контексте) в том, что \mathbf{x}_{1it} и \mathbf{x}_{2it} изменяются во времени и каким-то образом коррелированы с u_i , в то время как \mathbf{z}_{1i} и \mathbf{z}_{2i} не изменяются во времени и не коррелированы с u_i . В обеих формулировках можно использовать (11-61). Два подхода различаются по дополнительным моментным условиям $E[\text{variable} \times (u_i + \varepsilon_{it})] = 0$, которые используются для нахождения параметров α_1 и α_2 .

По построению любые МНК- или GLS-оценки этой модели несостоятельны, если модель содержит переменные, коррелированные со случайными эффектами. Хаусман и Тейлор предложили оценку инструментальных переменных, использующую только информацию внутри модели (т.е. уже уста-

новленную). Стратегия оценивания основана на следующей логике: во-первых, с помощью отклонений от групповых средних мы найдем

$$y_{it} - \bar{y}_i = (\mathbf{x}_{1it} - \bar{\mathbf{x}}_{1i})' \beta_1 + (\mathbf{x}_{2it} - \bar{\mathbf{x}}_{2i})' \beta_2 + \varepsilon_{it} - \bar{\varepsilon}_i, \quad (11-61)$$

а это подразумевает, что обе части β могут быть состоятельно оценены методом наименьших квадратов, *несмотря на корреляцию между x_2 и u* . Это уже знакомая нам оценка модели с фиксированными эффектами методом наименьших квадратов с фиктивными переменными — преобразование до отклонений от групповых средних устраняет из модели часть шока, коррелирующего с x_{2it} . В исходной модели Хаусман и Тейлор показали, что отклонения от групповых средних можно использовать как $(K_1 + K_2)$ инструментальных переменных при оценке (β, α) . В этом состоит смысл (11-61). Поскольку вектор z_1 не коррелирован с шоками, он может служить в качестве набора L_1 инструментальных переменных. Но необходимо еще L_2 инструментальных переменных. Авторы показали, что групповые средние значения для x_1 можно использовать в качестве инструментальных переменных и тогда модель будет идентифицируема, если K_1 больше или равно L_2 . Поэтому в целях идентификации K_1 должно быть, по крайней мере, таким же, как L_2 . Как и прежде, **допустимый GLS** является более предпочтительным, чем МНК, и он доступен. Подобно этому оценка по FGLS будет лучше, чем простая оценка инструментальных переменных, которая будет состоятельна, но неэффективна.

Авторы предлагают последовательность шагов для получения состоятельной и эффективной оценки:

Шаг 1. Получить LSDV-оценку (с фиксированными эффектами) для $\beta = (\beta_1', \beta_2')$, основанную на x_1 и x_2 . Оценка дисперсии остатков на этом шаге является состоятельной оценкой σ_ε^2 .

Шаг 2. Построить внутригрупповые остатки e_{it} исходя из LSDV-регрессии на шаге 1. Свести групповые средние значения этих остатков в вектор данных размерностью, равной числу элементов во всей выборке. Итак, $e_{it}^* = \bar{e}_i = \frac{1}{T} \sum_{t=1}^T (y_{it} - \mathbf{x}_{it}' \mathbf{b}_w)$, $t = 1, \dots, T, i = 1, \dots, n$. (Индивидуальный постоянный член a_i не включен в e_{it}^* .) Эти групповые средние используются в качестве зависимой переменной в регрессии инструментальных переменных на z_1 и z_2 с инструментальными переменными z_1 и x_1 . (Отметим требование идентификации о том, что K_1 , число переменных в x_1 , должно быть как минимум таким же, как L_2 , число переменных в z_2 .) Каждая не изменяющаяся во времени переменная повторяется T раз в матрице данных этой регрессии. Это позволяет получить состоятельную оценку для α .

Шаг 3. Дисперсия остатков в регрессии на шаге 2 представляет собой состоятельную оценку для $\sigma^{*2} = \sigma_u^2 + \sigma_\varepsilon^2/T$. Из этой оценки и оценки σ_ε^2 на шаге 1 мы выводим оценку для $\sigma_u^2 = \sigma^{*2} - \sigma_\varepsilon^2/T$. Затем находим вес для допустимого GLS в этой модели при помощи оценки для

$$\theta = 1 - \sqrt{\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_u^2}}$$

Шаг 4. Финальным шагом является получение взвешенной оценки инструментальных переменных. Пусть полный набор переменных в модели равен

$$\mathbf{w}'_{it} = (\mathbf{x}'_{1it}, \mathbf{x}'_{2it}, \mathbf{z}'_{1i}, \mathbf{z}'_{2i}).$$

Соберем эти nT наблюдений в строки матрицы данных \mathbf{W} . Преобразованные переменные для GLS, как и прежде, когда мы впервые строили модель со случайными эффектами, имеют вид

$$\mathbf{w}^{*'}_{it} = \mathbf{w}'_{it} - \hat{\theta} \bar{\mathbf{w}}'_i \quad \text{и} \quad y^*_{it} = y_{it} - \hat{\theta} \bar{y}_i,$$

где $\hat{\theta}$ обозначает выборочную оценку для θ . Преобразованные данные собраны в строках матрицы данных \mathbf{W}^* и в столбце вектора \mathbf{y}^* . Заметим, что в случае с не изменяющимися во времени переменными в \mathbf{w}_{it} групповое среднее является исходной переменной, а преобразование лишь умножает эту переменную на $1 - \hat{\theta}$. Инструментальные переменные имеют следующий вид:

$$\mathbf{v}'_{it} = [(\mathbf{x}_{1it} - \bar{\mathbf{x}}_{1i})', (\mathbf{x}_{2it} - \bar{\mathbf{x}}_{2i})', \mathbf{z}'_{1i} \bar{\mathbf{x}}'_{1i}].$$

Они собраны в строки матрицы \mathbf{V} размером $nT \times (K_1 + K_2 + L_1 + K_1)$. Заметим, что для третьего и четвертого множеств инструментов не изменяющиеся во времени переменные и групповые средние для каждого члена группы повторяются. Оценка методом инструментальных переменных будет равна

$$(\hat{\beta}', \hat{\alpha}')'_{IV} = [(\mathbf{W}^{*'} \mathbf{V})(\mathbf{V}' \mathbf{V})^{-1} (\mathbf{V}' \mathbf{W}^*)]^{-1} [(\mathbf{W}^{*'} \mathbf{V})(\mathbf{V}' \mathbf{V})^{-1} (\mathbf{V}' \mathbf{y}^*)]^{25}. \quad (11-62)$$

Оценка методом инструментальных переменных состоятельна, если данные не взвешиваются, т.е. если в вычислениях используется \mathbf{W} , а не \mathbf{W}^* . Но она неэффективна, так же, как и оценка по МНК, состоятельна, но неэффективна в оценивании более простой модели со случайными эффектами.

Пример 11.14. *Отдача от обучения*

Экономическая отдача от обучения является довольно распространенной темой исследований у эконометристов. Для этих целей PSID и NLS представляют собой богатый источник панельных данных. Из уравнений заработной платы (или логарифма заработной платы) ясно, что экономическая выгода от обучения коррелирована со скрытыми, неизмеряемыми характеристиками конкретного агента, как то врожденные способности, интеллект, стимулы или упорство. По сути, проблема того, что простые модели случайных эффектов, в основе которых лежат панельные данные, будут подвержены эффектам, отмеченным ранее, является далеко не главной. Модель с фиксированными эффектами служит очевидной альтернативой, однако эти ценные данные содержат много полезных переменных, среди которых раса, членство в профсоюзе и семейное положение, которые, как правило, не изменяются во времени. Что еще хуже, более всего интересующая нас переменная, годы обучения, также является не изменяющейся во времени. Хаусман и Тейлор (1981) предложили оценку, описываемую здесь, в качестве решения этих проблем.

²⁵Обратите внимание, что оценка FGLS со случайными эффектами имеет вид $(\hat{\beta}', \hat{\alpha}')'_{RE} = [\mathbf{W}^{*'} \mathbf{W}^*]^{-1} \mathbf{W}^{*'} \mathbf{y}^*$.

Авторы изучили влияние обучения на логарифм заработной платы при помощи случайной выборки из PSID по 750 человек в возрасте 25–55 лет, наблюдаемых в течение двух лет, в 1968 и 1972 гг. Два года были выбраны таким образом, чтобы минимизировать эффект серийной корреляции, не считая стойкие неизмеряемые индивидуальные эффекты. В их модели использовались следующие переменные:

опыт работы = возраст – число лет обучения – 5,

число лет обучения,

плохое здоровье = фиктивная переменная, характеризующая общее состояние здоровья,

раса = фиктивная переменная, обозначающая небелых (70 из 750 наблюдений),

профсоюз = фиктивная переменная, обозначающая членство в профсоюзе,

безработные = фиктивная переменная, обозначающая, что индивид был безработным в предыдущий год.

Модель также включает постоянный член и дамми на период. [Код последнего не приводится, однако два различных значения, включая 0 за 1968 г. и 1 за 1972 г., дадут тождественные результаты. (Почему?)]

Основное внимание в исследовании сосредоточено на значении коэффициента при числе лет обучения в уравнении логарифма заработной платы. Поскольку число лет обучения и, возможно, опыт работы и безработные коррелированы со скрытым эффектом, в традиционных оценках данного уравнения с большой долей вероятности присутствует серьезное смещение. В табл. 11.11 представлены некоторые результаты, полученные авторами. Результаты оценок МНК и GLS со случайными эффектами в первых двух столбцах задают стандарт для последующего анализа. Коэффициент при числе лет обучения оценен как 0,0669, что намного меньше значения, которое ожидали получить авторы. Как мы видели ранее, даже при наличии корреляции между измеряемыми и скрытыми эффектами в этой модели LSDV-оценка позволяет получить состоятельную оценку коэффициентов при изменяющихся во времени переменных. Таким образом, мы можем использовать ее в **тесте на спецификацию Хаусмана** для проверки наличия корреляции между включенными переменными и скрытой неоднородностью. Вычисления приведены в подразделе 11.5.4, в (11-42). Поскольку в LSDV-уравнении остаются три переменные, статистика хи-квадрат имеет три степени свободы. Приводимое значение 20,2 намного больше 95%-го критического значения 7,81, поэтому результаты показывают, что модель со случайными эффектами неверно специфицирована.

Хаусман и Тейлор продолжили переоценивать уравнение логарифма заработной платы, используя предложенную ими оценку. Четвертый и пятый столбцы с результатами в табл. 11.11 представляют оценки инструментальных переменных. Тест на спецификацию, приведенный в четвертом столбце, показывает, что процедура дала ожидаемый результат. Гипотеза о модифицированной модели со случайными эффектами теперь не отвергается; значение хи-квадрат, равное 2,24, намного меньше

критического. Переменная числа лет обучения рассматривается в качестве эндогенной (коррелированной с u_i) в обоих случаях. Различие между двумя случаями заключается в трактовке переменных «безработные» и «опыт работы». В более предпочтительном уравнении они включены в x_2 , а не в x_1 . Конечным результатом упражнения снова является коэффициент при числе лет обучения, который вырос с 0,0669 в худшей спецификации (МНК) до 0,2169 в последней, прирост составил более 200%. Как отмечают авторы, в то же время измеряемый эффект от расовой принадлежности практически исчезает.

Таблица 11.11. Оценки уравнения логарифма зарплаты

| | Переменные | МНК | GLS/RE | LSDV | HT/IV-GLS | HT/IV-GLS |
|-------|--|-----------------------|----------|----------|-----------|-----------|
| x_1 | Опыт | 0,0132 | 0,0133 | 0,0241 | 0,0217 | |
| | работы | (0,0011) ^a | (0,0017) | (0,0042) | (0,0031) | |
| | Плохое | -0,0843 | -0,0300 | -0,0388 | -0,0278 | -0,0388 |
| | здоровье | (0,0412) | (0,0363) | (0,0460) | (0,0307) | (0,0348) |
| | Безработные | -0,0015 | -0,0402 | -0,0560 | -0,0559 | |
| | | (0,0267) | (0,0207) | (0,0295) | (0,0246) | |
| | Время | NR ^b | NR | NR | NR | R |
| x_2 | Опыт | | | | | 0,0241 |
| | работы | | | | | |
| | Безработные | | | | | -0,0560 |
| z_1 | Раса | -0,0853 | -0,0878 | | -0,0278 | -0,0175 |
| | | (0,0328) | (0,0518) | | (0,0752) | (0,0764) |
| | Профсоюз | 0,0450 | 0,0374 | | 0,1227 | 0,2240 |
| | | (0,0191) | (0,0296) | | (0,0473) | (0,2863) |
| | Обучение | 0,0669 | 0,0676 | | | |
| | | (0,0033) | (0,0052) | | | |
| z_2 | Константа | NR | NR | NR | NR | R |
| | Обучение | | | | 0,1246 | 0,2169 |
| | | | | | (0,0434) | (0,0979) |
| | σ_ϵ | 0,321 | 0,192 | 0,160 | 0,190 | 0,629 |
| | $\rho =$ | | 0,632 | | 0,661 | 0,817 |
| | $\sqrt{\sigma_u^2 / (\sigma_u^2 + \sigma_\epsilon^2)}$ | | | | | |
| | Тест на | | 20,2 | | 2,24 | 0,00 |
| | спец. [3] | | | | | |

^a В скобках приведены оцененные асимптотические стандартные ошибки.

^b NR означает, что оценка коэффициента не была приведена в исследовании.

11.8.2. Состоятельная оценка динамических моделей панельных данных: IV-оценка Андерсона-Хсяо

Рассмотрим однородную динамическую модель панельных данных

$$y_{it} = \gamma y_{i,t-1} + x'_{it} \beta + c_i + \epsilon_{it}, \quad (11-63)$$

где c_i — индивидуальная неизмеряемая неоднородность, как и в предыдущих разделах этой главы, которая может быть (или нет) коррелирована с x_{it} . Мы рассматриваем методы оценки такой модели, когда T зафиксировано и относительно мало, а n может быть большим и возрастающим.

Объединенный МНК, очевидно, дает несостоятельные оценки. Перепишем (11-63) как

$$y_{it} = \gamma y_{i,t-1} + \mathbf{x}'_{it} \boldsymbol{\beta} + w_{it}.$$

Шок в этой объединенной регрессии может коррелировать с x_{it} , но в любом случае он обязательно коррелирован с $y_{i,t-1}$. После подстановки получим

$$\text{Cov}[y_{i,t-1}, (c_i + \varepsilon_{it})] = \sigma_c^2 + \gamma \text{Cov}[y_{i,t-2}, (c_i + \varepsilon_{it})]$$

и т.д. При повторной подстановке можно увидеть, что при $|\gamma| < 1$ и умеренно большом T

$$\text{Cov}[y_{i,t-1}, (c_i + \varepsilon_{it})] \approx \sigma_c^2 / (1 - \gamma). \quad (11-64)$$

[Будет полезно получить такой результат по другому направлению. Если статистический процесс, образующий (y_{it}, c_i) , стационарен, то $\text{Cov}[y_{i,t-1}, c_i] = \text{Cov}[y_{i,t-2}, c_i]$, из чего мы могли бы непосредственно получить (11-64). Для стационарности потребуется допущение о $|\gamma| < 1$.] Следовательно, МНК и GLS несостоятельны. Модель с фиксированными эффектами также не решает эту проблему. Если найти отклонения от индивидуальных средних значений, то получим

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' \boldsymbol{\beta} + \gamma (y_{i,t-1} - \bar{y}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i).$$

Андерсон и Хсяо (Anderson, Hsiao, 1981, 1982) показали, что

$$\begin{aligned} \text{Cov}[(y_{it} - \bar{y}_i), (\varepsilon_{it} - \bar{\varepsilon}_i)] &\approx \frac{-\sigma_\varepsilon^2}{T(1-\gamma)^2} \left[\frac{(T-1) - T\gamma + \gamma^T}{T} \right] \\ &= \frac{-\sigma_\varepsilon^2}{T(1-\gamma)^2} \left[(1-\gamma) - \frac{1-\gamma^T}{T} \right]. \end{aligned}$$

Последнее выражение сходится к нулю с увеличением T , однако мы снова рассматриваем случаи, в которых T мало или умеренно, скажем от 5 до 15, а в этой ситуации смещение в МНК-оценке могло бы составить от 15 до 60%. Из этого следует заключение, что «внутри» преобразование не позволяет получить состоятельную оценку.

Легко заметить, что использование первых разностей также неэффективно. Первые разности для наших наблюдений имеют вид

$$y_{it} - y_{i,t-1} = (\mathbf{x}_{it} - \mathbf{x}_{i,t-1})' \boldsymbol{\beta} + \gamma (y_{i,t-1} - y_{i,t-2}) + (\varepsilon_{it} - \varepsilon_{i,t-1}). \quad (11-65)$$

Как и прежде, корреляция между последним регрессором и шоком сохранится, поэтому оценки по МНК или GLS, в основе которых лежат первые разности, также будут несостоятельны. Существует другой подход. Запишем регрессию в измененной форме как

$$\Delta y_{it} = \Delta \mathbf{x}'_{it} \boldsymbol{\beta} + \gamma \Delta y_{i,t-1} + \Delta \varepsilon_{it}$$

или, определяя $\mathbf{x}_{it}^* = [\Delta x_{it}, \Delta y_{i,t-1}]$, $\varepsilon_{it}^* = \Delta \varepsilon_{it}$ и $\theta = [\beta', \gamma]'$

$$y_{it}^* = \mathbf{x}_{it}^{*\prime} \theta + \varepsilon_{it}^*.$$

Для объединенной выборки, начиная с $t = 3$, запишем

$$\mathbf{y}^* = \mathbf{X}^* \theta + \varepsilon^*.$$

Оценка методом наименьших квадратов, основанная на данных в первых разностях, имеет вид

$$\begin{aligned} \hat{\theta} &= \left[\frac{1}{n(T-3)} \mathbf{X}^{*\prime} \mathbf{X}^* \right]^{-1} \left(\frac{1}{n(T-3)} \mathbf{X}^{*\prime} \mathbf{y}^* \right) \\ &= \theta + \left[\frac{1}{n(T-3)} \mathbf{X}^{*\prime} \mathbf{X}^* \right]^{-1} \left(\frac{1}{n(T-3)} \mathbf{X}^{*\prime} \varepsilon^* \right). \end{aligned}$$

При допущении о том, что обратная матрица в скобках сходится к положительно определенной матрице — что еще предстоит показать, — в этой оценке возникает несостоятельность, поскольку вектор в скобках не сходится к нулю. Последним элементом является $\text{plim}_{n \rightarrow \infty} [1/(n(T-3))] \sum_{i=1}^n \sum_{t=3}^T (y_{i,t-1} - y_{i,t-2})(\varepsilon_{it} - \varepsilon_{i,t-1})$, который не равен нулю.

Предположим, что у нас была бы переменная \mathbf{z}^* такая, что $\text{plim} [1/(n(T-3))] \mathbf{z}^{*\prime} \varepsilon^* = 0$ и $\text{plim} [1/(n(T-3))] \mathbf{z}^{*\prime} \mathbf{X}^* \neq 0$. Пусть $\mathbf{Z} = [\Delta \mathbf{X}, \mathbf{z}^*]$; z_{it}^* заменит $\Delta y_{i,t-1}$ в \mathbf{x}_{it}^* . С таким построением мы, кажется, получим состоятельную оценку. Рассмотрим

$$\begin{aligned} \hat{\theta}_{IV} &= (\mathbf{Z}' \mathbf{X}^*)^{-1} \mathbf{Z}' \mathbf{y}^* \\ &= (\mathbf{Z}' \mathbf{X}^*)^{-1} \mathbf{Z}' (\mathbf{X}^* \theta + \varepsilon^*) \\ &= \theta + (\mathbf{Z}' \mathbf{X}^*)^{-1} \mathbf{Z}' \varepsilon^*. \end{aligned}$$

Далее после умножения на $1/(n(T-3))$, как и раньше, мы найдем

$$\text{plim} \hat{\theta}_{IV} = \theta + \text{plim} \{ [1/(n(T-3))] (\mathbf{Z}' \mathbf{X}^*) \}^{-1} \times 0,$$

что, как видно, решает проблему состоятельности оценки.

Переменная z^* — **инструментальная переменная**, а оценка — **оценка инструментальных переменных** (отсюда индекс у оценки). Нахождение подходящих, обоснованных инструментов, т.е. переменных, удовлетворяющих необходимым допущениям, в моделях, где переменные в правой части коррелированы с пропущенными факторами, часто представляет сложность. В данной ситуации существует подходящий вариант, но на самом деле их несколько. Из (11-65) в период $t = 3$ мы имеем

$$y_{i3} - y_{i2} = (\mathbf{x}_{i3} - \mathbf{x}_{i2})' \beta + \gamma(y_{i2} - y_{i1}) + (\varepsilon_{i3} - \varepsilon_{i2}).$$

Мы могли бы использовать y_{i1} как требуемую переменную, так как она не коррелирована с $\varepsilon_{i3} - \varepsilon_{i2}$. Продолжая в том же направлении, мы видим, что для $t = 3, 4, \dots, T$ $y_{i,t-2}$, видимо, удовлетворяет всем требованиям. Аналогично, начиная с периода $t = 4$, мы можем видеть, что $z_{it} = (y_{i,t-2} - y_{i,t-3})$

снова удовлетворяет нашим требованиям. Это и есть результат Андерсона и Хсяо (Anderson, Hsiao, 1981) для оценки динамической модели панельных данных с инструментальными переменными. Теперь под вопросом то, какой подход, в уровнях ($y_{i,t-2}, t = 3, \dots, T$) или в разностях ($y_{i,t-2} - y_{i,t-3}, t = 4, \dots, T$), является более предпочтительным. Ареллано (Arellano (1989)) и Кивье (Kiviet (1995)) получили результаты, предполагающие, что оценка в уровнях более эффективна.

11.8.3. Эффективная оценка динамических моделей панельных данных: метод Ареллано–Бонда

Ведущее направление в приложениях современных методов, рассматриваемых в данной главе, представлено **динамической моделью панельных данных**, которую мы запишем как

$$y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \delta y_{i,t-1} + c_i + \varepsilon_{it}.$$

В примере 11.21 описаны некоторые примеры ее приложения. Основными допущениями модели являются:

- 1) строгая экзогенность: $E[\varepsilon_{it} | \mathbf{X}_i, c_i] = 0,$
- 2) гомоскедастичность: $E[\varepsilon_{it}^2 | \mathbf{X}_i, c_i] = \sigma_\varepsilon^2,$
- 3) отсутствие автокорреляции: $E[\varepsilon_{it}\varepsilon_{is} | \mathbf{X}_i, c_i] = 0,$ если $t \neq s,$
- 4) некоррелированные наблюдения: $E[\varepsilon_{it}\varepsilon_{js} | \mathbf{X}_i, c_i, \mathbf{X}_j, c_j] = 0$ для $i \neq j$ и

для всех t и $s,$

где строки матрицы данных \mathbf{X}_i размером $T \times K$ — это \mathbf{x}'_{it} . Мы не будем допускать независимость средних значений. «Эффекты» могут быть фиксированными или случайными, поэтому мы допустим

$$E[c_i | \mathbf{X}_i] = g(\mathbf{X}_i).$$

(См. подраздел 11.2.1.) Мы также предположим для удобства, что число периодов фиксировано и равно T . Анализ здесь (и в литературе по теме) можно модифицировать, с тем чтобы распространить результаты на несбалансированные панели, но это несколько неудобно. (Это включает расстановку нулей в разных местах матриц данных, определенных ниже, и изменение периодических показателей в суммировании от 1 до T .)

Наличие запаздывающей зависимой переменной в этой модели является существенным препятствием для ее оценки. Сперва рассмотрим прямое приложение предположения А.13 из раздела 8.2. Составной шок в модели имеет вид $(c_i + \varepsilon_{it})$. Корреляция между $y_{i,t-1}$ и $(c_i + \varepsilon_{i,t})$, очевидно, ненулевая, так как $y_{i,t-1} = \mathbf{x}'_{i,t-1}\boldsymbol{\beta} + \delta y_{i,t-2} + c_i + \varepsilon_{i,t-1}$:

$$\text{Cov}[y_{i,t-1}, (c_i + \varepsilon_{it})] = \sigma_c^2 + \delta \text{Cov}[y_{i,t-2}, (c_i + \varepsilon_{it})].$$

Если T большое и $0 < \delta < 1$, тогда эта ковариация будет равна примерно $\sigma_c^2/(1 - \delta)$. В большинстве случаев допущение о большом T соблюдаться не будет. Однако, поскольку δ , как правило, будет положительным, мы можем ожидать, что эта ковариация будет как минимум больше, чем σ_c^2 . Из этого

следует, что обе оценки по (объединенному) МНК и GLS в этой модели будут несостоятельными. В отличие от случая статистической модели ($\delta = 0$) модель с фиксированными эффектами не решает эту проблему. Используя разности групповых средних, мы получим

$$y_{i,t} - \bar{y}_i = (\mathbf{x}_{i,t} - \bar{\mathbf{x}}_i)' \beta + \delta(y_{i,t-1} - \bar{y}_i) + (\varepsilon_{i,t} - \bar{\varepsilon}_i).$$

Как показали Андерсон и Хсяо (1981, 1981),

$$\text{Cov}[(y_{i,t-1} - \bar{y}_i), (\varepsilon_{i,t} - \bar{\varepsilon}_i)] \approx \frac{-\sigma_\varepsilon^2 (T-1) - T\delta + \delta^T}{T^2 (1-\delta)^2}.$$

Такой результат будет $O(1/T)$, что обычно не составляет проблемы, если асимптотики в нашей модели учитывают возрастающее T . Однако в этой модели панельных данных предполагается, что T фиксировано и относительно мало. Для обычных значений T , скажем от 5 до 15, соразмерное смещение в оценивании δ может быть порядка, скажем, 15–60%.

Ни МНК, ни GLS не будут подходящими оценками. Однако внутри структуры модели есть доступные инструментальные переменные. Андерсон и Хсяо (Anderson, Hsiao (1981, 1982)) предложили подход, основанный на первых разностях, а не на отклонениях от групповых средних:

$$y_{it} - y_{i,t-1} = (\mathbf{x}_{it} - \mathbf{x}_{i,t-1})' \beta + \delta(y_{i,t-1} - y_{i,t-2}) + \varepsilon_{it} - \varepsilon_{i,t-1}.$$

Для первого полного наблюдения

$$y_{i3} - y_{i2} = (\mathbf{x}_{i3} - \mathbf{x}_{i2})' \beta + \delta(y_{i2} - y_{i1}) + \varepsilon_{i3} - \varepsilon_{i2} \quad (11-66)$$

переменная y_{i1} (предполагая, что начальная точка $t = 0$ расположена там, откуда начинается процесс порождения данных) удовлетворяет требованиям, поскольку ε_{i1} заранее предопределена по отношению к $(\varepsilon_{i3} - \varepsilon_{i2})$. [То есть если бы мы использовали данные только для периодов 1 – 3, построенные как в (11-66), то инструментальные переменные для $(y_{i2} - y_{i1})$ были бы $\mathbf{z}_{i(3)}$, где $\mathbf{z}_{i(3)} = (y_{1,1}, y_{2,1}, \dots, y_{n,1})$ для n наблюдений.] Для следующего наблюдения

$$y_{i4} - y_{i3} = (\mathbf{x}_{i4} - \mathbf{x}_{i3})' \beta + \delta(y_{i3} - y_{i2}) + \varepsilon_{i4} - \varepsilon_{i3}$$

обе переменные, y_{i2} и $(y_{i2} - y_{i1})$, доступны.

На основании предыдущего параграфа можно было бы подумать, что здесь вместо недостатка инструментов существует, по сути, их избыток. В данном ограниченном направлении исследования перед нами стоит выбор между разностями и уровнями. На самом деле мы могли бы применить обе техники и, более того, в любом периоде, начиная с четвертого, в качестве инструмента доступен не только y_{i2} , но и y_{i1} и т.д. Это является существенным замечанием в дополнение к методам оценки Ареллано, Бовера и Бонда (Arellano, Bover, Bond (1991, 1995)), основанным на крайне большом числе претендентов на роль инструментальных переменных в модели панельных данных. Начнем с того, что в модели в форме первых разностей для $y_{i3} - y_{i2}$ доступна переменная y_{i1} . Для $y_{i4} - y_{i3}$ доступны обе, y_{i1} и y_{i2} ; для $y_{i5} - y_{i4}$ мы

имеем y_{i1} , y_{i2} и y_{i3} и т.д. Обратим внимание и на то, что мы не использовали экзогенные переменные. При строго экзогенных регрессорах в качестве инструментов доступны не только все запаздывающие значения y_{is} от s до $t-1$, но и все значения x_{it} . Например, для $y_{i4} - y_{i3}$ претендентами являются y_{i1} , y_{i2} и $(x'_{i1}, x'_{i2}, \dots, x'_{iT})$ для всех T периодов. Число претендентов на роль инструмента на самом деле крайне велико. [См. более детальный анализ у Ahn, Schmidt (1995).] Если экзогенные переменные лишь predeterminedены, а не строго экзогенны, тогда только $E[\varepsilon_{it} | x_{i,t}, x_{i,t-1}, \dots, x_{i1}] = 0$ и только векторы x_{is} от 1 до $t-1$ будут являться обоснованными (допустимыми) инструментами в уравнении в разностях, содержащем $\varepsilon_{it} - \varepsilon_{i,t-1}$. [См. пример у Baltagi, Levin (1986).] Едва ли это можно назвать ограничением, учитывая то, что в конечном счете для модели среднего размера нам, возможно, придется рассматривать сотни или тысячи потенциальных инструментальных переменных для оценки, как правило, небольшого числа параметров.

Теперь сформулируем модель в более знакомой нам форме, дабы иметь возможность применить оценку инструментальных переменных. В терминах данных в разностях основное уравнение будет иметь следующий вид:

$$y_{it} - y_{i,t-1} = (x_{it} - x_{i,t-1})' \beta + \delta(y_{i,t-1} - y_{i,t-2}) + \varepsilon_{it} - \varepsilon_{i,t-1},$$

или

$$\Delta y_{it} = (\Delta x_{it})' \beta + \delta(\Delta y_{i,t-1}) + \Delta \varepsilon_{it}, \quad (11-67)$$

где Δ — оператор первых разностей, $\Delta a_t = a_t - a_{t-1}$ для любого временно-го ряда (или вектора) a_t . (Следует отметить, что постоянный член и любые не изменяющиеся во времени переменные в x_{it} будут выпадать из первых разностей. Мы восстановим их далее после выведения оценки для β .) Параметры модели, подлежащие оценке, — это $\theta = (\beta', \delta)'$ и σ_ε^2 . Для удобства запишем модель так:

$$\tilde{y}_{it} = \tilde{x}'_{it} \theta + \tilde{\varepsilon}_{it}.$$

Мы намерены определить оценку инструментальных переменных по образцу (8-9) и (8-10). Так как наши данные являются панельными, аналогом для

$$Z' \tilde{X} = \sum_{i=1}^n z_i \tilde{x}'_i \quad (11-68)$$

в случае межобъектных данных будет

$$Z' \tilde{X} = \sum_{i=1}^n \sum_{t=3}^T z_{it} \tilde{x}'_{it} = \sum_{i=1}^n Z'_i \tilde{X}'_i, \quad (11-69)$$

$$\tilde{y}_i = \begin{bmatrix} \Delta y_{i3} \\ \Delta y_{i4} \\ \vdots \\ \Delta y_{iT_i} \end{bmatrix}, \quad \tilde{X}_i = \begin{bmatrix} \Delta x'_{i3} & \Delta y_{i2} \\ \Delta x'_{i4} & \Delta y_{i3} \\ \dots & \dots \\ \Delta x'_{iT} & \Delta y_{i,T-1} \end{bmatrix},$$

где есть $(T-2)$ наблюдений (строк) и $K+1$ столбцов в \tilde{X}_i . Однако это несколько осложняется тем, что число инструментов, которое мы определили, может изменяться в зависимости от периода, поэтому в матричных вычислениях в (11-69) суммироваться будут матрицы разной размерности.

Рассмотрим альтернативный подход. Если бы мы использовали только первые полные наблюдения, определенные в (11-67), тогда применялся бы межобъектный вариант, а множество инструментов Z в (11-68) со строго экзогенными переменными приобрело бы вид матрицы размером $n \times (1 + KT)$

$$Z_{(3)} = \begin{bmatrix} y_{1,1}, x'_{1,1}, x'_{1,2}, \dots, x'_{1,T} \\ y_{2,1}, x'_{2,1}, x'_{2,2}, \dots, x'_{2,T} \\ \vdots \\ y_{n,1}, x'_{n,1}, x'_{n,2}, \dots, x'_{n,T} \end{bmatrix}$$

и оценка (8-9) инструментальных переменных была бы основана на

$$\tilde{X}_{(3)} = \begin{bmatrix} x'_{1,3} - x'_{1,2} & y_{1,4} - y_{1,3} \\ x'_{2,3} - x'_{2,2} & y_{2,4} - y_{2,3} \\ \vdots & \vdots \\ x'_{n,3} - x'_{n,2} & y_{n,4} - y_{n,3} \end{bmatrix} \text{ и } \tilde{Y}_{(3)} = \begin{bmatrix} y_{1,3} - y_{1,2} \\ y_{2,3} - y_{2,2} \\ \vdots \\ y_{n,3} - y_{n,2} \end{bmatrix}.$$

Индексом (3) обозначено первое наблюдение, используемое для левой части уравнения. Не обращая внимания на остальные наблюдения, мы могли бы затем использовать эти данные для вывода IV-оценки в (8-9), которую мы здесь обозначим как $\hat{\theta}_{IV(3)}$. Теперь повторим построение с использованием следующего (четвертого) наблюдения как первого и, кроме того, с использованием лишь одного года из панели. Теперь матрицы данных имеют вид

$$\tilde{X}_{(4)} = \begin{bmatrix} x'_{1,4} - x'_{1,3} & y_{1,3} - y_{1,2} \\ x'_{2,4} - x'_{2,3} & y_{2,3} - y_{2,2} \\ \vdots & \vdots \\ x'_{n,4} - x'_{n,3} & y_{n,3} - y_{n,2} \end{bmatrix}, \tilde{Y}_{(4)} = \begin{bmatrix} y_{1,4} - y_{1,3} \\ y_{2,4} - y_{2,3} \\ \vdots \\ y_{n,4} - y_{n,3} \end{bmatrix} \text{ и} \quad (11-70)$$

$$Z_{(4)} = \begin{bmatrix} y_{1,1}, y_{1,2}, x'_{1,1}, x'_{1,2}, \dots, x'_{1,T} \\ y_{2,1}, y_{2,2}, x'_{2,1}, x'_{2,2}, \dots, x'_{2,T} \\ \vdots \\ y_{n,1}, y_{n,2}, x'_{n,1}, x'_{n,2}, \dots, x'_{n,T} \end{bmatrix},$$

и мы получаем вторую IV-оценку, $\hat{\theta}_{IV(4)}$, также основанную на n наблюдениях, но теперь с $2 + KT$ инструментами. И т.д.

Теперь нам необходимо согласовать $T - 2$ оценки для θ , которые мы построили: $\hat{\theta}_{IV(3)}, \hat{\theta}_{IV(4)}, \dots, \hat{\theta}_{IV(T)}$. Мы сталкивались с этой проблемой в подразделе 11.5.8, где рассматривалась модель с фиксированными эффектами Чемберлена. Оценка, основанная на минимизации расстояния, предложенная тогда и используемая в исследовании больничных затрат у Carey (1997) в примере 11.10, представляет собой средство эффективного «усреднения» многочисленных оценок вектора параметров. Мы вернемся к MDE в главе 13. Сейчас же мы вместо этого рассмотрим метод Ареллано–Бонда (Arellano, Bond (1991)) и Ареллано–Бовера (Arellano, Bover (1995)) применительно к этой проблеме. Мы соберем полный набор оценок по аналогии с (11-56) и (11-57). Во-первых, сведем наборы инструментов в единую матрицу Z ,

где для каждого отдельного объекта мы получим матрицу Z_i размером $(T - 2) \times L$. Описание строк в Z_i зависит от того, являются ли регрессоры строго экзогенными или predetermined. Для строго экзогенных переменных

$$Z_i = \begin{bmatrix} y_{i,1}, x'_{i,1}, x'_{i,2}, \dots, x'_{i,T} & 0 & \dots & 0 \\ 0 & y_{i,1}, y_{i,2}, x'_{i,1} & \dots & 0 \\ \dots & x'_{i,2}, \dots, x'_{i,T} & \dots & \dots \\ 0 & 0 & \dots & y_{i,1}, y_{i,2}, \dots, y_{i,T-2}, \\ & & & x'_{i,1}, x'_{i,2}, \dots, x'_{i,T} \end{bmatrix} \quad (11-71a)$$

$$\text{и } L = \sum_{i=1}^{T-2} (i + TK) = (T-2)(T-1)/2 + (T-2)TK.$$

Для predetermined переменных матрица инструментальных переменных имеет вид

$$Z_i = \begin{bmatrix} y_{i,1}, x'_{i,1}, x'_{i,2} & 0 & \dots & 0 \\ 0 & y_{i,1}, y_{i,2}, x'_{i,1}, x'_{i,2}, x'_{i,3} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & y_{i,1}, y_{i,2}, \dots, y_{i,T-2}, x'_{i,1}, \\ & & & x'_{i,2}, \dots, x'_{i,T-1} \end{bmatrix} \quad (11-71b)$$

$$\text{и } L = \sum_{i=1}^{T-2} (i(K+1) + K) = [(T-2)(T-1)/2](1+K) + (T-2)K.$$

Такое построение быстро увеличивает число инструментов (моментные условия, как мы увидим в главе 13). В приложении из примера 11.15 у нас есть небольшая панель только с $T = 7$ периодами и мы строим модель лишь с $K = 4$ регрессорами в x_{it} плюс запаздывающая зависимая переменная. Допущение о строгой экзогенности дает в данном случае матрицу Z_i размером (5×135) . С одним лишь допущением о predetermined x_{it} Z_i немного уменьшается — до (5×95) . Для иллюстрации мы использовали только два предшествующих наблюдения x_{it} . Это сокращает матрицу до

$$Z_i = \begin{bmatrix} y_{i,1}, x'_{i,1}, x'_{i,2} & 0 & \dots & 0 \\ 0 & y_{i,1}, y_{i,2}, x'_{i,2}, x'_{i,3} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & y_{i,1}, y_{i,2}, \dots, y_{i,T-2}, \\ & & & x'_{i,T-2}, x'_{i,T-1} \end{bmatrix}, \quad (11-71c)$$

которая теперь имеет размерность (5×55) при $T = 7$ и $K = 4$. [Baltagi (2005, глава 8) представил некоторые альтернативные конфигурации Z_i , которые допускают смешение строго экзогенных и predetermined переменных.]

Теперь мы можем вычислить оценку двухшаговым методом наименьших квадратов в (11-10), используя наши определения матриц данных Z_i , \tilde{X}_i и \tilde{y}_i и (11-69). Получим

$$\hat{\theta}_{IV} = \left[\left(\sum_{i=1}^n \tilde{X}'_i Z_i \right) \left(\sum_{i=1}^n Z'_i Z_i \right)^{-1} \left(\sum_{i=1}^n Z'_i \tilde{X}_i \right) \right]^{-1} \times \\ \times \left[\left(\sum_{i=1}^n \tilde{X}'_i Z_i \right) \left(\sum_{i=1}^n Z'_i Z_i \right)^{-1} \left(\sum_{i=1}^n Z'_i \tilde{y}_i \right) \right]. \quad (11-72)$$

Естественной оценкой асимптотической ковариационной матрицы для оценки является

$$\text{Est.Asy.Var} [\hat{\theta}_{IV}] = \hat{\sigma}_{\Delta\varepsilon}^2 \left[\left(\sum_{i=1}^n \tilde{\mathbf{X}}_i' \mathbf{Z}_i \right) \left(\sum_{i=1}^n \mathbf{Z}_i' \mathbf{Z}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{Z}_i' \mathbf{X}_i \right) \right]^{-1}, \quad (11-73)$$

где

$$\hat{\sigma}_{\Delta\varepsilon}^2 = \frac{\sum_{i=1}^n \sum_{t=3}^T [(y_{it} - y_{i,t-1}) - (\mathbf{x}_{it} - \mathbf{x}_{i,t-1})' \hat{\beta} - \hat{\delta}(y_{i,t-1} - y_{i,t-2})]^2}{n(T-2)}. \quad (11-74)$$

Однако эта оценка дисперсии, скорее всего, будет недооценивать истинную асимптотическую дисперсию, так как между наблюдениями есть автокорреляция первого порядка. Поскольку $(y_{it} - y_{i,t-1}) = \tilde{\mathbf{x}}_{it}' \boldsymbol{\theta} + (\varepsilon_{it} - \varepsilon_{i,t-1}) = \tilde{\mathbf{x}}_{it}' \boldsymbol{\theta} + v_{it}$,

$$\text{Cov}[v_{it}, v_{i,t-1}] = \text{Cov}[v_{it}, v_{i,t+1}] = -\sigma_\varepsilon^2.$$

Ковариации при больших лагах или опережениях равны нулю. В модели в разностях, несмотря на то что ковариационная матрица шоков не равна $\sigma_\varepsilon^2 \mathbf{I}$, она принимает достаточно простой вид:

$$\text{Cov} \begin{pmatrix} \varepsilon_{i,3} - \varepsilon_{i,2} \\ \varepsilon_{i,4} - \varepsilon_{i,3} \\ \varepsilon_{i,5} - \varepsilon_{i,4} \\ \dots \\ \varepsilon_{i,T} - \varepsilon_{i,T-1} \end{pmatrix} = \sigma_\varepsilon^2 \begin{bmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & \dots & 0 \\ 0 & -1 & 2 & \dots & 0 \\ \dots & \dots & -1 & \dots & -1 \\ 0 & 0 & \dots & -1 & 2 \end{bmatrix} = \sigma_\varepsilon^2 \boldsymbol{\Omega}_i. \quad (11-75)$$

Из этого следует то, что оценка (11-74) оценивает не σ_ε^2 , а $2\sigma_\varepsilon^2$. Однако простое деление оценки на два не позволяет получить корректную асимптотическую ковариационную матрицу, так как между самими наблюдениями есть автокорреляция. По сути, матрица в (11-73) является неподходящей. (Мы сталкивались с этой проблемой в теореме 9.1 и в подразделах 9.2.3, 9.4.3 и 11.3.2.) Необходимая корректировка может быть основана на аналоге оценки Уайта, который мы выводили в (11-3). Для простоты пусть

$$\hat{\mathbf{A}} = \left[\left(\sum_{i=1}^n \tilde{\mathbf{X}}_i' \mathbf{Z}_i \right) \left(\sum_{i=1}^n \mathbf{Z}_i' \mathbf{Z}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{Z}_i' \tilde{\mathbf{X}}_i \right) \right]^{-1}.$$

Тогда устойчивая к автокорреляции ковариационная матрица имеет вид

$$\hat{\mathbf{A}} \left[\left(\sum_{i=1}^n \tilde{\mathbf{X}}_i' \mathbf{Z}_i \right) \left(\sum_{i=1}^n \mathbf{Z}_i' \mathbf{Z}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{Z}_i' \hat{\mathbf{v}}_i \hat{\mathbf{v}}_i' \mathbf{Z}_i \right) \left(\sum_{i=1}^n \mathbf{Z}_i' \mathbf{Z}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{Z}_i' \tilde{\mathbf{X}}_i \right) \right] \hat{\mathbf{A}}. \quad (11-76)$$

[Можно было бы также заменить $\hat{\mathbf{v}}_i \hat{\mathbf{v}}_i'$ в (11-76) на $\hat{\sigma}_\varepsilon^2 \boldsymbol{\Omega}_i$ в (11-75), поскольку это является известным ожиданием.]

Сделаем небольшое отступление и рассмотрим оценку из (11-72). Вычисления здесь менее громоздкие, чем можно было представить. Заметим, что

строки \mathbf{Z}_i в (11-71a, b, c) ортогональны. Из этого следует, что матрица

$$\mathbf{F} = \sum_{i=1}^n \mathbf{Z}'_i \mathbf{Z}_i$$

в (11-72) блочно-диагональна с $T - 2$ блоками. Отдельные блоки в \mathbf{F} имеют вид

$$\mathbf{F}_t = \sum_{i=1}^n \mathbf{z}_{it} \mathbf{z}'_{it} = \mathbf{Z}'_{(t)} \mathbf{Z}_{(t)}$$

для $t = 3, \dots, T$. Поскольку число инструментов для каждого периода разное — см. (11-71), — эти блоки разного размера, скажем $(L_t \times L_t)$. Аналогичная конструкция показывает, что матрица $\sum_{i=1}^n \tilde{\mathbf{X}}'_i \mathbf{Z}_i$ на самом деле представляет собой разделенную на блоки матрицу вида

$$\sum_{i=1}^n \tilde{\mathbf{X}}'_i \mathbf{Z}_i = \left[\tilde{\mathbf{X}}'_{(3)} \mathbf{Z}_{(3)} \quad \tilde{\mathbf{X}}'_{(4)} \mathbf{Z}_{(4)} \quad \dots \quad \tilde{\mathbf{X}}'_{(T)} \mathbf{Z}_{(T)} \right],$$

где матрицы снова имеют разные размеры; в каждой из них $T - 2$ строк, но число столбцов разное. Из этого следует, что обратная матрица, $(\sum_{i=1}^n \mathbf{Z}'_i \mathbf{Z}_i)^{-1}$, также блочно-диагональна, а квадратичная форма матрицы в (11-72) может быть записана как

$$\begin{aligned} & \left(\sum_{i=1}^n \tilde{\mathbf{X}}'_i \mathbf{Z}_i \right) \left(\sum_{i=1}^n \tilde{\mathbf{Z}}'_i \mathbf{Z}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{Z}'_i \tilde{\mathbf{X}}_i \right) = \sum_{t=3}^T \left(\tilde{\mathbf{X}}'_{(t)} \mathbf{Z}_{(t)} \right) \left(\mathbf{Z}'_{(t)} \mathbf{Z}_{(t)} \right)^{-1} \times \\ & \times \left(\mathbf{Z}'_{(t)} \tilde{\mathbf{X}}_{(t)} \right) = \sum_{t=3}^T \left(\hat{\mathbf{X}}'_{(t)} \hat{\mathbf{X}}_{(t)} \right) = \sum_{t=3}^T \mathbf{W}_{(t)} \end{aligned}$$

[см. (8-9) и предыдущий результат]. Продолжая таким же образом, мы найдем

$$\left(\sum_{i=1}^n \tilde{\mathbf{X}}'_i \mathbf{Z}_i \right) \left(\sum_{i=1}^n \tilde{\mathbf{Z}}'_i \mathbf{Z}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{Z}'_i \tilde{\mathbf{y}}_i \right) = \sum_{t=3}^T \hat{\mathbf{X}}'_{(t)} \mathbf{y}_{(t)}.$$

Из (8-10) можно увидеть, что

$$\hat{\mathbf{X}}'_{(t)} \mathbf{y}_{(t)} = \left(\hat{\mathbf{X}}'_{(t)} \hat{\mathbf{X}}_{(t)} \right) \hat{\boldsymbol{\theta}}_{IV}(t) = \mathbf{W}_{(t)} \hat{\boldsymbol{\theta}}_{IV}(t).$$

Соединив все построенные на данный момент члены, мы обнаружим, что оценка в (11-72) может быть записана в виде

$$\hat{\boldsymbol{\theta}}_{IV} = \left(\sum_{t=3}^T \mathbf{W}_{(t)} \right)^{-1} \left(\sum_{t=3}^T \mathbf{W}_{(t)} \hat{\boldsymbol{\theta}}_{IV}(t) \right) = \sum_{t=3}^T \mathbf{R}_{(t)} \hat{\boldsymbol{\theta}}_{IV}(t),$$

где

$$\mathbf{R}_{(t)} = \left(\sum_{t=3}^T \mathbf{W}_{(t)} \right)^{-1} \mathbf{W}_{(t)} \text{ и } \sum_{t=3}^T \mathbf{R}_{(t)} = \mathbf{I}.$$

То есть мы обнаружим, что, как и ожидалось, оценка вектора параметров методом Ареллано–Бонда представляет собой матричное взвешенное среднее из специфических для периода $T - 2$ оценок двухшаговым методом наименьших квадратов, где инструменты, используемые в каждом периоде, могут различаться. Поскольку наша оценка представляет собой среднее значение из оценок, то возникает вопрос: является ли это среднее значение эффективным — выбраны ли веса таким образом, чтобы составить эффективную оценку? Наверное, неудивительно, что ответ для данного $\hat{\theta}$ — нет; существует более эффективный набор весов, который можно построить для этой модели. Мы составим его после того, как рассмотрим обобщенный метод моментов в главе 13.

В итоге получаем некоторую неопределенность. После (11-67) было отмечено, что рассматриваемый способ оценки исключает постоянный член и все не изменяющиеся во времени переменные, возникающие в этой модели. Метод Хаусмана–Тейлора (Hausman, Taylor (1981)), описанный в предыдущем подразделе, позволяет получить средство, с помощью которого модель может быть доработана так, чтобы примириться с такой возможностью. Расширим базовую формулировку, чтобы включить не изменяющиеся во времени эффекты:

$$y_{it} = \mathbf{x}'_{it}\beta + \delta y_{i,t-1} + \alpha + \mathbf{f}'_i\gamma + c_i + \varepsilon_{it},$$

где \mathbf{f}_i — набор не изменяющихся во времени переменных, а γ — вектор параметров, который необходимо оценить. Эта модель полностью согласуется с предшествующим анализом, так как компонента $\alpha + \mathbf{f}'_i\gamma$ выпала бы из уравнения в разностях вместе с c_i на первом шаге в (11-63). После того как мы описали состоятельную оценку для $\theta = (\beta', \delta)'$, обратимся к оценке $(\alpha, \gamma)'$. Остатки из IV-регрессии (11-72)

$$w_{it} = \mathbf{x}'_{it}\hat{\beta}_{IV} - \hat{\delta}_{IV}y_{i,t-1}$$

представляют собой точечные состоятельные оценки для

$$w_{it} = \alpha + \mathbf{f}'_i\gamma + c_i + \varepsilon_{it}.$$

Таким образом, групповые средние значения остатков могут сформировать основу для регрессии, оцениваемой на втором шаге:

$$\bar{w}_i = \alpha + \mathbf{f}'_i\gamma + c_i + \bar{\varepsilon}_i + \eta_i, \quad (11-77)$$

где $\eta_i = (\bar{w}_i - \bar{w}_i.)$ — ошибка оценивания, которая сходится к нулю по мере того, как $\hat{\theta}$ сходится к θ . Следствием этого будет то, что мы теперь сможем построить линейную регрессию этих остатков групповых средних значений на константу и не изменяющиеся во времени переменные \mathbf{f}_i , чтобы оценить α и γ . Недостатком такой стратегии, однако, является то, что исходные допущения модели не предполагают, что c_i не коррелирован с другими переменными модели, включая скрытые не изменяющиеся во времени члены \mathbf{f}_i . Таким образом, оценка методом наименьших квадратов не подходит для использования здесь, если только не предполагается модель со случайными эффектами, чего мы намеренно стремились избежать с самого начала. Как

и в исследовании Хаусмана и Тейлора, существует выполнимая стратегия в том случае, если можно допустить, что в модели есть несколько переменных, вероятно, включая некоторые из f_i и другие из x_{it} , которые не коррелированы с c_i и ε_{it} . В оценке по методу Хаусмана и Тейлора это те же z_1 и x_1 (см. шаг 2 в анализе из предыдущего подраздела). Допуская, что эти переменные доступны, в модель следует добавить предположение об идентификации — мы получим удобную оценку инструментальных переменных, в качестве инструментов используя постоянный член (1), любые переменные из f_i , не коррелированные со скрытыми эффектами или шоками (назовем их f_{i1}), и групповые средние значения любых переменных из x_{it} , которые также являются экзогенными. Этого должно хватить, чтобы получить достаточно большое множество инструментов, соответствующих всем параметрам в (11-76а). Это та же идентификация, которую мы видели на шаге 2 в оценке методом Хаусмана-Тейлора; K_1 , число экзогенных переменных в x_{it} , должно быть как минимум таким же, как L_2 , число эндогенных переменных в f_i . С учетом всего этого мы получаем оценку инструментальных переменных, в которой \bar{w}_i — зависимая переменная, $(1, f_i)$ — переменные из правой части, а $(1, f_{i1}, \bar{x}_{i1})$ — инструментальные переменные.

Существует и другое направление, следуя которому мы могли бы расширить этот метод оценки. В (11-76) мы уже косвенным образом допускали, чтобы более общая ковариационная матрица регулировала образование шоков ε_{it} , и вычислили устойчивую ковариационную матрицу для простой IV-оценки. Мы могли бы сделать еще один шаг и осуществить поиск более эффективной оценки. Как показали многочисленные современные исследования, базы панельных данных богаты информацией, позволяющей аналитику специфицировать очень общие модели и использовать предполагаемые взаимосвязи между переменными для построения намного более эффективных оценок обобщенным методом моментов (GMM). [См., в частности, Arellano, Bover (1995) и Blundell, Bond (1998).] Мы вернемся к этому в главе 13.

Пример 11.15. Динамическое уравнение предложения труда

В примере 8.5 мы использовали инструментальные переменные для оценки уравнения предложения труда

$$Wks_{it} = \gamma_1 + \gamma_2 \ln Wage_{it} + \gamma_3 Ed_i + \gamma_4 Union_{it} + \gamma_5 Fem_i + u_{it}.$$

Чтобы пояснить вычисления в данном разделе, мы расширим эту модель следующим образом:

$$Wks_{it} = \beta_1 \ln Wage_{it} + \beta_2 Union_{it} + \beta_3 Occ_{it} + \beta_4 Exp_{it} + \delta Wks_{i,t-1} + \alpha + \gamma_1 Ed_i + \gamma_2 Fem_i + c_i + \varepsilon_{it}.$$

(Мы поменяли переменные и названия параметров для соответствия обозначениям этого раздела.) Заметим, с теоретической точки зрения, как было предложено в примере ранее, рассмотрение переменной $\ln Wage_{it}$ как не коррелированной с ε_{it} и c_i может оказаться необоснованным.

Однако мы будем анализировать эту модель в первых разностях. Вполне допустимо рассматривать изменения заработной платы как экзогенные. Это будет зависеть от теоретических основ модели. Здесь и далее мы будем рассматривать переменную как предопределенную. В модели есть две не изменяющиеся во времени переменные: Fem_i , полностью экзогенная, и Ed_i , которая может быть эндогенной. Требование идентифицируемости для оценки $(\alpha, \gamma_1, \gamma_2)$ соблюдается ввиду наличия трех экзогенных переменных: $Union_{it}$, Occ_{it} и Exp_{it} ($K_1 = 3$ и $L_2 = 1$).

Уравнение в разностях, анализируемое на первом шаге, имеет вид

$$\begin{aligned} \Delta Wks_{it} &= \beta_1 \Delta \ln Wage_{it} + \beta_2 \Delta Union_{it} + \beta_3 \Delta Occ_{it} + \beta_4 \Delta Exp_{it} + \\ &+ \delta \Delta Wks_{i,t-1} + \Delta \varepsilon_{it}. \end{aligned}$$

Мы оценили параметры и асимптотическую ковариационную матрицу в соответствии с (11-72) и (11-76). Для спецификации инструментальных переменных мы использовали одно прошлое наблюдение x_{it} , как показано в тексте²⁶. В табл. 11.12 представлены результаты вычислений с использованием некоторых других несостоятельных оценок.

Различные оценки весьма далеки друг от друга. При отсутствии общих эффектов (и автокорреляции шоков) все пять приведенных оценок будут состоятельными. При крайне больших различиях можно было бы ожидать, что общие эффекты являются важными характеристиками данных. Вторые стандартные ошибки, приведенные вместе с IV-оценками, основаны на неправильной матрице из (11-73) с $\hat{\sigma}_{\Delta \varepsilon}^2$ из (11-74), разделенной на два. Мы обнаружили, что эта оценка весьма нестабильна, что видно в таблице. Оценка также очень чувствительна к выбору инструментов, входящих в состав Z_i . Использование (11-71a) вместо (11-71b) дает сильные колебания в оценках и, в сущности, влечет неправдоподобные результаты. Одним из возможных объяснений в данном конкретном примере является то, что инструментальные переменные, которые мы используем, представляют собой фиктивные переменные с относительно небольшой изменчивостью во времени.

11.8.4. Нестационарные временные ряды и модели панельных данных

Некоторая часть дискуссии, которую мы вели до сих пор, была сфокусирована (и будет сосредоточена далее) на статистических результатах при «малом T ». Применяются панели, содержащие наблюдения при фиксированном и малом T и при большом числе n объектов. В современных исследованиях с использованием данных по разным странам, как в Penn World

²⁶Эта оценка и GMM-оценки в главе 13 встроены в некоторые современные компьютерные программы, включая NLOGIT и Stata. Многие исследователи используют программу Gauss, распространяемую М. Ареллано: <http://www.cemfi.es>, или сами программируют вычисления, используя MatLab или R. Мы запрограммировали матричные вычисления непосредственно для данного примера, используя матричный комплекс в NLOGIT.

Таблица 11.12. Оценки динамической модели панельных данных с использованием оценки Арелано-Бонда

| Переменная | МНК, полное урав. | | МНК, дифференци- рованный | | IV, дифференци- рованный | | Случайные эффекты | | Фиксированные эффекты | |
|--------------------------|----------------------------------|---------------|----------------------------------|---------------|---|--------------------|----------------------------------|---------------|----------------------------------|---------------|
| | Оценка | Станд. ош. | Оценка | Станд. ош. | Оценка | Станд. ош. | Оценка | Станд. ош. | Оценка | Станд. ош. |
| <i>ln Wage</i> | 0,2966 | 0,2052 | -0,1100 | 0,4565 | -1,1402 | 0,2639 0,8768 | 0,2281 | 0,2405 | 0,5886 | 0,4790 |
| <i>Union</i> | -1,2945 | 0,1713 | 1,1640 | 0,4222 | 2,7089 | 0,3684 0,8676 | -1,4104 | 0,2199 | 0,1444 | 0,4369 |
| <i>Occ</i> | 0,4163 | 0,2005 | 0,8142 | 0,3924 | 2,2808 | 1,3105 0,7220 | 0,5191 | 0,2484 | 1,0064 | 0,4030 |
| <i>Exp</i> | -0,0295 | 0,00728 | -0,0742 | 0,0975 | -0,0208 | 0,1126 0,1104 | -0,0353 | 0,01021 | -0,1683 | 0,05954 |
| <i>Wks_{t-1}</i> | 0,3804 | 0,01477 | -0,3527 | 0,01609 | 0,1304 | 0,04760 0,02131 | 0,2100 | 0,01511 | 0,0148 | 0,01705 |
| <i>Constant</i> | 28,918 | 1,4490 | | | -0,4110 | 0,3364 | 37,461 | 1,6778 | | |
| <i>Ed</i> | -0,0690 | 0,03703 | | | 0,0321 | 0,02587 | -0,0657 | 0,04988 | | |
| <i>Fem</i> | -0,8607 | 0,2544 | | | -0,0122 | 0,1554 | -1,1463 | 0,3513 | | |
| <i>Выборка</i> | <i>t</i> = 2-7 <i>n</i> = 595 | | <i>t</i> = 3-7 <i>n</i> = 595 | | <i>t</i> = 3-7; <i>n</i> = 595 Для средних использовалось <i>t</i> = 7 | | <i>t</i> = 2-7 <i>n</i> = 595 | | <i>t</i> = 2-7 <i>n</i> = 595 | |

Tables (http://pwt.econ.upenn.edu/php_site/pwt_index.php), которые на данный момент включают наблюдения примерно по 200 странам в течение более 50 лет, анализируются панели с достаточно большим T , так что временные свойства данных становятся предметом пристального внимания. В частности, выявление и учет нестационарности, являющиеся сейчас традиционной частью анализа одномерных временных рядов (как в главе 21), в настоящее время рассматриваются как допустимые для широкомасштабных постранных исследований, таких как исследования роста доходов, основанные на Penn World Tables, постранные исследования расходов на здравоохранение и анализ паритета покупательной способности.

Анализ длинных панелей, как в литературе по росту и конвергенции, обычно включает динамические модели, такие как

$$y_{it} = \alpha_i + \gamma_i y_{i,t-1} + \mathbf{x}'_{it} \beta_i + \varepsilon_{it}. \quad (11-78)$$

При анализе одного временного ряда, включающего низкочастотные данные по макроэкономическим потокам, таким как доход, потребление, инвестирование, дефицит текущего счета и т.д., долгое время считалось, что оцененные регрессионные взаимосвязи могут быть искажены нестационарностью в данных. Кажущиеся устойчивыми и сильными регрессионные связи могут оказаться полностью ложными и появиться ввиду основных характеристик процессов во временных рядах, а не действительных связей между переменными. Проверка гипотезы о долгосрочных эффектах будет в значительной степени искажаться из-за наличия единичных корней в данных. В настоящее время стало очевидно, что то же влияние и те же вредоносные эффекты можно обнаружить в длинных панельных данных. Кроме того, анализ панельных данных осложняется возможной неоднородностью (гетерогенностью) параметров. Интересующие нас коэффициенты во многих исследованиях по странам представляют собой запаздывающие эффекты, такие как γ_i в (11-77), и в этом случае совершенно ясно, что именно результаты, полученные для нестационарных данных, открыли проблемы оценки и получения статистических выводов. Во многих работах предлагались обоснованные тесты на единичные корни в панельных данных. Три наиболее часто цитируемые из них — это работы Levin, Lin (1992), Im, Pesaran, Shin (2003) и Maddala, Wu (1999).

Существует множество эмпирических материалов о применении методов анализа нестационарных временных рядов в случае панелей, среди которых исследования паритета покупательной способности у Frankel, Rose (1996) и Pedroni (2001), стационарности реальной заработной платы у Fleissig, Strauss (1997), инфляции у Culver, Papell (1997), баланса счета текущих операций у Wu (2000), расходов на здравоохранение у McCoskey, Selden (1998), роста и конвергенции у Sala-i-Martin (1996), урбанизации и производства у McCoskey и Као (1999) и сбережений и инвестиций у Coakely et al. (1996). Обширный перечень работ приводится в книге Baltagi (2005, глава 12).

Коварная проблема возникает при выводе результатов, которые можно было использовать для того, чтобы охарактеризовать свойства оценок модели (11-77). Асимптотические результаты, основанные на большом n и большом T , необязательно получать одновременно, необходима большая осторожность, чтобы вывести асимптотическое поведение полезной статисти-

ки. Как правило, по этому вопросу ссылаются на работы Phillips, Moon (1999, 2000).

Мы вернемся к теме нестационарных данных в главе 21. Это появляющиеся работы, большая часть из которых выходит за рамки данного обсуждения. Мы будем обращаться к некоторым подробным общепризнанным исследованиям, таким как Bannerjee (1999), Smith (2000) и Baltagi, Kao (2000), для получения более подробной информации.

11.9. Нелинейная регрессия с панельными данными

Как это ни удивительно, распространить модели панельных данных на случай нелинейной регрессии непросто. До сих пор для согласования нелинейной модели мы, как правило, применяли уже известные результаты к линеаризованной регрессии. Такой подход мы будем переносить на случай кластеризованных данных. (См. подраздел 11.3.3.) К сожалению, его нельзя применить к стандартным методам анализа панельных данных. Нелинейная регрессия будет первым из рассматриваемых нами многочисленных примеров применения панельных данных, в которой разумность модели линейной регрессии нельзя распространить на более общую схему.

11.9.1. Устойчивая ковариационная матрица для нелинейного метода наименьших квадратов

В аналоге для (11-3) или (11-4) X_i было бы просто заменено на \hat{X}_i^0 , где строки представляют собой псевдорегрессоры для кластера i , как определено в (7-12), а « $\hat{}$ » показывает, что матрица вычислена с помощью оценки параметров нелинейным методом наименьших квадратов.

Пример 11.16. *Обращение за медицинской помощью*

Современная литература по экономике здравоохранения включает многочисленные исследования использования системы здравоохранения. Обычной мерой интересующей нас зависимой переменной является подсчет числа «встреч» с системой здравоохранения во время визитов к терапевту или в больницу. Подсчет этих событий обычно осуществляют при помощи модели пуассоновской регрессии, описанной в разделе 18.4. Нелинейная регрессионная модель имеет вид

$$E[y_i | \mathbf{x}_i] = \exp(\mathbf{x}_i' \boldsymbol{\beta}).$$

Недавнее исследование из этой серии — работа Riphahn, Wambach, Million (2003) «Побудительные эффекты в спросе на здравоохранение: оценка двумерных панельных счетных данных». Авторы интересовались подсчетом визитов к терапевту и посещений больницы, а в частности — влияние наличия индивидуального страхования на число посещений, т.е. вопрос, свидетельствуют ли данные о наличии риска недобросовестного поведения.

Необработанные данные опубликованы в архиве на веб-сайте Journal of Applied Econometrics URL для этих данных: [//qed.econ.queensu.ca/jae/2003-v18.4/riphahn-wambach-million/](http://qed.econ.queensu.ca/jae/2003-v18.4/riphahn-wambach-million/).

Переменные из файла данных приведены в табл. F7.1 приложения. Выборка представляет собой несбалансированную панель из 7,293 домохозяйств, данные из Социально-экономической панели Германии. Число наблюдений варьирует от одного до семи (1525; 1079; 825; 926; 1311; 1000; 887), общее число наблюдений составляет 27 326. Мы будем использовать эти данные в нескольких примерах здесь и далее в этой книге. В приведенной ниже модели используется простая спецификация для подсчета числа визитов к терапевту в год наблюдения:

$$x_{it} = (1, age_{it}, educ_{it}, income_{it}, kids_{it}).$$

В табл. 11.13 подробно изложены итерации нелинейного метода наименьших квадратов и полученные результаты. Критерий сходимости для итераций имеет вид $e^{0'}X^0(X^{0'}X^0)^{-1}X^0e^0 < 10^{-10}$. Несмотря на то что для сходимости потребуется 11 итераций, фактически функция достигает минимума в 7 из них. Оценки асимптотических стандартных ошибок вычислены с помощью классического метода $s^2(\hat{X}^{0'}\hat{X}^0)^{-1}$, а затем с кластерной корректировкой в (11-4). Скорректированные стандартные ошибки значительно больше, чем можно было ожидать, учитывая, что эти данные являются панельными.

Таблица 11.13. Оценки уравнения посещения нелинейным методом наименьших квадратов

| Начало итераций по НМНК. Линеаризированная регрессия | | | |
|--|----------|---------------|-----------------------------|
| Итерация = 1; сумма квадратов = 1 014 865,00; градиент = 156 281,794 | | | |
| Итерация = 2; сумма квадратов = 8 995 221,17; градиент = 8 131 951,67 | | | |
| Итерация = 3; сумма квадратов = 1 757 006,18; градиент = 897 066,012 | | | |
| Итерация = 4; сумма квадратов = 930 876,806; градиент = 73 036,2457 | | | |
| Итерация = 5; сумма квадратов = 860 068,332; градиент = 2 430,80472 | | | |
| Итерация = 6; сумма квадратов = 857 614,333; градиент = 12,8270683 | | | |
| Итерация = 7; сумма квадратов = 857 600,927; градиент = 0,411851239E-01 | | | |
| Итерация = 8; сумма квадратов = 857 600,883; градиент = 0,190628165E-03 | | | |
| Итерация = 9; сумма квадратов = 857 600,883; градиент = 0,904650588E-06 | | | |
| Итерация = 10; сумма квадратов = 857 600,883; градиент = 0,430441193E-08 | | | |
| Итерация = 11; сумма квадратов = 857 600,883; градиент = 0,204875467E-10 | | | |
| Переменная | Оценка | Станд. ошибка | Устойчивая станд. ошибка |
| Constant | 0,9801 | 0,08927 | 0,12522 |
| Возраст | 0,01873 | 0,001053 | 0,00142 |
| Образование | -0,03613 | 0,005732 | 0,00780 |
| Доход | -0,5911 | 0,07173 | 0,09702 |
| Дети | -0,1692 | 0,02642 | 0,03330 |

11.9.2. Фиксированные эффекты

Нелинейная модель регрессии панельных данных будет выглядеть как

$$y_{it} = h(\mathbf{x}_{it}, \boldsymbol{\beta}) + \varepsilon_{it}, t = 1, \dots, T_i, i = 1, \dots, n.$$

Рассмотрим модель со скрытой неоднородностью c_i . Сразу же возникает неоднозначность: как ввести неоднородность в модель? При построении линейной модели естественным выходом было бы включение аддитивного члена, как в

$$y_{it} = h(\mathbf{x}_{it}, \boldsymbol{\beta}) + c_i + \varepsilon_{it}, t = 1, \dots, T_i, i = 1, \dots, n. \quad (11-79)$$

Однако, как мы могли видеть в предшествующем примере, это, вероятнее всего, не подойдет. Логлинейная модель из предыдущего раздела ограничена необходимостью гарантии положительности $E[y_{it} | \mathbf{x}_{it}]$. Но наличие аддитивного случайного члена c_i , как в (11-78), не дает такой гарантии; за исключением случая, когда разброс c_i ограничен, условное среднее может быть отрицательным. Наиболее распространенным приложением нелинейных моделей является **модель индексной функции**

$$y_{it} = h(\mathbf{x}'_{it}\boldsymbol{\beta} + c_i) + \varepsilon_{it}.$$

Это естественное расширение линейной модели, но только при наличии условного среднего. Модели ни с фиксированными, ни со случайными эффектами нельзя оценить так, как это делалось в линейном случае.

Сначала рассмотрим модель с фиксированными эффектами. Запишем ее как

$$y_{it} = h(\mathbf{x}'_{it}\boldsymbol{\beta} + \alpha_i) + \varepsilon_{it}, \quad (11-80)$$

где параметры, которые следует оценить, это $\boldsymbol{\beta}$ и $\alpha_i, i = 1, \dots, n$. Преобразование данных в отклонения от групповых средних не устранил фиксированные эффекты из модели. Например,

$$y_{it} - \bar{y}_i = h(\mathbf{x}'_{it}\boldsymbol{\beta} + \alpha_i) - \frac{1}{T_i} \sum_{s=1}^{T_i} h(\mathbf{x}'_{is}\boldsymbol{\beta} + \alpha_i),$$

что совершенно не упрощает модель. Преобразование регрессоров в отклонения также бессмысленно. Чтобы оценить параметры, необходимо минимизировать сумму квадратов по отношению ко всем $n + K$ параметрам одновременно. Поскольку число коэффициентов при фиктивных переменных может быть огромным — предыдущий пример основывается на базе данных из 7293 групп, — то вычисления здесь могут быть сложными или неосуществимыми. Метод максимизации функций (таких как отрицательная сумма квадратов), содержащий неограниченное число коэффициентов при фиктивных переменных, описан в главе 17. Как мы увидим далее в этой книге, сложность в работе с нелинейными моделями, содержащими большое число коэффициентов при фиктивной переменной, необязательно заключается в практической стороне вычисления оценок. Как правило, эта проблема решаема. Сложность таких моделей представляет интригующий феномен,

известный как **проблема дополнительных параметров**. В большинстве (но не во всех, как мы обнаружим позже) нелинейных моделей панельных данных, содержащих n коэффициентов при фиктивных переменных, как в (11-79), следствием того, что число параметров возрастает с числом агентов в выборке, является смещенность и несостоятельность оценки β в той же степени, что и $O(1/T)$. Поскольку T в нашем приложении равно лишь 7 или меньше, то это и будет служить подходящим примером.

Пример 11.17. Экспоненциальная модель с фиксированными эффектами
Экспоненциальная модель из предыдущего примера на самом деле является одним из небольшого числа известных случаев, в которых возможно «поставить условие» выведения фиктивных переменных. Рассмотрим сумму квадратов остатков

$$S_n = \frac{1}{2} \sum_{i=1}^n \sum_{t=1}^{T_i} [y_{it} - \exp(\mathbf{x}'_{it}\beta + \alpha_i)]^2.$$

Условием первого порядка для минимизации S_n в отношении α_i является

$$\frac{\partial S_n}{\partial \alpha_i} = \sum_{t=1}^{T_i} - [y_{it} - \exp(\mathbf{x}'_{it}\beta + \alpha_i)] \exp(\mathbf{x}'_{it}\beta + \alpha_i) = 0. \quad (11-81)$$

Пусть $\gamma_i = \exp(\alpha_i)$. Тогда равносильным необходимым условием было бы

$$\frac{\partial S_n}{\partial \gamma_i} = \sum_{t=1}^{T_i} - [y_{it} - \gamma_i \exp(\mathbf{x}'_{it}\beta)] [\gamma_i \exp(\mathbf{x}'_{it}\beta)] = 0$$

или

$$\gamma_i \sum_{t=1}^{T_i} [y_{it} \exp(\mathbf{x}'_{it}\beta)] = \gamma_i^2 \sum_{t=1}^{T_i} [\exp(\mathbf{x}'_{it}\beta)]^2.$$

Очевидно, что если мы можем решить уравнение для γ_i , то мы можем получить $\alpha_i = \ln \gamma_i$. Предыдущее уравнение действительно можно решить для γ_i , по крайней мере условно. Для минимальной суммы квадратов будет верно, что

$$\hat{\gamma}_i = \frac{\sum_{t=1}^{T_i} y_{it} \exp(\mathbf{x}'_{it}\hat{\beta})}{\sum_{t=1}^{T_i} [\exp(\mathbf{x}'_{it}\hat{\beta})]^2}. \quad (11-82)$$

Теперь мы можем подставить (11-81) в (11-80), чтобы оценить α_i . (Это аналог взятия отклонений от средних в линейном случае. Как уже отмечалось, это возможно только для очень специальных моделей — наш случай именно такой. Этот процесс также называется «вытаскиванием» параметров γ_i . (Обратите внимание, что в данном решении $\hat{\gamma}_i$ был получен как коэффициент наклона в регрессии без постоянного члена y_{it} на $\hat{z}_{it} = \exp(\mathbf{x}'_{it}\hat{\beta})$ с использованием T_i наблюдений.) Результат в (11-81) должен сохраниться в этом решении. Итак, подстановка (11-81) в (11-80)

ограничивает поиск β до тех значений, которые удовлетворяют ограничениям в (11-81). Итоговая функция суммы квадратов теперь представляет собой функцию лишь из данных и β и может быть минимизирована по отношению к этому вектору из K параметров. Имея оценку β , мы можем оценить α_i взятием логарифма результата из (11-81) (а он положительный по построению).

В предыдущем примере представлена смешанная картина модели с фиксированными эффектами. В нелинейных случаях возникают две проблемы, о которых не говорилось ранее, первая из них практическая, связанная с вычислением параметров при фиктивных переменных, вторая теоретическая, проблема дополнительных параметров, ее нам еще предстоит исследовать, однако она, по всей видимости, и окажется значительным недостатком модели с фиксированными эффектами. Также заметим, что мы сосредотачиваемся на особой форме этой модели, функции «единого индекса», в которой условное среднее представляет собой нелинейную функцию линейной функции. В более общих случаях может оставаться неясным, как ненаблюдаемая гетерогенность должна включаться в функцию регрессии.

11.9.3. Случайные эффекты

Нелинейная модель со случайными эффектами также содержит некоторые сложности и для спецификации, и для оценивания. Начнем с общей модели

$$y_{it} = h(\mathbf{x}_{it}, \beta, u_i) + \varepsilon_{it}. \quad (11-83)$$

Допущение о случайных эффектах, как обычно, будет состоять в предположении независимости средних значений

$$E[u_i | \mathbf{X}_i] = 0.$$

В отличие от линейной модели нелинейная регрессия не может быть состоятельно оценена при помощи (нелинейного) метода наименьших квадратов. С практической точки зрения мы можем понять, почему так происходит, из (7-28)–(7-30). В линеаризированной регрессии и условное среднее в точке разложения β^0 [см. (7-28)], и псевдорегрессоры являются функциями ненаблюдаемой ошибки u_i . Это верно в как общем случае, так и в более простом случае модели единого индекса

$$y_{it} = h(\mathbf{x}'_{it}\beta + u_i) + \varepsilon_{it}. \quad (11-84)$$

Таким образом, вычислить итерации для нелинейного метода наименьших квадратов не представляется возможным. Как и в случае с фиксированными эффектами, ни отклонения от групповых средних, ни метод первых разностей не решают эту проблему. Игнорирование проблемы — т.е. просто вычисление оценки нелинейным методом наименьших квадратов без учета неоднородности — не позволяет получить состоятельную оценку по тем же причинам. Как правило, благотворное влияние скрытой неоднородности

(случайных эффектов), которое мы наблюдаем в линейной модели, распространяется лишь на очень небольшое число нелинейных моделей, и, к сожалению, наш случай не из их числа.

Проблема вычисления частных эффектов в модели со случайными эффектами вроде (11-83) состоит в том, что когда $E[y_{it} | \mathbf{x}_{it}, u_i]$ задано по (11-83), то

$$\frac{\partial E[y_{it} | \mathbf{x}'_{it}\boldsymbol{\beta} + u_i]}{\partial \mathbf{x}_{it}} = [h'(\mathbf{x}'_{it}\boldsymbol{\beta} + u_i)]\boldsymbol{\beta}$$

является функцией ненаблюдаемой ошибки u_i . Существует два выхода из этой ситуации: подход с фиксированными эффектами из предыдущего раздела и подход со случайными эффектами. Подход с фиксированными эффектами допустим, но его реализации может препятствовать проблема побочных параметров, отмеченная ранее. Подход со случайными эффектами мог бы оказаться более предпочтительным, однако он осуществляется ценой допущения того, что \mathbf{x}_{it} и u_i не коррелированы, а это может быть нерационально. Парке, Вулдридж (Parke, Wooldridge (2008)) рассмотрели несколько случаев и предложили метод Мундлака проекции u_i на групповые средние для \mathbf{x}_{it} . Отсюда работающая спецификация модели имеет вид

$$E^*[y_{it} | \mathbf{x}_{it}, \bar{\mathbf{x}}_i, v_i] = h(\mathbf{x}'_{it}\boldsymbol{\beta} + \alpha + \bar{\mathbf{x}}'_i\boldsymbol{\theta} + v_i).$$

При этом сохраняется практическая проблема вычисления оценки параметров и частных эффектов. Парке, Вулдредж (Parke, Wooldridge (2008)) предлагают вполне удобный результат при условии, что можно допустить, что v_i нормально распределены с нулевым средним и дисперсией σ_v^2 . В таком случае

$$\begin{aligned} E[y_{it} | \mathbf{x}_{it}, \bar{\mathbf{x}}] &= E_{v_i} E[y_{it} | \mathbf{x}_{it}, \bar{\mathbf{x}}, v_i] = h\left(\frac{\mathbf{x}'_{it}\boldsymbol{\beta} + \alpha + \bar{\mathbf{x}}'_i\boldsymbol{\theta}}{\sqrt{1 + \sigma_v^2}}\right) = \\ &= h(\mathbf{x}'_{it}\boldsymbol{\beta}_v + \alpha_v + \bar{\mathbf{x}}'_i\boldsymbol{\theta}_v). \end{aligned}$$

Из этого следует вывод, что в регрессии, оцененной нелинейным методом наименьших квадратов, оцениваться будут сопоставимые коэффициенты, после чего средний частный эффект можно оценить для конкретного значения ковариат \mathbf{x}_0 при

$$\hat{\Delta}(\mathbf{x}_0) = \frac{1}{n} \sum_{i=1}^n h'(\mathbf{x}'_0\hat{\boldsymbol{\beta}}_v + \hat{\alpha}_v + \bar{\mathbf{x}}'_i\hat{\boldsymbol{\theta}}_v) \hat{\boldsymbol{\beta}}_v.$$

Они применяли эту технику к случаю результатов тестов, которые представляют собой дробь, ограниченную нулем и единицей. Другим примером служит работа Лудермилка (Loudermilk (2007)), являющаяся расширением на случай динамической модели.

11.10. Системы уравнений

Расширение модели внешне не связанных регрессий на случай панельных данных может быть осуществлено в двух направлениях. Несколько исследо-

ваний возвели уже известный нам подход со случайными эффектами из раздела 11.5 на вершину обобщенной регрессии. Альтернативный подход к анализу моделей с фиксированными и случайными эффектами как виду модели внешне не связанных регрессий, предложенный Чемберленом (1982, 1984), заложил основу для современных исследований динамических моделей панельных данных, как в подразделах 11.8.2 и 11.8.3.

Эври (Avery (1977)) предложил естественное расширение модели со случайными эффектами на несколько уравнений

$$y_{it,j} = \mathbf{x}'_{it,j} \beta_j + \varepsilon_{it,j} + u_{i,j},$$

где j показывает уравнение, i указывает на объект, а t — индекс времени, как и прежде. Каждое уравнение может рассматриваться как модель со случайными эффектами. Однако в таких условиях, если уравнения действительно не связаны (т.е. $\text{Cov}[\varepsilon_{it,m}, \varepsilon_{it,l} | \mathbf{X}] = 0$ и $\text{Cov}[u_{i,m}, u_{i,l} | \mathbf{X}] = 0$), эффективной является GLS-оценка каждого уравнения в отдельности, как описывалось в разделе 11.5, а не МНК. То есть при отсутствии корреляции между уравнениями каждое уравнение образует модель со случайными эффектами. Корреляция между уравнениями принимает форму

$$E[\varepsilon_{it,j} \varepsilon_{it,l} | \mathbf{X}] = \sigma_{jl}$$

и

$$E[u_{i,j} u_{i,l} | \mathbf{X}] = \theta_{jl}.$$

Наблюдения также не коррелированы по объектам (индивидам) ($\varepsilon_{it,j}$, $\varepsilon_{rs,l}$) и (u_{ij} , $u_{r,l}$) при $i \neq r$. «Зашумляющие» члены $\varepsilon_{it,j}$ также не коррелированы во времени для каждого объекта и между объектами. Корреляция с течением времени возникает из-за влияния общего эффекта, который порождает устойчивые случайные эффекты для того или иного объекта и внутри уравнения, и между уравнениями через θ_{jl} . Эври разработал двухшаговую оценку для этой модели. На первом шаге, как обычно, оценки компонент дисперсии основываются на МНК-остатках. Второй шаг — это FGLS. Последующие исследования добавили в модель некоторые особенности. Мэгноус (Magnus (1982)) получил логарифмическую функцию правдоподобия для нормально распределенных шоков, уравнения правдоподобия для MLE и метод оценивания. Вербон (Verbon (1980)) добавил в модель гетероскедастичность.

Кроме того, было еще несколько примеров применения модели, включая анализ ценообразования на электроэнергию и влияние ставки в зависимости от времени суток у Хоури, Вэриана (Howrey, Varian (1984)), анализ формы модели ценообразования финансовых активов (CAPM) у Брауна и др. (Brown et al. (1983)), анализ расходов авиакомпаний у Сиклса (Sickles (1985)) и развитие нелинейной модели внешне не связанных регрессий для панельных данных по сельскохозяйственной продукции у Вэна и др. (Wan et al.'s (1992)).

Пример 11.18. *Спрос на электричество и газ*

Бирлайн, Данн, Макконнон (Beierlein, Dunn, McConnon (1981)) предложили динамическую модель внешне не связанных регрессий для панельных

данных при анализе спроса на электричество и природный газ на северо-востоке США. Центральное уравнение модели имеет вид

$$\begin{aligned} \ln Q_{it,j} &= \beta_0 + \beta_1 \ln P_{\text{natural gas}}_{it,j} + \beta_2 \ln P_{\text{electricity}}_{it,j} + \beta_3 \ln P_{\text{fuel oil}}_{it,j} \\ &\quad + \beta_4 \ln \text{per capita income}_{it,j} + \beta_5 \ln Q_{i,t-1,j} + w_{it,j} \\ w_{it,j} &= \varepsilon_{it,j} + u_{i,j} + v_{t,j}, \end{aligned}$$

где

j = потребительский сектор (природный газ, электричество) \times (жилой, коммерческий, промышленный),

i = штат (Новая Англия плюс Нью-Йорк, Нью-Джерси, Пенсильвания),

t = год (1957, ..., 1977).

Обратите внимание, что в этой модели есть и временные, и индивидуальные случайные эффекты, а также запаздывающая зависимая переменная в каждом уравнении.

11.11. Неоднородность параметров

До сих пор мы допускали, что параметры коэффициента угла наклона модели представляют собой фиксированные константы, а константа случайным образом меняется от группы к группе. Эквивалентная формулировка объединенных моделей, моделей с фиксированными эффектами и моделей со случайными эффектами имеет вид

$$y_{it} = (\alpha + u_i) + \mathbf{x}'_{it}\beta + \varepsilon_{it},$$

где u_i — конкретная для агента случайная переменная с условной дисперсией, равной нулю, в объединенной модели и положительной в других, а условное среднее зависит от X_i в модели с фиксированными эффектами и от константы в модели со случайными эффектами. В любой из них неоднородность в модели проявляет себя как изменчивость в постоянных членах в регрессионной модели. В исследованиях существует достаточно свидетельств — два из них мы рассмотрим позже, — которые предполагают, что другие параметры в модели также меняются по объектам. В динамической модели, которую мы рассмотрим в подразделе 11.11.3, основное внимание уделяется межстрановой изменчивости параметра угла наклона производственной функции. В этом разделе мы рассмотрим некоторые подходы к изучению неоднородности параметров в моделях панельных данных.

11.11.1. Модель со случайными коэффициентами

Неоднородность параметров модели по агентам/объектам или группам можно смоделировать как стохастическую изменчивость²⁷. Предположим,

²⁷ Наиболее широко цитируемые работы — это Hildreth, Houck (1968), Swamy (1970, 1971, 1974), Hsiao (1975) и Chow (1984). См. также Breusch, Pagan (1979). Некоторые современные работы отмечены у Swamy, Tavlas (1995, 2001) и Hsiao (2003). Модель имеет некоторое сходство с байесовским методом из главы 16. Однако это сходство лишь поверхностное. Мы будем придерживаться классического метода оценивания во всем.

что мы запишем

$$\begin{aligned} y_i &= \mathbf{X}_i \beta_i + \varepsilon_i, \\ E[\varepsilon_i | \mathbf{X}_i] &= \mathbf{0}, \\ E[\varepsilon_i \varepsilon_i' | \mathbf{X}_i] &= \sigma_\varepsilon^2 \mathbf{I}_T, \end{aligned} \quad (11-85)$$

где

$$\beta_i = \beta + \mathbf{u}_i \quad (11-86)$$

и

$$\begin{aligned} E[\mathbf{u}_i | \mathbf{X}_i] &= \mathbf{0}, \\ E[\mathbf{u}_i \mathbf{u}_i' | \mathbf{X}_i] &= \Gamma. \end{aligned} \quad (11-87)$$

(Обратите внимание, что в том случае, если постоянный член в β является случайным и остальные параметры зафиксированы, как и раньше, эта модель приводит к модели со случайными эффектами, которую мы рассматривали в разделе 11.5.) На данный момент предположим, что автокорреляция, или перекрестная корреляция, в ε_i отсутствует. Также допустим, что $T > K$, поэтому при необходимости можно вычислить линейную регрессию y_i на \mathbf{X}_i для каждой группы. Таким образом, β_i , используемый для конкретного объекта, является результатом некоторого случайного процесса с вектором среднего значения β и ковариационной матрицей Γ ²⁸. Подставив (11-85) в (11-84) и расширив результат, мы получим обобщенную модель регрессии для каждого блока наблюдений

$$y_i = \mathbf{X}_i \beta + (\varepsilon_i + \mathbf{X}_i \mathbf{u}_i),$$

отсюда

$$\Omega_{ii} = E[(y_i - \mathbf{X}_i \beta)(y_i - \mathbf{X}_i \beta)' | \mathbf{X}_i] = \sigma_\varepsilon^2 \mathbf{I}_T + \mathbf{X}_i \Gamma \mathbf{X}_i'.$$

Для системы в целом ковариационная матрица шоков блочно-диагональна с диагональным блоком Ω_{ii} размером $T \times T$. Мы можем записать GLS-оценку как матричное взвешенное среднее из МНК-оценок по отдельным группам:

$$\hat{\beta} = (\mathbf{X}' \Omega^{-1} \mathbf{X})^{-1} \mathbf{X}' \Omega^{-1} \mathbf{y} = \sum_{i=1}^n \mathbf{W}_i \mathbf{b}_i, \quad (11-88)$$

где

$$\mathbf{W}_i = \left[\sum_{i=1}^n \left(\Gamma + \sigma_\varepsilon^2 (\mathbf{X}_i' \mathbf{X}_i)^{-1} \right)^{-1} \right]^{-1} \left(\Gamma + \sigma_\varepsilon^2 (\mathbf{X}_i' \mathbf{X}_i)^{-1} \right)^{-1}.$$

Для эмпирического приложения этой модели необходима оценка Γ . Один способ [см., например, Swamy (1971)] заключается в использовании эмпирической дисперсии множества n оценок наименьших квадратов, \mathbf{b}_i минус среднее значение для $s_i^2 (\mathbf{X}_i' \mathbf{X}_i)^{-1}$:

$$\mathbf{G} = [1/(n-1)] \left[\sum_i \mathbf{b}_i \mathbf{b}_i' - n \bar{\mathbf{b}} \bar{\mathbf{b}}' \right] - (1/N) \sum_i \mathbf{V}_i, \quad (11-89)$$

²⁸Свэми и Тавлас (Swamy и Tavlás (2001)) называют эту моделью со случайными коэффициентами первого поколения (RCM). Мы будем рассматривать «второе поколение» (текущее поколение) моделей со случайными коэффициентами в следующем разделе.

где

$$\bar{\mathbf{b}} = (1/n)\sum_i \mathbf{b}_i$$

и

$$\mathbf{V}_i = s_i^2(\mathbf{X}'_i\mathbf{X}_i)^{-1}.$$

Эта матрица может не быть положительно определена, однако в таком случае [как предлагает Baltagi (2005)] можно было бы отказаться от второго члена.

Тест хи-квадрат для модели со случайными коэффициентами при альтернативе о классической регрессии (т.е. без случайных коэффициентов) может основываться на

$$C = \sum_i (\mathbf{b}_i - \mathbf{b}_*)' \mathbf{V}_i^{-1} (\mathbf{b}_i - \mathbf{b}_*),$$

где

$$\mathbf{b}_* = [\sum_i \mathbf{V}_i^{-1}]^{-1} \sum_i \mathbf{V}_i^{-1} \mathbf{b}_i.$$

В условиях нулевой гипотезы об однородности C имеет предельное распределение хи-квадрат с $(n - 1)K$ степенями свободы. Наилучшими отдельными линейными несмещенными предикторами векторов коэффициентов, специфических для группы, являются матричные взвешенные средние для GLS-оценки, $\hat{\beta}$, и специфических для группы МНК-оценок \mathbf{b}_i ²⁹:

$$\hat{\beta}_i = \mathbf{Q}_i \hat{\beta} + [\mathbf{I} - \mathbf{Q}_i] \mathbf{b}_i, \quad (11-90)$$

где $\mathbf{Q}_i = [(1/s_i^2)\mathbf{X}'_i\mathbf{X}_i + \mathbf{G}^{-1}]^{-1} \mathbf{G}^{-1}$.

Пример 11.19. Модель со случайными коэффициентами

В примере 10.1 мы рассматривали модель Маннелла производительности общественного капитала на уровне штатов:

$$\begin{aligned} \ln gsp_{it} &= \beta_1 + \beta_2 \ln pc_{it} + \beta_3 \ln hwy_{it} + \beta_4 \ln water_{it} + \\ &+ \beta_5 \ln util_{it} + \beta_6 \ln emp_{it} + \beta_7 unemp_{it} + \varepsilon_{it}, \\ i &= 1, \dots, 48; t = 1, \dots, 17. \end{aligned}$$

Панель состоит из данных по штатам за 17 лет. Модель в примере 10.1 (и у Маннелла) не дает никаких возможностей для учета неоднородности параметров, за исключением постоянного члена. Мы произвели переоценку модели, используя метод Hildreth и Houck. Оценки МНК и FGLS приведены в табл. 11.14. Статистика хи-квадрат для проверки нулевой гипотезы об однородности параметров модели равна 25 556,26 с $7(47) = 329$ степенями свободы. Критическое значение из таблицы равно 372,299, поэтому гипотеза будет отвергнута.

В отличие от других случаев, которые мы рассматривали в этой главе, FGLS-оценки сильно отличаются от МНК, несмотря на тот факт, что оба метода дают состоятельные оценки, а выборка весьма большая. Приведенные ниже стандартные отклонения вычислены с использованием

²⁹Hsiao (2003, p. 144–149).

G в качестве ковариационной матрицы. [Для этих данных вычитание второй матрицы привело **G** к тому, что она стала не положительно определенной, поэтому в таблице стандартные отклонения основаны на оценках с использованием лишь первого члена из (11-88).] Увеличение стандартных ошибок поразительно. Это предполагает, что параметры модели значительно различаются по штатам. Мы использовали (11-89), чтобы вычислить оценки специфических для штата коэффициентов. На рис. 11.1 представлена гистограмма для коэффициента частного капитала. Как и предполагалось, изменчивость в оценках велика.

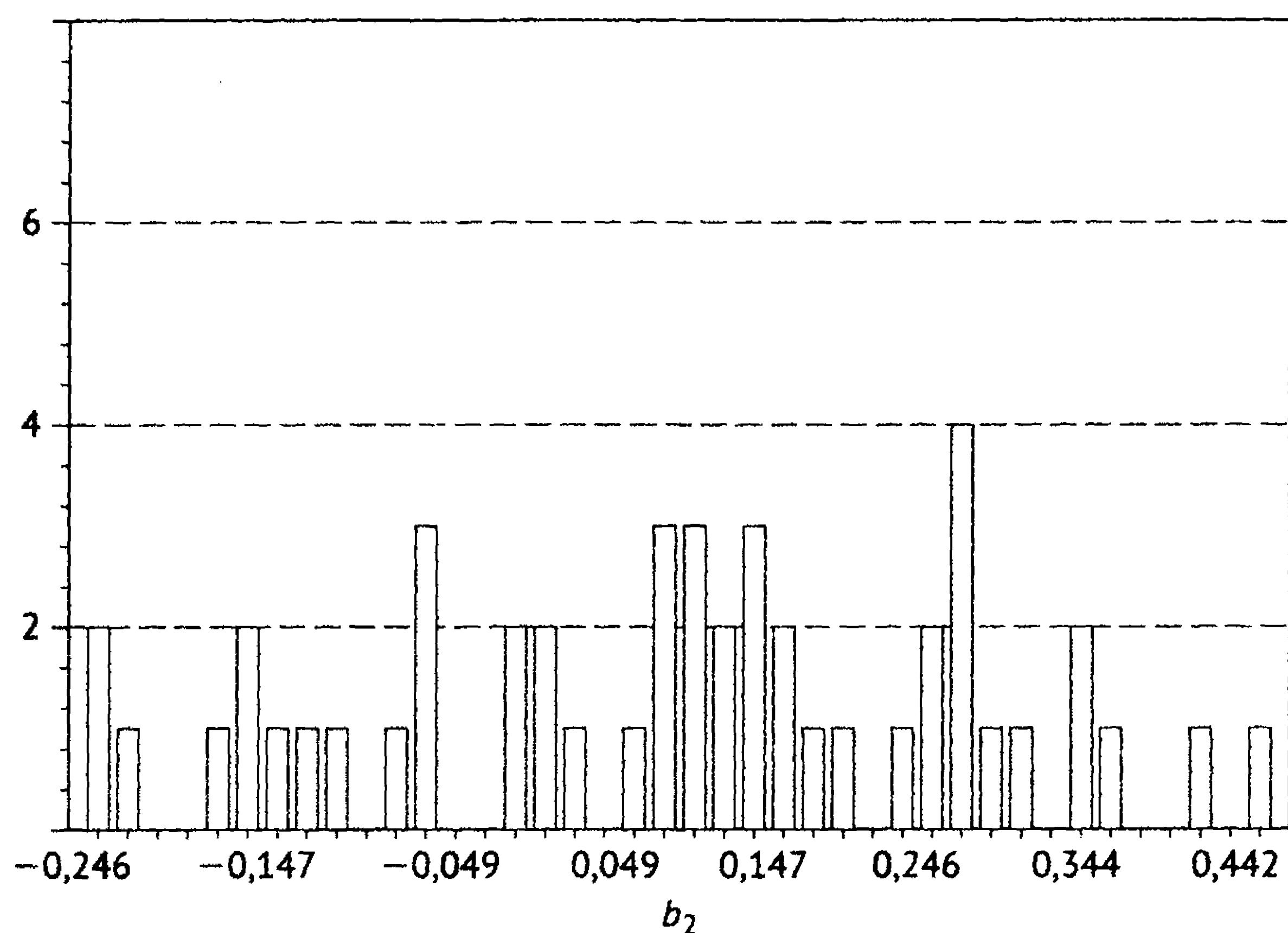


Рис. 11.1. Оценки коэффициентов частного капитала

Таблица 11.14. Оценки моделей со случайными коэффициентами

| Переменная | МНК | | FGLS | | Совокуп. станд. от- клонение |
|----------------------|----------|------------------|-----------|------------------|------------------------------------|
| | Оценка | Станд. ошибка | Оценка | Станд. ошибка | |
| Constant | 1,9260 | 0,05250 | 1,6533 | 1,08331 | 7,0782 |
| ln pc | 0,3120 | 0,01109 | 0,09409 | 0,05152 | 0,3036 |
| ln hwy | 0,05888 | 0,01541 | 0,1050 | 0,1736 | 1,1112 |
| ln water | 0,1186 | 0,01236 | 0,07672 | 0,06743 | 0,4340 |
| ln util | 0,00856 | 0,01235 | -0,01489 | 0,09886 | 0,6322 |
| ln emp | 0,5497 | 0,01554 | 0,9190 | 0,1044 | 0,6595 |
| unemp | -0,00727 | 0,001384 | -0,004706 | 0,002067 | 0,01266 |
| σ_ε | 0,08542 | | 0,2129 | | |
| ln L | 853,1372 | | | | |

11.11.2. Иерархическая линейная модель

Многие исследователи применяли двухшаговый метод для оценки двухуровневых моделей. В наиболее распространенных примерах панельные данные используются, чтобы оценить модель

$$y_{it} = \mathbf{x}'_{it}\beta_i + \varepsilon_{it}, i = 1, \dots, n, t = 1, \dots, T,$$

$$\beta_{i,k} = \mathbf{z}'_i\alpha_k + u_{i,k}, i = 1, \dots, n.$$

При допущении о том, что панель достаточно длинная, первое уравнение оценивается n раз, по одному разу для каждого отдельного i , а затем оцененный коэффициент по x_{itk} в каждой регрессии образует наблюдение для регрессии на втором шаге³⁰. (Это тот метод, который мы использовали в (11-16) в разделе 11.4; каждое a_i вычисляется при помощи линейной регрессии $y_i - \mathbf{X}_i\mathbf{b}_{LSDV}$ на столбец из единиц.)

Пример 11.20. Облигации Fannie Mae

Fannie Mae — широко известное название Федеральной национальной ипотечной ассоциации (FNMC). Fannie Mae является вторичным поставщиком ипотечных денег для почти всех ипотек малого и среднего размера в США. Кредиты на сумму менее 100 000 долл. в данном исследовании названы малыми. Кредит назван традиционным на языке, используемом на данном рынке, если (по состоянию на 2004 г.) его размер не превышал 333 700 долл. Кредит на сумму больше чем традиционный носит название «гигантская ипотека». Fannie Mae обеспечивает капиталом почти все традиционные кредиты и не обеспечивает нетрадиционные. Вопрос, поставленный в рамках описываемого здесь исследования, состоял в том, отражает ли четко наблюдаемая разница между ставками по гигантским и приспособительным кредитам стоимость привлечения капитала на рынке. Fannie Mae является «спонсируемым государством предприятием» (GSE). Она была создана Конгрессом США, однако не является правительственным рычагом: это частная корпорация. Несмотря на такие двусмысленные отношения с государством (или благодаря им), рынки капитала, по-видимому, считают, что в привлечении капитала для Fannie Mae есть некоторая выгода. Покупатели долговых ценных бумаг GSE, кажется, верят, что долг неявно субсидируется государством, несмотря на тот факт, что Fannie Mae в своих публикациях прямо говорит обратное. Возникает «преимущество финансирования» (GFA), оцениваемое авторами исследования примерно в 17 базисных пунктов (сотая доля одного процента). В исследованиях ипотечного рынка жилой недвижимости Passmore (2005) и Passmore, Sherlund, Burgess (2005) хотели определить, перешла ли эта косвенная субсидия для GSE залогодержателям или вместо этого перешла акционерам. В своем исследовании они использовали очень большой массив данных и двухуровневую, двухшаговую процедуру

³⁰Расширение модели, в котором u_i гетероскедастична, подробно описано у Saxonhouse (1976) и пересмотрено у Achen (2005).

оценки. Оцененное на первом шаге уравнение представляло собой уравнение ставки по ипотеке с использованием выборки из приблизительно 1 млн закрытых ипотек. Все они были обычными 30-летними кредитами по фиксированной ставке и были закрыты в период между апрелем 1997 г. и маем 2003 г. Интересующая нас зависимая переменная — это ставка по ипотечному кредиту RM_{it} . Уравнение первого уровня имеет вид

$$RM_{it} = \beta_{1i} + \beta_{2,i}J_{it} + \text{члены для «коэффициента ипотечной задолженности (LTV)», «фиктивной переменной для нового дома», «малой ипотеки»} \\ + \text{члены для «взимаемых сборов» и была ли ипотека начата ипотечной компанией} + \varepsilon_{it}.$$

Главная интересующая нас переменная в этой модели — это J_{it} , которая служит фиктивной переменной, показывающей, является ли кредит гигантской ипотекой. i в этой ситуации обозначает пару (штат, время) для Калифорнии, Нью-Джерси, Мэриленда, Вирджинии и всех других штатов и месяцев с апреля 1997-го по май 2003-го. В общей совокупности было 370 групп. Регрессионная модель оценивалась для каждой группы. На втором шаге — интересующий нас коэффициент $\beta_{2,i}$. По совокупному среднему разница между гигантскими и остальными займами в то время составила 16 базисных пунктов. Уравнение второго уровня имеет вид

$$\beta_{2,i} = \alpha_1 + \alpha_2 GFA_i \\ + \alpha_3 \text{ ставка по годовым гособлигациям} \\ + \alpha_4 \text{ ставка по десятилетним гособлигациям} \\ + \alpha_5 \text{ кредитный риск} \\ + \alpha_6 \text{ риск досрочного погашения} \\ + \text{риск, связанный с несовпадением сроков погашения кредита} \\ + \text{фиксированные эффекты на квартал и штат} \\ + \text{емкость рынка ипотечного кредитования} \\ + \text{развитие ипотечного рынка} \\ + u_i.$$

Интересующий нас в конечном счете результат — это коэффициент при GFA , α_2 , интерпретируемый как доля преимущества финансирования, которая доходит до держателей ипотек. Были получены четыре разные оценки α_2 , в основе которых лежали четыре разные меры ликвидности корпоративного долга; оцененные значения составили: $(\hat{\alpha}_2^1, \hat{\alpha}_2^2, \hat{\alpha}_2^3, \hat{\alpha}_2^4) = (0,07; 0,31; 0,17; 0,10)$. Четыре оценки были усреднены при помощи **оценки наименьшего расстояния** (MDE). Пусть $\hat{\Omega}$ обозначает оцененную асимптотическую ковариационную матрицу размером 4×4 для этих оценок. Обозначим вектор расстояния

$$\mathbf{d} = (\hat{\alpha}_2^1 - \alpha_2, \hat{\alpha}_2^2 - \alpha_2, \hat{\alpha}_2^3 - \alpha_2, \hat{\alpha}_2^4 - \alpha_2)'$$

Оценка наименьшего расстояния представляет собой значение для α_2 , минимизирующее $\mathbf{d}'\hat{\Omega}^{-1}\mathbf{d}$. В данном исследовании $\hat{\Omega}$ — диагональная матрица. Довольно просто показать, что в этом случае MDE будет

$$\hat{\alpha}_2 = \sum_{j=1}^4 \hat{\alpha}_2^j \left(\frac{1/\hat{\omega}_j}{\sum_{m=1}^4 1/\hat{\omega}_m} \right).$$

Конечный ответ составил приблизительно 16%. Далее по косвенным оценкам авторы определили, что $100 - 16 = 84\%$ преимуществ финансирования GSE оставались внутри компании или перетекали к акционерам.

11.11.3. Неоднородность параметров и динамические модели панельных данных

В этом разделе мы рассматривали статические модели и относительно очевидные проблемы оценивания. По ходу изложения мы видели, что неоднородность параметров модели вносит существенную сложность в проводимый анализ. Динамические эффекты в этой модели при отсутствии неоднородности или вместе с ней также порождают комплекс новых вопросов в оценивании и получении статистических выводов. Существует множество случаев, когда динамические эффекты и неоднородность параметров совмещаются в моделях панельных данных. В этом разделе мы изучим несколько спецификаций и примеров. Уже знакомые нам техники оценки (МНК, FGLS и т.д.) в этих случаях неэффективны. Предлагаемые решения представлены в главе 8, где мы рассматривали метод инструментальных переменных, и в главе 13, где будет рассмотрена GMM-оценка и ее приложение в динамических моделях панельных данных.

Пример 11.21. Динамические модели панельных данных

Предшественником большинства современных исследований с использованием панельных данных является анализ рынка природного газа у Балестры и Нерлова (Balestra и Nerlove (1966)). [См. также Nerlove (2002, глава 2).] Модель представляет собой описание спроса на средства производства топливного газа. Центральное уравнение представлено моделью совокупного спроса

$$G_{it} = G_{it}^* + (1 - r)G_{i,t-1},$$

где G_{it} — текущий совокупный спрос. Текущий спрос состоит из нового спроса G_{it}^* , который образован прибавлением к запасам оборудования, плюс старый спрос, который составляет долю спроса в предыдущем периоде, а r — норма амортизации оборудования. Новый спрос образуется за счет чистого увеличения запасов оборудования, что моделируется таким образом:

$$G_{it}^* = \beta_0 + \beta_1 Price_{it} + \beta_2 \Delta Pop_{it} + \beta_3 Pop_{it} + \beta_4 \Delta Income_{it} + \beta_5 Income_{it} + \varepsilon_{it},$$

где Δ — это оператор (изменения) первых разностей, $\Delta X_t = X_t - X_{t-1}$. Приведенная форма модели представляет собой динамическое уравнение

$$G_{it} = \beta_0 + \beta_1 Price_{it} + \beta_2 \Delta Pop_{it} + \beta_3 Pop_{it} + \beta_4 \Delta Income_{it} + \beta_5 Income_{it} + \gamma G_{i,t-1} + \varepsilon_{it}.$$

Авторы анализировали панель, состоящую из 36 штатов, за шестилетний период (1957–1962). Рассматривались оба подхода — с фиксированными и случайными эффектами.

Модель равновесия для состояния устойчивого роста применялась многими авторами [например, Robertson, Symons (1992), Pesaran, Smith (1995), Lee, Pesaran, Smith (1997), Pesaran, Shin, Smith (1999), Nerlove (2002) и Hsiao, Pesaran, Tahmiscioglu (2002)] для сравнений по отраслям или странам. Робертсон и Саймонс (Robertson и Symons) смоделировали реальную заработную плату в 13 странах ОЭСР за период с 1958 до 1986 г. с уравнением заработной платы

$$W_{it} = \alpha_i + \beta_{1i} k_{it} + \beta_{2i} \Delta wedge + \gamma_i W_{i,t-1} + \varepsilon_{it},$$

где W_{it} — это реальная зарплата по продукту для страны i в год t , k_{it} — фондовооруженность, а $wedge$ — это «клин» налогов и импортных цен.

Ли, Песаран, Смит (Lee, Pesaran, Smith (1997)) сравнили рост доходов по странам с помощью модели устойчивого роста доходов следующего вида:

$$\ln y_{it} = \alpha_i + \theta_i t + \lambda_i \ln y_{i,t-1} + \varepsilon_{it},$$

где $\theta_i = (1 - \lambda_i)\delta_i$, δ_i — темпы технологического роста для страны i , а λ_i — параметр конвергенции. Скорость конвергенции для устойчивого состояния равна $1 - \lambda_i$.

Песаран и Смит (Pesaran, Smith, 1995) анализировали занятость с помощью панели из 38 отраслей Соединенного Королевства, наблюдаемых в течение 29 лет, 1956–1984 гг. Главное оцениваемое уравнение имело вид

$$\ln e_{it} = \alpha_i + \beta_{1i} t + \beta_{2i} \ln y_{it} + \beta_{3i} \ln y_{i,t-1} + \beta_{4i} \ln \bar{y}_t + \beta_{5i} \ln \bar{y}_{t-1} + \beta_{6i} \ln w_{it} + \beta_{7i} \ln w_{i,t-1} + \gamma_{1i} \ln e_{i,t-1} + \gamma_{2i} \ln e_{i,t-2} + \varepsilon_{it},$$

где y_{it} — отраслевой выпуск, \bar{y}_t — совокупный (не усредненный) выпуск, а w_{it} — реальная заработная плата.

В моделях роста интересующая нас величина — это **долгосрочный мультипликатор**, или **долгосрочная эластичность**. Долгосрочные эффекты получены при помощи следующего абстрактного эксперимента. Неотъемлемой чертой рассмотренных выше моделей является динамическое уравнение вида

$$y_t = \alpha + \beta x_t + \gamma y_{t-1}.$$

Предположим, что в момент t x_t фиксируется дальше от этой точки по направлению к \bar{x} . Тогда значение y_t в этот момент станет $\alpha + \beta \bar{x} + \gamma y_{t-1}$ с учетом прошлого значения. Если продолжать этот процесс и если $|\gamma| < 1$, то

y_s в конце концов достигнет равновесия в значении $y_s = y_{s-1} = \bar{y}$. В таком случае $\bar{y} = \alpha + \beta\bar{x} + \gamma\bar{y}$, из чего мы можем вывести, что $\bar{y} = (\alpha + \beta\bar{x})/(1 - \gamma)$. Путь к этому равновесию от момента t в будущее определяется **уравнением приспособления**

$$y_s - \bar{y} = (y_t - \bar{y})\gamma^{s-t}, s \geq t.$$

Далее интересно спросить: как влияет на равновесие изменение во входных данных \bar{x} ? Результатом будет $\partial\bar{y}/\partial\bar{x} = \beta/(1 - \gamma)$. Это долгосрочный, или **равновесный, мультипликатор** в модели. В предшествующей модели Песарана–Смита входные данные представлены в логарифмах, поэтому мультипликаторы являются долгосрочными эластичностями. К примеру, при двух лагах у $\ln e_{it}$ в модели Песарана–Смита долгосрочные эффекты для заработной платы равны

$$\phi_i = (\beta_{6i} + \beta_{7i})/(1 - \gamma_{1i} - \gamma_{2i}).$$

В этой ситуации в отличие от предыдущих подходов число элементов n , как правило, считается фиксированным, хотя оно часто будет оказываться весьма большим. Penn World Tables (http://pwt.econ.upenn.edu/php_site/pwt_index.php) в которых представлены базы данных по многим исследованиям такого типа, в настоящее время содержат информацию почти по 200 странам более чем за 50 лет. Асимптотические результаты для оценок приводятся с учетом возрастающего T , хотя мы в общем будем рассматривать случаи, в которых T мало. Удивительно, но одновременный рост T и n необязательно упростит выводы.

Искомым параметром во многих исследованиях является средний долгосрочный эффект, скажем $\bar{\phi} = (1/n)\sum_i \phi_i$ в примере Песарана–Смита. Поскольку n фиксировано, «параметр» $\bar{\phi}$ становится определяемым объектом оценки, т.е. при фиксированном n мы можем говорить о $\bar{\phi}$ как о параметре, а не его оценке. Использовать можно множество методов. Для оценки можно использовать объединенные модели, модели с фиксированными или случайными эффектами, модели, оцененные при помощи групповых средних, или отдельные регрессии. (К сожалению, почти все они несостоятельны.) Кроме того, нужно делать выбор, вычислять среднее из долгосрочных эффектов или долгосрочный эффект из среднего параметров. Выбор среднего значения функций $\bar{\phi}$ против функции средних

$$\bar{\phi}^* = \frac{\frac{1}{n}\sum_{i=1}^n (\hat{\beta}_{6i} + \hat{\beta}_{7i})}{1 - \frac{1}{n}\sum_{i=1}^n (\hat{\gamma}_{1i} + \hat{\gamma}_{2i})}$$

оказывается существенным. В своем исследовании промышленности Соединенного Королевства Песаран и Смит представили такие оценки: $-0,33$ для $2\bar{\phi}$ и $-0,45$ для $2\bar{\phi}^*$. (Авторы не выражают предпочтения какой-либо из них.)

Анализ до этого момента косвенно основывался на оценке отдельных моделей для каждого элемента (страны, отрасли и проч.). Существует множество других стратегий оценивания, которые можно было бы рассматривать. В данный момент мы допустим, что ряды данных стационарны во времени. (См. главу 21.) Это явно ложное допущение, что обнаруживается путем простого взгляда на тренды в макроэкономических данных, однако его

предположение позволяет нам продвинуться дальше. Мы пересмотрим его позже.

Рассмотрим общую динамическую модель панельных данных

$$y_{it} = \alpha_i + \beta_i x_{it} + \gamma_i y_{i,t-1} + \varepsilon_{it}. \quad (11-91)$$

Допустим, что T велико настолько, что можно вычислить индивидуальные регрессии. При отсутствии автокорреляции в ε_{it} уже было показано [например, у Griliches (1961), Maddala, Rao (1973)], что МНК-оценка γ_i смещена вниз, но состоятельна по T . Таким образом, $E[\hat{\gamma}_i - \gamma_i] = \theta_i/T$ для некоторого θ_i . В этом случае вывод по индивидуальной оценке долгосрочного мультипликатора, $\phi_i = \beta_i/(1 - \gamma_i)$, становится неясным. Знаменатель завышен. Однако неясно, является ли оценка β_i завышенной или заниженной. Верно то, что все смещения здесь равны $O(1/T)$. В данном примере T фиксировано и, вероятно, достаточно мало. Конечным результатом является то, что индивидуальная оценка ϕ_i окажется несмещенной с крайне маленькой долей вероятности, а по построению она несостоятельна, так как нельзя предположить, что T возрастает. Если так происходит, то $\hat{\phi}$ так же несостоятельна для $\bar{\phi}$. Мы усредняем n оценок, смещение и дисперсия каждой из которых составляет $O(1/T)$. Дисперсия среднего, таким образом, равна $O(1/nT)$ и стремится к нулю, однако смещение остается $O(1/T)$. Из этого следует, что среднее арифметическое из n средних не сходится к $\bar{\phi}$; оно сходится в среднее любого параметра, который оценивается по этим смещенным оценкам. Эта проблема исчезает при большом T , но к данной ситуации это не относится. Однако в исследовании Песарана и Смита T было равно 29, что достаточно велико, чтобы эти эффекты были, скорее всего, умеренными. В макроэкономических исследованиях по странам, как и в тех, в основе которых лежат Penn World Tables, ряды данных могут быть и длиннее, чем здесь.

Для улучшения результатов можно было бы рассмотреть агрегирование данных. Песаран и Смит (Pesaran, Smith (1995)) предложили использовать среднюю величину из усредненных по времени значений для каждой страны. Усреднение наблюдений по T в (11-90) дает

$$\bar{y}_i = \alpha_i + \beta_i \bar{x}_i + \gamma_i \bar{y}_{-1,i} + \bar{\varepsilon}_i. \quad (11-92)$$

Линейная регрессия с использованием n наблюдений была бы несостоятельна по двум причинам: во-первых, $\bar{\varepsilon}_i$ и $\bar{y}_{-1,i}$ должны быть коррелированы друг с другом. Во-вторых, ввиду неоднородности параметров модели без новых допущений неясно, что оценивают угловые МНК-коэффициенты при ложном допущении о том, что все коэффициенты равны. Однако \bar{y}_i и $\bar{y}_{-1,i}$ различаются лишь первым и последним наблюдениями; $\bar{y}_{-1,i} = \bar{y}_i - (y_{iT} - y_{i0})/T = \bar{y}_i - [\Delta_T(y)/T]$. Подстановка этого выражения в (11-89) дает

$$\begin{aligned} \bar{y}_i &= \alpha_i + \beta_i \bar{x}_i + \gamma_i \bar{y}_i - \gamma_i [\Delta_T(y)/T] + \bar{\varepsilon}_i = \\ &= \frac{\alpha_i}{1 - \gamma_i} + \frac{\beta_i}{1 - \gamma_i} \bar{x}_i - \frac{\gamma_i}{1 - \gamma_i} [\Delta_T(y)/T] + \bar{\varepsilon}_i = \\ &= \delta_i + \phi_i \bar{x}_i + \tau_i [\Delta_T(y)/T] + \bar{\varepsilon}_i. \end{aligned} \quad (11-93)$$

Мы все еще намерены оценить $\bar{\phi}$. Формула (11-92) не решает проблему оценки, так как предлагаемая регрессия с использованием групповых средних

остается неоднородной. Если можно было бы допустить, что индивидуальные долгосрочные коэффициенты случайно отличаются от средних величин таким же образом, как в модели со случайными параметрами из раздела 11.1.1, так что $\delta_i = \bar{\delta} + u_{\delta,i}$, и так же для всех остальных параметров, тогда модель можно было бы записать как

$$\begin{aligned}\bar{y}_i &= \bar{\delta} + \bar{\phi}\bar{x}_i + \bar{\tau}[\Delta_T(y)/T]_i + \bar{\varepsilon}_i + \{u_{\delta,i} + u_{\phi,i}\bar{x}_i + u_{\tau,i}[\Delta_T(y)/T]_i\} = \\ &= \bar{\delta} + \bar{\phi}\bar{x}_i + \bar{\tau}[\Delta_T(y)/T]_i + \bar{\varepsilon}_i + w_i.\end{aligned}$$

На этом шаге уравнение представляет собой регрессию с гетероскедастичными ошибками, поддающуюся оцениванию методом наименьших квадратов, но с одной проблемой. Состоятельность возникает, если члены $[\Delta_T(y)/T]_i$ и $\bar{\varepsilon}_i$ не коррелированы друг с другом. Поскольку первое — скорость изменения, а второе представлено в уровнях, то это условие, как правило, выполняется. Другая трактовка, отвечающая этой же цели, состоит в том, что скорость изменения в $[\Delta_T(y)/T]_i$ не должна быть коррелирована с уровнями в \bar{x}_i , а в этом случае регрессию можно разделить на части и простая линейная регрессия средних значений y_{it} для стран на средние значения x_{it} для стран и константу позволит получить состоятельные оценки $\bar{\phi}$ и $\bar{\delta}$.

Рассмотрим в качестве альтернативы метод временных рядов. Мы усредняем наблюдение в (11-90) по странам для каждого периода, а не по времени для каждой страны. В таком случае мы получим

$$\bar{y}_{.t} = \bar{\alpha} + \frac{1}{n} \sum_{i=1}^n \beta_i x_{it} + \frac{1}{n} \sum_{i=1}^n \gamma_i y_{i,t-1} + \frac{1}{n} \sum_{i=1}^n \varepsilon_{it}.$$

Пусть $\bar{\gamma} = \frac{1}{n} \sum_{i=1}^n \gamma_i$, так что $\gamma_i = \bar{\gamma} + (\gamma_i - \bar{\gamma})$ и $\beta_i = \bar{\beta} + (\beta_i - \bar{\beta})$. Тогда

$$\begin{aligned}\bar{y}_{.t} &= \bar{\alpha} + \bar{\beta}\bar{x}_{.t} + \bar{\gamma}\bar{y}_{-1,t} + [\bar{\varepsilon}_{.t} + (\beta_i - \bar{\beta})\bar{x}_{.t} + (\gamma_i - \bar{\gamma})\bar{y}_{-1,t}] = \\ &= \bar{\alpha} + \bar{\beta}\bar{x}_{.t} + \bar{\gamma}\bar{y}_{-1,t} + \bar{\varepsilon}_{.t} + w_{.t}.\end{aligned}$$

К сожалению, регрессор $\bar{\gamma}\bar{y}_{-1,t}$, несомненно, коррелирован с $w_{.t}$, поэтому ни МНК, ни GLS не дадут состоятельной оценки для данной модели. (Можно было бы рассмотреть оценку методом инструментальных переменных, однако в этой модели, согласно построению, нет доступного подходящего инструмента.) Другой способ — объединить полный набор данных, возможно, с фиксированными или случайными эффектами для постоянных членов. Поскольку объединение даже со специфическими для страны постоянными членами налагает условие однородности на остальные параметры, то сохраняются те же проблемы.

Наконец, вернувшись к (11-90), можно было бы рассматривать это выражение как формальную модель со случайными параметрами

$$\begin{aligned}y_{it} &= \alpha_i + \beta_i x_{it} + \gamma_i y_{i,t-1} + \varepsilon_{it}, \\ \alpha_i &= \alpha + u_{\alpha,i}, \\ \beta_i &= \beta + u_{\beta,i}, \\ \gamma_i &= \gamma + u_{\gamma,i}.\end{aligned}\tag{11-94}$$

Допущения, необходимые для формулировки модели таким образом, те же, что и в предыдущей части. Как заметили Песаран и Смит (Pesaran, Smith

(1995)), эту модель можно оценить с использованием оценки Свэми (Swamy, 1971), которая представляет собой матричное взвешенное среднее из оценок наименьших квадратов, рассмотренное в подразделе 11.11.1. Для такой оценки необходимо настолько большое T , чтобы можно было оценить методом наименьших квадратов регрессию для каждой страны. Так и произошло в случае с нашими примерами. В самом деле в примерах, которые мы рассматривали, n и T относительно велики. Если же это условие не выполняется, то можно использовать метод смешанных моделей, рассматриваемый в главе 15. Компромиссом, который, по всей видимости, хорошо работает в панелях с n и T среднего размера, является «смешанно-фиксированная» модель, предложенная Хсяо (Hsiao (1986, 2003)) и Вайнхолдом (Weinhold (1999)). Динамическая модель в (11-92) задается как модель с частичными фиксированными эффектами:

$$\begin{aligned} y_{it} &= \alpha_i d_{it} + \beta_i x_{it} + \gamma_i d_{it} y_{i,t-1} + \varepsilon_{it}, \\ \beta_i &= \beta + u_{\beta,i}, \end{aligned}$$

где d_{it} — фиктивная переменная, равная единице для страны i в каждом периоде и нулю, если иначе (т.е. обычный метод фиксированных эффектов). Заметим, что d_{it} также возникает с $y_{i,t-1}$. Как было сказано, модель содержит «фиксированные эффекты», один случайный коэффициент и в общей сумме $2n + 1$ коэффициентов, которые нужно оценить, в дополнение к двум дисперсионным компонентам, σ_ε^2 и σ_u^2 . Модель могла быть оценена неэффективно обычным методом наименьших квадратов — случайный коэффициент вызывает наличие гетероскедастичности (см. подраздел 11.11.1), — методом Хилдрета–Хоука–Свэми или при помощи смешанной линейной модели, представленной в главе 15.

Пример 11.22. Смешанная модель роста для развивающихся стран с фиксированными эффектами

Вайнхолд (Weinhold (1996)) и Найр-Райнхарт, Вайнхолд (Nair-Reichert, Weinhold (2001)) анализировали рост и развитие в панели из 24 развивающихся стран, наблюдаемых в течение 25 лет (1971–1995 гг.). Модель, которую они применили, представляла собой вариант смешанной модели с фиксированными эффектами, предложенной Хсяо (Hsiao (1986, 2003)). По их спецификации

$$\begin{aligned} GGDP_{i,t} &= \alpha_i d_{it} + \gamma_i d_{it} GGDP_{i,t-1} \\ &+ \beta_{1i} GGDI_{i,t-1} + \beta_{2i} GFDI_{i,t-1} + \beta_{3i} GEXP_{i,t-1} + \\ &+ \beta_4 INFL_{i,t-1} + \varepsilon_{it}, \end{aligned}$$

где

- $GGDP$ = темпы роста валового внутреннего продукта,
- $GGDI$ = темпы роста валовых внутренних инвестиций,
- $GFDI$ = темпы роста прямых иностранных инвестиций (притоки),
- $GEXP$ = темпы роста экспорта товаров и услуг,
- $INFL$ = уровень инфляции.

11.12. Заключение и выводы

В этой главе были представлены некоторые примеры расширения классической модели, которые можно осуществить при наличии панельных данных. В принципе любую модель, которую мы рассматривали до этой главы, и все модели, которые мы будем рассматривать далее, включая системы уравнений, можно расширить таким же образом. Как мы отметили в самом начале, главным преимуществом является то, что с помощью панельных данных можно формально смоделировать динамические эффекты и неоднородность групп, что типично для микроэкономических данных.

Основные термины и понятия

- Adjustment equation** — Уравнение приспособления
- Autocorrelation** — Автокорреляция
- Arellano and Bond's estimator** — Оценка Ареллано-Бонда
- Balanced panel** — Сбалансированная панель
- Between groups** — Межгрупповые
- Cluster estimator** — Кластерная оценка
- Contiguity** — Смежность
- Contiguity matrix** — Матрица смежности
- Contrasts** — Противопоставления
- Dynamic panel data model** — Динамическая модель панельных данных
- Equilibrium multiplier** — Равновесный мультипликатор
- Error components model** — Модель компонент ошибки
- Estimator** — Оценка
- Feasible GLS** — Допустимый GLS
- First difference** — Первые разности
- Fixed effects** — Фиксированные эффекты
- Fixed effects vector decomposition** — Декомпозиция вектора фиксированных эффектов
- Fixed panel** — Фиксированная панель
- Group means** — Групповые средние
- Group means estimator** — Оценка групповых средних
- Hausman specification test** — Тест Хаусмана на спецификацию
- Heterogeneity** — Неоднородность
- Hierarchical linear model** — Иерархическая линейная модель
- Hierarchical model** — Иерархическая модель
- Hausman and Taylor's estimator** — Оценка Хаусмана-Тейлора
- Incidental parameters problem** — Проблема несущественных параметров
- Index function model** — Модель индексной функции
- Individual effect** — Индивидуальный эффект
- Instrumental variable** — Инструментальная переменная
- Instrumental variable estimator** — Оценка инструментальных переменных
- Lagrange multiplier test** — Тест множителей Лагранжа
- Least squares dummy variable model** — Модель наименьших квадратов с фиктивной переменной
- Long run elasticity** — Долгосрочная эластичность
- Long run multiplier** — Долгосрочный мультипликатор
- Longitudinal data set** — Лонгитюдные данные
- Matrix weighted average** — Матричное взвешенное среднее
- Mean independence** — Независимость средних значений
- Measurement error** — Ошибка техники измерения
- Minimum distance estimator** — Оценка метода наименьшего расстояния
- Mixed model** — Смешанная модель
- Mundlak's approach** — Метод Мундлака
- Nested random effects** — Вложенные случайные эффекты
- Panel data** — Панельные данные

| | |
|---|--|
| Parameter heterogeneity — Неоднородность параметров | малых T |
| Partial effects — Частные эффекты | Spatial autocorrelation — Пространственная автокорреляция |
| Pooled model — Объединенная модель | Spatial autoregression coefficient — Коэффициент пространственной авторегрессии |
| Pooled regression — Объединенная регрессия | Spatial error correlation — Пространственная корреляция ошибок |
| Population averaged model — Усредненная по генеральной совокупности модель | Spatial lags — Пространственные лаги |
| Projections — Проекции | Specification test — Тест на спецификацию |
| Random coefficients model — Модель со случайными коэффициентами | Strict exogeneity — Строгая экзогенность |
| Random effects — Случайные эффекты | Time-invariant — Не изменяющиеся во времени |
| Random parameters — Случайные параметры | Two-step estimation — Двухшаговая оценка |
| Robust covariance matrix — Устойчивая ковариационная матрица | Unbalanced panel — Несбалансированная панель |
| Rotating panel — Обновляемая панель | Variable addition test — Тест на включение переменных |
| Simulation based estimation — Оценка, основанная на имитационном моделировании | Within groups — Внутригрупповые |
| Small T asymptotics — Асимптотики при | |

Упражнения

1. Ниже представлена панель данных по инвестициям (y) и прибыли (x) для $n = 3$ фирм за $T = 10$ периодов.

| t | $i = 1$ | | $i = 2$ | | $i = 3$ | |
|-----|---------|-------|---------|-------|---------|-------|
| | y | x | y | x | y | x |
| 1 | 13,32 | 12,85 | 20,30 | 22,93 | 8,85 | 8,65 |
| 2 | 26,30 | 25,69 | 17,47 | 17,96 | 19,60 | 16,55 |
| 3 | 2,62 | 5,48 | 9,31 | 9,16 | 3,87 | 1,47 |
| 4 | 14,94 | 13,79 | 18,01 | 18,73 | 24,19 | 24,91 |
| 5 | 15,80 | 15,41 | 7,63 | 11,31 | 3,99 | 5,01 |
| 6 | 12,20 | 12,59 | 19,84 | 21,15 | 5,73 | 8,34 |
| 7 | 14,93 | 16,64 | 13,76 | 16,13 | 26,68 | 22,70 |
| 8 | 29,82 | 26,45 | 10,00 | 11,61 | 11,49 | 8,36 |
| 9 | 20,32 | 19,64 | 19,51 | 19,55 | 18,49 | 15,44 |
| 10 | 4,77 | 5,43 | 18,32 | 17,06 | 20,84 | 17,87 |

- Объедините данные и вычислите коэффициенты регрессии методом наименьших квадратов для модели $y_{it} = \alpha + \beta x_{it} + \varepsilon_{it}$.
- Оцените модель с фиксированными эффектами из (11-13), а затем проверьте гипотезу о том, что постоянный член будет одинаковым для всех трех фирм.
- Оцените модель со случайными эффектами из (11-28), а затем выполните тест множителей Лагранжа для проверки гипотезы о том, что нужно использовать классическую модель без общих эффектов.

- d. Выполните тест Хаусмана на спецификацию для модели со случайными эффектами против гипотезы о модели с фиксированными эффектами.
2. Предположим, что сформулирована модель с фиксированными эффектами с общим постоянным членом и $n - 1$ фиктивной переменной (опустим, скажем, последнюю). Проанализируйте эффект, который оказывает это предположение на множество коэффициентов при фиктивных переменных, а также на оценки угловых коэффициентов, полученных методом наименьших квадратов, в сравнении с (11-3).
3. *Модель со случайными эффектами в несбалансированной панели.* Предположим, что модель со случайными эффектами из раздела 11.5 следует оценить по панели, в группах которой разное число наблюдений. Пусть T_i будет числом наблюдений для группы i .
- a. Покажите, что оценка методом наименьших квадратов для объединенной модели является несмещенной и состоятельной, несмотря на это усложнение.
- b. Покажите, что оценка в (11-40), в основе которой лежит оценка β методом наименьших квадратов в объединенной регрессии (или, вообще говоря, *любая* состоятельная оценка β), является состоятельной оценкой для σ_ε^2 .
4. Чему равны пределы по вероятности для $(1/n)LM$, где критерий множителей Лагранжа (LM) определяется в (11-42) в условиях нулевой гипотезы о том, что $\sigma_u^2 = 0$, и при альтернативной гипотезе, что $\sigma_u^2 \neq 0$?
5. *Модель с двунаправленными фиксированными эффектами.* Предположим, что модель с фиксированными эффектами модифицирована так, чтобы включить фиктивную переменную, специфическую и по времени, и по агенту. Тогда $y_{it} = \alpha_i + \gamma_t + \mathbf{x}'_{it}\beta + \varepsilon_{it}$. Для каждого наблюдения специфические по агенту и времени переменные в сумме равны 1, поэтому здесь есть несколько лишних коэффициентов. Анализ из подраздела 11.4.4 показал, что один из способов удалить эту избыточность — включить общую константу и удалить одну из специфических по времени фиктивных переменных. Таким образом, модель имеет вид

$$y_{it} = \mu + (\alpha_i - \alpha_1) + (\gamma_t - \gamma_1) + \mathbf{x}'_{it}\beta + \varepsilon_{it}.$$

(Обратите внимание, что соответствующая специфическая по времени и агенту переменная равна нулю, когда t или i равно единице.) Оценки β простым методом наименьших квадратов далее получены при помощи регрессии $y_{it} - \bar{y}_i - \bar{y}_t + \bar{y}$ на $\mathbf{x}_{it} - \bar{\mathbf{x}}_i - \bar{\mathbf{x}}_t + \bar{\mathbf{x}}$. Тогда $(\alpha_i - \alpha_1)$ и $(\gamma_t - \gamma_1)$ оцениваются с использованием выражений из (11-25). Используя следующие данные, оцените полный набор коэффициентов для модели наименьших квадратов с фиктивной переменной:

| | $t = 1$ | $t = 2$ | $t = 3$ | $t = 4$ | $t = 5$ | $t = 6$ | $t = 7$ | $t = 8$ | $t = 9$ | $t = 10$ |
|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|----------|
| $i = 1$ | | | | | | | | | | |
| y | 21,7 | 10,9 | 33,5 | 22,0 | 17,6 | 16,1 | 19,0 | 18,1 | 14,9 | 23,2 |
| x_1 | 26,4 | 17,3 | 23,8 | 17,6 | 26,2 | 21,1 | 17,5 | 22,9 | 22,9 | 14,9 |
| x_2 | 5,79 | 2,60 | 8,36 | 5,50 | 5,26 | 1,03 | 3,11 | 4,87 | 3,79 | 7,24 |
| $i = 2$ | | | | | | | | | | |
| y | 21,8 | 21,0 | 33,8 | 18,0 | 12,2 | 30,0 | 21,7 | 24,9 | 21,9 | 23,6 |
| x_1 | 19,6 | 22,8 | 27,8 | 14,0 | 11,4 | 16,0 | 28,8 | 16,8 | 11,8 | 18,6 |
| x_2 | 3,36 | 1,59 | 6,19 | 3,75 | 1,59 | 9,87 | 1,31 | 5,42 | 6,32 | 5,35 |
| $i = 3$ | | | | | | | | | | |
| y | 25,2 | 41,9 | 31,3 | 27,8 | 13,2 | 27,9 | 33,3 | 20,5 | 16,7 | 20,7 |
| x_1 | 13,4 | 29,7 | 21,6 | 25,1 | 14,1 | 24,1 | 10,5 | 22,1 | 17,0 | 20,5 |
| x_2 | 9,57 | 9,62 | 6,61 | 7,24 | 1,64 | 5,99 | 9,00 | 1,75 | 1,74 | 1,82 |
| $i = 4$ | | | | | | | | | | |
| y | 15,3 | 25,9 | 21,9 | 15,5 | 16,7 | 26,1 | 34,8 | 22,6 | 29,0 | 37,1 |
| x_1 | 14,2 | 18,0 | 29,9 | 14,1 | 18,4 | 20,1 | 27,6 | 27,4 | 28,5 | 28,6 |
| x_2 | 4,09 | 9,56 | 2,18 | 5,43 | 6,33 | 8,27 | 9,16 | 5,24 | 7,92 | 9,63 |

Проверьте гипотезу о том, что (1) все эффекты «периода» равны нулю, (2) все эффекты «группы» равны нулю и (3) эффекты группы и периода вместе равны нулю. Используйте F -критерий в каждом случае.

6. Модель с двунаправленными случайными эффектами. Мы модифицируем модель со случайными эффектами при помощи включения специфических временных шоков. Итак,

$$y_{it} = \alpha + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it} + u_i + v_t,$$

где

$$\begin{aligned} E[\varepsilon_{it} | \mathbf{X}] &= E[u_i | \mathbf{X}] = E[v_t | \mathbf{X}] = 0, \\ E[\varepsilon_{it} u_j | \mathbf{X}] &= E[\varepsilon_{it} v_s | \mathbf{X}] = E[u_i v_t | \mathbf{X}] = 0 \quad \text{для всех } i, j, t, s, \\ \text{Var}[\varepsilon_{it} | \mathbf{X}] &= \sigma_\varepsilon^2, \quad \text{Cov}[\varepsilon_{it}, \varepsilon_{js} | \mathbf{X}] = 0 \quad \text{для всех } i, j, t, s, \\ \text{Var}[u_i | \mathbf{X}] &= \sigma_u^2, \quad \text{Cov}[u_i, u_j | \mathbf{X}] = 0 \quad \text{для всех } i, j, \\ \text{Var}[v_t | \mathbf{X}] &= \sigma_v^2, \quad \text{Cov}[v_t, v_s | \mathbf{X}] = 0 \quad \text{для всех } t, s. \end{aligned}$$

Выпишите полную ковариационную матрицу шоков для набора данных при $n = 2$ и $T = 2$.

7. Модель

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}$$

соответствует групповой модели регрессии с гетероскедастичностью ошибок из подраздела 9.7.2. Средние значения всех переменных равны нулю. Представленная ниже выборочная матрица вторых моментов получена по выборке из 20 наблюдений:

$$y_1 \quad y_2 \quad x_1 \quad x_2$$

$$\begin{matrix} y_1 \\ y_2 \\ x_1 \\ x_2 \end{matrix} \begin{bmatrix} 20 & 6 & 4 & 3 \\ 6 & 10 & 3 & 6 \\ 4 & 3 & 5 & 2 \\ 3 & 6 & 2 & 10 \end{bmatrix}.$$

- Вычислите две отдельные МНК-оценки для β , их выборочные дисперсии, оценки для σ_1^2 и σ_2^2 , а также R^2 для обеих регрессий.
 - Выполните тест множителей Лагранжа для гипотезы $\sigma_1^2 = \sigma_2^2$.
 - Вычислите двухшаговую FGLS-оценку для β и оценку ее выборочной дисперсии. Проверьте гипотезу, что β равно 1.
 - Проведите тест Вальда для равных дисперсий шоков.
 - Вычислите оценки β , σ_1^2 и σ_2^2 методом максимального правдоподобия, путем итерации FGLS-оценок в конвергенцию.
 - Проведите тест отношения максимального правдоподобия для проверки гипотезы о равенстве дисперсий шоков.
8. Предположим, что в модели с групповой гетероскедастичностью из подраздела 9.7.2 X_i одинакова для всех i . Чему равна оценка β , полученная при помощи обобщенного метода наименьших квадратов? Как бы вы вычислили оценку, если бы было необходимо оценить σ_i^2 ?
9. В приведенной ниже таблице представлена гипотетическая панель данных:

| | $i = 1$ | | $i = 2$ | | $i = 3$ | |
|----|---------|-------|---------|-------|---------|-------|
| t | y | x | y | x | y | x |
| 1 | 30,27 | 24,31 | 38,71 | 28,35 | 37,03 | 21,16 |
| 2 | 35,59 | 28,47 | 29,74 | 27,38 | 43,82 | 26,76 |
| 3 | 17,90 | 23,74 | 11,29 | 12,74 | 37,12 | 22,21 |
| 4 | 44,90 | 25,44 | 26,17 | 21,08 | 24,34 | 19,02 |
| 5 | 37,58 | 20,80 | 5,85 | 14,02 | 26,15 | 18,64 |
| 6 | 23,15 | 10,55 | 29,01 | 20,43 | 26,01 | 18,97 |
| 7 | 30,53 | 18,40 | 30,38 | 28,13 | 29,64 | 21,35 |
| 8 | 39,90 | 25,40 | 36,03 | 21,78 | 30,25 | 21,34 |
| 9 | 20,44 | 13,57 | 37,90 | 25,65 | 25,41 | 15,86 |
| 10 | 36,85 | 25,60 | 33,90 | 11,66 | 26,04 | 13,28 |

- Оцените модель с групповой гетероскедастичностью из подраздела 9.7.2. Включите оценку асимптотической дисперсии оценки коэффициентов наклона. Используйте двухшаговую процедуру, которая на втором шаге FGLS-оценки включает остатки, полученные из оценки методом наименьших квадратов объединенной регрессии.
- Выполните тесты Вальда и множителей Лагранжа для проверки гипотезы о том, что все дисперсии равны.

Эмпирические приложения

Как обычно, для выполнения представленных ниже примеров необходима эконометрическая компьютерная программа. Вычисления можно осуществить при помощи любой современной программы, поэтому мы не рекомендуем какую-либо конкретно.

1. Данные из табл. F10.4 приложения были использованы Грюнфельдом (Grunfeld (1958)) и многими другими исследователями позже, включая Zellner (1962, 1963) и Zellner и Huang (1962), чтобы изучить различные оценки для панельных данных и систем линейных регрессий. [См. Kleiber и Zeileis (2010).] Модель представляет собой уравнение инвестиций

$$I_{it} = \beta_1 + \beta_2 F_{it} + \beta_3 C_{it} + \varepsilon_{it}, \quad t = 1, \dots, 20, i = 1, \dots, 10,$$

где I_{it} — реальные валовые инвестиции фирмы i в год t , F_{it} — реальная стоимость фирмы — акции в обращении, C_{it} — реальная стоимость акционерного капитала.

Для целей текущей работы это сбалансированный набор панельных данных.

- a. Оцените объединенную регрессию.
 - b. Ссылаясь на результаты из пункта а, выясните, есть ли здесь свидетельства внутригрупповой корреляции. Вычислите устойчивые стандартные ошибки для нашей МНК-оценки объединенной модели и сравните их со стандартными.
 - c. Вычислите оценку с фиксированными эффектами для этих данных и далее, используя F -критерий, проверьте гипотезу о том, что все константы для 10 фирм одинаковы.
 - d. Используйте статистику множителей Лагранжа, чтобы проверить наличие общих эффектов в данных.
 - e. Вычислите оценку с однонаправленными случайными эффектами и приведите результаты всех оценок. Объясните разницу между этой спецификацией и спецификацией из пункта с.
 - f. Используйте тест Хаусмана, чтобы определить, какая из спецификаций (с фиксированными или случайными эффектами) более предпочтительна для этих данных.
2. Данные из табл. F6.1 приложения являются несбалансированной панелью по 25 авиакомпаниям США в дни до отмены государственного регулирования в 1970-х и 1980-х гг. Размеры группы колеблются от 2 до 15. В файле приведены данные по следующим переменным. (Названия переменных в файле данных созданы таким образом, чтобы отражать содержание переменных.)
- Общие затраты
 - Расходы на капитал, труд, топливо, материалы, имущество и оборудование
 - Меры цен для шести входных данных
 - Меры выпуска для шести входных данных
 - Выходные данные, измеренные годовой выручкой от пассажирских миль, преобразованные в порядковый номер авиакомпании
 - *Loadfactor* — средний процент использования производственных мощностей флота компании
 - *Stage* — средняя длина полета (перелета) в милях
 - *Points* — количество точек, обслуживаемых авиакомпанией

- $Year$ — календарный год
- $T = \text{Год} - 1969$
- TI — число наблюдений по авиакомпании, воспроизводимое каждый год

Используйте эти данные, чтобы построить модель затрат для услуг авиакомпаний. Допустите неоднородность в константах по авиакомпаниям в модели. Используйте спецификации с фиксированными и случайными эффектами и примените доступные статистические тесты, чтобы определить, какая модель будет более предпочтительной. Подходящей для начала анализа моделью затрат будет

$$\ln cost_{it} = \alpha_i + \sum_{k=1}^6 \beta_k \ln Price_{k,it} + \gamma \ln Output_{it} + \varepsilon_{it}.$$

Необходимо ввести линейную однородность для цен в функции затрат, что можно сделать, разделив пять из шести цен и общие затраты на шестую цену (выберите любую) и далее используя $\ln(cost/P_6)$ и $\ln(P_k/P_6)$ в регрессии. Вы также можете обобщить функцию затрат, включив квадратичный член в логарифм выпуска в функции. Транслогарифмическая модель включала бы квадраты и смешанные произведения цен и произведения логарифма выпуска с логарифмом цен. Данные включают три дополнительных фактора, которые могут оказать влияние на затраты, длину перелета, коэффициент нагрузки и число обслуживаемых точек. Включите их в вашу модель и примените подходящую тестовую статистику, чтобы проверить, действительно ли они нужны для определения (логарифма) общих затрат.

ГЛАВА 12

Подходы к оцениванию в эконометрике

12.1. Введение

С этой главы начинается обсуждение методов оценивания. Современная эконометрика предлагает практикующему исследователю широкий спектр методов оценивания — от подробно параметризованных техник, основанных на правдоподобии, на одном полюсе доступного спектра до свободно заданных непараметрических методов, предполагающих немногим более чем наличие взаимосвязи между переменными, — на другом и богатое разнообразие между ними. Даже опытному исследователю простительно сомнение относительно выбора из такого длинного списка. Конечно, исчерпывающий ответ на этот вопрос выходит за рамки данной книги, но несколько общих принципов могут быть выделены. Авторы недавних исследований склоняются к использованию минимально возможного числа необоснованных предположений. Это объясняет распространение GMM-оценки в ситуациях, когда подробной параметризации на основе правдоподобия можно избежать и перейти на робастные методы оценки при наличии гетероскедастичности и серийной корреляции. (Интерес представляет тот факт, что происходит это параллельно с развитием возможностей численного оценивания, которое приводит к *растущему* принятию очень сильно параметризованных байесовских методов.)

В общем случае продвижение от параметрических (через полупараметрические) к **непараметрическим методам оценки** связано с отказом от строгих предпосылок ценой ослабления выводов, которые могут быть сделаны на основе анализа данных. Это особенно заметно при анализе моделей дискретного выбора, которому посвящена наиболее активно развивающаяся часть эконометрической литературы. (Эта тема рассматривается в главе 17.) Формальная логит- или пробит-модель позволяет получать оценки вероятностей, предельных эффектов и некоторые другие полезные результаты ценой навязывания данным предположения нормального или логистического распределения. **Полупараметрические и непараметрические оценки** позволяют отказаться от этого предположения, но часто взамен дают лишь диапазоны для вероятностей, а во многих случаях и не позволяют оценивать вероятности или интересующие исследователя предельные эффекты. Однако получаемые выводы имеют преимущество с точки зрения устойчивости результатов. [См., например, симпозиум в Angrist (2001), посвященный обсуждению общих тенденций в этой области.]

Другой площадкой для сравнения различных подходов являются свойства получаемых оценок. В классе оценок можно определить «лучший» (наиболее эффективный) способ использования данных. (См. приложение в примере 12.2.) Иногда сравнения также допустимы между классами. Например,

когда оцениваются одни и те же параметры — этот результат еще следует получить, — лучшая параметрическая оценка превзойдет лучшую полупараметрическую. В этом, конечно, состоит ценность информации. Обратная сторона сравнения состоит в том, что полупараметрическая оценка окажется предпочтительной в том случае, если параметрическая модель специфицирована с ошибкой, к которой робастна полупараметрическая оценка (в отличие от параметрической).

Научные школы давно заняли свои позиции в этом споре. Сторонники **байесовских методов оценки** часто принимают почти теологическую точку зрения в своей критике коллег-классиков. [См., например, Poirier (1995).] Современные практикующие исследователи, как правило, более прагматичны. Байесовские методы оценки зарекомендовали себя как набор инструментов, который во многих случаях предоставляет элегантные и прозрачные решения не решенных до сих пор проблем. Так, **основанные на симуляциях методы оценки**, продвигаемые во множестве статей Chib, Greenberg (например, 1996), дали возможность решить множество сложных с точки зрения численной реализации проблем¹. Аргументам, относящимся к методологическим преимуществам того или иного подхода, уделяется гораздо меньше внимания, чем раньше.

Главы 2–7 этой книги были посвящены модели классической регрессии и конкретному методу ее оценки — методу (линейных и нелинейных) наименьших квадратов. В этой и четырех следующих главах мы рассмотрим несколько общих подходов к оцениванию, используемых в различных ситуациях. Эта глава содержит обзор некоторых методов, относящихся к трем приведенным нами общим подходам. Глава 13 посвящена **обобщенному методу моментов**, который является центральным среди полупараметрических методов оценки. В главе 14 представлен метод **максимального правдоподобия** — общее основание для параметрического, классического оценивания в эконометрике. Глава 15 посвящена методам оценки, основанным на симуляциях, и методам бутстрэпа (самонастройки). Это развивающийся в последнее время набор инструментов, которые стали доступны благодаря развитию компьютерных технологий и сделали простыми и прозрачными многие оценки, использовавшиеся ранее лишь изредка ввиду сложности, связанной с численными методами. Наконец, в главе 16 представлены методы байесовской эконометрики.

Приведенный тут набор инструментов далек от исчерпывающего. Мы выбрали методы, составляющие основное направление эконометрики. Конечно, могут быть рассмотрены и другие методы. [См., например, длинный их список в Mittelhammer, Judge, Miller (2000).] Практически всем методам посвящены прекрасные монографии. В этой главе мы приведем несколько при-

¹Распространение байесовских методов в эконометрике может быть переоценено. Они достаточно широко представлены в таких журналах, как *Journal of Econometrics*, *Journal of Applied Econometrics*, *Journal of Business and Economic Statistics* и т.д. С другой стороны, из шести важных изложений эконометрической теории, вышедших в 2000 г., четыре (Hayashi, Ruud, Patterson, Davidson) вообще не упоминают байесовские методы, в наборе из 32 эссе (Baltagi) этому вопросу посвящено только одно, а в единственном эссе, сопоставляющем подходы (Mittelhammer et al.), байесовской постановке посвящено около 10% (70) страниц, но все они посвящены метатеории линейной регрессионной модели, а не более замысловатым приложениям, получившим признание в профессиональной литературе.

ложений — собственных и из академической литературы, чтобы продемонстрировать набор инструментов, актуальный для эконометрической практики. Мы начнем с раздела 12.2, посвященного параметрическим подходам, преимущественно методу максимального правдоподобия. Поскольку данный метод обсуждается в оставшейся части книги, этот раздел краток. Раздел 12.2 также представляет байесовское оценивание, которое в традиционной форме так же сильно параметризовано, как и оценивание методом максимального правдоподобия. Раздел 12.3 посвящен полупараметрическим методам оценки. GMM-оценки являются предметом обсуждения в главе 13, так что тут они только вводятся. Представлен также метод наименьших абсолютных отклонений. Кроме того, дан обзор последних прикладных работ. В разделе 12.4 описывается непараметрический подход к оценке моделей. Подробно обсуждается основной инструмент — оценка ядерной плотности, который затем применяется в регрессионном анализе. Также приводятся два приложения. Фокусируясь на приложениях, данная глава содержит мало статистической теории, лежащей в основе этих инструментов, таких как их асимптотические свойства. (Эти результаты, конечно, широко представлены в литературе.) Мы обратимся к вопросу о свойствах оценок в конце главы, в разделе 12.5, затем более подробно в главах 13–16.

12.2. Параметрические методы оценки и тестирование статистических гипотез

Параметрические методы оценки начинаются с полной спецификации **плотности** или вероятностной модели, которая представляет собой **механизм порождения данных** интересующей нас случайной величиной. В рассмотренных выше приложениях мы можем рассматривать совместную плотность скалярной случайной величины y и случайного вектора \mathbf{x} как заданные через

$$f(y, \mathbf{x}) = g(y | \mathbf{x}, \beta) \times h(\mathbf{x} | \theta) \quad (12-1)$$

с неизвестными параметрами β и θ . Продолжая анализ приложения, начатый в главе 2, рассмотрим линейную регрессионную модель с нормально распределенными ошибками. Это предположение задает **условную плотность**, т.е. генеральную совокупность, из которой берется наблюдение

$$y_i | \mathbf{x}_i \sim N[\mathbf{x}'_i \beta, \sigma^2].$$

Для полного определения генеральной совокупности остается только узнать конкретные значения, которые принимают *неизвестные*, но *фиксированные* параметры. Имея их, мы получаем полностью заданное условное распределение — среднее, дисперсию, вероятности определенных событий и т.д. (Безусловная плотность для условных переменных обычно не представляет интереса.) Таким образом, отличительной особенностью этого подхода к моделированию является спецификация плотности и ее отдельных особенностей (параметров).

Пространство параметров для параметрической модели — это множество допустимых значений параметров, которые удовлетворяют некоторой

предварительной спецификации модели. Например, в специфицированной выше регрессионной модели K коэффициентов наклона могут принимать любые действительные значения, но дисперсия должна быть положительным числом. Таким образом, пространством параметров этой модели является $[\beta, \sigma^2] \in R^K \times R_+$. «Процедура оценки» в такой ситуации состоит в спецификации критерия для ранжирования точек в пространстве параметров, а затем в выборе точки (точечной оценки) или множества точек (интервальной оценки), которые оптимизируют этот критерий, т.е. имеют наилучший ранг. Таким образом, например, мы выбираем минимум квадратов ошибок как один из **критериев оценки** для линейной модели. «Выведение» в такой постановке — это процесс, в ходе которого некоторые области (уже специфицированного) пространства параметров признаются не содержащими неизвестных параметров. Однако на практике мы обычно определяем критерий, а затем утверждаем, что согласно этому критерию наличие истинных параметров в определенных областях *маловероятно*.

12.2.1. Классический подход к оценке, основанный на функции правдоподобия

Самым распространенным (с большим отрывом) классом параметрических оценок, используемых в эконометрике, являются оценки максимального правдоподобия. В основе их философии заложена идея «выборочной информации». Когда плотность выборки наблюдений полностью специфицирована, за исключением неизвестных параметров, совместная плотность этих наблюдений (предполагается их независимость) представляет собой **функцию правдоподобия**

$$f(y_1, y_2, \dots, x_1, x_2, \dots) = \prod_{i=1}^n f(y_i, x_i | \beta, \theta). \quad (12-2)$$

Эта функция включает всю содержащуюся в выборке информацию о генеральной совокупности, из которой были извлечены эти наблюдения. Стратегия использования этой информации в процессе оценивания характеризует оценку методом максимального правдоподобия. **Оценка методом максимального правдоподобия** [Fisher (1925)] — это функция данных, которая (как подразумевает ее название) максимизирует функцию правдоподобия (или, поскольку обычно это более удобно, ее логарифм). Мотивацию для этого подхода легко проиллюстрировать в случае дискретной случайной величины. В такой постановке функция правдоподобия задает совместную вероятность наблюдений из выборки, и оценка методом максимального правдоподобия — это функция информации из выборки, которая делает наблюдаемые данные наиболее вероятными (по крайней мере, по этому критерию). Хотя аналогия наиболее интуитивна для дискретной переменной, она переносится и на непрерывный случай. Поскольку эта оценка является предметом рассмотрения достаточно длинной главы 14, мы отложим формальное обсуждение до этого момента и рассмотрим вместо этого два приложения, чтобы проиллюстрировать технику и некоторые ее особенности.

Пример 12.1. Линейная регрессионная модель

Метод наименьших квадратов взвешивает отрицательные и положительные отклонения одинаково и придает при оценке непропорциональный вес большим отклонениям. Это свойство может быть как преимуществом, так и недостатком в зависимости от процесса, порождающего данные. Для нормально распределенных ошибок это именно тот метод, который требуется для наиболее эффективного использования данных. Если данные порождены нормальным распределением, логарифм функции правдоподобия равен

$$\ln L = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\beta)' (\mathbf{y} - \mathbf{X}\beta).$$

Вы легко можете показать, что метод наименьших квадратов является именно той оценкой, которая подходит для этой модели. Максимизация функции означает минимизацию экспоненты, чего можно добиться, минимизируя квадратичные отклонения для β , после чего $e'e/n$ следует в качестве оценки для σ^2 .

Если подходящим распределением является ненормальное, — возможно, на основе наблюдения, что хвосты распределения ошибок слишком толстые, — см. пример 4.7 и подраздел 14.9.5а, — то существует три пути для продолжения анализа. Во-первых, как мы отметили, состоятельность оценки методом наименьших квадратов устойчива к подобной ошибке спецификации постольку, поскольку условное среднее значение ошибок равно нулю. Однако некоторая коррекция стандартных ошибок требуется для получения правильных выводов. Во-вторых, может возникнуть желание использовать оценку с лучшими свойствами в малых выборках. В качестве такого варианта в подразделе 12.3.2 рассматривается оценка методом наименьших абсолютных отклонений. Наконец, можно рассмотреть другое распределение, учитывающее наблюдаемое расхождение. Например, Ruud (2000) подробно рассматривает линейную регрессионную модель с t -распределением ошибок и с ν степенями свободы. Если число степеней свободы ν конечно, эта случайная величина имеет большую дисперсию, чем нормальная. Какой вариант следует предпочесть? Третий подход наименее привлекателен. Если нормальное распределение неуместно, сложно представить себе механизм, в результате которого подойдет t -распределение. Оценку методом наименьших абсолютных отклонений (LAD) можно предпочесть при наличии небольшого числа наблюдений. В противном случае оценка методом наименьших квадратов останется предпочтительной с учетом того факта, что стандартные инструменты для статистических выводов, скорее всего, будут неточны. Текущая практика состоит в применении первой стратегии.

Пример 12.2. Модель стохастической границы

Модель стохастической границы, подробно обсуждаемая в главе 19, — это модель распределения ошибок, похожая на регрессионную и являющаяся

асимметричной и бесспорно ненормальной. Условная плотность для зависимой переменной в этой модели принимает вид

$$f(y | \mathbf{x}, \boldsymbol{\beta}, \sigma, \lambda) = \frac{\sqrt{2}}{\sigma\sqrt{\pi}} \exp \left[\frac{-(y - \alpha - \mathbf{x}'\boldsymbol{\beta})^2}{2\sigma^2} \right] \Phi \left(\frac{-\lambda(y - \alpha - \mathbf{x}'\boldsymbol{\beta})}{\sigma} \right).$$

Из этого следует логарифм функции правдоподобия для модели

$$\ln L = -n \ln \sigma - \frac{n}{2} \ln \frac{2}{\pi} - \frac{1}{2} \sum_{i=1}^n \left(\frac{\varepsilon_i}{\sigma} \right)^2 + \sum_{i=1}^n \ln \Phi \left(\frac{-\lambda \varepsilon_i}{\sigma} \right).$$

Для этой модели существует, по крайней мере, две параметрические оценки. Оценка максимального правдоподобия обсуждается в подразделе 19.2.4. В работе Грина (Greene (2007)) предлагается следующая оценка методом моментов: для коэффициентов наклона регрессии, за исключением константы, используем метод наименьших квадратов. Для параметров α , σ и λ на основе второго и третьего моментов остатков, полученных методом наименьших квадратов, и константы решаем

$$\begin{aligned} m_2 &= \sigma_v^2 + [1 - 2/\pi]\sigma_u^2, \\ m_3 &= (2/\pi)^{1/2}[1 - 4/\pi]\sigma_u^3, \\ a &= \alpha + (2/\pi)^2\sigma_u, \end{aligned}$$

где $\lambda = \sigma_u/\sigma_v$ и $\sigma^2 = \sigma_u^2 + \sigma_v^2$.

Обе оценки полностью параметрические. Оценка методом максимального правдоподобия — по вышеуказанным причинам. Оценка методом моментов (см. раздел 13.2) применима только для этого распределения. Какая оценка предпочтительнее? Как мы увидим в главе 19, обе оценки состоятельны и асимптотически нормально распределены. По теореме Рао–Крамера, оценка методом максимального правдоподобия имеет меньшую асимптотическую дисперсию. Никакая из них не обладает свойствами оптимальности в малых выборках. Таким образом, единственным преимуществом оценки, полученной методом моментов, является то, что ее можно получить с помощью любого стандартного компьютерного статистического пакета и калькулятора, тогда как оценка методом максимального правдоподобия требует специализированного приложения (однако довольно распространенного).

12.2.2. Моделирование совместных распределений копула-функциями

Спецификация функции правдоподобия заставляет исследователя принять достаточно строгие предположения о распределении интересующей случайной величины. Выгода, конечно, состоит в открывающихся в таком случае более широких возможностях для получения статистических выводов. Однако когда интерес представляют несколько случайных величин, как в случае с совместным решением домохозяйства об использовании меди-

цинского обслуживания, формулировка полного правдоподобия включает спецификацию частных распределений, что может быть удобно, и полную спецификацию совместного распределения, что, скорее всего, связано с меньшим удобством. В типичной ситуации модель может включать две похожие случайные величины и некорректно специфицированную корреляцию между ними. Этот случай неявно включает спецификацию частных распределений. Совместное распределение необходимо для моделирования ненулевой корреляции. Подход **копула-функций**, или **функций-связок**, или **связывающих функций**, представляет механизм который исследователь может использовать для избежания этой ситуации.

Триведи и Циммер (Trivedi, Zimmer (2007)) предлагают несколько приложений, подходящих под это описание.

- Финансовые организации часто озабочены ценами различных связанных (зависимых) активов. Стандартное предположение о многомерной нормальности проблематично ввиду GARCH-эффектов (см. раздел 20.13) и тяжелых хвостов распределений. Тогда как спецификация приемлемых частных распределений не вызывает вопросов, спецификация совместного распределения связана с существенными трудностями. Клугман и Парса (Klugman и Parsa (2000)) демонстрируют приложение.
- Во множестве микроэконометрических приложений типичные частные распределения не могут быть объединены в естественное совместное распределение. Двумерная модель числа исходов проанализирована в работе Манкина и Триведи (Munkin и Trivedi (1999)), и одно из приложений мы рассматриваем в следующем примере.
- В линейной модели самоотбора из главы 19 необходимое совместное распределение — это часть большей модели. Функция правдоподобия для наблюдаемого исхода включает совместное распределение интересующей переменной, отработанных часов, заработных плат, дохода и т.д. и вероятности наблюдения. Типичное приложение основывается на совместном нормальном распределении. Смит (Smith (2003, 2005)) предлагает некоторые приложения, в которых гибкое, основанное на копуле представление более уместно. [В одном из первых интересных приложений моделирования с помощью копул, где использовалась другая терминология, так как оно предшествует эконометрической литературе, Ли (Lee (1983)) моделировал переменную исхода как нормальную, вероятность наблюдения как логистическую, а связь между ними как то, что далее называется гауссовой копулой.]

Хотя первые упоминания в статистической литературе относятся к выкладкам в работе Шкляра (Sklar (1973)), приложения в эконометрике и финансах достаточно недавние — большинство из них появились после 2000 г. [См. исчерпывающее описание в прекрасном обзоре Триведи и Циммера (Trivedi, Zimmer (2007)).]

Рассмотрим задачу моделирования, в которой частные кумулятивные функции распределения (КФР) двух случайных величин специфицируются как $F_1(y_1 | \bullet)$ и $F_2(y_2 | \bullet)$, где условием являются информация из выборки (данные) и параметры, обозначаемые \bullet . На минуту предположим, что это непрерывные случайные величины, удовлетворяющие всем аксиомам вероятно-

сти. Совместная кумулятивная функция распределения обозначается $F_{12}(y_1, y_2 | \bullet)$. Бинарная копула-функция (результаты обобщаются на многомерный случай) — это функция $C(u_1, u_2)$, определенная на единичном квадрате $[(0 \leq u_1 \leq 1) \times (0 \leq u_2 \leq 1)]$, которая удовлетворяет

- (1) $C(1, u_2) = u_2$ и $C(u_1, 1) = u_1$,
- (2) $C(0, u_2) = C(u_1, 0) = 0$,
- (3) $\partial C(u_1, u_2)/\partial u_1 \geq 0$ и $\partial C(u_1, u_2)/\partial u_2 \geq 0$.

Это свойства двумерных кумулятивных функций распределения случайных величин u_1 и u_2 , ограниченных в единичном квадрате. Из этого следует, что копула-функция — это двумерная кумулятивная функция распределения, определенная на единичном квадрате со стандартными, равномерными, одномерными безусловными распределениями на единичном интервале [т.е. свойство (1)]. Чтобы выгодно использовать это соотношение, отметим, что кумулятивная функция распределения случайной величины $F_1(y_1 | \bullet)$ сама по себе равномерно распределенная случайная величина. Это **фундаментальное преобразование вероятности**, которое мы используем для генерирования случайных величин. (См. раздел 15.2.) В **теореме Шкляра** (Sklar (1973)) частные кумулятивные функции распределения играют роли u_1 и u_2 . Теорема говорит о том, что существует связывающая (копула) функция $C(\dots)$ такая, что

$$F_{12}(y_1, y_2 | \bullet) = C[F_1(y_1 | \bullet), F_2(y_2 | \bullet)].$$

Если $F_{12}(y_1, y_2 | \bullet) = C[F_1(y_1 | \bullet), F_2(y_2 | \bullet)]$ непрерывна и частные КФР имеют квантильные (обратные) функции $F_j^{-1}(u_j)$, где $0 \leq u_j \leq 1$, тогда копула-функция может быть представлена как

$$\begin{aligned} F_{12}(y_1, y_2 | \bullet) &= F_{12}[F_1^{-1}(u_1 | \bullet), F_2^{-1}(u_2 | \bullet)] \\ &= \text{Prob}[U_1 \leq u_1, U_2 \leq u_2] \\ &= C(u_1, u_2). \end{aligned}$$

Словом, теорема говорит о том, что совместная плотность может быть записана как связывающая функция, оцененная в двух кумулятивных функциях распределения.

Копула-функции позволяют аналитикам собирать совместные распределения, когда специфицированы могут быть только частные. Для того чтобы заполнить требуемый элемент корреляции между случайными величинами, копула-функция записывается как

$$F_{12}(y_1, y_2 | \bullet) = C[F_1(y_1 | \bullet), F_2(y_2 | \bullet), \theta],$$

где θ — это параметр зависимости. В таком случае для непрерывных случайных величин совместная функция плотности — это смешанная частная производная

$$\begin{aligned} f_{12}(y_1, y_2 | \bullet) &= c_{12}[F_1(y_1 | \bullet), F_2(y_2 | \bullet), \theta] \\ &= \partial^2 C[F_1(y_1 | \bullet), F_2(y_2 | \bullet), \theta] / \partial y_1 \partial y_2 \quad (12-3) \\ &= [\partial^2 C(., ., \theta) / \partial F_1 \partial F_2] f_1(y_1 | \bullet) f_2(y_2 | \bullet). \end{aligned}$$

Функция логарифма правдоподобия теперь может быть построена с использованием логарифмов правых частей (12-3). Взяв логарифмы (12-3), мы понимаем преимущества подхода связывающих функций. Вклад совместного наблюдения в логарифм правдоподобия — это

$$\ln f_{12}(y_1, y_2 | \bullet) = \ln[\partial^2 C(., ., \theta) / \partial F_1 \partial F_2] + \ln f_1(y_1 | \bullet) + \ln f_2(y_2 | \bullet).$$

Некоторые из часто используемых в приложениях копула-функций — это

$$\text{Product : } C[u_1, u_2, \theta] = u_1 \times u_2,$$

$$\text{FGM : } C[u_1, u_2, \theta] = u_1 u_2 [1 + \theta(1 - u_1)(1 - u_2)],$$

$$\text{Gaussian : } C[u_1, u_2, \theta] = \Phi_2[\Phi^{-1}(u_1), \Phi^{-1}(u_2), \theta],$$

$$\text{Clayton : } C[u_1, u_2, \theta] = [u_1^{-\theta} + u_2^{-\theta} - 1]^{-1/\theta},$$

$$\text{Frank : } C[u_1, u_2, \theta] = \frac{1}{\theta} \ln \left[1 + \frac{\exp(\theta u_1 - 1) \exp(\theta u_2 - 1)}{\exp(\theta) - 1} \right].$$

Копула произведения подразумевает, что случайные величины независимы, поскольку она означает, что совместная КФР равна произведению частных функций. В FGM (Fairlie, Gumbel, Morgenstern)-копуле можно заметить, что $\theta = 0$ приводит к копуле произведения, или независимости. Тот же результат может быть получен для копулы Клейтона. В гауссовой функции копула — это двумерная нормальная КФР, если частные распределения изначально нормальны. Ключевой момент состоит в том, что для построения копула-функций частные распределения не обязаны быть нормальными, поскольку частные КФР могут быть специфицированы. (Параметр зависимости — это не корреляция между переменными. Триведи и Циммер приводят преобразования θ , которые тесно связаны с корреляциями для каждой приведенной копулы.)

Суть копула-техники состоит в том, что исследователь может специфицировать и анализировать частные распределения и копула-функцию отдельно. Функция правдоподобия получается из КФР [или плотностей, поскольку дифференцирование в (12-3) сведет совместную плотность к удобной функции частных плотностей] и копулы.

Пример 12.3. Совместное моделирование пары числа событий

Стандартный подход регрессионного моделирования случайной величины y , представляющей собой число событий, — это пуассоновская регрессионная модель

$$\text{Prob}[Y = y | \mathbf{x}] = \exp(-\lambda) \lambda^y / y!, \text{ где } \lambda = \{\exp(\mathbf{x}'\boldsymbol{\beta}), y = 0, 1, \dots\}$$

Более замысловатые спецификации используют отрицательную биномиальную модель (версия 2, ОБ2)

$$\text{Prob}[Y = y | \mathbf{x}] = \frac{\Gamma(y + \alpha)}{\Gamma(\alpha)\Gamma(y + 1)} \left(\frac{\alpha}{\lambda + \alpha} \right)^\alpha \left(\frac{\lambda}{\lambda + \alpha} \right)^y, y = 0, 1, \dots$$

где α — это параметр избыточной дисперсии. (См. раздел 18.4.) Удовлетворительная, подходящая спецификация двумерных исходов продолжает оставаться темой для исследований. Первые предложения основывались на модели смешанных латентных (ненаблюдаемых) переменных

$$\begin{aligned}y_1 &= z + w_1, \\y_2 &= z + w_2,\end{aligned}$$

где w_1 и w_2 имеют распределения пуассона, или ОБ2, специфицированные ранее с условными средними λ_1 и λ_2 , и z рассматривается как ненаблюдаемая пуассоновская или ОБ-переменная. Такая формулировка приводит к корреляции между переменными, но является неудовлетворительной, поскольку корреляция обязана быть положительной. В очевидном приложении, y_1 это число посещений врача, а y_2 — это число посещений больницы. Они могут быть отрицательно коррелированы. Манкин, Триведи (Munkin, Trivedi (1999)) специфицировали совместность в функциях условных средних в форме ненаблюдаемой общей неоднородности

$$\lambda_j = \exp(\mathbf{x}'_j \beta_j + \varepsilon),$$

где ε — общая для двух функций. Кэмерон и др. (Cameron et al. (2004)) использовали двумерную копулу для анализа австралийских данных об опросных и фактических посещениях врачей (последние собирает Health Insurance Commission). Они сделали две поправки к предыдущей модели, изложенной выше. Во-первых, они адаптировали базовую формулировку копулы к этим дискретным переменным. Во-вторых, их интересовало не число фактических или опросных визитов, а разность между ними. И то и другое являются естественными модификациями базовой модели копулы.

12.3. Полупараметрические методы оценки

Полупараметрические методы оценки основываются на меньшем количестве предположений, чем параметрические. В целом предположения о распределениях исключаются и оценка выводится из более общих характеристик генеральной совокупности. Интуиция подсказывает два (верных) вывода. Во-первых, полупараметрические оценки будут более устойчивыми (робастными), чем параметрические, они сохранят свои свойства, в частности состоятельность, для более широкого набора спецификаций. Рассмотрим наиболее знакомый нам пример. Оценка коэффициента наклона методом наименьших квадратов состоятельна, когда данные являются доброкачественными и ошибки не коррелированы с регрессорами. Это верно даже для функции границы из примера 12.2, которая имеет асимметричные, ненормальные ошибки. Но, во-вторых, за эту устойчивость нужно расплачиваться. Предположение о распределении обычно делает эффективную оценку более предпочтительной по сравнению с устойчивой. Лучшая устой-

чивая оценка в своем классе обычно хуже параметрической в том случае, когда предположение о распределении выполняется. В постановке задачи о функции границы метод наименьших квадратов может быть устойчив для коэффициентов наклона, и это наиболее эффективная оценка, использующая только ортогональность ошибок и регрессоров, но она будет хуже оценки методом максимального правдоподобия, когда корректным является предположение о состоящем из двух частей нормальном распределении.

12.3.1. GMM-оценка в эконометрике

Многие из недавних экономических приложений основаны на оценке **методом моментов**. Обобщенный метод моментов берет начало от множества моделей, основанных на моментных уравнениях $E[m(y_i, x_i, \beta)] = 0$, где набор уравнений специфицирует соотношение, выполняющееся в генеральной совокупности. Мы использовали одно из них в предыдущем параграфе. Использование оценки наименьших квадратов может быть мотивировано тем, что ключевым является предположение $E[x_i(y_i - x_i'\beta)] = 0$. Оценка получается при поиске параметра \mathbf{b} , который воспроизводит теоретическое соотношение $(1/n)\sum_i[x_i(y_i - x_i'\mathbf{b})] = 0$. Это, конечно, нормальные уравнения для метода наименьших квадратов. Отметим, что оценка специфицирована без использования каких-либо предположений о распределениях. Оценка методом моментов является предметом изучения в главе 13, так что мы отложим ее подробный анализ до этого времени.

12.3.2. Оценка методом максимального эмпирического правдоподобия

Методы эмпирического правдоподобия предлагаются в качестве полупараметрической альтернативы методам максимального правдоподобия. Как мы скоро увидим, оценка тесно связана с оценкой GMM. Пусть π_i обозначает в общем случае вероятность того, что $y_i|x_i$ принимает в выборке реализовавшееся значение. Интуиция (верно) подсказывает, что за неимением другой информации π_i будет равно $1/n$. **Функция эмпирического правдоподобия** имеет вид

$$EL = \prod_{i=1}^n \pi_i^{1/n}.$$

Оценка методом максимального эмпирического правдоподобия максимизирует EL . Другими словами, мы максимизируем логарифм эмпирического правдоподобия

$$ELL = \frac{1}{n} \sum_{i=1}^n \ln \pi_i.$$

Как задаче максимизации, ей не хватает структуры, чтобы она имела решение, решения для π_i неограниченны. Если мы наложим ограничение о том, что π_i — это вероятности, которые суммируются в единицу, то сможем использовать лагранжиан для решения оптимизационной задачи

$$ELL = \left[\frac{1}{n} \sum_{i=1}^n \ln \pi_i \right] + \lambda \left[1 - \sum_{i=1}^n \pi_i \right].$$

Это слегка ограничивает задачу, поскольку при $0 < \pi_i < 1$ и $\sum_i \pi_i = 1$ решение, предложенное ранее, становится очевидным. (Ничто в задаче не отличает различные π_i , так что они все должны быть равны друг другу.) Подстановка этого результата в производную по любой конкретной π_i приводит к оставшемуся результату $\lambda = 1$.

Максимизационная задача становится осмысленной, когда мы накладываем структуру на данные. Для того чтобы подтвердить это, вспомним пример 7.6, уравнение нелинейной регрессии для *Income* для данных из немецкой социально-экономической панели (German Socioeconomic Panel), которую мы специфицировали как

$$E[\text{Income} | \text{Age}, \text{Sex}, \text{Education}] = \exp(\mathbf{x}'\boldsymbol{\beta}) = h(\mathbf{x}, \boldsymbol{\beta}).$$

Для примера допустим, что переменная *Education* может быть эндогенной в этом уравнении, но нам доступен набор инструментов \mathbf{z} , скажем (*Age, Health, Sex, MarketCondition*). Мы предположили, что инструментов больше (4), чем включенных переменных (3), так что параметры будут переопределены (и пример будет достаточно сложным, чтобы представлять интерес). (См. разделы 8.3.4 и 8.6.) Условия ортогональности для нелинейной оценки инструментальных переменных состоят в том, что ошибки не коррелированы с инструментальными переменными, так что

$$E\{\mathbf{z}_i[\text{Income}_i - h(\mathbf{x}_i, \boldsymbol{\beta})]\} = E[\mathbf{m}_i(\boldsymbol{\beta})] = \mathbf{0}.$$

Решение этой проблемы нелинейным методом наименьших квадратов было получено в разделе 8.6. Оценка GMM будет минимизировать целевую функцию по $\boldsymbol{\beta}$:

$$q = \bar{\mathbf{m}}'(\boldsymbol{\beta}) \mathbf{A} \bar{\mathbf{m}}(\boldsymbol{\beta}),$$

где \mathbf{A} — это выбранная матрица весов. Обратите внимание, что в нашем примере, учитывая константу, $\boldsymbol{\beta}$ включает четыре элемента, а моментных условий 5, так что параметры переопределены.

Если, однако, мы наложим ограничения, вытекающие из наших моментных условий, на функцию эмпирического правдоподобия, то получим моментное условие для генеральной совокупности

$$\left[\sum_{i=1}^n \pi_i \mathbf{z}_i (\text{Income}_i - h(\mathbf{x}_i, \boldsymbol{\beta})) \right] = \mathbf{0}.$$

(Вероятности характеризуют генеральную совокупность, так что это ожидаемое значение.) Из этого получаем логарифм ограниченного эмпирического правдоподобия

$$ELL = \left[\frac{1}{n} \sum_{i=1}^n \ln \pi_i \right] + \lambda \left[1 - \sum_{i=1}^n \pi_i \right] + \boldsymbol{\gamma}' \left[\sum_{i=1}^n \pi_i \mathbf{z}_i (\text{Income}_i - h(\mathbf{x}_i, \boldsymbol{\beta})) \right].$$

Теперь функция максимизируется по π_i , λ , $\boldsymbol{\beta}$ (K элементов) и $\boldsymbol{\gamma}$ (L элементов, число инструментальных переменных). В решении значения π_i представляют собой, по сути, набор весов. Кэмерон, Триведи (Cameron, Trivedi (2005,

р. 205)) приводят решение для $\hat{\pi}_i$, выраженное через (β, γ) , и снова демонстрируют, что $\lambda = 1$. Подставляя их в функцию ELL , получаем функцию от γ и β , которую еще нужно максимизировать.

Оценка методом эмпирического правдоподобия имеет те же асимптотические свойства, что и оценка GMM. (Это естественно, учитывая схожесть критериев оценивания, в конечном счете они обе фокусируются на моментных уравнениях.) Есть основание считать, что, по крайней мере в некоторых случаях, свойства оценки методом эмпирического правдоподобия в конечных выборках могут быть лучше, чем у GMM. Обзор представлен в работе Имбенса (Imbens (2002)). Одной рассматриваемой модификацией процедуры может быть замена функции в $(1/n)\sum_i \ln \pi_i$ мерой **энтропии**:

$$Entropy = (1/n)\sum_i \pi_i \ln \pi_i.$$

Оценка методом **максимальной энтропии** приводится в работах Голана, Джаджа, Миллера (Golan, Judge, Miller (1996)) и Голана (Golan (2009)).

12.3.3. Оценка методом наименьших абсолютных отклонений и квантильная регрессия

Оценка методом наименьших квадратов может быть сильно искажена в малых выборках из-за наличия в них выбросов. Например, недавние приложения в микроэкономике и финансовой экономике, рассматривающие распределения с толстыми хвостами, будут особенно сильно подвержены влиянию именно этих наблюдений. (Конечно, в становящихся обычными финансовых приложениях, рассматривающих сотни тысяч наблюдений, это обсуждение излишне.) Эти приложения привели к тому, что были предложены «устойчивые» оценки, не подверженные влиянию выбросов. Одна из них, оценка наименьших абсолютных отклонений, или оценка LAD, обсуждаемая в подразделе 7.3.1, также интересна сама по себе в качестве оценки функции условной медианы в модифицированной модели

$$Med[y|\mathbf{x}] = \mathbf{x}'\beta_{.50}.$$

Таким образом, вместо того чтобы представлять устойчивую альтернативу методу наименьших квадратов для оценки коэффициентов наклонов $E[y|\mathbf{x}]$, LAD оценивает другую характеристику генеральной совокупности. Это, в сущности, полупараметрическая спецификация в том отношении, что она задает только конкретную характеристику распределения, его медиану, но не само распределение. Она также специфицирует условную медиану как линейную функцию \mathbf{x} .

Медиана, в свою очередь, лишь один из представляющих интерес квантилей. Если обобщить модель для других квантилей условного распределения, мы получим

$$Q[y|\mathbf{x}, q] = \mathbf{x}'\beta_q \text{ так, что } Prob[y \leq \mathbf{x}'\beta_q|\mathbf{x}] = q, 0 < q < 1.$$

Это, по сути, непараметрическая спецификация. Никаких предположений не делается относительно распределения $y|\mathbf{x}$ или его условной дисперсии.

Тот факт, что q может непрерывно изменяться (строго) между нулем и единицей, означает, что существует бесконечное множество «векторов параметров». Представляется разумным рассмотреть коэффициенты, которые мы можем записать как $\beta(q)$, скорее не как «параметров», как мы поступаем в случае линейной регрессионной модели, а неформально, как *характеристики* распределения $y|x$. Например, вряд ли имеет смысл сравнивать $\beta(.49)$ как существенно отличающийся от $\beta(.50)$ или точно оценивать конкретную разность, как, например, $\beta(.5) - \beta(.3)$. С другой стороны, качественная разница или, возможно, ее отсутствие между $\beta(.3)$ и $\beta(.5)$ может оказаться интересной характеристикой генеральной совокупности. Модель квантильной регрессии рассматривается в подразделе 7.3.2.

12.3.4. Ядерные методы оценки

Ядерная оценка плотности — это принципиально непараметрический инструмент, так что он более уместен в следующем разделе. Но некоторые модели, использующие ядерные методы, не являются полностью непараметрическими. Примером служит модель частично линейной регрессии из раздела 7.4. Многие модели формулируются с использованием индексной функции, т.е. специфицируются вокруг линейной функции $x'\beta$, что делает их, по крайней мере, полупараметрическими, но все же следует избегать предположений о распределении путем использования ядерного оценивания. Другим примером будет оценка бинарного выбора в работе Льюбела (Lewbel (2000)).

Пример 12.4. Полупараметрическая оценка модели бинарного выбора

Основная модель бинарного выбора, рассматриваемая в разделе 17.3, пробит-модель, — это полностью параметрическая спецификация. В предположениях модели оценка методом максимального правдоподобия эффективна (и уместна). Однако, как неоднократно показано в литературе, оценка β подвержена нарушениям предпосылки о распределении. Мы рассмотрим несколько полупараметрических и непараметрических оценок в подразделе 17.4.7. Для того чтобы продемонстрировать природу процесса моделирования, рассмотрим оценку, предложенную в работе Льюбела (Lewbel (2000)). Пробит-модель основана на нормальном распределении, для которого $\text{Prob}[y_i = 1 | \mathbf{x}_i] = \text{Prob}[\mathbf{x}_i'\beta + \varepsilon_i > 0]$, где $\varepsilon_i \sim N[0, 1]$. Оценка β в этой спецификации будет несостоятельной, если распределение ненормально или если ε_i гетероскедастичны. Льюбел предлагает следующую процедуру: если (а) можно предположить, что \mathbf{x}_i содержит «специальную» переменную v_i , знак коэффициента которой известен — предлагается метод для определения знака, и (b) плотность ε_i не зависит от этой переменной, тогда состоятельная оценка β может быть получена регрессией $[y_i - s(v_i)]/f(v_i | \mathbf{x}_i)$ на \mathbf{x}_i , где $s(v_i) = 1$, если $v_i > 0$, и 0 в противном случае, и $f(v_i | \mathbf{x}_i)$ — это ядерная оценка плотности $v_i | \mathbf{x}_i$. Оценка Льюбела устойчива к гетероскедастичности и распределению. Также предлагается метод для оценки распределения ε_i . Обратите внимание, что оценка Льюбела полупараметрическая. Его

модель рассматривает функцию параметров β , но распределение не специфицируется.

12.3.5. Сравнение параметрического и полупараметрического анализа

Часто интерес представляет сравнение параметрической и полупараметрической моделей. Как мы отметили ранее, сильные предположения полностью параметрической модели даются большой ценой; выводы, принимаемые на основе модели, устойчивы настолько, насколько устойчивы ее предпосылки. Конечно, обратной стороной этой медали является то, что, когда предпосылки выполняются, параметрические модели предлагают эффективные стратегии для анализа данных. Альтернативные полупараметрические подходы ослабляют такие предпосылки, как нормальность и гомоскедастичность. Важно отметить, что обобщения моделей, к которым полупараметрические оценки обычно устойчивы, приводят к несостоятельности более подробно параметризованных моделей. Сравнение касается не только эффективности. Как следствие сравнение оценок параметров может быть некорректным — параметрические и полупараметрические оценки часто оценивают очень разные величины.

Пример 12.5. Модель расходов на отдых

Меленберг и Ван Соест (Melenberg, van Soest (1996)) проанализировали расходы на отдых выборки из 1143 голландских семей в 1981 г. Важной особенностью данных, усложнившей анализ, был тот факт, что 37% (423) семей указали нулевые расходы. Линейная регрессия, которая игнорирует эту особенность данных, сильно недооценила бы влияние на расходы различных факторов, таких как общие расходы семьи (бюджет), размер семьи, возраст или образование. (См. раздел 19.3.) Стандартным параметрическим подходом к анализу такого рода данных является «Тобит», или цензурированная регрессионная модель:

$$\begin{aligned} y_i^* &= \mathbf{x}_i' \beta + \varepsilon_i, \varepsilon_i \sim \mathbf{N}[0, \sigma^2], \\ y_i &= \max(0, y_i^*). \end{aligned}$$

(Оценки методом максимального правдоподобия этой модели подробно рассмотрены в разделе 19.3.) Модель основана на двух сильных предположениях t -нормальности и гомоскедастичности. Оба предположения могут быть ослаблены с помощью более подробной параметризации, но авторы обнаружили, что тестовые статистики стабильно отвергали одно или оба предположения даже в расширенных спецификациях. Альтернативным подходом, устойчивым к обоим предположениям, является изложенная в работах Пауэлла (Powell (1984, 1986a, b)) оценка цензурированных наименьших абсолютных отклонений, представляющая собой более технически изощренную версию оценки LAD из подраздела 7.3.1. Неудивительно, что оценки, получаемые с помощью двух подходов, сильно различались.

Авторы получили множество оценок β . Полезным упражнением, которое они не выполнили, было бы сравнение предельных эффектов в различных моделях. Это эталон, по которому можно иногда сопоставлять различия между оценками. В тобит-модели $\partial E[y_i | \mathbf{x}_i] / \partial \mathbf{x}_i = \Phi(\mathbf{x}_i' \beta / \sigma) \beta$ (см. раздел 19.3). Неясно, как оценить эквивалент в полупараметрической модели, поскольку спецификация фиксирует только $\text{Med}[\varepsilon_i | \mathbf{x}_i] = 0$. (Авторы указывают в архиве данных *Journal of Applied Econometrics*, что эти данные находятся в ограниченном доступе. Таким образом, мы не смогли расширить анализ и получить оценки предельных эффектов.) Все это показывает существенную сложность, связанную с полупараметрическим подходом к оцениванию. В нелинейных моделях вроде этой интерес часто представляют предельные эффекты, а не коэффициенты. Но одним из побочных продуктов более «устойчивой» спецификации является то, что предельные эффекты не определены.

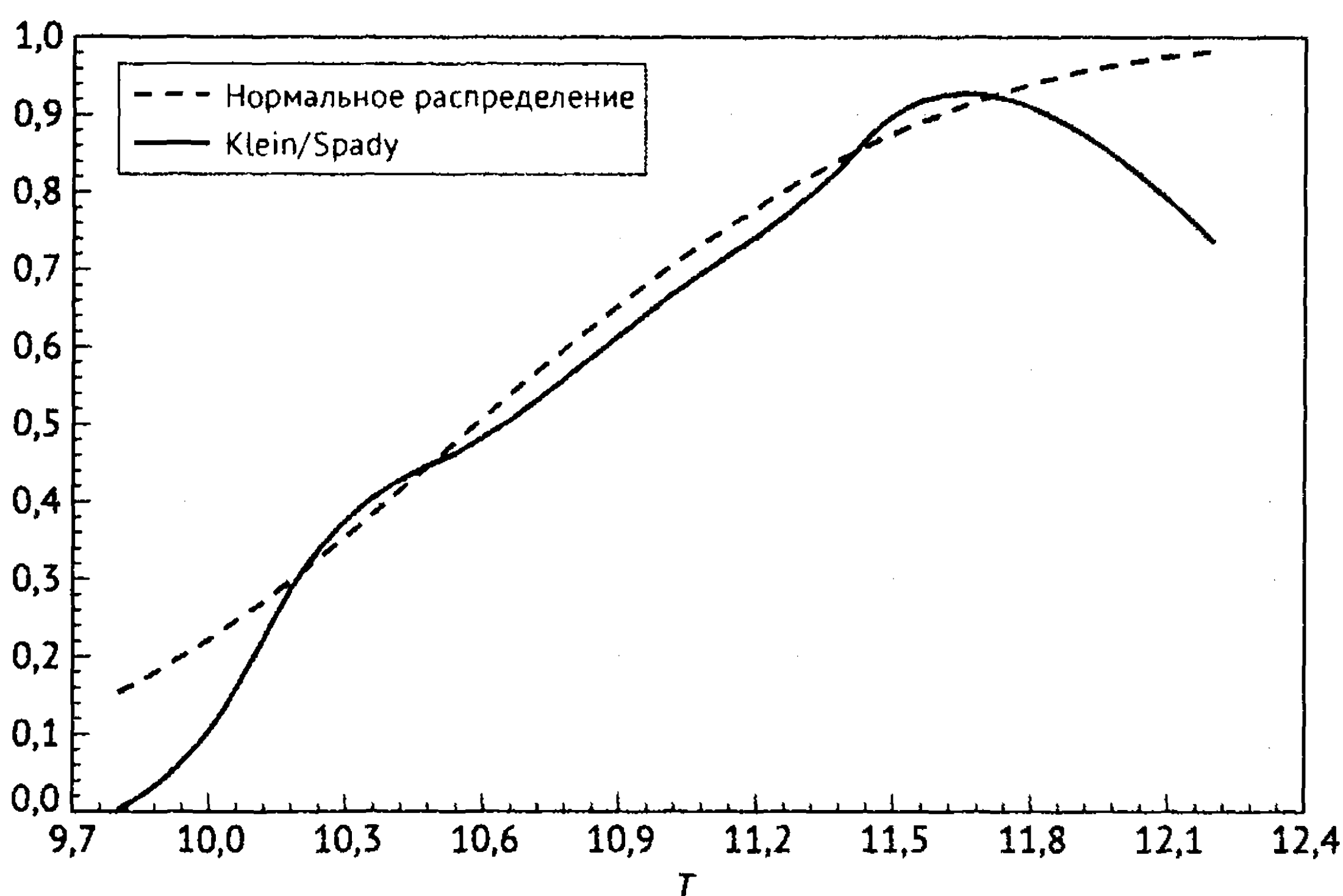


Рис. 12.1. Предсказанные вероятности положительных расходов

На втором этапе анализа авторы разделили уравнение расходов на уравнение «участия», моделирующее вероятность наступления бинарного исхода «*expenditure* = 0 или > 0», и условное уравнение расходов для имеющих положительные расходы. [В подразделе 18.4.8 мы назовем это моделью «барьера». См. Mullahy (1986).] Для этого этапа авторы в очередной раз использовали параметрическую модель, основанную на нормальном распределении (пробит-модель — см. раздел 17.3), и полупараметрическую модель, устойчивую к типу распределения и гетероскедастичности, предложенную в работе Клейна, Спэди (Klein, Spady (1993)). Как и прежде, оценки коэффициентов сильно различались. Однако в этом случае тесты спецификаций благоволили параметрической модели. На рис. 12.1, воспроизводящем их рис. 2, сравниваются вероятности, предсказанные двумя моделями. Штрихованная кривая соответствует пробит-модели. В диапазоне, охватывающем большую часть данных, модели приводят к похожим предсказаниям. Однако, как и прежде, сравнить

предельные эффекты невозможно. Интересным результатом этого этапа анализа является то, что нарушение параметрической спецификации относится в большей степени к моделированию непрерывной переменной расходов, нежели к модели, разделяющей две подвыборки на основе положительных или нулевых расходов.

12.4. Непараметрические методы оценки

Исследователи на протяжении долгого времени с сомнением относились к сильным предположениям, необходимым для параметрических моделей, оцениваемых методом максимального правдоподобия. Основным примером является линейная регрессионная модель с нормальными ошибками. Сплайны, транслогарифмические модели и полиномы представляют собой попытки обобщить функциональную форму. Тем не менее остаются вопросы относительно того, какого уровня общности можно добиться при помощи подобных аппроксимаций. Инструменты непараметрических методов оценки избавляют от необходимости делать любые предположения о функциональной форме или распределении. Учитывая их ограниченную структуру, непараметрические спецификации редко дают возможность делать очень точные выводы. Преимущество состоит в том, что получаемая информация чрезвычайно устойчива. Основным инструментом такого анализа является ядерная оценка плотности, которую мы использовали в предыдущих примерах. Мы рассмотрим некоторые примеры, а затем приложение к парной регрессии².

12.4.1. Ядерная оценка плотности

Выборочные статистики, такие как среднее, дисперсия и разброс, представляют обобщающую информацию о значениях, которые может принимать случайная величина. Но их недостаточно, чтобы показать, как распределяются значения, которые принимает случайная величина, что также может представлять интерес. Для этой цели используется плотность случайной величины. Полностью параметрический подход к оценке плотности начинается с предположения о форме распределения. Оценка плотности проводится путем оценки параметров распределения. Канонический пример — нормальное распределение со средним μ и дисперсией σ^2 , которые полностью характеризуют плотность. Из этого следует, что

$$\hat{f}(x) = f(x | \hat{\mu}, \hat{\sigma}^2) = \frac{1}{\hat{\sigma}} \frac{1}{\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{x - \hat{\mu}}{\hat{\sigma}} \right)^2 \right].$$

Предположение о конкретной функции распределения может быть нежелательным. Обычным подходом в таком случае будет представление **гисто-**

²Литература, посвященная этому разделу эконометрики, обширна и быстро растет. Важными источниками, содержащими теоретические и прикладные основы непараметрических методов, являются работы Härdle (1990), Pagan, Ullah (1999) и Li, Racine (2007).

граммы в качестве описания. Рассмотрим пример. В примерах 15.17 и в работе Грина (Greene (2004a)) мы оценивали модель, дающую условную оценку вектора коэффициентов наклона для каждой из 1270 фирм в нашей выборке. Нас может интересовать распределение этих оценок по фирмам. В частности, условная оценка коэффициента наклона для *ln sales* среди 1270 фирм имеет выборочное среднее, равное 0,3428, стандартное отклонение 0,08919, минимум 0,2361 и максимум 0,5664. Это мало что говорит нам о распределении значений, но тот факт, что среднее существенно меньше середины покрываемого интервала, равного 0,4013, может указывать на некоторую скошенность. Гистограмма, изображенная на рис. 12.2, гораздо более информативна. На основании увиденного нами предположение о нормальности не представляется уместным. Распределение выглядит бимодальным, и никакая конкретная функциональная форма не видится естественной в качестве его описания.

Гистограмма — это грубая оценка плотности. Прямоугольники на рисунке называются сегментами. По построению они имеют одинаковую ширину. (Параметрами гистограммы являются число сегментов, их ширина, а также левая начальная точка. Каждый оказывает важное влияние на форму результата.) Поскольку частоты в сегментах в сумме дают размер выборки, разделив их на n , мы получаем оценку плотности, которая удовлетворяет очевидному требованию к плотности; она суммируется (интегрируется) в единицу. Мы можем формализовать это, выписав метод, с помощью которого получают частоты. Пусть x_k — это середина k -го сегмента и пусть h — ширина сегмента (ниже мы переименуем h в ширину окна оценки плотности). Расстояния до левой и правой границ сегмента равны $h/2$. Частота в каждом сегменте — это число наблюдений в выборке, которые попадают в интервал $x_k \pm h/2$. Приводя подобные слагаемые, мы получаем нашу «оценку»:

$$\hat{f}(x) = \frac{1 \text{ частота в сегменте}_x}{n \text{ ширина сегмента}_x} = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} \mathbf{1} \left(x - \frac{h}{2} < x_i < x + \frac{h}{2} \right),$$

где $\mathbf{1}$ (*утверждение*) обозначает индикаторную функцию, которая равна 1, если утверждение верно, и 0 в противном случае, а bin_x (сегмент x) — сегмент, для которого x является серединой. Теперь мы видим, что гистограмма — это оценка, в некоторых отношениях напоминающая оценки, которые мы встречали ранее. Событие в индикаторе может быть преобразовано, чтобы получить эквивалентную форму

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} \mathbf{1} \left(-\frac{1}{2} < \frac{x_i - x}{h} < \frac{1}{2} \right).$$

Эта форма оценки просто подсчитывает число точек, находящихся в пределах половины сегмента от x_k .

Хотя и достаточно грубая, эта «наивная» (ее формальное название в литературе) оценка имеет форму **ядерных оценок плотности**, которые мы уже встречали:

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h} K \left[\frac{x_i - x}{h} \right], \quad \text{где } K[z] = \mathbf{1}[-1/2 < z < 1/2].$$

Таблица 12.1. Ядерные функции для оценки плотности

| Ядро | Формула $K[z]$ |
|------------------|--|
| Епанечникова | $0,75(1 - 0,2z^2)/2,236$, если $ z \leq 2,236$, 0 в противном случае |
| Нормальное | $\phi(z)$ (нормальная плотность) |
| Логистическое | $\Lambda(z)[1 - \Lambda(z)]$ (логистическая плотность) |
| Равномерное | 0,5, если $ z \leq 1$, 0 в противном случае |
| Бета | $0,75(1 - z)(1 + z)$, если $ z \leq 1$, 0 в противном случае |
| Косинусоидальное | $1 + \cos(2\pi z)$, если $ z \leq 0,5$, 0 в противном случае |
| Треугольное | $1 - z $, если $ z \leq 1$, 0 в противном случае |
| Парзена | $4/3 - 8z^2 + 8 z ^3$, если $ z \leq 0,5$; $8(1 - z)^3/3$, если $0,5 < z \leq 1$, 0 в противном случае. |

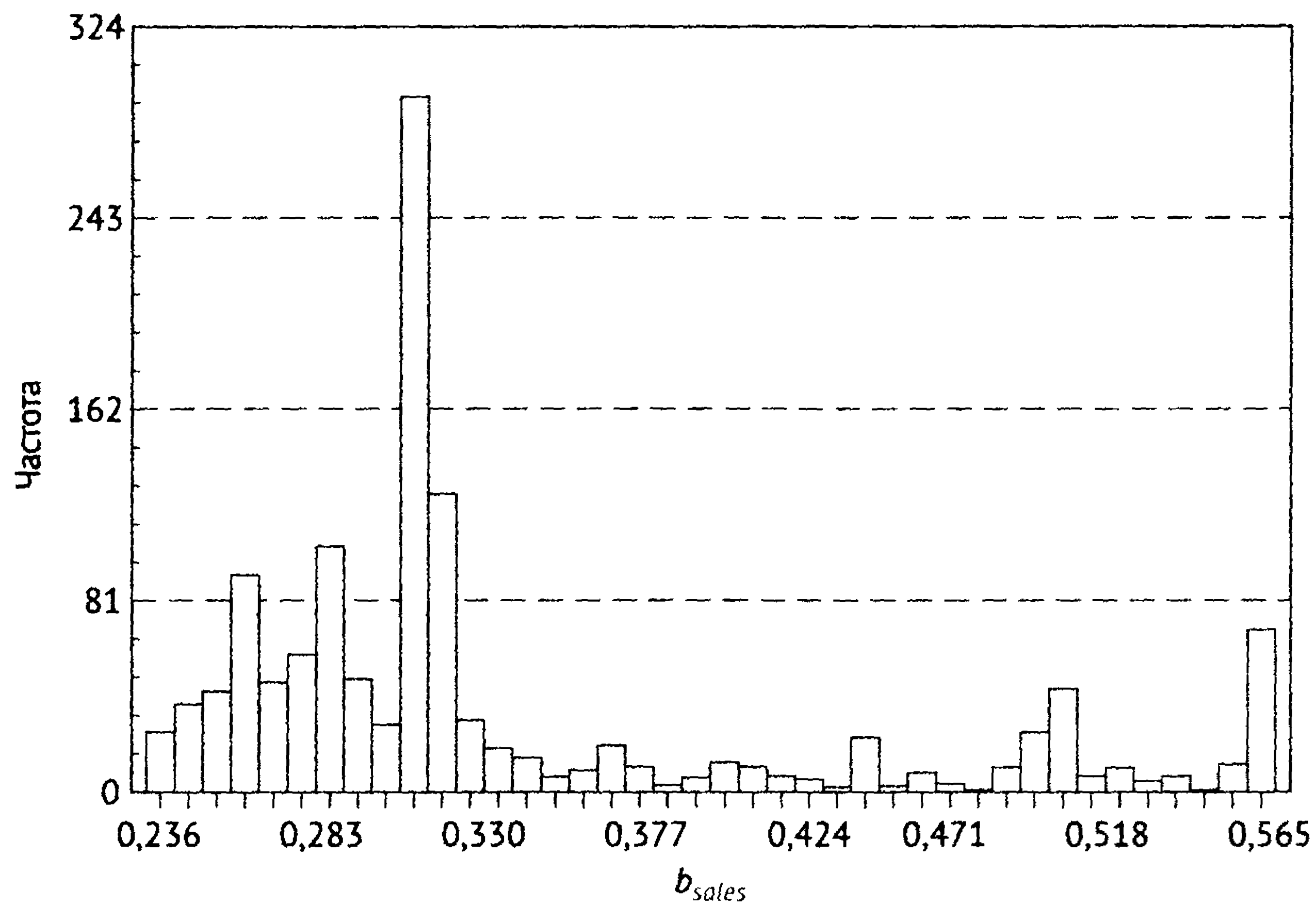
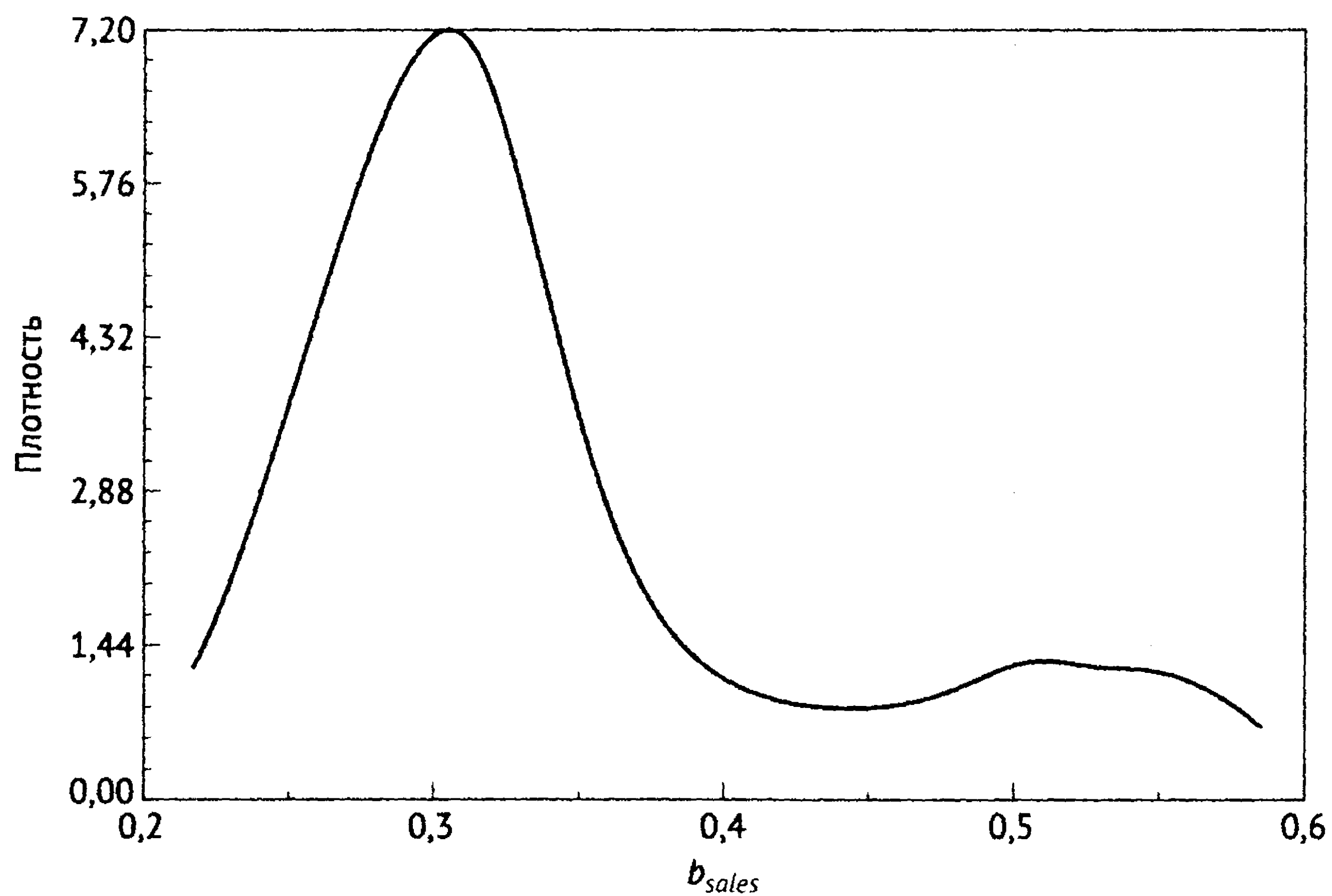
«Наивная» оценка имеет ряд недостатков. Она не является ни гладкой, ни непрерывной. Ее форма отчасти задается тем, где зафиксированы левая и правая границы гистограммы. (При построении гистограммы ширина сегмента часто выбирается как фиксированная доля разброса выборки. В таком случае концы левого и правого сегментов будут равны минимальному и максимальному выборочным значениям, и это повлияет на форму гистограммы. Если же ширина сегмента устанавливается независимо от выборочных значений, эта проблема решается.) Что более важно, форма гистограммы будет сильно зависеть от самой ширины окна. (К сожалению, эта проблема присуща и более сложным спецификациям.)

Грубость взвешивающей функции в оценке легко исправить. Предположение Розенблатта (Rosenblatt (1956)) состоит в том, чтобы заменить «наивную» оценку другой, непрерывной и интегрируемой в 1 взвешивающей функцией. Было предложено несколько вариантов, включая (длинный) список, представленный в табл. 12.1. Каждая из них — это гладкая, непрерывная, симметричная функция, и все они одинаково привлекательны. Логистическое и нормальное ядра определяются таким образом, чтобы веса сходились у 0 лишь асимптотически, тогда как остальные достигают 0 в конкретных значениях. Было отмечено, что при построении оценки плотности выбор ядерной функции редко является определяющим и обычно менее важен по сравнению с более сложной проблемой выбора ширины окна. (Логистическое и нормальное ядра представляют собой стандартный вариант в большинстве приложений.)

Функция ядерной плотности — это оценка для каждого конкретного x , $\hat{f}(x)$ — это выборочная статистика

$$\hat{f}(z) = \frac{1}{n} \sum_{i=1}^n g(x_i | z, h).$$

Поскольку $g(x_i | z, h)$ нелинейна, нам стоит ожидать смещения в конечных выборках. Хотелось бы применить наши обычные результаты для вы-

Рис. 12.2. Гистограмма оценок коэффициентов b_{sales} Рис. 12.3. Ядерная оценка плотности для коэффициентов b_{sales}

борочных моментов, но требуется более сложный анализ, поскольку ширина окна является функцией n . Пэган, Уллах (Pagan, Ullah (1999)) подробно рассмотрели свойства ядерных оценок и выяснили, что при некоторых предположениях оценка состоятельна и асимптотически нормально распределена, но смещена в конечных выборках. Смещение является функцией ширины окна, но при правильном выборе h асимптотически смещение пропадает. Как подсказывает интуиция, чем больше ширина окна, тем больше смещение, но, с другой стороны, меньше дисперсия оценки. Это может указывать на наличие оптимальной ширины окна. Однако в результате продолжительного анализа этого вопроса выводы авторов не дают прямых указаний на то, как ее найти. Одно соображение представляется полезным. Для того чтобы доля наблюдений, попавших в сегмент, сходилась к соответствующей площади под плотностью, ширина должна сокращаться медленнее, чем $1/n$. В связи с этим в приложениях обычно используется **ширина окна**, пропорциональная $n^{-1/5}$. Таким образом, нами использовалась $h = 0,9 \times s/n^{1/5}$. Чтобы завершить ранее начатую иллюстрацию, рис. 12.3 — это основанная на логистической ядерной функции оценка плотности для распределения оценок коэффициента наклона в ранее оцененной модели. Схожесть с гистограммой, изображенной на рис. 12.2, вполне ожидаема.

12.5. Свойства оценок

До сих пор мы рассматривали методы оценки. Мы сделали обзор различных инструментов, использовавшихся в прикладной литературе. Пока что мы не анализировали статистические свойства этих оценок. Хотя, как было отмечено ранее, мы оставим подробное рассмотрение асимптотической теории до более продвинутого изложения, стоит потратить некоторое время на рассмотрение основ фундаментальной теории, на которой построены эти инструменты.

12.5.1. Статистические свойства оценок

Мы рассмотрели следующие свойства:

- **Несмещенность.** Это свойство в конечных выборках, которое может быть установлено лишь в небольшом числе случаев. Строгая несмещенность редко играет центральную роль за пределами линейной регрессионной модели. Однако асимптотическая несмещенность (когда математическое ожидание оценки сходится к истинному значению параметра с ростом выборки) может представлять интерес. [См., например, работу Пэгана, Уллаха (Pagan, Ullah (1999), раздел 2.5.1, посвященный ядерной оценке плотности).] В большинстве случаев, однако, обсуждение асимптотической несмещенности направлено на состоятельность, которая является более важным свойством.
- **Состоятельность.** Это куда более важное свойство. Эконометристы редко доверяют оценке, для которой состоятельность не может быть установлена.

- **Асимптотическая нормальность.** Это свойство формирует основание для большей части статистических выводов, осуществляемых с помощью стандартных оценок. Когда асимптотическая нормальность не может быть установлена, иногда становится сложно найти способ продвинуться дальше, чем презентация численных значений оценок (с оговорками). Однако большая часть современной литературы, посвященной макроэкономике и анализу временных рядов, фокусируется на оценках, которые асимптотически не распределены нормально. Таким образом, это свойство имеет важное значение лишь в контексте, а не в качестве абсолютной ценности.
- **Асимптотическая эффективность.** Эффективность редко может быть установлена в абсолютном смысле. Однако может быть установлена эффективность в некотором классе. Таким образом, например, многое может быть сказано об относительной эффективности оценок максимального правдоподобия и GMM в классе состоятельных и асимптотически нормальных (CAN) оценок. В этом отношении есть два важных практических соображения. Во-первых, исследователь захочет узнать, использует ли он данные очевидно неоптимальным образом. (В литературе встречаются обсуждения GMM-оценок полностью специфицированных параметрических пробит-моделей — GMM-оценки в этом случае строго хуже оценок максимального правдоподобия.) Таким образом, когда возможно, предпочтительно избегать очевидно неэффективных оценок. С другой стороны, обычно исследователь не выбирает из списка доступных оценок, в его или ее распоряжении имеется одна и вопрос об относительной эффективности не стоит.

12.5.2. Экстремальные оценки

Экстремальная оценка — это оценка, получаемая путем оптимизации целевой функции $q(\theta | \text{data})$. Три метода, на которые была потрачена большая часть усилий, — это

- метод наименьших квадратов:

$$\hat{\theta}_{LS} = \text{Argmax} \left[-\left(1/n\right) \sum_{i=1}^n (y_i - h(\mathbf{x}_i, \theta_{LS}))^2 \right],$$
- метод максимального правдоподобия:

$$\hat{\theta}_{ML} = \text{Argmax} \left[\left(1/n\right) \sum_{i=1}^n \ln f(y_i | \mathbf{x}_i, \theta_{ML}) \right],$$
- GMM: $\hat{\theta}_{GMM} = \text{Argmax} \left[-\bar{\mathbf{m}}(\text{data}, \theta_{GMM})' \mathbf{W} \bar{\mathbf{m}}(\text{data}, \theta_{GMM}) \right].$

(Мы поменяли знаки первого и третьего только для удобства, чтобы все три могли быть представлены как оптимизационные задачи одного типа.) Методы наименьших квадратов и максимального правдоподобия являются примерами **M-оценок**, которые определяются путем оптимизации суммы слагаемых. Большая часть известных теоретических результатов, приведенных здесь и в других книгах, относится к поведению экстремальных оценок, но некоторые — включая байесовские оценки, некоторые полупараметрические оценки и все непараметрические оценки — не относятся к ним. Тем не менее интерес представляет установление, когда это возможно, свойств оценок во всех трех случаях. Конечным результатом для практикующего исследователя будет набор статистических свойств, которые позволят ему или

ей уверенно делать выводы о процессах, порождающих данные, которые послужили поводом для анализа.

Способы получения экстремальных оценок приведены в литературе на разных уровнях. (См., например, любой из источников, приведенных в сноске 1 этой главы.) Доступное изложение представлено в Amemiya (1985) и Davidson, MacKinnon (2004). Ньюи и Макфадден (Newey, McFadden (1994)) приводят строгий анализ, и их работа является стандартным актуальным источником. Наше обсуждение до сих пор включало только отдельные элементы анализа. Читателю следует обратиться к одному из этих источников за доказательствами и выкладками.

12.5.3. Предположения для асимптотических свойств экстремальных оценок

Некоторые общие результаты необходимы для того, чтобы установить асимптотические свойства классических (небайесовских) стандартных экстремальных оценок, приведенных выше.

1. **Пространство параметров** (см. раздел 12.2) должно быть выпуклым и вектор параметров, который является предметом оценивания, должен быть его внутренней точкой. Первое требование исключает плохо поставленные задачи оценивания, например те, в которых параметр может принимать только одно значение из дискретного множества. Таким образом, поиск даты структурного изменения как обычного параметра в модели временных рядов приводит к невыпуклости. Некоторые доказательства в этом контексте упрощаются предположением о компактности пространства параметров. (Компактное множество замкнуто и ограничено.) Однако предположение компактности обычно достаточно сильное, так что мы ограничимся более слабым требованием.
2. **Целевая функция** должна быть вогнутой по параметрам. (См. раздел А.8.2.) Это предположение подразумевает, что для каждого набора данных целевая функция имеет внутренний оптимум и что мы можем его найти. Целевые функции не должны быть «глобально вогнутыми», у них может быть несколько оптимумов. Но если они даже не «локально вогнуты», мы не можем содержательно обсуждать оптимизацию. Эта проблема может возникнуть в плохо структурированной задаче, но возможно сформулировать модель, в которой целевая функция монотонно возрастает или убывает по параметру. Такая модель приводит к невогнутой целевой функции³. Тут становится важным различие между компактностью и вогнутостью в предыдущем условии. Если целевая функция строго непрерывна на компактном пространстве параметров, то она достигает

³В своем упражнении 23.6 Griffiths, Hill, Judge (1993), основываясь (увы) на первом издании этого текста, предлагают пробит-модель для результатов голосования по штатам, которая включает фиктивные переменные для регионов Северо-Восток, Юго-Восток, Запад и Горы. В обычной ситуации стоило бы включить три из четырех фиктивных переменных в модель, но авторы аккуратно исключили две из них, поскольку помимо ловушки фиктивных переменных переменная Юго-Восток всегда равна нулю, когда зависимая переменная равна нулю. Включение этой переменной приводит к невогнутости функции правдоподобия — параметр при этой переменной расходится. Анализ аналогичной ситуации приводится в качестве предостережения на с. 272 в Amemiya (1985).

на этом множестве максимума и вогнутость не требуется. Проблема с оцениванием, однако, состоит в том, что это не исключает наличия этого максимума на (предполагаемой) границе пространства параметров. Этот случай вступает в противоречие с доказательствами состоятельности и асимптотической нормальности. Общая проблема решается предположением о том, что целевая функция вогнута в окрестности истинного вектора параметров.

3. **Идентифицируемость параметров.** Любая фраза, начинающаяся со слов «истинные параметры модели θ_0 идентифицированы, если...», связана с проблемами, поскольку если параметры «не идентифицированы», то они вообще не являются параметрами (какой бы то ни было) модели. (Например, не существует «истинного» вектора параметров в неидентифицированной модели из примера 2.5.) Плодотворный подход к этой проблеме, который позволяет избежать неопределенности, связанной с попытками дать определение *истинного* вектора параметров и затем задаться вопросом о его идентифицируемости (оцениваемости), изложен в книге Дэвидсона и Маккиннона (Davidson, MacKinnon (1993, p. 591)) и выглядит так: рассмотрим параметризованную модель M и набор процессов, порождающих данные для модели μ . При конкретной параметризации μ пусть найдется предполагаемый «истинный» вектор параметров $\theta(\mu)$. Рассмотрим любой вектор параметров θ в пространстве параметров Θ . Определим

$$q_\mu(\mu, \theta) = \text{plim}_\mu q_n(\theta | \text{data}).$$

Эта функция — это предел по вероятности целевой функции при предполагаемой параметризации μ . Если этот предел по вероятности существует (является конечной константой) и, кроме того,

$$q_\mu[\mu, \theta(\mu)] > q_\mu(\mu, \theta), \quad \text{если } \theta \neq \theta(\mu),$$

тогда, если пространство параметров компактно, вектор параметров идентифицирован целевой функцией. Мы не предполагали компактность. Для выпуклого пространства параметров нам потребуется дополнительное условие, состоящее в том, что не существует последовательностей без предельных точек θ^m таких, что $q(\mu, \theta^m)$ сходятся к $q[\mu, \theta(\mu)]$.

Изложенный здесь подход состоит в том, чтобы сначала предположить, что у модели есть *некоторый* набор параметров. Критерий идентифицируемости означает, что, если это предположение выполняется, предел по вероятности целевой функции максимизируется при этих параметрах. Этот результат зависит от сходимости целевой функции к конечному значению в любой точке внутри пространства параметров. Поскольку целевая функция — это функция данных, эта сходимость требует некоторого утверждения о свойствах данных. Оставляем этот вопрос в стороне на некоторое время, интерес для нас представляет то, что приведенные выше результаты уже достаточны для состоятельности M -оценки. Используя крайне емкую формулировку, Амэмия (1985) определил идентифицируемость просто как «существование состоятельной оценки». Мы видим, что идентификация и условия для состоятельности

М-оценки в сущности представляют собой одно и то же.

Такая форма идентификации необходима теоретически для того, чтобы получить условия для состоятельности. Однако, за исключением простейших случаев, ее будет очень сложно проверить на практике. К счастью, существуют более простые и интуитивные подходы к проверке идентификации.

- Для оценки методом наименьших квадратов достаточным условием идентификации является то, что два различных вектора параметров θ и θ_0 могут давать различные значения функции условного среднего. Вы должны проверить, что для линейной модели это предположение А.2 о полноте ранга. В случае модели из примера 2.5 мы рассматриваем регрессию, в которой $x_2 = x_3 + x_4$. В таком случае любой вектор параметров вида $(\beta_1, \beta_2 - a, \beta_3 + a, \beta_4 + a)$ приводит к тому же значению условного среднего, что и $(\beta_1, \beta_2, \beta_3, \beta_4)$, независимо от x_i , так что модель неидентифицируема. Предположение о полноте ранга необходимо, чтобы избежать этой проблемы. Для нелинейной регрессии проблема куда сложнее, и общего простого решения не существует. В примере 7.2 приведена нелинейная регрессионная модель, для которой не выполнено условие идентификации и не указано, как отсутствие идентификации исправляется.
 - Для оценки методом максимального правдоподобия требуется условие, похожее на условие, используемое для регрессионной модели. Для любых двух векторов параметров $\theta \neq \theta_0$ должны быть допустимы различные значения плотности $f(y_i | x_i, \theta)$ для некоторого вектора данных (y_i, x_i) . Многие эконометрические модели, оцениваемые с помощью метода максимального правдоподобия, являются моделями «индексной функции», включающими плотности вида $f(y_i | x_i, \theta) = f(y_i | x_i' \theta)$. В таких случаях достаточным может оказаться то же условие полноты ранга, что и в случае регрессионной модели. (Оно будет достаточным, если в модели нет других параметров.)
 - Для GMM-оценки особенного упрощения не получается. Достаточное условие идентификации — это $E[\bar{m}(\text{data}, \theta)] \neq 0$, если $\theta \neq \theta_0$.
4. **Поведение данных** обсуждалось выше в различных контекстах. Оценки основываются на средних значениях функций, зависящих от наблюдений. (Вы можете отметить это во всех трех предыдущих определениях. Производные этих целевых функций будут точно так же средними функций от наблюдений.) Анализ их асимптотического поведения будет сконцентрирован на выявлении условий, при которых для некоторых выборочных средних функций от наблюдений будут выполняться законы больших чисел, такие как теоремы Хинчина (D.5) или Чебышева (D.6), и на том, какие требуются предположения, чтобы убедиться, что произведение «корня из n » и выборочного среднего функций будет подчиняться центральному предельным теоремам, таким как теоремы Линдберга–Феллера (D.19) или Ляпунова (D.20) для межобъектных данных, или центральной предельной теореме для мартингальных разностных последовательностей для зависимых наблюдений (теорема 20.3). В конечном счете это вопрос получения статистических свойств. Свойство

сходимости, заявленное выше, должно проявляться в контексте данных. Эти условия обсуждались в подразделах 4.4.1 и 4.4.2 под заголовком «Хорошие данные». На данный момент мы предположим, что данные хорошие.

12.5.4. Асимптотические свойства оценок

Учитывая весь этот формальный аппарат, далее следуют стандартные результаты, касающиеся асимптотических свойств М-оценок.

Теорема 12.1. *Состоятельность М-оценок*

Если (а) пространство параметров выпукло и истинный параметр — это точка внутри него, (б) целевая функция вогнута, (с) параметры идентифицируются целевой функцией и (д) данные доброкачественные, то М-оценка сходится по вероятности к истинному значению вектора параметров.

Доказательства сходимости М-оценок основаны на фундаментальном результате, связанном со сходимостью, который, в свою очередь, основан на предположениях (а)–(д) из теоремы 12.1. Мы предполагаем идентификацию. Фундаментальным ходом является следующее: ввиду своей зависимости от данных $q(\theta | \text{data})$ является случайной величиной. Мы предположили в (с), что $\text{plim } q(\theta | \text{data}) = q_0(\theta)$ для любой точки в пространстве параметров. Предположение (с) означает, что максимум $q_0(\theta)$ достигается в $q_0(\theta_0)$, так что θ_0 максимизирует предел по вероятности. По определению оценка $\hat{\theta}$ максимизирует $q(\theta | \text{data})$. Следовательно, состоятельность требует, чтобы предел максимизирующего значения $\hat{\theta}$ был равен значению, максимизирующему предел θ_0 . Это следует из нашего условия идентификации. Мы будем более подробно использовать этот подход в подразделе 14.4.5а, где установим состоятельность оценки методом максимального правдоподобия.

Теорема 12.2. *Асимптотическая нормальность М-оценок*

Если

(i) $\hat{\theta}$ — это состоятельная оценка θ_0 , где θ_0 — это внутренняя точка пространства параметров,

(ii) $q(\theta | \text{data})$ вогнута и дважды непрерывно дифференцируема в точке θ в окрестности θ_0 ,

(iii) $\sqrt{n}[\partial q(\theta_0 | \text{data})/\partial \theta_0] \xrightarrow{d} N[0, \Phi]$,

(iv) для любых θ в Θ , $\lim_{n \rightarrow \infty} \text{Pr}[|(\partial^2 q(\theta | \text{data})/\partial \theta_k \partial \theta_m) - h_{km}(\theta)| > \varepsilon] = 0 \forall \varepsilon >$

0, где $h_{km}(\theta)$ — это непрерывная конечная функция θ ,

(v) матрица элементов $H(\theta)$ невырождена в θ_0 ,

то

$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N\{0, [H^{-1}(\theta_0)\Phi H^{-1}(\theta_0)]\}$.

Доказательство асимптотической нормальности основывается на теореме о среднем из математического анализа и разложении Тейлора производных максимизируемой целевой функции около истинного вектора параметров

$$\sqrt{n} \frac{\partial q(\hat{\theta} | \text{data})}{\partial \hat{\theta}} = \mathbf{0} = \sqrt{n} \frac{\partial q(\theta_0 | \text{data})}{\partial \theta_0} + \frac{\partial^2 q(\bar{\theta} | \text{data})}{\partial \bar{\theta} \partial \bar{\theta}'} \sqrt{n}(\hat{\theta} - \theta_0).$$

Вторая производная оценивается в точке $\bar{\theta}$, т.е. между $\hat{\theta}$ и θ_0 , $\bar{\theta} = w\hat{\theta} + (1-w)\theta_0$ для некоторых $0 < w < 1$. Поскольку мы предположили $\text{plim } \hat{\theta} = \theta_0$, то видим, что матрица во втором слагаемом справа должна сходиться к $\mathbf{H}(\theta_0)$. Предположения теоремы могут быть объединены, чтобы получить заявленное нормальное распределение. Формальное доказательство этого набора результатов приводится в работе Ньюи и Макфаддена (Newey, McFadden (1994)). Несколько более подробный анализ на основе этой теоремы приводится в подразделе 14.4.5b, где мы устанавливаем асимптотическую нормальность оценки максимального правдоподобия.

Предшествующее обсуждение относилось к М-оценкам, так что нам остается еще получить соответствующие результаты для важной GMM-оценки. Состоятельность следует из аргументов, похожих на использованные выше, но установить асимптотическую нормальность несколько сложнее. Мы вернемся к этому вопросу в главе 13, где опять же опишем формальные результаты и предложим читателю за формальным изложением обратиться к источнику, такому как работа Ньюи и Макфаддена (Newey, McFadden (1994)).

Применимость приведенных выше результатов одинаково очевидна не для всех задач оценки. Например, оценка методом наименьших абсолютных отклонений (НАО) не была упомянута выше, хотя и является М-оценкой и для нее актуальны приведенные результаты. Анализ усложнен тем, что целевая функция не является непрерывно дифференцируемой. Тем не менее состоятельность и асимптотическая нормальность были установлены. [См. Koenker и Bassett (1982) и Amemiya (1985, p. 152–154).] Некоторые из полупараметрических оценок и все упомянутые непараметрические оценки требуют несколько большего внимания. Например, Pagan, Ullah (разделы 2.5 и 2.6) получают нужные свойства ядерной оценки плотности $f(x^*)$, но это требует более тщательного анализа функции и данных, чем необходимо, скажем, для линейной регрессионной или бинарной логит-модели. Заинтересованный читатель может найти развернутый анализ асимптотических свойств оценок, например, в Amemiya (1985), Newey, McFadden (1994), Davidson, MacKinnon (2004) и Hayashi (2000). На практике нечасто удается проверить выполнение условий для имеющейся задачи оценивания и обычно они просто предполагаются. Однако поиск нарушения условий иногда более прост и потраченные на него усилия вознаграждаются. Например, отсутствие параметрической идентификации часто может быть выявлено при анализе самой модели.

12.5.5. Тестирование гипотез

Выше мы привели набор результатов, которые (в большей или меньшей степени) унифицируют теоретическую основу трех важных классов эконо-

метрических оценок: методов наименьших квадратов, максимального правдоподобия и GMM. Подобная теоретическая основа была разработана для знакомых тестовых статистик Вальда, отношения правдоподобия (LR) и множителей Лагранжа (LM). [См. Newey, McFadden (1994).] Их практическое изложение приведено в других разделах этой книги, так что в интересах краткости мы рекомендуем заинтересованному читателю обратиться за техническими деталями к указанным источникам.

12.6. Заключение и выводы

В этой главе сделан краткий обзор методов оценки в эконометрике. Существуют разные подходы к тому, как написать такой обзор. Актуальная литература может быть сгруппирована по трем важным типам оценок — параметрическим, полупараметрическим и непараметрическим. Было отмечено, что литература развивается от первого из них в направлении третьего, но при более внимательном рассмотрении мы обнаруживаем, что это не так. Метод максимального правдоподобия все еще остается основным в разных контекстах. Новые приложения были найдены для GMM-оценки, но в то же время быстро развиваются полностью параметрические байесовские и симуляционные методы оценки. Естественно, набор инструментов, применимых в любом контексте, стабильно растет.

Основные термины и понятия

| | |
|---|--|
| Bandwidth — Ширина окна | Kernel density estimator — Ядерная оценка плотности |
| Bayesian estimation — Байесовское оценивание | Least absolute deviations (LAD) — Метод наименьших абсолютных отклонений (LAD) |
| Bootstrap — Бутстрэп | Likelihood function — Функция правдоподобия |
| Conditional density — Условная плотность | M estimator — M-оценка |
| Copula function — Копула-функция | Maximum empirical likelihood estimator — Оценка максимального эмпирического правдоподобия |
| Criterion function — Целевая функция | Maximum entropy — Оценка максимума энтропии |
| Data generating mechanism — Процесс, порождающий данные | Maximum likelihood estimator — Оценка максимального правдоподобия |
| Density — Плотность | Method of moments — Метод моментов |
| Empirical likelihood function — Функция эмпирического правдоподобия | Nearest neighbor — Ближайший сосед |
| Entropy — Энтропия | Nonparametric estimators — Непараметрическая оценка |
| Estimation criterion — Критерий оценки | Parameter space — Пространство параметров |
| Extremum estimator — Экстремальная оценка | Parametric estimation — Параметрические методы оценки |
| Fundamental probability transform — Фундаментальное преобразование вероятности | |
| Generalized method of moments — Обобщенный метод моментов | |
| Histogram — Гистограмма | |
| Identifiability — Идентифицируемость | |

| | |
|--|--|
| Partially linear model — Частично линейная модель | ционные методы оценки |
| Quantile regression — Квантильная регрессия | Sklar's theorem — Теорема Шкляра |
| Semiparametric estimation — Полупараметрические методы оценки | Smoothing function — Сглаживающая функция |
| Simulation-based estimation — Симуля- | Stochastic frontier model — Модель стохастической границы |

Упражнение

1. Сравните полностью параметрический и полупараметрический подходы к оценке модели дискретного выбора, например множественной логит-модели, описанной в главе 17. В чем состоят плюсы и минусы полупараметрического подхода?

ГЛАВА 13

Оценка методом наименьшего расстояния и обобщенным методом моментов

13.1. Введение

Оценка методом максимального правдоподобия, рассматриваемая в главе 14, эффективна среди состоятельных и асимптотически нормальных оценок в контексте специфицированной параметрической модели. Возможным недостатком этого результата является то, что для достижения этой эффективности необходимы, возможно, сильные, ограничивающие предпосылки о распределении или о процессе, порождающем данные. Оценки обобщенным методом моментов (ОММ, GMM), обсуждаемые в этой главе, отходят от параметрических предположений в случае оценок, устойчивых к некоторой вариации в процессе, порождающем данные.

В этой главе представлены некоторые достаточно общие результаты, касающиеся параметрических методов оценки. Мы начнем с, пожалуй, старейшей из формально изложенных теорий оценивания — с классического метода моментов. Развитие метода моментов началось в 1925 г. с пионерной работы Фишера (Fisher (1925)). Использование выборочных моментов в качестве основы для оцениваемых уравнений является центральной идеей в эконометрике. GMM — это развитие этого подхода, которое, как мы скоро увидим, включает практически все оценки, представленные в данной книге. В разделе 13.2 вводятся основные определения и свойства оценки методом моментов. Метод наименьшего расстояния изложен в разделе 13.3. Формальная сторона оценки GMM представлена в разделе 13.4. В разделе 13.5 обсуждается тестирование гипотез на основе моментных уравнений. Важные приложения, включая динамические модели панельных данных, описаны в разделе 13.6.

Пример 13.1. Уравнения Эйлера и жизненный цикл потребления

Одним из наиболее цитируемых приложений принципа GMM для оценки эконометрических моделей является модель постоянного дохода, представленная в работе Холла (Hall (1978)). В базовую модель (с некоторыми изменениями в обозначениях) заложена гипотеза об оптимизирующем поведении потребителя на протяжении жизненного цикла. Предполагается, что потребители ведут себя согласно следующей модели:

Потребитель максимизирует

$$E_t \left[\sum_{\tau=0}^{T-t} \left(\frac{1}{1+\delta} \right)^\tau U(c_{t+\tau}) \mid \Omega_t \right] \text{ при условии } \sum_{\tau=0}^{T-t} \left(\frac{1}{1+r} \right)^\tau (c_{t+\tau} - w_{t+\tau}) = A_t.$$

Информация, доступная в момент t , обозначается Ω_t , так что E_t обозначает ожидание, сформированное в момент t на основе информационного множества Ω_t . Максимизируемое выражение — это дисконтированный ожидаемый поток будущей полезности с момента t до конца жизни в момент T . Субъективная норма межвременного предпочтения индивида — это $\beta = 1/(1 + \delta)$. Реальная ставка процента $r \geq \delta$ предполагается постоянной. Функция полезности $U(c_t)$ предполагается строго вогнутой и сепарабельной по времени (как указано в модели). Потребление в одном периоде — это c_t . Межвременное бюджетное ограничение означает, что текущее дисконтированное превышение c_t над заработками w_t за всю жизнь равно общим активам A_t , не включающим человеческий капитал. В этой модели утверждается, что единственным источником неопределенности является w_t . Относительно стохастических свойств w_t предположений не делается, за исключением того, что существуют ожидаемые будущие заработки $E_t[w_{t+\tau} | \Omega_t]$. Последовательные значения не предполагаются независимыми, и процесс для w_t не предполагается стационарным. Важная «теорема» Холла состоит в решении оптимизационной задачи, которая записывается как

$$E_t[U'(c_{t+1}) | \Omega_t] = \frac{1 + \delta}{1 + r} U'(c_t).$$

Для наших целей важным выводом статьи является «Следствие 1», которое гласит, что «никакая информация, доступная в момент t , кроме уровня потребления c_t , не помогает предсказать будущее потребление c_{t+1} через воздействие на ожидаемое значение предельной полезности. В частности, доход или богатство в периодах t или ранее становятся несущественными, как только реализуется c_t ». Мы можем использовать это как основание для модели, которая может быть сформулирована в контексте GMM. Для этого необходимо предположить форму функции полезности. Обычная (удобная) форма функции полезности — это $U(c_t) = c_t^{1-\alpha}/(1-\alpha)$, монотонная $U' = c_t^{-\alpha} > 0$ и вогнутая функция $U''/U' = -\alpha/c_t < 0$. Подставляя все это в решение, переставляя слагаемые и переобозначая их для удобства, получаем

$$E_t \left[(1 + r) \left(\frac{1}{1 + \delta} \right) \left(\frac{c_{t+1}}{c_t} \right)^{-\alpha} - 1 | \Omega_t \right] = E_t [\beta(1 + r)R_{t+1}^\lambda - 1 | \Omega_t] = 0,$$

где $R_{t+1} = c_{t+1}/c_t$ и $\lambda = -\alpha$.

Холл предположил, что r постоянна во времени. Другие приложения этого подхода к моделированию [например, Hansen, Singleton (1982)] модифицировали модель, включив в нее прогноз процентной ставки r_{t+1} . Дальнейшие шаги зависят от того, что находится в информационном множестве. Безусловное среднее не идентифицирует два параметра. Следствие гласит, что вся существенная информация в информационном множестве — это c_t . Учитывая форму модели, более естественным инструментом может быть R_t . Это предположение точно идентифицирует два параметра в модели:

$$E_t \left[(\beta(1 + r_{t+1})R_{t+1}^\lambda - 1) \begin{pmatrix} 1 \\ R_t \end{pmatrix} \right] = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

В представленном виде из модели не следуют тестируемые выводы. Эти два моментных уравнения точно идентифицировали бы два неизвестных параметра. Холл предложил гипотезы в виде нескольких моделей, включающих доход и потребление, которые были бы переопределены и установили бы, таким образом, ограничения на модель.

13.2. Состоятельная оценка: метод моментов

Выборочные статистики, такие как среднее и дисперсия, могут служить простыми описательными показателями. В нашем обсуждении методов оценки в приложении С, однако, утверждается, что в общем случае выборочные статистики имеют теоретические аналоги, например соотношение между выборочным средним и математическим ожиданием в генеральной совокупности. Естественным (возможно, очевидным) следующим шагом в анализе может стать использование этой аналогии для обоснования использования выборочных «моментов» в качестве оценок соответствующих им параметров генеральной совокупности. Остается только понять, является ли этот подход лучшим или даже просто хорошим способом использования выборочных данных для определения характеристик генеральной совокупности.

Метод моментов основан на следующем предположении: в результате случайного эксперимента при достаточно безобидных предположениях выборочная статистика сойдется по вероятности к некоторой константе. Например, в случае независимых, одинаково распределенных случайных величин $\bar{m}'_2 = (1/n) \sum_{i=1}^n y_i^2$ сойдется в среднеквадратичном к сумме дисперсии и квадрата среднего случайной величины y_i . Эта константа, в свою очередь, будет функцией неизвестных параметров распределения. Для оценки K параметров $\theta_1, \dots, \theta_K$ мы можем оценить K таких статистик, $\bar{m}_1, \dots, \bar{m}_K$, **пределы по вероятности** которых являются известными функциями параметров. Эти моменты приравниваются к K -функциям, а функции обращаются, чтобы выразить параметры как функции моментов. Моменты будут состоятельно оцениваться благодаря закону больших чисел (теоремы D.4–D.9). Асимптотически они будут нормально распределены вследствие **центральной предельной теоремы** Линдберга–Леви (D.18). Полученные таким образом оценки унаследуют состоятельность благодаря теореме Слуцкого (D.12), а асимптотическую нормальность — благодаря дельта-методу (теорема D.21).

В этом разделе перечисленные теоремы будут приведены подробно отчасти как представляющие ценность сами по себе, а отчасти в качестве предисловия к оценке обобщенным методом моментов, или GMM, обсуждаемой в разделе 13.4.

13.2.1. Случайные выборки и оценка параметров распределений

Рассмотрим случайную выборку из распределения $f(y | \theta_1, \dots, \theta_K)$ с конечными моментами до $E[y^{2K}]$. Случайная выборка состоит из n наблюдений y_1, \dots, y_n . k -й «чистый», или **нецентрированный, момент** — это

$$\bar{m}'_k = \frac{1}{n} \sum_{i=1}^n y_i^k.$$

По теореме D.4

$$E[\bar{m}'_k] = \mu'_k = E[y_i^k]$$

и

$$\text{Var}[\bar{m}'_k] = \frac{1}{n} \text{Var}[y_i^k] = \frac{1}{n} (\mu'_{2k} - \mu_k'^2).$$

Традиционно $\mu'_1 = E[y_i] = \mu$. По теореме Хинчина D.5

$$\text{plim } \bar{m}'_k = \mu'_k = E[y_i^k].$$

Наконец, по центральной предельной теореме Линдберга–Леви

$$\sqrt{n}(\bar{m}'_k - \mu'_k) \xrightarrow{d} N[0, \mu'_{2k} - \mu_k'^2].$$

В общем случае μ'_k будет функцией исходных параметров. Приравнивая K исходных моментов к этим функциям, мы получаем K уравнений, которые могут быть разрешимы для получения оценок K неизвестных параметров.

Пример 13.2. Оценка методом моментов для $N[\mu, \sigma^2]$

Для случайной выборки из $N[\mu, \sigma^2]$

$$\text{plim } \frac{1}{n} \sum_{i=1}^n y_i = \text{plim } \bar{m}'_1 = E[y_i] = \mu$$

и

$$\text{plim } \frac{1}{n} \sum_{i=1}^n y_i^2 = \text{plim } \bar{m}'_2 = \text{Var}[y_i] + \mu^2 = \sigma^2 + \mu^2.$$

Приравнивая правую и левую части пределов по вероятности, получаем оценки методом моментов:

$$\hat{\mu} = \bar{m}'_1 = \bar{y}$$

и

$$\hat{\sigma}^2 = \bar{m}'_2 - \bar{m}'_1^2 = \left(\frac{1}{n} \sum_{i=1}^n y_i^2 \right) - \left(\frac{1}{n} \sum_{i=1}^n y_i \right)^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2.$$

Следует обратить внимание на то, что $\hat{\sigma}^2$ смещена, хотя обе оценки являются состоятельными.

Несмотря на то что моменты, основанные на степенях y , представляют собой естественный источник информации о параметрах, другие функции данных также могут оказаться полезными. Пусть $m_k(\cdot)$ является непрерывной и дифференцируемой функцией, не зависящей от размера выборки n , и пусть

$$\bar{m}_k = \frac{1}{n} \sum_{i=1}^n m_k(y_i), \quad k = 1, 2, \dots, K.$$

Это тоже «моменты» данных. Из теоремы D.4 и следствия из нее (D-5) следует, что

$$\text{plim } \bar{m}_k = E[m_k(y_i)] = \mu_k(\theta_1, \dots, \theta_K).$$

Мы предполагаем что $\mu_k(\cdot)$ включает некоторые или все параметры распределения. При оценке K параметров K **моментных уравнений**

$$\bar{m}_1 - \mu_1(\theta_1, \dots, \theta_K) = 0,$$

$$\bar{m}_2 - \mu_2(\theta_1, \dots, \theta_K) = 0,$$

...

$$\bar{m}_K - \mu_K(\theta_1, \dots, \theta_K) = 0$$

дают K уравнений с K неизвестными, $\theta_1, \dots, \theta_K$. Если уравнения непрерывны и функционально независимы, **оценки методом моментов** могут быть получены при решении системы уравнений для

$$\hat{\theta}_k = \hat{\theta}_k[\bar{m}_1, \dots, \bar{m}_K].$$

Как было отмечено, наборов моментов, которые могут быть использованы для оценки параметров, может быть больше одного, и доступных моментных уравнений может оказаться больше, чем необходимо.

Пример 13.3. Обратное гауссовское распределение (распределение Вальда)

Обратное гауссовское распределение используется для моделирования периодов дожития или промежутков времени с начального момента до того момента, когда происходит некоторый переход. Стандартная форма плотности для этой случайной величины имеет вид

$$f(y) = \sqrt{\frac{\lambda}{2\pi y^3}} \exp\left[-\frac{\lambda(y - \mu)^2}{2\mu^2 y}\right], \quad y > 0, \lambda > 0, \mu > 0.$$

Среднее — это μ , а дисперсия — это μ^3/λ . Эффективная оценка максимального правдоподобия двух параметров основана на $(1/n) \sum_{i=1}^n y_i$ и $(1/n) \sum_{i=1}^n (1/y_i)$. Поскольку среднее и дисперсия — это простые функции параметров, мы можем также использовать выборочное среднее и выборочную дисперсию в качестве оценок этих функций методом моментов. Таким образом, альтернативная пара оценок параметров распределения Вальда методом моментов может быть основана на $(1/n) \sum_{i=1}^n y_i$ и $(1/n) \sum_{i=1}^n y_i^2$. Точные формулы для этих двух пар оценок оставляются в качестве упражнения.

Пример 13.4. Смесь нормальных распределений

Квандт и Рамсей (Quandt, Ramsey (1978)) изучали задачу оценки параметров смеси нормальных распределений. Предположим, каждое наблюдение в случайной выборке порождено одним из двух различных нормальных распределений. Вероятность, что наблюдение является реализацией из первого распределения $N[\mu_1, \sigma_1^2]$, равна λ , из второго — $(1 - \lambda)$. Функция плотности для наблюдаемого y равна

$$\begin{aligned} f(y) &= \lambda N[\mu_1, \sigma_1^2] + (1 - \lambda)N[\mu_2, \sigma_2^2], \quad 0 \leq \lambda \leq 1 \\ &= \frac{\lambda}{(2\pi\sigma_1^2)^{1/2}} e^{-1/2[(y-\mu_1)/\sigma_1]^2} + \frac{1-\lambda}{(2\pi\sigma_2^2)^{1/2}} e^{-1/2[(y-\mu_2)/\sigma_2]^2}. \end{aligned}$$

Прежде чем продолжить, отметим, что эта плотность в точности совпадает с моделью конечной смеси, приведенной в подразделе 14.10.1. Оценка максимального правдоподобия модели с помощью описанного там метода несколько проще метода производящих функций моментов, приведенного здесь.

Выборочное среднее и второй–пятый центральные моменты

$$\bar{m}_k = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^k, \quad k = 2, 3, 4, 5$$

дают пять уравнений для пяти неизвестных, которые могут быть преобразованы (с помощью многочлена 9-го порядка) в состоятельные оценки пяти параметров. Поскольку $2\bar{y}$ сходится по вероятности к $E[y_i] = \mu$, приведенные выше теоремы для \bar{m}'_k как оценки μ'_k так же применимы к \bar{m}_k , как оценки

$$\mu_k = E[(y_i - \mu)^k].$$

Для смеси нормальных распределений среднее и дисперсия имеют вид

$$\mu = E[y_i] = \lambda\mu_1 + (1 - \lambda)\mu_2$$

и

$$\sigma^2 = \text{Var}[y_i] = \lambda\sigma_1^2 + (1 - \lambda)\sigma_2^2 + 2\lambda(1 - \lambda)(\mu_1 - \mu_2)^2,$$

что говорит о том, насколько сложным может стать знакомый метод моментов. Предложенный авторами альтернативный метод оценки основан на

$$E[e^{ty_i}] = \lambda e^{t\mu_1 + t^2\sigma_1^2/2} + (1 - \lambda)e^{t\mu_2 + t^2\sigma_2^2/2} = \Lambda_t,$$

где t — это любое, необязательно целое число. Квандт и Рамсей (Quandt, Ramsey (1978)) предлагают выбрать пять значений t не слишком похожих и использовать статистики

$$\bar{M}_t = \frac{1}{n} \sum_{i=1}^n e^{ty_i}$$

для оценки параметров. Моментные уравнения примут вид $\bar{M}_t - \Lambda_t(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \lambda) = 0$. Авторы назвали эту процедуру методом производящих функций моментов. (См. определение производящих функций моментов в разделе В.6.)

В большинстве случаев оценки методом моментов неэффективны. Исключением является оценка на основе случайных выборок из **экспоненциальных семейств** распределений.

Определение 13.1. Экспоненциальное семейство

Экспоненциальным (параметрическим) семейством распределений называется семейство, логарифм правдоподобия которого имеет форму

$$\ln L(\theta | \text{data}) = a(\text{data}) + b(\theta) + \sum_{k=1}^K c_k(\text{data}) s_k(\theta),$$

где $a(\cdot)$, $b(\cdot)$, $c_k(\cdot)$ и $s_k(\cdot)$ — это функции. Члены «семейства» определяются различными значениями параметров.

Если логарифм правдоподобия принимает такую форму, функции $c_k(\cdot)$ называются **достаточными статистиками**¹. Когда достаточные статистики существуют, оценки методом моментов могут являться их функциями. В этом случае оценки методом моментов будут также являться оценками методом максимального правдоподобия и будут, конечно, эффективными, по крайней мере асимптотически. Следует подчеркнуть в таком случае, что распределение вероятности полностью специфицировано. Поскольку нормальное распределение относится к экспоненциальному семейству с достаточными статистиками \bar{m}'_1 и \bar{m}'_2 , оценки, приведенные в примере 13.2, полностью эффективны. (Они являются оценками максимального правдоподобия.) Смеси нормальных распределений не относятся к экспоненциальному семейству. Можно показать в качестве упражнения, что распределение Вальда из примера 13.3 относится к этому семейству. Вы должны показать, что статистики, предложенные в примере 13.3 в качестве основы для оценок методом максимального правдоподобия параметров μ и λ , являются достаточными.

Пример 13.5. Гамма-распределение

Гамма-распределение (см. раздел В.4.5) имеет вид

$$f(y) = \frac{\lambda^P}{\Gamma(P)} e^{-\lambda y} y^{P-1}, \quad y \geq 0, P > 0, \lambda > 0.$$

Функция логарифма правдоподобия для этого распределения принимает вид

$$\frac{1}{n} \ln L = [P \ln \lambda - \ln \Gamma(P)] - \lambda \frac{1}{n} \sum_{i=1}^n y_i + (P - 1) \frac{1}{n} \sum_{i=1}^n \ln y_i.$$

¹Stuart, Ord (1989, p. 1-29) обсуждают достаточные статистики и экспоненциальные семейства распределений. Результат, который мы будем использовать в главе 17, состоит в том, что если статистики $c_k(\text{data})$ являются достаточными, то условная плотность $f\{y_1, \dots, y_n | c_k(\text{data}), k = 1, \dots, K\}$ не зависит от параметров.

Эта функция принадлежит к экспоненциальному семейству с $a(\text{data}) = 0$, $b(\theta) = n[P \ln \lambda - \ln \Gamma(P)]$ и двумя достаточными статистиками $\frac{1}{n} \sum_{i=1}^n y_i$ и $\frac{1}{n} \sum_{i=1}^n \ln y_i$. Оценки методом моментов основываются на $\frac{1}{n} \sum_{i=1}^n y_i$ и $\frac{1}{n} \sum_{i=1}^n \ln y_i$ и будут оценками максимального правдоподобия. Но мы также имеем

$$\text{plim} \frac{1}{n} \sum_{i=1}^n \begin{bmatrix} y_i \\ y_i^2 \\ \ln y_i \\ 1/y_i \end{bmatrix} = \begin{bmatrix} P/\lambda \\ P(P+1)/\lambda^2 \\ \Psi(P) - \ln \lambda \\ \lambda/(P-1) \end{bmatrix}.$$

(Функции $\Gamma(P)$ и $\Psi(P) = d \ln \Gamma(P)/dP$ обсуждаются в разделе E.2.3.) Любая из этих двух оценок может быть использована для оценки λ и P .

Для данных о доходе из примера C.1 четыре приведенных ранее момента равны

$$\begin{aligned} (\bar{m}'_1, \bar{m}'_2, \bar{m}'_*, \bar{m}'_{-1}) &= \frac{1}{n} \sum_{i=1}^n \left[y_i, y_i^2, \ln y_i, \frac{1}{y_i} \right] = \\ &= [31, 278; 1453, 96; 3, 22139; 0, 050014]. \end{aligned}$$

Оценки методом моментов $\theta = (P, \lambda)$, основанные на шести возможных парах этих моментов, выглядят следующим образом:

$$(\hat{P}, \hat{\lambda}) = \begin{bmatrix} & \bar{m}'_1 & \bar{m}'_2 & \bar{m}'_{-1} \\ \bar{m}'_2 & 2,05682; 0,065759 & & \\ \bar{m}'_{-1} & 2,77198; 0,0886239 & 2,60905; 0,080475 & \\ \bar{m}'_* & 2,4106; 0,0770702 & 2,26450; 0,071304 & 3,03580; 0,1018202 \end{bmatrix}.$$

Оценки методом максимального правдоподобия — это $\hat{\theta}(\bar{m}'_1, \bar{m}'_*) = (2,4106; 0,0770702)$.

13.2.2. Асимптотические свойства оценки метода моментов

В некоторых случаях мы можем получить точное распределение оценки методом моментов. Например, для выборки из нормального распределения $\hat{\mu}$ имеет среднее μ и дисперсию σ^2/n и распределена нормально, тогда как $\hat{\sigma}^2$ имеет среднее $[(n-1)/n]\sigma^2$ и дисперсию $[(n-1)/n]^2 2\sigma^4/(n-1)$ и точное распределение, пропорциональное случайной величине хи-квадрат с $(n-1)$ степенями свободы. Если выборка не из нормального распределения, точная дисперсия выборочного среднего все еще будет равняться $\text{Var}[y]/n$, тогда как асимптотическая дисперсия оценки момента популяционной дисперсии может быть основана на первом слагаемом в (D-27), в примере D.10, но точное распределение может быть слишком сложным.

В некоторых случаях явное выражение для дисперсии выборочного момента недоступно. Так, в примере 13.4 выборочная статистика равна

$$\bar{M}_t = \frac{1}{n} \sum_{i=1}^n e^{ty_i} = \frac{1}{n} \sum_{i=1}^n M_{it}.$$

Точная дисперсия \bar{M}_t известна, только если t — это целое число. Но если выборка случайная и если \bar{M}_t — это выборочное среднее, то мы можем оценить его дисперсию как произведение $1/n$ и выборочной дисперсии наблюдений M_{it} . Мы можем также построить оценку ковариации \bar{M}_t и \bar{M}_s :

$$\text{Est.Asy.Cov}[\bar{M}_t, \bar{M}_s] = \frac{1}{n} \left\{ \frac{1}{n} \sum_{i=1}^n [(e^{ty_i} - \bar{M}_t)(e^{sy_i} - \bar{M}_s)] \right\}.$$

В общем когда моменты вычисляются как

$$\bar{m}_{n,k} = \frac{1}{n} \sum_{i=1}^n m_k(\mathbf{y}_i), \quad k = 1, \dots, K,$$

где \mathbf{y}_i — это реализация вектора переменных, подходящая оценка асимптотической ковариационной матрицы $\bar{\mathbf{m}}_n = [\bar{m}_{n,1}, \dots, \bar{m}_{n,k}]$ может быть получена, используя

$$\frac{1}{n} \mathbf{F}_{jk} = \frac{1}{n} \left\{ \frac{1}{n} \sum_{i=1}^n [(m_j(\mathbf{y}_i) - \bar{m}_j)(m_k(\mathbf{y}_i) - \bar{m}_k)] \right\}, \quad j, k = 1, \dots, K.$$

(Можно поделить внутреннюю сумму на $n - 1$, а не на n . Асимптотически это приведет к одному результату.) Эта оценка дает асимптотическую ковариационную матрицу моментов, использованных при подсчете оцененных параметров. Предполагая случайную выборку независимых, одинаково распределенных случайных величин из распределения с конечными моментами, \mathbf{F} сойдется по вероятности к соответствующей ковариационной матрице нормализованного вектора моментов $\Phi = \text{Asy.Var}[\sqrt{n} \bar{\mathbf{m}}_n(\theta)]$. Наконец, при нашем предположении случайности выборки, хотя точное распределение, скорее всего, будет неизвестно, мы можем сослаться на центральную предельную теорему Линдберга–Леви (D.18) для получения асимптотического приближения.

Чтобы формализовать оставшуюся часть выкладок, обратитесь снова к моментным уравнениям, которые мы теперь запишем как

$$\bar{m}_{n,k}(\theta_1, \theta_2, \dots, \theta_K) = 0, \quad k = 1, \dots, K.$$

Индекс n указывает на зависимость от выборки из n наблюдений. Мы также объединили выборочную статистику (сумму) и функцию параметров $\mu(\theta_1, \dots, \theta_K)$ в этой общей форме моментного уравнения. Пусть $\bar{\mathbf{G}}_n(\theta)$ является матрицей $K \times K$, чья k -я строка — это вектор частных производных

$$\bar{\mathbf{G}}'_{n,k} = \frac{\partial \bar{m}_{n,k}}{\partial \theta'}.$$

Теперь разложим набор разрешенных моментных уравнений около истинных значений параметров θ_0 в линейный ряд Тейлора. Линейное приближение — это

$$\mathbf{0} \approx [\bar{\mathbf{m}}_n(\theta_0)] + \bar{\mathbf{G}}'_n(\theta_0)(\hat{\theta} - \theta_0).$$

Таким образом,

$$\sqrt{n}(\hat{\theta} - \theta_0) \approx -[\bar{\mathbf{G}}_n(\theta_0)]^{-1} \sqrt{n}[\bar{\mathbf{m}}_n(\theta_0)]. \quad (13-1)$$

(Мы считаем это выражение приближением, поскольку не приводим формально член более высокого порядка в ряде Тейлора. Мы обозначим это явным образом в ходе обсуждения GMM-оценки в разделе 13.4.) Аргумент, необходимый для описания асимптотического поведения оценки $\hat{\theta}$, обсуждается в приложении D. Из теоремы D.18 (центральной предельной теоремы) следует, что $\sqrt{n} \bar{m}_n(\theta_0)$ стремится к нормальному распределению с вектором средних $\mathbf{0}$ и ковариационной матрицей, равной Φ . Предполагая, что функции в моментном уравнении непрерывны и функционально независимы, мы можем ожидать, что $\bar{G}_n(\theta_0)$ сойдется к невырожденной матрице констант $\Gamma(\theta_0)$. При достаточно общих условиях предельное распределение правой части (13-1) будет линейной функцией нормально распределенного вектора. Переходя к выводам, мы можем ожидать, что асимптотическое распределение $\hat{\theta}$ будет нормальным с вектором средних θ_0 и ковариационной матрицей $(1/n) \times \{-[\Gamma(\theta_0)]^{-1}\} \Phi \{-[\Gamma'(\theta_0)]^{-1}\}$. Таким образом, асимптотическая ковариационная матрица для оценки методом моментов может быть оценена при помощи

$$\text{Est. Asy. Var} [\hat{\theta}] = \frac{1}{n} [\bar{G}'_n(\hat{\theta}) \mathbf{F}^{-1} \bar{G}_n(\hat{\theta})]^{-1}.$$

Пример 13.5. (Продолжение)

Используя оценки $\hat{\theta}(m'_1, m'_*) = (2, 4106; 0, 0770702)$,

$$\hat{G} = \begin{bmatrix} -1/\hat{\lambda} & \hat{P}/\hat{\lambda}^2 \\ -\hat{\Psi}' & 1/\hat{\lambda} \end{bmatrix} = \begin{bmatrix} -12,97515 & 405,8353 \\ -0,51241 & 12,97515 \end{bmatrix}.$$

[Функция Ψ' — это $d^2 \ln \Gamma(P)/dP^2 = (\Gamma\Gamma'' - \Gamma'^2)/\Gamma^2$. С $\hat{P} = 2,4106$, $\hat{\Gamma} = 1,250832$, $\hat{\Psi} = 0,658347$ и $\hat{\Psi}' = 0,512408$]². Матрица \mathbf{F} — это выборочная ковариационная матрица y и $\ln y$ (используя 19 в качестве делителя):

$$\mathbf{F} = \begin{bmatrix} 500,68 & 14,31 \\ 14,31 & 0,47746 \end{bmatrix}.$$

Произведение — это

$$\frac{1}{n} [\hat{G}' \mathbf{F}^{-1} \hat{G}]^{-1} = \begin{bmatrix} 0,38978 & 0,014605 \\ 0,014605 & 0,00068747 \end{bmatrix}.$$

Для оценки методом максимального правдоподобия оценка асимптотической ковариационной матрицы, основанная на ожидаемом (и реализовавшемся) гесссиане, — это

$$[-\mathbf{H}]^{-1} = \frac{1}{n} \begin{bmatrix} \Psi' & -1/\lambda \\ -1/\lambda & P/\lambda^2 \end{bmatrix}^{-1} = \begin{bmatrix} 0,51243 & 0,01638 \\ 0,01638 & 0,00064654 \end{bmatrix}.$$

² Ψ' — это тригамма-функция. Значения для $\Gamma(P)$, $\Psi(P)$ и $\Psi'(P)$ затабулированы в Abramovitz, Stegun (1971). Эти значения были получены с использованием программной библиотеки IMSL.

Гессиан содержит те же элементы, что и у G , поскольку мы решили использовать достаточные статистики для оценок моментов, так что моментные уравнения, которые мы дифференцировали с точностью до знака, также являются производными логарифма правдоподобия. Оценки двух дисперсий равны 0,51203 и 0,00064654 соответственно, что хорошо соотносится с оценками методом моментов. Различие объясняется выборочной вариативностью в конечных выборках и присутствием F в первой оценке дисперсии.

13.2.3. Заключение — метод моментов

В простейших случаях метод моментов устойчив к различиям в спецификации процесса, порождающего данные (ППД). Выборочное среднее или дисперсия оценивают свои теоретические аналоги (преполагая их существование) независимо от порождающего их процесса. Именно эта свобода от необоснованных предположений о распределениях сделала в последнее время этот метод таким популярным. Однако за это приходится платить. Если о ППД известно больше, например его конкретное распределение, тогда метод моментов не может использовать всю доступную информацию. Таким образом, в примере 13.3 естественные оценки параметров распределения, основанные на выборочном среднем и дисперсии, оказываются неэффективными. Метод максимального правдоподобия, который лежит в основе существенной части эконометрических исследований, становится альтернативным подходом, используя эту не содержащуюся в выборке информацию и являясь, таким образом, более эффективным.

13.3. Оценка методом наименьшего расстояния

Приведенный выше анализ рассматривал случаи точной идентификации. В каждом примере было K оцениваемых параметров и мы использовали K моментов для их оценки. В примере 13.5 мы проанализировали гамма-распределение, семейство с двумя параметрами и рассмотрели различные пары моментов, которые могли быть использованы для оценки этих двух параметров. (Наиболее эффективная оценка для параметров этого распределения будет основана на $(1/n)\sum_i y_i$ и $(1/n)\sum_i \ln y_i$.) Это поднимает следующий общий вопрос: как стоит поступить, имея большее, чем необходимо, количество моментов? Было бы непродуктивно просто избавиться от дополнительной информации. В таком случае, следуя логике, информация в выборке дает возможность получить более одной оценки параметров модели и встает вопрос о том, как согласовать эти конкурирующие оценки.

Мы сталкивались с этой ситуацией в некоторых предыдущих примерах: в примере 11.29 исследования Fannie Mae, представленном в работе Passmore (2005), было четыре независимые оценки одного параметра $\hat{\alpha}_j$ с оценкой асимптотической дисперсии $\hat{V}_j, j = 1, \dots, 4$. Оценки комбинировали, ис-

пользуя целевую функцию

$$\text{минимум по } \alpha : q = \sum_{j=1}^4 \frac{(\hat{\alpha}_j - \alpha)^2}{\hat{V}_j}.$$

Решение этой минимизационной задачи имеет вид

$$\hat{\alpha}_{\text{MDE}} = \sum_{j=1}^4 w_j \hat{\alpha}_j, w_j = \frac{1/\hat{V}_j}{\sum_{s=1}^4 (1/\hat{V}_s)}, j = 1, \dots, 4 \text{ и } \sum_{j=1}^4 w_j = 1.$$

Формируя двухшаговую оценку наименьших квадратов параметров динамической модели панельных данных в подразделе 11.11.3, мы получили $T - 2$ -оценки инструментальных переменных вектора параметров θ , формируя различные инструменты для каждого периода, для которого у нас было достаточно данных. $T - 2$ -оценки одного вектора параметров равны $\hat{\theta}_{\text{IV}(t)}$. Оценка Ареллано–Бонда для единственного вектора параметров в этой постановке равна

$$\begin{aligned} \hat{\theta}_{\text{IV}} &= \left(\sum_{t=3}^T \mathbf{W}_{(t)} \right)^{-1} \left(\sum_{t=3}^T \mathbf{W}_{(t)} \hat{\theta}_{\text{IV}(t)} \right) \\ &= \sum_{t=3}^T \mathbf{R}_{(t)} \hat{\theta}_{\text{IV}(t)}, \end{aligned}$$

где

$$\mathbf{W}_{(t)} = \left(\hat{\mathbf{X}}'_{(t)} \hat{\mathbf{X}}_{(t)} \right)$$

и

$$\mathbf{R}_{(t)} = \left(\sum_{t=3}^T \mathbf{W}_{(t)} \right)^{-1} \mathbf{W}_{(t)} \text{ и } \sum_{t=3}^T \mathbf{R}_{(t)} = \mathbf{I}.$$

Наконец, анализ Кэри (Carey (1997)) издержек больниц, который мы рассматривали в примере 11.10, включал модель внешне не связанных регрессий, которая давала несколько оценок параметров модели. Мы вернемся к этому приложению в примере 13.6.

Оценка методом наименьшего расстояния (MDE) определяется следующим образом: пусть $\bar{m}_{n,l}$ обозначает выборочную статистику, основанную на n наблюдениях, такую, что

$$\text{plim } \bar{m}_{n,l} = g_l(\theta_0), l = 1, \dots, L,$$

где θ_0 — это вектор $K \leq L$ оцениваемых параметров. Представим эти моменты и функции как $L \times 1$ -векторы \bar{m}_n и $g(\theta_0)$ и предположим также, что статистики совместно нормально распределены с $\text{plim } \bar{m}_n = g(\theta)$ и $\text{Asy.Var}[\bar{m}_n] = (1/n)\Phi$. Определим целевую функцию

$$q = [\bar{m}_n - g(\theta)]' \mathbf{W} [\bar{m}_n - g(\theta)]$$

для положительно определенной матрицы весов \mathbf{W} . Оценка наименьшего расстояния — это $\hat{\theta}_{\text{MDE}}$, которая минимизирует q . Различные варианты \mathbf{W}

приведут к разным оценкам, но оценка обладает следующими свойствами для любой W :

Теорема 13.1. *Асимптотическое распределение оценки методом наименьшего расстояния*

В предположении, что $\sqrt{n}[\bar{m}_n - g(\theta_0)] \xrightarrow{d} N[0, \Phi]$, оценка методом наименьшего расстояния обладает следующими асимптотическими свойствами:

$$\text{plim } \hat{\theta}_{\text{MDE}} = \theta_0,$$

$$\begin{aligned} \text{Asy. Var } [\hat{\theta}_{\text{MDE}}] &= \frac{1}{n} [\Gamma(\theta_0)' W \Gamma(\theta_0)]^{-1} [\Gamma(\theta_0)' W \Phi W \Gamma(\theta_0)] \times \\ &\times [\Gamma(\theta_0)' W \Gamma(\theta_0)]^{-1} = \frac{1}{n} V, \end{aligned}$$

где

$$\Gamma(\theta_0) = \text{plim } G(\hat{\theta}_{\text{MDE}}) = \text{plim } \frac{\partial g(\hat{\theta}_{\text{MDE}})}{\partial \hat{\theta}'_{\text{MDE}}}$$

и

$$\hat{\theta}_{\text{MDE}} \xrightarrow{a} N \left[\theta_0, \frac{1}{n} V \right].$$

Доказательства могут быть найдены в Malinvaud (1970) и Amemiya (1985). Для наших целей мы можем отметить, что MDE – это обобщение метода моментов, представленного в предыдущем разделе. Одним из следствий является то, что оценка состоятельна для любой W , но асимптотическая ковариационная матрица является функцией W . Это говорит о том, что выбор W может быть сделан с учетом размера ковариационной матрицы и что может существовать оптимальный вариант. Это оказывается именно так. Для оценки наименьшего расстояния взвешивающая матрица, минимизирующая дисперсию, – это

$$\begin{aligned} \text{оптимальная матрица весов : } W^* &= [\text{Asy. Var. } \sqrt{n}\{\bar{m}_n - g(\theta)\}]^{-1} \\ &= \Phi^{-1}. \end{aligned}$$

[См. обсуждение в Hansen (1982).] При таком выборе W

$$\text{Asy. Var } [\hat{\theta}_{\text{MDE}}] = \frac{1}{n} [\Gamma(\theta_0)' \Phi^{-1} \Gamma(\theta_0)]^{-1},$$

что также является результатом, полученным нами ранее для оценки методом моментов.

Решение задачи оценки MDE получается путем нахождения $\hat{\theta}_{\text{MDE}}$ такой, что

$$\frac{\partial q}{\partial \hat{\theta}_{\text{MDE}}} = -G(\hat{\theta}_{\text{MDE}})' W [\bar{m}_n - g(\hat{\theta}_{\text{MDE}})] = 0.$$

Важная особенность MDE проявляется в случае точной идентификации. Если K равняется L и если функции $g_l(\theta)$ функционально независимы, т.е. $G(\theta)$

имеет полный ранг K , тогда можно точно решить моментные уравнения. То есть минимизационная задача становится задачей решения K моментных уравнений $\bar{m}_{n,l} = g_l(\theta_0)$ от K неизвестных $\hat{\theta}_{\text{MDE}}$. Это оценка методом моментов, рассмотренная в предыдущем разделе. В таком случае матрица весов \mathbf{W} не оказывает влияния на решение, поскольку MDE будет удовлетворять моментным уравнениям

$$\left[\bar{\mathbf{m}}_n - \mathbf{g}(\hat{\theta}_{\text{MDE}}) \right] = \mathbf{0}.$$

Для примеров, приведенных ранее, относящихся к случаям **переопределенных** систем, оценки методом наименьшего расстояния определяются как

$$q = ((\hat{\alpha}_1 - \alpha)(\hat{\alpha}_2 - \alpha)(\hat{\alpha}_3 - \alpha)(\hat{\alpha}_4 - \alpha)) \begin{bmatrix} \hat{V}_1 & 0 & 0 & 0 \\ 0 & \hat{V}_2 & 0 & 0 \\ 0 & 0 & \hat{V}_3 & 0 \\ 0 & 0 & 0 & \hat{V}_4 \end{bmatrix}^{-1} \begin{pmatrix} (\hat{\alpha}_1 - \alpha) \\ (\hat{\alpha}_2 - \alpha) \\ (\hat{\alpha}_3 - \alpha) \\ (\hat{\alpha}_4 - \alpha) \end{pmatrix}$$

для представленного Passmore (2005) анализа Fannie Mae и

$$q = ((\mathbf{b}_{\text{IV}(3)} - \boldsymbol{\theta}) \dots (\mathbf{b}_{\text{IV}(T)} - \boldsymbol{\theta}))' \begin{bmatrix} (\hat{\mathbf{X}}'_{(3)} \hat{\mathbf{X}}_{(3)}) & \dots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \dots & (\hat{\mathbf{X}}'_{(T)} \hat{\mathbf{X}}_{(T)}) \end{bmatrix}^{-1} \times \\ \times \begin{pmatrix} (\mathbf{b}_{\text{IV}(3)} - \boldsymbol{\theta}) \\ \vdots \\ (\mathbf{b}_{\text{IV}(T)} - \boldsymbol{\theta}) \end{pmatrix}$$

для оценки Ареллано–Бонда динамической модели панельных данных.

Пример 13.6. Оценка методом наименьшего расстояния функции издержек больницы

В исследовании издержек больниц Кэри (Carey (1997)) из примера 11.10 подход внешне не связанных регрессий (SUR) к панельным данным, представленный в Chamberlain (1984), дает пять оценок вектора параметров для каждого периода θ_t . Одни параметры относятся к конкретному году, тогда как другие (исходя из гипотезы) постоянны для всех пяти лет. Интерес представляют два конкретных параметра β_D и β_O , которые меняются по годам, но оцениваются несколько раз в модели SUR. Мы остановимся только на этих параметрах. Модель говорит, что

$$y_{it} = \alpha_i + A_{it} + \beta_{D,t} DIS_{it} + \beta_{O,t} OUT_{it} + \varepsilon_{it},$$

где

$$\alpha_i = B_i + \sum_t \gamma_{D,t} DIS_{it} + \sum_t \gamma_{O,t} OUT_{it} + u_i, t = 1987, \dots, 1991,$$

DIS_{it} — число выписанных из больниц пациентов и OUT_{it} — число посещений амбулаторных отделений больниц. (Мы слегка меняем обозначения Кэри и опускаем части модели, внешние по отношению к рассматриваемой нами проблеме. Члены A_{it} и B_i содержат эти дополнительные компоненты.) Представленная модель оценивается подстановкой выражения для α_i в основное уравнение и последующей оценкой неограниченной системы внешне не связанных регрессий доступным FGLS. Данные доступны для пяти лет, что дает пять наборов оценок. Отметим, однако, по отношению к переменной число выписанных из больниц пациентов DIS , несмотря на то что каждое уравнение дает свою оценку $(\gamma_{D,1}, \dots, (\beta_{D,t} + \gamma_{D,t}), \dots, \gamma_{D,5})$, т.е. у нас есть всего пять оценок параметров в каждом уравнении (в каждом году), оценивается только 10, а не 25 параметров. Параметры при OUT_{it} также переопределены. Таблица 13.1 воспроизводит оценки из табл. 11.7 для коэффициентов при числе выписанных из больниц пациентов и добавляет оценки для переменной посещений амбулаторных отделений больниц.

Таблица 13.1а. Оценки коэффициентов DIS в модели SUR для издержек госпиталей

| Equation | Коэффициент при переменной в уравнении | | | | |
|----------|--|--|--|--|--|
| | DIS87 | DIS88 | DIS89 | DIS90 | DIS91 |
| SUR87 | $\beta_{D,87} + \gamma_{D,87}$ 1,76 | $\gamma_{D,88}$ 0,116 | $\gamma_{D,89}$ -0,0881 | $\gamma_{D,90}$ 0,0570 | $\gamma_{D,91}$ -0,0617 |
| SUR88 | $\gamma_{D,87}$ 0,254 | $\beta_{D,88} + \gamma_{D,88}$ 1,61 | $\gamma_{D,89}$ -0,0934 | $\gamma_{D,90}$ 0,0610 | $\gamma_{D,91}$ -0,0514 |
| SUR89 | $\gamma_{D,87}$ 0,217 | $\gamma_{D,88}$ 0,0846 | $\beta_{D,89} + \gamma_{D,89}$ 1,51 | $\gamma_{D,90}$ 0,0454 | $\gamma_{D,91}$ -0,0253 |
| SUR90 | $\gamma_{D,87}$ 0,179 | $\gamma_{D,88}$ 0,0822 | $\gamma_{D,89}$ 0,0295 | $\beta_{D,90} + \gamma_{D,90}$ 1,57 | $\gamma_{D,91}$ 0,0244 |
| SUR91 | $\gamma_{D,87}$ 0,153 | $\gamma_{D,88}$ 0,0363 | $\gamma_{D,89}$ -0,0422 | $\gamma_{D,90}$ 0,0813 | $\beta_{D,91} + \gamma_{D,91}$ 1,70 |
| MDE | $\beta = 1,50$ $\gamma = 0,219$ | $\beta = 1,58$ $\gamma = 0,0666$ | $\beta = 1,54$ $\gamma = -0,0539$ | $\beta = 1,57$ $\gamma = 0,0690$ | $\beta = 1,63$ $\gamma = -0,0213$ |

Из таблицы мы видим, что модель SUR дает четыре прямые оценки $\gamma_{D,87}$, основанные на уравнениях для 1988–1991 гг. Она также неявно дает четыре оценки $\beta_{D,87}$, поскольку любая из четырех оценок $\gamma_{D,87}$ из последних четырех уравнений может быть вычтена из коэффициента при DIS в уравнении для 1987 г. для оценки $\beta_{D,87}$. Всего мы имеем 50 оценок параметров различных функций от 20 параметров:

$$\theta = (\beta_{D,87}, \dots, \beta_{D,91}), (\gamma_{D,87}, \dots, \gamma_{D,91}), (\beta_{O,87}, \dots, \beta_{O,91}), (\gamma_{O,87}, \dots, \gamma_{O,91})$$

и, таким образом, 30 ограничений для поиска общей ограниченной оценки. MDE была использована для объединения конкурирующих оценок. Пусть $\hat{\beta}_t$ обозначает 10×1 оценку параметров модели для конкретного периода. В отличие от других рассмотренных нами случаев представленные

здесь индивидуальные оценки некоррелированы. В модели SUR оценка асимптотической ковариационной матрицы – это разделенная на блоки матрица, приведенная в (10-7). Для оценок двух уравнений

$$\text{Est. Asy. Cov} [\hat{\beta}_t, \hat{\beta}_s] = t, s$$

блок матрицы

$$\begin{bmatrix} \hat{\sigma}^{11} \mathbf{X}'_1 \mathbf{X}_1 & \hat{\sigma}^{12} \mathbf{X}'_1 \mathbf{X}_2 & \dots & \hat{\sigma}^{15} \mathbf{X}'_1 \mathbf{X}_5 \\ \hat{\sigma}^{21} \mathbf{X}'_2 \mathbf{X}_1 & \hat{\sigma}^{22} \mathbf{X}'_2 \mathbf{X}_2 & \dots & \hat{\sigma}^{25} \mathbf{X}'_2 \mathbf{X}_5 \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}^{51} \mathbf{X}'_5 \mathbf{X}_1 & \hat{\sigma}^{52} \mathbf{X}'_5 \mathbf{X}_2 & \dots & \hat{\sigma}^{55} \mathbf{X}'_5 \mathbf{X}_5 \end{bmatrix}^{-1} = \hat{\mathbf{V}}_{ts},$$

где $\hat{\sigma}^{ts}$ – это элемент t, s $\hat{\Sigma}^{-1}$. (Мы выделяем подматрицу релевантных матриц, поскольку модель SUR у Кэри содержала 26 других переменных в каждом уравнении в дополнение к пяти периодам DIS и OUT.) 50×50 матрица весов для MDE имеет вид

$$\mathbf{W} = \begin{bmatrix} \hat{\mathbf{V}}_{87,87} & \hat{\mathbf{V}}_{87,88} & \hat{\mathbf{V}}_{87,89} & \hat{\mathbf{V}}_{87,90} & \hat{\mathbf{V}}_{87,91} \\ \hat{\mathbf{V}}_{88,87} & \hat{\mathbf{V}}_{88,88} & \hat{\mathbf{V}}_{88,89} & \hat{\mathbf{V}}_{88,90} & \hat{\mathbf{V}}_{88,91} \\ \hat{\mathbf{V}}_{89,87} & \hat{\mathbf{V}}_{89,88} & \hat{\mathbf{V}}_{89,89} & \hat{\mathbf{V}}_{89,90} & \hat{\mathbf{V}}_{89,91} \\ \hat{\mathbf{V}}_{90,87} & \hat{\mathbf{V}}_{90,88} & \hat{\mathbf{V}}_{90,89} & \hat{\mathbf{V}}_{90,90} & \hat{\mathbf{V}}_{90,91} \\ \hat{\mathbf{V}}_{91,87} & \hat{\mathbf{V}}_{91,88} & \hat{\mathbf{V}}_{91,89} & \hat{\mathbf{V}}_{91,90} & \hat{\mathbf{V}}_{91,91} \end{bmatrix}^{-1} = [\hat{\mathbf{V}}^{ts}].$$

Вектор квадратичной формы – это набор пяти 10×1 векторов, первый из которых имеет вид

$$\begin{aligned} & \bar{\mathbf{m}}_{n,87} - \mathbf{g}_{87}(\theta) \\ & = \begin{bmatrix} \left\{ \hat{\beta}_{D,87}^{87} - (\beta_{D,87} + \gamma_{D,87}) \right\}, \left\{ \hat{\beta}_{D,88}^{87} - \gamma_{D,88} \right\}, \left\{ \hat{\beta}_{D,89}^{87} - \gamma_{D,89} \right\}, \\ \left\{ \hat{\beta}_{D,90}^{87} - \gamma_{D,90} \right\}, \left\{ 3\hat{\beta}_{D,91}^{87} - \gamma_{D,90} \right\}, \\ \left\{ \hat{\beta}_{O,87}^{87} - (\beta_{O,87} + \gamma_{O,87}) \right\}, \left\{ \hat{\beta}_{O,88}^{87} - \gamma_{O,88} \right\}, \left\{ \hat{\beta}_{O,89}^{87} - \gamma_{D,89} \right\}, \\ \left\{ 3\hat{\beta}_{O,90}^{87} - \gamma_{O,90} \right\}, \left\{ 3\hat{\beta}_{O,91}^{87} - \gamma_{O,90} \right\} \end{bmatrix}' \end{aligned}$$

для уравнения 1987 г. и так далее для четырех других уравнений. Целевая функция MDE для этой модели – это

$$q = \sum_{t=1987}^{1991} \sum_{s=1987}^{1991} [\bar{\mathbf{m}}_t - \mathbf{g}_t(\theta)]' \hat{\mathbf{V}}^{ts} [\bar{\mathbf{m}}_s - \mathbf{g}_s(\theta)].$$

Следует отметить, что всего в целевой функции 50 оценок параметров из уравнений SUR (они приведены в табл. 13.1) и 20 неизвестных параметров для калибровки. Приведенные оценки методом наименьшего расстояния указаны в последних строках каждой таблицы.

Таблица 13.1b. Оценки коэффициентов OUT в модели SUR для издержек госпиталей

| Equation | Коэффициент при переменной в уравнении | | | | |
|----------|--|--|--|---|--|
| | OUT87 | OUT88 | OUT89 | OUT90 | OUT91 |
| SUR87 | $\beta_{0,87} + \gamma_{D,87}$ 0,0139 | $\gamma_{0,88}$ 0,00292 | $\gamma_{0,89}$ 0,00157 | $\gamma_{0,90}$ 0,000951 | $\gamma_{0,91}$ 0,000678 |
| SUR88 | $\gamma_{0,87}$ 0,00347 | $\beta_{0,88} + \gamma_{0,88}$ 0,0125 | $\gamma_{0,89}$ 0,00501 | $\gamma_{0,90}$ 0,00550 | $\gamma_{0,91}$ 0,00503 |
| SUR89 | $\gamma_{0,87}$ 0,00118 | $\gamma_{0,88}$ 0,00159 | $\beta_{0,89} + \gamma_{0,89}$ 0,00832 | $\gamma_{0,90}$ -0,00220 | $\gamma_{0,91}$ -0,00156 |
| SUR90 | $\gamma_{0,87}$ -0,00226 | $\gamma_{0,88}$ -0,00155 | $\gamma_{0,89}$ 0,000401 | $\beta_{0,90} + \gamma_{0,90}$ 0,00897 | $\gamma_{0,91}$ 0,000450 |
| SUR91 | $\gamma_{0,87}$ 0,00278 | $\gamma_{0,88}$ 0,00255 | $\gamma_{0,89}$ 0,00233 | $\gamma_{0,90}$ 0,00305 | $\beta_{0,91} + \gamma_{0,91}$ 0,0105 |
| MDE | $\beta = 0,0112$ $\gamma = 0,00177$ | $\beta = 0,00999$ $\gamma = 0,00408$ | $\beta = 0,0100$ $\gamma =$ -0,00011 | $\beta = 0,00915$ $\gamma =$ -0,00073 | $\beta = 0,00793$ $\gamma = 0,00267$ |

13.4. Оценка обобщенным методом моментов (GMM)

Большая часть авторов недавних эмпирических работ в эконометрике, в частности в макроэкономике и финансах, использовали GMM-оценки. Как мы увидим, этот широкий класс оценок на самом деле включает большую часть оценок, обсуждаемых в этой книге. Метод оценивания GMM является развитием метода наименьшего расстояния, описанного в разделе 13.3³. Ниже мы расширим обобщенный метод моментов до других моделей помимо обобщенной линейной регрессии и заполним некоторые пробелы в выкладках из раздела 13.2.

13.4.1. Оценивание на основе условий ортогональности

Рассмотрим оценку методом наименьших квадратов параметров классической регрессионной модели. Важным предположением модели является

$$E[x_i \varepsilon_i] = E[x_i(y_i - x_i' \beta)] = 0.$$

³Формальное изложение результатов, требуемых для этого анализа, приведено в Hansen (1982), Hansen, Singleton (1988), Chamberlain (1987), Cumby, Huizinga, Obstfeld (1983), Newey (1984, 1985a, 1985b), Davidson, MacKinnon (1993) и Newey, McFadden (1994). Полезные обзоры GMM-оценивания и других нововведений в эконометрике приведены в Pagan, Wickens (1989) и Matyas (1999). Приложение некоторых из этих методов, содержащее полезные обзоры, — это Pagan и Vella (1989). Некоторое дальнейшее обсуждение приведено в Davidson и MacKinnon (2004). Ruud (2000) приводит много теоретических деталей. Hayashi (2000) представляет собой другое развернутое изложение теории оценивания, основанное на GMM-оценках.

Выборочный аналог имеет вид

$$\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \hat{\varepsilon}_i = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i (y_i - \mathbf{x}_i' \hat{\beta}) = \mathbf{0}.$$

Оценка β удовлетворяет этим моментным уравнениям, являющимся обычными уравнениями оценки методом наименьших квадратов. Таким образом, мы видим, что оценка МНК является оценкой методом моментов. Для оценки инструментальных переменных из главы 8 мы использовали асимптотический аналог моментного условия

$$\text{plim} \left(\frac{1}{n} \sum_{i=1}^n \mathbf{z}_i \varepsilon_i \right) = \text{plim} \left(\frac{1}{n} \sum_{i=1}^n \mathbf{z}_i (y_i - \mathbf{x}_i' \beta) \right) = \mathbf{0}.$$

Мы решили проблему превышения числа доступных инструментов над числом параметров, решив уравнения

$$\left(\frac{1}{n} \mathbf{X}' \mathbf{Z} \right) \left(\frac{1}{n} \mathbf{Z}' \mathbf{Z} \right)^{-1} \left(\frac{1}{n} \mathbf{Z}' \hat{\varepsilon} \right) = \frac{1}{n} \hat{\mathbf{X}}' \hat{\varepsilon} = \frac{1}{n} \sum_{i=1}^n \hat{\mathbf{x}}_i \hat{\varepsilon}_i = \mathbf{0},$$

где столбцы $\hat{\mathbf{X}}$ — это оцененные значения регрессий на все столбцы \mathbf{Z} (т.е. проекция этих столбцов \mathbf{X} на пространство столбцов \mathbf{Z}). (См. подробное обсуждение в подразделе 8.3.4.)

Оценка нелинейным методом наименьших квадратов определялась похожим образом, однако в этом случае обычные уравнения более сложны, поскольку оценка неявная. **Условия ортогональности** в генеральной совокупности для нелинейной регрессионной модели — это $E[\mathbf{x}_i^0 \varepsilon_i] = \mathbf{0}$. **Эмпирическое моментное уравнение** имеет вид

$$\frac{1}{n} \sum_{i=1}^n \left(\frac{\partial E[y_i | \mathbf{x}_i, \beta]}{\partial \beta} \right) (y_i - E[y_i | \mathbf{x}_i, \beta]) = \mathbf{0}.$$

Оценки методом максимального правдоподобия получаются приравниванием производных логарифма функции правдоподобия к нулю. Нормированная функция логарифма правдоподобия равна

$$\frac{1}{n} \ln L = \frac{1}{n} \sum_{i=1}^n \ln f(y_i | \mathbf{x}_i, \theta),$$

где $f(\cdot)$ — функция плотности, θ — вектор параметров. Для плотностей, которые удовлетворяют условиям регулярности [см. раздел 14.4.1], имеем

$$E \left[\frac{\partial \ln f(y_i | \mathbf{x}_i, \theta)}{\partial \theta} \right] = \mathbf{0}.$$

Оценка методом максимального правдоподобия получается приравниванием выборочного аналога к нулю:

$$\frac{1}{n} \frac{\partial \ln L}{\partial \hat{\theta}} = \frac{1}{n} \sum_{i=1}^n \frac{\partial \ln f(y_i | \mathbf{x}_i, \hat{\theta})}{\partial \hat{\theta}} = \mathbf{0}.$$

(Деление на n для сопоставимости с вышеприведенными примерами не меняет решения.) Вывод состоит в том, что практически все приведенные нами и встречающиеся впоследствии оценки могут интерпретироваться как оценки методом моментов. [В изложении Манским (Manski (1992)) **аналогового оценивания** приводятся некоторые интересные обобщения и методологический обзор.]

По ходу развития этого рассуждения станет ясно, что большинство определяемых в этой книге оценок могут рассматриваться как оценки обобщенным методом моментов.

13.4.2. Обобщение метода моментов

Представленные выше примеры объединяет одна особенность. В каждом приведенном случае, за исключением общей постановки для оценки методом инструментальных переменных, моментных уравнений ровно столько же, сколько и оцениваемых параметров. Таким образом, каждый из представленных случаев является случаем **точной идентификации**. Решение для моментных уравнений будет единственным, и в этом решении уравнения выполняются в точности⁴. Но встречаются случаи, в которых моментных уравнений больше, чем параметров, так что система переопределена. В примере 13.5 мы определили четыре выборочных момента

$$\bar{g} = \frac{1}{n} \sum_{i=1}^n \left[y_i, y_i^2, \frac{1}{y_i}, \ln y_i \right]$$

с пределами по вероятности P/λ , $P(P+1)/\lambda^2$, $\lambda/(P-1)$ и $\psi(P) - \ln \lambda$ соответственно. Любая пара может быть использована для оценки двух параметров, но, как было показано в примере выше, шесть пар дают шесть несколько различающихся оценок $\theta = (P, \lambda)$.

В таком случае для использования всей выборочной информации необходимо найти способ соотнесения различных оценок, которые могут получиться из переопределенной системы. Обобщая, предположим, что модель включает K параметров $\theta = (\theta_1, \theta_2, \dots, \theta_K)'$ и что теория предоставляет набор из $L > K$ моментных условий

$$E [m_l(y_i, \mathbf{x}_i, \mathbf{z}_i, \theta)] = E [m_{il}(\theta)] = 0,$$

где y_i , \mathbf{x}_i и \mathbf{z}_i — это переменные, входящие в модель, и индекс i у $m_{il}(\theta)$ указывает на зависимость от $(y_i, \mathbf{x}_i, \mathbf{z}_i)$. Обозначим соответствующие выборочные средние как

$$\bar{m}_l(y, \mathbf{X}, \mathbf{Z}, \theta) = \frac{1}{n} \sum_{i=1}^n m_l(y_i, \mathbf{x}_i, \mathbf{z}_i, \theta) = \frac{1}{n} \sum_{i=1}^n m_{il}(\theta).$$

Если уравнения функционально независимы, система L уравнений с K неизвестными параметрами,

$$\bar{m}_l(\theta) = \frac{1}{n} \sum_{i=1}^n m_l(y_i, \mathbf{x}_i, \mathbf{z}_i, \theta) = 0, \quad l = 1, \dots, L,$$

⁴Конечно, если существует *любое* решение. В регрессионной модели с мультиколлинеарностью параметров K , но независимых моментных уравнений меньше, чем K .

не будет иметь единственного решения⁵. Для удобства моментные уравнения определяются здесь неявно в отличие от равенства моментов функциям, как в разделе 13.3. Будет необходимо привести в соответствие $\binom{L}{K}$ различных наборов оценок, которые могут получиться. Одним возможным подходом может быть минимизация целевой функции, такой как сумма квадратов⁶

$$q = \sum_{l=1}^L \bar{m}_l^2 = \bar{\mathbf{m}}(\theta)' \bar{\mathbf{m}}(\theta). \quad (13-2)$$

Можно показать [см., например, Hansen (1982)], что в сделанных выше предположениях, в частности, что $\text{plim} \bar{\mathbf{m}}(\theta) = E[\bar{\mathbf{m}}(\theta)] = \mathbf{0}$, значение, минимизирующее q в (13-2), дает состоятельную (хотя, как мы увидим, возможно неэффективную) оценку θ . На самом деле мы можем использовать в качестве критерия взвешенную сумму квадратов

$$q = \bar{\mathbf{m}}(\theta)' \mathbf{W}_n \bar{\mathbf{m}}(\theta),$$

где \mathbf{W}_n — любая положительно определенная матрица, которая может зависеть от данных, но не является функцией θ , например \mathbf{I} в (13-2), для получения состоятельной оценки θ ⁷. Например, мы можем использовать диагональную матрицу весов, если доступна какая-либо информация о важности (по какому-либо критерию) различных моментов. Мы делаем дополнительное предположение о том, что $\text{plim} \mathbf{W}_n = \mathbf{a}$ — положительно определенная матрица \mathbf{W} .

По той же логике, что делает обобщенный метод наименьших квадратов предпочтительным по отношению к обыкновенному методу наименьших квадратов, должно быть выгодно использовать взвешенный критерий, где веса обратно пропорциональны дисперсиям моментов. Пусть \mathbf{W} будет диагональной матрицей, диагональные элементы которой обратно пропорциональны дисперсиям отдельных моментов:

$$w_{ll} = \frac{1}{\text{Asy. Var}[\sqrt{n} \bar{m}_l]} = \frac{1}{\phi_{ll}}.$$

(Запись в такой форме подчеркивает, что правая часть включает дисперсию выборочного среднего, которое имеет порядок $(1/n)$.) Тогда **оценка взвешенным методом наименьших квадратов** минимизирует

$$q = \bar{\mathbf{m}}(\theta)' \Phi^{-1} \bar{\mathbf{m}}(\theta). \quad (13-3)$$

В общем L элементов $\bar{\mathbf{m}}$ свободно коррелированы. В (13-3) мы использовали диагональную \mathbf{W} , которая игнорирует эту корреляцию. Для использования обобщенного метода наименьших квадратов мы определили бы полную

⁵Это возможно, если L превышает размер выборки n . Мы предполагаем что L строго меньше, чем n .

⁶Этот подход предложен Quandt, Ramsey (1978) для задачи из примера 13.4.

⁷В принципе взвешивающая матрица может также быть функцией от параметров. См. обсуждение в Hansen, Heaton, Yaron (1996). Маловероятно, что это каким-либо образом улучшает асимптотические свойства оценки. Одним из преимуществ, отмеченных авторами, является то, что некоторые оценки становятся инвариантными по отношению к нормализации, обсуждавшейся в примере 14.1. На практике это может быть актуально в небольшом числе случаев.

матрицу

$$\mathbf{W} = \{ \text{Asy. Var}[\sqrt{n} \bar{\mathbf{m}}] \}^{-1} = \Phi^{-1}. \quad (13-4)$$

Оценки, определяемые выбором θ с целью минимизации

$$q = \bar{\mathbf{m}}(\theta)' \mathbf{W}_n \bar{\mathbf{m}}(\theta),$$

являются оценками методом наименьшего расстояния по определению из раздела 13.3. Общим результатом будет то, что если \mathbf{W}_n — положительно определенная матрица и если

$$\text{plim } \bar{\mathbf{m}}(\theta) = \mathbf{0},$$

то оценка методом наименьшего расстояния (обобщенным методом моментов, или GMM) θ состоятельна⁸. Поскольку МНК-критерий в (13-2) использует \mathbf{I} , этот метод дает состоятельную оценку, как и оценка взвешенным методом наименьших квадратов и полная ОМНК-оценка. Остается определить наилучшую \mathbf{W} . Интуиция подсказывает (верно), что матрица, определенная в (13-4), будет оптимальной опять же на основании логики, лежащей в основе обобщенного метода наименьших квадратов. Этот знаменательный результат принадлежит Хансену (Hansen (1982)).

Асимптотическая ковариационная матрица **оценки обобщенным методом моментов (ОММ)** равна

$$\mathbf{V}_{GMM} = \frac{1}{n} [\mathbf{\Gamma}' \mathbf{W} \mathbf{\Gamma}]^{-1} = \frac{1}{n} [\mathbf{\Gamma}' \Phi^{-1} \mathbf{\Gamma}]^{-1}, \quad (13-5)$$

где $\mathbf{\Gamma}$ — это матрица производных, в которой j -я строка равна

$$\mathbf{\Gamma}^j = \text{plim } \frac{\partial \bar{m}_j(\theta)}{\partial \theta'},$$

и $\Phi = \text{Asy. Var}[\sqrt{n} \bar{\mathbf{m}}]$. Наконец, по центральной предельной теореме, примененной к выборочным моментам, и **теореме Слуцкого**, примененной к этому преобразованию, мы можем ожидать, что оценка будет асимптотически нормальной. Мы рассмотрим асимптотические свойства оценки в подразделе 13.4.3.

Пример 13.7. GMM-оценка нелинейной регрессионной модели

В примере 7.6 мы рассматривали нелинейную регрессионную модель для доходов, использующую данные Немецкой социоэкономической панели (German Socioeconomic Panel). Регрессионная модель имела вид

$$\text{Income} = h(1, \text{Age}, \text{Education}, \text{Female}, \gamma) + \varepsilon,$$

где $h(\cdot)$ — это экспоненциальная функция переменных. В этом примере мы использовали несколько перекрестных произведений. В этом приложении

⁸В наиболее общих случаях некоторые специфические условия должны выполняться, чтобы обеспечить состоятельность и другие обсуждаемые нами свойства. Для наших целей будет достаточно приведенных условий. Оценки методом наименьшего расстояния обсуждаются в работах Malinvaud (1970), Hansen (1982) и Amemiya (1985).

мы несколько упростим функцию условного среднего и будем использовать

$$Income = \exp(\gamma_1 + \gamma_2 Age + \gamma_3 Education + \gamma_4 Female) + \varepsilon,$$

которую для удобства запишем как

$$y_i = \exp(\mathbf{x}_i' \boldsymbol{\gamma}) + \varepsilon_i = \mu_i + \varepsilon_i^9.$$

Выборка состоит из волны панели, относящейся к 1988 г., за исключением двух наблюдений, для которых переменная *Income* равна нулю. Итоговая выборка насчитывает 4481 наблюдение. Описательные статистики для выборочных данных приведены в табл. 7.2.

Сначала мы рассмотрим оценку параметров нелинейным методом наименьших квадратов. Уравнениями необходимых условий нелинейного метода наименьших квадратов будут

$$(1/n) \sum_i [(y_i - \mu_i) \mu_i \mathbf{x}_i] = (1/n) \sum_i [\varepsilon_i \mu_i \mathbf{x}_i] = \mathbf{0}.$$

Отметьте, что условие ортогональности включает псевдорегрессоры $\partial \mu_i / \partial \boldsymbol{\gamma} = \mathbf{x}_i^0 = \mu_i \mathbf{x}_i$. Соответствующее моментное уравнение в генеральной совокупности имеет вид

$$E[\varepsilon_i (\mu_i \mathbf{x}_i)] = \mathbf{0}.$$

Получение оценки нелинейным методом наименьших квадратов обсуждается в подразделе 7.2.6. Оценка асимптотической ковариационной матрицы равна

$$\text{Est. Asy. Var}[\hat{\boldsymbol{\gamma}}_{\text{NLSQ}}] = \frac{\sum_{i=1}^n (y_i - \hat{\mu}_i)^2}{(4481 - 4)} \left[\sum_{i=1}^{4481} (\hat{\mu}_i \mathbf{x}_i) (\hat{\mu}_i \mathbf{x}_i)' \right]^{-1},$$

где $\hat{\mu}_i = \exp(\mathbf{x}_i' \hat{\boldsymbol{\gamma}})$.

Простая оценка методом моментов может быть построена на основании гипотезы о том, что \mathbf{x}_i (не \mathbf{x}_i^0) ортогонален ε_i . Тогда

$$E[\varepsilon_i \mathbf{x}_i] = E \left[\varepsilon_i \begin{pmatrix} 1 \\ Age_i \\ Education_i \\ Female_i \end{pmatrix} \right] = \mathbf{0}$$

приводит к четырем моментным уравнениям. Выборочные аналоги равны

$$\bar{m}_k(\boldsymbol{\gamma}) = \frac{1}{n} \sum_{i=1}^n (y_i - \mu_i) x_{ik} = \frac{1}{n} \sum_{i=1}^n \varepsilon_i x_{ik}.$$

Для получения оценки методом моментов мы минимизируем сумму квадратов

$$\bar{\mathbf{m}}'(\boldsymbol{\gamma}) \bar{\mathbf{m}}(\boldsymbol{\gamma}) = \sum_{k=1}^4 \bar{m}_k^2(\boldsymbol{\gamma}).$$

Это нелинейная оптимизационная задача, которая решается итеративными методами, описанными в разделе Е.3.

Имея оценки параметров, полученные на первом шаге $\hat{\gamma}^0$, ковариационная матрица оценивается с помощью (13-5):

$$\hat{\Phi} = \left\{ \frac{1}{4481} \sum_{i=1}^{4481} \mathbf{m}_i(\hat{\gamma}^0) \mathbf{m}_i'(\hat{\gamma}^0) \right\} = \left\{ \frac{1}{4481} \sum_{i=1}^{4481} (\hat{\varepsilon}_i^0 \mathbf{x}_i) (\hat{\varepsilon}_i^0 \mathbf{x}_i)' \right\}$$

$$\bar{\mathbf{G}} = \left\{ \frac{1}{4481} \sum_{i=1}^n (\mathbf{x}_i) (-\hat{\mu}_i^0 \mathbf{x}_i)' \right\}.$$

Асимптотическая ковариационная матрица для оценки методом моментов оценивается, используя (13-5):

$$\text{Est. Asy. Var}[\hat{\gamma}_{\text{MOM}}] = \frac{1}{n} [\bar{\mathbf{G}} \hat{\Phi}^{-1} \bar{\mathbf{G}}']^{-1}.$$

Предположим, у нас есть дополнительные переменные *Health Satisfaction* и *Marital Status* такие, что хотя функция условного среднего остается той же, что и прежде, мы будем использовать их для построения GMM-оценки. Это дает еще два дополнительных моментных уравнения

$$E \left[\varepsilon_i \begin{pmatrix} \text{Health Satisfaction}_i \\ \text{Marital Status}_i \end{pmatrix} \right]$$

и всего шесть моментных уравнений для оценки четырех параметров. Мы строим оценку обобщенным методом моментов следующим образом: начальный шаг тот же, что и прежде, за исключением того, что сумма квадратов моментов $\bar{\mathbf{m}}'(\gamma) \bar{\mathbf{m}}(\gamma)$ складывает шесть, а не четыре члена. Затем мы строим

$$\hat{\Phi} = \left\{ \frac{1}{4481} \sum_{i=1}^{4481} \mathbf{m}_i(\hat{\gamma}) \mathbf{m}_i'(\hat{\gamma}) \right\} = \left\{ \frac{1}{4481} \sum_{i=1}^{4481} (\hat{\varepsilon}_i \mathbf{z}_i) (\hat{\varepsilon}_i \mathbf{z}_i)' \right\},$$

где теперь \mathbf{z}_i во второй части — это шесть экзогенных переменных, а не исходные четыре (включая константу). Таким образом, $\hat{\Phi}$ — это теперь 6×6 матрица моментов. Оптимальная матрица весов для оценки (получаемая в следующем разделе) — это $\hat{\Phi}^{-1}$. Оценка GMM получается минимизацией при условии, что γ

$$q = \bar{\mathbf{m}}'(\gamma) \hat{\Phi}^{-1} \bar{\mathbf{m}}(\gamma).$$

Асимптотическая ковариационная матрица получается с использованием (13-5), как и в случае простого метода моментов.

В табл. 13.2 приведены четыре набора оценок: нелинейным методом наименьших квадратов, методом моментов, первым шагом GMM и GMM с использованием оптимальной взвешивающей матрицы. Отмечается два сопоставления. Метод моментов дает несколько отличные от нелинейного метода наименьших квадратов оценки. Этого следует ожидать, поскольку они используют различные критерии. Судя по стандартным ошибкам,

GMM-оценка имеет очень небольшое преимущество перед оценкой нелинейным методом наименьших квадратов и оценкой методом момента. В конечном счете, однако, можно сделать вывод, что два дополнительных момента не добавляют информации для оценки параметров.

Таблица 13.2. Оценки нелинейной регрессии (стандартные ошибки в скобках)

| Оценка | Нелинейный МНК | Метод моментов | Первый шаг GMM | GMM |
|-------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Константа | -1,69331 (0,04408) | -1,62969 (0,04214) | -1,45551 (0,10102) | -1,61192 (0,04163) |
| Возраст | 0,00207 (0,00061) | 0,00178 (0,00057) | -0,00028 (0,00100) | 0,00092 (0,00056) |
| Образование | 0,04792 (0,00247) | 0,04861 (0,00262) | 0,03731 (0,00518) | 0,04647 (0,00262) |
| Женщины | -0,00658 (0,01373) | 0,00070 (0,01384) | -0,02205 (0,01445) | -0,01517 (0,01357) |

13.4.3. Свойства GMM-оценки

Теперь мы рассмотрим свойства GMM-оценки более подробно. Поскольку GMM-оценка включает другие знакомые нам оценки, в том числе методы наименьших квадратов и инструментальных переменных, эти результаты будут применимы для этих случаев. Здесь будут приведены лишь отдельные элементы формальных доказательств. Предположения, которые мы будем делать, несколько менее абстрактны, чем может позволить самое общее изложение, но они достаточно общи, чтобы включать ситуации, встречающиеся на практике. Более детальное и строгое изложение может быть найдено, например, в Newey, McFadden (1994), White (2001), Hayashi (2000), Mittelhammer et al. (2000) или Davidson (2000).

GMM-оценка основана на наборе теоретических условий ортогональности

$$E[m_i(\theta_0)] = 0,$$

где мы обозначаем истинный вектор параметров как θ_0 . Индекс i в левой части обозначает зависимость от наблюдаемых данных (y_i, x_i, z_i) . Усреднение по выборке дает выборочное моментное уравнение

$$E[\bar{m}_n(\theta_0)] = 0,$$

где

$$\bar{m}_n(\theta_0) = \frac{1}{n} \sum_{i=1}^n m_i(\theta_0).$$

⁹ Отметим, что в этой модели вероятно, что Education эндогенна. Это очень просто отразить в OMM-оценке. Однако, для целей простого численного примера, мы продолжим, предполагая, что Education экзогенна.

Этот момент — это набор L уравнений, включающих K параметров. Мы будем предполагать, что это математическое ожидание существует, что выборочный аналог сходится к нему. Определения формулируются в терминах теоретических параметров и индексируются размером выборки. Для закрепления основных идей рассмотрим еще раз эмпирические моментные уравнения, которые определяют оценку методом инструментальных переменных для линейной или нелинейной регрессионной модели.

Пример 13.8. *Эмпирическое моментное уравнение для инструментальных переменных*

Для оценки методом инструментальных переменных в линейной или нелинейной регрессионной модели мы предполагаем

$$E[\bar{\mathbf{m}}_n(\boldsymbol{\beta})] = E\left[\frac{1}{n} \sum_{i=1}^n \mathbf{z}_i [y_i - h(\mathbf{x}_i, \boldsymbol{\beta})]\right] = \mathbf{0}.$$

Всего L инструментальных переменных в \mathbf{z}_i и K параметров $\boldsymbol{\beta}$. Это утверждение определяет L моментных уравнений, по одному для каждой инструментальной переменной.

Мы делаем следующие предположения о модели и этих эмпирических моментах:

Предположение 13.1. *Сходимость эмпирических моментов*

Предполагается, что процесс, порождающий данные, удовлетворяет условиям, необходимым для применимости закона больших чисел, так что можно предположить, что эмпирические моменты сходятся по вероятности к своим математическим ожиданиям. В приложении D приведены различные законы больших чисел по возрастанию общности. Для этого предположения требуется следующее:

$$\bar{\mathbf{m}}_n(\boldsymbol{\theta}_0) = \frac{1}{n} \sum_{i=1}^n \mathbf{m}_i(\boldsymbol{\theta}_0) \xrightarrow{p} \mathbf{0}.$$

Законы больших чисел, которые мы рассматриваем в приложении D, относятся к случаю независимых наблюдений. Случаи зависимых, или коррелированных, наблюдений могут быть отнесены к **эргодической теореме** (20.1). В этом более общем случае мы предположили бы, что последовательность наблюдений $\mathbf{m}(\boldsymbol{\theta})$ представляет собой совместно $(L \times 1)$ стационарный и эргодический процесс.

Предполагается, что эмпирические моменты являются непрерывными и непрерывно дифференцируемыми функциями параметров. В нашем предыдущем примере это означало бы, что функция условного среднего $h(\mathbf{x}_i, \boldsymbol{\beta})$ — это непрерывная функция $\boldsymbol{\beta}$ (хотя необязательно \mathbf{x}_i). С непрерывностью и

дифференцируемостью мы также можем предположить, что производные моментов

$$\bar{G}_n(\theta_0) = \frac{\partial \bar{m}_n(\theta_0)}{\partial \theta_0'} = \frac{1}{n} \sum_{i=1}^n \frac{\partial m_{i,n}(\theta_0)}{\partial \theta_0'}$$

сходятся к пределу по вероятности, скажем $\text{plim } \bar{G}_n(\theta_0) = \bar{G}(\theta_0)$. [См. (13-1), (13-5) и теорему 13.1.] Для наборов *независимых* наблюдений непрерывность функций и производных позволит нам применить теорему Слуцкого для получения этого результата. Для более общего случая последовательностей *зависимых* наблюдений теорема 20.2 об эргодичности функций дает аналог теоремы Слуцкого для временных рядов. Подводя итог, если моменты следуют закону больших чисел, разумно предположить, что ему следуют и производные.

Предположение 13.2. Идентификация

Для любого $n \geq K$, если θ_1 и θ_2 — два различных вектора параметров, существует набор данных такой, что $\bar{m}_n(\theta_1) \neq \bar{m}_n(\theta_2)$. Формально в подразделе 12.5.3 идентификация определяется так, чтобы из нее следовало, что предел по вероятности целевой функции GMM единственным образом минимизировался при истинных значениях параметров θ_0 .

Предположение 13.2 — это практическая рекомендация в отношении идентификации. Более формальные условия обсуждаются в подразделе 12.5.3. Мы рассмотрели два нарушения этого важного предположения. В линейной регрессионной модели одним из предположений является полный ранг матрицы экзогенных переменных — отсутствие мультиколлинеарности в X . В ходе обсуждения оценки методом максимального правдоподобия мы встретимся со случаем (пример 14.1), в котором нормализация необходима для идентификации вектора параметров. [См. обсуждение этого случая в Hansen et al. (1996).] Оба этих случая включены в данное предположение. Условие идентификации имеет три важных следствия:

1. **Условие порядка.** Число моментных условий не меньше числа параметров $L \geq K$. Это необходимое, но недостаточное условие идентификации.
2. **Условие ранга.** $L \times K$ матрица производных, $\bar{G}_n(\theta_0)$ имеет ранг по строкам, равный K . (Снова отметим, что число строк должно быть равно числу столбцов или превышать его.)
3. **Единственность.** В предположении непрерывности из предположения идентификации следует, что вектор параметров, удовлетворяющий теоретическим моментным условиям, единственный.

Мы знаем, что в истинном значении вектора параметров $\text{plim } \bar{m}_n(\theta_0) = 0$. Если θ_1 — любой вектор параметров, удовлетворяющий этому условию, то θ_1 должен быть равен θ_0 .

Предположения 13.1 и 13.2 характеризуют параметризацию модели. Вместе они означают, что вектор параметров можно будет оценить. Теперь мы делаем статистическое предположение, которое позволит нам установить свойства GMM-оценки.

Предположение 13.3. *Асимптотическое распределение эмпирических моментов*

Мы предполагаем, что эмпирические моменты следуют центральной предельной теореме. Это предполагает, что моменты имеют конечную асимптотическую ковариационную матрицу $(1/n)\Phi$, так что

$$\sqrt{n} \bar{\mathbf{m}}_n(\boldsymbol{\theta}_0) \xrightarrow{d} N[\mathbf{0}, \Phi].$$

Требования к данным, необходимые для выполнения этого предположения, будут различаться и могут быть достаточно замысловатыми, если наблюдения, составляющие эмпирический момент, не являются независимыми. Для выборок независимых наблюдений мы предполагаем, что условия, лежащие в основе центральных предельных теорем Линдберга–Феллера (D.19) или Ляпунова (D.20), будут достаточными. Для более общего случая опять же необходимо делать некоторые предположения относительно данных. Мы предположили, что

$$E[\mathbf{m}_i(\boldsymbol{\theta}_0)] = \mathbf{0}.$$

Если мы можем сделать еще один шаг и предположить, что функции $\mathbf{m}_i(\boldsymbol{\theta}_0)$ являются эргодичной, стационарной **мартингальной разностью**

$$E[\mathbf{m}_i(\boldsymbol{\theta}_0) | \mathbf{m}_{i-1}(\boldsymbol{\theta}_0), \mathbf{m}_{i-2}(\boldsymbol{\theta}_0) \dots] = \mathbf{0},$$

мы сможем применить теорему 20.3 — центральную предельную теорему для мартингальных разностей. В общем случае проверить это предположение для нелинейных моделей будет непросто, так что обычно его выполнение будет просто предполагаться. С другой стороны, в типичном приложении эти предположения, скорее всего, будут не очень существенными. Для регрессионных моделей предположение принимает форму

$$E[\mathbf{z}_i \varepsilon_i | \mathbf{z}_{i-1} \varepsilon_{i-1}, \dots] = \mathbf{0},$$

что часто будет представлять центральную структуру модели.

С этими предположениями мы имеем

Теорема 13.2. *Асимптотическое распределение GMM-оценки*
При изложенных выше предположениях

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{GMM} &\xrightarrow{p} \boldsymbol{\theta}_0, \\ \hat{\boldsymbol{\theta}}_{GMM} &\overset{a}{\sim} N[\boldsymbol{\theta}_0, \mathbf{V}_{GMM}], \end{aligned} \quad (13-6)$$

где \mathbf{V}_{GMM} определена в (13-5).

Теперь приведем краткое доказательство теоремы 13.2. GMM-оценка получается минимизацией целевой функции

$$q_n(\theta) = \bar{m}_n(\theta)' W_n \bar{m}_n(\theta),$$

где W_n — это использованная матрица весов. Состоятельность оценки, минимизирующей этот критерий, может быть установлена с помощью той же логики, что использовалась для оценки методом максимального правдоподобия. Сначала нужно установить, что $q_n(\theta)$ сходится к значению $q_0(\theta)$. По нашим предположениям о строгой непрерывности и предположению 13.1 $q_n(\theta_0)$ сходится к 0. (Для получения этого результата следует применить теорему Слуцкого.) Мы предположим, что $q_n(\theta)$ сходится к $q_0(\theta)$, и для других точек в пространстве параметров. Поскольку W_n положительно определена для любого конечного n , мы знаем, что

$$0 \leq q_n(\hat{\theta}_{GMM}) \leq q_n(\theta_0). \quad (13-7)$$

То есть в конечных выборках $\hat{\theta}_{GMM}$ действительно минимизирует функцию, так что выборочное значение целевой функции не больше в $\hat{\theta}_{GMM}$, чем в любом другом значении, включая истинное значение параметра. Но в истинном значении $q_n(\theta_0) \xrightarrow{p} 0$. Таким образом, из выполнения (13-7) должно следовать, что и $q_n(\hat{\theta}_{GMM}) \xrightarrow{p} 0$ ввиду предположения об идентификации 13.2. Когда $n \rightarrow \infty$, $q_n(\hat{\theta}_{GMM})$ и $q_n(\theta)$ сходятся к одному пределу. Тогда должно быть верно, что при $n \rightarrow \infty$ $\bar{m}_n(2\hat{\theta}_{GMM}) \rightarrow \bar{m}_n(\theta_0)$, поскольку функция квадратична и W положительно определена. Условие идентификации, которое мы предположили ранее, теперь обеспечивает, что, когда $n \rightarrow \infty$, $\hat{\theta}_{GMM}$ должен быть равен θ_0 . Из этого следует состоятельность оценки.

Теперь мы изложим краткое доказательство асимптотической нормальности оценки: для GMM-оценки условия первого порядка имеют вид

$$\frac{\partial q_n(\hat{\theta}_{GMM})}{\partial \hat{\theta}_{GMM}} = 2\bar{G}_n(\hat{\theta}_{GMM})' W_n \bar{m}_n(\hat{\theta}_{GMM}) = 0. \quad (13-8)$$

(2 в начале не оказывает влияния на решение, так что с этого момента будет опускаться.) Уравнения ортогональности предполагаются непрерывными и непрерывно дифференцируемыми. Это позволяет нам применить **теорему о среднем значении**, когда мы разложим эмпирические моменты в линейный ряд Тейлора около истинного значения θ_0 :

$$\bar{m}_n(2\hat{\theta}_{GMM}) = \bar{m}_n(\theta_0) + \bar{G}_n(\bar{\theta})(\hat{\theta}_{GMM} - \theta_0), \quad (13-9)$$

где $\bar{\theta}$ — точка между $\hat{\theta}_{GMM}$ и истинными параметрами θ_0 . Таким образом, для каждого элемента $\bar{\theta}_k = w_k \hat{\theta}_{k,GMM} + (1-w_k)\theta_{0,k}$ для некоторого w_k такого, что $0 < w_k < 1$. Подставим (13-9) в (13-8), чтобы получить

$$\bar{G}_n(\hat{\theta}_{GMM})' W_n \bar{m}_n(\theta_0) + \bar{G}_n(\hat{\theta}_{GMM})' W_n \bar{G}_n(\bar{\theta})(\hat{\theta}_{GMM} - \theta_0) = 0.$$

Решим это уравнение относительно ошибки оценивания и умножим на \sqrt{n} . Таким образом, получаем

$$\sqrt{n}(\hat{\theta}_{GMM} - \theta_0) = -[\bar{G}_n(\hat{\theta}_{GMM})' W_n \bar{G}_n(\bar{\theta})]^{-1} \bar{G}_n(2\hat{\theta}_{GMM})' W_n \sqrt{n} \bar{m}_n(\theta_0).$$

Предполагая, что они у них есть, значения в левой и правой частях имеют одинаковые предельные распределения. Из состоятельности $\hat{\theta}_{GMM}$ мы знаем $\hat{\theta}_{GMM}$, и $\bar{\theta}$ оба сойдутся к θ_0 . По предположению строгой непрерывности также должно выполняться

$$\bar{G}_n(\bar{\theta}) \xrightarrow{p} \bar{G}(\theta_0) \text{ и } \bar{G}_n(\hat{\theta}_{GMM}) \xrightarrow{p} \bar{G}(\theta_0).$$

Мы также предположили, что взвешивающая матрица W_n сходится к матрице констант W . Приводя подобные слагаемые, мы получаем, что предельное распределение вектора в левой части должно совпадать с предельным распределением вектора в правой части (13-10):

$$\sqrt{n}(\hat{\theta}_{GMM} - \theta_0) \xrightarrow{d} \{-[\bar{G}(\theta_0)'W\bar{G}(\theta_0)]^{-1}\bar{G}(\theta_0)'W\}\sqrt{n}\bar{m}_n(\theta_0). \quad (13-10)$$

Теперь воспользуемся предположением 13.3. Матрица в фигурных скобках — это набор констант. Последний член сходится к нормальному распределению по предположению 13.3. Среднее и дисперсия этого предельного распределения — это ноль и Φ соответственно. Приводя подобные слагаемые, мы получаем утверждение теоремы 13.2, где

$$V_{GMM} = \frac{1}{n} [\bar{G}(\theta_0)'W\bar{G}(\theta_0)]^{-1} \bar{G}(\theta_0)'W\Phi W\bar{G}(\theta_0) \times \\ \times [\bar{G}(\theta_0)'W\bar{G}(\theta_0)]^{-1}. \quad (13-11)$$

Финальный результат — это функция выбора взвешивающей матрицы W . Если используется оптимальная взвешивающая матрица $W = \Phi^{-1}$, выражение принимает вид

$$V_{GMM, optimal} = \frac{1}{n} [\bar{G}(\theta_0)'\Phi^{-1}\bar{G}(\theta_0)]^{-1}. \quad (13-12)$$

Возвращаясь к (13-11), один случай представляет отдельный интерес. Если мы используем инструментальные переменные с $W = I$, то

$$V_{GMM} = \frac{1}{n} (\bar{G}'\bar{G})^{-1} \bar{G}'\Phi\bar{G}(\bar{G}'\bar{G})^{-1}.$$

Это уравнение представляет собой оценку Уайта или **оценку Ньюи–Веста**, которая возвращает нас к началу и иллюстрирует удобную симметрию принципа GMM. Мы формализуем его в подразделе 13.6.1.

13.5. Тестирование гипотез в контексте GMM

Подход к оцениванию, изложенный в предыдущем разделе, предоставляет основу для формирования удобного набора статистик при тестировании гипотез. Мы рассмотрим три группы тестов. Первая — это пара статистик, которые используются для тестирования обоснованности ограничений, из которых получаются моментные уравнения. Вторая — это трио тестов, соответствующих знакомым тестам Вальда, множителей Лагранжа (LM) и отношения правдоподобия (LR). Третья — это набор тестов, основанных на теоретических тонкостях условных моментов, которые мы использовали ранее для получения GMM-оценки.

13.5.1. Тестирование обоснованности моментных ограничений

В случаях точной идентификации, которые мы рассматривали ранее (МНК, инструментальные переменные, максимальное правдоподобие), целевая функция для GMM-оценивания,

$$q = \bar{m}(\theta)' W \bar{m}(\theta),$$

была бы в точности равна нулю, поскольку мы можем найти набор оценок, для которых $\bar{m}(\theta)$ в точности нуль. Таким образом, в случае точной идентификации, когда моментных уравнений столько же, сколько и оцениваемых параметров, взвешивающая матрица W не оказывает влияния на решение. Но если параметры избыточно идентифицированы моментными уравнениями, то эти уравнения подразумевают существенные ограничения. Таким образом, если гипотеза модели, из которой изначально были получены моментные условия, неверна, по крайней мере, некоторые выборочные моментные ограничения будут систематически нарушаться. Этот вывод лежит в основе теста на **сверхидентифицирующие ограничения**. По построению, когда используется оптимальная взвешивающая матрица,

$$nq = \left[\sqrt{n} \bar{m}(\hat{\theta})' \right] \left\{ \text{Est. Asy. Var}[\sqrt{n} \bar{m}(\hat{\theta})] \right\}^{-1} \left[\sqrt{n} \bar{m}(\hat{\theta}) \right],$$

так что nq — это статистика Вальда. Таким образом, при нулевой гипотезе соответствующей модели

$$nq \xrightarrow{d} \chi^2[L - K].$$

(Для случая точной идентификации степеней свободы ноль и $q = 0$.)

Пример 13.9. Сверхидентифицирующие ограничения

В модели потребления Холла два условия ортогональности, отмеченные в примере 13.1, точно идентифицируют два параметра. Но его анализ модели указывает на возможность протестировать спецификацию. Вывод: «Никакая информация, доступная в момент t , помимо уровня потребления c_t не помогает предсказать будущее потребление c_{t+1} в смысле влияния на ожидаемую ценность предельной полезности. В частности, доход или богатство в период t или ранее не важны, когда известно c_t » — предлагает подход к тестированию модели. Если лагированные значения дохода (Y_t может быть равен отношению текущего дохода к доходу предыдущего периода) добавляются к набору инструментов, модель становится переопределенной условиями ортогональности:

$$E_t \left[(\beta(1 + r_{t+1})R_{t+1}^\lambda - 1) \times \begin{pmatrix} 1 \\ R_t \\ Y_{t-1} \\ Y_{t-2} \end{pmatrix} \right] = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Простой тест на сверхидентифицирующие ограничения указал бы на обоснованность следствия. Отвержение ограничений бросает тень на

исходную модель. Тесты, предложенные Холлом для различения моделей жизненного цикла — перманентного дохода и других теорий потребления, включали добавление двух лагов дохода к информационному множеству. Hansen, Singleton (1982) работали непосредственно с этой формой модели. Другие авторы, например Campbell, Mankiw's (1989), так же как и Холл, использовали следствия из модели для формулирования более привычных регрессионных моделей инструментальных переменных.

Выше изложен **тест на спецификацию**, а не тест на параметрические ограничения. Однако между моментными ограничениями и ограничениями на параметры есть симметрия. Предположим, на θ накладывается J ограничений (линейных или нелинейных), которые сокращают число свободных параметров с K до $K - J$. (То есть ограничивают размерность пространства параметров с K до $K - J$.) Природа задачи GMM-оценивания, поставленной нами, совершенно не меняется этими ограничениями. Ограниченная задача может быть сформулирована в терминах

$$q_R = \bar{m}(\theta_R)' W \bar{m}(\theta_R).$$

Отметьте, что взвешивающая матрица W не меняется. Конкретная природа метода решения может измениться — ограничения требуют оптимизации с ограничениями. Однако целевая функция в сущности не меняется. Тогда из этого следует, что

$$nq_R \xrightarrow{d} \chi^2[L - (K - J)].$$

Этот результат указывает на способ тестирования ограничений, однако распределение такого теста неочевидно. Взвешенная сумма квадратов с установленными ограничениями nq_R должна быть больше, чем взвешенная сумма квадратов, полученная без ограничений nq . Разность равна

$$(nq_R - nq) \xrightarrow{d} \chi^2[J]. \quad (13-13)$$

Авторство теста приписывается Ньюи и Весту (Newey, West (1987b)). Таким может быть один подход к тестированию набора ограничений. (Свойствам этого теста в малых выборках будет уделено основное внимание в приложении, обсуждаемом в подразделе 13.6.5.) Теперь рассмотрим несколько альтернатив.

13.5.2. GMM-аналоги для тестов Вальда, множителей Лагранжа и отношения правдоподобия

В разделе 14.6 описывается тройка тестовых процедур, которые могут быть применены для проверки гипотезы в контексте оценивания методом максимального правдоподобия. Пусть тестируемая гипотеза — это набор J возможно нелинейных ограничений, накладываемых на K параметров θ в форме $H_0: r(\theta) = 0$. Пусть s_1 будет оценкой методом максимального правдоподобия θ , полученной без ограничений, и пусть s_0 обозначает оценки методом максимального правдоподобия с ограничениями, т.е. оценки, полученные в предположении выполнения нулевой гипотезы. Три статистики,

которые оказываются асимптотически эквивалентными, получаются следующим образом:

$$LR = \text{отношение правдоподобия} = -2(\ln L_0 - \ln L_1),$$

где

$$\ln L_j = \log \text{ функция правдоподобия в точке } c_j, \quad j = 0, 1.$$

Статистика отношения правдоподобия требует получения обеих оценок. Статистика Вальда — это

$$W = \text{Wald} = [\mathbf{r}(c_1)]' \{ \text{Est. Asy. Var}[\mathbf{r}(c_1)] \}^{-1} [\mathbf{r}(c_1)]. \quad (13-14)$$

Статистика Вальда — это мера расстояния для степени невыполнения ограничения моделью без ограничений. Обычной оценкой для асимптотической ковариационной матрицы будет

$$\text{Est. Asy. Var}[\mathbf{r}(c_1)] = \mathbf{R}_1 \{ \text{Est. Asy. Var}[c_1] \} \mathbf{R}_1', \quad (13-15)$$

где

$$\mathbf{R}_1 = \partial \mathbf{r}(c_1) / \partial c_1' \quad (\mathbf{R}_1 \text{ является } J \times K \text{ матрицей}).$$

Статистика Вальда может быть получена с использованием только оценки без ограничений. Статистика LM — это

$$LM = \text{Lagrange multiplier} = \mathbf{g}_1'(c_0) \{ \text{Est. Asy. Var}[\mathbf{g}_1(c_0)] \}^{-1} \mathbf{g}_1(c_0), \quad (13-16)$$

где $\mathbf{g}_1(c_0) = \partial \ln L_1(c_0) / \partial c_0$, т.е. первые производные неограниченного логарифма функции правдоподобия в точках, соответствующих оценкам с ограничениями. Член $\text{Est. Asy. Var}[\mathbf{g}_1(c_0)]$ — это матрица, обратная любой из обычных оценок асимптотической ковариационной матрицы оценок максимального правдоподобия, оцененной с использованием оценки с ограничениями. Наиболее удобный выбор обычно это оценка ВНН. Статистика LM основана на ограниченных оценках.

Ньюи и Вест (Newey, West (1987b)) получили аналоги этих тестовых статистик для GMM-оценки. Статистика Вальда оценивается идентично с использованием результатов GMM-оценивания, а не максимального правдоподобия¹⁰, т.е. в (13-14) мы использовали бы GMM-оценку без ограничений θ . Уместная асимптотическая ковариационная матрица — это (13-12). Процедура получения оценки в точности та же. Аналог статистики LM — это разность в значениях nq в (13-13). Необходимо использовать одну и ту же взвешивающую матрицу \mathbf{W} и для оценки с ограничениями, и для оценки без ограничений. Поскольку оценка без ограничений состоятельна и при H_0 , и при H_1 , состоятельная оценка без ограничений θ используется для получения \mathbf{W} . Обозначим ее $\Phi_1^{-1} = \{ \text{Asy. Var}[\sqrt{n} \bar{m}_1(c_1)] \}^{-1}$. Индекс 1 указывает на оценку без ограничений. Тогда q минимизируется без ограничений для получения q_1 и затем при условии ограничений для получения q_0 . Тогда статистика равна $(nq_0 - nq_1)$ ¹¹. Поскольку мы используем одну и ту же матрицу

¹⁰ Некоторые результаты для этой процедуры с малыми выборками см. в Burnside, Eichenbaum (1996). Newey, McFadden (1994) показали асимптотическую эквивалентность трех процедур.

¹¹ Ньюи и Вест обозначают это как D -тест.

W в обоих случаях, эта статистика обязательно неотрицательна. (Эта статистика обсуждается в подразделе 13.5.1.)

Наконец, аналогом LM-статистики будет

$$LM_{GMM} = n[\bar{\mathbf{m}}_1(\mathbf{c}_0)' \hat{\Phi}_1^{-1} \bar{\mathbf{G}}_1(\mathbf{c}_0)] [\bar{\mathbf{G}}_1(\mathbf{c}_0)' \hat{\Phi}_1^{-1} \bar{\mathbf{G}}_1(\mathbf{c}_0)]^{-1} [\bar{\mathbf{G}}_1(\mathbf{c}_0)' \hat{\Phi}_1^{-1} \bar{\mathbf{m}}_1(\mathbf{c}_0)].$$

Логика построения этой LM-статистики такая же, что и в случае MLE. Производные минимизирующего критерия q в (13-3) в точке, соответствующей оценке с ограничениями:

$$\mathbf{g}_1(\mathbf{c}_0) = \frac{\partial q}{\partial \mathbf{c}_0} = 2\bar{\mathbf{G}}_1(\mathbf{c}_0)' \hat{\Phi}_1^{-1} \bar{\mathbf{m}}(\mathbf{c}_{t0}).$$

LM-статистика, LM_{GMM} , — это статистика Вальда для тестирования гипотезы о том, что вектор равняется нулю при выполнении ограничений нулевой гипотезы. Из наших предыдущих результатов следует

$$\text{Est. Asy. Var}[\mathbf{g}_1(\mathbf{c}_0)] = \frac{4}{n} \bar{\mathbf{G}}_1(\mathbf{c}_0)' \hat{\Phi}_1^{-1} \{ \text{Est. Asy. Var}[\sqrt{n} \bar{\mathbf{m}}(\mathbf{c}_0)] \} \hat{\Phi}_1^{-1} \bar{\mathbf{G}}_1(\mathbf{c}_0).$$

Оценка асимптотической дисперсии $\sqrt{n} \bar{\mathbf{m}}(\mathbf{c}_{t0})$ равна $\hat{\Phi}_1$, так что

$$\text{Est. Asy. Var}[\mathbf{g}_1(\mathbf{c}_0)] = \frac{4}{n} \bar{\mathbf{G}}_1(\mathbf{c}_0)' \hat{\Phi}_1^{-1} \bar{\mathbf{G}}_1(\mathbf{c}_0).$$

Статистика Вальда примет вид

$$\begin{aligned} \text{Wald} &= \mathbf{g}_1(\mathbf{c}_0)' \{ \text{Est. Asy. Var}[\mathbf{g}_1(\mathbf{c}_0)] \}^{-1} \mathbf{g}_1(\mathbf{c}_0) \\ &= n \bar{\mathbf{m}}_1'(\mathbf{c}_0) \hat{\Phi}_1^{-1} \bar{\mathbf{G}}_1(\mathbf{c}_0) \{ \bar{\mathbf{G}}_1(\mathbf{c}_0)' \hat{\Phi}_1^{-1} \bar{\mathbf{G}}_1(\mathbf{c}_0) \}^{-1} \bar{\mathbf{G}}_1(\mathbf{c}_0)' \hat{\Phi}_1^{-1} \bar{\mathbf{m}}_1(\mathbf{c}_{t0}). \end{aligned} \quad (13-17)$$

13.6. GMM-оценка эконометрических моделей

Выше мы увидели, что GMM-оценка включает большую часть оценок, которые встречаются в этой книге. Параллельно мы неявным образом рассмотрели метод наименьших квадратов и общий подход к инструментальным оценкам. В этом разделе мы более конкретно формализуем GMM-оценки для некоторых оценок из предыдущих глав. В подразделе 13.6.1 рассматривается обобщенная регрессионная модель из главы 9. В подразделе 13.6.2 описывается довольно незначительное расширение GMM/IV-оценки до случая нелинейной регрессии. В подразделах 13.6.3 и 13.6.4 описываются GMM-оценки для наших моделей систем уравнений, внешне не связанных регрессий (SUR) и моделей одновременных уравнений. В последней, как мы поступили и в главе 10, мы рассмотрим оценки с ограниченной информацией (для одного уравнения) и с полной информацией (для нескольких уравнений). Наконец, в подразделе 13.6.5 мы представим одно из важных приложений GMM-оценок, оценку Ареллано–Бонда–Бовера для динамических моделей панельных данных.

13.6.1. Линейные модели для одного уравнения

Полезно обратить внимание на случай инструментальных переменных, так как он достаточно общий и мы легко можем применить его для более простых регрессионных моделей, если это будет уместно. Таким образом, мы отталкиваемся от обычной линейной модели (8-1), но больше не требуем, чтобы $E[\varepsilon_i | \mathbf{x}_i] = 0$. Вместо этого мы используем модель инструментальных переменных из главы 8, т.е. наша модель имеет вид

$$\begin{aligned} y_i &= \mathbf{x}'_i \boldsymbol{\beta} + \varepsilon_i \\ E[\mathbf{z}_i \varepsilon_i] &= \mathbf{0} \end{aligned}$$

для K переменных в \mathbf{x}_i и для некоторого набора из L инструментальных переменных \mathbf{z}_i , где $L \geq K$. Предыдущий случай обобщенной регрессионной модели возникает, если $\mathbf{z}_i = \mathbf{x}_i$, и классическая регрессионная форма появляется, если мы также добавим $\boldsymbol{\Omega} = \mathbf{I}$, так что это удобная, достаточно общая формулировка.

В главе 9, посвященной обобщенному методу наименьших квадратов, мы рассматривали два случая: первый с известной $\boldsymbol{\Omega}$, затем еще один — с неизвестной $\boldsymbol{\Omega}$, которая должна быть оценена. При оценивании обобщенным методом моментов ни один из этих случаев не подходит, поскольку мы работаем с гораздо меньшим количеством (предположенной) информации о процессе, порождающем данные. Мы рассмотрим три случая:

- классическая регрессия: $\text{Var}[\varepsilon_i | \mathbf{X}, \mathbf{Z}] = \sigma^2$,
- гетероскедастичность: $\text{Var}[\varepsilon_i | \mathbf{X}, \mathbf{Z}] = \sigma_i^2$,
- обобщенная модель: $\text{Cov}[\varepsilon_t, \varepsilon_s | \mathbf{X}, \mathbf{Z}] = \sigma^2 \omega_{ts}$,

где \mathbf{Z} и \mathbf{X} — это $n \times L$ и $n \times K$ матрицы наблюдаемых данных. (Мы предполагаем, как это бывает в действительности, что наиболее общий случай будет применен к временным рядам. Отсюда изменение в индексах.) Никакое конкретное распределение ошибок, условное или безусловное, не предполагается.

Предположение $E[\mathbf{z}_i \varepsilon_i] = \mathbf{0}$ влечет следующее **условие ортогональности**:

$$\text{Cov}[\mathbf{z}_i, \varepsilon_i] = \mathbf{0}, \quad \text{или} \quad E[\mathbf{z}_i (y_i - \mathbf{x}'_i \boldsymbol{\beta})] = \mathbf{0}.$$

Суммируя слагаемые, мы получаем, что из этого следует **моментное уравнение** для генеральной совокупности

$$E \left[\frac{1}{n} \sum_{i=1}^n \mathbf{z}_i (y_i - \mathbf{x}'_i \boldsymbol{\beta}) \right] = E[\bar{\mathbf{m}}(\boldsymbol{\beta})] = \mathbf{0}. \quad (13-18)$$

Это соотношение указывает, как мы можем оценить $\boldsymbol{\beta}$. Действительно, отметьте, что если $\mathbf{z}_i = \mathbf{x}_i$, тогда это просто теоретический аналог необходимых условий метода наименьших квадратов. Предположим, мы заменяем это теоретическое математическое ожидание на его выборочный аналог и используем его в основе нашей оценки. То есть если теоретическое соотношение выполняется для истинного вектора параметров $\boldsymbol{\beta}$, предположим,

что мы попытаемся воспроизвести этот результат с выборочным аналогом или эмпирическим моментным уравнением

$$\left[\frac{1}{n} \sum_{i=1}^n \mathbf{z}_i (y_i - \mathbf{x}'_i \hat{\beta}) \right] = \left[\frac{1}{n} \sum_{i=1}^n \mathbf{m}_i(\hat{\beta}) \right] = \bar{\mathbf{m}}(\hat{\beta}) = \mathbf{0}. \quad (13-19)$$

В отсутствие другой информации о процессе, порождающем данные, мы можем использовать эмпирическое моментное уравнение в основе нашей стратегии оценивания. Эмпирическое моментное условие — это L уравнений (число переменных в \mathbf{Z}) в K неизвестных (число параметров, которые мы оцениваем). Необходимо рассмотреть три возможности:

1. Недоопределенность. $L < K$. Если моментных уравнений меньше, чем параметров, найти решение системы в (13-19) будет невозможно. В отсутствие другой информации, такой как ограничения, которые сократили бы число свободных параметров, продолжать анализ этого случая не требуется. Для идентифицированных случаев удобно записывать (13-19) как

$$\bar{\mathbf{m}}(\hat{\beta}) = \left(\frac{1}{n} \mathbf{Z}' \mathbf{y} \right) - \left(\frac{1}{n} \mathbf{Z}' \mathbf{X} \right) \hat{\beta}. \quad (13-20)$$

2. Точная идентификация. Если $L = K$, то легко можно показать (это остается в качестве упражнения), что единственным решением нашей системы уравнений является уже знакомая оценка методом инструментальных переменных из подраздела 8.3.2

$$\hat{\beta} = (\mathbf{Z}' \mathbf{X})^{-1} \mathbf{Z}' \mathbf{y}. \quad (13-21)$$

3. Переопределенность. Если $L > K$, то не существует единственного решения системы уравнений $\bar{\mathbf{m}}(\hat{\beta}) = \mathbf{0}$. В таком случае нам нужно сформулировать стратегию для выбора оценки. Интуитивно понятным вариантом, хорошо работавшим до сих пор, была бы «минимизация квадратов». В таком случае это бы означало выбор оценки на основе целевой функции

$$\text{Min}_{\beta} q = \bar{\mathbf{m}}(\hat{\beta})' \bar{\mathbf{m}}(\hat{\beta}).$$

Нужно помнить, что мы сможем минимизировать ее в каком-то положительном значении; точного решения для (13-19) в случае переопределенности не существует. Кроме того, вы можете убедиться, что, если мы рассмотрим случай точной идентификации как случай переопределенности, т.е. все равно будем минимизировать сумму квадратов, мы все равно получим оценку инструментальных переменных, приведенную в (13-21) в качестве решения для случая (2). Для случая переопределенности необходимые условия имеют вид

$$\begin{aligned} \frac{\partial q}{\partial \beta} &= 2 \left(\frac{\partial \bar{\mathbf{m}}'(\hat{\beta})}{\partial \beta} \right) \bar{\mathbf{m}}(\hat{\beta}) = 2 \bar{\mathbf{G}}(\hat{\beta})' \bar{\mathbf{m}}(\hat{\beta}) \\ &= 2 \left(\frac{1}{n} \mathbf{X}' \mathbf{Z} \right) \left(\frac{1}{n} \mathbf{Z}' \mathbf{y} - \frac{1}{n} \mathbf{Z}' \mathbf{X} \hat{\beta} \right) = \mathbf{0}. \end{aligned} \quad (13-22)$$

Мы оставляем в качестве упражнения демонстрацию того, что решение в случаях (2) и (3) теперь

$$\hat{\beta} = [(\mathbf{X}' \mathbf{Z})(\mathbf{Z}' \mathbf{X})]^{-1} (\mathbf{X}' \mathbf{Z})(\mathbf{Z}' \mathbf{y}). \quad (13-23)$$

Оценка в (13-23) — это гибрид, который нам еще не встречался, однако, если $L = K$, он сводится к предыдущему из (13-21). (В случае переопределенности (13-21) — это не оценка методом инструментальных переменных, это, как мы и хотели, **оценка методом моментов**.)

Остается показать состоятельность и найти асимптотическое распределение и асимптотическую ковариационную матрицу для оценки. Необходимые нам промежуточные результаты приведены в предположениях 13.1–13.3 в подразделе 13.4.3:

- **Сходимость моментов.** Выборочные моменты сходятся по вероятности к своим теоретическим аналогам, т.е. $\bar{m}(\beta) \rightarrow 0$. В различных ситуациях сходимость будет различная, но в той или иной форме она потребуется. В простейших случаях, таких как модель с гетероскедастичностью, это будет сходимость в среднеквадратичном. Некоторые модели временных рядов, которые применяются к коррелированным наблюдениям, требуют какой-то другой формы сходимости. Но в любом из рассматриваемых нами случаев потребуется следующий общий результат: $\text{plim } \bar{m}(\beta) = 0$.
- **Идентификация.** Параметры идентифицируются моментными уравнениями. Идентификация, в сущности, означает, что большая выборка будет содержать информацию, достаточную для получения состоятельной оценки β при использовании выборочных моментов. Два условия должны быть выполнены — **условие порядка**, которое мы уже предположили ($L \geq K$), и **условие ранга**, которое означает, что моментные уравнения не являются избыточными. Из условия ранга следует условие порядка, так что нам требуется только формализовать это.
- **Условие идентификации для GMM-оценки.** $L \times K$ матрица

$$\Gamma(\beta) = E[\bar{G}(\beta)] = \text{plim } \bar{G}(\beta) = \text{plim } \frac{\partial \bar{m}}{\partial \beta'} = \text{plim } \frac{1}{n} \sum_{i=1}^n \frac{\partial m_i}{\partial \beta'}$$

должна иметь ранг по строкам, равный K ¹². Поскольку для этого требуется $L \geq K$, условие порядка является следствием этого утверждения. Данное предположение означает, что матрица производных сходится по вероятности к своему ожидаемому значению. Заметим, что мы также предположили, что производные, как и сами моменты, следуют закону больших чисел — они сходятся по вероятностям к своим ожидаемым значениям.

- **Предельное нормальное распределение для выборочных моментов.** Моменты генеральной совокупности следуют центральной предельной теореме или какому-то подобному утверждению. Поскольку мы изучаем обобщенную регрессионную модель, теорема Линдеберга–Леви (D.18) будет неприменима — наблюдения будут иметь различные дисперсии. Теорема Линдеберга–Феллера (D.19.A) достаточна для случая гетероскедастичности, но в общем случае нам потребуется что-то более общее. См. подраздел 13.4.3.

¹²Нам требуется, чтобы ранг по строкам был не меньше чем K . Могут существовать избыточные, т.е. функционально зависимые, моменты, только если существует по крайней мере K функционально независимых.

Из предположений 13.1–13.3 будет следовать (мы снова представляем это без доказательства), что GMM-оценки, которые мы получим, на самом деле состоятельны. По теореме Слуцкого мы можем перенести наши предельные результаты на эмпирические моментные уравнения.

Для получения асимптотической ковариационной матрицы мы просто воспользуемся общим результатом для GMM-оценок из подраздела 13.4.3, т.е.

$$\text{Asy. Var}[\hat{\beta}] = \frac{1}{n} [\Gamma' \Gamma]^{-1} \Gamma' \{ \text{Asy. Var}[\sqrt{n} \bar{m}(\beta)] \} \Gamma [\Gamma' \Gamma]^{-1}.$$

Для конкретной модели, которую мы рассматриваем,

$$\begin{aligned} \bar{m}(\beta) &= (1/n)(\mathbf{Z}'\mathbf{y} - \mathbf{Z}'\mathbf{X}\beta), \\ \bar{\mathbf{G}}(\beta) &= (1/n)\mathbf{Z}'\mathbf{X}, \\ \Gamma(\beta) &= \mathbf{Q}_{\mathbf{Z}\mathbf{X}} \text{ (см. подраздел 8.3.2)}. \end{aligned}$$

(Вам следует проверить, что размерности отдельных матриц и произведений действительно дают правильно сконфигурированную матрицу, которую мы ищем.) Оставшаяся тонкость, важная для рассматриваемой модели, состоит в том, чтобы показать

$$\mathbf{V} = \text{Asy. Var}[\sqrt{n} \bar{m}(\beta)].$$

Зная форму $\bar{m}(\beta)$,

$$\mathbf{V} = \frac{1}{n} \text{Var} \left[\sum_{i=1}^n \mathbf{z}_i \varepsilon_i \right] = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \sigma^2 \omega_{ij} \mathbf{z}_i \mathbf{z}_j' = \sigma^2 \frac{\mathbf{Z}' \Omega \mathbf{Z}}{n}$$

для наиболее общего случая. Заметим, что это точно выражение, возникающее в (9-6), так что уже вставивший в том контексте вопрос снова актуален. То есть при каких условиях эта матрица сойдется к постоянной матрице? Мы примем это обсуждение за данное. Единственной оставшейся деталью является оценивание этой матрицы. Ответ указан в подразделе 9.2.3, где мы искали ответ на этот же вопрос в связи с устойчивым оцениванием асимптотической ковариационной матрицы оценки методом наименьших квадратов. Таким образом, на данный момент мы изложили теоретическую основу оценки методом инструментальных переменных. Как было отмечено ранее, это относится только к оценке методом наименьших квадратов. Оценки \mathbf{V} для трех наших случаев будут:

- классическая регрессия:

$$\hat{\mathbf{V}} = \frac{(\mathbf{e}'\mathbf{e}/n)}{n} \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i' = \frac{(\mathbf{e}'\mathbf{e}/n)}{n} \mathbf{Z}'\mathbf{Z},$$

- регрессия с гетероскедастичностью:

$$\hat{\mathbf{V}} = \frac{1}{n} \sum_{i=1}^n e_i^2 \mathbf{z}_i \mathbf{z}_i', \quad (13-24)$$

- обобщенная регрессия:

$$\hat{V} = \frac{1}{n} \left[\sum_{t=1}^n e_t^2 \mathbf{z}_t \mathbf{z}_t' + \sum_{\ell=1}^p \left(1 - \frac{\ell}{(p+1)} \right) \sum_{t=\ell+1}^n e_t e_{t-\ell} (\mathbf{z}_t \mathbf{z}_{t-\ell}' + \mathbf{z}_{t-\ell} \mathbf{z}_t') \right].$$

Стоит отметить, что в каждом из этих случаев мы использовали некоторую информацию о структуре Ω . Если известно только то, что члены $\bar{\mathbf{m}}(\beta)$ не коррелированы, тогда доступна удобная оценка

$$\hat{V} = \frac{1}{n} \sum_{i=1}^n \mathbf{m}_i(\hat{\beta}) \mathbf{m}_i(\hat{\beta})',$$

которая является естественной эмпирической оценкой дисперсии. Заметим, что именно она использовалась в предыдущем случае при наличии гетероскедастичности.

Приводя подобные слагаемые, мы имеем

$$\begin{aligned} \text{Est. Asy. Var}[\hat{\beta}] &= \frac{1}{n} [\bar{\mathbf{G}}(\hat{\beta})' \bar{\mathbf{G}}(\hat{\beta})]^{-1} \bar{\mathbf{G}}(\hat{\beta})' \hat{V} \bar{\mathbf{G}}(\hat{\beta}) [\bar{\mathbf{G}}(\hat{\beta})' \bar{\mathbf{G}}(\hat{\beta})]^{-1} \\ &= n[(\mathbf{X}'\mathbf{Z})(\mathbf{Z}'\mathbf{X})]^{-1} (\mathbf{X}'\mathbf{Z}) \hat{V} (\mathbf{Z}'\mathbf{X}) [(\mathbf{X}'\mathbf{Z})(\mathbf{Z}'\mathbf{X})]^{-1}. \end{aligned} \quad (13-25)$$

Может показаться, что предыдущие рассуждения придают методу наименьших квадратов или методу моментов некоторую степень оптимальности, но это не так. Мы всего лишь придали им иную статистическую мотивацию (и установили состоятельность). Теперь мы рассмотрим вопрос о том, есть ли более эффективный способ использования данных, поскольку это обобщенная регрессионная модель.

Класс оценок методом наименьшего расстояния для этой модели определяется решениями целевой функции

$$\text{Min}_{\beta} q = \bar{\mathbf{m}}(\beta)' \mathbf{W} \bar{\mathbf{m}}(\beta),$$

где \mathbf{W} — любая положительно определенная взвешивающая матрица (или матрица весов). На основе только что сделанных предположений мы можем применить теорему 13.1 для того, чтобы получить

$$\text{Asy. Var} [\hat{\beta}_{MD}] = \frac{1}{n} [\bar{\mathbf{G}}' \mathbf{W} \bar{\mathbf{G}}]^{-1} \bar{\mathbf{G}}' \mathbf{W} \hat{V} \mathbf{W} \bar{\mathbf{G}} [\bar{\mathbf{G}}' \mathbf{W} \bar{\mathbf{G}}]^{-1}.$$

Заметим, что выше при анализе рассматривалась самая простая оценка методом наименьшего расстояния, у которой $\mathbf{W} = \mathbf{I}$. Теперь встает очевидный вопрос: если любая \mathbf{W} дает состоятельную оценку, есть ли преимущество у конкретной \mathbf{W} перед любой другой или это остается на усмотрение исследователя? Существует точный ответ, который требует рассмотрения двух случаев.

- **Случай точной идентификации.** Если $L = K$, т.е. если число моментных условий совпадает с числом параметров для оценивания, то \mathbf{W} не оказывает влияния на решение, так что исключительно для простоты оптимальная \mathbf{W} — это \mathbf{I} .

- **Случай переопределенности.** В таком случае «оптимальная» взвешивающая матрица, т.е. матрица \mathbf{W} , которая дает наиболее эффективную оценку, — это $\mathbf{W} = \mathbf{V}^{-1}$. Наилучшая взвешивающая матрица — это матрица, обратная к асимптотической ковариационной матрице вектора моментов. В таком случае MDE будет GMM-оценкой с

$$\hat{\beta}_{GMM} = [(\mathbf{X}'\mathbf{Z})\hat{\mathbf{V}}^{-1}(\mathbf{Z}'\mathbf{X})]^{-1}(\mathbf{X}'\mathbf{Z})\hat{\mathbf{V}}^{-1}(\mathbf{Z}'\mathbf{y})$$

и

$$\begin{aligned} \text{Asy. Var}[\hat{\beta}_{GMM}] &= \frac{1}{n}[\bar{\mathbf{G}}'\mathbf{V}^{-1}\bar{\mathbf{G}}]^{-1} \\ &= n[(\mathbf{X}'\mathbf{Z})\mathbf{V}^{-1}(\mathbf{Z}'\mathbf{X})]^{-1}. \end{aligned}$$

Мы подводим итог этому обсуждению, объединяя то, что может казаться оборванными рассуждениями. GMM-оценка получается как решение

$$\text{Min}_{\beta} q = \bar{\mathbf{m}}(\beta)' \{ \text{Asy. Var}[\sqrt{n} \bar{\mathbf{m}}(\beta)] \}^{-1} \bar{\mathbf{m}}(\beta),$$

что может указывать на то, что взвешивающая матрица — это функция того, что мы пытаемся оценить. Схема GMM-оценивания состоит из двух шагов: шага 1 — получение оценки \mathbf{V} ; шага 2 — использование матрицы, обратной к этой \mathbf{V} , в качестве взвешивающей матрицы при получении GMM-оценки. Так выглядит обычная процедура:

Шаг 1. Используйте $\mathbf{W} = \mathbf{I}$ для получения состоятельной оценки β . Затем оцените \mathbf{V} , где

$$\hat{\mathbf{V}} = \frac{1}{n} \sum_{i=1}^n e_i^2 \mathbf{z}_i \mathbf{z}_i'$$

в случае гетероскедастичности (т.е. оценку Уайта) или в более общем случае оценку Ньюи–Веста.

Шаг 2. Используйте $\mathbf{W} = \hat{\mathbf{V}}^{-1}$ для получения GMM-оценки.

К настоящему моменту наблюдательный читатель должен был обратить внимание на то, что в предыдущих рассуждениях мы не встречали оценку двухшаговым методом наименьших квадратов, описанным в подразделе 8.3.4. Для получения этой оценки мы должны вернуться к классическому случаю гомоскедастичных и неавтокоррелированных ошибок. В такой постановке взвешивающая матрица из теоремы 13.2 примет вид $\mathbf{W} = (\mathbf{Z}'\mathbf{Z})^{-1}$ и мы получим пропущенный результат.

GMM-оценка регрессионной модели с гетероскедастичностью получается из эмпирических моментных уравнений

$$\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i (y_i - \mathbf{x}_i' \hat{\beta}_{GMM}) = \frac{1}{n} \mathbf{X}' \hat{\boldsymbol{\varepsilon}}(\hat{\beta}_{GMM}) = \bar{\mathbf{m}}(\hat{\beta}_{GMM}) = \mathbf{0}. \quad (13-26)$$

Оценка получается минимизацией

$$q = \bar{\mathbf{m}}'(\hat{\beta}_{GMM}) \mathbf{W} \bar{\mathbf{m}}(\hat{\beta}_{GMM}),$$

где \mathbf{W} — положительно определенная взвешивающая матрица. Оптимальная взвешивающая матрица примет вид

$$\mathbf{W} = \{ \text{Asy. Var}[\sqrt{n} \bar{\mathbf{m}}(\beta)] \}^{-1},$$

которая является обратной к

$$\text{Asy. Var}[\sqrt{n}\bar{m}(\beta)] = \text{Asy. Var} \left[\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i \boldsymbol{\varepsilon}_i \right] = \text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sigma^2 \omega_i \mathbf{x}_i \mathbf{x}_i' = \sigma^2 \mathbf{Q}^*.$$

[См. подраздел 9.4.1.] Оптимальная взвешивающая матрица примет вид $[\sigma^2 \mathbf{Q}^*]^{-1}$. Но вспомните, что эта минимизационная задача встречается в случае точной идентификации, так что взвешивающая матрица не оказывает влияния на результат. Вы можете увидеть результат в моментном уравнении, которое представляет собой просто необходимые условия для метода наименьших квадратов. Мы можем получить точные решения моментных уравнений, так что взвешивающая матрица не требуется. *Независимо от ковариационной матрицы моментов GMM-оценка для регрессионной модели с гетероскедастичностью совпадает с обычной МНК-оценкой.* Мы можем использовать уже полученные результаты для получения ее асимптотической ковариационной матрицы. Соответствующая оценка – это оценка Уайта, приведенная в (9-27). [Опять же см. теорему 13.2.] Доступный на данный момент вывод состоит в том, что, пока мы не сделали конкретные предположения о дисперсиях, нам недоступны оценки более эффективные, чем метод наименьших квадратов, но мы должны модифицировать оценку асимптотической ковариационной матрицы.

13.6.2. Нелинейные модели регрессии

Предположим, что на основе теории получено соотношение

$$y_i = h(\mathbf{x}_i, \beta) + \varepsilon_i,$$

где β – это $K \times 1$ вектор оцениваемых параметров. Это может не быть регрессионным соотношением, поскольку возможен случай

$$\text{Cov}[\varepsilon_i, h(\mathbf{x}_i, \beta)] \neq 0$$

или даже

$$\text{Cov}[\varepsilon_i, \mathbf{x}_j] \neq 0 \text{ для всех } i \text{ и } j.$$

Рассмотрим, например, модель, содержащую лагированные зависимые переменные и автокоррелированные ошибки. (См. подраздел 20.9.3.) Предположим, что

$$E[\boldsymbol{\varepsilon} | \mathbf{X}] \neq \mathbf{0},$$

и

$$E[\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}' | \mathbf{X}] = \sigma^2 \boldsymbol{\Omega} = \boldsymbol{\Sigma},$$

где $\boldsymbol{\Sigma}$ – симметричная и положительно определенная матрица. Ошибки могут быть гетероскедастичными и/или автокоррелированными. Учитывая возможность корреляции между регрессорами и ошибками, это будет обобщенная, возможно, нелинейная регрессионная модель. Предположим, что для каждой реализации i мы наблюдаем L переменных \mathbf{z}_i таких, что \mathbf{z}_i не коррелированы с ε_i . Вы узнаете в \mathbf{z}_i набор **инструментальных переменных**. Из указанных предположений следует набор условий ортогональности

$$E[\mathbf{z}_i \varepsilon_i] = \mathbf{0},$$

которых может быть достаточно для точной идентификации (если $L = K$) или переопределенности (если $L > K$) параметров модели. (См. подраздел 8.3.4.)

Для удобства определим

$$\mathbf{e}(\mathbf{X}, \hat{\boldsymbol{\beta}}) = y_i - h(\mathbf{x}_i, \hat{\boldsymbol{\beta}}), \quad i = 1, \dots, n$$

и

$\mathbf{Z} = n \times L$ матрица, i -я строка которой – это \mathbf{z}'_i .

В качестве очевидного следствия наших предыдущих результатов мы получаем GMM-оценку $\boldsymbol{\beta}$. Выборочными моментами будут

$$\bar{\mathbf{m}}_n(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n \mathbf{z}_i e(\mathbf{x}_i, \boldsymbol{\beta}) = \frac{1}{n} \mathbf{Z}' \mathbf{e}(\mathbf{X}, \boldsymbol{\beta}).$$

Оценкой методом наименьшего расстояния будет $\hat{\boldsymbol{\beta}}$, минимизирующая

$$q = \bar{\mathbf{m}}_n(\hat{\boldsymbol{\beta}})' \mathbf{W} \bar{\mathbf{m}}_n(\hat{\boldsymbol{\beta}}) = \left(\frac{1}{n} [\mathbf{e}(\mathbf{X}, \hat{\boldsymbol{\beta}})' \mathbf{Z}] \right) \mathbf{W} \left(\frac{1}{n} [\mathbf{Z}' \mathbf{e}(\mathbf{X}, \hat{\boldsymbol{\beta}})] \right) \quad (13-27)$$

для некоторой матрицы \mathbf{W} , которую нам предстоит определить. Из целевой функции, приведенной ранее, получается **оценка нелинейным методом инструментальных переменных**. Если мы используем $\mathbf{W} = (\mathbf{Z}'\mathbf{Z})^{-1}$, то получаем в точности критерий оценивания, который мы использовали в разделе 8.6, когда определяли оценку нелинейным методом инструментальных переменных. Нетрудно заметить, что (13-27) соответствует более общему случаю, поскольку мы не ограничивали себя выбором этой \mathbf{W} . Для любого заданного выбора \mathbf{W} , если условий ортогональности достаточно для идентификации параметров, оценка путем минимизации q является, по крайней мере в принципе, стандартной задачей нелинейной оптимизации. Оптимальная матрица \mathbf{W} для этой оценки имеет вид

$$\begin{aligned} \mathbf{W}_{\text{GMM}} &= \left\{ \text{Asy. Var}[\sqrt{n} \bar{\mathbf{m}}_n(\boldsymbol{\beta})] \right\}^{-1} \\ &= \left\{ \text{Asy. Var} \left[\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{z}_i \varepsilon_i \right] \right\}^{-1} = \left\{ \text{Asy. Var} \left[\frac{1}{\sqrt{n}} \mathbf{Z}' \mathbf{e}(\mathbf{X}, \boldsymbol{\beta}) \right] \right\}^{-1}. \end{aligned} \quad (13-28)$$

Для нашей модели имеем

$$\mathbf{W} = \left[\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \text{Cov}[\mathbf{z}_i \varepsilon_i, \mathbf{z}_j \varepsilon_j] \right]^{-1} = \left[\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij} \mathbf{z}_i \mathbf{z}'_j \right]^{-1} = \left[\frac{\mathbf{Z}' \boldsymbol{\Sigma} \mathbf{Z}}{n} \right]^{-1}.$$

Если мы подставим этот результат в (13-27), то получим целевую функцию для GMM-оценивания

$$q = \left[\left(\frac{1}{n} \right) \mathbf{e}(\mathbf{X}, \hat{\boldsymbol{\beta}})' \mathbf{Z} \right] \left(\frac{\mathbf{Z}' \boldsymbol{\Sigma} \mathbf{Z}}{n} \right)^{-1} \left[\left(\frac{1}{n} \right) \mathbf{Z}' \mathbf{e}(\mathbf{X}, \hat{\boldsymbol{\beta}}) \right].$$

Одна тонкость может привести к трудностям. GMM-оценка включает

$$\begin{aligned} \frac{1}{n} \mathbf{Z}' \Sigma \mathbf{Z} &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \mathbf{z}_i \mathbf{z}_j' \text{Cov}[\varepsilon_i, \varepsilon_j] = \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n \mathbf{z}_i \mathbf{z}_j' \text{Cov}[(y_i - h(\mathbf{x}_i, \beta)), (y_j - h(\mathbf{x}_j, \beta))]. \end{aligned}$$

Условия, при которых такая двойная сумма может сойтись к положительно определенной матрице, приведены в подразделе 9.2.2. Если эти условия выполняются, процедура оценивания предполагает, что оценка β уже есть в распоряжении, несмотря на то что она является предметом оценивания. Возможна ситуация, в которой доступна состоятельная, но неэффективная оценка β . Предположим, что так и есть. Если наблюдения не коррелированы, произведения, включающие разные наблюдения, могут быть опущены, при этом требуется

$$\frac{1}{n} \mathbf{Z}' \Sigma \mathbf{Z} = \frac{1}{n} \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i' \text{Var}[(y_i - h(\mathbf{x}_i, \beta))].$$

Мы можем использовать аналог оценки Уайта (White (1980)), обсуждавшейся в подразделе 9.4.4, для этого случая:

$$\mathbf{S}_0 = \frac{1}{n} \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i' (y_i - h(\mathbf{x}_i, \hat{\beta}))^2. \quad (13-29)$$

Если наблюдения автокоррелированы, но процесс стационарен, то доступна оценка Ньюи–Веста (Newey, West (1987a)) (предполагая, что автокорреляции достаточно малы для разумного лага p):

$$\mathbf{S} = \left[\mathbf{S}_0 + \frac{1}{n} \sum_{\ell=1}^p w(\ell) \sum_{i=\ell+1}^n e_i e_{i-\ell} (\mathbf{z}_i \mathbf{z}_{i-\ell}' + \mathbf{z}_{i-\ell} \mathbf{z}_i') \right] = \sum_{\ell=0}^p w(\ell) \mathbf{S}_\ell, \quad (13-30)$$

где

$$w(\ell) = 1 - \frac{\ell}{p+1}.$$

Максимальная длина лага p должна определяться заранее. Мы потребуем, чтобы наблюдения, далеко отстоящие во времени, т.е. для которых $|i - \ell|$ велико, имели убывающие ковариации, чтобы получить результаты сходимости, обуславливающие МНК-, GLS-, а теперь и GMM-оценки. Выбор p отражает то, как далеко во времени нужно вернуться, чтобы рассматривать автокорреляцию незначительной для целей оценивания $(1/n) \mathbf{Z}' \Sigma \mathbf{Z}$. Практические исследования подсказывают, что можно использовать наименьшее целое число, превышающее $n^{1/4}$.

Остается открытым вопрос о том, как получить начальную состоятельную оценку. Одним вариантом является получение неэффективной, но состоятельной GMM-оценки, используя $\mathbf{W} = \mathbf{I}$ в (13-27), т.е. оценку нелинейным (или линейным, если уравнение линейное) методом инструментальных переменных. Эта оценка на первом шаге может быть использована для

построения \mathbf{W} , а затем для получения GMM-оценки. Другой возможностью является использование какой-либо отличной от GMM простой состоятельной оценки β .

Как только GMM-оценка получена, ее асимптотическая ковариационная матрица и асимптотическое распределение могут быть оценены исходя из теоремы 13.2. Вспомним, что

$$\bar{\mathbf{m}}_n(\beta) = \frac{1}{n} \sum_{i=1}^n \mathbf{z}_i \varepsilon_i -$$

это сумма $L \times 1$ векторов. Производная $\partial \bar{\mathbf{m}}_n(\beta) / \partial \beta'$ — это сумма $L \times K$ матриц, так что

$$\bar{\mathbf{G}}(\beta) = \partial \bar{\mathbf{m}}(\beta) / \partial \beta' = \frac{1}{n} \sum_{i=1}^n \mathbf{G}_i(\beta) = \frac{1}{n} \sum_{i=1}^n \mathbf{z}_i \left[\frac{\partial \varepsilon_i}{\partial \beta'} \right]. \quad (13-31)$$

В рассматриваемой нами модели

$$\frac{\partial \varepsilon_i}{\partial \beta'} = \frac{-\partial h(\mathbf{x}_i, \beta)}{\partial \beta'}.$$

Производные являются псевдорегрессорами в линеаризированной регрессионной модели, которую мы рассматривали в подразделе 7.2.3. Используя определенные там обозначения, получаем

$$\frac{\partial \varepsilon_i}{\partial \beta} = -\mathbf{x}_i^0,$$

таким образом,

$$\bar{\mathbf{G}}(\beta) = \frac{1}{n} \sum_{i=1}^n \mathbf{G}_i(\beta) = \frac{1}{n} \sum_{i=1}^n -\mathbf{z}_i \mathbf{x}_i^{0'} = -\frac{1}{n} \mathbf{Z}' \mathbf{X}^0. \quad (13-32)$$

Тогда оценка асимптотической ковариационной матрицы GMM-оценки имеет вид

$$\begin{aligned} \text{Est. Asy. Var}[\hat{\beta}] &= \left[\bar{\mathbf{G}}(\hat{\beta})' \left(\frac{1}{n} \mathbf{Z}' \hat{\Sigma} \mathbf{Z} \right)^{-1} \bar{\mathbf{G}}(\hat{\beta}) \right]^{-1} \\ &= [(\mathbf{X}^{0'} \mathbf{Z})(\mathbf{Z}' \hat{\Sigma} \mathbf{Z})^{-1} (\mathbf{Z}' \mathbf{X}^0)]^{-1}. \end{aligned} \quad (13-33)$$

(Два знака минус, $1/n^2$ и n^2 , в результате сокращаются.)

Если бы Σ , участвующая в (13-33), была равна $\sigma^2 \mathbf{I}$, то (13-33) была бы точно асимптотической ковариационной матрицей, которая появляется в теореме 8.1 для линейных моделей и теореме 8.2 для нелинейных моделей. Но между этой оценкой и оценками методом инструментальных переменных, обсуждавшимися ранее, есть интересное различие. В предыдущих случаях, когда инструментов было больше, чем параметров, переопределенность разрешалась путем выбора K инструментов, K проекций столбцов \mathbf{X} или \mathbf{X}^0 на пространство столбцов \mathbf{Z} . В данном случае, напротив, мы не будем

решать проблему переопределенности; мы просто используем все инструменты и минимизируем целевую функцию GMM. Теперь вы должны суметь показать, что, когда $\Sigma = \sigma^2 \mathbf{I}$ и мы используем эту информацию, будут получены те же самые оценки параметров. Но если мы будем использовать взвешивающую матрицу, отличную от $\mathbf{W} = (\mathbf{Z}'\mathbf{Z}/n)^{-1}$, оценки будут различаться.

13.6.3. Системы внешне не связанных уравнений

В разделе 10.4 мы рассматривали доступную FGLS-оценке систему уравнений

$$\begin{aligned} y_1 &= h_1(\mathbf{X}, \beta) + \varepsilon_1, \\ y_2 &= h_2(\mathbf{X}, \beta) + \varepsilon_2, \\ &\vdots \\ y_M &= h_M(\mathbf{X}, \beta) + \varepsilon_M. \end{aligned}$$

Все, что мы обсудили в разделе 10.4, актуально и для линейной модели. Однако никакая оценка не является состоятельной, если псевдорегрессоры \mathbf{x}_{tm}^0 или сами регрессоры \mathbf{x}_{tm} для линейной модели коррелированы с ошибками ε_{tm} . Предположим, мы допустим эту корреляцию и в отдельных уравнениях, и между ними. (Если на самом деле она отсутствует, то изложенная здесь GMM-оценка останется состоятельной.) Для простоты в этом разделе мы будем обозначать наблюдения индексом t и уравнения индексами i и j . Предположим также, что в наличии есть набор инструментальных переменных \mathbf{z}_t такой, что

$$E[\mathbf{z}_t \varepsilon_{tm}] = \mathbf{0}, t = 1, \dots, T \text{ и } m = 1, \dots, M. \quad (13-34)$$

(Мы могли бы рассмотреть отдельные наборы инструментов для каждого уравнения, но это усложнило бы изложение.) При этих предположениях оценки нелинейным FGLS и ML, приведенные ранее, будут несостоятельны. Но в данном случае может быть использовано небольшое расширение метода инструментальных переменных, приведенное для отдельного уравнения в разделе 8.4. Выборочный аналог (13-34) имеет вид

$$\frac{1}{T} \sum_{t=1}^T \mathbf{z}_t [y_{ti} - h_i(\mathbf{x}_t, \beta)] = \mathbf{0}, \quad i = 1, \dots, M.$$

Если мы используем этот результат для каждого отдельного уравнения в системе, то получим GMM-оценку, обсуждаемую в подразделе 13.6.2. Но в дополнение к потере эффективности за счет отсутствия возможности установить ограничения на β между уравнениями мы также упустим корреляцию между ошибками. Пусть

$$\frac{1}{T} \mathbf{Z}' \Omega_{ij} \mathbf{Z} = E \left[\frac{\mathbf{Z}' \varepsilon_i \varepsilon_j' \mathbf{Z}}{T} \right]. \quad (13-35)$$

Целевая функция для GMM-оценки в этой постановке имеет вид

$$\begin{aligned}
 q &= \sum_{i=1}^M \sum_{j=1}^M [(y_i - h_i(\mathbf{X}, \beta))' \mathbf{Z}/T] [\mathbf{Z}' \Omega_{ij} \mathbf{Z}/T]^{ij} [\mathbf{Z}' (y_j - h_j(\mathbf{X}, \beta))/T] \\
 &= \sum_{i=1}^M \sum_{j=1}^M [\varepsilon_i(\beta)' \mathbf{Z}/T] [\mathbf{Z}' \Omega_{ij} \mathbf{Z}/T]^{ij} [\mathbf{Z}' \varepsilon_j(\beta)/T],
 \end{aligned} \tag{13-36}$$

где $[\mathbf{Z}' \Omega_{ij} \mathbf{Z}/T]^{ij}$ обозначает ij -й блок обратной матрицы с ij -м блоком, равным $\mathbf{Z}' \Omega_{ij} \mathbf{Z}/T$. (Эта матрица приведена полностью в подразделе 13.6.4.)

GMM-оценка может быть получена в несколько этапов. Для оценки любого параметра дисперсии нам потребуется начальная состоятельная оценка β . Этот шаг может быть выполнен с помощью нелинейного метода инструментальных переменных, примененного к каждому отдельному уравнению — см. раздел 8.6, — хотя, если у уравнений будут общие параметры, должен быть сделан выбор, какие из них использовать. На следующем шаге знакомая процедура Уайта или Ньюи–Веста используется для оценки блок за блоком матрицы в (13-35). Поскольку она основана на состоятельной оценке β (по предположению), эта матрица не должна оцениваться заново. Теперь, имея этот результат, итеративное решение максимизационной задачи в (13-36) может быть найдено, например, с помощью методов, изложенных в приложении Е. Необходимые условия имеют вид

$$\frac{\partial q}{\partial \beta} = -2 \sum_{i=1}^M \sum_{j=1}^M [\mathbf{x}_i^0(\beta)' \mathbf{Z}/T] [\mathbf{Z}' \mathbf{W}_{ij} \mathbf{Z}/T]^{ij} [\mathbf{Z}' \varepsilon_j(\beta)/T] = 0. \tag{13-37}$$

Отметьте снова, что блоки обратной матрицы в центре получены из построенной большей матрицы *после обращения*. [Это краткое обсуждение может недооценить сложности оптимизационной задачи в (13-36), но это присуще самой процедуре.] Наконец, асимптотическая ковариационная матрица GMM-оценки получается из

$$\mathbf{V}_{\text{GMM}} = \frac{1}{T} \left[\sum_{i=1}^M \sum_{j=1}^M [\mathbf{x}_i^0(\beta)' \mathbf{Z}/T] [\mathbf{Z}' \mathbf{W}_{ij} \mathbf{Z}/T]^{ij} [\mathbf{Z}' \mathbf{x}_j^0(\beta)/T] \right]^{-1}.$$

13.6.4. Модели одновременных уравнений с гетероскедастичностью

GMM-оценка из подраздела 13.6.1 — это с минимальными изменениями в обозначениях в точности набор процедур, который мы использовали в подразделах 10.6.4 и 10.6.5 для оценки уравнений в модели одновременных уравнений. Использование GMM-оценки, однако, позволит нам обобщить ковариационную структуру ошибок. Предположим, что

$$y_{tj} = \mathbf{z}'_{tj} \delta_j + \varepsilon_{tj}, \quad t = 1, \dots, T,$$

где $\mathbf{z}_{tj} = [\mathbf{Y}_{tj}, \mathbf{x}_{tj}]$. (Мы используем заглавную \mathbf{Y}_{tj} для обозначения L_j включенных эндогенных переменных. Отметьте также, что для сопоставимости

с главой 10 значения символов \mathbf{x} и \mathbf{z} тут меняются местами; теперь \mathbf{x} — это вектор экзогенных переменных.) Мы предположили, что ошибка ε_{tj} в j -м уравнении не гетероскедастична и не автокоррелирована. Теперь нет нужды накладывать это ограничение. Однако автокорреляция в контексте одновременных уравнений представляет значительную сложность. В данный момент мы рассмотрим исключительно случай гетероскедастичности.

Предположения модели дают условия ортогональности

$$E[\mathbf{x}_t \varepsilon_{tj}] = E[\mathbf{x}_t (y_{tj} - \mathbf{z}'_{tj} \delta_j)] = \mathbf{0}.$$

Если \mathbf{x}_t рассматривается как полный набор экзогенных переменных в модели, то мы получаем целевую функцию для GMM-оценки для j -го уравнения:

$$\begin{aligned} q &= \left[\frac{\mathbf{e}(\mathbf{z}_t, \delta_j)' \mathbf{X}}{T} \right] \mathbf{W}_{jj}^{-1} \left[\frac{\mathbf{X}' \mathbf{e}(\mathbf{z}_t, \delta_j)}{T} \right] \\ &= \bar{\mathbf{m}}(\delta_j)' \mathbf{W}_{jj}^{-1} \bar{\mathbf{m}}(\delta_j), \end{aligned}$$

где

$$\bar{\mathbf{m}}(\delta_j) = \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t (y_{tj} - \mathbf{z}'_{tj} \delta_j) \quad \text{и} \quad \mathbf{W}_{jj}^{-1} = \text{взвешивающая матрица GMM.}$$

Это снова в точности оценка, изложенная в подразделе 13.6.1. Если ошибки предполагаются гомоскедастичными и не автокоррелированными, то оптимальной взвешивающей матрицей будет оценка матрицы, обратной к

$$\begin{aligned} \mathbf{W}_{jj} &= \text{Asy. Var}[\sqrt{T} \bar{\mathbf{m}}(\delta_j)] \\ &= \text{plim} \left[\frac{1}{T} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}'_t (y_{tj} - \mathbf{z}'_{tj} \delta_j)^2 \right] \\ &= \text{plim} \frac{1}{T} \sum_{t=1}^T \sigma_{jj} \mathbf{x}_t \mathbf{x}'_t \\ &= \text{plim} \sigma_{jj} \left(\frac{\mathbf{X}' \mathbf{X}}{T} \right). \end{aligned}$$

Константа σ_{jj} не влияет на решение. Если мы используем $(\mathbf{X}' \mathbf{X})^{-1}$ в качестве взвешивающей матрицы, то GMM-оценкой, минимизирующей q , будет оценка 2SLS.

Мы можем расширить модель на случай гетероскедастичности в общей форме. Выводить предыдущий результат не требуется. Если ошибки гетероскедастичны, то

$$\mathbf{W}_{jj} = \text{plim} \frac{1}{T} \sum_{t=1}^T \omega_{jj,t} \mathbf{x}_t \mathbf{x}'_t = \text{plim} \frac{\mathbf{X}' \boldsymbol{\Omega}_{jj} \mathbf{X}}{T}.$$

Взвешивающая матрица может быть оценена с помощью оценки Уайта — см. (13-24), — если доступна состоятельная оценка δ_j для получения остатков. И она доступна, поскольку 2SLS, игнорирующая гетероскедастичность,

состоятельна, хотя и неэффективна. Таким образом, вывод состоит в том, что в этих предположениях можно улучшить 2SLS, добавив дополнительный шаг. Для оценок такого вида зарезервировано обозначение 3SLS (3МНК). Выбирая между 2,5-шаговым МНК и предложенным Дэвидсоном и Маккинном «гетероскедастичным 2SLS», или H2SLS, мы предпочитаем последний. Оценка основана на исходной процедуре минимизации квадратов остатков. Таким образом,

$$\hat{\delta}_{j,H2SLS} = [\mathbf{Z}'_j \mathbf{X}(\mathbf{S}_{0,jj})^{-1} \mathbf{X}' \mathbf{Z}_j]^{-1} [\mathbf{Z}'_j \mathbf{X}(\mathbf{S}_{0,jj})^{-1} \mathbf{X}' \mathbf{y}_j],$$

где

$$\mathbf{S}_{0,jj} = \sum_{t=1}^T \mathbf{x}_t \mathbf{x}'_t (y_{tj} - \mathbf{z}'_{tj} \hat{\delta}_{j,2SLS})^2.$$

Асимптотическая ковариационная матрица оценивается с помощью

$$\text{Est.Asy.Var}[\hat{\delta}_{j,H2SLS}] = [\mathbf{Z}'_j \mathbf{X}(\mathbf{S}_{0,jj})^{-1} \mathbf{X}' \mathbf{Z}_j]^{-1}.$$

Расширения подобного рода были предложены в работах Cragg (1983) и Cummy, Huizinga, Obstfeld (1983).

GMM-оценка для системы уравнений рассмотрена в подразделе 13.6.3. Как и в случае с отдельным уравнением, незначительное изменение обозначений приводит к оценке для систем одновременных уравнений. Как и прежде, мы рассмотрим только случай с гетероскедастичностью в общей форме. Расширение на случай автокорреляции достаточно сложно. [См. Cummy, Huizinga, Obstfeld (1983).] Условия ортогональности, определенные в (13-34), имеют вид

$$E[\mathbf{x}_t \varepsilon_{tj}] = E[\mathbf{x}_t (y_{tj} - \mathbf{z}'_{tj} \delta_j)] = 0.$$

Если мы рассмотрим все уравнения совместно, то получим целевую функцию для оценки всех параметров модели:

$$\begin{aligned} q &= \sum_{j=1}^M \sum_{l=1}^M \left[\frac{\mathbf{e}(\mathbf{z}_t, \delta_j)' \mathbf{X}}{T} \right] [\mathbf{W}]^{jl} \left[\frac{\mathbf{X}' \mathbf{e}(\mathbf{z}_t, \delta_l)}{T} \right] \\ &= \sum_{j=1}^M \sum_{l=1}^M \bar{\mathbf{m}}(\delta_j)' [\mathbf{W}]^{jl} \bar{\mathbf{m}}(\delta_l), \end{aligned}$$

где

$$\bar{\mathbf{m}}(\delta_j) = \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t (y_{tj} - \mathbf{z}'_{tj} \delta_j),$$

и

$$[\mathbf{W}]^{jl} = \text{блок } jl \text{ взвешивающей матрицы } \mathbf{W}^{-1}.$$

Как и прежде, мы рассматриваем оптимальную взвешивающую матрицу, полученную как асимптотическую ковариационную матрицу эмпирических

моментов $\bar{m}(\delta_j)$. Эти моменты собраны в один вектор $\bar{m}(\delta)$. Затем jl -й блок $\text{Asy. Var}[\sqrt{T} \bar{m}(\delta)]$ равен

$$\Phi_{jl} = \text{plim} \left\{ \frac{1}{T} \sum_{t=1}^T [\mathbf{x}_t \mathbf{x}'_t (y_{tj} - \mathbf{z}'_{tj} \delta_j)(y_{tl} - \mathbf{z}'_{tl} \delta_l)] \right\} = \text{plim} \left(\frac{1}{T} \sum_{t=1}^T \omega_{jl,t} \mathbf{x}_t \mathbf{x}'_t \right).$$

Если ошибки гомоскедастичны, то получается $\Phi_{jl} = \sigma_{jl} [\text{plim}(\mathbf{X}'\mathbf{X}/T)]$. В противном случае мы получаем матрицу вида $\Phi_{jl} = \text{plim}[\mathbf{X}'\Omega_{jl}\mathbf{X}/T]$. Тогда, приводя подобные слагаемые, получаем целевую функцию для GMM-оценки:

$$q = \begin{bmatrix} [\mathbf{X}'(\mathbf{y}_1 - \mathbf{Z}_1 \delta_1)]/T \\ [\mathbf{X}'(\mathbf{y}_2 - \mathbf{Z}_2 \delta_2)]/T \\ \vdots \\ [\mathbf{X}'(\mathbf{y}_M - \mathbf{Z}_M \delta_M)]/T \end{bmatrix}' \begin{bmatrix} \Phi_{11} & \Phi_{12} & \cdots & \Phi_{1M} \\ \Phi_{21} & \Phi_{22} & \cdots & \Phi_{2M} \\ \vdots & \vdots & \cdots & \vdots \\ \Phi_{M1} & \Phi_{M2} & \cdots & \Phi_{MM} \end{bmatrix}^{-1} \begin{bmatrix} [\mathbf{X}'(\mathbf{y}_1 - \mathbf{Z}_1 \delta_1)]/T \\ [\mathbf{X}'(\mathbf{y}_2 - \mathbf{Z}_2 \delta_2)]/T \\ \vdots \\ [\mathbf{X}'(\mathbf{y}_M - \mathbf{Z}_M \delta_M)]/T \end{bmatrix}.$$

На практике Φ_{jl} может быть оценена с помощью

$$\hat{\Phi}_{jl} = \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}'_t (y_{tj} - \mathbf{z}'_{tj} \mathbf{d}_j)(y_{tl} - \mathbf{z}'_{tl} \mathbf{d}_l),$$

где \mathbf{d}_j — это состоятельная оценка δ_j . Естественным выбором будет двухшаговый метод наименьших квадратов. Для диагональных блоков стандартный выбор — это оценка Уайта. Для внедиагональных блоков это простое расширение. Имея этот результат, необходимыми условиями для GMM-оценки являются

$$\frac{\partial \hat{q}}{\partial \delta_j} = -2 \sum_{l=1}^M \left(\frac{\mathbf{Z}'_j \mathbf{X}}{T} \right) \hat{\Phi}^{jl} \left[\frac{\mathbf{X}'(\mathbf{y}_l - \mathbf{Z}_l \delta_l)}{T} \right],$$

где $\hat{\Phi}^{jl}$ — это jl -й блок в матрице, обратной оценке центральной матрицы в q .

Решение — это:

$$\begin{bmatrix} \hat{\delta}_{1,\text{GMM}} \\ \hat{\delta}_{2,\text{GMM}} \\ \vdots \\ \hat{\delta}_{M,\text{GMM}} \end{bmatrix} = \begin{bmatrix} \mathbf{Z}'_1 \mathbf{X} \hat{\Phi}^{11} \mathbf{X}' \mathbf{Z}_1 & \mathbf{Z}'_1 \mathbf{X} \hat{\Phi}^{12} \mathbf{X}' \mathbf{Z}_2 & \cdots & \mathbf{Z}'_1 \mathbf{X} \hat{\Phi}^{1M} \mathbf{X}' \mathbf{Z}_M \\ \mathbf{Z}'_2 \mathbf{X} \hat{\Phi}^{21} \mathbf{X}' \mathbf{Z}_1 & \mathbf{Z}'_2 \mathbf{X} \hat{\Phi}^{22} \mathbf{X}' \mathbf{Z}_2 & \cdots & \mathbf{Z}'_2 \mathbf{X} \hat{\Phi}^{2M} \mathbf{X}' \mathbf{Z}_M \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{Z}'_M \mathbf{X} \hat{\Phi}^{M1} \mathbf{X}' \mathbf{Z}_1 & \mathbf{Z}'_M \mathbf{X} \hat{\Phi}^{M2} \mathbf{X}' \mathbf{Z}_2 & \cdots & \mathbf{Z}'_M \mathbf{X} \hat{\Phi}^{MM} \mathbf{X}' \mathbf{Z}_M \end{bmatrix} \times$$

$$\times \begin{bmatrix} \sum_{j=1}^M \mathbf{Z}'_1 \mathbf{X} \hat{\Phi}^{1j} \mathbf{X}' \mathbf{y}_j \\ \sum_{j=1}^M \mathbf{Z}'_2 \mathbf{X} \hat{\Phi}^{2j} \mathbf{X}' \mathbf{y}_j \\ \vdots \\ \sum_{j=1}^M \mathbf{Z}'_M \mathbf{X} \hat{\Phi}^{Mj} \mathbf{X}' \mathbf{y}_j \end{bmatrix}^{-1}.$$

Асимптотическая ковариационная матрица для оценки получается с помощью матрицы в T раз большей, чем обратная матрица в скобках.

Некоторые уже рассмотренные оценки являются специальными случаями:

- если $\hat{\Phi}_{jj} = \hat{\sigma}_{jj}(\mathbf{X}'\mathbf{X}/T)$ и $\hat{\Phi}_{jl} = \mathbf{0}$ для $j \neq l$, то $\hat{\delta}_j$ — это 2SLS;
- если $\hat{\Phi}_{jl} = \mathbf{0}$ для $j \neq l$, то $\hat{\delta}_j$ — это H2SLS, оценка GMM для отдельного уравнения;
- если $\hat{\Phi}_{jl} = \hat{\sigma}_{jl}(\mathbf{X}'\mathbf{X}/T)$, то $\hat{\delta}_j$ — это 3SLS.

Как и прежде, GMM-оценка эффективна в случае гетероскедастичности ошибок. Если ошибки гомоскедастичны, она асимптотически эквивалентна 3SLS, [хотя в конечных выборках она будет численно отличаться, поскольку S_{jl} не будет идентична $\hat{\sigma}_{jl}(\mathbf{X}'\mathbf{X})$].

13.6.5. GMM-оценка динамических моделей панельных данных

Панельные данные хорошо подходят для рассмотрения динамических эффектов, как, например, в модели первого порядка:

$$\begin{aligned} y_{it} &= \mathbf{x}'_{it}\boldsymbol{\beta} + \delta y_{i,t-1} + c_i + \varepsilon_{it} \\ &= \mathbf{w}'_{it}\boldsymbol{\theta} + \alpha_i + \varepsilon_{it}, \end{aligned}$$

где набор переменных в правой части, \mathbf{w}_{it} , теперь включает запаздывающую зависимую переменную $y_{i,t-1}$. Включение таким образом динамики в модель существенным образом меняет интерпретацию уравнения. Без лагированной переменной «независимые переменные» отражали весь набор информации, который приводил к наблюдаемому исходу y_{it} . С запаздывающей переменной теперь у нас в уравнении вся история переменных из правой части, так что любой измеренный эффект обусловлен этой историей; в таком случае любое влияние \mathbf{x}_{it} отражает эффект *новой* информации. Значительные осложнения возникают при оценке такой модели. В случае и фиксированных, и случайных эффектов сложность состоит в том, что лагированная зависимая переменная коррелирована с ошибкой, даже если предполагается, что сама ε_{it} не автокоррелирована. Рассмотрим модель с фиксированными эффектами как обычную регрессию с лагированной зависимой переменной, которая коррелирована между наблюдениями. В такой динамической регрессионной модели оценка, основанная на T наблюдениях, смещена в конечных выборках, но состоятельна по T . Смещение в конечных выборках имеет порядок $1/T$. Тот же результат применим и здесь, но отличие состоит в том, что, если ранее мы получали асимптотические результаты, позволяя T расти, в этой постановке T предполагается малым и фиксированным и асимптотические результаты выводятся из роста n , а не T . Оценка $\boldsymbol{\theta} = [\boldsymbol{\beta}, \delta]$ с фиксированными эффектами может рассматриваться как среднее для n таких оценок. Теперь предположим, что $T \geq K + 1$, где

K — это набор переменных в \mathbf{x}_{it} . Тогда из (11-14)

$$\begin{aligned}\hat{\theta} &= \left[\sum_{i=1}^n \mathbf{W}_i' \mathbf{M}^0 \mathbf{W}_i \right]^{-1} \left[\sum_{i=1}^n \mathbf{W}_i' \mathbf{M}^0 \mathbf{y}_i \right] \\ &= \left[\sum_{i=1}^n \mathbf{W}_i' \mathbf{M}^0 \mathbf{W}_i \right]^{-1} \left[\sum_{i=1}^n \mathbf{W}_i' \mathbf{M}^0 \mathbf{W}_i \mathbf{d}_i \right] \\ &= \sum_{i=1}^n \mathbf{F}_i \mathbf{d}_i,\end{aligned}$$

где строки $T \times (K + 1)$ матрицы \mathbf{W}_i — это \mathbf{w}_{it}' и \mathbf{M}^0 — это $T \times T$ матрица, которая оценивает отклонения от групповых средних [см. (11-14)]. Каждая внутригрупповая оценка \mathbf{d}_i несостоятельна, так как она смещена в конечных выборках и ее дисперсия не стремится к нулю с ростом n . Эта матрица взвешенных средних n несостоятельных оценок также будет несостоятельной. (Этот анализ лишь эвристический. Если $T < K + 1$, то отдельные векторы коэффициентов не могут быть оценены¹³.)

Проблема более прозрачна в модели со случайными эффектами. В модели

$$y_{it} = \mathbf{x}_{it}' \boldsymbol{\beta} + \delta y_{i,t-1} + u_i + \varepsilon_{it}$$

лагированная зависимая переменная коррелирована с общей ошибкой модели, поскольку одна и та же u_i входит в уравнение для всех наблюдений в группе i .

Эти результаты не делают невозможной оценку модели, но они требуют применения других методов, нежели знакомые метод наименьших квадратов с фиктивными переменными и FGLS. Общий подход, который был развит в литературе в несколько этапов¹⁴, основан на оценках методом инструментальных переменных и в недавних работах Ареллано и Бонда (Arellano, Bond (1991)) и Ареллано и Бовера (Arellano, Bover (1995)) на GMM-оценке. Например, в случаях фиксированных или случайных эффектов неоднородность может быть убрана из модели взятием первых разностей, что приводит к

$$y_{it} - y_{i,t-1} = (\mathbf{x}_{it} - \mathbf{x}_{i,t-1})' \boldsymbol{\beta} + \delta (y_{i,t-1} - y_{i,t-2}) + (\varepsilon_{it} - \varepsilon_{i,t-1}).$$

Эта модель все еще усложняется корреляцией между запаздывающей зависимой переменной и ошибкой (и ее ошибкой в форме скользящего среднего первого порядка). Но без групповых эффектов доступна простая оценка методом инструментальных переменных. Предполагая, что временные ряды достаточно длинные, можно использовать лагированные разности ($y_{i,t-2} - y_{i,t-3}$) или лагированные уровни $y_{i,t-2}$ и $y_{i,t-3}$ как одну или две инструментальные переменные для ($y_{i,t-1} - y_{i,t-2}$). (Другие переменные могут служить

¹³ Более подробное обсуждение приведено в работах Nickell (1981), Ridder, Wansbeek (1990) и Kiviet (1995).

¹⁴ Модель в такой форме была впервые предложена Balestra, Nerlove (1966). См., например, Anderson, Hsiao (1981, 1982), Bhargava, Sargan (1983), Arellano (1989), Arellano, Bond (1991), Arellano, Bover (1995), Ahn, Schmidt (1995) и Nerlove (2003).

собственными инструментами.) Это оценка Андерсона и Хсяо, рассмотренная для этой модели в подразделе 11.8.2. При таком построении работа с моделью основана на стандартном применении метода инструментальных переменных, который был рассмотрен в разделе 11.8¹⁵. Этот пример иллюстрирует особенность метода инструментальных переменных. Но, как показали Ареллано и др. и Эн, Шмидт (Arellano et al. и Ahn, Schmidt (1995)), в выборке есть дополнительная информация, которая может быть использована в оценивании в контексте GMM-оценки, что будет рассмотрено ниже.

Мы расширим модель со случайными эффектами Хаусмана и Тейлора (НТ) в подразделе 11.8.1, включив лагированную зависимую переменную:

$$\begin{aligned} y_{it} &= \delta y_{i,t-1} + \mathbf{x}'_{1it}\beta_1 + \mathbf{x}'_{2it}\beta_2 + \mathbf{z}'_{1i}\alpha_1 + \mathbf{z}'_{2i}\alpha_2 + \varepsilon_{it} + u_i \\ &= \boldsymbol{\theta}'\mathbf{w}_{it} + \varepsilon_{it} + u_i \\ &= \boldsymbol{\theta}'\mathbf{w}_{it} + \eta_{it}, \end{aligned}$$

где $\mathbf{w}_{it} = [y_{i,t-1}, \mathbf{x}'_{1it}, \mathbf{x}'_{2it}, \mathbf{z}'_{1i}, \mathbf{z}'_{2i}]'$ — это теперь $(1 + K_1 + K_2 + L_1 + L_2) \times 1$ вектор. Слагаемые в уравнении те же, что и в модели Хаусмана и Тейлора. Оценка методом инструментальных переменных модели без лагированной зависимой переменной обсуждается в подразделе 11.8.1, посвященном оценке НТ. Кроме того, включая в \mathbf{x}_{2it} только $y_{i,t-1}$, мы видим, что подход НТ обобщается и на эту постановку, в сущности, без модификаций. Ареллано и др. предлагают GMM-оценку и показывают, что увеличение эффективности доступно при использовании большего числа моментных условий. В предыдущем изложении мы использовали GMM-оценку, построенную следующим образом: набор моментных условий, который мы использовали для формирования инструментальных переменных, имел вид

$$E \left[\begin{pmatrix} \mathbf{x}_{1it} \\ \mathbf{x}_{2it} \\ \mathbf{z}_{1i} \\ \bar{\mathbf{x}}_{1i} \end{pmatrix} (\eta_{it} - \bar{\eta}_i) \right] = E \left[\begin{pmatrix} \mathbf{x}_{1it} \\ \mathbf{x}_{2it} \\ \mathbf{z}_{1i} \\ \bar{\mathbf{x}}_{1i} \end{pmatrix} (\varepsilon_{it} - \bar{\varepsilon}_i) \right] = \mathbf{0}.$$

Это моментное условие использовалось для формирования оценки методом инструментальных переменных. Мы могли проигнорировать не скалярную дисперсию η_{it} и использовать простой метод инструментальных переменных. Однако, учитывая формулировку модели со случайными эффектами и используя аналог FGLS, мы получили в разделе 11.8 более эффективную оценку. Как обычно, это можно сделать в два шага. Неэффективная оценка вычисляется для получения остатков, требуемых для оценки компонент дисперсии. Это шаги 1 и 2 у Хаусмана и Тейлора. Шаги 3 и 4 состоят в получении GMM-оценки на основе этих компонент дисперсии.

Ареллано и др. утверждают, что описанная процедура не использует всей доступной информации из выборки. Выражаясь проще, среди T наблюде-

¹⁵Остается вопрос о том, следует ли использовать разности или уровни в качестве инструментов. Arellano (1989) и Kiviet (1995) приводят аргументы в пользу предпочтительности последних.

ний в группе i мы не использовали тот факт, что

$$E \left[\begin{pmatrix} \mathbf{x}_{1it} \\ \mathbf{x}_{2it} \\ \mathbf{z}_{1i} \\ \bar{\mathbf{x}}_{1i.} \end{pmatrix} (\eta_{is} - \bar{\eta}_i) \right] = \mathbf{0} \quad \text{для некоторых } s \neq t.$$

Таким образом, например, не только ошибки в период t не коррелированы с этими переменными в период t , но можно считать, что они не коррелированы с этими же переменными в периоды $t-1$, $t-2$, возможно, $t+1$ и т.д. В принципе число доступных инструментов потенциально крайне велико. Предположим, например, что приведенный выше набор инструментов строго экзогенен по отношению к η_{it} в каждом периоде, включая текущий, предыдущий и будущий. Тогда всего доступно $[T(K_1 + K_2) + L_1 + K_1]$ моментных условий для каждого наблюдения. Рассмотрим, например, панель с двумя периодами. Для двух периодов

$$E \left[\begin{pmatrix} \mathbf{x}_{1i1} \\ \mathbf{x}_{2i1} \\ \mathbf{x}_{1i2} \\ \mathbf{x}_{2i2} \\ \mathbf{z}_{1i} \\ \bar{\mathbf{x}}_{1i.} \end{pmatrix} (\eta_{i1} - \bar{\eta}_i) \right] = \mathbf{0} \quad \text{и} \quad E \left[\begin{pmatrix} \mathbf{x}_{1i1} \\ \mathbf{x}_{2i1} \\ \mathbf{x}_{1i2} \\ \mathbf{x}_{2i2} \\ \mathbf{z}_{1i} \\ \bar{\mathbf{x}}_{1i.} \end{pmatrix} (\eta_{i2} - \bar{\eta}_i) \right] = \mathbf{0}. \quad (13-38)$$

Насколько много полезной для оценки параметров информации привнесено, сказать сложно, так как это зависит от корреляции инструментов с включенными в уравнение экзогенными переменными. Чем дальше во времени отстоят эти наборы переменных, тем меньше информации они будут содержать. (Литература, посвященная этому вопросу, содержит указания на «сильные» инструментальные переменные против «слабых»¹⁶.) Для продолжения, как было отмечено, мы можем включить лагированную зависимую переменную в \mathbf{x}_{2i} . Этот набор инструментальных переменных может быть использован для построения оценки независимо от наличия запаздывающей зависимой переменной. Отметим, что на этом основании оценка Хаусмана и Тейлора не использовала всей доступной в выборке информации. Теперь мы получили элементы оценки Ареллано и др.; остаются, в сущности, только (к сожалению, достаточно замысловатые) алгебраические преобразования, которые приведены ниже.

Пусть

$$\mathbf{W}_i = \begin{bmatrix} \mathbf{w}'_{i1} \\ \mathbf{w}'_{i2} \\ \vdots \\ \mathbf{w}'_{iT} \end{bmatrix} = \text{полный набор данных для правой части в группе } i$$

и

$$\mathbf{y}_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix}.$$

¹⁶West (2001).

Заметим, что W_i по предположению матрица размерностью $T \times (1 + K_1 + K_2 + L_1 + L_2)$. Поскольку в модели присутствует лагированная зависимая переменная, предполагается что на самом деле только $T + 1$ наблюдений доступно для y_{it} . Чтобы избежать путаницы в обозначениях, мы оставим это различие включенным в обозначение. Позднее по мере необходимости мы сделаем его явным. Оно появится снова в формулировке модели инструментальных переменных. Всего для построения IV-оценки будет доступно T наблюдений. Теперь мы составляем матрицу инструментальных переменных. [Различные подходы к этому рассматривались Hausman, Taylor (1981), Arellano et al. (1991, 1995, 1999), Ahn, Schmidt (1995) и Amemiya, MaCurdy (1986) среди прочих.] Мы построим матрицу V_i , состоящую из $T_i - 1$ строк, составленных одним и тем же образом для $T_i - 1$ наблюдений и последней строки, которая, как будет обсуждаться позднее, будет отличаться. [Специально для использования полезного алгебраического результата, обсуждаемого Arellano, Bover (1995).] Матрица примет форму

$$V_i = \begin{bmatrix} v'_{i1} & 0' & \dots & 0' \\ 0' & v'_{i2} & \dots & 0' \\ \vdots & \vdots & \ddots & \vdots \\ 0' & 0' & \dots & a'_i \end{bmatrix}. \quad (13-39)$$

Предложенные наборы инструментальных переменных, содержащиеся в v'_{it} , могут включать полученные из модели:

x_{it} и $x_{i,t-1}$ (т.е. текущие и лагированные уровни всех меняющихся во времени переменных),

x_{i1}, \dots, x_{iT} (т.е. все текущие, прошлые и будущие значения всех меняющихся во времени переменных),

x_{i1}, \dots, x_{it} (т.е. все текущие и прошлые значения всех меняющихся во времени переменных).

Постоянные во времени переменные, не коррелированные с u_i , т.е. z_{1i} добавляются в конце ненулевых частей каждой из первых $T - 1$ строк. Может показаться, что включение x_2 в число инструментов будет ошибочным. Однако мы преобразуем ошибки в отклонения от средних по группам, которые свободны от скрытых эффектов, т.е. этот набор моментных условий будет в конечном счете преобразован в то, что представлено в (13-38). Хотя переменные коррелированы с u_i по построению, они не коррелированы с $\varepsilon_{it} - \bar{\varepsilon}_i$. Последняя строка V_i важна для построения. Были предложены две возможности:

$$a'_i = [z'_{1i} \quad \bar{x}_{i1}] \text{ (дает оценку Хаусмана и Тейлора),}$$

$$a'_i = [z'_{1i} \quad x'_{i11}, x'_{i12}, \dots, x_{i1T}] \text{ (дает оценку Амемии и Макариди).}$$

Заметим, что переменные a — это экзогенные и не меняющиеся во времени переменные z_{1i} и экзогенные, меняющиеся во времени переменные, или включенные в общее групповое среднее или в исходной форме с полным набором T наблюдений.

Для построения оценки нам потребуется трансформационная матрица, \mathbf{H} , построенная следующим образом. Пусть \mathbf{M}^{01} обозначает первые $T - 1$ строк \mathbf{M}^0 , матрицы, дающей отклонения от групповых средних. Тогда

$$\mathbf{H} = \begin{bmatrix} \mathbf{M}^{01} \\ \frac{1}{T} \mathbf{i}'_T \end{bmatrix}.$$

Таким образом, \mathbf{H} заменяет последний ряд \mathbf{M}^0 рядом из $1/T$. Получается следующий эффект: если \mathbf{q} — это T наблюдений переменной, то \mathbf{Hq} дает вектор \mathbf{q}^* , в котором первые $T - 1$ наблюдений обращаются в отклонения от групповых средних, а последнее наблюдение — это групповое среднее. В частности, пусть вектор-столбец размерностью $T \times 1$ ошибок равен

$$\boldsymbol{\eta}_i = [\eta_{i1}, \eta_{i2}, \dots, \eta_{iT}] = [(\varepsilon_{i1} + u_i), (\varepsilon_{i2} + u_i), \dots, (\varepsilon_{iT} + u_i)]',$$

тогда

$$\mathbf{H}\boldsymbol{\eta}_i = \begin{bmatrix} \eta_{i1} - \bar{\eta}_i \\ \vdots \\ \eta_{i,T-1} - \bar{\eta}_i \\ \bar{\eta}_i \end{bmatrix}.$$

Теперь мы можем составить моментные условия. С помощью всех этих преобразований мы получаем результат, приведенный в (13-40), т.е.

$$E[\mathbf{V}'_i \mathbf{H}\boldsymbol{\eta}_i] = E[\mathbf{g}_i] = \mathbf{0}.$$

Полезно развернуть это выражение для конкретного случая. Пусть $T = 3$ и в качестве инструментов мы используем текущие значения в период 1, текущие и предыдущие значения в период 2 и форму Хаусмана и Тейлора для инструментальных переменных. Тогда предыдущее выражение принимает форму

$$E \left[\begin{pmatrix} \mathbf{x}_{1i1} & \mathbf{0} & \mathbf{0} \\ \mathbf{x}_{2i1} & \mathbf{0} & \mathbf{0} \\ \mathbf{z}_{1i} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_{1i1} & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_{2i1} & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_{1i2} & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_{2i2} & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_{1i} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{z}_{1i} \\ \mathbf{0} & \mathbf{0} & \bar{\mathbf{x}}_{1i} \end{pmatrix} \begin{pmatrix} \eta_{i1} - \bar{\eta}_i \\ \eta_{i2} - \bar{\eta}_i \\ \bar{\eta}_i \end{pmatrix} \right] = \mathbf{0}. \quad (13-40)$$

Это то же самое, что и (13-38)¹⁷. Из этого следует эмпирическое моментное

¹⁷В некоторых источниках [напр., Blundell, Bond (1998)] дополнительное условие накладывается на начальное значение y_{i0} , а именно $E[y_{i0} | \text{экзогенные данные}] = \mu_0$. Это добавляет строку вверху матрицы в (13-40), содержащую $[(y_{i0} - \mu_0), 0, 0]$.

условие

$$\begin{aligned} & \text{plim} \frac{1}{n} \sum_{i=1}^n \mathbf{V}'_i \mathbf{H} \eta_i \\ &= \text{plim} \frac{1}{n} \sum_{i=1}^n \mathbf{V}'_i \mathbf{H} \begin{pmatrix} y_{i1} - \delta y_{i0} - \mathbf{x}'_{1i1} \beta_1 - \mathbf{x}'_{2i1} \beta_2 - \mathbf{z}'_{1i} \alpha_1 - \mathbf{z}'_{2i} \alpha_2 \\ y_{i2} - \delta y_{i1} - \mathbf{x}'_{1i2} \beta_1 - \mathbf{x}'_{2i2} \beta_2 - \mathbf{z}'_{1i} \alpha_1 - \mathbf{z}'_{2i} \alpha_2 \\ \vdots \\ y_{iT} - \delta y_{i,T-1} - \mathbf{x}'_{1iT} \beta_1 - \mathbf{x}'_{2iT} \beta_2 - \mathbf{z}'_{1i} \alpha_1 - \mathbf{z}'_{2i} \alpha_2 \end{pmatrix} = 0. \end{aligned}$$

Запишем это как

$$\text{plim} \frac{1}{n} \sum_{i=1}^n \mathbf{m}_i = \text{plim} \bar{\mathbf{m}} = \mathbf{0}.$$

Тогда GMM-оценка $\hat{\theta}$ получается минимизацией

$$q = \bar{\mathbf{m}}' \mathbf{A} \bar{\mathbf{m}}$$

с подобранной соответствующим образом взвешивающей матрицей \mathbf{A} . Оптимальная взвешивающая матрица будет обратной к асимптотической ковариационной матрице $\sqrt{n} \bar{\mathbf{m}}$. Имея состоятельную оценку θ , эта оценка может быть получена эмпирически, используя

$$\text{Est. Asy. Var}[\sqrt{n} \bar{\mathbf{m}}] = \frac{1}{n} \sum_{i=1}^n \hat{\mathbf{m}}_i \hat{\mathbf{m}}'_i = \frac{1}{n} \sum_{i=1}^n \mathbf{V}'_i \mathbf{H} \hat{\eta}_i \hat{\eta}'_i \mathbf{H}' \mathbf{V}_i.$$

Это устойчивая оценка, позволяющая получить неограниченную ковариационную матрицу размерностью $T \times T$ для T ошибок $\varepsilon_{it} + u_i$. Но мы предположили, что эта ковариационная матрица — это Σ , определенная в (11-31) для модели со случайными эффектами. Для включения этой информации мы использовали бы остатки из

$$\hat{\eta}_i = \mathbf{y}_i - \mathbf{W}_i \hat{\theta}$$

для оценки σ_u^2 и σ_ε^2 и затем Σ , что дает

$$\text{Est. Asy. Var}[\sqrt{n} \bar{\mathbf{m}}] = \frac{1}{n} \sum_{i=1}^n \mathbf{V}'_i \mathbf{H} \hat{\Sigma} \mathbf{H}' \mathbf{V}_i.$$

Теперь у нас в распоряжении полный набор результатов, необходимых для получения GMM-оценки. Решение оптимизационной задачи минимизации q по вектору параметров θ — это

$$\begin{aligned} \hat{\theta}_{GMM} &= \left[\left(\sum_{i=1}^n \mathbf{W}'_i \mathbf{H} \mathbf{V}_i \right) \left(\sum_{i=1}^n \mathbf{V}'_i \mathbf{H}' \hat{\Sigma} \mathbf{H} \mathbf{V}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{V}'_i \mathbf{H}' \mathbf{W}_i \right) \right]^{-1} \\ &\times \left(\sum_{i=1}^n \mathbf{W}'_i \mathbf{H} \mathbf{V}_i \right) \left(\sum_{i=1}^n \mathbf{V}'_i \mathbf{H}' \hat{\Sigma} \mathbf{H} \mathbf{V}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{V}'_i \mathbf{H}' \mathbf{y}_i \right). \quad (13-41) \end{aligned}$$

Оценка асимптотической ковариационной матрицы для $\hat{\theta}_{GMM}$ — это обратная матрица в скобках.

Осталось выяснить, как получить состоятельную оценку $\hat{\theta}$ для оценки Σ . Вспомним, что GMM-оценка состоятельна с любой положительно определенной взвешивающей матрицей A в нашем предыдущем выражении. Следовательно, для начальной оценки мы можем взять $A = I$ и использовать простую оценку методом инструментальных переменных:

$$\hat{\theta}_{IV} = \left[\left(\sum_{i=1}^n \mathbf{w}'_i \mathbf{H} \mathbf{v}_i \right) \left(\sum_{i=1}^n \mathbf{v}'_i \mathbf{H}' \mathbf{w}_i \right) \right]^{-1} \left(\sum_{i=1}^n \mathbf{w}'_i \mathbf{H} \mathbf{v}_i \right) \left(\sum_{i=1}^n \mathbf{v}'_i \mathbf{H}' \mathbf{y}_i \right).$$

Более привычно сразу продолжить применять двухшаговый метод наименьших квадратов (см. подразделы 8.3.4 и 11.8.2), которые используют

$$A = \left(\frac{1}{n} \sum_{i=1}^n \mathbf{v}'_i \mathbf{H}' \mathbf{H} \mathbf{v}_i \right)^{-1}.$$

В результате выражение для оценки будет тем же, что приводилось ранее в (13-41) с $\hat{\Sigma}$, замененной на I_T . Любая из этих оценок является функцией только выборочных данных и представляет начальную оценку, которая нам требуется.

Эн и Шмидт (среди прочих) обратили внимание на то, что оценка методом инструментальных переменных, предложенная здесь, несмотря на использование обширной информации, все еще упускает ее часть и, следовательно, является (относительно) неэффективной. Например, в модели в первых разностях

$$E[y_{is}(\varepsilon_{it} - \varepsilon_{i,t-1})] = 0, \quad s = 0, \dots, t-2, \quad t = 2, \dots, T.$$

То есть уровень y_{is} не коррелирован с разностями ошибок, следующих, по крайней мере, через два периода¹⁸. (Преобразование взятия первых разностей, как и преобразование в отклонения от групповых средних, убирает индивидуальный эффект.) Соответствующие моментные уравнения, которые могут участвовать в построении GMM-оценки, имеют вид

$$\frac{1}{n} \sum_{i=1}^n y_{is} [(y_{it} - y_{i,t-1}) - \delta(y_{i,t-1} - y_{i,t-2}) - (\mathbf{x}_{it} - \mathbf{x}_{i,t-1})' \beta] = 0, \\ s = 0, \dots, t-2, \quad t = 2, \dots, T.$$

Всего Эн и Шмидт определяют $T(T-1)/2 + T - 2$ таких уравнений, которые включают смеси уровней и разностей переменных. Основной их вывод состоит в том, что в динамической модели большое количество информации получается не только из знакомых соотношений между уровнями переменных, но также из подразумеваемых соотношений между уровнями и первыми разностями. Вопрос корреляции между преобразованным y_{it} и отклонениями ε_{it} обсуждается в цитируемых статьях. [Как показывают Эн и Шмидт,

¹⁸Этот подход был предложен Holtz-Eakin (1988) и Holtz-Eakin, Newey, Rosen (1988).

в этой модели потенциально огромное число дополнительных условий ортогональности ввиду наличия зависимости первых разностей и вторых моментов. Мы не рассматриваем такие условия. Матрица V_i может быть огромной. Рассмотрим модель с 10 меняющимися во времени переменными справа и предположим, что T_i равно 15. Тогда в ней 15 строк и примерно $15 \times (10 \times 15)$, или 2250, столбцов. Оценка Эна и Шмидта, которая потенциально включает тысячи инструментов в модели, содержащей лишь дюжину параметров, может перестать быть практичной в такой ситуации. Общий подход состоит в использовании только малого подмножества доступных инструментальных переменных. Порядок вычислительной сложности растет, как число параметров, умноженное на квадрат T .]

Число условий ортогональности (инструментальных переменных), используемых для оценивания параметров модели, определяется числом переменных в v_{it} и a_i в (13-39). В большинстве случаев модель сильно переопределена — условий ортогональности гораздо больше, чем параметров. Как обычно в GMM-оценивании, тест на сверхидентифицирующие ограничения может быть основан на q , критерии оценивания. В точке минимума предельное распределение nq — это хи-квадрат со степенями свободы, равными общему числу инструментальных переменных за вычетом $(1 + K_1 + K_2 + L_1 + L_2)^{19}$.

Пример 13.10. *GMM-оценка динамической модели панельных данных для расходов местных органов власти*

Dahlberg, Johansson (2000) оценили модель расходов местных органов власти нескольких сотен муниципальных округов в Швеции, наблюдавшихся в течение девяти лет между $t = 1979$ и 1987 гг. Центральное уравнение — это

$$S_{i,t} = \alpha_t + \sum_{j=1}^m \beta_j S_{i,t-j} + \sum_{j=1}^m \gamma_j R_{i,t-j} + \sum_{j=1}^m \delta_j G_{i,t-j} + f_i + \varepsilon_{it}$$

для $i = 1, \dots, n = 265$ и $t = m + 1, \dots, 9$. (Мы слегка изменили обозначения, чтобы сделать их более удобными.) $S_{i,t}$, $R_{i,t}$ и $G_{i,t}$ — это муниципальные расходы, доходы (налоги и пошлины) и гранты центрального правительства соответственно. Аналогичные уравнения специфицируются для текущих значений $R_{i,t}$ и $G_{i,t}$. Подходящая глубина запаздывания m является одной из интересных особенностей, которые предстоит выяснить в ходе исследования. Модель содержит эффект муниципалитета f_i , который не специфицируется как фиксированный или случайный. Для исключения индивидуального эффекта модель рассматривается в первых разностях. Итоговое уравнение — это

$$\Delta S_{i,t} = \lambda_t + \sum_{j=1}^m \beta_j \Delta S_{i,t-j} + \sum_{j=1}^m \gamma_j \Delta R_{i,t-j} + \sum_{j=1}^m \delta_j \Delta G_{i,t-j} + u_{it},$$

или

$$y_{i,t} = \mathbf{x}'_{i,t} \boldsymbol{\theta} + u_{i,t},$$

¹⁹Это верно в общем случае GMM-оценивания. В контексте динамической модели панельных данных этот вопрос был рассмотрен в работе Bhargava, Sargan (1983).

где $\Delta S_{i,t} = S_{i,t} - S_{i,t-1}$ и т.д. и $u_{i,t} = \varepsilon_{i,t} - \varepsilon_{i,t-1}$. Это преобразование убирает групповой эффект и оставляет временной. Поскольку временной эффект был неограничен, $\Delta \alpha_t = \lambda_t$ остается неограниченным временным эффектом, который оценивается как фиксированный и моделируется при помощи фиктивных переменных для отдельных периодов. Максимальная длина лага составляет $m = 3$. Поскольку рассматриваются данные за 9 лет, после всех преобразований для оценки остаются наблюдения с 1983 по 1987 г., т.е. $t = m + 2, \dots, 9$. Похожие уравнения оцениваются для $R_{i,t}$ и $G_{i,t}$.

Условия ортогональности, заявленные авторами, имеют вид

$$E[S_{i,s}u_{i,t}] = E[R_{i,s}u_{i,t}] = E[G_{i,s}u_{i,t}] = 0, \quad s = 1, \dots, t - 2.$$

Условия ортогональности сформулированы в терминах уровней финансовых переменных и разностей ошибок. Проблема с такой формулировкой в отличие от, например, $E[\Delta S_{i,s} \Delta \varepsilon_{i,t}] = 0$ (которая из нее следует) обсуждается в работе Эна и Шмидта (Ahn, Schmidt (1995)). Как мы увидим, этот набор условий ортогональности подразумевает общее число инструментальных переменных, равное 80. Авторы используют только первый из трех приведенных наборов, что дает общее число, равное 30. Для пяти наблюдений, используя постановку, приведенную в подразделе 13.6.5, мы получаем следующую матрицу инструментальных переменных в условиях ортогональности:

$$Z_i = \begin{bmatrix} S_{81-79} & d_{83} & \mathbf{0}' & 0 & \mathbf{0}' & 0 & \mathbf{0}' & 0 & \mathbf{0}' & 0 & 0 \\ \mathbf{0}' & 0 & S_{82-79} & d_{84} & \mathbf{0}' & 0 & \mathbf{0}' & 0 & \mathbf{0}' & 0 & 0 \\ \mathbf{0}' & 0 & \mathbf{0}' & 0 & S_{83-79} & d_{85} & \mathbf{0}' & 0 & \mathbf{0}' & 0 & 0 \\ \mathbf{0}' & 0 & \mathbf{0}' & 0 & \mathbf{0}' & 0 & S_{84-79} & d_{86} & \mathbf{0}' & 0 & 0 \\ \mathbf{0}' & 0 & \mathbf{0}' & 0 & \mathbf{0}' & 0 & \mathbf{0}' & 0 & \mathbf{0}' & 0 & S_{85-79} & d_{87} \end{bmatrix} \begin{matrix} 1983 \\ 1984 \\ 1985 \\ 1986 \\ 1987, \end{matrix}$$

где обозначение S_{t1-t0} указывает диапазон лет для этой переменной. Например, S_{83-79} обозначает $[S_{i,1983}, S_{i,1982}, S_{i,1981}, S_{i,1980}, S_{i,1979}]$, а d_{year} — фиктивные переменные для конкретных лет. Подсчитывая столбцы Z_i , мы видим, что, используя только лагированные значения зависимой переменной и фиктивные переменные периодов, получаем $(3+1) + (4+1) + (5+1) + (6+1) + (7+1) = 30$ инструментальных переменных. Использование лагированных значений других двух переменных в каждом уравнении добавило бы дополнительно 50 для получения итоговых 80, если бы использовались все предложенные ранее условия ортогональности. Учитывая предыдущее построение, условия ортогональности теперь принимают вид

$$E[Z_i' u_i] = 0,$$

где $u_i = [u_{i,1983}, u_{i,1984}, u_{i,1985}, u_{i,1986}, u_{i,1987}]'$. Эмпирическое моментное

условие — это

$$\text{plim} \left[\frac{1}{n} \sum_{i=1}^n \mathbf{Z}'_i \mathbf{u}_i \right] = \text{plim} \bar{\mathbf{m}}(\boldsymbol{\theta}) = \mathbf{0}.$$

Параметры сильно переопределены. Используя только запаздывающие значения зависимой переменной в каждом из трех оцененных уравнений, число моментных условий равно 30 при оценивании 14 параметров, когда $m = 3$, 11, когда $m = 2$, 8, когда $m = 1$, и 5, когда $m = 0$. (Поскольку мы оцениваем в каждом из этих случаев, мы также будем использовать одну и ту же матрицу инструментальных переменных в каждом из них.) GMM-оценка может быть получена в два этапа. На первом, базовом, невзвешанные инструментальные переменные оцениваются с использованием

$$\hat{\boldsymbol{\theta}}'_{IV} = \left[\left(\sum_{i=1}^n \mathbf{X}'_i \mathbf{Z}_i \right) \left(\sum_{i=1}^n \mathbf{Z}'_i \mathbf{Z}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{Z}'_i \mathbf{X}_i \right) \right]^{-1} \left(\sum_{i=1}^n \mathbf{X}'_i \mathbf{Z}_i \right) \left(\sum_{i=1}^n \mathbf{Z}'_i \mathbf{Z}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{Z}'_i \mathbf{y}_i \right),$$

где $\mathbf{y}'_i = (\Delta S_{83} \ \Delta S_{84} \ \Delta S_{85} \ \Delta S_{86} \ \Delta S_{87})$ и $\mathbf{X} =$

$$\begin{bmatrix} \Delta S_{82} & \Delta S_{81} & \Delta S_{80} & \Delta S_{82} & \Delta S_{81} & \Delta S_{80} & \Delta S_{82} & \Delta S_{81} & \Delta S_{80} & 1 & 0 & 0 & 0 & 0 \\ \Delta S_{83} & \Delta S_{82} & \Delta S_{81} & \Delta S_{83} & \Delta S_{82} & \Delta S_{81} & \Delta S_{83} & \Delta S_{82} & \Delta S_{81} & 0 & 1 & 0 & 0 & 0 \\ \Delta S_{84} & \Delta S_{83} & \Delta S_{82} & \Delta S_{84} & \Delta S_{83} & \Delta S_{82} & \Delta S_{84} & \Delta S_{83} & \Delta S_{82} & 0 & 0 & 1 & 0 & 0 \\ \Delta S_{85} & \Delta S_{84} & \Delta S_{83} & \Delta S_{85} & \Delta S_{84} & \Delta S_{83} & \Delta S_{85} & \Delta S_{84} & \Delta S_{83} & 0 & 0 & 0 & 1 & 0 \\ \Delta S_{86} & \Delta S_{85} & \Delta S_{84} & \Delta S_{86} & \Delta S_{85} & \Delta S_{84} & \Delta S_{86} & \Delta S_{85} & \Delta S_{84} & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Второй этап начинается с вычисления новой матрицы весов

$$\hat{\Phi} = \text{Est. Asy. Var}[\sqrt{n}\bar{\mathbf{m}}] = \frac{1}{N} \sum_{i=1}^n \mathbf{Z}'_i \hat{\mathbf{u}}_i \hat{\mathbf{u}}'_i \mathbf{Z}_i.$$

После умножения и деления на неявные $(1/n)$ во внешних матрицах мы получаем оценку

$$\begin{aligned} \boldsymbol{\theta}'_{GMM} &= \left[\left(\sum_{i=1}^n \mathbf{X}'_i \mathbf{Z}_i \right) \left(\sum_{i=1}^n \mathbf{Z}'_i \hat{\mathbf{u}}_i \hat{\mathbf{u}}'_i \mathbf{Z}_i \right)^{-1} \left(\sum_{i=1}^n \mathbf{Z}'_i \mathbf{X}_i \right) \right]^{-1} \times \\ &\times \left(\sum_{i=1}^n \mathbf{X}'_i \mathbf{Z}_i \right) \left(\sum_{i=1}^n \{ \mathbf{Z}'_i \hat{\mathbf{u}}_i \hat{\mathbf{u}}'_i \mathbf{Z}_i \} \right)^{-1} \left(\sum_{i=1}^n \mathbf{Z}'_i \mathbf{y}_i \right) = \\ &= \left[\left(\sum_{i=1}^n \mathbf{X}'_i \mathbf{Z}_i \right) \mathbf{W} \left(\sum_{i=1}^n \mathbf{Z}'_i \mathbf{X}_i \right) \right]^{-1} \left(\sum_{i=1}^n \mathbf{X}'_i \mathbf{Z}_i \right) \mathbf{W} \left(\sum_{i=1}^n \mathbf{Z}'_i \mathbf{y}_i \right). \end{aligned}$$

Оценкой асимптотической ковариационной матрицы оценки является обратная матрица в квадратных скобках в первой строке результата.

Основной интерес в исследовании представляла не сама оценка, а длина лага и наличие лагированных значений независимых переменных в каждом уравнении. Эти ограничения могли быть протестированы с использованием целевой функции GMM, которая в такой постановке приняла бы вид (основанный на пересчете остатков после GMM-оценки)

$$nq = \left(\sum_{i=1}^n \hat{\mathbf{u}}'_i \mathbf{Z}_i \right) \mathbf{W} \left(\sum_{i=1}^n \mathbf{Z}'_i \hat{\mathbf{u}}_i \right).$$

Отметьте, что матрица весов не (обязательно) переоценивается. Для тестирования гипотез должна использоваться одна и та же взвешивающая матрица.

Теперь мы рассмотрим подходящую длину лага m . Спецификация может быть сокращена простым переопределением X , изменяя длину лага. Для тестирования спецификации взвешивающая матрица должна оставаться постоянной для всех ограниченных ($m = 2$ и $m = 1$) версий модели.

Данные Дальберга и Йохансона могут быть получены на веб-сайте *Journal of Applied Econometrics* — см. табл. F13.1 приложения. Авторы представляют описательную статистику для исходных данных, которая приведена в табл. 13.3. Данные, использованные в исследовании и представленные в онлайн-источнике, — это номинальные значения в шведских кронах, продефлированные с помощью ценовых индексов отдельных муниципалитетов и затем преобразованные в уровни на душу населения. Описательная статистика для исходных данных приведена в табл. 13.3²⁰. Уравнения оценивались для всех трех переменных с максимальными длинами лагов $m = 1, 2$ и 3 . (Авторы не предоставили сами оценки.) Оценивание, как и было описано, проводилось методами Эна и Шмидта (Ahn, Schmidt (1995)), Ареллано и Бовера (Arellano, Bover (1995)) и Holtz-Eakin, Newey и Rosen (1988). Оценки первой представленной спецификации приведены в табл. 13.4.

Таблица 13.3. Описательные статистики данных о расходах местных органов власти

| Переменная | Среднее | Станд. отклонение | Минимум | Максимум |
|--------------------|----------|-------------------|----------|----------|
| Spending (расходы) | 18478,51 | 3174,36 | 12225,68 | 33883,25 |
| Revenues (доходы) | 13422,56 | 3004,16 | 6228,54 | 29141,62 |
| Grants (гранты) | 5236,03 | 1260,97 | 1570,64 | 12589,14 |

Таблица 13.5 содержит оценки параметров модели для каждого из трех уравнений, для трех длин лагов и значение целевой функции GMM для каждой оцененной модели. Базовый случай для каждой модели включал $m = 3$. Три ограничения следовали из каждого сокращения длины лага. Критическое значение хи-квадрат-распределения для трех степеней свободы равно 7,81 при уровне значимости 95%, так что на этом уровне значимости мы получаем, что двухуровневая модель может быть рассмотрена как подходящая для уравнения затрат, но хорошо подходит и для двух других — разность между двумя критериями равна 7,62. Модель доходов, оцененная только для $m = 2$, отвергает ограничение $m = 1$. В качестве финального теста мы можем выяснить, указывают ли данные на то, что лаговая структура вообще является излишней. Значения GMM-критерия для трех уравнений с исключительно

²⁰ Данные, представленные на сайте и использованные в наших вычислениях, были также поделены на 100,000.

фиктивными переменными составили 45,840, 57,908, и 62,042 соответственно. Следовательно, все три модели без лагов отвергаются.

Таблица 13.4. Оценка уравнения расходов

| Переменная | Оценка | Станд. ошибка | <i>t</i> -отношение |
|--------------------------|-------------|---------------|---------------------|
| Year 1983 | -0,0036578 | 0,0002969 | -12,32 |
| Year 1984 | -0,00049670 | 0,0004128 | -1,20 |
| Year 1985 | 0,00038085 | 0,0003094 | 1,23 |
| Year 1986 | 0,00031469 | 0,0003282 | 0,96 |
| Year 1987 | 0,00086878 | 0,0001480 | 5,87 |
| Spending (<i>t</i> - 1) | 1,15493 | 0,34409 | 3,36 |
| Revenues (<i>t</i> - 1) | -1,23801 | 0,36171 | -3,42 |
| Grants (<i>t</i> - 1) | 0,016310 | 0,82419 | 0,02 |
| Spending (<i>t</i> - 2) | -0,0376625 | 0,22676 | -0,17 |
| Revenues (<i>t</i> - 2) | 0,0770075 | 0,27179 | 0,28 |
| Grants (<i>t</i> - 2) | 1,55379 | 0,75841 | 2,05 |
| Spending (<i>t</i> - 3) | -0,56441 | 0,21796 | -2,59 |
| Revenues (<i>t</i> - 3) | 0,64978 | 0,26930 | 2,41 |
| Grants (<i>t</i> - 3) | 1,78918 | 0,69297 | 2,58 |

Интерес в этом исследовании представлял выбор подходящих для использования в тестах на спецификацию моментных условий критических значений. С 16 степенями свободы критическое значение хи-квадрат-распределения для 95%-го уровня значимости равно 26,3, что указывает на неправильную спецификацию уравнения доходов. Используя процедуру самонастройки (бутстрэп), авторы выяснили, что более подходящее критическое значение указывает на разумность спецификации. Наконец, заметим, что модель из трех уравнений в столбцах $m = 3$ табл. 13.5 подразумевает векторную авторегрессию вида

$$y_t = \Gamma_1 y_{t-1} + \Gamma_2 y_{t-2} + \Gamma_3 y_{t-3} + v_t,$$

где $y_t = (\Delta S_t, \Delta R_t, \Delta G_t)'$.

13.7. Заключение и выводы

Обобщенный метод моментов является подходом к оценке, включающим метод наименьших квадратов, нелинейный метод наименьших квадратов, методы инструментальных переменных и максимального правдоподобия, а также общий класс оценок помимо приведенных здесь. Но он представляет больше, чем просто теоретическое обобщение. GMM указывает способ формулировки моделей и соответствующих оценок, не требующий сильных предположений о распределении. Модель потребления Холла является полезным примером, который показывает, как условия оптимальности,

Таблица 13.5. Оценка уравнений с лагами для расходов, доходов и грантов

| | Модель расходов | | | Модель доходов | | | Модель грантов | | |
|-----------|-----------------|---------|---------|----------------|---------|---------|----------------|---------|---------|
| | $m = 3$ | $m = 2$ | $m = 1$ | $m = 3$ | $m = 2$ | $m = 1$ | $m = 3$ | $m = 2$ | $m = 1$ |
| S_{t-1} | 1,155 | 0,8742 | 0,5562 | -0,1715 | -0,3117 | -0,1242 | -0,1675 | -0,1461 | -0,1958 |
| S_{t-2} | -0,0377 | 0,2493 | — | 0,1621 | -0,0773 | — | -0,0303 | -0,0304 | — |
| S_{t-3} | -0,5644 | — | — | -0,1772 | — | — | -0,0955 | — | — |
| R_{t-1} | -1,2380 | -0,8745 | -0,5328 | -0,0176 | 0,1863 | -0,0245 | 0,1578 | 0,1453 | 0,2343 |
| R_{t-2} | 0,0770 | -0,2776 | — | -0,0309 | 0,1368 | — | 0,0485 | 0,0175 | — |
| R_{t-3} | 0,6497 | — | — | 0,0034 | — | — | 0,0319 | — | — |
| G_{t-1} | 0,0163 | -0,4203 | 0,1275 | -0,3683 | 0,5425 | -0,0808 | -0,2381 | -0,2066 | -0,0559 |
| G_{t-2} | 1,5538 | 0,1866 | — | 2,7152 | 2,4621 | — | -0,0492 | -0,0804 | — |
| G_{t-3} | 1,7892 | — | — | 0,0948 | — | — | 0,0598 | — | — |
| n_q | 22,8287 | 30,4526 | 34,4986 | 30,5398 | 34,2590 | 53,2506 | 17,5810 | 20,5416 | 27,5927 |

следующие из экономической теории, позволяют получить уравнения для оценки, не зависящие от распределений. В этой главе мы сначала рассмотрели классический метод моментов. GMM как оценка представляет собой расширение этой стратегии, дает исследователю возможность оптимальным образом использовать дополнительную информацию помимо необходимой для идентификации модели. После определения и установления свойств оценки мы перешли к процедурам проверки статистических гипотез. Очень удобно, что GMM представляет аналоги известного трио тестовых статистик: Вальда, LM и LR. В заключительном разделе мы специфицировали GMM-оценки для линейных и нелинейных уравнений и моделей из нескольких уравнений, затем привели пример, который часто встречается в актуальной прикладной литературе, — динамическую модель панельных данных с индивидуальными эффектами и лагированными значениями зависимой переменной.

Основные термины и понятия

Analog estimation — Аналоговое оценивание

Arellano and Bond — Ареллано и Бонд

Arellano and Bover estimator — Оценка Ареллано и Бовера

Central limit theorem — Центральная предельная теорема

Criterion function — Целевая функция

Dynamic panel data model — Динамическая модель панельных данных

Empirical moment equation — Эмпирическое моментное условие

Ergodic theorem — Эргодическая теорема

Euler equation — Уравнение Эйлера

Exactly identified cases — Случай точной идентификации

Exponential family — Экспоненциальное семейство

Generalized method of moments — Обобщенный метод моментов

GMM estimator — GMM-оценка

H2SLS — H2GLS

Instrumental variables — Инструментальные переменные

Likelihood ratio statistic — Статистика отношения правдоподобия

LM statistic — LM-статистика

Martingale difference series — Ряд мартингальных разностей

| | |
|--|--|
| Maximum likelihood estimator — Оценка методом максимального правдоподобия | Orthogonality conditions — Условие ортогональности |
| Mean value theorem — Теорема о среднем | Overidentifying restrictions — Сверхидентифицирующие ограничения |
| Method of moment generating functions — Метод производящих функций моментов | Overidentified cases — Случаи переопределенности |
| Method of moments — Метод моментов | Population moment equation — Моментные уравнения в генеральной совокупности |
| Method of moments estimators — Оценка методом моментов | Probability limit — Предел по вероятности |
| Minimum distance estimator (MDE) — Оценка методом наименьшего расстояния (MDE) | Random sample — Случайная выборка |
| Moment equation — Моментное уравнение | Rank condition — Условие ранга |
| Newey-West estimator — Оценка Ньюи-Веста | Slutsky theorem — Теорема Слуцкого |
| Nonlinear instrumental variable estimator — Оценка нелинейным методом инструментальных переменных | Specification test — Тест спецификации |
| Optimal weighting matrix — Оптимальная взвешивающая матрица | Sufficient statistic — Достаточная статистика |
| Order condition — Условие порядка | Taylor series — Ряд Тейлора |
| | Uncentered moment — Нецентрированный момент |
| | Wald statistic — Статистика Вальда |
| | Weighted least squares — Взвешенный метод наименьших квадратов |
| | Weighting matrix — Взвешивающая матрица (матрица весов) |

Упражнения

1. Для нормального распределения $\mu_{2k} = \sigma^{2k}(2k)!/(k!2^k)$ и $\mu_{2k+1} = 0$, $k = 0, 1, \dots$. Используйте этот результат для анализа двух оценок

$$\sqrt{b_1} = \frac{m_3}{m_2^{3/2}} \quad \text{и} \quad b_2 = \frac{m_4}{m_2^2},$$

где $m_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k$. Полезным будет следующий результат:

$$\text{Asy. Cov}[\sqrt{nm_j}, \sqrt{nm_k}] = \mu_{j+k} - \mu_j \mu_k + jk \mu_2 \mu_{j-1} \mu_{k-1} - \\ - j \mu_{j-1} \mu_{k+1} - k \mu_{k-1} \mu_{j+1}.$$

Используйте дельта-метод для получения асимптотических дисперсии и ковариации этих двух функций, предполагая, что данные порождаются нормальным распределением со средним μ и дисперсией σ^2 . (Подсказка: исходя из предположений, выборочное среднее — это состоятельная оценка μ , так что для получения асимптотических результатов разность между \bar{x} и μ может быть проигнорирована. Таким образом, без потери общности можно предположить, что среднее равно нулю.) Получите V , ковариационную матрицу размерностью 3×3 для трех моментов, и затем используйте дельта-метод, чтобы показать, что ковариационная матрица для двух оценок равна

$$JVJ' = \begin{bmatrix} 6/n & 0 \\ 0 & 24/n \end{bmatrix},$$

где J — матрица производных размерностью 2×3 .

2. Используя результаты из примера 13.5, оцените асимптотическую ковариационную матрицу оценок методом моментов P и λ на основе m'_1 и m'_2 . [Замечание: для оценки V вам потребуются данные из примера С.1.]
3. **Экспоненциальные семейства распределений.** Для каждого из следующих распределений определите, относятся ли они к экспоненциальному семейству, рассмотрев функцию логарифма правдоподобия. Затем определите достаточные статистики:
 - a. нормальное распределение со средним μ и дисперсией σ^2 ;
 - b. распределение Уэйбулла (Weibull) из упражнения 4 из главы 14;
 - c. смешанное распределение из упражнения 3 из главы 14.
4. Какой метод более эффективен в классической регрессионной модели с гетероскедастичностью: метод наименьших квадратов или GMM? Получите две оценки и соответствующие асимптотические ковариационные матрицы, затем докажите ваше утверждение.
5. Рассмотрите пробит-модель из главы 17. В модели утверждается, что для заданного вектора независимых переменных

$$\text{Prob}[y_i = 1 | \mathbf{x}_i] = \Phi[\mathbf{x}'_i \boldsymbol{\beta}], \quad \text{Prob}[y_i = 0 | \mathbf{x}_i] = 1 - \text{Prob}[y_i = 1 | \mathbf{x}_i].$$

Рассмотрим GMM-оценку, основанную на том, что

$$E[y_i | \mathbf{x}_i] = \Phi(\mathbf{x}'_i \boldsymbol{\beta}).$$

Это указывает на то, что оценка может опираться на условия ортогональности

$$E[(y_i - \Phi(\mathbf{x}'_i \boldsymbol{\beta})) \mathbf{x}_i] = 0.$$

Постройте GMM-оценку на основе этих результатов. Заметим, что она не является оценкой нелинейным методом инструментальных переменных. Объясните, как выглядели бы условия ортогональности для нелинейного метода наименьших квадратов в этой модели.

6. Рассмотрим GMM-оценку регрессионной модели, приведенной в начале примера 13.8. Пусть W_1 будет оптимальной взвешивающей матрицей, основанной на моментных уравнениях. Пусть W_2 будет некоторой другой положительно определенной матрицей. Сравните асимптотические ковариационные матрицы двух предложенных оценок. Покажите, что асимптотическая ковариационная матрица оценки, основанной на W_1 , не больше чем матрица оценки, основанной на W_2 .

ГЛАВА 14

Метод максимального правдоподобия

14.1. Введение

Обобщенный метод моментов, рассмотренный в главе 13, и полупараметрические, непараметрические и байесовские оценки, приведенные в главах 12 и 16, находят все большее применение при построении моделей. Тем не менее оценки методом максимального правдоподобия, представленные в данной главе, остаются более предпочтительными. Исходя из этого мы сконцентрируемся на распространенных способах оценивания с помощью этой методики. Разделы 14.2–14.6 содержат основные статистические результаты, необходимые для проверки гипотез методом максимального правдоподобия. В разделах 14.7 и 14.8 мы рассматриваем два обобщения этого метода: двухшаговую оценку и оценку методом квазimaxимального правдоподобия. После получения общих результатов для этого метода оценивания мы применяем его в более привычной постановке эконометрических моделей. В разделах 14.9 и 14.10 применим метод максимального правдоподобия к большей части моделей, встречавшихся нам ранее, а также к некоторым другим, чтобы продемонстрировать разнообразное использование этого метода оценивания.

14.2. Функция правдоподобия и идентификация параметров

Функция плотности вероятности (probability density function, pdf) для случайной переменной y условно по набору параметров θ обозначается как $f(y | \theta)$ ¹. Эта функция описывает процесс порождения данных, на основе которого получилась наблюдаемая выборка, и в то же время дает математическое описание данных, порожденных процессом. Плотность совместного распределения n независимых и одинаково распределенных (н.о.р., i.i.d., independent and identically distributed) наблюдений, порожденных этим процессом, представляет собой произведение отдельных плотностей

$$f(y_1, \dots, y_n | \theta) = \prod_{i=1}^n f(y_i | \theta) = L(\theta | y). \quad (14-1)$$

Эта совместная плотность является **функцией правдоподобия** (likelihood function), определяемой как функция от вектора неизвестных параметров

¹ В дальнейшем мы будем рассматривать случайный вектор y с многомерной плотностью, но на данном этапе это лишь усложнит изложение, не добавляя ничего существенного в изучение вопроса.

θ , где y обозначает имеющийся набор данных. Обратим внимание, что мы выражали совместную плотность как функцию от данных условно по параметрам, в то время как функцию правдоподобия записали, наоборот, как функцию от параметров, условно по имеющимся данным. Хотя обе функции — это одно и то же, функция правдоподобия записана таким образом для того, чтобы подчеркнуть, что нас интересуют именно параметры и та информация о них, которая содержится в наблюдаемых данных. Тем не менее требуется понимать, что мы не имеем в виду то, что функция правдоподобия отображает плотность вероятности параметров, как это имеет место в главе 16. В этой классической постановке задачи оценивания предполагается, что параметры являются фиксированными постоянными, которые мы исследуем по имеющимся данным.

Обычно проще работать с логарифмом функции правдоподобия

$$\ln L(\theta | y) = \sum_{i=1}^n \ln f(y_i | \theta). \quad (14-2)$$

Здесь также, чтобы подчеркнуть, что нас интересуют параметры при заданных наблюдаемых данных, мы обозначаем эту функцию как $L(\theta | \text{data}) = L(\theta | y)$. Функция правдоподобия и ее логарифм, вычисленные в точке θ , иногда обозначаются как $L(\theta)$ и $\ln L(\theta)$ соответственно или, если нет возможности появления неясности, просто как L или $\ln L$.

Обычно требуется обобщать понятие функции правдоподобия, для того чтобы условная плотность могла зависеть от других переменных. Чтобы сразу перейти к одной из наших основных моделей, предположим, что шоки в классической модели линейной регрессии являются нормально распределенными. Тогда условно по заданному x_i величина y_i распределена нормально со средним $\mu_i = x_i' \beta$ и дисперсией σ^2 . Таким образом, наблюдаемые случайные переменные не являются н.о.р.: они имеют разные средние. Однако наблюдения считаются независимыми и, как мы выведем далее,

$$\ln L(\theta | y, \mathbf{X}) = \sum_{i=1}^n \ln f(y_i | x_i, \theta) = -\frac{1}{2} \sum_{i=1}^n [\ln \sigma^2 + \ln(2\pi) + (y_i - x_i' \beta)^2 / \sigma^2], \quad (14-3)$$

где \mathbf{X} — $n \times K$ матрица данных с i -й строкой, равной x_i' .

Оставшаяся часть этой главы будет посвящена получению оценок параметров θ и проверке гипотезы об этих параметрах и процессе порождения данных. Перед тем как приступить к этому, рассмотрим вопрос о том, могут ли параметры быть оценены в принципе. Этот вопрос называют проблемой **идентификации** (identification). Вопрос идентификации связан с самой постановкой модели. Он должен быть решен до того, как приступить к самой процедуре оценивания. Смысл проблемы, по сути, состоит в следующем. Допустим, мы располагаем бесконечно большой выборкой, т. е. всей необходимой нам информацией о параметрах. Можно ли в этом случае однозначно определить значения параметров θ ? Как вскоре мы увидим, иногда ответ на этот вопрос отрицательный.

Определение 14.1. Идентификация

Вектор параметров θ идентифицируем (может быть оценен), если для любого другого вектора параметров $\theta^* \neq \theta$ при некотором наборе данных y $L(\theta^* | y) \neq L(\theta | y)$.

Этот результат будет очень важен для некоторых моментов, излагаемых далее. Рассмотрим два примера, первый из которых нам уже хорошо известен.

Пример 14.1. Идентификация параметров

Предположим, что для регрессионной модели (14-3) существует отличный от нуля вектор a такой, что $x_i' a = 0$ для любого x_i . Тогда существует другой вектор «параметров» $\gamma = \beta + a \neq \beta$ такой, что $x_i' \beta = x_i' \gamma$ для любого x_i . Как видно из (14-3), в этом случае логарифм правдоподобия одинаков независимо от того, вычисляется ли он в точке β или в точке γ . Таким образом, рассматривать оценивание β в этой модели невозможно, поскольку β нельзя отличить от γ . Это пример полной коллинеарности в регрессионной модели, которого раньше мы избегали в модели линейной регрессии, сделав «Предположение 2. Идентифицируемость параметров модели».

Предыдущий абзац был связан с некоторыми необходимыми характеристиками данных в выборке. Теперь опишем модель, в которой идентификация гарантируется ее спецификацией. (Подробнее это будет рассмотрено в главе 17.) Представим простую форму регрессионной модели, которая была описана выше: $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$, где $\varepsilon_i | x_i$ имеет нормальное распределение с нулевым средним и дисперсией σ^2 . Чтобы иметь некоторую реальную ситуацию, рассмотрим покупку потребителем большого товара, такого как машина, где x_i — потребительский доход и y_i — разница между суммой p_i^* , которую потребитель готов заплатить за машину, и реальной ценой машины p_i . Предположим, что вместо наблюдения величин p_i^* или p_i мы наблюдаем сам факт покупки машины, который происходит, как мы считаем, когда $y_i = p_i^* - p_i$ является величиной положительной. Таким образом, наша модель утверждает, что машина покупается, если $y_i > 0$, и не покупается, если $y_i \leq 0$. Выразим функцию правдоподобия для наблюдаемых данных, в которые входит факт покупки (или факт ее отсутствия), а также доход покупателя. Зависимая переменная в этой модели — это «покупка» или «непокупка», т. е. существует только два исхода. Вероятность покупки (purchase) определяется следующим образом:

$$\begin{aligned} \text{Prob}(\text{покупки} | \beta_1, \beta_2, \sigma, x_i) &= \text{Prob}(y_i > 0 | \beta_1, \beta_2, \sigma, x_i) \\ &= \text{Prob}(\beta_1 + \beta_2 x_i + \varepsilon_i > 0 | \beta_1, \beta_2, \sigma, x_i) \\ &= \text{Prob}[\varepsilon_i > -(\beta_1 + \beta_2 x_i) | \beta_1, \beta_2, \sigma, x_i] \\ &= \text{Prob}[\varepsilon_i / \sigma > -(\beta_1 + \beta_2 x_i) / \sigma | \beta_1, \beta_2, \sigma, x_i] \\ &= \text{Prob}[z_i > -(\beta_1 + \beta_2 x_i) / \sigma | \beta_1, \beta_2, \sigma, x_i], \end{aligned}$$

где z_i имеет стандартное нормальное распределение. Вероятность того, что покупки не будет, равна просто 1 минус вероятность покупки. Функция правдоподобия записывается как

$$\prod_{\substack{i = \\ \text{куплено}}} [\text{Prob}(\text{покупки} | \beta_1, \beta_2, \sigma, x_i)] \prod_{\substack{i = \text{не} \\ \text{куплено}}} [1 - \text{Prob}(\text{покупки} | \beta_1, \beta_2, \sigma, x_i)].$$

Из последней формулы видно, что параметры этой модели не идентифицируются. Если β_1 , β_2 и σ умножить на одну и ту же отличную от нуля постоянную независимо от ее величины, то $\text{Prob}(\text{purchase})$, равно как и $1 - \text{Prob}(\text{purchase})$, остается прежней и, значит, функция правдоподобия не изменится. В данной модели требуется **нормировка** (normalization). Обычно для этого используют $\sigma = 1$, но ряд авторов [например, Horowitz (1993)] вместо этого используют $\beta_1 = 1$.

14.3. Эффективное оценивание: принцип максимального правдоподобия

Принцип **максимального правдоподобия** (maximum likelihood) дает способ построения асимптотически эффективной оценки для параметра или набора параметров. Логику этой методики можно легко продемонстрировать в случае дискретного распределения. Рассмотрим случайную выборку из следующих 10 наблюдений из распределения Пуассона: 5, 0, 1, 1, 0, 3, 2, 3, 4 и 1. Плотность для каждого наблюдения записывается как

$$f(y_i | \theta) = \frac{e^{-\theta} \theta^{y_i}}{y_i!}.$$

Ввиду того что наблюдения независимы, их совместная плотность, которая является правдоподобием для данной выборки, выглядит как

$$f(y_1, y_2, \dots, y_{10} | \theta) = \prod_{i=1}^{10} f(y_i | \theta) = \frac{e^{-10\theta} \theta^{\sum_{i=1}^{10} y_i}}{\prod_{i=1}^{10} y_i!} = \frac{e^{-10\theta} \theta^{20}}{207,360}.$$

Последний результат дает вероятность наблюдения *этой конкретной выборкой* при предположении, что данные получены из распределения Пуассона с еще пока неизвестным параметром θ . Возникает вопрос: какое значение θ сделало бы эту выборку наиболее вероятной? Рисунок 14.1 графически изображает эту функцию для разных значений θ . Она имеет единственный пик (моду) при $\theta = 2$, и это значение в данном случае является **оценкой максимального правдоподобия** (maximum likelihood estimate, MLE) параметра θ .

Рассмотрим максимизацию функции $L(\theta | y)$ относительно θ . Поскольку логарифмическая функция монотонно возрастает и с ней легче работать,

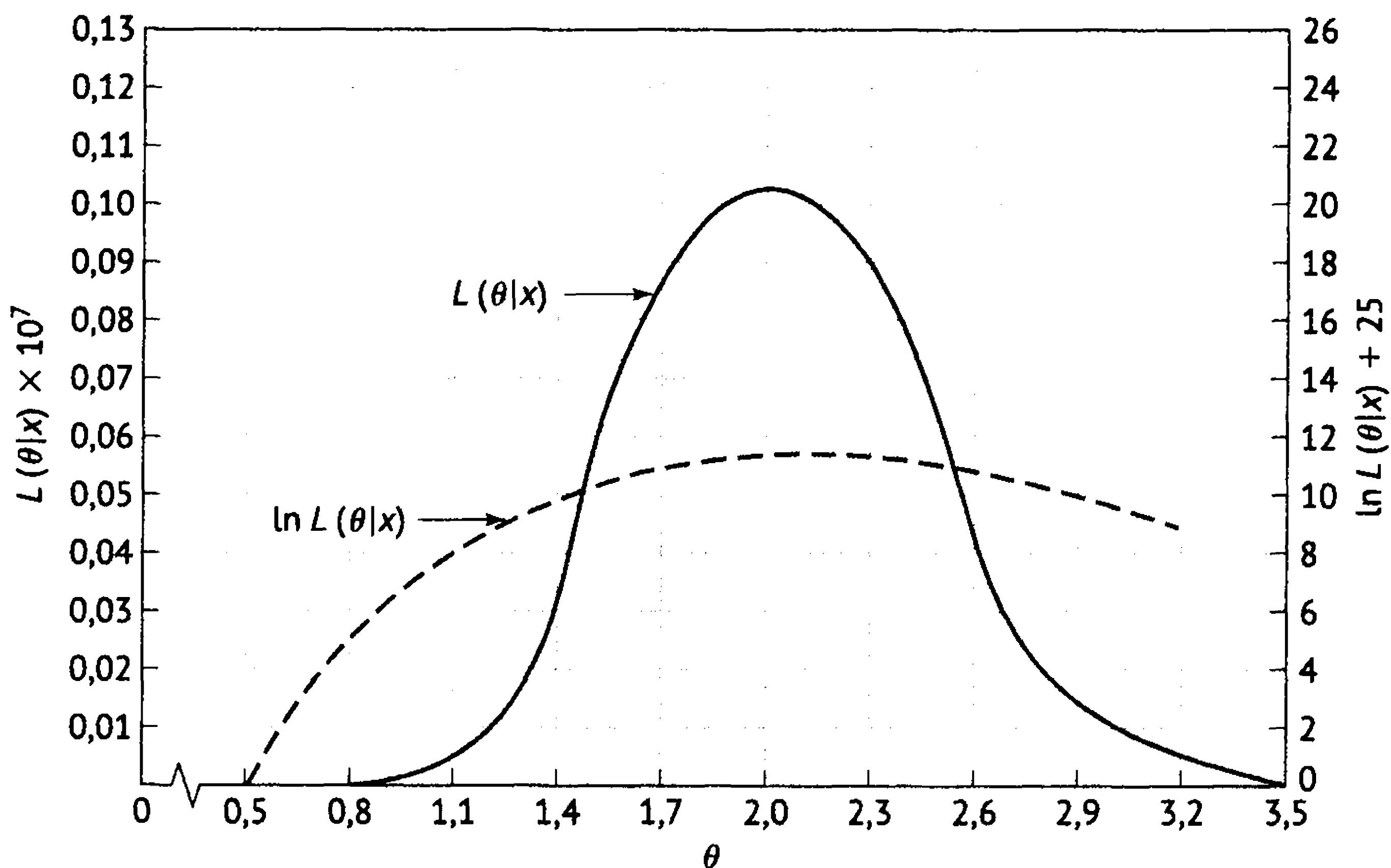


Рис. 14.1. Функция правдоподобия и логарифмическая функция правдоподобия для распределения Пуассона

обычно максимизируют функцию $\ln L(\theta | \mathbf{y})$. Для выборки из распределения Пуассона

$$\ln L(\theta | \mathbf{y}) = -n\theta + \ln \theta \sum_{i=1}^n y_i - \sum_{i=1}^n \ln(y_i!),$$

$$\frac{\partial \ln L(\theta | \mathbf{y})}{\partial \theta} = -n + \frac{1}{\theta} \sum_{i=1}^n y_i = 0 \Rightarrow \hat{\theta}_{\text{ML}} = \bar{y}_n.$$

Для имеющейся у нас выборки

$$\ln L(\theta | \mathbf{y}) = -10\theta + 20 \ln \theta - 12,242,$$

$$\frac{d \ln L(\theta | \mathbf{y})}{d\theta} = -10 + \frac{20}{\theta} = 0 \Rightarrow \hat{\theta} = 2$$

и

$$\frac{d^2 \ln L(\theta | \mathbf{y})}{d\theta^2} = \frac{-20}{\theta^2} < 0 \Rightarrow \text{это максимальное значение.}$$

Мы получили то же решение, что и ранее. Чтобы проиллюстрировать этот результат, на рис. 14.1 также графически представлен логарифм функции $L(\theta | \mathbf{y})$.

Интерпретация вероятности заданной выборки не является корректной для непрерывного распределения, поскольку конкретная выборка имеет вероятность, равную нулю. Тем не менее принцип остается тем же. Значения параметров, которые максимизируют функцию $L(\theta | \text{data})$ или ее логарифм, представляют собой оценки максимального правдоподобия, обозначаемые как $\hat{\theta}$. Логарифм является монотонной функцией, поэтому значения, которые максимизируют функцию $L(\theta | \text{данные})$, максимизируют и

$\ln L(\theta | \text{data})$. Необходимым условием для максимизации функции $\ln L(\theta | \text{data})$ будет уравнение

$$\frac{\partial \ln L(\theta | \text{данные})}{\partial \theta} = 0. \quad (14-4)$$

Оно называется **уравнением правдоподобия** (likelihood equation). Тогда общий результат состоит в том, что оценка MLE есть корень уравнения правдоподобия. Применение этого метода к параметрам дискретного DGP (data generating processes, процесса порождения данных) дает основание предполагать, что этот метод является «хорошим» способом обработки данных. Нам требуется установить это формально в общем случае. Мы вернемся к этому в следующем разделе.

Пример 14.2. *Логарифмическая функция правдоподобия и уравнения правдоподобия для нормального распределения*

Для выборки из нормального распределения со средним μ и дисперсией σ^2 логарифмическая функция правдоподобия и уравнения правдоподобия для μ и σ^2 записываются следующим образом:

$$\ln L(\mu, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 - \frac{1}{2} \sum_{i=1}^n \left[\frac{(y_i - \mu)^2}{\sigma^2} \right], \quad (14-5)$$

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \mu) = 0, \quad (14-6)$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \mu)^2 = 0. \quad (14-7)$$

Для решения этих уравнений умножим (14-6) на σ^2 и найдем $\hat{\mu}$. Затем подставим получившееся решение в (14-7) и найдем σ^2 . Решения выглядят как

$$\hat{\mu}_{\text{ML}} = \frac{1}{n} \sum_{i=1}^n y_i = \bar{y}_n \quad \text{и} \quad \hat{\sigma}_{\text{ML}}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}_n)^2. \quad (14-8)$$

14.4. Свойства оценок максимального правдоподобия

Оценки максимального правдоподобия (MLE, maximum likelihood estimator, ML-оценка) очень привлекательны ввиду их свойств, которые они проявляют на больших выборках, или, иначе, их асимптотических свойств.

Определение 14.2. Асимптотическая эффективность

Оценка называется асимптотически эффективной, если она является состоятельной, асимптотически нормально распределенной (consistent and asymptotically normally distributed, CAN) и имеет асимптотическую ковариационную матрицу, которая не больше, чем асимптотическая ковариационная матрица любой другой состоятельной, асимптотически нормально распределенной оценки².

При выполнении некоторых условий регулярности MLE будет обладать этими свойствами. Свойства выборки конечного размера иногда менее оптимальны. Так, например, оценка MLE может быть смещенной: оценка MLE параметра σ^2 в примере 14.2 смещена вниз. Однако утверждать, что свойства MLE будут оптимальными *только* при больших выборках, тем не менее будет неверно. Можно показать, что если выборка формируется из экспоненциального семейства распределений (см. определение 13.1), то будут существовать достаточные статистики. В этом случае MLE будут функциями от этих статистик, откуда следует, что если существуют несмещенные оценки с минимальной дисперсией, то они являются MLE. [См. работу Stuart and Ord (1989).] Большинство моделей в эконометрике не используют экспоненциальные семейства, поэтому основная привлекательность ML-оценок все же заключается в их асимптотических свойствах.

Мы будем использовать следующие обозначения: $\hat{\theta}$ — это оценка методом максимального правдоподобия, θ_0 — истинное значение вектора параметров, θ — другое возможное значение вектора параметров, не MLE и необязательно истинное значение. Математическое ожидание, основанное на истинных значениях параметров, обозначается как $E_0[\cdot]$. Если мы предположим, что для $f(x, \theta_0)$ выполняются условия регулярности, о которых сейчас шла речь, то можно сформулировать следующую теорему.

Теорема 14.1. Свойства ML-оценки

При условиях регулярности оценка методом максимального правдоподобия (MLE) имеет следующие асимптотические свойства:

M1. Состоятельность (Consistency): $\text{plim } \hat{\theta} = \theta_0$.

M2. Асимптотическая нормальность (Asymptotic normality): $\hat{\theta} \stackrel{a}{\sim} N[\theta_0, \{I(\theta_0)\}^{-1}]$, где

$$I(\theta_0) = -E_0[\partial^2 \ln L / \partial \theta_0 \partial \theta_0'].$$

M3. Асимптотическая эффективность (Asymptotic efficiency): $\hat{\theta}$ асимптотически эффективна и достигает нижней границы Крамера–Рао для состоятельных оценок, задаваемой в условии M2 и теореме С.2.

²«Не больше» понимается в смысле (А-118): ковариационная матрица менее эффективной оценки равна ковариационной матрице эффективной оценки плюс неотрицательно определенная матрица.

М4. Инвариантность (Invariance): оценка методом максимального правдоподобия параметра $\gamma_0 = c(\theta_0)$ равна $c(\hat{\theta})$, если $c(\theta_0)$ является непрерывной и непрерывно дифференцируемой функцией.

14.4.1. Условия регулярности

Чтобы в общих чертах описать доказательства этих результатов, нужно сначала обратиться к некоторым полезным свойствам функций плотностей распределения. Мы предположим, что (y_1, \dots, y_n) представляет собой случайную выборку из генеральной совокупности с функцией плотности $f(y_i | \theta_0)$ и что выполняются следующие **условия регулярности** (regularity conditions). [Наше изложение является несколько неформальным. Более строгую трактовку можно найти в работах Stuart, Ord (1989) или Davidson, MacKinnon (2004).]

Определение 14.3. Условия регулярности

R1. Первые три производные функции $\ln f(y_i | \theta)$ относительно θ непрерывны и конечны для почти всех y_i для любого θ . Это условие гарантирует существование определенной аппроксимации рядом Тейлора и конечной дисперсии производных функции $\ln L$.

R2. Выполнены условия, необходимые для получения математических ожиданий первой и второй производных функции $\ln f(y_i | \theta)$.

R3. Для всех значений параметра θ функция $|\partial^3 \ln f(y_i | \theta) / \partial \theta_j \partial \theta_k \partial \theta_l|$ меньше, чем функция, имеющая конечное значение математического ожидания. Это условие позволит нам рассматривать усеченные ряды Тейлора.

Используя эти условия регулярности, мы можем получить нижеследующие основные характеристики функции $f(y_i | \theta)$. D1 является просто следствием из определения функции правдоподобия. D2 ведет к моментному условию, которое определяет оценку максимального правдоподобия. С одной стороны, MLE определяется как точка максимума функции, что дает возможность просто найти вектор, приравнивающий градиент к нулю. С другой стороны, D2 отражает более существенную зависимость, утверждая, что MLE относится к классу оценок обобщенным методом моментов. D3 выводит то, что известно как **информационное равенство** (information matrix equality). Из этого соотношения можно получить асимптотическую ковариационную матрицу для MLE.

14.4.2. Свойства регулярных распределений

Плотности, являющиеся «регулярными» по определению 14.3, обладают тремя свойствами, которые используются для установления свойств оценок методом максимального правдоподобия.

Теорема 14.2. *Моменты производных логарифмического правдоподобия*
D1. Все величины $\ln f(y_i | \theta)$, $\mathbf{g}_i = \partial \ln f(y_i | \theta) / \partial \theta$ и $\mathbf{H}_i = \partial^2 \ln f(y_i | \theta) / \partial \theta \partial \theta'$, $i = 1, \dots, n$, представляют собой случайные выборки случайных величин. Это утверждение следует из нашего предположения о случайной выборке. Обозначения $\mathbf{g}_i(\theta_0)$ и $\mathbf{H}_i(\theta_0)$ относятся к производной, вычисленной в точке θ_0 .

D2. $E_0[\mathbf{g}_i(\theta_0)] = \mathbf{0}$.

D3. $\text{Var}[\mathbf{g}_i(\theta_0)] = -E[\mathbf{H}_i(\theta_0)]$.

Условие D1 просто является следствием из определения плотности вероятности.

На данном этапе мы допускаем, что носитель распределения y_i зависит от параметров $A(\theta_0) \leq y_i \leq B(\theta_0)$. (Рассмотрим, например, нахождение оценки методом максимального правдоподобия параметра θ_0 для непрерывного равномерного распределения на отрезке $[0, \theta_0]$.) (В дальнейшем, если необходимо, обычный интеграл $\int \dots dy_i$ будет использоваться для обозначения многократного интегрирования по всем элементам многомерного y_i .) По определению

$$\int_{A(\theta_0)}^{B(\theta_0)} f(y_i | \theta_0) dy_i = 1.$$

Теперь продифференцируем это выражение по θ_0 . Из теоремы Лейбница следует

$$\begin{aligned} \frac{\partial \int_{A(\theta_0)}^{B(\theta_0)} f(y_i | \theta_0) dy_i}{\partial \theta_0} &= \int_{A(\theta_0)}^{B(\theta_0)} \frac{\partial f(y_i | \theta_0)}{\partial \theta_0} dy_i + f(B(\theta_0) | \theta_0) \frac{\partial B(\theta_0)}{\partial \theta_0} - \\ &- f(A(\theta_0) | \theta_0) \frac{\partial A(\theta_0)}{\partial \theta_0} = 0. \end{aligned}$$

Если второе и третье выражения стремятся к нулю, то мы можем поменять местами операции дифференцирования и интегрирования. Необходимое условие заключается в том, что $\lim_{y_i \rightarrow A(\theta_0)} f(y_i | \theta_0) = \lim_{y_i \rightarrow B(\theta_0)} f(y_i | \theta_0) = 0$. (Обратим внимание, что для равномерного распределения, предложенного выше, это условие нарушается.) Достаточное условие состоит в том, что носитель наблюдаемой случайной переменной y_i не зависит от параметров, т. е. либо $\partial A(\theta_0) / \partial \theta_0 = \partial B(\theta_0) / \partial \theta_0 = 0$, либо плотность равна нулю в конечных точках. Тогда это условие является условием регулярности R2. Обычно предполагается выполнение второго условия, и мы в дальнейшем также будем его придерживаться. Таким образом,

$$\begin{aligned} \frac{\partial \int f(y_i | \theta_0) dy_i}{\partial \theta_0} &= \int \frac{\partial f(y_i | \theta_0)}{\partial \theta_0} dy_i = \int \frac{\partial \ln f(y_i | \theta_0)}{\partial \theta_0} f(y_i | \theta_0) dy_i = \\ &= E_0 \left[\frac{\partial \ln f(y_i | \theta_0)}{\partial \theta_0} \right] = 0. \end{aligned}$$

Это доказывает утверждение D2.

Поскольку операции дифференцирования и интегрирования можно поменять местами, мы проводим дифференцирование под интегралом еще раз, чтобы получить

$$\int \left[\frac{\partial^2 \ln f(y_i | \theta_0)}{\partial \theta_0 \partial \theta'_0} f(y_i | \theta_0) + \frac{\partial \ln f(y_i | \theta_0)}{\partial \theta_0} \frac{\partial f(y_i | \theta_0)}{\partial \theta'_0} \right] dy_i = 0.$$

Но

$$\frac{\partial f(y_i | \theta_0)}{\partial \theta'_0} = f(y_i | \theta_0) \frac{\partial \ln f(y_i | \theta_0)}{\partial \theta'_0}$$

и интеграл суммы есть сумма интегралов. Таким образом,

$$-\int \left[\frac{\partial^2 \ln f(y_i | \theta_0)}{\partial \theta_0 \partial \theta'_0} \right] f(y_i | \theta_0) dy_i = \int \left[\frac{\partial \ln f(y_i | \theta_0)}{\partial \theta_0} \frac{\partial \ln f(y_i | \theta_0)}{\partial \theta'_0} \right] f(y_i | \theta_0) dy_i.$$

Левая сторона уравнения есть взятое со знаком минус математическое ожидание матрицы вторых производных. Правая сторона — это математическое ожидание квадрата (векторное произведение) вектора первых производных. Но поскольку математическое ожидание этого вектора есть $\mathbf{0}$ (мы уже показали это), правая сторона представляет собой дисперсию вектора первых производных, что доказывает D3:

$$\begin{aligned} \text{Var}_0 \left[\frac{\partial \ln f(y_i | \theta_0)}{\partial \theta_0} \right] &= E_0 \left[\left(\frac{\partial \ln f(y_i | \theta_0)}{\partial \theta_0} \right) \left(\frac{\partial \ln f(y_i | \theta_0)}{\partial \theta'_0} \right) \right] = \\ &= -E \left[\frac{\partial^2 \ln f(y_i | \theta_0)}{\partial \theta_0 \partial \theta'_0} \right]. \end{aligned}$$

14.4.3. Уравнение правдоподобия

Логарифмическая функция правдоподобия записывается как

$$\ln L(\theta | \mathbf{y}) = \sum_{i=1}^n \ln f(y_i | \theta).$$

Вектор первых производных, или скор-вектор (score vector), или **градиент**, равен

$$\mathbf{g} = \frac{\partial \ln L(\theta | \mathbf{y})}{\partial \theta} = \sum_{i=1}^n \frac{\partial \ln f(y_i | \theta)}{\partial \theta} = \sum_{i=1}^n \mathbf{g}_i. \quad (14-9)$$

Поскольку мы просто добавляем слагаемые, из D1 и D2 следует, что в точке θ_0

$$E_0 \left[\frac{\partial \ln L(\theta_0 | \mathbf{y})}{\partial \theta_0} \right] = E_0[\mathbf{g}_0] = \mathbf{0}, \quad (14-10)$$

что и является упомянутым выше **уравнением правдоподобия** (likelihood equation).

14.4.4. Информационное равенство

Гессиан логарифмического правдоподобия записывается как

$$\mathbf{H} = \frac{\partial^2 \ln L(\boldsymbol{\theta} | \mathbf{y})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \sum_{i=1}^n \frac{\partial^2 \ln f(y_i | \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \sum_{i=1}^n \mathbf{H}_i.$$

Вычисляя его все в той же точке $\boldsymbol{\theta}_0$, используя

$$E_0[\mathbf{g}_0 \mathbf{g}_0'] = E_0 \left[\sum_{i=1}^n \sum_{j=1}^n \mathbf{g}_{0i} \mathbf{g}_{0j}' \right]$$

и, как следует из D1, отбрасывая члены с несовпадающими индексами, мы получаем

$$E_0[\mathbf{g}_0 \mathbf{g}_0'] = E_0 \left[\sum_{i=1}^n \mathbf{g}_{0i} \mathbf{g}_{0i}' \right] = E_0 \left[\sum_{i=1}^n (-\mathbf{H}_{0i}) \right] = -E_{t0}[\mathbf{H}_0],$$

так что

$$\begin{aligned} \text{Var}_0 \left[\frac{\partial \ln L(\boldsymbol{\theta}_0 | \mathbf{y})}{\partial \boldsymbol{\theta}_0} \right] &= E_0 \left[\left(\frac{\partial \ln L(\boldsymbol{\theta}_0 | \mathbf{y})}{\partial \boldsymbol{\theta}_0} \right) \left(\frac{\partial \ln L(\boldsymbol{\theta}_0 | \mathbf{y})}{\partial \boldsymbol{\theta}_0'} \right) \right] \\ &= -E_0 \left[\frac{\partial^2 \ln L(\boldsymbol{\theta}_0 | \mathbf{y})}{\partial \boldsymbol{\theta}_0 \partial \boldsymbol{\theta}_0'} \right]. \end{aligned} \quad (14-11)$$

Этот очень полезный результат известен как **информационное равенство** (information matrix equality).

14.4.5. Асимптотические свойства оценки методом максимального правдоподобия

Теперь мы можем в общих чертах описать, как получаются асимптотические свойства MLE. Формальные доказательства этих результатов требуют довольно сложных математических конструкций. Две наиболее широко цитируемые работы, содержащие этот вывод, — это работы Крамера (Cramér (1948)) и Амеии (Amemiya (1985)). Чтобы почувствовать вкус исследования, кратко изложим анализ, сделанный в работе Стюарта и Орда (Stuart, Ord (1989)) для простого случая, и укажем, где необходимо обобщить рассуждения, чтобы получить наиболее общий результат.

14.4.5а. Состоятельность

Мы предположим, что $f(y_i | \boldsymbol{\theta}_0)$, возможно, является многомерной плотностью, но на данном этапе не зависит от независимых переменных x_i . Таким образом, это пример н.о.р. случайной выборки. Поскольку $\hat{\boldsymbol{\theta}}$ представляет собой MLE, в любой выборке конечного объема для любого $\boldsymbol{\theta} \neq \frac{1}{\theta}$ (включая истинное значение $\boldsymbol{\theta}_0$) выполняется

$$\ln L(\hat{\boldsymbol{\theta}}) \geq \ln L(\boldsymbol{\theta}). \quad (14-12)$$

Рассмотрим случайную величину $L(\theta)/L(\theta_0)$. Поскольку логарифмическая функция является строго вогнутой (выпуклой кверху), из неравенства Йенсена (теорема D.13.) следует, что

$$E_0 \left[\ln \frac{L(\theta)}{L(\theta_0)} \right] < \ln E_0 \left[\frac{L(\theta)}{L(\theta_0)} \right]. \quad (14-13)$$

Математическое ожидание правой части в точности равно единице, так как

$$E_0 \left[\frac{L(\theta)}{L(\theta_0)} \right] = \int \left(\frac{L(\theta)}{L(\theta_0)} \right) L(\theta_0) dy = 1 \quad (14-14)$$

просто представляет собой интеграл от совместной плотности. Таким образом, правая часть уравнения (14-13) равна нулю. Поделив левую часть уравнения (14-13) на n , получаем

$$E_0[1/n \ln L(\theta)] - E_0[1/n \ln L(\theta_0)] < 0.$$

Отсюда получаем основной результат.

Теорема 14.3. Неравенство правдоподобия

$$E_0[(1/n) \ln L(\theta_0)] > E_0[(1/n) \ln L(\theta)] \quad \text{для любого } \theta \neq \theta_0 \text{ (включая } \hat{\theta}).$$

Это означает, что математическое значение логарифмического правдоподобия максимизируется при истинном значении параметров.

Для любого значения θ , включая $\hat{\theta}$, выражение

$$[(1/n) \ln L(\theta)] = (1/n) \sum_{i=1}^n \ln f(y_i | \theta)$$

представляет собой выборочное среднее n н.о.р. случайных переменных с математическим ожиданием $E_0[(1/n) \ln L(\theta)]$. Поскольку выборка является н.о.р. согласно условиям регулярности, мы можем применить теорему Хинчина D.5: выборочное среднее значение сходится по вероятности к среднему значению генеральной совокупности. Взяв $\theta = \hat{\theta}$ из теоремы 14.3, мы получаем, что при $n \rightarrow \infty$, $\lim \text{Prob}\{[(1/n) \ln L(\hat{\theta})] < [(1/n) \ln L(\theta_0)]\} = 1$ для $\hat{\theta} \neq \theta_0$. Но $\hat{\theta}$ является MLE, поэтому для каждого n $(1/n) \ln L(\hat{\theta}) \geq (1/n) \ln L(\theta_0)$. Единственный вариант, при котором могут выполняться оба этих соотношения, состоит в том, что вычисленное в MLE логарифмическое правдоподобие выборки, умноженное на $(1/n)$, сходится к математическому ожиданию логарифмического правдоподобия генеральной совокупности, оцененному в истинных параметрах и умноженному на $(1/n)$. Остается один последний шаг. Следует ли из того, что $(1/n) \ln L(\hat{\theta}) \rightarrow (1/n) \ln L(\theta_0)$, что $\hat{\theta} \rightarrow \theta_0$? Если параметр всего один, а функция правдоподобия взаимно однозначна, то, очевидно, так и есть. Для более общих случаев требуется дальнейшая характеристика функции правдоподобия. Если функция правдоподобия строго

непрерывна и дважды дифференцируема, что мы предполагали в условиях регулярности, и если параметры модели идентифицируются, что мы предположили в начале нашего обсуждения, то ответ на поставленный вопрос положительный и центральный результат верен.

Это доказательство является эвристическим. Как мы уже заметили выше, строгое изложение можно найти в более сложных текстах, чем эта книга. Также следует отметить, что в нескольких местах мы предполагали, что выборочные средние сходятся к математическим ожиданиям их распределений. Это обычно верно для моделей, рассматриваемых в эконометрике, но полностью общий результат требовал бы отнестись к этому условию более аккуратно. Далее мы сделали предположение о н.о.р.- выборке, т. е. о том, что плотность y_i не зависит от любых других переменных x_i . На практике это почти никогда не выполняется. В доказательство потребуется включить предположения о поведении этих переменных. Например, при изучении асимптотического поведения оценки методом наименьших квадратов мы предполагали, что данные являются доброкачественными. Подобные предположения требуются и здесь. Вскоре мы вернемся к этому вопросу. Имея в виду все эти соображения, мы получаем свойство M1: $\text{plim } \hat{\theta} = \theta_0$.

14.4.5b. Асимптотическая нормальность

В оценке методом максимального правдоподобия градиент логарифмического правдоподобия равен нулю (по определению), т. е.

$$g(\hat{\theta}) = 0.$$

(Это выборочная статистика, но не математическое ожидание.) Разложим эту систему уравнений с помощью рядов Тейлора около истинных параметров θ_0 . Используем теорему о среднем значении, чтобы урезать ряды Тейлора во втором члене:

$$g(\hat{\theta}) = g(\theta_0) + \mathbf{H}(\bar{\theta})(\hat{\theta} - \theta_0) = 0.$$

Гессиан вычисляется в точке $\bar{\theta}$, которая находится между $\hat{\theta}$ и θ_0 [$\bar{\theta} = w\hat{\theta} + (1-w)\theta_0$ для некоторого $0 < w < 1$]. Затем преобразуем эту функцию и умножим ее на \sqrt{n} , чтобы получить

$$\sqrt{n}(\hat{\theta} - \theta_0) = [-\mathbf{H}(\bar{\theta})]^{-1}[\sqrt{n}g(\theta_0)].$$

Поскольку $\text{plim}(\hat{\theta} - \theta_0) = 0$, так же и $\text{plim}(\bar{\theta} - \theta_0) = 0$. Вторые производные являются непрерывными функциями. Таким образом, если предельное распределение существует, то

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} [-\mathbf{H}(\theta_0)]^{-1}[\sqrt{n}g(\theta_0)].$$

Поделив $\mathbf{H}(\theta_0)$ и $g(\theta_0)$ на n , получаем

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} \left[-\frac{1}{n}\mathbf{H}(\theta_0)\right]^{-1}[\sqrt{n}\bar{g}(\theta_0)]. \quad (14-15)$$

Мы можем применить здесь центральную предельную теорему Линдберга–Леви (D.18) к величине $[\sqrt{n}\bar{g}(\theta_0)]$, поскольку она является средним значением случайной выборки, умноженным на \sqrt{n} : здесь мы снова используем D1. Предельная дисперсия $[\sqrt{n}\bar{g}(\theta_0)]$ равняется $-E_0[(1/n)\mathbf{H}(\theta_0)]$, поэтому

$$\sqrt{n}\bar{g}(\theta_0) \xrightarrow{d} N\{\mathbf{0}, -E_0[\frac{1}{n}\mathbf{H}(\theta_0)]\}.$$

На основании теоремы D.2 $\text{plim}[-(1/n)\mathbf{H}(\theta_0)] = -E_0[(1/n)\mathbf{H}(\theta_0)]$. Этот результат представляет собой постоянную матрицу, поэтому, объединяя эти результаты, получаем

$$\begin{aligned} [-\frac{1}{n}\mathbf{H}(\theta_0)]^{-1}\sqrt{n}\bar{g}(\theta_0) &\xrightarrow{d} N[\mathbf{0}, \{-E_0[\frac{1}{n}\mathbf{H}(\theta_0)]\}^{-1}\{-E_0[\frac{1}{n}\mathbf{H}(\theta_0)]\} \times \\ &\times \{-E_{t_0}[\frac{1}{n}\mathbf{H}(\theta_0)]\}^{-1}], \end{aligned}$$

или

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N[\mathbf{0}, \{-E_0[\frac{1}{n}\mathbf{H}(\theta_0)]\}^{-1}],$$

что приводит к асимптотическому распределению оценки ML:

$$\hat{\theta} \stackrel{a}{\sim} N[\theta_0, \{\mathbf{I}(\theta_0)\}^{-1}].$$

Этот последний шаг и завершает доказательство M2.

Пример 14.3. Информационная матрица для нормального распределения
Для функции правдоподобия в примере 14.2 вторыми производными являются

$$\begin{aligned} \frac{\partial^2 \ln L}{\partial \mu^2} &= \frac{-n}{\sigma^2}, \\ \frac{\partial^2 \ln L}{\partial (\sigma^2)^2} &= \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^n (y_i - \mu)^2, \\ \frac{\partial^2 \ln L}{\partial \mu \partial \sigma^2} &= \frac{-1}{\sigma^4} \sum_{i=1}^n (y_i - \mu). \end{aligned}$$

Для вычисления **асимптотической дисперсии** (asymptotic variance) оценки методом максимального правдоподобия нам нужны математические ожидания этих производных. Первая из них является неслучайной, а третья имеет математическое ожидание, равное нулю, как $E[y_i] = \mu$. Остается вторая, которая, как можно проверить, имеет математическое ожидание $-n/(2\sigma^4)$, поскольку каждое из n выражений $(y_i - \mu)^2$ имеет математическое ожидание σ^2 . Подставляя их в информационную матрицу, поменяв знак и обратив матрицу, мы получим асимптотическую ковариационную матрицу для оценки методом максимального правдоподобия:

$$\left\{ -E_0 \left[\frac{\partial^2 \ln L}{\partial \theta_0 \partial \theta_0'} \right] \right\}^{-1} = \begin{bmatrix} \sigma^2/n & 0 \\ 0 & 2\sigma^4/n \end{bmatrix}.$$

14.4.5c. Асимптотическая эффективность

Теорема С.2 дает нижнюю границу для дисперсии несмещенной оценки. Поскольку асимптотическая дисперсия MLE достигает этой границы, кажется естественным сформулировать соответствующий результат для этой оценки. Однако проблема здесь заключается в том, что оценка ML почти никогда не является несмещенной. Таким образом, нам нужен асимптотический вариант этой границы, который был представлен в работах Cramér (1948) и Rao (1945) (откуда и произошло название).

Теорема 14.4. Нижняя граница Крамера–Рао

Предполагая, что плотность y_i удовлетворяет условиям регулярности R1–R3, асимптотическая дисперсия состоятельной и асимптотически нормально распределенной оценки вектора параметров θ_0 всегда будет не менее чем

$$[\mathbf{I}(\theta_0)]^{-1} = \left(-E_0 \left[\frac{\partial^2 \ln L(\theta_0)}{\partial \theta_0 \partial \theta_0'} \right] \right)^{-1} = \left(E_0 \left[\left(\frac{\partial \ln L(\theta_0)}{\partial \theta_0} \right) \left(\frac{\partial \ln L(\theta_0)}{\partial \theta_0} \right)' \right] \right)^{-1}.$$

Асимптотическая дисперсия MLE и в самом деле равна нижней границе Крамера–Рао для дисперсии состоятельной, асимптотически нормально распределенной оценки, что завершает доказательство³.

14.4.5d. Инвариантность

Последнее свойство, свойство инвариантности M4, представляет собой математический результат расчета оценок методом ML. В принципе это не статистический результат. Более точно, оценка ML инвариантна к взаимно однозначным преобразованиям параметра θ . Любое преобразование, не являющееся взаимно однозначным, либо приводит модель, не поддающуюся оценке (случай, когда одному значению соответствует несколько значений), либо налагает некоторые ограничения (случай, когда много значений соответствуют одному). Некоторые теоретические моменты, касающиеся этого свойства, рассмотрены в книге Дэвидсона и Маккиннона (Davidson, MacKinnon (2004, p. 446, 539–540)). Для специалистов-практиков этот результат может быть чрезвычайно полезным. Так, например, если параметр участвует в функции правдоподобия в выражении $1/\theta_j$, обычно удобно перепараметризовать модель в терминах $\gamma_j = 1/\theta_j$. Важным практическим примером, где использовался этот прием, является работа Олсена (Olsen (1978)). (См. подраздел 19.3.3.) Пусть нормальное логарифмическое правдоподобие из

³Результат, впервые представленный в работе LeCam (1953) и подробно изложенный в работе Amemiya (1985, p. 124), гласит, что в принципе существуют CAN-функции от данных, имеющие меньшую дисперсию, чем MLE. Однако подобные случаи являются очень узкими и не имеют практического применения. Для практических целей можно довольствоваться сформулированным здесь утверждением.

примера 14.2 параметризовано в терминах **параметра точности** (precision parameter) $\theta^2 = 1/\sigma^2$. Тогда логарифмическое правдоподобие имеет вид

$$\ln L(\mu, \theta^2) = -(n/2) \ln(2\pi) + (n/2) \ln \theta^2 - \frac{\theta^2}{2} \sum_{i=1}^n (y_i - \mu)^2.$$

ML-оценка для μ , очевидно, по-прежнему будет равна \bar{x} . Но уравнение правдоподобия для θ^2 теперь будет иметь вид

$$\partial \ln L(\mu, \theta^2) / \partial \theta^2 = \frac{1}{2} \left[n/\theta^2 - \sum_{i=1}^n (y_i - \mu)^2 \right] = 0,$$

что, как и ожидалось, приводит к решению $\hat{\theta}^2 = n / \sum_{i=1}^n (y_i - \hat{\mu})^2 = 1/\hat{\sigma}^2$. Существует и второе следствие из этого свойства. Если требуется проанализировать функцию от MLE, то она сама также будет ML-оценкой.

14.4.5e. Заключение

Эти четыре свойства объясняют, почему оценки методом максимального правдоподобия преобладают в эконометрике. Второе свойство сильно способствует проверке гипотез и построению интервальных оценок. Третье свойство представляет собой особенно сильный результат. MLE обладает наименьшей дисперсией, которая возможна для состоятельной и асимптотически нормально распределенной оценки.

14.4.6. Оценка асимптотической дисперсии оценки методом максимального правдоподобия

Асимптотическая ковариационная матрица оценки методом максимального правдоподобия является матрицей параметров, которые требуется оценить (т. е. представляет собой функцию от θ_0 , которая оценивается). Если вид математических ожиданий значений вторых производных логарифмического правдоподобия известен, то матрица

$$[\mathbf{I}(\theta_0)]^{-1} = \left\{ -E_0 \left[\frac{\partial^2 \ln L(\theta_0)}{\partial \theta_0 \partial \theta_0'} \right] \right\}^{-1} \quad (14-16)$$

может быть вычислена в точке $\hat{\theta}$ для получения оценки ковариационной матрицы для MLE. Однако эта оценка редко доступна. Вторые производные функции логарифмического правдоподобия почти всегда являются сложными нелинейными функциями от данных, математические ожидания которых неизвестны. Тем не менее существует две альтернативы. Вторая оценка матрицы имеет вид

$$[\hat{\mathbf{I}}(1\hat{\theta})]^{-1} = \left(-\frac{\partial^2 \ln L(1\hat{\theta})}{\partial 1\hat{\theta} \partial 1\hat{\theta}'} \right)^{-1}. \quad (14-17)$$

Эта оценка вычисляется с помощью матрицы существующих (а не ожидаемых) вторых производных логарифмической функции правдоподобия

в оценке методом максимального правдоподобия. Можно непосредственно показать, что она равна оценке ожидаемых вторых производных плотности с выборочным средним этой величины. Это можно проверить с помощью теоремы D.4 и результата (D-5). Единственный недостаток этой оценки заключается в том, что вторые производные может оказаться сложно вывести или находить численно на компьютере. Третья оценка основывается на следствии D3 к теореме 14.2, согласно которому матрица математических ожиданий вторых производных является ковариационной матрицей вектора первых производных. Эта оценка представляет собой

$$[\hat{\mathbf{I}}(\hat{\theta})]^{-1} = \left[\sum_{i=1}^n \hat{\mathbf{g}}_i \hat{\mathbf{g}}_i' \right]^{-1} = [\hat{\mathbf{G}}' \hat{\mathbf{G}}]^{-1}, \quad (14-18)$$

где

$$\hat{\mathbf{g}}_i = \frac{\partial \ln f(\mathbf{x}_i, \hat{\theta})}{\partial \hat{\theta}}$$

и

$$\hat{\mathbf{G}} = [\hat{\mathbf{g}}_1, \hat{\mathbf{g}}_2, \dots, \hat{\mathbf{g}}_n]'.$$

$\hat{\mathbf{G}}$ является $n \times K$ матрицей с i -й строкой, равной транспонированному i -му вектору производных логарифмической функции правдоподобия. В случае единственного параметра эта оценка будет просто обратной величиной для суммы квадратов первых производных. Эта оценка чрезвычайно удобна в большинстве случаев, так как не требует никаких расчетов помимо тех, которые требуются для решения уравнения правдоподобия. Она также обладает хорошим свойством — всегда является неотрицательно определенной. Для некоторых очень сложных логарифмических функций правдоподобия, иногда в связи с ошибкой округления, наблюдаемый гессиан может быть неопределенным, даже будет вычисленным в максимуме функции. Оценка (14-18) известна как **ВНН-оценка**⁴ и оценка **внешнего произведения градиентов** (outer product of gradients, **OPG-оценка**).

Ни одна из трех представленных здесь оценок не является более предпочтительной по сравнению с другими в статистическом смысле: все они асимптотически эквивалентны. Почти во всех случаях легче всего рассчитать ВНН-оценку. Однако требуется сделать одну оговорку. Как показывает следующий пример, эти оценки могут давать разные результаты в конечной выборке. Это неизбежная проблема конечных выборок, которая иногда может приводить к разным статистическим выводам. Пример является подходящим случаем, чтобы продемонстрировать это. Используя обычную процедуру, мы должны были бы отвергнуть гипотезу $\beta = 0$, если бы использовали любую из первых двух оценок дисперсии, но не отвергнуть в случае использования третьей. Оценку (14-16), как правило, получить невозможно, так как точное математическое ожидание гессиана редко бывает известно. Из имеющегося примера следует, что в малых и ограниченных выборках оценка (14-17) () является более предпочтительной.

⁴Впервые плюсы этой оценки обосновывались в эконометрической работе Berndt et al. (1974).

Пример 14.4. Оценки для дисперсии MLE

Данные в примере С.1 были сгенерированы при помощи модели вида

$$f(y_i, x_i, \beta) = \frac{1}{\beta + x_i} e^{-y_i/(\beta + x_i)},$$

где $y = \text{income}$ (доход), а $x = \text{education}$ (образование). Чтобы вычислить оценку максимального правдоподобия β , мы должны максимизировать

$$\ln L(\beta) = - \sum_{i=1}^n \ln(\beta + x_i) - \sum_{i=1}^n \frac{y_i}{\beta + x_i}.$$

Уравнение правдоподобия представляет собой

$$\frac{\partial \ln L(\beta)}{\partial \beta} = - \sum_{i=1}^n \frac{1}{\beta + x_i} + \sum_{i=1}^n \frac{y_i}{(\beta + x_i)^2} = 0, \quad (14-19)$$

откуда получаем решение $\hat{\beta} = 15,602727$. Чтобы вычислить асимптотическую дисперсию MLE, нам потребуется

$$\frac{\partial^2 \ln L(\beta)}{\partial \beta^2} = \sum_{i=1}^n \frac{1}{(\beta + x_i)^2} - 2 \sum_{i=1}^n \frac{y_i}{(\beta + x_i)^3}. \quad (14-20)$$

Поскольку функция $E(y_i) = \beta + x_i$ известна, известен и точный вид математического ожидания величины (14-20). Подставляя $\hat{\beta} + x_i$ вместо y_i в (14-20), смена знака и обращение получившейся величины приводит к первой оценке дисперсии 44,2546. Простая подстановка $\hat{\beta} = 15,602727$ в (14-20), смена знака и обращение получившейся величины дают вторую оценку 46,16337. Наконец, взяв обратную величину к сумме квадратов первых производных плотностей, вычисленных в точке $\hat{\beta}$, т. е.

$$[\hat{\mathbf{I}}(\hat{\beta})]^{-1} = \frac{1}{\sum_{i=1}^n [-1/(\hat{\beta} + x_i) + y_i/(\hat{\beta} + x_i)^2]^2},$$

мы получаем ВНН-оценку, равную 100,5116.

14.5. Условные функции правдоподобия, эконометрические модели и GMM-оценка

Все предыдущие результаты формируют статистическую базу для оценивания методом максимального правдоподобия. Но для наших целей не хватает одного важного элемента. Мы провели анализ, рассуждая в терминах плотности наблюдаемой случайной величины и вектора параметров $f(y_i | \alpha)$. Но эконометрические модели включают также экзогенные или предопределенные переменные x_i , поэтому требуется обобщить имеющиеся результа-

ты. Подход здесь состоит в том, чтобы рассматривать эту постановку задачи так же, как было уже проделано в главе 4, где нами рассматривались асимптотические свойства линейной регрессионной модели. Таким образом, за \mathbf{x}_i мы обозначим ряд постоянных и случайных переменных, которые определяют условную плотность y_i . Если разделить совместную плотность y_i и \mathbf{x}_i на произведение условной и маргинальной плотностей, то логарифмическая функция правдоподобия может быть записана как

$$\ln L(\boldsymbol{\alpha} \mid \text{data}) = \sum_{i=1}^n \ln f(y_i, \mathbf{x}_i \mid \boldsymbol{\alpha}) = \sum_{i=1}^n \ln f(y_i \mid \mathbf{x}_i, \boldsymbol{\alpha}) + \sum_{i=1}^n \ln g(\mathbf{x}_i \mid \boldsymbol{\alpha}),$$

где любые нестохастические элементы \mathbf{x}_i , такие как временной тренд или дамми-переменная, считаются константами. Далее, как и прежде, мы предполагаем, что процесс, порождающий \mathbf{x}_i , задается вне интересующей нас модели. Для нас это означает, что параметры, которые появляются в функции $g(\mathbf{x}_i \mid \boldsymbol{\alpha})$, не пересекаются с теми, которые содержатся в функции $f(y_i \mid \mathbf{x}_i, \boldsymbol{\alpha})$. Таким образом, мы разбиваем $\boldsymbol{\alpha}$ на две части $[\boldsymbol{\theta}, \boldsymbol{\delta}]$ так, чтобы логарифмическая функция правдоподобия могла быть записана в виде

$$\ln L(\boldsymbol{\theta}, \boldsymbol{\delta} \mid \text{data}) = \sum_{i=1}^n \ln f(y_i, \mathbf{x}_i \mid \boldsymbol{\alpha}) = \sum_{i=1}^n \ln f(y_i \mid \mathbf{x}_i, \boldsymbol{\theta}) + \sum_{i=1}^n \ln g(\mathbf{x}_i \mid \boldsymbol{\delta}).$$

Пока у векторов $\boldsymbol{\theta}$ и $\boldsymbol{\delta}$ нет общих элементов и ограничений (таких как $\boldsymbol{\theta} + \boldsymbol{\delta} = 1$), две части логарифмического правдоподобия могут быть исследованы по отдельности. В большинстве случаев маргинальное распределение \mathbf{x}_i представляет лишь второстепенный интерес (или вообще не представляет интереса).

Асимптотические результаты для условной оценки максимального правдоподобия теперь должны учитывать наличие \mathbf{x}_i в функции $\ln f(y_i \mid \mathbf{x}_i, \boldsymbol{\theta})$ и ее производных. В дальнейшем изложении мы будем исходить из «доброкачества» данных, поэтому выборочные средние, такие как величина

$$(1/n) \ln L(\boldsymbol{\theta} \mid \mathbf{y}, \mathbf{X}) = \frac{1}{n} \sum_{i=1}^n \ln f(y_i \mid \mathbf{x}_i, \boldsymbol{\theta})$$

и ее градиент по $\boldsymbol{\theta}$, сходятся по вероятности к математическим ожиданиям своих популяционных аналогов. Нам также потребуется использовать центральные предельные теоремы, чтобы установить асимптотическую нормальность градиента функции логарифмического правдоподобия и дать описание самой оценки ML. Мы опускаем этот момент, отсылая читателя к более сложным изложениям этого вопроса, таким как работа Амемии (Amemiya (1985)) или Ньюи, Макфаддена (Newey, McFadden (1994)). В них установлены точные условия и тонкие моменты, которые требуются для установления «обычных» свойств оценок методом максимального правдоподобия. Для нашего изложения (а также очень большого количества эмпирических приложений) будет достаточно следующих предположений.

- **Пространство параметров (Parameter space).** Пространства параметров, которые имеют промежутки и невыпуклости, вообще говоря, не да-

ют возможности осуществлять рассматриваемые процедуры. К подобным задачам относится «оценивание» параметра, который может принимать только одно значение среди некоторого дискретного множества. Так, например, этим методом невозможно «оценивание» момента структурного сдвига модели. Функция правдоподобия должна быть непрерывной функцией от выпуклого пространства параметров. Возможны, однако, неограниченные пространства параметров, такие, например, как $\sigma > 0$ в регрессионной модели.

- **Идентифицируемость (Identifiability).** Оценивание должно быть возможным. Это аспект, связанный с определением 14.1, которое затрагивает идентификацию и связанные с этим вопросы.
- **Доброкачественность данных (Well-behaved data).** Законы больших чисел применяются к выборочным средним имеющихся данных, а некоторые формы центральной предельной теоремы (в основном теоремы Ляпунова) применяются к градиенту. Эргодическая стационарность является достаточно широким условием, чтобы включать любую ситуацию, которая вероятна на практике. Более того, она даже более широкая, чем нам необходимо на данном этапе, поскольку пока мы не рассматриваем зависимые наблюдения. Предполагается, что без изменений остаются все определения главы 4.

Имея в виду вышеуказанные предположения, дальнейший анализ, по сути, очень похож на использованный нами для модели линейной регрессии в главе 4 и строго следует содержанию раздела 12.5.

14.6. Тестирование гипотез и спецификации модели и показатели качества подгонки модели

В следующих нескольких подразделах мы рассматриваем широко используемые методы проверки гипотез: критерий отношения правдоподобия, критерий Вальда и критерий множителей Лагранжа. [Исчерпывающее описание этих методов приведено в работе Годфри (Godfrey (1988)).] Мы рассматриваем оценивание параметра θ методом максимального правдоподобия и проверку гипотезы $H_0: c(\theta) = 0$. Общая логика всех трех тестов представлена на рис. 14.2⁵. Рисунок графически изображает логарифмическую функцию правдоподобия $\ln L(\theta)$, ее производную по θ ($d \ln L(\theta)/d\theta$) и ограничение $c(\theta)$. Существует три критерия для проверки гипотез, представленные на этом рисунке.

- **Критерий отношения правдоподобия (Likelihood ratio test).** Если ограничение $c(\theta) = 0$ является верным, то его наложение не должно приводить к значительному уменьшению логарифмической функции правдоподобия. Таким образом, в основе критерия лежит разность $\ln L_U - \ln L_R$, где L_U — значение функции правдоподобия при оценке θ ,

⁵Buse (1982). Обратим внимание, что масштаб вертикальной оси должен быть разным для каждой кривой. Таким образом, точки пересечения не имеют смысла.

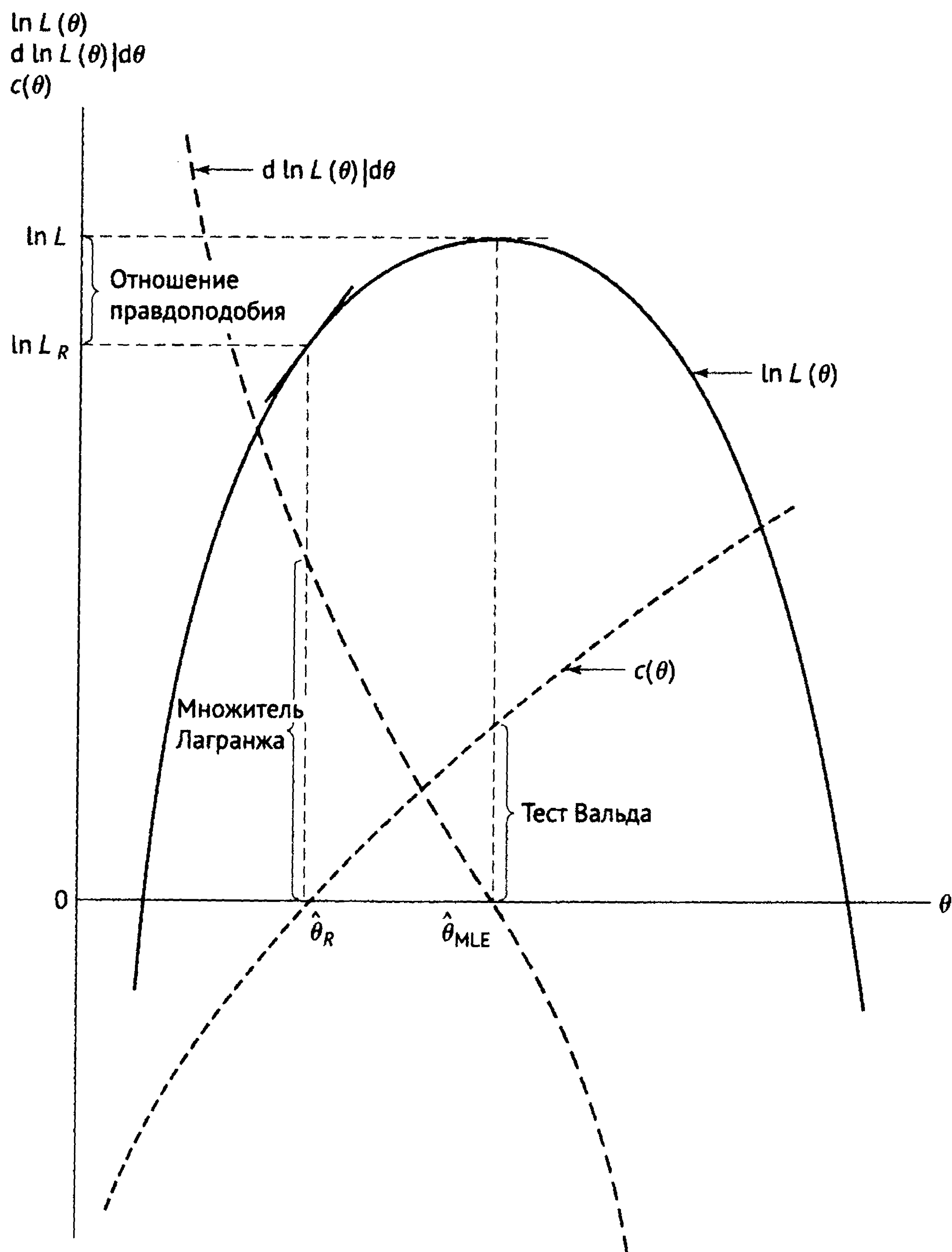


Рис. 14.2. Три критерия проверки гипотез

полученной без ограничения, а L_R — значение функции правдоподобия при оценке, полученной при ограничении.

- **Критерий Вальда (Wald test).** Если ограничение является верным, то величина $c(\hat{\theta}_{MLE})$ должна быть близка к нулю, поскольку ML-оценка состоятельна. Таким образом, критерий основан на величине $c(\hat{\theta}_{MLE})$. Мы отвергаем гипотезу, если это значение существенно отличается от нуля.
- **Критерий множителей Лагранжа (Lagrange multiplier test).** Если ограничение является верным, то оценка, вычисленная при этом ограничении, должна быть рядом с точкой, которая максимизирует логарифмическое правдоподобие. Поэтому наклон логарифмической функции правдоподобия в этой точке должен быть около нуля. Таким образом, в основе критерия лежит значение градиента логарифмического правдоподобия в точке, где функция максимизируется при наличии ограничения.

Эти три критерия асимптотически эквивалентны при нулевой гипотезе, но могут значительно различаться в малой выборке. К сожалению, их свойства в малых выборках неизвестны, за исключением нескольких частных случаев. Как следствие выбор среди них обычно делается исходя из простоты требуемых вычислений. Критерий отношения правдоподобия требует вычисления оценок как при ограничениях, так и без них. Если обе оценки просто вычислить, то этот способ является весьма удобным. Критерий Вальда требует вычисления оценок только без ограничений, а критерий множителей Лагранжа, напротив, требует вычисления оценок при ограничениях. В некоторых задачах может оказаться так, что одну из этих оценок рассчитать намного легче, чем другую. Например, линейная модель, оцениваемая просто, обращается в нелинейную и громоздкую при наложении нелинейного ограничения. В этом случае статистика Вальда может оказаться предпочтительнее. В свою очередь, ограничения иногда приводят к устранению нелинейностей. Это относится к критерию множителей Лагранжа, который в данном случае окажется более простым методом.

14.6.1. Критерий отношения правдоподобия

Пусть θ — вектор параметров, который требуется оценить, и пусть H_0 задает некоторое ограничение на эти параметры. Пусть $\hat{\theta}_U$ будет оценкой максимального правдоподобия θ , полученной без учета этого ограничения, а $\hat{\theta}_R$ — оценкой максимального правдоподобия при этом ограничении. Если \hat{L}_U и \hat{L}_R являются функциями правдоподобия, вычисленными в этих двух оценках, то **отношение правдоподобия** (likelihood ratio) представляет собой

$$\lambda = \frac{\hat{L}_R}{\hat{L}_U}. \quad (14-21)$$

Эта функция должна находиться между нулем и единицей. Оба правдоподобия имеют положительную величину, и \hat{L}_R не может быть больше, чем \hat{L}_U . (Оптимум при ограничениях не может превосходить оптимум без них.) Если λ слишком мала, тогда ограничение ставится под сомнение.

Пример, рассматривающий дискретное распределение, помогает прояснить эту мысль. При оценивании по выборке из 10 точек, взятых из пуассоновского распределения, в начале раздела 14.3 мы получили, что ML-оценка

параметра θ равна 2. При этом значении правдоподобие, которое представляет собой вероятность того, что выборка будет именно такой, какой мы ее наблюдаем, равно $0,104 \times 10^{-7}$. Согласуются ли эти данные с гипотезой $H_0: \theta = 1,8$? $L_R = 0,936 \times 10^{-8}$ и, как и ожидали, меньше L_U . Эта конкретная выборка менее вероятна при выполнении рассматриваемой гипотезы.

Формальная тестовая процедура основывается на следующем результате.

Теорема 14.5. *Предельное распределение статистики отношения правдоподобий*

При условиях регулярности и выполнении H_0 предельное распределение величины $-2 \ln \lambda$ есть распределение хи-квадрат с числом степеней свободы, равным числу наложенных ограничений.

Нулевая гипотеза отвергается, если это значение превышает соответствующее критическое значение из таблиц распределения хи-квадрат. Таким образом, для примера с распределением Пуассона

$$-2 \ln \lambda = -2 \ln \left(\frac{0,0936}{0,104} \right) = 0,21072.$$

Эта статистика распределения хи-квадрат с одной степенью свободы не является значимой при любом стандартном уровне значимости, поэтому на основании данного критерия мы не отвергаем гипотезу о том, что $\theta = 1,8$ ⁶.

Критерий отношения правдоподобий хотелось бы использовать для проверки простой нулевой гипотезы против простой альтернативы. Например, в предыдущей задаче нас могла бы интересовать проверка гипотезы $H_0: \theta = 1,8$ против альтернативы $H_1: \theta = 2,2$. Однако критерий не может быть использован в этом случае. Число степеней свободы статистики хи-квадрат для критерия отношения правдоподобий равно уменьшению порядка размерности пространства параметров, которое получается из наложения ограничений. В проверке простой нулевой гипотезы против простой альтернативы это значение равно нулю⁷. Иногда встречаются попытки протестировать вид предполагаемого распределения с помощью критерия отношения правдоподобий. Например, конкретная модель оценивается при предположении нормального распределения, а затем t -распределения. После этого сравнивается соотношение двух правдоподобий, чтобы установить, какое из распределений предпочтительнее. Этот способ не является подходящим. Пространства параметров, а следовательно, и функции правдоподобия для этих двух случаев не связаны между собой.

⁶Конечно, законность использования этого асимптотического результата на выборке, состоящей всего из 10 наблюдений, является спорной.

⁷Заметим, что, поскольку оба правдоподобия в этом примере вычислены при ограничениях, мы не можем гарантировать неотрицательность величины $-2 \ln \lambda$.

14.6.2. Критерий Вальда

Практический недостаток критерия отношения правдоподобий состоит в том, что при его использовании обычно требуется оценивать вектор параметров, вычисленный как при ограничениях, так и без них. В сложных моделях ту или иную из этих оценок иногда очень трудно рассчитать. К счастью, есть два альтернативных метода проверки — критерий Вальда и критерий множителей Лагранжа, которые позволяют обойти эту проблему. Оба критерия основаны на оценке, которая распределена асимптотически.

Эти два критерия основаны на распределении квадратичной формы полного ранга, рассмотренной в разделе В.11.6. Точнее,

$$\text{если } \mathbf{x} \sim N_J[\boldsymbol{\mu}, \boldsymbol{\Sigma}], \text{ то } (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \sim \quad (14-22)$$

В постановке задачи проверки гипотезы, если $E(\mathbf{x}) = \boldsymbol{\mu}$, квадратичная форма имеет распределение хи-квадрат. Если гипотеза о том, что $E(\mathbf{x}) = \boldsymbol{\mu}$, является ложной, то исходная квадратичная форма будет в среднем больше, чем в случае верной гипотезы⁸. Это условие формирует основу для статистик критериев, которые рассматриваются в этом и следующем разделах.

Пусть $\hat{\boldsymbol{\theta}}$ является вектором оценок параметров, полученных без ограничений. Мы тестируем наличие совокупности ограничений

$$H_0: \mathbf{c}(\boldsymbol{\theta}) = \mathbf{q}.$$

Если ограничения справедливы, то оценка $\hat{\boldsymbol{\theta}}$ должна, по крайней мере приблизительно, удовлетворять им. Однако если гипотеза ошибочна, то величина $\mathbf{c}(\hat{\boldsymbol{\theta}}) - \mathbf{q}$ должна быть дальше от $\mathbf{0}$, чем это могло бы объясняться случайностью выборки. Техника, которая используется для формализации этой мысли, есть критерий Вальда.

Теорема 14.6. *Предельное распределение статистики Вальда*
Статистика Вальда представляет собой

$$W = [\mathbf{c}(\hat{\boldsymbol{\theta}}) - \mathbf{q}]' (\text{Asy. Var}[\mathbf{c}(\hat{\boldsymbol{\theta}}) - \mathbf{q}])^{-1} [\mathbf{c}(\hat{\boldsymbol{\theta}}) - \mathbf{q}].$$

При гипотезе H_0 W имеет предельное распределение хи-квадрат с числом степеней свободы, равным числу ограничений [т. е. количеству уравнений в выражении $\mathbf{c}(\hat{\boldsymbol{\theta}}) - \mathbf{q} = \mathbf{0}$]. Вывод этого предельного распределения можно найти в теореме 5.1.

Этот критерий аналогичен статистике хи-квадрат в (14-22), если $\mathbf{c}(\hat{\boldsymbol{\theta}}) - \mathbf{q}$ распределено нормально с предположенным средним значением, равным $\mathbf{0}$.

⁸Если среднее значение не равно $\boldsymbol{\mu}$, то статистика (14-22) будет иметь **нецентральное распределение хи-квадрат** (noncentral chi-squared distribution). Это распределение имеет в целом тот же вид, что и центральное распределение хи-квадрат с тем же числом степеней свободы, но находится правее от него. Таким образом, случайное наблюдение из нецентрального распределения будет в среднем больше, чем случайное наблюдение из центрального распределения.

Большое значение статистики W ведет к отклонению гипотезы. Наконец, обратим внимание, что статистика W требует вычисления только модели без учета ограничений. Здесь требуется вычислять ковариационную матрицу, входящую в рассматриваемую квадратичную форму. Этот результат представляет собой дисперсию, вообще говоря, нелинейной функции, которую мы рассматривали ранее:

$$\begin{aligned} \text{Est. Asy. Var}[c(\hat{\theta}) - \mathbf{q}] &= \hat{\mathbf{C}} \text{Est. Asy. Var}[\hat{\theta}] \hat{\mathbf{C}}', \\ \hat{\mathbf{C}} &= \left[\frac{\partial c(\hat{\theta})}{\partial \hat{\theta}'} \right]. \end{aligned} \quad (14-23)$$

То есть \mathbf{C} представляет собой $J \times K$ матрицу, j -я строка которой состоит из производных j -го ограничения по K элементам вектора θ . На практике часто имеет место проверка набора линейных ограничений.

Для проверки набора линейных ограничений $\mathbf{R}\theta = \mathbf{q}$ критерий Вальда основывается на

$$\begin{aligned} H_0: c(\theta) - \mathbf{q} = \mathbf{R}\theta - \mathbf{q} = \mathbf{0}, \\ \hat{\mathbf{C}} = \left[\frac{\partial c(\hat{\theta})}{\partial \hat{\theta}'} \right] = \mathbf{R}, \end{aligned} \quad (14-24)$$

$$\text{Est. Asy. Var}[c(\hat{\theta}) - \mathbf{q}] = \mathbf{R} \text{Est. Asy. Var}[\hat{\theta}] \mathbf{R},$$

и

$$W = [\mathbf{R}\hat{\theta} - \mathbf{q}]' [\mathbf{R} \text{Est. Asy. Var}(\hat{\theta}) \mathbf{R}']^{-1} [\mathbf{R}\hat{\theta} - \mathbf{q}].$$

Число степеней свободы равно числу строк в матрице \mathbf{R} .

Если мы имеем дело с единственным ограничением $c(\theta) = \mathbf{q}$, то критерий Вальда превращается в критерий, основанный на доверительном интервале, который мы рассматривали ранее. Если мы тестируем

$$H_0: \theta = \theta_0 \quad \text{против} \quad H_1: \theta \neq \theta_0,$$

то первый критерий будет основан на статистике

$$z = \frac{|\hat{\theta} - \theta_0|}{s(\hat{\theta})}, \quad (14-25)$$

где $s(\hat{\theta})$ — это оцененная асимптотическая стандартная ошибка. Тестовая статистика сравнивается с соответствующим значением из таблицы нормального распределения. Критерий Вальда будет основан на статистике

$$W = [(\hat{\theta} - \theta_0) - 0] (\text{Asy. Var}[(\hat{\theta} - \theta_0) - 0])^{-1} [(\hat{\theta} - \theta_0) - 0] = \frac{(\hat{\theta} - \theta_0)^2}{\text{Asy. Var}[\hat{\theta}]} = z^2. \quad (14-26)$$

Здесь W имеет предельное распределение хи-квадрат с одной степенью свободы, которое представляет собой распределение квадрата стандартной нормальной тестовой статистики (14-25).

Подводя итог, следует отметить, что критерий Вальда основан на измерении того, насколько оценки, вычисленные без учета ограничений, не соответствуют предполагаемым ограничениям. У критерия Вальда есть два

недостатка. Первый — он является чистым тестом значимости нулевой гипотезы, не зависящим от конкретной альтернативной гипотезы. В связи с этим в некоторых задачах его возможности могут быть ограничены. В действительности статистика, лежащая в основе критерия, имеет тенденцию быть, скорее, больше в моделях. Вторым его недостатком многие рассматриваемые здесь тесты не обладают. Статистика Вальда не является инвариантной к формулировке ограничений. Например, для проверки гипотезы о том, что величина $\theta = \beta/(1 - \gamma)$ равна некоторому значению q , можно выбрать один из двух способов. Критерий Вальда, основанный непосредственно на $\theta - q = 0$, будет использовать статистику, основанную на дисперсии этой нелинейной функции. Другой способ состоит в рассмотрении линейного ограничения $\beta - q(1 - \gamma) = 0$, которое хоть и является эквивалентным, но уже является линейным. Статистики Вальда в этих двух случаях могут быть различными и привести к разным выводам. Эти два недостатка представляют собой частые аргументы против использования критерия Вальда. Но к достоинствам критерия Вальда относится тот факт, что он не зависит от сильного предположения о распределении в отличие от критерия отношения правдоподобий и критерия множителей Лагранжа. Современная литература по эконометрике часто рассматривает модели, которые основаны на процедурах оценивания, независимых от распределения, таких как GMM-метод. В связи с этим в последние годы популярность критерия Вальда возрастает.

14.6.3. Критерий множителей Лагранжа

Третья рассматриваемая процедура — это **критерий множителей Лагранжа** (Lagrange multiplier, LM), или **критерий эффективного сора** (или просто **сора**) (efficient score (score) test). Он основан на модели с ограничениями, а не на модели без них. Предположим, что мы максимизируем логарифмическое правдоподобие при условии наличия некоторого набора ограничений $c(\theta) - q = 0$. Пусть λ — вектор множителей Лагранжа, мы определим функцию Лагранжа

$$\ln L^*(\theta) = \ln L(\theta) + \lambda'(c(\theta) - q).$$

Решением задачи условной максимизации являются корни системы уравнений

$$\begin{aligned} \frac{\partial \ln L^*}{\partial \theta} &= \frac{\partial \ln L(\theta)}{\partial \theta} + C' \lambda = 0, \\ \frac{\partial \ln L^*}{\partial \lambda} &= c(\theta) - q = 0, \end{aligned} \tag{14-27}$$

где C' — транспонированная матрица производных из второй строки (14-23). Если ограничения верны, то их наложение не должно приводить к значимому различию в точке максимума функции правдоподобия. Если рассуждать в терминах условий первого порядка, то второй член в векторе производных должен быть небольшим. В частности, величина λ будет незначительной. Мы можем проверить это непосредственно, т. е. проверить гипотезу $H_0: \lambda = 0$, которая и приведет к критерию множителей Лагранжа. Однако существует более простая эквивалентная формулировка. В точке условного

максимума производные логарифмической функции правдоподобия представляют собой

$$\frac{\partial \ln L(\hat{\theta}_R)}{\partial \hat{\theta}_R} = -\hat{C}' \hat{\lambda} = \hat{g}_R. \quad (14-28)$$

Если ограничения являются верными по крайней мере внутри диапазона выборки, то тогда $\hat{g}_R = 0$. Это значит, что производные логарифмического правдоподобия, вычисленные в точке параметров при ограничениях, будут близки к нулю. Вектор первых производных логарифмического правдоподобия является **эффективным скор-вектором**. Поскольку критерий множителей Лагранжа основан на этом векторе, его часто называют **скор-критерием**. Дисперсия вектора первых производных есть информационная матрица, которую мы использовали для вычисления асимптотической ковариационной матрицы ML-оценки. Тестовая статистика основана на подходе, аналогичном статистике критерия Вальда.

Теорема 14.7. *Предельное распределение статистики множителей Лагранжа*

Статистика множителей Лагранжа имеет вид

$$LM = \left(\frac{\partial \ln L(\hat{\theta}_R)}{\partial \hat{\theta}_R} \right)' [\mathbf{I}(\hat{\theta}_R)]^{-1} \left(\frac{\partial \ln L(\hat{\theta}_R)}{\partial \hat{\theta}_R} \right).$$

При нулевой гипотезе LM имеет предельное распределение хи-квадрат с числом степеней свободы, равным числу ограничений. Все выражения вычисляются в оценке, полученной при ограничениях.

LM-статистика имеет следующий удобный вид. Пусть \hat{g}_{iR} обозначает i -й элемент градиента логарифмической функции правдоподобия. Тогда

$$\hat{g}_R = \sum_{i=1}^n \hat{g}_{iR} = \hat{G}'_R \mathbf{i},$$

где \hat{G}_R — матрица размером $n \times K$ с i -й строкой, равной \hat{g}'_{iR} , а \mathbf{i} — столбец единиц. Если мы используем ВНН-оценку (векторное произведение градиентов) в (14-18), чтобы оценить гессиан, то

$$[\hat{\mathbf{I}}(\hat{\theta})]^{-1} = [\hat{G}'_R \hat{G}_R]^{-1}$$

и

$$LM = \mathbf{i}' \hat{G}_R [\hat{G}'_R \hat{G}_R]^{-1} \hat{G}'_R \mathbf{i}.$$

Теперь, поскольку $\mathbf{i}'\mathbf{i}$ равняется n , $LM = n(\mathbf{i}' \hat{G}_R [\hat{G}'_R \hat{G}_R]^{-1} \hat{G}'_R \mathbf{i} / n) = nR_i^2$, что равно произведению n и нецентрированного коэффициента квадрата множественной корреляции, умноженного на n , в линейной регрессии столбца единиц производных логарифмической функции правдоподобия, вычисленной в оценке при ограничениях. Мы вернемся к этому результату в различных видах несколько раз на протяжении книги.

14.6.4. Применение критериев, основанных на правдоподобии

Рассмотрим снова данные примера С.1. В примере 14.4 параметр β в модели

$$f(y_i | x_i, \beta) = \frac{1}{\beta + x_i} e^{-y_i/(\beta + x_i)} \quad (14-29)$$

был оценен методом максимального правдоподобия. Для удобства обозначим $\beta_i = 1/(\beta + x_i)$. Эта экспоненциальная плотность есть частный случай более общего гамма-распределения

$$f(y_i | x_i, \beta, \rho) = \frac{\beta_i^\rho}{\Gamma(\rho)} y_i^{\rho-1} e^{-y_i \beta_i}. \quad (14-30)$$

Экспоненциальная плотность получается при $\rho = 1$. Рассмотрим проверку гипотезы

$$H_0: \rho = 1 \quad \text{против} \quad H_1: \rho \neq 1,$$

используя различные методы, описанные выше. Логарифмическое правдоподобие и его производные имеют вид

$$\ln L(\beta, \rho) = \rho \sum_{i=1}^n \ln \beta_i - n \ln \Gamma(\rho) + (\rho - 1) \sum_{i=1}^n \ln y_i - \sum_{i=1}^n y_i \beta_i,$$

$$\frac{\partial L}{\partial \beta} = -\rho \sum_{i=1}^n \ln \beta_i + \sum_{i=1}^n \ln y_i \beta_i^2, \quad \frac{\partial L}{\partial \rho} = \sum_{i=1}^n \ln \beta_i - n \Psi(\rho) + \sum_{i=1}^n \ln y_i, \quad (14-31)$$

$$\frac{\partial^2 L}{\partial \beta^2} = \rho \sum_{i=1}^n \ln \beta_i^2 - 2 \sum_{i=1}^n \ln y_i \beta_i^3, \quad \frac{\partial^2 L}{\partial \rho^2} = -n \Psi'(\rho), \quad \frac{\partial^2 \ln L}{\partial \beta \partial \rho} = -\sum_{i=1}^n \beta_i.$$

[Напомним, что $\Psi(\rho) = d \ln \Gamma(\rho) / d\rho$ и $\Psi'(\rho) = d^2 \ln \Gamma(\rho) / d\rho^2$.] Оценки максимального правдоподобия при отсутствии ограничений параметров β и ρ получаются, если приравнять две первые производные к нулю. Оценка максимального правдоподобия при ограничениях параметра β получена приравниванием $\partial \ln L / \partial \beta$ к нулю при ρ , равном единице. Результаты представлены в табл. 14.1. Для асимптотической ковариационной матрицы оценки $\theta = (\beta, \rho)'$ есть три способа оценивания. Используя фактический гессиан (14-17), мы можем вычислить $V = [-\sum_i \partial^2 \ln f(y_i | x_i, \beta, \rho) / \partial \theta \partial \theta']^{-1}$ в оценках максимального правдоподобия. Для этой модели можно легко показать, что $E[y_i | x_i] = \rho(\beta + x_i)$ (либо путем прямого интегрирования, либо, что проще, используя то, что $E[\partial \ln L / \partial \beta] = 0$). Вследствие этого мы также можем использовать математическое ожидание гессиана (14-16), чтобы вычислить $V_E = \{-\sum_i E[\partial^2 \ln f(y_i | x_i, \beta, \rho) / \partial \theta \partial \theta']\}^{-1}$. Наконец, используя суммы квадратов и попарные произведения первых производных, мы можем получить ВНН-оценку (14-18): $V_B = [\sum_i (\partial \ln f(y_i | x_i, \beta, \rho) / \partial \theta) (\partial \ln f(y_i | x_i, \beta, \rho) / \partial \theta)']^{-1}$. Результаты, представленные в табл. 14.1, основываются на матрице V .

Таблица 14.1. Оценки максимального правдоподобия

| Величина | Оценка без учета ограничений ^a | Оценка при ограничениях |
|---|---|-------------------------|
| β | -4,7185 (2,345) | 15,6027 (6,794) |
| ρ | 3,1509 (0,794) | 1,0000 (0,000) |
| $\ln L$ | -82,91605 | -88,43626 |
| $\partial \ln L / \partial \beta$ | 0,0000 | 0,0000 |
| $\partial \ln L / \partial \rho$ | 0,0000 | 7,9145 |
| $\partial^2 \ln L / \partial \beta^2$ | -0,85570 | -0,02166 |
| $\partial^2 \ln L / \partial \rho^2$ | -7,4592 | -32,8987 |
| $\partial^2 \ln L / \partial \beta \partial \rho$ | -2,2420 | -0,66891 |

^a Оценки асимптотических стандартных ошибок, основанных на матрице \mathbf{V} , даны в скобках.

Эти три оценки асимптотической ковариационной матрицы приводят к весьма различающимся результатам:

$$\mathbf{V} = \begin{bmatrix} 5,499 & -1,653 \\ -1,653 & 0,6309 \end{bmatrix}, \quad \mathbf{V}_E = \begin{bmatrix} 4,900 & -1,473 \\ -1,473 & 0,5768 \end{bmatrix},$$

$$\mathbf{V}_B = \begin{bmatrix} 13,37 & -4,322 \\ -4,322 & 1,537 \end{bmatrix}.$$

При имеющемся малом размере выборке эти различия должны были ожидать. Однако заметное отличие ВНН-оценки есть типичное проявление ее плохого качества в малых выборках.

- **Критерий доверительного интервала (Confidence interval test):** 95%-й доверительный интервал для параметра ρ , основанный на оценках без ограничений, равен $3,1509 \pm 1,96\sqrt{0,6309} = [1,5941; 4,7076]$. Этот интервал не содержит $\rho = 1$, поэтому гипотеза отвергается.
- **Критерий отношения правдоподобия (Likelihood ration test):** статистика LR выражается как $\lambda = -2[-88,43626 - (-82,91604)] = 11,0404$. Критическое значение для распределения хи-квадрат с одной степенью свободы равно 3,842. Рассчитанное значение больше, чем это критическое значение, поэтому гипотеза снова отвергается.
- **Критерий Вальда (Wald test):** критерий Вальда основан на оценках без ограничений. Для тестируемого ограничения $c(\theta) - q = \rho - 1$, $dc(\hat{\rho})/d2\hat{\rho} = 1$, $\text{Est. Asy. Var}[c(\hat{\rho}) - q] = \text{Est. Asy. Var}[\hat{\rho}] = 0,6309$, поэтому $W = (3,1517 - 1)^2 / [0,6309] = 7,3384$. Критическое значение здесь то же, что и в предыдущем случае. Следовательно, H_0 еще раз отвергается. Обратим внимание, что статистика Вальда есть квадрат статистики, которая использовалась бы в критерии доверительного интервала для проверки этой гипотезы: $|3,1509 - 1| / \sqrt{0,6309} = 2,73335$.
- **Критерий множителей Лагранжа (Lagrange multiplier test):** критерий множителей Лагранжа основан на оценках при ограничениях. Оцениваемая асимптотическая ковариационная матрица производ-

⁹Гамма-функция $\Gamma(\rho)$ и гамма-распределение описаны в разделах В.4.5 и Е2.3.

ных, используемая для вычисления статистики, может быть любой из трех ранее рассмотренных оценок. ВНН-оценка V_B является эмпирической оценкой дисперсии градиента и обычно используется на практике. Вычисления приводят к

$$LM = [0,0000 \quad 7,9145] \begin{bmatrix} 0,00995 & 0,26776 \\ 0,26776 & 11,199 \end{bmatrix}^{-1} \begin{bmatrix} 0,0000 \\ 7,9145 \end{bmatrix} = 15,687.$$

Замечание 14.1. Мы приходим к тому же выводу, что и прежде. Обратим внимание на то, что тот же расчет, выполненный с использованием V , а не V_B , приводит к значению 5,1162. Как и ранее, мы видим существенный разброс оценок, полученных на малой выборке.

Статистики, полученные в трех последних случаях, имеют значительно различающиеся значения. Возможно прийти к разным выводам в зависимости от того, какая из этих статистик используется. Так, например, если рассматривать 1%-й уровень значимости вместо 5%-го и вычислять статистику LM с помощью матрицы V , то соответствующее критическое значение распределения хи-квадрат было бы равно 6,635 и гипотеза не была бы отвергнута LM-тестом. Асимптотически все три критерия эквивалентны. Но в выборке конечного объема, подобной этой, следует ожидать различий¹⁰. К сожалению, не существует четких правил, как действовать в такой ситуации. Эта ситуация представляет собой пример проблемы, связанной с выводом конкретного уровня значимости и принятием твердого решения в вопросе о том, отклонить или принять вывод, основанный на данных выборки.

14.6.5. Сравнение моделей и вычисление качества подгонки модели

Тестовые статистики, рассмотренные нами в подразделах 14.6.1–14.6.3, могут быть использованы для определения обоснованности ограничений на параметры модели. Если модели являются вложенными, то может быть использован любой из трех упомянутых методов проверки. Если же модели являются невложенными, то вычисление того, какая модель более предпочтительна, должно основываться на некотором оценочном критерии. Двумя общими показателями, которые основаны на том же принципе, что и скорректированный R -квадрат для линейной модели, выступают:

информационный критерий Акаике (AIC, Akaike information criterion) $= -2 \ln L + 2K$,

байесовский (Шварца) информационный критерий

(BIC, Bayes (Schwarz) information criterion) $= -2 \ln L + K \ln n$,

где K — число параметров в модели. Выбор модели, основанной по наименьшему AIC, логически обосновывается так же, как и использование \bar{R}^2 в линейной модели: он не является статистическим, но тем не менее широко распространен.

¹⁰Для дальнейшего обсуждения этой проблемы см. работу Berndt, Savin (1977).

AIC и BIC считаются информационными критериями, но, в сущности, не показателями подгонки модели. Соответственно остается открытым вопрос о том, как определить степень подгонки модели. Только в случае оценивания линейной регрессии методом наименьших квадратов в модели со свободным членом мы получаем R^2 , который измеряет часть вариации, объясненную регрессией. Неопределенность R^2 как характеристики подгонки возникает сразу же, когда мы переходим от модели линейной регрессии к модели обобщенной регрессии в главе 9. Эта проблема становится еще более актуальной при рассмотрении моделей этой главы. Так, например, для оценок моделей порядковых данных в примере 14.10 не используется «вариация» зависимой переменной и нет ясного показателя «объясненной вариации».

Показатель подгонки, который был первоначально предложен для моделей дискретного выбора в работе Макфаддена (McFadden (1974)), но неожиданно приобрел широкую популярность в эмпирической литературе, — это **индекс отношения правдоподобий** (likelihood ratio index), который стал известен как **псевдо R^2** (Pseudo R^2). Он рассчитывается как

$$\text{Pseudo}R^2 = 1 - (\ln L)/(\ln L_0),$$

где $\ln L$ — логарифмическое правдоподобие для вычисленной модели, а $\ln L_0$ — логарифмическое правдоподобие для той же модели, но содержащей только постоянный член. Статистика имеет сходство с R^2 в линейной регрессии. Выбор существующего названия для этой статистики, однако, сложно назвать удачным, поскольку даже в рамках моделей дискретного выбора, для которых эта статистика была предложена, она не имеет связи с качеством подгонки данных моделью. В рамках моделей дискретного выбора, в которых логарифмические правдоподобия должны быть отрицательными, псевдо R^2 должен находиться между нулем и единицей и возрастать по мере добавления переменных в модель. Очевидно, он может быть равен нулю, но обычно ограничен уровнем, меньшим единицы. В линейной модели с нормально распределенными шоками максимизированное логарифмическое правдоподобие представляет собой

$$\ln L = (-n/2)[1 + \ln 2\pi + \ln(e'e/n)].$$

Преобразовывая, мы находим, что псевдо R^2 для модели линейной регрессии может быть записан как

$$\text{Pseudo}R^2 = \frac{-\ln(1 - R^2)}{1 + \ln 2\pi + \ln s_y^2},$$

в то время как «истинный» R^2 — это $1 - e'e/e_0'e_0$. Поскольку s_y^2 может варьировать независимо от R^2 (умножение y на любую постоянную A оставляет R^2 без изменений, но увеличивает s_y^2 в A^2 раз), верхний предел есть единица, но нижнего предела у этого показателя нет. Эта же проблема возникает в любой модели, в которой используется информация о масштабе зависимой переменной, например в тобит-модели (глава 19). Эти вычисления имеют еще меньше смысла с точки зрения качества подгонки в мультиномиальных моделях, таких как порядковая пробит-модель (глава 18) или

мультиномиальная логит-модель. Для моделей дискретного выбора существует целый набор таких показателей, которые будут рассмотрены нами в главе 17. Для моделей ограниченной зависимой переменной и многих логлинейных моделей более пригоден будет другой показатель, основывающийся на корреляции между предсказанным и действительным значениями. Тем не менее именно этот показатель, по всей видимости, наиболее широко используется в современной литературе. [Популярное программное обеспечение *Stata* приводит псевдо R^2 при оценивании любой модели с помощью MLE, но в то же время предостерегает пользователей от трактования этого показателя как имеющего какой-нибудь смысл. См., например, <http://www.stata.com/support/faqs/stat/pseudor2.html>. В работе Cameron, Trivedi (2005) авторы обстоятельно рассматривают псевдо R^2 , приходят к аналогичным предостережениям относительно него и советуют читателю воспользоваться более рациональным показателем корреляции между предсказанным моделью значением и наблюдаемым в действительности. В работе Wooldridge (2002a) автор без долгих рассуждений отвергает этот показатель и советует больше внимания уделять коэффициентам модели.]

14.6.6. Тест Вуонга и информационный критерий Кульбака–Лейблера

Подход, предложенный в работе Wuong (1989) для тестирования **невложенных моделей** (nonnested models), также основывается на статистике отношения правдоподобий. Идея подхода схожа с общим принципом, лежащим в основе критерия отношения правдоподобий. Предположим, что $f(y_i | \mathbf{Z}_i, \theta)$ и $g(y_i | \mathbf{Z}_i, \gamma)$ есть две конкурирующие модели для плотности случайной переменной y_i , причем f соответствует нулевой гипотезе H_0 , а g — альтернативе H_1 . Так, в примере 5.7 обе плотности являются (по выдвигаемому предположению) нормальными, y_i — потребление C_t , \mathbf{Z}_i — вектор $[1, Y_t, Y_{t-1}, C_{t-1}]$, θ — вектор $(\beta_1, \beta_2, \beta_3, 0, \sigma^2)$, γ — вектор $(\gamma_1, \gamma_2, 0, \gamma_3, \omega^2)$, а σ^2 и ω^2 — соответствующие условные дисперсии шоков ε_{0t} и ε_{1t} . Существенным моментом в анализе является то, что нет необходимости, чтобы одна из двух конкурирующих моделей была «истинной»: они обе могут быть ложными. То, что мы хотим сделать, — это попытка использовать данные для определения того, какая из моделей находится ближе к истине, т. е. ближе к правильной (неизвестной) модели.

Мы предположим, что наблюдения в выборке (шоки) условно независимы. Обозначим за $L_{i,0}$ вклад i -го наблюдения в функцию правдоподобия при нулевой гипотезе. Таким образом, логарифмическая функция правдоподобия при нулевой гипотезе есть $\sum_i \ln L_{i,0}$. Обозначим за $L_{i,1}$ ту же функцию при альтернативной модели. Наконец, пусть m_i равняется $\ln L_{i,1} - \ln L_{i,0}$. Если бы мы использовали знакомый критерий отношения правдоподобий, то статистика имела бы вид $LR = 2\sum_i m_i = 2n\bar{m}$, где $L_{i,0}$ и $L_{i,1}$ рассчитаны в соответствующих оценках максимального правдоподобия. В случае если конкурирующие модели являются вложенными, т. е. модель H_0 есть ограничение на модель H_1 , мы знаем, что $\sum_i m_i \geq 0$. Ограничения, налагаемые нулевой гипотезой, никогда не увеличат значение функции правдоподобия. (В модели линейной регрессии с нормально распределенными шоками, ко-

торая уже была нами исследована, и логарифмическое правдоподобие, и эти результаты основываются на сумме квадратов остатков и, как мы уже видели, наложение ограничений не может уменьшить эту сумму квадратов.) Предельное распределение статистики LR при нулевой гипотезе есть распределение хи-квадрат с числом степеней свободы, равным уменьшению количества измерений в пространстве параметров альтернативной гипотезы, которое получается вследствие ограничений.

Анализ, проведенный в работе Wuong (1989), связан с невложенными моделями, для которых величина $\sum_i m_i$ необязательно должна быть положительной. Для формализации этого критерия требуется более тщательно разобраться, что подразумевается под «истинной» моделью (это даст нам также удобную базу для обсуждения следующих двух разделов). В контексте невложенных моделей автор допускает возможность, что ни одна из моделей не является «истинной» в абсолютном смысле. Мы предполагаем, следуя классической схеме, что существует «истинная» модель $h(y_i | \mathbf{Z}_i, \alpha)$, где α — это «истинный» вектор параметров, но, возможно, ни одна из рассматриваемых моделей не является истинной. **Информационный критерий Кульбака–Лейблера** (KLIC, Kullback–Leibler Information Criterion) измеряет разницу между истинной моделью (распределением) и предполагаемой моделью в терминах функции правдоподобия. Грубо говоря, KLIC — это логарифмическая функция правдоподобия для истинной модели минус логарифмическая функция правдоподобия для (неверной) рассматриваемой модели при справедливости истинной модели. Формально для модели нулевой гипотезы

$$KLIC = E[\ln h(y_i | \mathbf{Z}_i, \alpha) | h \text{ истинно}] - E[\ln f(y_i | \mathbf{Z}_i, \theta) | h \text{ истинно}].$$

Первый член в правой части уравнения — это то, что мы вычислили бы с помощью $(1/n) \ln L$, если бы максимизировали логарифмическое правдоподобие для истинной модели $h(y_i | \mathbf{Z}_i, \alpha)$. Вторым членом — это то, что оценивается выражением $(1/n) \ln L$ при предположении (неверном), что $f(y_i | \mathbf{Z}_i, \theta)$ есть верная модель. Обратим внимание, что выражение $f(y_i | \mathbf{Z}_i, \theta)$ записано относительно вектора параметров θ . Поскольку α есть «истинный» вектор параметров, то не совсем ясно, что имеется в виду под параметризацией с помощью параметра θ . Автор называет этот вектор (Wuong (1989), p. 310) псевдоистинным вектором параметров. Это вектор постоянных, к которому сходится оценка, полученная с помощью $f(y_i | \mathbf{Z}_i, \theta)$. В примере 5.7, если H_0 является верной моделью, эта формулировка означает, что оценка, полученная методом наименьших квадратов при альтернативе H_1 , будет стремиться к некоторому вектору псевдоистинных параметров. Эти параметры не являются параметрами правильной модели: они будут коэффициентами при линейной проекции генеральной совокупности C_t на регрессоры $[1, Y_t, C_{t-1}]$.

Предположим, что «истинная» модель имеет вид $y = \mathbf{X}\beta + \varepsilon$ с нормально распределенными шоками и $y = \mathbf{Z}\delta + w$ есть предложенная конкурирующая модель. KLIC будет ожидаемой логарифмической функцией правдоподобия для истинной модели минус ожидаемая логарифмическая функция правдоподобия для второй модели, полученная при предположении, что первая модель истинна. По построению KLIC является положительным. Определим, что одна модель «лучше», чем другая, если она ближе к «истине»,

основываясь на KLIC. Если мы рассмотрим разность KLIC для двух моделей, то истинная логарифмическая функция правдоподобия сокращается и мы приходим к

$$KLIC_1 - KLIC_0 = E[\ln f(y_i | \mathbf{Z}_i, \theta) | h \text{ истинно}] - E[\ln g(y_i | \mathbf{Z}_i, \gamma) | h \text{ истинно}].$$

Для того чтобы вычислить это выражение на основе выборки, требуется просто снова вычислить статистику отношения правдоподобия $n\bar{m}$ (без умножения на 2). Таким образом, это дает интерпретацию статистики LR. Но при этом полученная статистика может оказаться отрицательной: мы не знаем, какая из двух конкурирующих моделей ближе к истине.

Общий результат Вуонга для невложенных моделей (теорема 5.1 работы Wuong (1989)) описывает поведение статистики

$$V = \frac{\sqrt{n} \left(\frac{1}{n} \sum_{i=1}^n m_i \right)}{\sqrt{\frac{1}{n} \sum_{i=1}^n (m_i - \bar{m})^2}} = \sqrt{n}(\bar{m}/s_m), \quad m_i = \ln L_{i,1} - \ln L_{i,0}.$$

Автор получает, что:

- 1) при гипотезе, что модели «эквивалентны», $V \xrightarrow{D} N[0, 1]$,
- 2) при гипотезе, что $f(y_i | \mathbf{Z}_i, \theta)$ «лучше», $V \xrightarrow{A.S.} +\infty$,
- 3) при гипотезе, что $g(y_i | \mathbf{Z}_i, \gamma)$ «лучше», $V \xrightarrow{A.S.} -\infty$.

Этот критерий является направляющим. Большие положительные значения свидетельствуют в пользу нулевой модели, тогда как большие отрицательные значения — в пользу альтернативы. Промежуточные значения (например, между $-1,96$ и $+1,96$ для 95-го уровня значимости) считаются областью неопределенности. Применение этого критерия в прикладной задаче рассмотрено в примере 14.10.

14.7. Двухшаговая оценка методом максимального правдоподобия

Прикладная литература содержит большое и все увеличивающееся количество моделей, в которых элементы одной модели включены в другую, что приводит к так называемым проблемам «двухшагового» оценивания. [Среди наиболее известных моделей — модель Хекмана (1979) выборочного отбора, рассматриваемая в примере 1.1 и в главе 19.] Пусть есть два вектора параметров: θ_1 и θ_2 . Первый появляется во второй модели, но не наоборот. В такой ситуации есть два пути. Оценивание методом **максимального правдоподобия с полной информацией (FIML, full information maximum likelihood)** требует формирования совместного распределения $f(y_1, y_2 | \mathbf{x}_1, \mathbf{x}_2, \theta_1, \theta_2)$ двух случайных величин, а затем максимизации полной логарифмической функции правдоподобия

$$\ln L(\theta_1, \theta_2) = \sum_{i=1}^n \ln f(y_{i1}, y_{i2} | \mathbf{x}_{i1}, \mathbf{x}_{i2}, \theta_1, \theta_2).$$

Двухшаговая процедура для подобной модели могла бы быть проведена следующим образом. Сначала оценивают параметры модели 1 путем максимизации

$$\ln L_1(\theta_1) = \sum_{i=1}^n \ln f_1(y_{i1} | \mathbf{x}_{i1}, \theta_1),$$

а затем максимизируют маргинальную функцию правдоподобия для y_2 , подставляя состоятельную оценку θ_1 и рассматривая ее как заданный параметр. Таким образом, второй шаг требует максимизации

$$\ln L_2(\hat{\theta}_1, \theta_2) = \sum_{i=1}^n \ln f_2(y_{i2} | \mathbf{x}_{i1}, \mathbf{x}_{i2}, \hat{\theta}_1, \theta_2).$$

Существует по меньшей мере две причины для того, чтобы воспользоваться именно этим способом. Во-первых, может оказаться так, что выразить непосредственно формулой два отдельных логарифмических правдоподобия не представляет сложности, однако очень сложно вывести их совместное распределение. Эта ситуация часто возникает, когда две моделируемые переменные принадлежат к двум разным видам распределений: одно является дискретным, а другое — непрерывным (что распространено в подобных задачах). Вторая причина заключается в том, что максимизация отдельных логарифмических правдоподобий может быть несложной, но максимизация совместного логарифмического правдоподобия может быть сложной или численно трудоемкой¹¹. Эти результаты могут быть найдены в важном источнике по этой тематике — работе Мёрфи, Топеля (Murphy, Topel) (2002 г., впервые опубликована в 1985 г.).

Теперь предположим, что наша модель состоит из двух предельных распределений: $f_1(y_1 | \mathbf{x}_1, \theta_1)$ и $f_2(y_2 | \mathbf{x}_1, \mathbf{x}_2, \theta_1, \theta_2)$. Оценивание мы будем осуществлять в два шага.

1. Оцениваем θ_1 методом максимального правдоподобия для модели 1. Пусть \hat{V}_1 — это любая из оценок асимптотической ковариационной матрицы этой оценки, которые были рассмотрены в подразделе 14.4.6, умноженная на n .
2. Оцениваем θ_2 методом максимального правдоподобия для модели 2, заменяя θ_1 на $\hat{\theta}_1$, как будто это известный параметр. Пусть \hat{V}_2 — любая подходящая оценка асимптотической ковариационной матрицы $\hat{\theta}_2$, умноженная на n .

Доказательство состоятельности оценки $\hat{\theta}_2$, по сути, в том, что если бы θ_1 была известна, тогда все наши результаты для MLE применялись бы для оценивания θ_2 , а поскольку $\text{plim } \hat{\theta}_1 = \theta_1$, то асимптотически эта цепочка рассуждений остается верной. (См. п. 3 теоремы D.16.) Но этой цепочки рассуждений недостаточно, чтобы обосновать использование $(1/n)\hat{V}_2$ в качестве

¹¹ Существует третья возможная мотивация. Если одна из моделей является неверной, то тогда FIML-оценки обеих моделей будут несостоятельными. Но если только вторая модель специфицирована неверно, то, по крайней мере, первая модель будет оценена состоятельно. Конечно, этот результат есть только половина дела, но это, может быть, лучше, чем ничего.

оценки асимптотической ковариационной матрицы $\hat{\theta}_2$. Требуется некоторая поправка, учитывающая оценку θ_1 , которая использовалась при оценивании θ_2 . Общий результат состоит в следующем.

Теорема 14.8. *Асимптотическое распределение двухшаговой ML-оценки [Murphy, Torpe (2002)]*

Если стандартные условия регулярности выполняются для обеих логарифмических функций правдоподобия, то полученная на втором шаге оценка методом максимального правдоподобия θ_2 является состоятельной и асимптотически нормально распределенной с асимптотической ковариационной матрицей

$$\mathbf{V}_2^* = \frac{1}{n} [\mathbf{V}_2 + \mathbf{V}_2[\mathbf{C}\mathbf{V}_1\mathbf{C}' - \mathbf{R}\mathbf{V}_1\mathbf{C}' - \mathbf{C}\mathbf{V}_1\mathbf{R}']\mathbf{V}_2],$$

где

$$\mathbf{V}_1 = \text{Asy. Var}[\sqrt{n}(\hat{\theta}_1 - \theta_1)], \text{ основанная на } \ln L_1,$$

$$\mathbf{V}_2 = \text{Asy. Var}[\sqrt{n}(\hat{\theta}_2 - \theta_2)], \text{ основанная на } \ln L_2 | \theta_1,$$

$$\mathbf{C} = E \left[\frac{1}{n} \left(\frac{\partial \ln L_2}{\partial \theta_2} \right) \left(\frac{\partial \ln L_2}{\partial \theta_1'} \right) \right], \quad \mathbf{R} = E \left[\frac{1}{n} \left(\frac{\partial \ln L_2}{\partial \theta_2} \right) \left(\frac{\partial \ln L_1}{\partial \theta_1'} \right) \right].$$

Для коррекции асимптотической ковариационной матрицы на втором шаге процедуры требуются некоторые дополнительные расчеты. Матрицы \mathbf{V}_1 и \mathbf{V}_2 оцениваются по соответствующим нескорректированным ковариационным матрицам. Как правило, используются оценки ВННН

$$\hat{\mathbf{V}}_1 = \left[\frac{1}{n} \sum_{i=1}^n \left(\frac{\partial \ln f_{i1}}{\partial \hat{\theta}_1} \right) \left(\frac{\partial \ln f_{i1}}{\partial \hat{\theta}_1'} \right) \right]^{-1}$$

и $\hat{\mathbf{V}}_2 = \left[\frac{1}{n} \sum_{i=1}^n \left(\frac{\partial \ln f_{i2}}{\partial \hat{\theta}_2} \right) \left(\frac{\partial \ln f_{i2}}{\partial \hat{\theta}_2'} \right) \right]^{-1}$. Матрицы \mathbf{R} и \mathbf{C} получаются суммированием индивидуальных наблюдений попарных произведений производных. Они оцениваются с помощью

$$\hat{\mathbf{C}} = \frac{1}{n} \sum_{i=1}^n \left(\frac{\partial \ln f_{i2}}{\partial \hat{\theta}_2} \right) \left(\frac{\partial \ln f_{i2}}{\partial \hat{\theta}_1'} \right)$$

и

$$\hat{\mathbf{R}} = \frac{1}{n} \sum_{i=1}^n \left(\frac{\partial \ln f_{i2}}{\partial \hat{\theta}_2} \right) \left(\frac{\partial \ln f_{i1}}{\partial \hat{\theta}_1'} \right).$$

Вывод этого результата весьма поучителен. Мы основываемся на (14-11) и результатах подраздела 14.4.5b, где доказывалась асимптотическая нормальность оценки максимального правдоподобия. На первом шаге ML-оценка

параметра θ_1 определяется как

$$\frac{1}{n} \frac{\partial \ln L_1(\hat{\theta}_1)}{\partial \hat{\theta}_1} = \frac{1}{n} \sum_{i=1}^n \frac{\partial \ln f_1(y_{i1} | \mathbf{x}_{i1}, \hat{\theta}_1)}{\partial \hat{\theta}_1} = \frac{1}{n} \sum_{i=1}^n \mathbf{g}_{i1}(\hat{\theta}_1) = \bar{\mathbf{g}}_1(\hat{\theta}_1) = \mathbf{0}.$$

Используя результаты упомянутого раздела, мы получаем из (14-15), что асимптотическое распределение оценки — это

$$\sqrt{n}(\hat{\theta}_1 - \theta_1)d \rightarrow \left[-\mathbf{H}_{11}^{(1)}(\theta_1) \right]^{-1} \sqrt{n} \bar{\mathbf{g}}_1(\theta_1),$$

где выражение означает, что предельное распределение двух случайных векторов является таким же, как

$$\mathbf{H}_{11}^{(1)} = E \left[\frac{1}{n} \frac{\partial^2 \ln L_1(\theta_1)}{\partial \theta_1 \partial \theta_1'} \right].$$

На втором шаге ML-оценка θ_2 определяется как

$$\begin{aligned} \frac{1}{n} \frac{\partial \ln L_2(\hat{\theta}_1, \hat{\theta}_2)}{\partial \hat{\theta}_2} &= \frac{1}{n} \sum_{i=1}^n \frac{\partial \ln f_2(y_{i2} | \mathbf{x}_{i1}, \mathbf{x}_{i2}, \hat{\theta}_1, \hat{\theta}_2)}{\partial \hat{\theta}_2} \\ &= \frac{1}{n} \sum_{i=1}^n \mathbf{g}_{i2}(\hat{\theta}_1, \hat{\theta}_2) = \bar{\mathbf{g}}_2(\hat{\theta}_1, \hat{\theta}_2) = \mathbf{0}. \end{aligned}$$

Разложим вектор производных $\bar{\mathbf{g}}_2(\hat{\theta}_1, \hat{\theta}_2)$ в линейный ряд Тейлора, как обычно, и используем результаты подраздела 14.4.5b еще раз:

$$\begin{aligned} \bar{\mathbf{g}}_2(\hat{\theta}_1, \hat{\theta}_2) &= \bar{\mathbf{g}}_2(\theta_1, \theta_2) + \left[\mathbf{H}_{22}^{(2)}(\theta_1, \theta_2) \right] (\hat{\theta}_2 - \theta_2) \\ &+ \left[\mathbf{H}_{21}^{(2)}(\theta_1, \theta_2) \right] (\hat{\theta}_1 - \theta_1) + o(1/n) = \mathbf{0}, \end{aligned}$$

где

$$\mathbf{H}_{21}^{(2)}(\theta_1, \theta_2) = E \left[\frac{1}{n} \frac{\partial^2 \ln L_2(\theta_1, \theta_2)}{\partial \theta_2 \partial \theta_1'} \right] \text{ и } \mathbf{H}_{22}^{(2)}(\theta_1, \theta_2) = E \left[\frac{1}{n} \frac{\partial^2 \ln L_2(\theta_1, \theta_2)}{\partial \theta_2 \partial \theta_2'} \right].$$

Чтобы получить асимптотическое распределение, мы используем тот же метод, что и прежде:

$$\begin{aligned} \sqrt{n}(\hat{\theta}_2 - \theta_2) &\xrightarrow{d} \left[-\mathbf{H}_{22}^{(2)}(\theta_1, \theta_2) \right]^{-1} \sqrt{n} \bar{\mathbf{g}}_2(\theta_1, \theta_2) + \\ &+ \left[-\mathbf{H}_{22}^{(2)}(\theta_1, \theta_2) \right]^{-1} \left[\mathbf{H}_{21}^{(2)}(\theta_1, \theta_2) \right] \sqrt{n}(\hat{\theta}_1 - \theta_1). \end{aligned}$$

Для удобства введем обозначения $\mathbf{H}_{22}^{(2)} = \mathbf{H}_{22}^{(2)}(\theta_1, \theta_2)$, $\mathbf{H}_{21}^{(2)} = \mathbf{H}_{21}^{(2)}(\theta_1, \theta_2)$ и $\mathbf{H}_{11}^{(1)} = \mathbf{H}_{11}^{(1)}(\theta_1)$. Теперь заменим θ_1 на ее оценку из первого шага в этом выражении и получим

$$\begin{aligned} \sqrt{n}(\hat{\theta}_2 - \theta_2) &\xrightarrow{d} \left[-\mathbf{H}_{22}^{(2)} \right]^{-1} \sqrt{n} \bar{\mathbf{g}}_2(\theta_1, \theta_2) + \\ &+ \left[-\mathbf{H}_{22}^{(2)} \right]^{-1} \left[\mathbf{H}_{21}^{(2)} \right] \left[-\mathbf{H}_{11}^{(1)} \right]^{-1} \sqrt{n} \bar{\mathbf{g}}_1(\theta_1). \end{aligned}$$

Состоятельность и асимптотическая нормальность эти двух оценок следуют из полученных ранее результатов. Для того чтобы получить асимптотическую ковариационную матрицу для оценки $\hat{\theta}_2$, вычислим предельную дисперсию случайного вектора из предыдущего выражения. Совместное нормальное распределение двух векторов первых производных имеет нулевые средние и

$$\text{Var} \begin{bmatrix} \sqrt{n}\bar{g}_1(\theta_1) \\ \sqrt{n}\bar{g}_2(\theta_2, \theta_1) \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

Тогда искомая асимптотическая ковариационная матрица есть

$$\begin{aligned} \text{Var} \left[\sqrt{n}(\hat{\theta}_2 - \theta_2) \right] &= \left[-\mathbf{H}_{22}^{(2)} \right]^{-1} \Sigma_{22} \left[-\mathbf{H}_{22}^{(2)} \right]^{-1} + \\ &+ \left[-\mathbf{H}_{22}^{(2)} \right]^{-1} \left[\mathbf{H}_{21}^{(2)} \right] \left[-\mathbf{H}_{11}^{(1)} \right]^{-1} \Sigma_{11} \left[-\mathbf{H}_{11}^{(1)} \right]^{-1} \left[\mathbf{H}_{21}^{(2)} \right]' \left[-\mathbf{H}_{22}^{(2)} \right]^{-1} + \\ &+ \left[-\mathbf{H}_{22}^{(2)} \right]^{-1} \Sigma_{21} \left[-\mathbf{H}_{11}^{(1)} \right]^{-1} \left[\mathbf{H}_{21}^{(2)} \right]' \left[-\mathbf{H}_{22}^{(2)} \right]^{-1} + \\ &+ \left[-\mathbf{H}_{22}^{(2)} \right]^{-1} \left[\mathbf{H}_{21}^{(2)} \right] \left[-\mathbf{H}_{11}^{(1)} \right]^{-1} \Sigma_{12} \left[-\mathbf{H}_{22}^{(2)} \right]^{-1}. \end{aligned}$$

Как мы уже знаем, дисперсия вектора первых производных функции логарифмического правдоподобия есть математическое ожидание матрицы вторых производных, взятое со знаком минус [см. (14-11)]. Вследствие этого $\Sigma_{22} = \left[-\mathbf{H}_{22}^{(2)} \right]$ и $\Sigma_{11} = \left[-\mathbf{H}_{11}^{(1)} \right]$. Подставляя, получаем

$$\begin{aligned} \text{Var} \left[\sqrt{n}(\hat{\theta}_2 - \theta_2) \right] &= \left[-\mathbf{H}_{22}^{(2)} \right]^{-1} + \left[-\mathbf{H}_{22}^{(2)} \right]^{-1} \left[\mathbf{H}_{21}^{(2)} \right] \left[-\mathbf{H}_{11}^{(1)} \right]^{-1} \left[\mathbf{H}_{21}^{(2)} \right]' \left[-\mathbf{H}_{22}^{(2)} \right]^{-1} \\ &+ \left[-\mathbf{H}_{22}^{(2)} \right]^{-1} \Sigma_{21} \left[-\mathbf{H}_{11}^{(1)} \right]^{-1} \left[\mathbf{H}_{21}^{(2)} \right]' \left[-\mathbf{H}_{22}^{(2)} \right]^{-1} \\ &+ \left[-\mathbf{H}_{22}^{(2)} \right]^{-1} \left[\mathbf{H}_{21}^{(2)} \right] \left[-\mathbf{H}_{11}^{(1)} \right]^{-1} \Sigma_{12} \left[-\mathbf{H}_{22}^{(2)} \right]^{-1}. \end{aligned}$$

Из (14-15) $\left[-\mathbf{H}_{11}^{(1)} \right]^{-1}$ и $\left[-\mathbf{H}_{22}^{(2)} \right]^{-1}$ есть те самые матрицы \mathbf{V}_1 и \mathbf{V}_2 , которые появляются в теореме 14.8, что позволяет нам сократить выражение

$$\begin{aligned} \text{Var} \left[\sqrt{n}(\hat{\theta}_2 - \theta_2) \right] &= \mathbf{V}_2 + \mathbf{V}_2 \left[\mathbf{H}_{21}^{(2)} \right] \mathbf{V}_1 \left[\mathbf{H}_{21}^{(2)} \right]' \mathbf{V}_2 - \mathbf{V}_2 \Sigma_{21} \mathbf{V}_1 \left[\mathbf{H}_{21}^{(2)} \right]' \mathbf{V}_2 - \\ &- \mathbf{V}_2 \left[\mathbf{H}_{21}^{(2)} \right] \mathbf{V}_1 \Sigma_{12} \mathbf{V}_2. \end{aligned}$$

Два оставшихся члена — это матрица $\mathbf{H}_{21}^{(2)}$, равная $E[\partial^2 \ln L_2(\theta_1, \theta_2) / \partial \theta_2 \partial \theta_1']$, которая оценивается как $-\mathbf{C}$ в утверждении теоремы [снова обратим внимание на изменение знака из-за (14-11)], и матрица Σ_{21} , которая представляет собой ковариацию двух векторов первых производных. Она оценивается как \mathbf{R} в теореме 14.8. Подставляя, находим

$$\text{Var} \left[\sqrt{n}(\hat{\theta}_2 - \theta_2) \right] = \mathbf{V}_2 + \mathbf{V}_2 \mathbf{C} \mathbf{V}_1 \mathbf{C}' \mathbf{V}_2 - \mathbf{V}_2 \mathbf{R} \mathbf{V}_1 \mathbf{C}' \mathbf{V}_2 - \mathbf{V}_2 \mathbf{C} \mathbf{V}_1 \mathbf{R}' \mathbf{V}_2,$$

что завершает доказательство теоремы.

Пример 14.5. Двухшаговая ML-оценка

Распространенное применение двухшагового метода связано с дисперсией построенного регрессора на втором шаге модели. В этом случае построенная переменная часто становится оценкой ожидаемого значения переменной, которая, вероятно, является эндогенной для второго шага. В этом примере мы построим простую модель для иллюстрации этих расчетов.

В работе Рипхана, Вомбаха, Мийона (Riphahn, Wambach, Million) (RWM, 2003) авторы исследовали использование немецкой системы здравоохранения. Они выясняли, объясняет ли это использование (по крайней мере частично) факт покупки индивидами конкретного вида дополнительного страхования здоровья. Мы уже использовали набор данных Немецкой социоэкономической панели (German Socioeconomic Panel (GSOEP)) в нескольких местах по ходу книги (см., например, пример 7.6). Одной из интересующих нас переменных в исследовании является *DocVis* — количество посещений индивидом врача в течение года медицинского обследования. Авторы предположили, что наличие дополнительного (*Addon*) страхования могло оказать влияние на число посещений. Наша простая модель формулируется следующим образом: число посещений описывается пуассоновской регрессией (см. подраздел 18.4.1). Это логлинейная модель, которую мы задаем как

$$E[DocVis|x_2, P_{Addon}] = \mu(x'_2\beta, \gamma, x'_1\alpha) = \exp[x'_2\beta + \gamma\Lambda(x'_1\alpha)].$$

Мы не включаем в модель дамми-переменную, равную 1, если человек имеет *Addon*-страховку, и 0 в противном случае, поскольку она, по всей видимости, является эндогенной для этого уравнения. Вместо этого модель содержит оценку $E[Addon|x_1]$, полученную из **логистической модели вероятности** (logistic probability model) (см. раздел 17.2) того, имел ли индивид эту страховку:

$$\Lambda(x'_1\alpha) = \frac{\exp(x'_1\alpha)}{1 + \exp(x'_1\alpha)} = \text{Prob} [\text{индивид приобрел } Addon\text{-страховку} | x_1].$$

Для этого примера определим

$$\begin{aligned} (y_1 = Addon)x_1 &= (\text{constant}, \text{Age}, \text{Education}, \text{Married}, \text{Kids}), \\ (y_2 = DocVis)x_2 &= (\text{constant}, \text{Age}, \text{Education}, \text{Income}, \text{Female}). \end{aligned}$$

Как и ранее, чтобы обойти проблемы, связанные с характером панельных данных, мы будем использовать 4483 наблюдения по данным 1988 г. и исключим два наблюдения, для которых *Income* (доход) равен нулю.

Логарифмическое правдоподобие для логистической модели вероятности записывается как

$$\ln L_1(\alpha) = \sum_i \{(1 - y_{i1}) \ln[1 - \Lambda(x'_{i1}\alpha)] + y_{i1} \ln \Lambda(x'_{i1}\alpha)\}.$$

Производные этого логарифмического правдоподобия выражаются следующим образом:

$$\mathbf{g}_{i1}(\alpha) = \partial \ln f_1(y_{i1} | \mathbf{x}_{i1}, \alpha) / \partial \alpha = [y_{i1} - \Lambda(\mathbf{x}'_{i1} \alpha)] \mathbf{x}_{i1}.$$

Мы будем максимизировать это логарифмическое правдоподобие по α , а затем вычислим \mathbf{V}_1 с помощью ВНН-оценки, следуя теореме 14.8. Для вычисления \mathbf{R} мы будем использовать $\mathbf{g}_{i1}(\alpha)$.

Логарифмическое правдоподобие для модели пуассоновской регрессии записывается как

$$\ln L_2 = \sum_i [-\mu(\mathbf{x}'_{i2} \beta, \gamma, \mathbf{x}'_{i1} \alpha) + y_{i2} \ln \mu(\mathbf{x}'_{i2} \beta, \gamma, \mathbf{x}'_{i1} \alpha) - \ln y'_{i2}].$$

Производные этого логарифмического правдоподобия выглядят следующим образом:

$$\begin{aligned} \mathbf{g}_{i2}^{(2)}(\beta, \gamma, \alpha) &= \partial \ln f_2(y_{i2}, \mathbf{x}_{i1}, \mathbf{x}_{i2}, \beta, \gamma, \alpha) / \partial (\beta', \gamma)' = \\ &= [y_{i2} - \mu(\mathbf{x}'_{i2} \beta, \gamma, \mathbf{x}'_{i1} \alpha)] [\mathbf{x}'_{i2}, \Lambda(\mathbf{x}'_{i1} \alpha)]' \\ \mathbf{g}_{i1}^{(2)}(\beta, \gamma, \alpha) &= \partial \ln f_2(y_{i2}, \mathbf{x}_{i1}, \mathbf{x}_{i2}, \beta, \gamma, \alpha) / \partial \alpha = \\ &= [y_{i2} - \mu(\mathbf{x}'_{i2} \beta, \gamma, \mathbf{x}'_{i1} \alpha)] \gamma \Lambda(\mathbf{x}'_{i1} \alpha) [1 - \Lambda(\mathbf{x}'_{i1} \alpha)] \mathbf{x}_{i1}. \end{aligned}$$

Мы будем использовать $\mathbf{g}_{i2}^{(2)}$ для вычисления \mathbf{V}_2 , а также \mathbf{R} , \mathbf{C} и $\mathbf{g}_{i1}^{(2)}$ для вычисления \mathbf{C} . В частности,

$$\begin{aligned} \mathbf{V}_1 &= [(1/n) \sum_i \mathbf{g}_{i1}(\alpha) \mathbf{g}_{i1}(\alpha)']^{-1}, \\ \mathbf{V}_2 &= [(1/n) \sum_i \mathbf{g}_{i2}^{(2)}(\beta, \gamma, \alpha) \mathbf{g}_{i2}^{(2)}(\beta, \gamma, \alpha)']^{-1}, \\ \mathbf{C} &= [(1/n) \sum_i \mathbf{g}_{i2}^{(2)}(\beta, \gamma, \alpha) \mathbf{g}_{i1}^{(2)}(\beta, \gamma, \alpha)'], \\ \mathbf{R} &= [(1/n) \sum_i \mathbf{g}_{i2}^{(2)}(\beta, \gamma, \alpha) \mathbf{g}_{i1}(\alpha)']. \end{aligned}$$

В табл. 14.2 приводятся оценки двухшаговым методом максимального правдоподобия для параметров моделей вместе с оцененными стандартными ошибками. Для логистической модели, используемой на первом шаге, стандартные ошибки, обозначенные как \mathbf{H}_1 и \mathbf{V}_1 , сравнивают со значениями, вычисленными с использованием обратной матрицы вторых производных (со знаком минус) (\mathbf{H}_1) и с использованием внешнего произведения первых производных (\mathbf{V}_1). Как и ожидалось, для выборки такого объема различие является незначительным. Вторая матрица использовалась при вычислении скорректированной ковариационной матрицы на втором шаге. В пуассоновской модели, сравнивая \mathbf{V}_2 с \mathbf{V}_2^* , мы видим, что учет наличия $\hat{\alpha}$ в построенном регрессоре имеет существенное влияние на стандартные ошибки даже для этой относительно большой выборки. Обратим внимание, что результатом коррекции является удвоение стандартных ошибок коэффициентов для переменных, содержащихся в обоих уравнениях, но изменение довольно незначительно для переменных *Income* и *Female*, которые встречаются лишь во втором уравнении.

Таблица 14.2. Оценки логистической и пуассоновской моделей

| | Логистическая модель для Addon | | | Пуассоновская модель для DocVis | | |
|------------------------|-----------------------------------|-------------------------------|-------------------------------|------------------------------------|-------------------------------|---------------------------------|
| | Кэф- фициент | Станд. ошибка (H_1) | Станд. ошибка (V_1) | Кэф- фициент | Станд. ошибка (V_2) | Станд. ошибка (V_2^*) |
| Постоянная | -6,19246 | 0,60228 | 0,58287 | 0,77808 | 0,04884 | 0,09319 |
| Возраст | 0,01486 | 0,00912 | 0,00924 | 0,01752 | 0,00044 | 0,00111 |
| Образование | 0,16091 | 0,03003 | 0,03326 | -0,03858 | 0,00462 | 0,00980 |
| Женат/за- мужем | 0,22206 | 0,23584 | 0,23523 | | | |
| Дети | -0,10822 | 0,21591 | 0,21993 | | | |
| Доход | | | | -0,80298 | 0,02339 | 0,02719 |
| Женщины | | | | 0,16409 | 0,00601 | 0,00770 |
| $\Lambda(x_1' \alpha)$ | | | | 3,91140 | 0,77283 | 1,87014 |

Ковариация двух градиентов \mathbf{R} может стремиться к нулю в конкретной модели. Если оценки на первом и втором шагах основываются на различных выборках, то \mathbf{R} является в точности нулевым. Например, в модели, рассмотренной нами выше, \mathbf{R} основан на двух остатках:

$$g_{i1} = \{Addon_i - E[Addon_i | x_{i1}]\} \text{ и } g_{i2}^{(2)} = \{DocVis_i - E[DocVis_i | x_{i2}, \Lambda_{i1}]\}.$$

Эти остатки могут быть некоррелированными. Это предположение может быть проверено, но в таком случае третий и четвертый члены в V_2 асимптотически стремятся к нулю и остается более простое альтернативное выражение

$$V_2^{**} = (1/n)[V_2 + V_2 C V_1 C' V_2].$$

(В нашем примере выборочная корреляция между g_{i1} и $g_{i2}^{(2)}$ составляет лишь 0,015658 и элементы оценки \mathbf{R} равны около 0,01, умноженных на соответствующие элементы \mathbf{C} : по сути, около 99% корректировки матрицы V_2^* происходит из-за \mathbf{C} .)

Есть мнение, что этот набор процедур является более сложным, чем это необходимо. [Например, Cameron, Trivedi (2005, p. 202).] Существует два альтернативных подхода, которые можно использовать. Во-первых, при достаточно широких условиях асимптотическая ковариационная матрица оценки на втором шаге может быть аппроксимирована с использованием процедуры бутстрэпа, которая будет нами рассмотрена в разделе 15.4. Заметим, однако, что если применяется этот подход, то весьма важно, чтобы оба шага были «бутстрэпированы». В противном случае, расценивая $\hat{\theta}_1$ как данную и фиксированную величину, мы будем оценивать $(1/n)V_2$, а не нужную нам ковариационную матрицу. Задача здесь состоит в том, чтобы принять во внимание вариацию $\hat{\theta}_1$. Вторая возможность состоит в том, чтобы оценивать сразу всю модель. Таким образом, требуется использовать одношаговую оценку максимального правдоподобия с полной информацией и оценить θ_1 и θ_2 одновременно. Конечно, обычно это тот самый метод, кото-

рого мы и пытались избежать вначале. Используя современное программное обеспечение, рассмотренный двухшаговый метод часто можно применить непосредственно. И, несмотря на его недостатки, иногда он оказывается весьма полезен. Так, напомним еще раз, одним из наиболее классических случаев является знаменитая модель выборочной селективности Хекмана (Heckman, 1979). Двухшаговые оценки и оценки с полной информацией для модели Хекмана рассматриваются нами в подразделе 19.5.3.

14.8. Оценка методом квазикасимального правдоподобия и устойчивые асимптотические ковариационные матрицы

Оценивание методом максимального правдоподобия требует полного описания распределения наблюдаемой случайной переменной. Если истинное распределение не соответствует нашему предположению о нем, то функция правдоподобия задается неправильно и желаемые свойства ML-оценки могут не выполняться. В этом разделе мы рассматриваем ряд результатов, касающихся подхода к оцениванию, который является робастным (устойчивым) для некоторых видов неправильной спецификации моделей. Так, например, мы определили, что в модели функции условного среднего вида $E[y | \mathbf{x}] = \mathbf{x}'\beta$ некоторые оценки, такие как оценка методом наименьших квадратов, являются робастными к неверной спецификации распределения шоков. Таким образом, LS-оценка есть ML-оценка, если шоки распределены нормально, но тем не менее мы все еще можем говорить о наличии некоторых желаемых свойств LS, включая состоятельность, даже в том случае, если шоки не распределены нормально. В этом разделе мы рассматриваем некоторые результаты, относящиеся к случаю, если мы максимизируем «неверную» логарифмическую функцию правдоподобия, и изучаем ситуации, при которых оценка является состоятельной несмотря на это, и решаем, как вычислить соответствующую асимптотическую ковариационную матрицу в этом случае¹².

14.8.1. Максимальное правдоподобие и GMM-оценки

Пусть $f(y_i | \mathbf{x}_i, \beta)$ есть истинная плотность вероятности для случайной величины y_i при заданном наборе независимых переменных \mathbf{x}_i и векторе параметров β . Логарифмическая функция правдоподобия есть

$$(1/n) \ln L(\beta | \mathbf{y}, \mathbf{X}) = (1/n) \sum_{i=1}^n \ln f(y_i | \mathbf{x}_i, \beta).$$

¹²Дальнейшее изложение описывает в общих чертах ряд результатов, относящихся к этой задаче оценивания. Важными источниками по данной теме являются работы White (1982a), Gourieroux, Monfort, Trognon (1984), Huber (1967) и Amemiya (1985). Работа Mittelhammer et al. (2000) является недавним исследованием, содержащим значительный материал по данной теме. Выводы результатов в этих работах сложные, и наша цель состоит лишь в попытке дать интуитивное введение в эту тему.

ML-оценка $\hat{\beta}_{ML}$ есть та выборочная статистика, которая максимизирует эту функцию. (Деление $\ln L$ на n не влияет на решение.) Мы максимизируем логарифмическую функцию правдоподобия путем приравнивания ее производных к нулю, поэтому ML-оценка получается с помощью решения системы уравнений эмпирических (выборочных) моментов

$$\frac{1}{n} \sum_{i=1}^n \frac{\partial \ln f(y_i | \mathbf{x}_i, \hat{\beta}_{ML})}{\partial \hat{\beta}_{ML}} = \frac{1}{n} \sum_{i=1}^n \mathbf{d}_i(\hat{\beta}_{ML}) = \bar{\mathbf{d}}(\hat{\beta}_{ML}) = \mathbf{0}.$$

Популяционный аналог для уравнения выборочных моментов выглядит как

$$E \left[\frac{1}{n} \frac{\partial \ln L}{\partial \beta} \right] = E \left[\frac{1}{n} \sum_{i=1}^n \mathbf{d}_i(\beta) \right] = E [\bar{\mathbf{d}}(\beta)] = \mathbf{0}.$$

Как мы знаем из свойств GMM-оценок, если $E [\bar{\mathbf{d}}(\beta)] = \mathbf{0}$, то тогда $\hat{\beta}_{ML}$ является состоятельной и асимптотически нормально распределенной с асимптотической ковариационной матрицей, равной

$$\mathbf{V}_{ML} = [\mathbf{G}(\beta)' \mathbf{G}(\beta)]^{-1} \mathbf{G}(\beta)' \{ \text{Var}[3\bar{\mathbf{d}}(\beta)] \} \mathbf{G}(\beta) [\mathbf{G}(\beta)' \mathbf{G}(\beta)]^{-1},$$

где $\mathbf{G}(\beta) = \text{plim } \partial \bar{\mathbf{d}}(\beta) / \partial \beta'$. Поскольку $\bar{\mathbf{d}}(\beta)$ есть вектор производных, $\mathbf{G}(\beta)$ есть умноженное на $1/n$ математическое ожидание гессиана функции $\ln L$, т.е. $(1/n)E [\mathbf{H}(\beta)] = \bar{\mathbf{H}}(\beta)$. Как мы знаем, $\text{Var}[\partial \ln L / \partial \beta] = -E [\mathbf{H}(\beta)]$. Приводя все семь матриц $(1/n)E [\mathbf{H}(\beta)]$, мы приходим к известному нам результату: $\mathbf{V}_{ML} = \{ -E [\mathbf{H}(\beta)] \}^{-1}$. [Все n сокращаются, а $\text{Var}[\bar{\mathbf{d}}] = 1/n \bar{\mathbf{H}}(\beta)$.] Обратим внимание, что этот результат существенно опирается на тот факт, что $\text{Var}[\partial \ln L / \partial \beta] = -E [\mathbf{H}(\beta)]$.

14.8.2. Максимальное правдоподобие и M-оценки

Оценка методом максимального правдоподобия получается путем максимизации функции $\bar{h}_n(\mathbf{y}, \mathbf{X}, \beta) = (1/n) \sum_{i=1}^n \ln f(y_i, \mathbf{x}_i, \beta)$. Эта функция сходится к своему математическому ожиданию при $n \rightarrow \infty$. Поскольку эта функция является логарифмическим правдоподобием для выборки, так же верно (здесь мы это не доказываем), что при $n \rightarrow \infty$ она достигает своего единственного максимума в векторе истинных параметров β . (Мы использовали этот результат при доказательстве состоятельности оценки методом максимального правдоподобия.) Поскольку $\text{plim } \bar{h}_n(\mathbf{y}, \mathbf{X}, \beta) = E [\bar{h}_n(\mathbf{y}, \mathbf{X}, \beta)]$, отсюда следует (если поменять местами операторы дифференцирования и математического ожидания), что $\text{plim } \partial \bar{h}_n(\mathbf{y}, \mathbf{X}, \beta) / \partial \beta = E [\partial \bar{h}_n(\mathbf{y}, \mathbf{X}, \beta) / \partial \beta]$. Однако, поскольку эта функция достигает своего максимума в точке β , это значит, что

$$\text{plim } \partial \bar{h}_n(\mathbf{y}, \mathbf{X}, \beta) / \partial \beta = \mathbf{0}.$$

Оценка, которая получается путем максимизации некоторого функционального критерия, называется *M-оценкой* [Huber (1967)] или экстремальной оценкой [Amemiya (1985)]. Предположим, что мы построили некоторую

оценку путем максимизации некоторой другой функции, $M_n(\mathbf{y}, \mathbf{X}, \beta)$, которая, хотя и не является логарифмической функцией правдоподобия, также достигает своего единственного максимума в истинном векторе β при $n \rightarrow \infty$. Тогда на основании предыдущего рассуждения можно получить состоятельную оценку с известным асимптотическим распределением. Например, логарифмическое правдоподобие для модели линейной регрессии с нормально распределенными шоками с разными дисперсиями, $\sigma^2 \omega_i$, может быть записано как

$$\bar{h}_n(\mathbf{y}, \mathbf{X}, \beta) = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{-1}{2} \left[\ln(2\pi\sigma^2\omega_i) + \frac{(y_i - \mathbf{x}'_i\beta)^2}{\sigma^2\omega_i} \right] \right\}.$$

Максимизируя эту функцию, мы получаем оценку методом максимального правдоподобия. Но мы также знакомы с другой оценкой — простым методом наименьших квадратов, которая максимизирует $M_n(\mathbf{y}, \mathbf{X}, \beta) = -(1/n) \sum_{i=1}^n (y_i - \mathbf{x}'_i\beta)^2$. Как уже было доказано, оценка методом наименьших квадратов является состоятельной и асимптотически нормально распределенной даже при этом обобщении, поэтому она квалифицируется как M -оценка в том смысле, в котором мы его здесь понимаем.

Рассмотрим теперь общий случай. Предположим, что мы оцениваем β при помощи максимизации следующей функции критерия:

$$M_n(\mathbf{y} | \mathbf{X}, \beta) = \frac{1}{n} \sum_{i=1}^n \ln g(y_i | \mathbf{x}_i, \beta).$$

Также предположим, что $\text{plim} M_n(\mathbf{y}, \mathbf{X}, \beta) = E[M_n(\mathbf{y} | \mathbf{X}, \beta)]$ и что при $n \rightarrow \infty$, $E[M_n(\mathbf{y} | \mathbf{X}, \beta)]$ достигает своего единственного максимума в точке β . Тогда на основании тех же рассуждений, которые мы ранее использовали для ML оценки, $\text{plim} \partial M_n(\mathbf{y} | \mathbf{X}, \beta) / \partial \beta = E[\partial M_n(\mathbf{y} | \mathbf{X}, \beta) / \partial \beta] = \mathbf{0}$. Как и ранее, у нас есть система уравнений для моментов оценивания. Пусть $\hat{\beta}_E$ — это оценка, максимизирующая $M_n(\mathbf{y} | \mathbf{X}, \beta)$. Тогда эта оценка определяется как

$$\frac{\partial M_n(\mathbf{y} | \mathbf{X}, \hat{\beta}_E)}{\partial \hat{\beta}_E} = \frac{1}{n} \sum_{i=1}^n \frac{\partial \ln g(y_i | \mathbf{x}_i, \hat{\beta}_E)}{\partial \hat{\beta}_E} = \bar{m}(\hat{\beta}_E) = \mathbf{0}.$$

Таким образом, $\hat{\beta}_E$ является GMM-оценкой. Используя систему обозначений, введенную ранее, $\mathbf{G}(\hat{\beta}_E)$ есть симметричный гессиан функции $E[M_n(\mathbf{y}, \mathbf{X}, \beta)]$, который мы обозначим как $(1/n)E[\mathbf{H}_M(\hat{\beta}_E)] = \bar{\mathbf{H}}_M(\hat{\beta}_E)$. Применяв выкладки, аналогичные тем, которые использовались для получения \mathbf{V}_{ML} , мы находим, что соответствующая асимптотическая ковариационная матрица для экстремальной оценки будет иметь вид

$$\mathbf{V}_E = [\bar{\mathbf{H}}_M(\beta)]^{-1} \left(\frac{1}{n} \Phi \right) [\bar{\mathbf{H}}_M(\beta)]^{-1},$$

где $\Phi = \text{Var}[\partial \log g(y_i | \mathbf{x}_i, \beta) / \partial \beta]$ и, как и ранее, асимптотическое распределение является нормальным.

Гессиан \mathbf{V}_E может легко быть оценен с помощью своего эмпирического аналога

$$\text{Est.}[\bar{\mathbf{H}}_M(\hat{\beta}_E)] = \frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \ln g(y_i | \mathbf{x}_i, \hat{\beta}_E)}{\partial \hat{\beta}_E \partial \hat{\beta}'_E}.$$

Однако остается необходимость в определении Φ , и маловероятно, что возникнет ситуация, в которой мы будем знать, какую функцию следует использовать. Важное отличие этого случая состоит в том, что здесь дисперсия вектора первых производных необязательно будет равняться гесссиану, поэтому матрица V_E не упрощается. Тем не менее мы можем состоятельно оценить Φ , используя выборочную дисперсию первых производных:

$$\hat{\Phi} = \frac{1}{n} \sum_{i=1}^n \left[\frac{\partial \ln g(y_i | \mathbf{x}_i, \hat{\beta})}{\partial \hat{\beta}} \right] \left[\frac{\partial \ln g(y_i | \mathbf{x}_i, \hat{\beta})}{\partial \hat{\beta}'} \right]'$$

Если бы это была оценка методом максимального правдоподобия, то $\hat{\Phi}$ была бы ОРГ-оценкой, которую мы использовали в нескольких местах ранее. Например, для оценки методом наименьших квадратов в гетероскедастичной модели линейной регрессии критерием является $M_n(\mathbf{y}, \mathbf{X}, \beta) = -(1/n) \sum_{i=1}^n (y_i - \mathbf{x}_i' \beta)^2$, а решением — \mathbf{b} , $\mathbf{G}(\mathbf{b}) = (-2/n) \mathbf{X}' \mathbf{X}$, и

$$\hat{\Phi} = \frac{1}{n} \sum_{i=1}^n [2\mathbf{x}_i (y_i - \mathbf{x}_i' \beta)] [2\mathbf{x}_i (y_i - \mathbf{x}_i' \beta)]' = \frac{4}{n} \sum_{i=1}^n e_i^2 \mathbf{x}_i \mathbf{x}_i'$$

После приведения подобных членов и сокращения 4 мы имеем в точности оценку Уайта (9-27)!

14.8.3. «Сэндвич»-оценки

Теперь рассмотрим мотивацию, лежащую в основе всей этой громоздкой теории. Один из недостатков метода максимального правдоподобия заключается в наличии требования, что плотность наблюдаемой случайной переменной (или переменных) должна быть полностью специфицирована. Ранее мы видели, что в некоторых ситуациях можем обойтись меньшими предположениями о распределении по сравнению с полным описанием. Экстремальная оценка становится устойчивой к некоторым видам ошибок спецификации. Одним полезным и необходимым результатом является оценка для асимптотической ковариационной матрицы экстремальной оценки, которая была бы устойчивой, по крайней мере, к некоторой неправильной спецификации. В частности, если мы получаем $\hat{\beta}_E$ путем максимизации некоторого критерия, который удовлетворяет остальным нашим предположениям, то соответствующая оценка асимптотической ковариационной матрицы представляет собой

$$\text{Est. } V_E = \frac{1}{n} [\bar{\mathbf{H}}(2\hat{\beta}_E)]^{-1} \hat{\Phi}(\hat{\beta}_E) [\bar{\mathbf{H}}(2\hat{\beta}_E)]^{-1}$$

Если $\hat{\beta}_E$ есть истинная ML-оценка, то V_E упрощается до $\{-[\mathbf{H}(\hat{\beta}_E)]\}^{-1}$. В современной литературе ее часто называют «сэндвич»-оценкой (sandwich estimator). Тенденция, обозначившаяся в современной литературе, состоит в том, чтобы рассчитывать эту оценку непосредственно, т. е. без учета функции правдоподобия. Стоит заметить, что если логарифмическое правдоподобие задано неправильно, то, по всей видимости, оценки параметров будут

несостоятельными, за исключением случаев, рассмотренных ранее, поэтому робастное оценивание асимптотической ковариационной матрицы может быть совершенно излишним. В то же время если функция правдоподобия задана верно, то «сэндвич»-оценка является лишней. Этот метод не есть общий метод борьбы с неправильно специфицированными моделями. Не каждую функцию правдоподобия можно рассматривать как состоятельную экстремальную оценку *интересующих нас параметров в модели*.

Здесь можно задаться вопросом, насколько реально, что будут соблюдаться условия, необходимые для применения этой теории. В литературе, использующей этот механизм, есть модели, которые, вероятно, не удовлетворяют этим условиям, такие как тобит-модель в главе 19. Один важный случай мы уже рассмотрели. Метод наименьших квадратов в обобщенной регрессионной модели подходит под эти условия. Другой важный пример — это модели «индивидуальной неоднородности» в межобъектных данных. Есть свидетельства, что в простых моделях часто не учитываются ненаблюдаемые источники вариации индивидов в межобъектных данных, такие как неизмеримые «эффекты семьи» в исследованиях доходов или занятости. Предположим, что корректная модель есть $h(y_i | \mathbf{x}_i, v_i, \beta, \theta)$, где v_i является ненаблюдаемым случайным членом, а θ — это параметр распределения v . Корректная логарифмическая функция правдоподобия выглядит как $\sum_i \ln f(y_i | \mathbf{x}_i, \beta, \theta) = \sum_i \ln \int_v h(y_i | \mathbf{x}_i, v_i, \beta, \theta) f(v_i) dv_i$. Допустим, что мы максимизируем некоторую другую **квазилогарифмическую функцию правдоподобия** (pseudo-log-likelihood function), $\sum_i \ln g(y_i | \mathbf{x}_i, \beta)$, а затем используем «сэндвич»-оценку, чтобы оценить асимптотическую ковариационную матрицу вектора $\hat{\beta}$. Даст ли это состоятельную оценку вектора истинных значений параметров? Как ни удивительно, иногда это так, хотя зашумляющий параметр θ не был принят нами во внимание. Мы уже видели один случай, когда использовали OLS в GR-модели с гетероскедастичными шоками. Некорректное использование модели Пуассона, когда верной является отрицательная биномиальная модель (см. подраздел 18.4.4), представляет другой случай. Для некоторых спецификаций использование ошибочной функции правдоподобия в пробит-модели с данными о количественных соотношениях является третьим случаем. [Эти примеры вместе с несколькими другими предложены в работе Gourieroux, Monfort, Trognon (1984).] Подчеркнем еще раз, что «сэндвич»-оценка сама по себе необязательно имеет какие-то достоинства, если функция правдоподобия задана неправильно, а другие условия для M -оценки не соблюдены.

14.8.4. Кластерные оценки

Данные на микроуровне, или индивидуальные данные, часто группируются или разбиваются на кластеры. Модель производства или экономического успеха на уровне фирмы может основываться на группе отраслей с разнообразными фирмами в каждой отрасли. Анализ уровня образования студентов может быть основан на выборках целых классов или школ или средних значений школ внутри школьных округов в масштабе штата. И конечно, такая «кластеризация» является определяющим признаком панельного набора данных. Мы рассматривали некоторые такие модели при на-

шем исследовании панельных данных в главе 11. Современная литература содержит много исследований с «кластеризованными» данными, в которых исследователь оценивает объединенную модель (модель пула), но стремится учесть ожидаемую корреляцию между наблюдениями с помощью коррекции асимптотической ковариационной матрицы. Мы использовали этот подход при вычислении робастной ковариационной матрицы для оценки пула методом наименьших квадратов в модели панельных данных [см. (11-3) и пример 11.1 в подразделе 11.6.4].

Для нормальной модели линейной регрессии логарифмическое правдоподобие, которое мы максимизируем с помощью оценки методом наименьших квадратов модели пула, может быть записано как

$$\ln L = \sum_{i=1}^n \sum_{t=1}^{T_i} \left[-\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln \sigma^2 - \frac{1}{2} \frac{(y_{it} - \mathbf{x}'_{it}\boldsymbol{\beta})^2}{\sigma^2} \right].$$

[См. (14-34).] «Кластерно-робастная» оценка (11-3) может быть записана как

$$\begin{aligned} \mathbf{W} &= \left(\sum_{i=1}^n \mathbf{X}'_i \mathbf{X}_i \right)^{-1} \left[\sum_{i=1}^n (\mathbf{X}'_i \mathbf{e}_i)(\mathbf{e}'_i \mathbf{X}_i) \right] \left(\sum_{i=1}^n \mathbf{X}'_i \mathbf{X}_i \right)^{-1} \\ &= \left(-\frac{1}{\hat{\sigma}^2} \sum_{i=1}^n \sum_{t=1}^{T_i} \mathbf{x}_{it} \mathbf{x}'_{it} \right)^{-1} \left[\sum_{i=1}^n \left(\sum_{t=1}^{T_i} \frac{1}{\hat{\sigma}^2} \mathbf{x}_{it} e_{it} \right) \left(\sum_{t=1}^{T_i} \frac{1}{\hat{\sigma}^2} e_{it} \mathbf{x}'_{it} \right) \right] \times \\ &\times \left(-\frac{1}{\hat{\sigma}^2} \sum_{i=1}^n \sum_{t=1}^{T_i} \mathbf{x}_{it} \mathbf{x}'_{it} \right)^{-1} = \left(\sum_{i=1}^n \sum_{t=1}^{T_i} \frac{\partial^2 \ln f_{it}}{\partial \hat{\boldsymbol{\beta}} \partial \hat{\boldsymbol{\beta}}'} \right)^{-1} \times \\ &\times \left[\sum_{i=1}^n \left(\sum_{t=1}^{T_i} \frac{\partial \ln f_{it}}{\partial \hat{\boldsymbol{\beta}}} \right) \left(\sum_{t=1}^{T_i} \frac{\partial \ln f_{it}}{\partial \hat{\boldsymbol{\beta}}'} \right) \right] \left(\sum_{i=1}^n \sum_{t=1}^{T_i} \frac{\partial^2 \ln f_{it}}{\partial \hat{\boldsymbol{\beta}} \partial \hat{\boldsymbol{\beta}}'} \right)^{-1}, \end{aligned}$$

где f_{it} — плотность нормального распределения со средним $\mathbf{x}'_{it}\boldsymbol{\beta}$ и дисперсией σ^2 . Это в точности та самая «кластерно-скорректированная» робастная ковариационная матрица, которая повсеместно используется в литературе, [кроме специальной (ad hoc) «поправки на конечность выборки», как в (11-4)].

В обобщенной модели линейной регрессии (как и в других) OLS-оценка является состоятельной и имеет асимптотическую ковариационную матрицу, равную

$$\text{Asy. Var}[\mathbf{b}] = (\mathbf{X}'\mathbf{X})^{-1} [\mathbf{X}'(\sigma^2\boldsymbol{\Omega})\mathbf{X}](\mathbf{X}'\mathbf{X})^{-1}.$$

(См. теорему 9.1.) Центральная матрица в «сэндвиче» для случая панельных данных может быть записана как

$$\mathbf{X}'(\sigma^2\boldsymbol{\Omega})\mathbf{X} = \sum_{i=1}^n \mathbf{X}'_i \boldsymbol{\Sigma} \mathbf{X}_i,$$

что объясняет вид предыдущей робастной оценки. Когда мы впервые столкнулись с ней, то обосновали кластерную оценку, пользуясь той же логикой, которая привела нас к оценке Уайта для гетероскедастичности, а сейчас у нас

есть дополнительный результат, который позволяет обосновать эту оценку в терминах функции правдоподобия.

Рассмотрим ошибку спецификации модели, которую должна учитывать наша оценка. Предположим, что наблюдения в группе i распределены многомерно нормально с вектором математического ожидания шоков, равным $\mathbf{0}$, и $T_i \times T_i$ ковариационной матрицей Σ_i , на которую не накладываются никакие ограничения. В этом случае соответствующая логарифмическая функция правдоподобия имеет вид

$$\ln L = \sum_{i=1}^n \left(-T_i/2 \ln 2\pi - \frac{1}{2} \ln |\Sigma_i| - \frac{1}{2} \boldsymbol{\varepsilon}_i' \Sigma_i^{-1} \boldsymbol{\varepsilon}_i \right),$$

где $\boldsymbol{\varepsilon}_i$ есть $T_i \times 1$ вектор шоков для индивида i . Таким образом, мы максимизировали ошибочную функцию правдоподобия. Действительно, β , которая максимизирует логарифмическую функцию правдоподобия, есть GLS-оценка, а не OLS-оценка. OLS- и кластерно скорректированная оценка, рассмотренная выше, «работают» в том смысле, что (1) оценка методом наименьших квадратов состоятельна, несмотря на неверную спецификацию, и (2) робастная оценка действительно оценивает нужную асимптотическую ковариационную матрицу.

Теперь рассмотрим более общий случай. Предположим, что набор данных состоит из n многомерных наблюдений $[y_{i,1}, \dots, y_{i,T_i}]$, $i = 1, \dots, n$. Каждый кластер был взят из генеральной совокупности с совместной плотностью $f_i(\mathbf{y}_i | \mathbf{X}_i, \boldsymbol{\theta})$. Вновь, чтобы сохранить общий характер результата, мы предполагаем, что размеры кластеров могут быть различны. Соответствующее логарифмическое правдоподобие для выборки есть

$$\ln L = \sum_{i=1}^n \ln f_i(\mathbf{y}_i | \mathbf{X}_i, \boldsymbol{\theta}).$$

Вместо максимизации $\ln L$ мы максимизируем квазилогарифмическое правдоподобие

$$\ln L_P = \sum_{i=1}^n \sum_{t=1}^{T_i} \ln g(y_{it} | \mathbf{x}_{it}, \boldsymbol{\theta}),$$

где предполагаем (возможно, необоснованно), что тот же самый вектор параметров $\boldsymbol{\theta}$ входит и в квазилогарифмическое правдоподобие, так же как он входит и в корректное правдоподобие. Предположим, что это верно. Используя хорошо известную нам асимптотику первого порядка, **оценка методом квазикасимимального правдоподобия** (pseudo-maximum likelihood estimator) (PML) будет удовлетворять

$$\begin{aligned} (\hat{\boldsymbol{\theta}}_{\text{P.ML}} - \boldsymbol{\theta}) &\approx \left(\frac{-1}{\sum_{i=1}^n T_i} \sum_{i=1}^n \sum_{t=1}^{T_i} \frac{\partial^2 \ln f_{it}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right)^{-1} \left(\frac{1}{\sum_{i=1}^n T_i} \sum_{i=1}^n \sum_{t=1}^{T_i} \frac{\partial \ln f_{it}}{\partial \boldsymbol{\theta}} \right) + \\ &+ (\boldsymbol{\theta} - \boldsymbol{\beta}) = \left(\frac{-1}{\sum_{i=1}^n T_i} \sum_{i=1}^n \sum_{t=1}^{T_i} H_{it} \right)^{-1} \left(\sum_{i=1}^n w_i \bar{\mathbf{g}}_i \right) + (\boldsymbol{\theta} - \boldsymbol{\beta}), \end{aligned}$$

где $w_i = T_i / \sum_{i=1}^n T_i$ и $\bar{g}_i = (1/T_i) \sum_{t=1}^{T_i} \partial \ln f_{it} / \partial \theta$. Последний член включен в выражение, для того чтобы рассмотреть возможный случай $\text{plim } \hat{\theta}_{P,ML} = \beta$, которое может не быть равно θ . [Обратим внимание, что, например, в работе Кэмерона и Триверди (Cameron, Trivedi (2005, p. 842)) авторы специально предполагают состоятельность в общей модели, которую они рассматривают.] Взятие математического ожидания внешнего произведения этого выражения, что необходимо для оценки квадрата отклонения асимптотического среднего, дает два члена: перекрестный член обращается в ноль. Первый член есть кластерно-скорректированная матрица, которая повсеместно встречается в современной литературе. Второй есть квадратичная ошибка, которая может продолжать существовать при возрастающем n по той причине, что при квазиML-оценивании может не оценивать параметры интересующей нас модели.

Итак, можно сделать два вывода. Мы можем обосновать кластерную оценку на этой аппроксимации. В общем случае она будет оценивать математическое ожидание квадрата вариации квазиML-оценки около ее предела по вероятности. Измеряет ли она вариацию вблизи соответствующих параметров модели, зависит от того, равняется ли второй элемент нулю. Таким образом, что не удивительно, этот инструментарий работает, только если оценка состоятельна. Насколько это на самом деле так? Разумеется, это не так, если модель пула не учитывает ненаблюдаемые постоянные эффекты. Более того, она будет несостоятельной в большинстве случаев, в которых неверная спецификация касается скрытых случайных эффектов. КвазиML-оценка является состоятельной для случайных эффектов лишь в нескольких особых случаях, таких как линейная модель, модель Пуассона и отрицательная биномиальная модель, рассматриваемые нами в главе 18. Она несостоятельна в пробит- и логит-моделях, в которых зачастую используется этот подход. В конечном счете случаи, когда оценка состоятельна, редки, их в общем-то можно перечислить. Конечный вывод был лаконично изложен в работе Freedman (2006, p. 302): «Сэндвич-алгоритм при строгих условиях регулярности дает дисперсии для ML-оценок, которые являются асимптотически верными, даже когда спецификация — и как следствие функция правдоподобия — ошибочна. Тем не менее игнорирование смещения — это совсем другое дело. Совершенно непонятно, почему прикладного исследователя должна заботить дисперсия оценки ошибочного параметра».

14.9. Модели, оцениваемые методом максимального правдоподобия

В этом разделе мы рассмотрим несколько моделей, в которых используются оценки методом максимального правдоподобия (MLE, ML-оценки). Мы начнем с исследования ML-аналогов большинства оценок для классической и обобщенной моделей регрессии, представленных в главах 4–11. (Вообще говоря, рассмотрение динамических моделей требует несколько больших знаний, чем те, которыми мы пока обладаем. Одним исключением, которое мы рассмотрим, является стандартная модель автокорреляции.) Следует подчеркнуть, что для каждого изучаемого случая мы уже вывели эф-

фективную оценку — обобщенным методом моментов, которая обладает такими же асимптотическими свойствами, как ML-оценка при предположении о нормальном распределении. В более общих случаях мы будем иногда приходить к выводу о том, что GMM-оценка действительно является более предпочтительной по сравнению с ML-оценкой из-за ее устойчивости к ошибочным предположениям о распределении или из-за отсутствия необходимости делать эти предположения в первую очередь. Однако для обобщений классической модели, основанных на обобщенном методе наименьших квадратов, которые мы здесь рассматриваем, это не так. В принципе можно утверждать, что оценка ML здесь является ненужной. Есть случаи, когда ML-оценка будет предпочтительнее по другим причинам, таким как ее инвариантность к преобразованиям в нелинейных моделях или, возможно, ее поведение при малых выборках (хотя обычно это не так). Мы рассматриваем некоторые нелинейные модели, в которых нет линейного аналога метода моментов, поэтому MLE является естественной оценкой. Наконец, в каждом случае мы найдем некоторые полезные свойства оценки, включая разработку алгоритмов, таких как метод Ньютона и EM-метод для моделей скрытых классов.

14.9.1. Нормальная модель линейной регрессии

Модель линейной регрессии выглядит как

$$y_i = \mathbf{x}'_i \boldsymbol{\beta} + \varepsilon_i.$$

Функция правдоподобия для выборки из n независимых, одинаково и нормально распределенных шоков представляет собой

$$L = (2\pi\sigma^2)^{-n/2} e^{-\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}/(2\sigma^2)}. \quad (14-32)$$

Преобразование ε_i в y_i имеет вид $\varepsilon_i = y_i - \mathbf{x}'_i \boldsymbol{\beta}$, поэтому **якобиан** (Jacobian) для каждого наблюдения, $|\partial \varepsilon_i / \partial y_i|$, равен единице¹³. Преобразовывая, мы получаем, что функция правдоподобия для n наблюдений рассматриваемых случайных величин есть

$$L = (2\pi\sigma^2)^{-n/2} e^{(-1/(2\sigma^2))(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}. \quad (14-33)$$

Чтобы максимизировать эту функцию по $\boldsymbol{\beta}$, нам необходимо максимизировать экспоненту или минимизировать привычную нам сумму квадратов. Беря логарифмы, мы приходим к логарифмической функции правдоподобия для классической регрессионной модели

$$\ln L = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 - \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2\sigma^2}. \quad (14-34)$$

¹³См. (В-41) в разделе В.5. Последующий анализ проводится условно по \mathbf{X} . Чтобы избежать беспорядка в системе обозначений, мы не будем явно отображать этот аспект в наших результатах. Как уже отмечалось, мы предполагаем, что процесс порождения данных для \mathbf{X} не зависит от $\boldsymbol{\beta}$ или σ^2 и что данные являются доброкачественными, следуя определению главы 4.

Необходимыми условиями для максимизации этого логарифмического правдоподобия являются

$$\begin{bmatrix} \frac{\partial \ln L}{\partial \beta} \\ \frac{\partial \ln L}{\partial \sigma^2} \end{bmatrix} = \begin{bmatrix} \frac{\mathbf{X}'(\mathbf{y} - \mathbf{X}\beta)}{\sigma^2} \\ \frac{-n}{2\sigma^2} + \frac{(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)}{2\sigma^4} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}. \quad (14-35)$$

Значениями, которые удовлетворяют этим уравнениям, являются

$$\hat{\beta}_{ML} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{b} \quad \text{и} \quad \hat{\sigma}_{ML}^2 = \frac{\mathbf{e}'\mathbf{e}}{n}. \quad (14-36)$$

Оценка коэффициента наклона нам хороша известна, в то время как оценка дисперсии отличается от оценки методом наименьших квадратов знаменателем: n вместо $n - K$ ¹⁴.

Граница Крамера–Рао для дисперсии несмещенной оценки есть минус обратная матрица от математического ожидания

$$\begin{bmatrix} \frac{\partial^2 \ln L}{\partial \beta \partial \beta'} & \frac{\partial^2 \ln L}{\partial \beta \partial \sigma^2} \\ \frac{\partial^2 \ln L}{\partial \sigma^2 \partial \beta'} & \frac{\partial^2 \ln L}{\partial (\sigma^2)^2} \end{bmatrix} = \begin{bmatrix} -\frac{\mathbf{X}'\mathbf{X}}{\sigma^2} & -\frac{\mathbf{X}'\boldsymbol{\epsilon}}{\sigma^4} \\ -\frac{\boldsymbol{\epsilon}'\mathbf{X}}{\sigma^4} & \frac{n}{2\sigma^4} - \frac{\boldsymbol{\epsilon}'\boldsymbol{\epsilon}}{\sigma^6} \end{bmatrix}. \quad (14-37)$$

При взятии математического ожидания недиагональные элементы обращаются в ноль, оставляя

$$[\mathbf{I}(\beta, \sigma^2)]^{-1} = \begin{bmatrix} \sigma^2(\mathbf{X}'\mathbf{X})^{-1} & \mathbf{0} \\ \mathbf{0}' & 2\sigma^4/n \end{bmatrix}. \quad (14-38)$$

Оценка коэффициентов методом наименьших квадратов есть оценка методом максимального правдоподобия в этой модели. По этой причине она перенимает все желаемые *асимптотические* свойства оценок методом максимального правдоподобия.

Ранее мы показали, что $s^2 = \mathbf{e}'\mathbf{e}/(n - K)$ является несмещенной оценкой σ^2 . Следовательно, оценка методом максимального правдоподобия является смещенной к нулю:

$$E[\hat{\sigma}_{ML}^2] = \frac{n - K}{n}\sigma^2 = \left(1 - \frac{K}{n}\right)\sigma^2 < \sigma^2. \quad (14-39)$$

Несмотря на ее смещение при малых выборках, оценка максимального правдоподобия параметра σ^2 обладает теми же самыми желаемыми асимптотическими свойствами. (14-39) показывает, что s^2 и $\hat{\sigma}^2$ отличаются только множителем $-K/n$, который стремится к нулю в больших выборках. Поучительно формализовать асимптотическую эквивалентность двух этих оценок. Из (14-38) мы знаем, что

$$\sqrt{n}(\hat{\sigma}_{ML}^2 - \sigma^2) \xrightarrow{d} N[0, 2\sigma^4].$$

¹⁴Как правило, оценки методом максимального правдоподобия не используют корректировку на число степеней свободы.

Отсюда следует, что

$$z_n = \left(1 - \frac{K}{n}\right) \sqrt{n}(\hat{\sigma}_{ML}^2 - \sigma^2) + \frac{K}{\sqrt{n}}\sigma^2 \xrightarrow{d} \left(1 - \frac{K}{n}\right) N[0, 2\sigma^4] + \frac{K}{\sqrt{n}}\sigma^2.$$

Но K/\sqrt{n} и K/n обращаются в ноль при $n \rightarrow \infty$, поэтому предельное распределение z_n также равно $N[0, 2\sigma^4]$. Поскольку $z_n = \sqrt{n}(s^2 - \sigma^2)$, мы показали, что асимптотическое распределение оценки s^2 является в точности таким же, как и распределение оценки методом максимального правдоподобия.

Стандартная тестовая статистика для проверки нескольких линейных ограничений в линейной модели $\mathbf{R}\beta - \mathbf{q} = \mathbf{0}$ представляет собой F -отношение

$$F[J, n - K] = \frac{(\mathbf{e}'_*\mathbf{e}_* - \mathbf{e}'\mathbf{e})/J}{\mathbf{e}'\mathbf{e}/(n - K)} = \frac{(\mathbf{R}\mathbf{b} - \mathbf{q})'[\mathbf{R}\mathbf{s}^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b} - \mathbf{q})}{J}.$$

При нормально распределенных шоках F -критерий применим в выборках любого объема. Но остается проблема с нелинейными ограничениями вида $\mathbf{c}(\beta) = \mathbf{0}$, поскольку аналог F -статистики, который мы здесь будем рассматривать, является справедливым только асимптотически, даже если шоки распределены нормально. В этом разделе мы повторно рассмотрим статистику Вальда и проверим две связанные с ней статистики: статистику отношения правдоподобия и статистику множителей Лагранжа. Обе эти статистики основаны на функции правдоподобия и, подобно статистике Вальда, вообще говоря, справедливы лишь асимптотически.

В данном случае линейные ограничения ничем не проще нелинейных, и поэтому мы будем рассматривать общие гипотезы вида

$$\begin{aligned} H_0: \mathbf{c}(\beta) &= \mathbf{0}, \\ H_1: \mathbf{c}(\beta) &\neq \mathbf{0}. \end{aligned}$$

Статистика Вальда (Wald statistic) для проверки этой гипотезы и ее предельное распределение при гипотезе H_0 задаются как

$$W = \mathbf{c}(\mathbf{b})' \{ \mathbf{C}(\mathbf{b}) [\hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}] \mathbf{C}(\mathbf{b})' \}^{-1} \mathbf{c}(\mathbf{b}) \xrightarrow{d} \chi^2[J], \quad (14-40)$$

где

$$\mathbf{C}(\mathbf{b}) = [\partial \mathbf{c}(\mathbf{b}) / \partial \mathbf{b}']. \quad (14-41)$$

Критерий отношения правдоподобия (LR) получается путем сравнения значений логарифмической функции правдоподобия с наложенными ограничениями и без них. На данном этапе мы не рассматриваем вопрос вычисления оценки при ограничениях \mathbf{b}_* (кроме линейной модели, для которой это было рассмотрено ранее). Тестовая статистика и ее предельное распределение при гипотезе H_0 имеют вид

$$LR = -2[\ln L_* - \ln L] \xrightarrow{d} \chi^2[J]. \quad (14-42)$$

Логарифмическое правдоподобие для регрессионной модели задается (14-34). Из условий первого порядка следует, что независимо от метода вычисления коэффициентов наклона оценка σ^2 без ограничений на β будет равна

$\hat{\sigma}^2 = (\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b})/n$, а оценка при ограничениях аналогично будет равна $\hat{\sigma}_*^2 = (\mathbf{y} - \mathbf{X}\mathbf{b}_*)'(\mathbf{y} - \mathbf{X}\mathbf{b}_*)/n = \mathbf{e}'_*\mathbf{e}_*/n$. **Концентрированное логарифмическое правдоподобие** (concentrated log-likelihood)¹⁵ будет равно

$$\ln L_c = -\frac{n}{2}[1 + \ln 2\pi + \ln(\mathbf{e}'\mathbf{e}/n)]$$

и аналогично выписывается для случая с ограничениями. Если подставить эти выражения в выражение для LR-статистики, то мы получим

$$LR = n \ln[\mathbf{e}'_*\mathbf{e}_*/\mathbf{e}'\mathbf{e}] = n(\ln \hat{\sigma}_*^2 - \ln \hat{\sigma}^2) = n \ln(\hat{\sigma}_*^2/\hat{\sigma}^2). \quad (14-43)$$

Критерий **множителей Лагранжа (LM)** основан на градиенте логарифмической функции правдоподобия. Суть критерия заключается в том, что если гипотеза является верной, то производные логарифмической функции правдоподобия в оценке, вычисленной при ограничениях, должны быть близки к нулю. Есть два способа выполнить LM-тест. Логарифмическая функция правдоподобия может быть максимизирована при ограничениях с помощью

$$\ln L_{LM} = -\frac{n}{2} \left[\ln 2\pi + \ln \sigma^2 + \frac{[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})]/n}{\sigma^2} \right] + \boldsymbol{\lambda}'\mathbf{c}(\boldsymbol{\beta}).$$

Условиями первого порядка для решения являются

$$\begin{bmatrix} \frac{\partial \ln L_{LM}}{\partial \boldsymbol{\beta}} \\ \frac{\partial \ln L_{LM}}{\partial \sigma^2} \\ \frac{\partial \ln L_{LM}}{\partial \boldsymbol{\lambda}} \end{bmatrix} = \begin{bmatrix} \frac{\mathbf{X}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{\sigma^2} + \mathbf{C}(\boldsymbol{\beta})'\boldsymbol{\lambda} \\ \frac{-n}{2\sigma^2} + \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2\sigma^4} \\ \mathbf{c}(\boldsymbol{\beta}) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \quad (14-44)$$

Решения этих уравнений дают оценку методом наименьших квадратов при ограничениях \mathbf{b}_* , обычную оценку дисперсии, имеющую вид $\mathbf{e}'_*\mathbf{e}_*/n$, и множители Лагранжа. На данном этапе существует два пути вычисления тестовой статистики. В постановке классической модели линейной регрессии, когда мы вычисляем множители Лагранжа, удобным способом является проверка гипотезы о том, что эти множители равны нулю. В этой модели решением для $\boldsymbol{\lambda}_*$ является $\boldsymbol{\lambda}_* = [\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}']^{-1}(\mathbf{R}\mathbf{b} - \mathbf{q})$. Это уравнение представляет собой линейную функцию от оценки методом наименьших квадратов. Если мы применяем критерий *Вальда* для проверки гипотезы о том, что $\boldsymbol{\lambda}_*$ равна $\mathbf{0}$, то искомой статистикой будет

$$LM = \boldsymbol{\lambda}'_* \{\text{Est. Var}[\boldsymbol{\lambda}_*]\}^{-1} \boldsymbol{\lambda}_* = (\mathbf{R}\mathbf{b} - \mathbf{q})' [\mathbf{R} s_*^2 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{R}']^{-1} (\mathbf{R}\mathbf{b} - \mathbf{q}). \quad (14-45)$$

Оценка дисперсии шоков s_*^2 , основанная на коэффициентах, вычисленных при ограничениях, есть $\mathbf{e}'_*\mathbf{e}_*/n$.

Другой способ расчета LM-статистики часто дает интересные результаты. В большинстве ситуаций мы максимизируем логарифмическую функцию правдоподобия без непосредственного вычисления вектора множите-

¹⁵См. раздел E4.3.

лей Лагранжа. (Ограничения обычно накладываются некоторым другим способом.) Альтернативный способ вычисления статистики основывается на (общем) результате, согласно которому при проверяемой гипотезе

$$E[\partial \ln L / \partial \beta] = E[(1/\sigma^2)X'\epsilon] = 0$$

и¹⁶

$$\text{Asy. Var}[\partial \ln L / \partial \beta] = -E[\partial^2 \ln L / \partial \beta \partial \beta']^{-1} = \sigma^2(X'X)^{-1}. \quad (14-46)$$

Мы можем проверить гипотезу о том, что в оценке, вычисленной при ограничениях, производные равны нулю. Искомая статистика будет иметь вид

$$\text{LM} = \frac{e_*'X(X'X)^{-1}X'e_*}{e_*'e_*/n} = nR_*^2. \quad (14-47)$$

В этом виде LM-статистика представляет собой умноженный на n коэффициент детерминации в регрессии остатков $e_{i*} = (y_i - x_i'b_*)$ на полный набор регрессоров.

После некоторых преобразований мы можем получить, что $W = [n/(n-K)]JF$ и LR и LM приблизительно равны этой функции от F ¹⁷. Все три статистики сходятся к JF при увеличении n . Линейная модель представляет собой особый случай, в котором LR-статистика основывается только на оценке без ограничений и не требует вычисления оценки при ограничениях метода наименьших квадратов, хотя вычисление F и включает большинство расчетов, необходимых для b_* . Поскольку логарифмическая функция является вогнутой и $W/n \geq \ln(1 + W/n)$, в работе Годфри (Godfrey (1988)) также показано, что $W \geq LR \geq LM$, поэтому для линейной модели существует четкая упорядоченность этих трех статистик.

Вычисления на данных показывают, что асимптотические результаты для этих статистик малоприменимы в малых выборках или выборках умеренного объема. [См., например, книгу Дэвидсона и Манкиннона (Davidson, MacKinnon (2004, p. 424–428)).] Истинные распределения всех трех статистик зависят от данных и неизвестных параметров и, как можно показать алгебраически, сходятся к F -распределению сверху. Таким образом, критические значения распределения хи-квадрат будут являться слишком малыми, т. е. использование предельного распределения хи-квадрат в малых выборках будет преувеличивать значимость полученных эмпирических результатов. Таким образом, в приложениях предпочтительнее использовать более традиционную F -статистику (или t -статистику в случае одного ограничения), если только мы не имеем дело с избытком данных.

14.9.2. Обобщенная регрессионная модель

Для обобщенной регрессионной модели, рассмотренной в разделе 9.1,

$$\begin{aligned} y_i &= x_i'\beta + \epsilon_i, i = 1, \dots, n, \\ E[\epsilon | X] &= 0, \\ E[\epsilon\epsilon' | X] &= \sigma^2\Omega, \end{aligned}$$

¹⁶Здесь мы используем тот факт, что гесссиан есть блочно-диагональная матрица.

¹⁷Godfrey (1988, p. 49–51).

как и прежде, сначала мы предположим, что Ω — это известная постоянная матрица. Если шоки имеют многомерное нормальное распределение, то логарифмическая функция правдоподобия для выборки имеет вид

$$\ln L = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\beta)' \Omega^{-1} (\mathbf{y} - \mathbf{X}\beta) - \frac{1}{2} \ln |\Omega|. \quad (14-48)$$

Поскольку Ω — это матрица известных постоянных, оценка максимального правдоподобия параметра β представляет собой вектор, который минимизирует **обобщенную сумму квадратов** (generalized sum of squares):

$$S_*(\beta) = (\mathbf{y} - \mathbf{X}\beta)' \Omega^{-1} (\mathbf{y} - \mathbf{X}\beta)$$

(откуда и название метода — *обобщенный метод наименьших квадратов*). Необходимые условия для максимизации L имеют вид

$$\begin{aligned} \frac{\partial \ln L}{\partial \beta} &= \frac{1}{\sigma^2} \mathbf{X}' \Omega^{-1} (\mathbf{y} - \mathbf{X}\beta) = \frac{1}{\sigma^2} \mathbf{X}'_* (\mathbf{y}_* - \mathbf{X}_* \beta) = 0, \\ \frac{\partial \ln L}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} (\mathbf{y} - \mathbf{X}\beta)' \Omega^{-1} (\mathbf{y} - \mathbf{X}\beta) \\ &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} (\mathbf{y}_* - \mathbf{X}_* \beta)' (\mathbf{y}_* - \mathbf{X}_* \beta) = 0. \end{aligned} \quad (14-49)$$

Решения представляют собой OLS-оценки, использующие преобразованные данные:

$$\hat{\beta}_{ML} = (\mathbf{X}'_* \mathbf{X}_*)^{-1} \mathbf{X}'_* \mathbf{y}_* = (\mathbf{X}' \Omega^{-1} \mathbf{X})^{-1} \mathbf{X}' \Omega^{-1} \mathbf{y}, \quad (14-50)$$

$$\begin{aligned} \hat{\sigma}_{ML}^2 &= \frac{1}{n} (\mathbf{y}_* - \mathbf{X}_* \beta)' (\mathbf{y}_* - \mathbf{X}_* \beta) \\ &= \frac{1}{n} (\mathbf{y} - \mathbf{X}\beta)' \Omega^{-1} (\mathbf{y} - \mathbf{X}\beta), \end{aligned} \quad (14-51)$$

откуда следует, что при нормально распределенных шоках оценки обобщенного метода наименьших квадратов также являются оценками максимального правдоподобия. Как в классической регрессионной модели, оценка максимального правдоподобия σ^2 является смещенной. Несмещенная оценка задается формулой (9-14). Мы приходим к ожидаемому выводу, что, когда Ω известна, оценка максимального правдоподобия совпадает с оценкой обобщенного метода наименьших квадратов.

Когда Ω неизвестна и должна быть оценена, тогда необходимо максимизировать логарифмическое правдоподобие в (14-48) относительно всего набора параметров $[\beta, \sigma^2, \Omega]$ одновременно. Поскольку лишь матрица Ω , рассматриваемая без ограничений, включает $n(n+1)/2 - 1$ параметров, ясно, что для оценивания требуется наложить на структуру Ω некоторое ограничение. В нескольких следующих разделах мы рассмотрим несколько ситуаций, в которых $\Omega = \Omega(\theta)$ для некоторого небольшого вектора параметров. Здесь мы лишь упомянем несколько общих результатов.

1. Для заданного значения θ оценка β называется оценкой доступного GLS, а оценка σ^2 задается формулой (14-51).

2. Уравнения правдоподобия для θ обычно будут представлять собой сложные функции от β и σ^2 , поэтому необходимо проводить совместное оценивание. Тем не менее во многих случаях для заданных значений β и σ^2 оценку θ можно получить непосредственным образом. Так, например, в модели в (9-15) итеративная оценка θ при заданных β , σ^2 и начальном значении θ вычисляется как сумма прежнего значения и оценки коэффициента в регрессии $(e_i^2/\hat{\sigma}_i^2 - 1)$ на z_i .

На втором шаге предлагается выполнение ряда итераций, которое применимо в большом числе ситуаций. Начиная, скажем, с OLS-оценки, выполняются итерации вперед-назад между пунктами 1 и 2 до тех пор, пока процесс не сойдется и не даст совместную оценку максимального правдоподобия.

Эта ситуация была исследована в работе Oberhofer, Kmenta (1974), в которой было показано, что при некоторых довольно слабых требованиях, наиболее важным из которых является условие, что θ не влияет на σ^2 и на любой из параметров в векторе β , этот метод приведет к оценке максимального правдоподобия. Другой вывод, который можно достаточно просто доказать (мы оставляем это читателю в качестве упражнения), заключается в том, что при рассматриваемых предположениях асимптотическая ковариационная матрица оценки совпадает с GLS-оценкой. Она одинакова независимо от того, известна ли Ω или ее требуется оценивать, что означает, что если θ и β не имеют общих параметров, то *точное знание матрицы Ω не дает никакой выгоды в асимптотической эффективности при оценивании β по сравнению со случаем оценивания β , имея состоятельную оценку матрицы Ω .*

Теперь мы рассмотрим две базовые ситуации, включающие лишь одно уравнение, — гетероскедастичность и автокорреляцию.

14.9.2а. Мультипликативная гетероскедастичность

Модель мультипликативной гетероскедастичности, рассмотренная в работе Харви (Harvey (1976)), является очень гибкой и включает большинство интересных постановок задач в качестве частных случаев. Общая постановка задачи выглядит следующим образом:

$$\sigma_i^2 = \sigma^2 \exp(z_i' \alpha). \quad (14-52)$$

Гетероскедастичность вида

$$\sigma_i^2 = \sigma^2 \prod_{m=1}^M z_{im}^{\alpha_m} \quad (14-53)$$

получается, если вместо z_i взять логарифмы переменных. Групповая модель гетероскедастичности, представленная в подразделе 9.7.2, получается, если рассматривать вместо z_i набор групповых дамми-переменных (одну из них требуется пропустить). В этом случае σ^2 является дисперсией шоков для базовой группы, тогда как для других групп $\sigma_g^2 = \sigma^2 \exp(\alpha_g)$.

Мы начинаем с полезного упрощения. Пусть z_i включает постоянный член, так что $z_i' = [1, q_i']$, где q_i — исходный набор переменных, и пусть $\gamma' = [\ln \sigma^2, \alpha']$. Тогда модель выглядит просто как $\sigma_i^2 = \exp(z_i' \gamma)$. Как только

оценен полный вектор параметров, $\exp(\gamma_1)$ дает оценку σ^2 . (Эта оценка использует результат инвариантности оценок методом максимального правдоподобия. См. подраздел 14.4.5d.)

Логарифмическое правдоподобие представляет собой

$$\begin{aligned}\ln L &= -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^n \ln \sigma_i^2 - \frac{1}{2} \sum_{i=1}^n \frac{\varepsilon_i^2}{\sigma_i^2} \\ &= -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^n \mathbf{z}'_i \boldsymbol{\gamma} - \frac{1}{2} \sum_{i=1}^n \frac{\varepsilon_i^2}{\exp(\mathbf{z}'_i \boldsymbol{\gamma})}.\end{aligned}\quad (14-54)$$

Уравнения правдоподобия записываются как

$$\begin{aligned}\frac{\partial \ln L}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^n \mathbf{x}_i \frac{\varepsilon_i}{\exp(\mathbf{z}'_i \boldsymbol{\gamma})} = \mathbf{X}' \boldsymbol{\Omega}^{-1} \boldsymbol{\varepsilon} = \mathbf{0}, \\ \frac{\partial \ln L}{\partial \boldsymbol{\gamma}} &= \frac{1}{2} \sum_{i=1}^n \mathbf{z}_i \left(\frac{\varepsilon_i^2}{\exp(\mathbf{z}'_i \boldsymbol{\gamma})} - 1 \right) = \mathbf{0}.\end{aligned}\quad (14-55)$$

Для этой модели метод скоры оказывается особенно удобным для максимизации логарифмической функции правдоподобия. Выражения в гессiane имеют вид

$$\frac{\partial^2 \ln L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = - \sum_{i=1}^n \frac{1}{\exp(\mathbf{z}'_i \boldsymbol{\gamma})} \mathbf{x}_i \mathbf{x}'_i = -\mathbf{X}' \boldsymbol{\Omega}^{-1} \mathbf{X}, \quad (14-56)$$

$$\frac{\partial^2 \ln L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}'} = - \sum_{i=1}^n \frac{\varepsilon_i}{\exp(\mathbf{z}'_i \boldsymbol{\gamma})} \mathbf{x}_i \mathbf{z}'_i, \quad (14-57)$$

$$\frac{\partial^2 \ln L}{\partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'} = - \frac{1}{2} \sum_{i=1}^n \frac{\varepsilon_i^2}{\exp(\mathbf{z}'_i \boldsymbol{\gamma})} \mathbf{z}_i \mathbf{z}'_i. \quad (14-58)$$

Математическое ожидание $\partial^2 \ln L / \partial \boldsymbol{\beta} \partial \boldsymbol{\gamma}'$ равно $\mathbf{0}$, поскольку $E[\varepsilon_i | \mathbf{x}_i, \mathbf{z}_i] = 0$. Математическое ожидание дроби в выражении для $\partial^2 \ln L / \partial \boldsymbol{\gamma} \partial \boldsymbol{\gamma}'$ равно $E[\varepsilon_i^2 / \sigma_i^2 | \mathbf{x}_i, \mathbf{z}_i] = 1$. Пусть $\boldsymbol{\delta} = [\boldsymbol{\beta}, \boldsymbol{\gamma}]$. Тогда

$$-E \left(\frac{\partial^2 \ln L}{\partial \boldsymbol{\delta} \partial \boldsymbol{\delta}'} \right) = \begin{bmatrix} \mathbf{X}' \boldsymbol{\Omega}^{-1} \mathbf{X} & \mathbf{0} \\ \mathbf{0}' & \frac{1}{2} \mathbf{Z}' \mathbf{Z} \end{bmatrix} = -\bar{\mathbf{H}}. \quad (14-59)$$

Метод скоры (method of scoring) представляет собой алгоритм для нахождения итеративного решения уравнений правдоподобия. Итерация выглядит следующим образом:

$$\boldsymbol{\delta}_{t+1} = \boldsymbol{\delta}_t - \bar{\mathbf{H}}^{-1} \mathbf{g}_t,$$

где $\boldsymbol{\delta}_t$ (т. е. $\boldsymbol{\beta}_t$, $\boldsymbol{\gamma}_t$ и $\boldsymbol{\Omega}_t$) — оценка при итерации t , \mathbf{g}_t — вектор первых производных $[\partial \ln L / \partial \boldsymbol{\beta}'_t, \partial \ln L / \partial \boldsymbol{\gamma}'_t]'$, состоящий из двух частей, а матрица $\bar{\mathbf{H}}$ разделена аналогичным образом. В **методе Ньютона** (Newton's method) используются фактически вторые производные, вычисляемые по формулам (14-56)–(14-58), а не их математические ожидания, вычисляемые по формуле (14-59). В методе скоры используется тот удобный факт, что математическое

ожидание недиагонального блока матрицы (смешанной производной) равно нулю в (14-57).] Поскольку \bar{H} является блочно-диагональной матрицей, итерация может быть записана как отдельные уравнения:

$$\begin{aligned}\beta_{t+1} &= \beta_t + (\mathbf{X}'\Omega_t^{-1}\mathbf{X})^{-1}(\mathbf{X}'\Omega_t^{-1}\varepsilon_t) \\ &= \beta_t + (\mathbf{X}'\Omega_t^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega_t^{-1}(y - \mathbf{X}\beta_t) \\ &= (\mathbf{X}'\Omega_t^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega_t^{-1}y \text{ (несомненно)}.\end{aligned}\quad (14-60)$$

Следовательно, обновленный вектор коэффициентов β_{t+1} рассчитывается при помощи FGLS с использованием полученной ранее оценки γ для вычисления Ω . Мы используем тот же самый подход для γ :

$$\gamma_{t+1} = \gamma_t + [2(\mathbf{Z}'\mathbf{Z})^{-1}] \left[\frac{1}{2} \sum_{i=1}^n z_i \left(\frac{\varepsilon_i^2}{\exp(z_i'\gamma)} - 1 \right) \right]. \quad (14-61)$$

2 и $\frac{1}{2}$ сокращаются. Обновленное значение γ вычисляется путем добавления вектора коэффициентов в регрессию метода наименьших квадратов $[\varepsilon_i^2 / \exp(z_i'\gamma) - 1]$ на z_i к старому значению. Обратим внимание, что коррекция представляет собой $2(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\partial \ln L / \partial \gamma)$, поэтому процесс сойдется, когда производная будет равна нулю.

Нам осталось определить начальное значение для итерации. Поскольку любая состоятельная оценка подойдет, простейшим методом является использование OLS для β и коэффициентов (наклона) в регрессии логарифмов квадратов остатков метода наименьших квадратов на z_i для γ . В работе Харви (Harvey (1976)) показывается, что этот метод дает несостоятельную оценку $\gamma_1 = \ln \sigma^2$, но эта несостоятельность может быть скорректирована просто путем добавления 1,2704 к полученному значению¹⁸. Таким образом, итерация имеет следующий простой вид:

- 1) оценить дисперсию шоков σ_i^2 с помощью $\exp(z_i'\gamma)$;
- 2) вычислить β_{t+1} при помощи FGLS¹⁹;
- 3) обновить γ_t с использованием регрессии, описанной в предыдущем абзаце;
- 4) вычислить $d_{t+1} = [\beta_{t+1}, \gamma_{t+1}] - [\beta_t, \gamma_t]$. Если d_{t+1} велико, то вернуться к шагу 1.

Если d_{t+1} на шаге 4 достаточно мало, то получится выход из итерации. Асимптотическая ковариационная матрица — это просто $-\mathbf{H}^{-1}$, который представляет собой блочно-диагональную матрицу с блоками

$$\begin{aligned}\text{Asy. Var}[\hat{\beta}_{\text{ML}}] &= (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}, \\ \text{Asy. Var}[\hat{\gamma}_{\text{ML}}] &= 2(\mathbf{Z}'\mathbf{Z})^{-1}.\end{aligned}$$

При желании можно вычислить $\hat{\sigma}^2 = \exp(\hat{\gamma}_1)$. Асимптотическая дисперсия в этом случае составит $[\exp(\gamma_1)]^2(\text{Asy. Var}[\hat{\gamma}_{1,\text{ML}}])$.

¹⁸В работе также представлена коррекция асимптотической ковариационной матрицы для этой оценки γ на первом шаге.

¹⁹Если остановиться на данном этапе, то полученная двухшаговая оценка была бы полностью эффективной, если бы начальное значение для γ было состоятельным, однако не была бы оценкой максимального правдоподобия.

Проверка нулевой гипотезы гомоскедастичности в этой модели,

$$H_0: \alpha = 0$$

в (14-52), особенно проста. Для этого требуется применить критерий Вальда к проверке гипотезы, что последние M элементов γ равны нулю. Таким образом, статистика будет иметь вид

$$\lambda_{WALD} = \hat{\alpha}' \left\{ \begin{bmatrix} \mathbf{0} & \mathbf{I} \end{bmatrix} [2(\mathbf{Z}'\mathbf{Z})]^{-1} \begin{bmatrix} \mathbf{0} \\ \mathbf{I} \end{bmatrix} \right\}^{-1} \hat{\alpha}.$$

Поскольку первый столбец в \mathbf{Z} является постоянной, это сокращается до

$$\lambda_{WALD} = \frac{1}{2} \hat{\alpha}' (\mathbf{Z}'_1 \mathbf{M}^0 \mathbf{Z}_1) \hat{\alpha},$$

где матрица \mathbf{Z}_1 есть M последних столбцов матрицы \mathbf{Z} , не включая столбец единиц, а \mathbf{M}^0 представляет собой отклонения от средних. Статистика отношения правдоподобий вычисляется с помощью (14-54). И при нулевой гипотезе (гомоскедастичность, используя OLS), и при альтернативной (гетероскедастичность, используя MLE) третий член в $\ln L$ сокращается до $-n/2$. Следовательно, статистика имеет следующий простой вид:

$$\lambda_{LR} = 2(\ln L_1 - \ln L_0) = n \ln s^2 - \sum_{i=1}^n \ln \hat{\sigma}_i^2,$$

где $s^2 = \mathbf{e}'\mathbf{e}/n$, основанная на OLS-остатках. Чтобы вычислить LM-статистику, мы будем использовать математическое ожидание гессииана в (14-59). При нулевой гипотезе та часть вектора производных в (14-55), которая соответствует β , представляет собой $(1/s^2)\mathbf{X}'\mathbf{e} = 0$. Вследствие этого, используя (14-55), получаем, что LM-статистика — это

$$\lambda_{LM} = \left[\frac{1}{2} \sum_{i=1}^n \left(\frac{e_i^2}{s^2} - 1 \right) \begin{pmatrix} 1 \\ \mathbf{z}_{i1} \end{pmatrix} \right]' \left[\frac{1}{2} (\mathbf{Z}'\mathbf{Z}) \right]^{-1} \left[\frac{1}{2} \sum_{i=1}^n \left(\frac{e_i^2}{s^2} - 1 \right) \begin{pmatrix} 1 \\ \mathbf{z}_{i1} \end{pmatrix} \right].$$

Первый элемент в векторе производных равен нулю, потому что $\sum_i e_i^2 = ns^2$. По этой причине выражение сокращается до

$$\lambda_{LM} = \frac{1}{2} \left[\sum_{i=1}^n \left(\frac{e_i^2}{s^2} - 1 \right) \mathbf{z}_{i1} \right]' (\mathbf{Z}'_1 \mathbf{M}^0 \mathbf{Z}_1)^{-1} \left[\sum_{i=1}^n \left(\frac{e_i^2}{s^2} - 1 \right) \mathbf{z}_{i1} \right].$$

Это половина объясненной суммы квадратов в линейной регрессии переменной $h_i = (e_i^2/s^2 - 1)$ на \mathbf{Z} , что является LM-статистикой Бройша-Пэгана/Годфри из подраздела 9.5.2.

Пример 14.6. Мультипликативная гетероскедастичность

В примере 6.4 мы подбирали функцию издержек для отрасли авиаперевозок США вида

$$\ln C_{it} = \beta_1 + \beta_2 \ln Q_{it} + \beta_3 [\ln Q_{it}]^2 + \beta_4 \ln P_{fuel,i,t} + \beta_5 Loadfactor_{i,t} + \varepsilon_{i,t},$$

где $C_{i,t}$ — полная себестоимость, $Q_{i,t}$ — производительность, $Loadfactor$ — коэффициент загрузки, а $P_{fuel,i,t}$ — цена на топливо и где 90 наблюдений представляют собой данные для шести фирм, за которыми велось наблюдение в течение 15 лет. (Модель также включала фиктивные переменные для фирмы и года, которые мы здесь опустим для простоты.) В примере 9.4 мы подбирали скорректированную модель, в которой коэффициент загрузки появляется в дисперсии $\varepsilon_{i,t}$, а не в функции регрессии. Модель представляет собой следующее:

$$\begin{aligned}\sigma_{i,t}^2 &= \sigma^2 \exp(\alpha Loadfactor_{i,t}) \\ &= \exp(\gamma_1 + \gamma_2 Loadfactor_{i,t}).\end{aligned}$$

Оценки были получены путем итеративной процедуры вычисления взвешенным методом наименьших квадратов, использующим веса $W_{i,t} = \exp(-c_1 - c_2 Loadfactor_{i,t})$. Оценки γ_1 и γ_2 получались на каждой итерации путем регрессирования логарифмов квадратов остатков на постоянную и $Loadfactor_{i,t}$. В конце примера [это следует из (14-61)] было отмечено, что эти веса были бы ошибочными, если бы мы хотели вычислить MLE. В табл. 14.3 воспроизведены результаты из примера 9.4 и добавлены MLE, полученные с использованием методов работы Harvey (1976). ML-оценка параметра γ_2 существенно отличается от предшествующего результата. Статистика Вальда для проверки ограничения гомоскедастичности ($\alpha = 0$) составляет $(9,78076/2,839)^2 = 11,869$, что больше, чем 3,84, поэтому нулевая гипотеза будет отвергнута. Статистика отношения правдоподобия составляет $-2(54,2747 - 57,3122) = 6,075$, что приводит к тому же выводу. Однако LM-статистика равна 2,96, что вступает в противоречие с полученными результатами. Это результат выборки конечного объема, который не является редкостью.

14.9.2b. Автокорреляция

В разных местах в предыдущих разделах мы рассматривали модели, в которых имеет место корреляция между наблюдениями, включая случай пространственной автокорреляции в разделе 11.7, автокоррелированные шоки в моделях панельных данных [раздел 11.6.3 и (11-28)] и модели внешне не связанных регрессий (SUR) в разделе 10.3. Рассматриваемая здесь модель авторегрессии первого порядка будет детально описана в главе 20. Мы кратко рассмотрим ее здесь, с тем чтобы прояснить некоторые полезные моменты, касающиеся оценки методом максимального правдоподобия.

Модель линейной регрессии с авторегрессионными [AR(1)] шоками первого порядка представляет собой

$$\begin{aligned}y_t &= \mathbf{x}'_t \boldsymbol{\beta} + \varepsilon_t, t = 1, \dots, T, \\ \varepsilon_t &= \rho \varepsilon_{t-1} + u_t, |\rho| < 1, \\ E[u_t | \mathbf{X}] &= 0, \\ E[u_t u_s | \mathbf{X}] &= \sigma_u^2, \quad \text{если } t = s, \quad \text{и } 0 \text{ в противном случае.}\end{aligned}$$

Таблица 14.3. Модель мультипликативной гетероскедастичности

| | Посто- янная | Ln Q | Ln ² Q | Ln P _f | R ² | Сумма квадратов |
|---------------------------------|----------------------|----------|-------------------|-------------------|------------------------|-----------------------|
| OLS | 9,1382 | 0,92615 | 0,029145 | 0,41006 | | |
| ln L = 54,2747 | 0,24507 ^a | 0,032306 | 0,012304 | 0,018807 | 0,9861674 ^c | 1,577479 ^d |
| | 0,22595 ^b | 0,030128 | 0,011346 | 0,017524 | | |
| Двухшаговая | 9,2463 | 0,92136 | 0,024450 | 0,40352 | | |
| | 0,21896 | 0,033028 | 0,011412 | 0,016974 | 0,986119 | 1,612938 |
| Итериро- ванная ^e | 9,2774 | 0,91609 | 0,021643 | 0,40174 | | |
| | 0,20977 | 0,032993 | 0,011017 | 0,016332 | 0,986071 | 1,645693 |
| MLE ^f | 9,2611 | 0,91931 | 0,023281 | 0,40266 | | |
| ln L = 57,3122 | 0,2099 | 0,032295 | 0,010987 | 0,016304 | 0,986100 | 1,626301 |

^a Стандартные ошибки МНК.

^b Устойчивые стандартные ошибки Уайта.

^c Квадрат коэффициента корреляции между фактическими и подогнанными значениями.

^d Сумма квадратов остатков.

^e Значения c_2 по итерациям: 8,254344, 11,622473, 11,705029, 11,710618, 11,711012, 11,711040, 11,711042.

^f Оценка γ_2 равна 9,78076 (2,839).

Доступное GLS-оценивание параметров этой модели детально рассмотрено в главе 20. Здесь мы добавим предположение о нормальном распределении $u_t \sim N[0, \sigma_u^2]$ и построим оценку методом максимального правдоподобия.

Поскольку каждое наблюдение y_t коррелирует с любым другим наблюдением, в принципе для того, чтобы выразить функцию правдоподобия, требуется выразить плотность совместного распределения одной случайной величины размерностью T . Преобразование Прайса–Уинстена (Prais, Winsten (1954)) — см. формулу (20-28) — дает удобный способ переформулировать эту плотность. Мы можем записать

$$f(y_1, y_2, \dots, y_T) = f(y_1)f(y_2 | y_1), f(y_3 | y_2), \dots, f(y_T | y_{T-1}).$$

Поскольку

$$\begin{aligned} \sqrt{1 - \rho^2} y_1 &= \sqrt{1 - \rho^2} \mathbf{x}'_1 \boldsymbol{\beta} + u_1, \\ y_t | y_{t-1} &= \rho y_{t-1} + (\mathbf{x}_t - \rho \mathbf{x}_{t-1})' \boldsymbol{\beta} + u_t \end{aligned} \quad (14-62)$$

и наблюдения u_t независимо нормально распределены, мы можем использовать эти результаты, чтобы выразить логарифмическую функцию правдоподобия

$$\begin{aligned} \ln L &= \left[-\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln \sigma_u^2 + \frac{1}{2} \ln(1 - \rho^2) - \frac{(1 - \rho^2)(y_1 - \mathbf{x}'_1 \boldsymbol{\beta})^2}{2\sigma_u^2} \right] + \\ &+ \sum_{t=2}^T \left[-\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln \sigma_u^2 - \frac{[(y_t - \rho y_{t-1}) - (\mathbf{x}_t - \rho \mathbf{x}_{t-1})' \boldsymbol{\beta}]^2}{2\sigma_u^2} \right]. \end{aligned} \quad (14-63)$$

Как обычно, ML-оценка параметра β представляет собой GLS-оценку, основанную на ML-оценках параметров σ_u^2 и ρ , а MLE для σ_u^2 будет равняться $u'u/T$ при заданных β и ρ . Сложность состоит в том, как вычислить ρ . Как мы отметим далее в главе 20, существует удивительно большое число способов для состоятельного оценивания ρ в AR(1)-модели. Соблазнительной стратегией кажется выбрать из этих способов наиболее удобный, а затем проводить итерации вперед-назад между β и (σ_u^2, ρ) , чтобы получить ML-оценку. Тем не менее эта методика не будет (в общем случае) приводить к MLE, если только промежуточные оценки параметров дисперсии также не будут удовлетворять уравнению правдоподобия, что для ρ представляет собой

$$\frac{\partial \ln L}{\partial \rho} = \frac{\rho \varepsilon_1^2}{\sigma_u^2} - \frac{\rho}{1 - \rho^2} + \sum_{t=2}^T \frac{u_t \varepsilon_{t-1}}{\sigma_u^2}.$$

Эту проблему можно обойти просто путем просматривания всех ρ , принадлежащих области $(-1, +1)$, и расчета других оценок в каждой точке, чтобы локализовать максимум функции правдоподобия вручную. С современными компьютерами даже при длинных временных рядах этот объем расчетов был бы незначительным (правда, немного нескладным и неэффективным). В работе Бича и МакКиннона (Beach, MacKinnon (1978a)) был предложен более систематичный алгоритм для поиска ρ в этой модели. После этого итерация далее задается между оцениванием параметров ρ и (β, σ_u^2) обычным образом.

Информационная матрица для этого логарифмического правдоподобия есть

$$-E \left[\frac{\partial^2 \ln L}{\partial \begin{pmatrix} \beta \\ \sigma_u^2 \\ \rho \end{pmatrix} \partial (\beta' \sigma_u^2 \rho)} \right] = \begin{bmatrix} \frac{1}{\sigma_u^2} \mathbf{X}' \Omega^{-1} \mathbf{X} & \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & \frac{T}{2\sigma_u^4} & \frac{\rho}{\sigma_u^2(1-\rho^2)} \\ \mathbf{0}' & \frac{\rho}{\sigma_u^2(1-\rho^2)} & \frac{T-2}{1-\rho^2} + \frac{1+\rho^2}{(1-\rho^2)^2} \end{bmatrix}. \quad (14-64)$$

Обратим внимание, что диагональные элементы в матрице являются $O(T)$. Но элементы (2, 3) и (3, 2) — это постоянные $O(1)$, которые будут, подобно второй части элемента (3, 3), становиться минимальными по мере увеличения T . Отбрасывание этих «концевых эффектов» (и рассмотрение $T - 2$ не отличающимся от T по мере увеличения T) дает диагональную матрицу, из которой мы извлекаем стандартные отклонения для ML-оценок в этой модели:

$$\begin{aligned} \text{Asy. Var}[\hat{\beta}] &= \sigma_u^2 (\mathbf{X}' \Omega^{-1} \mathbf{X})^{-1}, \\ \text{Asy. Var}[\hat{\sigma}_u^2] &= \frac{2\sigma_u^4}{T}, \\ \text{Asy. Var}[\hat{\rho}] &= \frac{1 - \rho^2}{T}. \end{aligned} \quad (14-65)$$

Пример 14.7. Автокорреляция в уравнении спроса на деньги

Используя макроэкономические данные из табл. F5.2, мы оценим следующее уравнение спроса на деньги:

$$\ln(M1/CPI_u)_t = \beta_1 + \beta_2 \ln RealGDP_t + \beta_3 \ln T - billrate_t + \varepsilon_t.$$

Остатки метода наименьших квадратов, показанные на рис. 14.3, отображают типичный пример сильно автокоррелированных рядов.

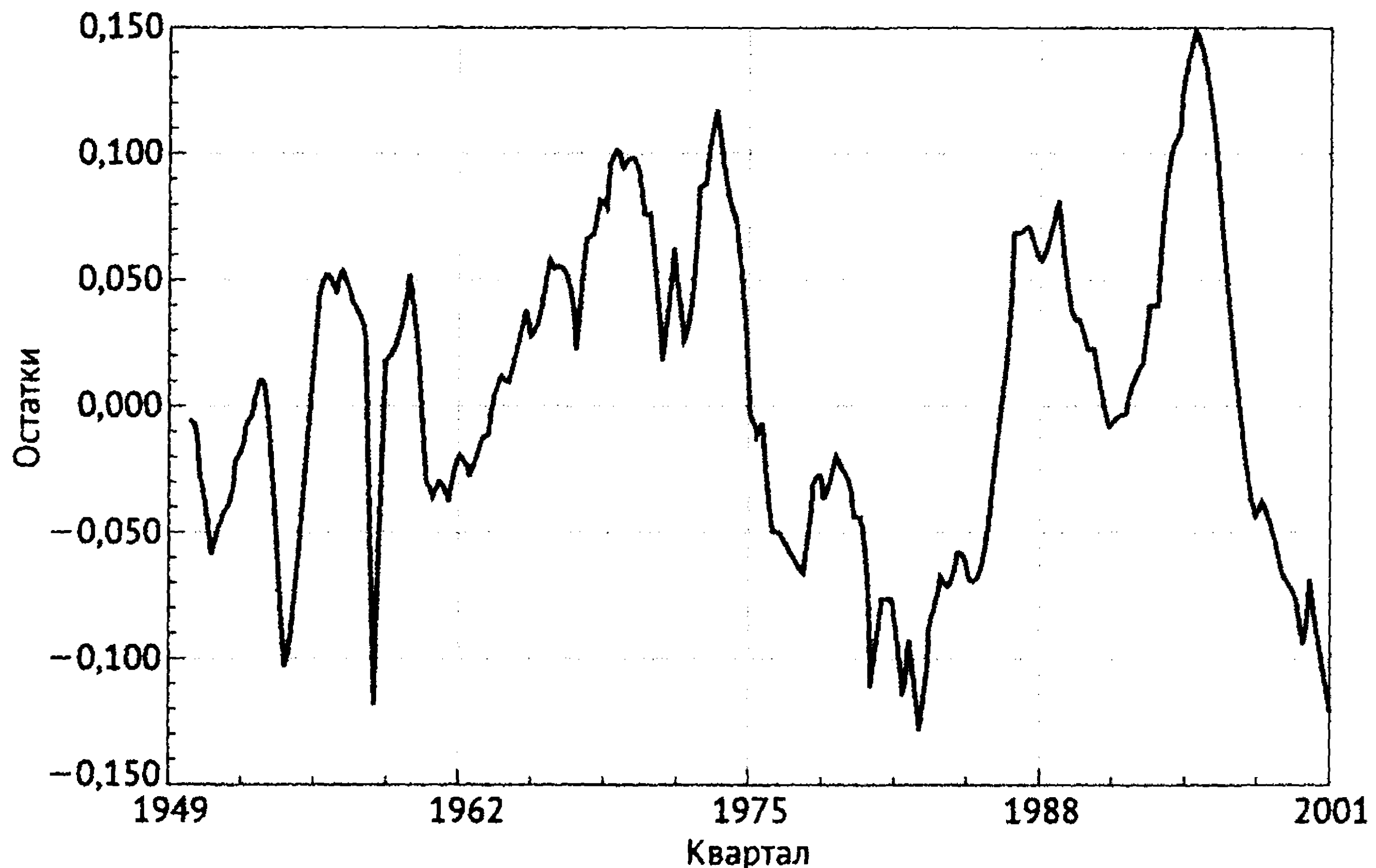


Рис. 14.3. Остатки оцененного уравнения спроса на деньги

Таблица 14.4. Оценки уравнения спроса на деньги: $T = 204$

| Переменная | OLS | | Прайса–Уинстона | | Максимального правдоподобия | |
|-----------------------------|---------|---------------|-----------------|---------------|-----------------------------|---------------|
| | Оценка | Станд. ошибка | Оценка | Станд. ошибка | Оценка | Станд. ошибка |
| Постоянная | -2,1316 | 0,09100 | -1,4755 | 0,2550 | -1,6319 | 0,4296 |
| Ln реального ВВП | 0,3519 | 0,01205 | 0,2549 | 0,03097 | 0,2731 | 0,0518 |
| Ln T — вексельная ставка | -0,1249 | 0,009841 | -0,02666 | 0,007007 | -0,02522 | 0,006941 |
| σ_ε | | 0,06185 | | 0,07767 | | 0,07571 |
| σ_u | | 0,06185 | | 0,01298 | | 0,01273 |
| ρ | 0,0 | 0,0 | 0,9557 | 0,02061 | 0,9858 | 0,01180 |

Простая автокорреляция первого порядка остатков обычного метода наименьших квадратов выражается как $r = 1 - d/2 = 0,9557$, где d —

это статистика Дарбина–Уотсона (см. 20-23). Далее мы оцениваем модель заново, используя FGLS-оценку Прайса–Уинстона, а также оценку методом максимального правдоподобия, полученную с помощью алгоритма Бича–МакКиннона. Результаты представлены в табл. 14.4. Несмотря на то что OLS-оценка в этой модели является состоятельной, FGLS- и ML-оценки сильно отличаются.

14.9.3. Модель внешне не связанных регрессий

Общее выражение модели внешне не связанных регрессий (SUR, seemingly unrelated regression) дано в (10-1)–(10-3):

$$\begin{aligned} y_i &= \mathbf{X}_i \boldsymbol{\beta}_i + \varepsilon_i, i = 1, \dots, M, \\ E[\varepsilon_i | \mathbf{X}_1, \dots, \mathbf{X}_M] &= 0, \\ E[\varepsilon_i \varepsilon_j' | \mathbf{X}_1, \dots, \mathbf{X}_M] &= \sigma_{ij} \mathbf{I}. \end{aligned} \quad (14-66)$$

FGLS-оценка этой модели подробно разобрана в подразделе 10.2.3. Теперь мы добавим предположение о нормальном распределении шоков в этой модели и построим оценки методом максимального правдоподобия. При данной ковариационной структуре, определенной в (14-66), предположение о совместном нормальном распределении должно относиться к вектору из M шоков, наблюдаемых в момент t , что можно записать как

$$\varepsilon_t | \mathbf{X}_1, \dots, \mathbf{X}_M \sim N[\mathbf{0}, \boldsymbol{\Sigma}], t = 1, \dots, T. \quad (14-67)$$

14.9.3а. Объединенная модель

Объединенная модель (pooled model, модель пула), в которой совпадают все векторы коэффициентов, представляет удобную отправную точку нашего исследования. При предположении об одинаковых векторах коэффициентов модель регрессии приобретает вид

$$\begin{aligned} y_{it} &= \mathbf{x}'_{it} \boldsymbol{\beta} + \varepsilon_{it}, \\ E[\varepsilon_{it} | \mathbf{X}_1, \dots, \mathbf{X}_M] &= 0, \\ E[\varepsilon_{it} \varepsilon_{js}] &= \sigma_{ij}, \text{ если } t = s, \text{ и } 0, \text{ если } t \neq s. \end{aligned} \quad (14-68)$$

Это модель гетероскедастичности и межобъектной корреляции. При многомерном нормальном распределении логарифмическое правдоподобие представляет собой

$$\ln L = \sum_{t=1}^T \left[-\frac{M}{2} \ln 2\pi - \frac{1}{2} \ln |\boldsymbol{\Sigma}| - \frac{1}{2} \boldsymbol{\varepsilon}'_t \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}_t \right]. \quad (14-69)$$

Как мы видели ранее, эффективной оценкой для этой модели является GLS, как показано в (10-21). Поскольку элементы $\boldsymbol{\Sigma}$ должны быть оценены, используется FGLS-оценка, основанная на (10-9).

Как мы уже видели в нескольких моделях, оценка методом максимального правдоподобия β при заданной Σ есть GLS-оценка, основанная на (10-21). Оценка методом максимального правдоподобия Σ представляет собой

$$\hat{\sigma}_{ij} = \frac{(y'_i - X_i \hat{\beta}_{ML})' (y_j - X_j t - \hat{2}_{pt} \beta_{ML})}{T} = \frac{\hat{\epsilon}'_i \hat{\epsilon}_j}{T} \quad (14-70)$$

и основывается на ML-оценке β . Если каждая из ML-оценок нуждается в другой, как тогда требуется действовать, чтобы получить обе оценки? Ответ дан в работе Оберхофера и Кменты (Oberhofer, Kmenta (1974)), в которой показано, что в ряде случаев, включая и этот, можно выполнять итерации туда и обратно между двумя оценками. Таким образом, ML-оценки могут быть получены итерационным способом, повторяемым до сходимости, между (14-70) и

$$\hat{\beta} = [X' \hat{\Omega}^{-1} X]^{-1} [X' \hat{\Omega}^{-1} y]. \quad (14-71)$$

Процесс может начинаться с (состоятельной) оценки обычным методом наименьших квадратов, затем (14-70) и т.д. Вычисления не являются сложными и задействуют лишь основы линейной алгебры. Критерии для проверки гипотез о параметре β могут быть построены с помощью хорошо известной нам статистики Вальда. Подходящей оценкой асимптотической ковариационной матрицы является обратная матрица в квадратных скобках в формуле (10-21).

Для проверки гипотезы о том, что недиагональные элементы Σ равны нулю, т. е. о том, что между фирмами нет корреляции, могут быть использованы три подхода. Критерий отношения правдоподобия основывается на статистике

$$\lambda_{LR} = T(\ln |\hat{\Sigma}_{heteroskedastic}| - \ln |\hat{\Sigma}_{general}|) = T \left(\sum_{i=1}^M \ln \hat{\sigma}_i^2 - \ln |\hat{\Sigma}| \right), \quad (14-72)$$

где $\hat{\sigma}_i^2$ — оценки σ_i^2 , полученные из оценок методом максимального правдоподобия групповой гетероскедастичной модели, а $\hat{\Sigma}$ — оценка максимального правдоподобия в модели без ограничений. (Обратим внимание, как избыточная дисперсия, появившаяся в модели с ограничениями, используется для построения критерия.) Асимптотическое распределение этой статистики есть хи-квадрат с $M(M-1)/2$ степенями свободы. Другой вариант — это критерий множителей Лагранжа, разобранный в работе Бройша и Пэгана (Breusch, Pagan (1980)). Общий вид этой статистики представляет собой

$$\lambda_{LM} = T \sum_{i=2}^n \sum_{j=1}^{i-1} r_{ij}^2, \quad (14-73)$$

где r_{ij}^2 — коэффициент корреляции между i -м и j -м остатками. Если бы каждое уравнение имело свой вектор параметров, то обычный метод наименьших квадратов, примененный к каждому уравнению, был бы эффективным (и совпадал бы с ML) и мы бы рассчитывали r_{ij} исходя из OLS-остатков (при

допущении, что есть достаточно много наблюдений для вычислений). Однако здесь мы предполагаем общий вектор параметров. Следовательно, подходящей основой для вычисления корреляций здесь являются остатки, использующие итерационную оценку в групповой гетероскедастичной модели, т. е. те же самые остатки, которые уже использовались для вычисления $\hat{\sigma}_i^2$. (Асимптотически верное приближение к этому критерию может быть основано на FGLS-остатках.) Обратим внимание, что это не метод для проверки полностью классической гомоскедастичной модели регрессии. Для этого случая необходимо включить другие LM- и LR-статистики, основанные на групповой гетероскедастичной модели. Если либо LR-статистика в (14-72), либо LM-статистика в (14-73) меньше, чем критическое значение из таблицы, вывод, основанный на этом критерии, заключается в том, что подходящей моделью является групповая гетероскедастичная модель.

14.9.3b. SUR-модель

Условия работы Оберхофера и Кменты (Oberhofer, Kmenta (1974)) соблюдаются для модели внешне не связанных регрессий, поэтому оценки максимального правдоподобия могут быть получены путем итерирования FGLS-метода. Еще раз обратим внимание, что этот метод предполагает использование (10-9) для оценивания σ_{ij} на каждой итерации. Максимальное правдоподобие не обладает преимуществами над FGLS в смысле асимптотических свойств²⁰. Вопрос о том, является ли он более предпочтительным при малых выборках, остается открытым: ответ на него будет зависеть от конкретного набора данных.

14.9.3с. Исключающие ограничения

Просто сделав подстановку матрицы Ω конкретного вида в логарифмическую функцию правдоподобия для обобщенной регрессионной модели в (14-48), мы можем провести непосредственную максимизацию вместо итерированной FGLS. Тем не менее полезно пересмотреть модель в несколько иной форме. Это альтернативное построение функции правдоподобия появляется также во многих других моделях в ряде источников.

Рассмотрим одно наблюдение по каждой из M зависимых переменных и связанных с ними регрессоров. Запишем это наблюдение горизонтально, а не вертикально. Модель для этого наблюдения запишется как

$$\begin{aligned} [y_1 y_2 y_M]_t &= [\mathbf{x}_t^*]' [\pi_1 \pi_2 \pi_M] + [\varepsilon_1 \varepsilon_2 \varepsilon_M]_t \\ &= [\mathbf{x}_t^*]' \Pi + \mathbf{E}', \end{aligned} \quad (14-74)$$

где \mathbf{x}_t^* — полный набор всех K^* различных независимых переменных, которые участвуют в модели. Тогда матрица параметров имеет один столбец для каждого уравнения, но столбцы не совпадают с β_i в (14-66), за исключением случая, когда каждая переменная появляется в каждом уравнении. В противном случае в i -м уравнении π_i будет включать несколько нулей,

²⁰В работе Jensen (1995) был обнаружен некоторый разброс при вычислении асимптотической ковариационной матрицы для оценки, что допускает возможность нарушения предположения о нормальном распределении.

каждый из которых является **исключающим ограничением**. Рассмотрим, например, модель издержек производства с двумя уравнениями для двух авиаперевозчиков:

$$\begin{aligned} C_{1t} &= \alpha_1 + \beta_{1P}P_{1t} + \beta_{1L}LF_{1t} + \varepsilon_{1t}, \\ C_{2t} &= \alpha_2 + \beta_{2P}P_{2t} + \beta_{2L}LF_{2t} + \varepsilon_{2t}, \end{aligned}$$

где C — издержки, P — цена на топливо, а LF — коэффициент загрузки. t -е наблюдение будет записываться как

$$[C_1 \quad C_2]_t = [1P_1LF_1P_2LF_2]_t \begin{bmatrix} \alpha_1 & \alpha_2 \\ \beta_{1P} & 0 \\ \beta_{1L} & 0 \\ 0 & \beta_{2P} \\ 0 & \beta_{2L} \end{bmatrix} + [\varepsilon_1\varepsilon_2]_t.$$

Этот вектор представляет собой одно наблюдение. Пусть ε_t является вектором M шоков для этого наблюдения, записанного (на данный момент) в виде столбца. Тогда $E[\varepsilon_t\varepsilon_t'] = \Sigma$. Логарифм плотности совместного нормального распределения этих M шоков есть

$$\ln L_t = -\frac{M}{2} \ln(2\pi) - \frac{1}{2} \ln|\Sigma| - \frac{1}{2} \varepsilon_t' \Sigma^{-1} \varepsilon_t. \quad (14-75)$$

Логарифмическое правдоподобие для выборки из T совместных наблюдений представляет собой сумму подобных членов по t :

$$\ln L = \sum_{t=1}^T \ln L_t = -\frac{MT}{2} \ln(2\pi) - \frac{T}{2} \ln|\Sigma| - \frac{1}{2} \sum_{t=1}^T \varepsilon_t' \Sigma^{-1} \varepsilon_t. \quad (14-76)$$

Выражение под знаком суммирования в (14-76) — это скалярная величина, которая равна своему следу. Мы всегда можем переставлять матрицы под оператором следа, поэтому

$$\sum_{t=1}^T \varepsilon_t' \Sigma^{-1} \varepsilon_t = \sum_{t=1}^T \text{tr}(\varepsilon_t' \Sigma^{-1} \varepsilon_t) = \sum_{t=1}^T \text{tr}(\Sigma^{-1} \varepsilon_t \varepsilon_t'). \quad (14-77)$$

Это может быть упрощено далее. Сумма следов T -матриц равна следу суммы матриц [см. (A-91)]. Теперь мы также сможем вынести постоянную матрицу Σ^{-1} за знак суммы. Наконец, нам пригодится умножить и поделить все выражение на T . Объединяя все три шага, мы получаем

$$\sum_{t=1}^T \text{tr}(\Sigma^{-1} \varepsilon_t \varepsilon_t') = T \text{tr} \left[\Sigma^{-1} \left(\frac{1}{T} \sum_{t=1}^T \varepsilon_t \varepsilon_t' \right) \right] = T \text{tr}(\Sigma^{-1} \mathbf{W}), \quad (14-78)$$

где

$$\mathbf{W}_{ij} = \frac{1}{T} \sum_{t=1}^T \varepsilon_{ti} \varepsilon_{tj}.$$

Поскольку на этом шаге используются фактические шоки $E[\mathbf{W}_{ij}] = \sigma_{ij}$, \mathbf{W} — это $M \times M$ матрица, которую мы бы использовали для оценки Σ , если бы ε 's были наблюдаемы. Подставляя этот результат в логарифмическое правдоподобие, мы получаем

$$\ln L = -\frac{T}{2} [M \ln(2\pi) + \ln|\Sigma| + \text{tr}(\Sigma^{-1}\mathbf{W})]. \quad (14-79)$$

Теперь рассмотрим максимизацию этой функции.

Было показано²¹, что

$$\begin{aligned} \frac{\partial \ln L}{\partial \Pi'} &= \frac{T}{2} \mathbf{X}^{*'} \mathbf{E} \Sigma^{-1}, \\ \frac{\partial \ln L}{\partial \Sigma} &= -\frac{T}{2} \Sigma^{-1} (\Sigma - \mathbf{W}) \Sigma^{-1}, \end{aligned} \quad (14-80)$$

где x_t^{*} в (14-74) есть строка t матрицы \mathbf{X}^* . Приравнивая вторую из этих производных к нулевой матрице, мы видим, что при заданных оценках максимального правдоподобия параметров коэффициентов оценка максимального правдоподобия Σ — это \mathbf{W} , матрица средних сумм квадратов остатков и попарных произведений, т. е. матрица, которую мы использовали для FGLS. [Заметим, что не проводится корректировка числа степеней свободы; из условия $\partial \ln L / \partial \Sigma = 0$ следует (10-9).]

Мы также знаем, что, поскольку эта модель является обобщенной регрессионной моделью, оценка максимального правдоподобия матрицы параметров $[\beta]$ должна быть эквивалентна FGLS-оценке, которую мы рассматривали ранее²². Полезно посмотреть, что из этого следует. Если мы подставляем наше решение для Σ в функцию правдоподобия, то получаемое **концентрированное логарифмическое правдоподобие** (concentrated log-likelihood)

$$\ln L_c = -\frac{T}{2} [M(1 + \ln(2\pi)) + \ln|\mathbf{W}|]. \quad (14-81)$$

Следовательно, мы показали, что критерием для выбора оценки максимального правдоподобия β является

$$\hat{\beta}_{ML} = \text{Min}_{\beta} \frac{1}{2} \ln|\mathbf{W}| \quad (14-82)$$

при условии соблюдения исключаящих ограничений. Этот важный результат появляется во многих других моделях и условиях. Эта минимизация должна быть сделана с учетом ограничивающих условий на матрицу параметров. В нашем примере с двумя уравнениями в матрице параметров есть два блока нулей, которые также будут присутствовать в MLE. Оценка β представляет собой множество ненулевых элементов в матрице параметров в формуле (14-74).

Статистика отношения правдоподобий (likelihood ratio statistic) является альтернативой F -статистики, о которой говорилось ранее, для проверки гипотез о параметре β . Статистика отношения правдоподобий выражается как²³

$$\lambda = -2(\ln L_r - \ln L_u) = T(\ln|\hat{\mathbf{W}}_r| - \ln|\hat{\mathbf{W}}_u|), \quad (14-83)$$

²¹См., например, работу Joreskog (1973).

²²Эта эквивалентность устанавливается из условий Оберхофера-Кменты.

²³В работе Attfield (1998) эти вычисления проведены немного аккуратнее для улучшения поведения статистики в малых выборках.

где \hat{W}_r и \hat{W}_u — матрицы сумм квадратов остатков и их попарных произведений, использующие оценки при ограничениях и без ограничений соответственно. При нулевой гипотезе о верности ограничений предельное распределение статистики отношения правдоподобия представляет собой хи-квадрат с числом степеней свободы, равным числу ограничений. Этот метод может быть также использован для проверки ограничения однородности в многомерной регрессионной модели. Модель с ограничениями — это модель пула, рассмотренная в предыдущем разделе.

Интерес также может представлять проверка того, является ли Σ диагональной матрицей. Два возможных подхода были предложены в подразделе 14.9.3а [см. (14-72) и (14-73)]. Модель без ограничений — это модель, которая используется нами здесь, в то время как модель при ограничениях — это групповая гетероскедастичная модель из подраздела 9.8.2 (пример 9.5), без ограничения о векторе совпадающих параметров. Таким образом, модель с ограничениями представляет собой отдельные регрессионные модели, допускающие оценивание обычным методом наименьших квадратов. Статистика отношения правдоподобий будет иметь вид

$$\lambda_{LR} = T \left[\sum_{i=1}^M \log \hat{\sigma}_i^2 - \log |\hat{\Sigma}| \right], \quad (14-84)$$

где $\hat{\sigma}_i^2$ — это $e_i' e_i / T$ из отдельных регрессий метода наименьших квадратов, а $\hat{\Sigma}$ — это оценка максимального правдоподобия Σ . Эта статистика имеет при проверяемой гипотезе предельное хи-квадрат-распределение с $M(M-1)/2$ степенями свободы. Другим способом, предложенным в работе Бройша и Пэгана (Breusch, Pagan (1980)), является **статистика множителей Лагранжа** (Lagrange multiplier statistic):

$$\lambda_{LM} = T \sum_{i=2}^M \sum_{j=1}^{i-1} r_{ij}^2, \quad (14-85)$$

где r_{ij} есть оцененная корреляция $\hat{\sigma}_{ij} / [\hat{\sigma}_{ii} \hat{\sigma}_{jj}]^{1/2}$. Эта статистика также имеет предельное распределение хи-квадрат с $M(M-1)/2$ степенями свободы. Этот критерий имеет то преимущество, что не требует вычисления оценки максимального правдоподобия Σ , поскольку основан на OLS-остатках.

Пример 14.8. ML-оценки модели внешне не связанных регрессий

Хотя немного устаревшие, данные Grunfeld, использованные в модели 11.2, выдержали проверку временем и до сих пор являются стандартным набором данных, который используется для демонстрации SUR-модели. Данные в табл. F10.4 приложения представлены для 10 фирм за 20 лет (1935–1954). Для наших демонстрационных целей мы будем использовать первые четыре фирмы. [Данные загружены с веб-сайта книги Baltagi (2005) по адресу: <http://www.wiley.com/legacy/wileychi/baltagi/supp/Grunfeld.fil>. См. также работу Kleiber, Zeileis (2010).]

Модель представляет собой уравнение инвестиций

$$I_{it} = \beta_{1i} + \beta_{2i} F_{it} + \beta_{3i} C_{it} + \varepsilon_{it}, t = 1, \dots, 20, i = 1, \dots, 10,$$

где I_{it} — реальные валовые инвестиции для фирмы i в году t , F_{it} — реальная стоимость выпущенных в обращение акций фирмы}}, C_{it} — реальная стоимость акционерного капитала.

OLS-оценки для четырех уравнений показаны в левой части табл. 14.5. Корреляционная матрица для четырех OLS-векторов остатков имеет вид

$$\mathbf{R}_e = \begin{bmatrix} 1 & -0,261 & 0,279 & -0,273 \\ -0,261 & 1 & 0,428 & 0,338 \\ 0,279 & 0,428 & 1 & -0,0679 \\ -0,273 & 0,338 & -0,0679 & 1 \end{bmatrix}.$$

Прежде чем перейти к FGLS- и ML-оценкам, мы применяем LM-критерий для проверки нулевой гипотезы о том, что регрессии действительно являются несвязанными. В качестве упражнения мы оставляем задачу показать, что LM-статистика в (14-85) может быть рассчитана как

$$\lambda_{LM} = (T/2)[\text{trace}(\mathbf{R}'_e \mathbf{R}_e) - M] = 10,451.$$

95%-е критическое значение из распределения хи-квадрат с 6 степенями свободы составляет 12,59, поэтому на этой стадии оказывается, что нулевая гипотеза не отвергается. Будем строить дальнейшее исследование, опираясь на этот результат.

Следующий шаг состоит в том, чтобы рассчитать ковариационную матрицу для OLS-остатков, используя

$$\mathbf{W} = (1/T)\mathbf{E}'\mathbf{E} = \begin{bmatrix} 7160,29 & -1967,05 & 607,533 & -282,756 \\ -1967,05 & 7904,66 & 978,45 & 367,84 \\ 607,533 & 978,45 & 660,829 & -21,3757 \\ -282,756 & 367,84 & -21,3757 & 149,872 \end{bmatrix},$$

где \mathbf{E} — 20×4 матрица OLS-остатков. Собирая данные в блочные матрицы

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & 0 & 0 & 0 \\ 0 & \mathbf{X}_2 & 0 & 0 \\ 0 & 0 & \mathbf{X}_3 & 0 \\ 0 & 0 & 0 & \mathbf{X}_4 \end{bmatrix} \quad \text{и} \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix},$$

мы можем вычислить $\hat{\Omega} = \mathbf{W} \otimes \mathbf{I}_{20}$ и FGLS-оценку

$$\hat{\beta} = [\mathbf{X}'\hat{\Omega}^{-1}\mathbf{X}]^{-1}\mathbf{X}'\hat{\Omega}^{-1}\mathbf{y}.$$

Оцененная асимптотическая ковариационная матрица для FGLS-оценок — это заключенная в скобки обратная матрица. Эти результаты представлены в центре табл. 14.5.

Для вычисления MLE мы воспользуемся результатом Оберхоффера-Кменты и выполним итерации FGLS-оценки. Используя FGLS-вектор коэффициентов, мы пересчитываем остатки, затем \mathbf{W} , далее $\hat{\beta}$. Итерации повторяются до тех пор, пока вектор оценок параметров не сойдется. В качестве показателя сходимости мы используем следующий критерий,

основанный на изменении в оценке параметра от итерации $(s - 1)$ к итерации (s) :

$$\delta = [\hat{\beta}(s) - \hat{\beta}(s - 1)]' [\mathbf{X}' [\hat{\Omega}(s)]^{-1} \mathbf{X}] [\hat{\beta}(s) - \hat{\beta}(s - 1)].$$

Последовательность значений этого критерия следующая: 0,21922, 0,16318, 0,00662, 0,00037, 0,00002367825, 0,000001563348, $0,1041980 \times 10^{-6}$. Мы заканчиваем процесс после итерации 7. ML-оценки показаны в правой части табл. 14.5.

Таблица 14.5. Эмпирические уравнения инвестиций

| Фирма | Переменная | OLS | | FGLS | | SMLE | |
|-------|------------|---------|------------|---------|------------|---------|------------|
| | | Оценка | Станд. ош. | Оценка | Станд. ош. | Оценка | Станд. ош. |
| 1 | Постоянная | -149,78 | 97,58 | -160,68 | 90,41 | -179,41 | 6,66 |
| | F | 0,1192 | 0,02382 | 0,1205 | 0,02187 | 0,1248 | 0,02086 |
| | C | 0,3714 | 0,03418 | 0,3800 | 0,03311 | 0,3802 | 0,03266 |
| 2 | Постоянная | -49,19 | 136,52 | 1,16 | 116,18 | 6,46 | 106,18 |
| | F | 0,1749 | 0,06841 | 0,1304 | 0,05737 | 0,1244 | 0,05191 |
| | C | 0,3896 | 0,1312 | 0,4485 | 0,1225 | 0,4367 | 0,1171 |
| 3 | Постоянная | -9,956 | 8,92 | -19,72 | 6,58 | -24,10 | 5,80 |
| | F | 0,02655 | 0,01435 | 0,03464 | 0,01279 | 0,03808 | 0,01217 |
| | C | 0,1517 | 0,02370 | 0,1368 | 0,02249 | 0,1311 | 0,02223 |
| 4 | Постоянная | -6,190 | 2,45 | 0,9366 | 1,59 | 2,581 | 1,54 |
| | F | 0,07795 | 0,01841 | 0,06785 | 0,01705 | 0,06564 | 0,01698 |
| | C | 0,3157 | 0,02656 | 0,3146 | 0,02606 | 0,3137 | 0,02617 |

Далее мы проверим гипотезу о том, что ковариационная матрица является диагональной. Оценка максимального правдоподобия Σ записывается как

$$\hat{\Sigma} = \begin{bmatrix} \mathbf{7235,46} & -2455,13 & 615,167 & -325,413 \\ -2455,13 & \mathbf{8146,41} & 1288,66 & 427,011 \\ 615,167 & 1288,66 & \mathbf{702,268} & 2,51786 \\ -325,413 & 427,011 & 2,51786 & \mathbf{153,889} \end{bmatrix}.$$

Оценкой для модели при ограничениях является диагональная матрица, сформированная из диагонали \mathbf{W} , вычисленной ранее для OLS-оценки. (Оценки выделены жирным шрифтом в предыдущей матрице.) Тогда статистика критерия представляет собой

$$LR = T(\ln |\text{diag}(\mathbf{W})| - \ln |\hat{\Sigma}|) = 18,55.$$

Напомним, что критическое значение равно 12,59. Результаты противоречат результатам LM-статистики. Гипотеза о диагональности ковариационной матрицы теперь отвергается.

Обратим внимание, что помимо постоянных четыре набора оценок коэффициентов являются весьма схожими. Из-за постоянных почти нет сомнений в том, что гипотеза пула будет отвергнута. Чтобы проверить это, вычислим статистику Вальда, основанную на MLE-результатах. Для проверки

$$H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4$$

переформулируем эту гипотезу как

$$H_0: \beta_1 - \beta_4 = 0, \beta_2 - \beta_4 = 0, \beta_3 - \beta_4 = 0.$$

Статистика Вальда есть

$$\lambda_W = (\mathbf{R}\hat{\beta} - \mathbf{q})' [\mathbf{RVR}']^{-1} (\mathbf{R}\hat{\beta} - \mathbf{q}) = 2190,96,$$

где $\mathbf{R} = \begin{bmatrix} \mathbf{I}_3 & \mathbf{0} & \mathbf{0} & -\mathbf{I}_3 \\ \mathbf{0} & \mathbf{I}_3 & \mathbf{0} & -\mathbf{I}_3 \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & -\mathbf{I}_3 \end{bmatrix}$, $\mathbf{q} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}$ и $\mathbf{V} = [\{\mathbf{X}'\hat{\Omega}^{-1}\mathbf{X}\}]^{-1}$. При нулевой

гипотезе статистика Вальда имеет предельное распределение хи-квадрат с 9 степенями свободы. Критическое значение равно 16,92, поэтому, как мы и предполагали, гипотеза отвергается. Возможно, разница обусловлена лишь различными постоянными членами. Чтобы проверить гипотезу о том, что четыре пары коэффициентов наклона равны, заменим \mathbf{I}_3 в \mathbf{R} на $[\mathbf{0}, \mathbf{I}_2]$, $\mathbf{0}$ s на 2×3 нулевые матрицы, а \mathbf{q} на 6×1 нулевой вектор. Результирующая статистика хи-квадрат равна 229,005. Критическое значение составляет 12,59, поэтому и эта гипотеза отвергается.

14.9.4. Модели одновременных уравнений

В главе 10 мы отметили два подхода к оцениванию системы уравнений методом максимального правдоподобия:

$$\begin{aligned} y_t' \Gamma + x_t' \mathbf{B} &= \varepsilon_t', \\ \varepsilon_t | \mathbf{X} &\sim N[\mathbf{0}, \Sigma]. \end{aligned} \quad (14-86)$$

Оценка методом максимального правдоподобия на ограниченной информации (limited information maximum likelihood, LIML) — это подход, который основан на одном уравнении и оценивает параметры лишь одного уравнения единовременно. Оценка методом максимального правдоподобия на полной информации (full information maximum likelihood, FIML) сразу анализирует полную систему уравнений.

Вывод LIML-оценки достаточно сложен. Подробные рассуждения на эту тему можно найти в работах Андерсона и Рубина (Anderson, Rubin (1948)), Тейла (Theil (1971)) и Дэвидсона, МакКиннона (Davidson, MacKinnon (1993, глава 18)). Техника же расчетов, как уже было показано выше (подраздел 10.6.4), оказывается на удивление простой. LIML-оценки для модели I Клейна участвуют в примере 10,6 наряду с другими оценками, основанными как на одном уравнении, так и на системе. Для практика полезный результат заключается в том, что асимптотическая дисперсия оценки двухшаговым ме-

тодом наименьших квадратов (2SLS, two-stage least squares), которую проще вычислить, является такой же, как для LIML-оценки. Для практических целей в целом это сделало бы LIML-оценку с ее дополнительным допущением о нормальном распределении спорной. Плюсы LIML-оценки большей частью теоретические: она обеспечивает хорошую точку отсчета для анализа свойств оценок, построенных на одном уравнении. Единственным исключением будет инвариантность оценки к нормализации уравнения (т.е. к тому, какая переменная появляется слева от знака равенства). Это становится полезным в контексте анализа проблемы слабых инструментов. (См. разделы 8.7 и 10.6.6.)

FIML-оценку гораздо проще вывести, чем LIML-оценку, но значительно труднее вычислить. Чтобы получить необходимые результаты, сначала мы рассмотрим приведенную форму модели

$$\begin{aligned} y'_t &= x'_t \Pi + v'_t, \\ v_t | \mathbf{X} &\sim N[\mathbf{0}, \Omega], \end{aligned} \quad (14-87)$$

которая является моделью кажущихся несвязанными регрессий, детально проанализированной нами в главе 10 и подразделе 14.9.3. Сложность связана с ограничениями, накладываемыми на параметры:

$$\Pi = -\mathbf{B}\Gamma^{-1} \quad \text{и} \quad \Omega = (\Gamma^{-1})' \Sigma (\Gamma^{-1}). \quad (14-88)$$

Как нам хорошо известно из нескольких рассматриваемых моделей, при заданных оценках Γ и \mathbf{B} в (14-86) оценкой Σ будет выражение $(1/T)\mathbf{E}'\mathbf{E}$, основанное на остатках. Сравнительно легко можно показать, что при заданных Γ и Σ оценка $(-\mathbf{B})$ в (14-86) получается из результатов для SUR-модели в подразделе 14.9.3с (где мы оцениваем модель при наличии исключаящих ограничений в матрице коэффициентов). Сложность в оценивание вносит Γ , определитель Якоби. Член $\ln |\Gamma|$ появляется в логарифмической функции правдоподобия. Нелинейная оптимизация по ненулевым элементам в функции, которая включает этот член, является безумно сложной. В то же время трехшаговый метод наименьших квадратов (3SLS, three-stage least squares) обладает такой же асимптотической эффективностью, как FIML-оценка, при этом не делая предположения о нормальном распределении и не сталкиваясь с практическими сложностями.

В конечном счете для практического применения LIML- и FIML-оценки были вытеснены в литературе намного более простыми GMM-, 2SLS-, H2SLS-, 3SLS- и H3SLS-оценками. LIML- и FIML-оценки в значительной степени интересны как часть теоретических исследований.

14.9.5. Оценка моделей нелинейной регрессии методом максимального правдоподобия

В главе 7 мы рассмотрели модели нелинейной регрессии, в которых нелинейность в параметрах проявляется в правой части уравнения. Метод максимального правдоподобия используется, когда шоки регрессии или, более обще, зависимая переменная не являются нормально распределенными. Примером такой ситуации служит геометрическая регрессионная модель.

Пример 14.9. Идентификация в логлинейной регрессионной модели

В примере 7.6 мы оценили экспоненциальную модель регрессии вида

$$E[Income|Age, Education, Female] = \exp(\gamma_1^* + \gamma_2 Age + \gamma_3 Education + \gamma_4 Female).$$

Это логлинейное условное среднее согласуется с несколькими различными распределениями, включая логнормальное, распределение Вейбулла, гамма-распределение и экспоненциальное. В каждом из этих случаев функция условного среднего имеет следующий вид:

$$\begin{aligned} E[Income|\mathbf{x}] &= g(\theta) \exp(\gamma_1 + \mathbf{x}'\gamma_2) \\ &= \exp(\gamma_1^* + \mathbf{x}'\gamma_2), \end{aligned}$$

где θ — дополнительный параметр распределения, а $\gamma_1^* = \ln g(\theta) + \gamma_1$. Отсюда можно сделать два вывода.

1. Нелинейный метод наименьших квадратов (NLS, nonlinear least squares) является устойчивым, по крайней мере, к некоторым неверным предположениям о распределении. Оценка γ_2 нелинейного метода наименьших квадратов будет состоятельной и асимптотически нормально распределенной во всех случаях, для которых $E[Income|\mathbf{x}] = \exp(\gamma_1^* + \mathbf{x}'\gamma_2)$.

2. NLS-оценка не может дать состоятельную оценку параметра γ_1 : $\text{plim } c_1 = \gamma_1^*$, которая зависит от корректного распределения. В функции условного среднего любая пара значений, для которой совпадают $\gamma_1' = \ln g(\theta) + \gamma_1$, приведет к такой же сумме квадратов. Это вид мультиколлинеарности: квазирегрессором для θ является $\partial E[Income|\mathbf{x}]/\partial \theta = \exp(\gamma_1^* + \mathbf{x}'\gamma_2)[g'(\theta)/g(\theta)]$, тогда как для γ_1 — это $\partial E[Income|\mathbf{x}]/\partial \gamma_1 = \exp(\gamma_1^* + \mathbf{x}'\gamma_2)$. Первое выражение отличается от второго на постоянный множитель.

В отличие от MLE, NLS не может представить отдельные оценки θ и γ_1 — см. следующий пример. Далее, NLS может быть менее эффективным по сравнению с MLE, поскольку не использует информацию о распределении зависимой переменной. Это второе соображение неоднозначно. Для оценивания γ_2 NLS-оценка является менее эффективной из-за неиспользования информации о распределении. Тем не менее этот недостаток мог быть компенсирован, потому что NLS-оценка не пытается рассчитывать независимую оценку дополнительного параметра θ .

В качестве иллюстрации рассмотрим еще раз оценку из примера 7.6. Модель гамма-регрессии задается как

$$f(y|\mathbf{x}) = \frac{\mu(\mathbf{x})^\theta}{\Gamma(\theta)} \exp[-\mu(\mathbf{x})y] y^{\theta-1}, y > 0, \theta > 0, \mu(\mathbf{x}) = \exp(\gamma_1 + \mathbf{x}'\gamma_2).$$

Функция условного среднего для этой модели есть

$$E[y|\mathbf{x}] = \theta/\mu(\mathbf{x}) = \theta \exp(-\gamma_1 - \mathbf{x}'\gamma_2) = \exp(-\gamma_1^* - \mathbf{x}'\gamma_2).$$

В табл. 14.6 представлены оценки θ и (γ_1, γ_2) . В круглых скобках указаны стандартные ошибки. Все оценки в столбцах (1), (2) и (4) рассчитаны с использованием нелинейного метода наименьших квадратов. В (1)

сделана попытка отдельно оценить θ и γ_1 . Алгоритм «сошелся» на некоторых двух значениях. Однако оцениваемые стандартные ошибки по существу являются бесконечно большими величинами. Сходимость к какой-то величине обусловлена ошибкой округления в компьютере. Результаты в столбце (2) относятся к γ_1^* и γ_2 . Суммы квадратов для этих двух оценок, так же как и для столбца (4), составляют 112,19688, указывая, что три результата только показывают три различных набора результатов, для которых γ_1^* является таким же. Оценки максимального правдоподобия представлены в столбце (3). Обратим внимание, что оценка θ получена здесь, потому что предполагаемое гамма-распределение дает другое независимое моментное уравнение для этого параметра, $\partial \ln L / \partial \theta = -n \ln \Psi(\theta) + \sum_i (\ln y_i - \ln \mu(\{x\})) = 0$, в то время как нормальные уравнения для суммы квадратов дают то же нормальное уравнение для θ и γ_1 .

Таблица 14.6. Оценка модели гамма-регрессии

| | (1) NLS | (2) Ограниченная NLS | (3) MLE | (4) NLS/MLE |
|-------------|-------------------------|----------------------------|-----------------------|-----------------------|
| Постоянная | 1,22468 (∞) | 1,69331 (0,04408) | 3,36826 (0,05048) | 3,36380 (0,04408) |
| Возраст | -0,00207 (0,00061) | -0,00207 (0,00061) | -0,00153 (0,00061) | -0,00207 (0,00061) |
| Образование | -0,04792 (0,00247) | -0,04792 (0,00247) | -0,04975 (0,00286) | -0,04792 (0,00247) |
| Женщины | 0,00658 (0,01373) | 0,00658 (0,01373) | -0,00696 (0,01322) | 0,00658 (0,08677) |
| θ | 0,62699 (∞) | — — | 5,31474 (0,10894) | 5,31474 (0,00000) |

Стандартный подход к построению модели количества событий начинается с модели регрессии Пуассона

$$\text{Prob}[Y = y_i | \mathbf{x}_i] = \frac{\exp(-\lambda_i) \lambda_i^{y_i}}{y_i!}, \lambda_i = \exp(\mathbf{x}_i' \boldsymbol{\beta}), y_i = 0, 1, \dots,$$

которая имеет **логлинейную функцию условного среднего** (loglinear conditional mean function) $E[y_i | \mathbf{x}_i] = \lambda_i$. (Модель регрессии Пуассона и другие модели количественных данных обсуждаются в главе 18. Мы рассматриваем здесь этот пример, чтобы начать исследовать MLE непосредственным образом в типичной нелинейной постановке.) В табл. F7.1 приложения представлены данные работы Riphahn et al. (2003), которые мы будем использовать, чтобы проанализировать порядковую переменную *DocVis*, число посещений работников медицины в году исследования. Гистограмма на рис. 14.4 показывает отчетливый всплеск на нуле с последующими резко убывающими частотами. Несмотря на то что распределение Пуассона, которое является типично горбообразным, может обеспечивать эту форму при условии λ_i

меньше единицы, сама форма все-таки является скорее «непуассоновской». [Так называемые модели нулевой инфляции (рассмотренные в главе 18) часто используются в этой ситуации.]

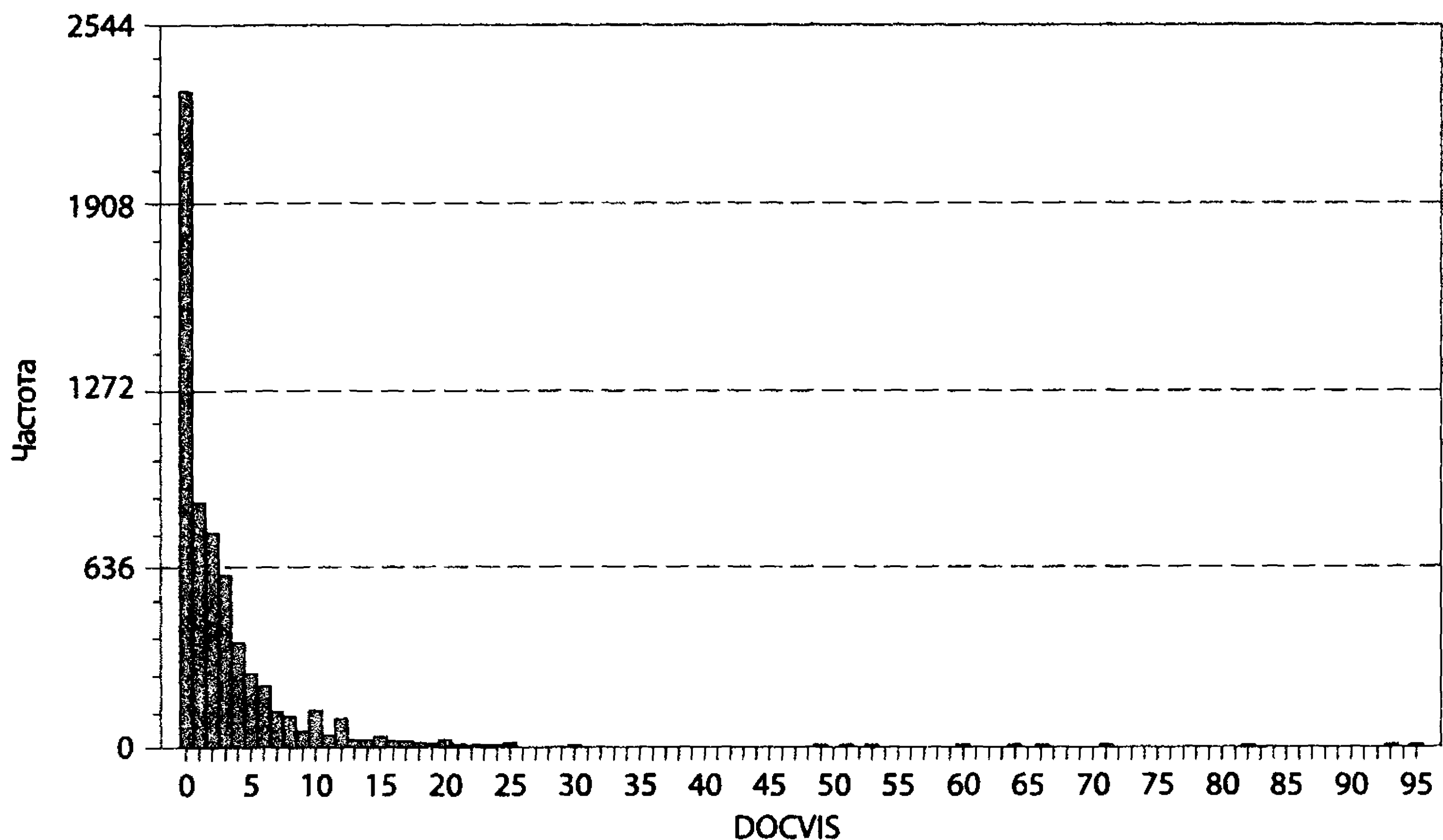


Рис. 14.4. Гистограмма числа посещений врача

Геометрическое распределение

$$f(y_i | \mathbf{x}_i) = \theta_i(1 - \theta_i)^{y_i}, \theta_i = 1/(1 + \lambda_i), \lambda_i = \exp(\mathbf{x}_i' \boldsymbol{\beta}), y_i = 0, 1, \dots$$

является удобной спецификацией для описания эффекта, показанного на рис. 14.4. (Обратим внимание, что формально это описание используется, чтобы смоделировать число неудач перед первой удачей в последовательности независимых испытаний, каждое из которых имеет вероятность успеха θ_i , поэтому фактически в качестве модели для количественной переменной она не подходит. Тем не менее эта модель дает удобную и полезную иллюстрацию.) Функция условного среднего также есть $E[y_i | \mathbf{x}_i] = \lambda_i$. Частные влияния в этой модели выражаются как

$$\frac{\partial E[y_i | \mathbf{x}_i]}{\partial \mathbf{x}_i} = \lambda_i \boldsymbol{\beta},$$

поэтому это, бесспорно, модель нелинейной регрессии. Мы построим оценку максимального правдоподобия, а затем сравним MLE с оценками **нелинейного метода наименьших квадратов** (nonlinear least squares) и (взятыми из неверной спецификации) оценками линейного метода наименьших квадратов.

Логарифмическая функция правдоподобия есть

$$\ln L = \sum_{i=1}^n \ln f(y_i | \mathbf{x}_i, \boldsymbol{\beta}) = \sum_{i=1}^n \ln \theta_i + y_i \ln(1 - \theta_i).$$

Уравнения правдоподобия выглядят как

$$\frac{\partial \ln L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^n \left(\frac{1}{\theta_i} - \frac{y_i}{1 - \theta_i} \right) \frac{d\theta_i}{d\lambda_i} \frac{\partial \lambda_i}{\partial \boldsymbol{\beta}} = \mathbf{0}.$$

Поскольку

$$\frac{d\theta_i}{d\lambda_i} \frac{\partial \lambda_i}{\partial \boldsymbol{\beta}} = \left(\frac{-1}{(1 + \lambda_i)^2} \right) \lambda_i \mathbf{x}_i = -\theta_i(1 - \theta_i) \mathbf{x}_i,$$

уравнения правдоподобия упрощаются до

$$\begin{aligned} \frac{\partial \ln L}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^n (\theta_i y_i - (1 - \theta_i)) \mathbf{x}_i \\ &= \sum_{i=1}^n (\theta_i(1 + y_i) - 1) \mathbf{x}_i. \end{aligned}$$

Для того чтобы оценить асимптотическую ковариационную матрицу, мы можем использовать любую из трех оценок $\text{Asy. Var} [\hat{\boldsymbol{\beta}}_{\text{MLE}}]$. ВНН-оценка выглядит как

$$\begin{aligned} \text{Est. Asy. Var}_{\text{ВНН}}[\hat{\boldsymbol{\beta}}_{\text{MLE}}] &= \left[\sum_{i=1}^n \left(\frac{\partial \ln f(y_i | \mathbf{x}_i, \hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}} \right) \left(\frac{\partial \ln f(y_i | \mathbf{x}_i, \hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}} \right)' \right]^{-1} \\ &= \left[\sum_{i=1}^n (\hat{\theta}_i(1 + y_i) - 1)^2 \mathbf{x}_i \mathbf{x}_i' \right]. \end{aligned}$$

Отрицательная обратная матрица вторых производных, вычисленная в оценке ML, равняется

$$\left[-\frac{\partial^2 \ln L}{\partial \hat{\boldsymbol{\beta}} \partial \hat{\boldsymbol{\beta}}'} \right]^{-1} = \left[\sum_{i=1}^n (1 + y_i) \hat{\theta}_i (1 - \hat{\theta}_i) \mathbf{x}_i \mathbf{x}_i' \right]^{-1}.$$

Наконец, как уже было отмечено, $E[y_i | \mathbf{x}_i] = \lambda_i = (1 - \theta_i)/\theta_i$ известно, поэтому мы также можем выразить математическое ожидание обратной матрицы вторых производных как

$$\left[-E \left(\frac{\partial^2 \ln L}{\partial \hat{\boldsymbol{\beta}} \partial \hat{\boldsymbol{\beta}}'} \right) \right]^{-1} = \left[\sum_{i=1}^n (1 - \hat{\theta}_i) \mathbf{x}_i \mathbf{x}_i' \right]^{-1}.$$

Чтобы вычислить оценки параметров, может быть использован либо **метод Ньютона** (Newton's method)

$$\hat{\boldsymbol{\beta}}^{t+1} = \hat{\boldsymbol{\beta}}^t - [\hat{\mathbf{H}}^t]^{-1} \hat{\mathbf{g}}^t,$$

либо скор-метод

$$\hat{\beta}^{t+1} = \hat{\beta}^t - \{E[\hat{H}^t]\}^{-1} \hat{g}^t,$$

где \mathbf{H} и \mathbf{g} — это вторая и первая производные, вычисленные в текущих оценках параметров. Как и во многих подобных моделях, здесь есть удобный набор начальных значений, если предположить, что модель содержит постоянный член. Поскольку $E[y_i | \mathbf{x}_i] = \lambda_i$, если мы положим начальные значения параметров наклона равными нулю, то естественное начальное значение для постоянного члена — это логарифм \bar{y} .

Пример 14.10. Геометрическая регрессионная модель для числа посещений врача

В примере 7.6 мы рассматривали оценивание логлинейной модели для числа переменной посещений врача, представленной на рис. 14.4, нелинейным методом наименьших квадратов. Данные взяты из набора данных Рипкана и др. (Riphahn et al. (2003)) в табл. F7.1 приложения. Здесь мы продолжим тот анализ путем подбора более подробной модели для количественной переменной *DocVis*. Условное среднее, рассматриваемое сейчас, имеет вид

$$\ln E[\text{DocVis}_{it} | \mathbf{x}_{it}] = \beta_1 + \beta_2 \text{Age}_{it} + \beta_3 \text{Education}_{it} + \beta_4 \text{Income}_{it} + \beta_5 \text{it}.$$

(Это немного отличается от модели в примере 11.16. Для этой задачи, предполагая в будущем рассмотреть модель с фиксированными эффектами (пример 14.13), мы будем рассматривать модель, не содержащую никаких не меняющихся со временем переменных, таких как Female_i .) Выборочные средние значения переменных модели размещены в табл. 14.7. Обращаем внимание, что эти данные представляют собой панель. В этом примере мы пренебрежем этим фактом и будем подбирать модель пула. Мы вернемся к изучению панельных данных в следующем разделе и рассмотрим этот пример еще раз.

Для оптимизации использовался метод Ньютона с начальными значениями, предложенными ранее. Пять итераций выглядят следующим образом:

| Переменная | Постоянная | Возраст | Образование | Доход | Дети |
|---------------------|--------------|---|--------------|--------------|--------------|
| Начальные значения: | 0,11580e+01 | 0,00000e+00 | 0,00000e+00 | 0,00000e+00 | 0,00000e+00 |
| 1-е производные | -0,25191e-08 | -0,61777e+05 | 0,73202e+04 | 0,42575e+04 | 0,16464e+04 |
| Параметры: | 0,11580e+01 | 0,00000e+00 | 0,00000e+00 | 0,00000e+00 | 0,00000e+00 |
| Итерация 1 F = | 0,6287e+05 | $\mathbf{g}'\text{inv}(\mathbf{H})\mathbf{g}$ | 0,4367e+02 | | |
| 1-е производные | 0,48616e+03 | -0,22449e+05 | -0,57162e+04 | -0,17112e+04 | -0,16521e+03 |
| Параметры: | 0,11186e+01 | 0,17563e-01 | -0,50263e-01 | -0,46274e-01 | -0,15609e+00 |
| Итерация 2 F = | 0,6192e+05 | $\mathbf{g}'\text{inv}(\mathbf{H})\mathbf{g}$ | 0,3547e+01 | | |
| 1-е производные | -0,31284e+01 | -0,15595e+03 | -0,37197e+02 | -0,10630e+02 | -0,77186e+00 |

| Переменная | Постоянная | Возраст | Образование | Доход | Дети |
|-----------------|--------------|--------------|--------------|--------------|--------------|
| Параметры: | 0,10922e+01 | 0,17981e-01 | -0,47303e-01 | -0,46739e-01 | -0,15683e+00 |
| Итерация | 0,6192e+05 | $g'inv(H)g$ | 0,2598e-01 | | |
| 3 F= | | = | | | |
| 1-е производные | -0,18417e-03 | -0,99368e-02 | -0,21992e-02 | -0,59354e-03 | -0,25994e-04 |
| Параметры: | 0,10918e+01 | 0,17988e-01 | -0,47274e-01 | -0,46751e-01 | -0,15686e+00 |
| Итерация | 0,6192e+05 | $g'inv(H)g$ | 0,1831e-05 | | |
| 4 F= | | = | | | |
| 1-е производные | -0,35727e-11 | 0,86745e-10 | -0,26302e-10 | -0,61006e-11 | -0,15620e-11 |
| Параметры: | 0,10918e+01 | 0,17988e-01 | -0,47274e-01 | -0,46751e-01 | -0,15686e+00 |
| Итерация | 0,6192e+05 | $g'inv(H)g$ | 0,1772e-12 | | |
| 5 F= | | = | | | |

Таблица 14.7. Оцененная геометрическая регрессионная модель
 Зависимая переменная: DocVis: среднее значение = 3,18352, среднеквадратическое отклонение = 5,68969

| Переменная | Оценка | Станд. ош. H | Станд. ош. E[H] | Станд. ош. ВНН | Станд. ош. кластерной оценки |
|-------------|---------|-----------------|--------------------|-------------------|------------------------------------|
| Постоянная | 1,0918 | 0,0524 | 0,0524 | 0,0354 | 0,1112 |
| Возраст | 0,0180 | 0,0007 | 0,0007 | 0,0005 | 0,0013 |
| Образование | -0,0473 | 0,0033 | 0,0033 | 0,0023 | 0,0069 |
| Доход | -0,0468 | 0,0041 | 0,0042 | 0,0023 | 0,0075 |
| Дети | -0,1569 | 0,0156 | 0,0155 | 0,0103 | 0,0319 |

| Переменная | APE | PE | | |
|-------------|--------|---------------------|--------|---------------------|
| | | Среднее значение | OLS | Среднее значение |
| Постоянная | — | — | 2,656 | |
| Возраст | 0,0572 | 0,0547 | 0,061 | 43,52 |
| Образование | -0,150 | -0,144 | -0,121 | 11,32 |
| Доход | -0,149 | -0,142 | -0,162 | 3,52 |
| Дети | -0,499 | -0,477 | -0,517 | 0,40 |

Сходимость, основанная на LM-критерии $g'H^{-1}g$, достигается после четвертой итерации. Обратим внимание, что производные на этой стадии чрезвычайно малы, хотя и не являются строго нулевыми. В табл. 14.7 представлены оценки максимального правдоподобия для параметров. Представлены также несколько вариантов стандартных ошибок. Сперва даются три набора, основанные на различных оценках информационной матрицы. Четвертый набор основан на кластерной скорректированной ковариационной матрице, рассмотренной в подразделе 14.8.4. Поскольку мы имеем дело фактически с (несбалансированной) панелью, то ожидаем наличия корреляции между наблюдениями. Неудивительно, что

стандартные ошибки существенно возрастают. Частные влияния, приведенные далее в таблице, рассчитываются двумя способами. «Среднее частное влияние» (Average, Partial Effect, APE) рассчитывается путем усреднения $\lambda_i\beta$ между отдельными индивидами в выборке. «Частичный эффект» (Partial Effect, PE) рассчитывается для среднего индивида путем вычисления λ по средним значениям данных. Предпоследний столбец содержит оценки коэффициентов обычного метода наименьших квадратов. В этой модели нет причины ожидать, что обычный метод наименьших квадратов даст состоятельную оценку β . Возникает вопрос: что же оценивает обычный метод наименьших квадратов? Ответ: он оценивает коэффициенты наклона линейной проекции DocVis на x_{it} . Сходство OLS-коэффициентов с оцененными частными влияниями не является случайным и подсказывает ответ на поставленный выше вопрос.

Таблица 14.8. Оценки трех моделей для DocVis

| Переменная | Геометрическая модель | | Модель Пуассона | | Нелинейная регрессия | |
|-------------------------|-----------------------|------------|-----------------|------------|----------------------|------------|
| | Оценка | Станд. ош. | Оценка | Станд. ош. | Оценка | Станд. ош. |
| Постоянная | 1,0918 | 0,0524 | 1,0480 | 0,0272 | 0,9801 | 0,0893 |
| Возраст (Age) | 0,0180 | 0,0007 | 0,0184 | 0,0003 | 0,0187 | 0,0011 |
| Образование (Education) | -0,0473 | 0,0033 | -0,0433 | 0,0017 | -0,0361 | 0,0057 |
| Доход (Income) | -0,0468 | 0,0041 | -0,0520 | 0,0022 | -0,0591 | 0,0072 |
| Дети (Kids) | -0,1569 | 0,0156 | -0,1609 | 0,0080 | -0,1692 | 0,0264 |

В табл. 14.8 представлены три конкурирующих подхода к моделированию DocVis. Результаты для геометрической регрессионной модели представлены в табл. 14.7. В начале этого раздела мы отметили, что более традиционный подход к построению модели для количественной переменной, подобной DocVis, опирается на модель пуассоновской регрессии. Логарифмическая функция правдоподобия и ее производные в этой модели являются даже более простыми, чем в геометрической модели:

$$\ln L = \sum_{i=1}^n y_i \ln \lambda_i - \lambda_i - \ln y_i!,$$

$$\partial \ln L / \partial \beta = \sum_{i=1}^n (y_i - \lambda_i) x_i,$$

$$\partial^2 \ln L / \partial \beta \partial \beta' = \sum_{i=1}^n -\lambda_i x_i x_i'.$$

В качестве третьего подхода можно рассмотреть полупараметрическую нелинейную модель регрессии

$$y_{it} = \exp(\mathbf{x}'_{it}\boldsymbol{\beta}) + \varepsilon_{it}.$$

В действительности это модель, которая включает как геометрический, так и пуассоновский случаи. При условии любого из двух предположений о распределении нелинейный метод наименьших квадратов неэффективен по сравнению с MLE. Однако предположение о распределении может быть полностью отброшено и модель может быть оценена как простая экспоненциальная регрессия. В табл. 14.8 приведены все три набора оценок.

Выбор между этими альтернативами неочевиден. Из трех моделей более часто используется модель Пуассона. Пуассоновская и геометрическая модели не являются вложенными, поэтому мы не можем использовать стандартный параметрический критерий, чтобы выбрать одну из них. Тем не менее эти две модели удовлетворяют условиям критерия Вуонга, описанного в подразделе 14.6.6. Чтобы провести тест, сначала вычислим

$$V_{it} = \ln f_{it} | \text{геометрическая} - \ln f_{it} | \text{Пуассона},$$

используя соответствующие ML-оценки параметров. Тестовая статистика из подраздела 14.6.6 в этом случае выражается как

$$V = \frac{\left(\sqrt{\sum_{i=1}^n T_i}\right) \bar{V}}{s_V}.$$

Эта статистика сходится к стандартному нормальному распределению при соответствующих предположениях. Большое положительное значение свидетельствует в пользу геометрической модели. Вычисленное значение составляет 37,885, что явно свидетельствует в пользу геометрической модели в сравнении с моделью Пуассона.

14.9.6. Модели панельных данных

Применение метода максимального правдоподобия к линейным моделям панельных данных, которые мы рассматривали до настоящего момента, не имеет каких-то особенностей. Для линейной модели случайных эффектов, рассматриваемой в подразделе 14.9.6а, ML-оценка $\boldsymbol{\beta}$, как обычно, представляет собой FGLS при заданных ML-оценках параметров дисперсии. Как мы увидим, последнее создает довольно существенную сложность. Этот пример демонстрирует интересную модель, позволяющую увидеть преимущества свойства инвариантности MLE. Мы перепараметризуем достаточно сложную логарифмическую функцию правдоподобия, чтобы представить ее в простом виде. Ситуация, где метод максимального правдоподобия становится действительно нужным, — это анализ фиксированных и случайных эффектов в нелинейных моделях. Мы опишем два метода в общем виде в подразделах 14.9.6b и 14.9.6c, а затем применим их в нескольких моделях

далее.

14.9.6а. ML-оценивание линейной модели со случайными эффектами

Вклад i -го индивида в логарифмическое правдоподобие в модели со случайными эффектами [см. (11-28) – (11-31)] с нормально распределенными шоками есть

$$\begin{aligned} \ln L_i(\beta, \sigma_\varepsilon^2, \sigma_u^2) &= \frac{-1}{2} [T_i \ln 2\pi + \ln |\Omega_i| + (\mathbf{y}_i - \mathbf{X}_i \beta)' \Omega_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \beta)] \\ &= \frac{-1}{2} [T_i \ln 2\pi + \ln |\Omega_i| + \varepsilon_i' \Omega_i^{-1} \varepsilon_i], \end{aligned} \quad (14-89)$$

где

$$\Omega_i = \sigma_\varepsilon^2 \mathbf{I}_{T_i} + \sigma_u^2 \mathbf{i} \mathbf{i}'$$

и \mathbf{i} обозначает $T_i \times 1$ столбец единиц. Обратим внимание, что Ω_i варьирует по i , поскольку она представляет собой $T_i \times T_i$ матрицу. В работе Baltagi (2005, р. 19–20) предлагается удобная и компактная оценка для этой модели, которая требует итераций между оценкой $\phi^2 = [\sigma_\varepsilon^2 / (\sigma_\varepsilon^2 + T \sigma_u^2)]$, основанной на суммах квадратов остатков, и $(\alpha, \beta, \sigma_\varepsilon^2)$ (α — это постоянный член), использующей FGLS. К сожалению, удобство и компактность исчезают в случае несбалансированной панели. Вместо этого мы рассматриваем то, что в указанной работе называется подходом «в лоб», т. е. прямую максимизацию логарифмической функции правдоподобия в (14-89). (См. там же, р. 169–170.)

Используя (А-66), получаем (из (11-28)), что

$$\Omega_i^{-1} = \frac{1}{\sigma_\varepsilon^2} \left[\mathbf{I}_{T_i} - \frac{\sigma_u^2}{\sigma_\varepsilon^2 + T_i \sigma_u^2} \mathbf{i} \mathbf{i}' \right].$$

Нам также потребуется определитель матрицы Ω_i . Чтобы его получить, мы рассмотрим произведение ее характеристических корней. Сперва запишем

$$|\Omega_i| = (\sigma_\varepsilon^2)^{T_i} |\mathbf{I} + \gamma \mathbf{i} \mathbf{i}'|,$$

где $\gamma = \sigma_u^2 / \sigma_\varepsilon^2$. Чтобы найти характеристические корни матрицы, используем определение

$$[\mathbf{I} + \gamma \mathbf{i} \mathbf{i}'] \mathbf{c} = \lambda \mathbf{c},$$

где \mathbf{c} — характеристический вектор, а λ — соответствующий характеристический корень. Из этого уравнения следует, что $\gamma \mathbf{i} \mathbf{i}' \mathbf{c} = (\lambda - 1) \mathbf{c}$. Умножим слева на \mathbf{i}' и получим $\gamma (\mathbf{i}' \mathbf{i}) (\mathbf{i}' \mathbf{c}) = (\lambda - 1) (\mathbf{i}' \mathbf{c})$. Любой вектор \mathbf{c} , состоящий из элементов, которые в сумме дают ноль, будет удовлетворять этому равенству. Существует $T_i - 1$ таких векторов, а соответствующие характеристические корни будут $(\lambda - 1) = 0$, т. е. $\lambda = 1$. Для остающегося корня поделим на ненулевой член $(\mathbf{i}' \mathbf{c})$ и заметим, что $\mathbf{i}' \mathbf{i} = T_i$, поэтому последний корень — это $T_i \gamma = \lambda - 1$, т. е. $\lambda = (1 + T_i \gamma)^{24}$. Следовательно, определитель равен

$$\ln |\Omega_i| = T_i \ln \sigma_\varepsilon^2 + \ln(1 + T_i \gamma).$$

²⁴С помощью этих преобразований мы установили полезный общий результат. Характеристические корни $T \times T$ матрицы вида $\mathbf{A} = (\mathbf{I} + a \mathbf{b} \mathbf{b}')$ равны 1 с кратностью $(T - 1)$ и $a \mathbf{b}' \mathbf{b}$ с кратностью 1. Доказательство проводится точно так же, как и этот вывод.

Раскрывая скобки в третьем члене, мы получаем

$$\begin{aligned} \ln L &= \sum_{i=1}^n \ln L_i = -\frac{1}{2} \left[(\ln 2\pi + \ln \sigma_\varepsilon^2) \sum_{i=1}^n T_i + \sum_{i=1}^n \ln(1 + T_i \gamma) \right] - \\ &- \frac{1}{2\sigma_\varepsilon^2} \sum_{i=1}^n \left[\varepsilon_i' \varepsilon_i - \frac{\sigma_u^2 (T_i \bar{\varepsilon}_i)^2}{\sigma_\varepsilon^2 + T_i \sigma_u^2} \right]. \end{aligned}$$

Обратим внимание, что в третьем члене мы можем записать, что $\sigma_\varepsilon^2 + T_i \sigma_u^2 = \sigma_\varepsilon^2 (1 + T_i \gamma)$ и $\sigma_u^2 = \sigma_\varepsilon^2 \gamma$. После подстановки этого в выражение, σ_ε^2 в квадратных скобках сокращается и остается

$$\ln L = -\frac{1}{2} \sum_{i=1}^n \left(T_i (\ln 2\pi + \ln \sigma_\varepsilon^2) + \ln(1 + T_i \gamma) + \frac{1}{\sigma_\varepsilon^2} \left[\varepsilon_i' \varepsilon_i - \frac{\gamma (T_i \bar{\varepsilon}_i)^2}{1 + T_i \gamma} \right] \right).$$

Пусть теперь $\theta = 1/\sigma_\varepsilon^2$, $R_i = 1 + T_i \gamma$, а $Q_i = \gamma/R_i$. Индивидуальный вклад в логарифмическое правдоподобие становится равным

$$\ln L_i = -\frac{1}{2} [\theta (\varepsilon_i' \varepsilon_i - Q_i (T_i \bar{\varepsilon}_i)^2) + \ln R_i - T_i \ln \theta + T_i \ln 2\pi].$$

Уравнения правдоподобия записываются как

$$\begin{aligned} \frac{\partial \ln L_i}{\partial \beta} &= \theta \left[\sum_{t=1}^{T_i} \mathbf{x}_{it} \varepsilon_{it} \right] - \theta \left[Q_i \left(\sum_{t=1}^{T_i} \mathbf{x}_{it} \right) \left(\sum_{t=1}^{T_i} \varepsilon_{it} \right) \right], \\ \frac{\partial \ln L_i}{\partial \theta} &= -\frac{1}{2} \left[\left(\sum_{t=1}^{T_i} \varepsilon_{it}^2 \right) - Q_i \left(\sum_{t=1}^{T_i} \varepsilon_{it} \right)^2 - \frac{T_i}{\theta} \right], \\ \frac{\partial \ln L_i}{\partial \gamma} &= \frac{1}{2} \left[\theta \left(\frac{1}{R_i^2} \left(\sum_{t=1}^{T_i} \varepsilon_{it} \right)^2 \right) - \frac{T_i}{R_i} \right]. \end{aligned}$$

Этого достаточно для проведения алгоритма оптимизации, такого как DFP или BFGS. (См. раздел E3.3.) Можно также получить вторые производные для вычисления асимптотической ковариационной матрицы, но это является излишним. Для $\hat{\beta}_{\text{MLE}}$ мы знаем, что, поскольку это обобщенная регрессионная модель, соответствующая асимптотическая ковариационная матрица имеет вид

$$\text{Asy. Var}[\hat{\beta}_{\text{MLE}}] = \left[\sum_{i=1}^n \mathbf{x}_i' \hat{\Omega}_i^{-1} \mathbf{x}_i \right]^{-1}.$$

(См. подраздел 11.5.1.) Мы также знаем, что ML-оценки компонент дисперсии будут асимптотически некоррелированными с ML-оценками β . В принципе можно также оценить асимптотические дисперсии ML-оценок σ_ε^2 и σ_u^2 . Для этого необходимо, как всегда в таких случаях, вывести их из оценок θ и γ . Однако статистические выводы о дисперсии шоков σ_ε^2 в регрессионной модели обычно не представляют интереса. С другой стороны, можно было попытаться проверить гипотезу о том, что σ_u^2 равна нулю или $\gamma = 0$.

LM-статистика Бройша–Пэгона (Breusch, Pagan(1979)), задающаяся формулой (11-42), обобщенная на случай несбалансированной панели, рассматриваемой здесь, имеет следующий вид:

$$LM = \frac{\left(\sum_{i=1}^N T_i\right)^2}{\left[2 \sum_{i=1}^N T_i(T_i - 1)\right]} \left[\frac{\sum_{i=1}^N (T_i \bar{e}_i)^2}{\sum_{i=1}^N \sum_{t=1}^{T_i} e_{it}^2} - 1 \right]^2$$

$$= \frac{\left(\sum_{i=1}^N T_i\right)^2}{\left[2 \sum_{i=1}^N T_i(T_i - 1)\right]} \left[\frac{\sum_{i=1}^N [(T_i \bar{e}_i)^2 - \mathbf{e}'_i \mathbf{e}_i]}{\sum_{i=1}^N \mathbf{e}'_i \mathbf{e}_i} \right]^2$$

Пример 14.11. *Оценки максимального правдоподобия и FGLS-оценки для уравнения заработной платы*

Таблица 14.9. Оценки уравнения заработной платы

| Переменная | Суммарный метод наименьших квадратов | | Случайные эффекты MLE | | Случайные эффекты FGLS | |
|------------------|--------------------------------------|----------------------------|----------------------------------|---------------|------------------------|---------------|
| | Оценка | Станд. ошибка ^a | Оценка | Станд. ошибка | Оценка | Станд. ошибка |
| Exp | 0,0361 | 0,004533 | 0,1078 | 0,002480 | 0,08906 | 0,002280 |
| Exp ² | -0,000655 | 0,0001016 | -0,0005054 | 0,00005452 | -0,0007577 | 0,00005036 |
| Wks | | 0,001728 | | 0,0006031 | | 0,0005939 |
| | 0,004461 | | 0,0008663 | | 0,001066 | |
| Occ | -0,3176 | 0,02726 | -0,03954 | 0,01374 | -0,1067 | 0,01269 |
| Ind | 0,03213 | 0,02526 | | 0,01531 | -0,01637 | 0,01391 |
| | | | 0,008807 | | | |
| South | -0,1137 | 0,02868 | -0,01615 | 0,03201 | -0,06899 | 0,02354 |
| SMSA | 0,1586 | 0,02602 | -0,04019 | 0,01901 | -0,01530 | 0,01649 |
| MS | 0,3203 | 0,03494 | -0,03540 | 0,01880 | -0,02398 | 0,01711 |
| Union | 0,06975 | 0,02667 | 0,03306 | 0,01482 | 0,03597 | 0,01367 |
| Постоянная | 5,8802 | 0,09673 | 4,8197 | 0,06035 | 5,3455 | 0,04361 |
| σ_e^2 | 0,146119 | | 0,023436 ($\theta = 42,66926$) | | 0,023102 | |
| σ_u^2 | 0 | | 0,876517 ($\gamma = 37,40035$) | | 0,12301719 | |
| ln L | -1899,537 | | 249,25 | | - | |

^a Оценка $\sigma_u = 0,9542921$.

В примерах 11.5 и 11.6 представлены FGLS-оценки для уравнения заработной платы при использовании панельных данных работы Cornwell, Rupert (1988). Мы оценим это уравнение заново, используя максимальное правдоподобие вместо FGLS-оценивания. Оценки параметров представлены в табл. 14.9, включая также FGLS-оценки и объединенные OLS-оценки (OLS-оценки модели пула). Также приводятся оценки дисперсии.

Ввиду большого объема выборки мы ожидаем увидеть схожесть ML- и FGLS-оценок. LM-статистика для проверки наличия общих эффектов составляет 3881,34, что значительно больше критического значения 3,84. Для MLE возможно также использовать LR-критерий, чтобы проверить гипотезу о наличии случайных эффектов против нулевой гипотезы об их отсутствии. Статистика хи-квадрат, основанная на двух логарифмических правдоподобиях, равна 4297,57, что приводит к такому же выводу.

14.9.6b. Вложенные случайные эффекты

Рассмотрим набор данных по результатам экзаменов для разных школьных округов одного штата. Чтобы установить систему обозначений для этой сложной модели, определим четырехуровневую несбалансированную структуру следующего вида:

- Z_{ijkt} = оценка за экзамен для учащегося t , учителя k , школы j , округа i ,
- L = школьные округа, $i = 1, \dots, L$,
- M_i = школы в каждом округе, $j = 1, \dots, M_i$,
- N_{ij} = учителя в каждой школе, $k = 1, \dots, N_{ij}$,
- T_{ijk} = учащиеся в каждом классе, $t = 1, \dots, T_{ijk}$.

Таким образом, в данных обозначениях мы допускаем, что модель может быть несбалансированной на всех уровнях. В общем виде регрессионная модель со случайными эффектами может иметь вид

$$y_{ijkt} = \mathbf{x}'_{ijkt} \boldsymbol{\beta} + u_{ijk} + v_{ij} + w_i + \varepsilon_{ijkt}.$$

Строгая экзогенность всех регрессоров предполагается на всех уровнях модели. Мы также предполагаем, что все составляющие шока не коррелированы друг с другом. (В дальнейшем мы также добавим предположение о нормальности.) Из структуры шоков ясно, что полная ковариационная матрица Ω является блочно-диагональной по i , причем каждый ее блок, в свою очередь, блочно-диагональный по j , а каждый из составляющих блоков блочно-диагональный по k и на самом нижнем уровне блоки (в нашем примере это класс) имеют вид модели со случайными эффектами, которая рассматривалась ранее.

Обобщенный метод наименьших квадратов был хорошо разработан для сбалансированного случая. [См., например, работу Baltagi, Song, Jung (2001), в которой также есть результаты для трехуровневого несбалансированного случая.] Определим следующие величины, получающиеся из компонент дисперсии, σ_ε^2 , σ_u^2 , σ_v^2 и σ_w^2 :

$$\begin{aligned} \sigma_1^2 &= T\sigma_u^2 + \sigma_\varepsilon^2, \\ \sigma_2^2 &= NT\sigma_v^2 + T\sigma_u^2 + \sigma_\varepsilon^2 = \sigma_1^2 + NT\sigma_v^2, \\ \sigma_3^2 &= MNT\sigma_w^2 + NT\sigma_v^2 + T\sigma_u^2 + \sigma_\varepsilon^2 = \sigma_2^2 + MNT\sigma_w^2. \end{aligned}$$

Тогда полный обобщенный метод наименьших квадратов будет эквивалентен OLS-оцениванию регрессии

$$\tilde{y}_{ijkt} = y_{ijkt} - \left(1 - \frac{\sigma_\varepsilon}{\sigma_1}\right) \bar{y}_{ijk\cdot} - \left(\frac{\sigma_\varepsilon}{\sigma_1} - \frac{\sigma_\varepsilon}{\sigma_2}\right) \bar{y}_{ij\cdot\cdot} - \left(\frac{\sigma_\varepsilon}{\sigma_2} - \frac{\sigma_\varepsilon}{\sigma_3}\right) \bar{y}_i \dots$$

на таким же образом преобразованные x_{ijkt} . FGLS-оценки получаются, если комбинировать три межгрупповые between-оценки и within-оценку для самого последнего уровня модели.

Можно вывести аналогичную оценку для несбалансированного случая [см. работу Балтаги и др. (Baltagi et al. (2001))], но степень сложности этого метода резко возрастает. Однако, как показано в работе Антвилера (Antwiler (2001)), если можно предположить нормальность распределений, то логарифмическое правдоподобие очень легко поддается обработке. (Отмечаем пересечение практичности с неустойчивостью.) Определим следующие дисперсионные отношения:

$$\rho_u = \frac{\sigma_u^2}{\sigma_\varepsilon^2}, \rho_v = \frac{\sigma_v^2}{\sigma_\varepsilon^2}, \rho_w = \frac{\sigma_w^2}{\sigma_\varepsilon^2}.$$

Построим следующие промежуточные результаты:

$$\theta_{ijk} = 1 + T_{ijk} \rho_u, \phi_{ij} = \sum_{k=1}^{N_{ij}} \frac{T_{ijk}}{\theta_{ijk}}, \theta_{ij} = 1 + \phi_{ij} \rho_v, \phi_i = \sum_{j=1}^{M_i} \frac{\phi_{ij}}{\theta_{ij}}, \theta_i = 1 + \rho_w \phi_i$$

и суммы квадратов шоков $e_{ijkt} = y_{ijkt} - \mathbf{x}'_{ijkt} \boldsymbol{\beta}$,

$$A_{ijk} = \sum_{t=1}^{T_{ijk}} e_{ijkt}^2, \\ B_{ijk} = \sum_{t=1}^{T_{ijk}} e_{ijkt}, B_{ij} = \sum_{k=1}^{N_{ij}} \frac{B_{ijk}}{\theta_{ijk}}, B_i = \sum_{j=1}^{M_i} \frac{B_{ij}}{\theta_{ij}}.$$

Логарифмическое правдоподобие записывается как

$$\ln L = -\frac{1}{2} H \ln(2\pi\sigma_\varepsilon^2) - \frac{1}{2} \left[\sum_{i=1}^L \left\{ \ln \theta_i + \sum_{j=1}^{M_i} \left\{ \ln \theta_{ij} + \sum_{k=1}^{N_{ij}} \left\{ \ln \theta_{ijk} + \frac{A_{ijk}}{\sigma_\varepsilon^2} - \frac{\rho_u B_{ijk}^2}{\theta_{ijk} \sigma_\varepsilon^2} \right\} - \frac{\rho_v B_{ij}^2}{\theta_{ij} \sigma_\varepsilon^2} \right\} - \frac{\rho_w B_i^2}{\theta_i \sigma_\varepsilon^2} \right\} \right],$$

где H — общее количество наблюдений. (Для трехуровневой модели $L = 1$ и $\rho_w = 0$.) В работе Antwiler (2001) выводятся первые производные логарифмической функции правдоподобия, необходимые, чтобы максимизировать $\ln L$. Тем не менее в работе отмечается, что сложность результатов делает заманчивым численное дифференцирование. С другой стороны, в работе говорится, что нахождение вторых производных функции очень сложно и

автор прибегает к числовым вторым производным в своей модели. Сложная часть гесса состоит в смешанных производных по β и параметрах дисперсии, а также в нижней правой части для самих параметров дисперсии. Тем не менее они не являются необходимыми. Как и в любой обобщенной регрессионной модели, оценки дисперсии и оценки коэффициентов являются асимптотически некоррелированными. Таким образом, нужно обращать только часть матрицы, относящуюся к β , чтобы получить необходимую асимптотическую ковариационную матрицу. Соответствующий блок матрицы представляет собой

$$\begin{aligned}
 -\frac{\partial^2 \ln L}{\partial \beta \partial \beta'} &= \frac{1}{\sigma_\varepsilon^2} \sum_{i=1}^L \sum_{j=1}^{M_i} \sum_{k=1}^{N_{ij}} \sum_{t=1}^{T_{ijk}} \mathbf{x}_{ijkt} \mathbf{x}'_{ijkt} - \frac{\rho_w}{\sigma_\varepsilon^2} \sum_{i=1}^L \sum_{j=1}^{M_i} \sum_{k=1}^{N_{ij}} \frac{1}{\theta_{ijk}} \left(\sum_{t=1}^{T_{ijk}} \mathbf{x}_{ijkt} \right) \times \\
 &\quad \times \left(\sum_{t=1}^{T_{ijk}} \mathbf{x}'_{ijkt} \right) - \frac{\rho_v}{\sigma_\varepsilon^2} \sum_{i=1}^L \sum_{j=1}^{M_i} \frac{1}{\theta_{ij}} \left(\sum_{k=1}^{N_{ij}} \frac{1}{\theta_{ijk}} \left(\sum_{t=1}^{T_{ijk}} \mathbf{x}_{ijkt} \right) \right) \times \\
 &\quad \times \left(\sum_{k=1}^{N_{ij}} \frac{1}{\theta_{ijk}} \left(\sum_{t=1}^{T_{ijk}} \mathbf{x}'_{ijkt} \right) \right) - \frac{\rho_u}{\sigma_\varepsilon^2} \sum_{i=1}^L \left(\sum_{j=1}^{M_i} \frac{1}{\theta_{ij}} \left(\sum_{k=1}^{N_{ij}} \frac{1}{\theta_{ijk}} \left(\sum_{t=1}^{T_{ijk}} \mathbf{x}_{ijkt} \right) \right) \right) \times \\
 &\quad \times \left(\sum_{j=1}^{M_i} \frac{1}{\theta_{ij}} \left(\sum_{k=1}^{N_{ij}} \frac{1}{\theta_{ijk}} \left(\sum_{t=1}^{T_{ijk}} \mathbf{x}'_{ijkt} \right) \right) \right).
 \end{aligned}$$

Оценка максимального правдоподобия параметра β есть FGLS, основанный на оценках максимального правдоподобия параметров дисперсии. Таким образом, выражение (14-90) дает подходящую ковариационную матрицу для GLS или оценки максимального правдоподобия. Разница заключается лишь в том, как рассчитаны компоненты дисперсии. В работе Baltagi et al. (2001) предложено множество методов для анализа трехуровневой модели. Для случаев с более чем тремя уровнями MLE становится более привлекательной.

Учитывая сложность результатов, можно было бы просто использовать OLS-оценку, несмотря на ее неэффективность. Как и следует ожидать, стандартные ошибки будут смещены по причине корреляции между наблюдениями; есть основания полагать, что смещение имеет место в сторону уменьшения. [См. Moulton (1986).] В этом случае робастная оценка, задаваемая (11-4), будет естественной заменой. В приведенном ранее примере вложенная структура была очевидной. В других случаях, таких как модель в примере 11.12, это может оказаться неверным. В примере 14.12 [и в модели в работе Балтаги (Baltagi (2005))] наблюдения по штату были сгруппированы по регионам интуитивно. Влияние неверной группировки на оценивание модели представляет собой открытый вопрос. И OLS, и FGLS останутся состоятельными: обе они эквивалентны GLS-оценке, взятой с ошибочными весами, как мы видели ранее. Тем не менее влияние на асимптотическую ковариационную матрицу оценки требует дальнейшего исследования.

Пример 14.12. Производительность по штатам

В работе Манелла (Munnell (1990)) была проанализирована производительность общественного капитала на уровне штата с использованием производственной функции Кобба–Дугласа. Мы будем использовать данные этого исследования, чтобы оценить трехуровневую модель логлинейной регрессии

$$\ln gsp_{jkt} = \alpha + \beta_1 \ln pc_{jkt} + \beta_2 \ln hwy_{jkt} + \beta_3 \ln water_{jkt} + \beta_4 \ln util_{jkt} + \beta_5 \ln emp_{jkt} + \beta_6 unemp_{jkt} + \varepsilon_{jkt} + u_{jk} + v_j, \\ j = 1, \dots, 9; t = 1, \dots, 17, k = 1, \dots, N_j,$$

где переменными в модели являются:

- gsp = валовой продукт штата (grossstate product),
- pc = общественный капитал (public capital) = $hwy + water + util$,
- hwy = дорожный фонд (highway capital),
- $water$ = фонд коммунального водоснабжения (water utility capital),
- $util$ = коммунальный фонд (utility capital),
- pc = частный капитал (private capital),
- emp = занятость (трудовые ресурсы) (employment (labor)),
- $unemp$ = уровень безработицы (unemployment rate),

и определим $M = 9$ регионов, каждый из которых состоит из группы, включающей 48 континентальных штатов:

| | |
|--|-------------------------------|
| Залив (<i>Gulf</i>) | = AL, FL, LA, MS, |
| Средний Запад (<i>Midwest</i>) | = IL, IN, KY, MI, MN, OH, WI, |
| Средняя Атлантика (<i>MidAtlantic</i>) | = DE, MD, NJ, NY, PA, VA, |
| Горный (<i>Mountain</i>) | = CO, ID, MT, ND, SD, WY, |
| Новая Англия (<i>New England</i>) | = CT, ME, MA, NH, RI, VT, |
| Юг (<i>South</i>) | = GA, NC, SC, TN, WV, |
| Юго-Запад (<i>Southwest</i>) | = AZ, NV, NM, TX, UT, |
| Аллея Торнадо (<i>Tornado Alley</i>) | = AR, IA, KS, MO, NE, OK, |
| Западное Побережье (<i>West Coast</i>) | = CA, OR, WA. |

По каждому штату у нас есть данные за 17 лет — с 1970 по 1986 г.²⁵ Двух- и трехуровневые модели со случайными эффектами были оценены методом максимального правдоподобия. Двухуровневая модель была также оценена при помощи FGLS с использованием методов, рассмотренных в разделе 11.5.3.

Таблица 14.10. Оценка производственной функции по штатам

| | OLS | | Постоянные эффекты | Случайные эффекты FGLS | Случайные эффекты ML | Вложенные случайные эффекты |
|----------------------|----------|----------------------------|-------------------------|------------------------------|----------------------------|-----------------------------------|
| | Оценка | Станд. ош. ^a | Оценка | Оценка | Оценка | Оценка |
| | | | (Станд. ош.) | (Станд. ош.) | (Станд. ош.) | (Станд. ош.) |
| α | 1,9260 | 0,05250 (0,2143) | | 2,1608 (0,1380) | 2,1759 (0,1477) | 2,1348 (0,1514) |
| β_1 | 0,3120 | 0,01109 (0,04678) | 0,2350 (0,02621) | 0,2755 (0,01972) | 0,2703 (0,02110) | 0,2724 (0,02141) |
| β_2 | 0,05888 | 0,01541 (0,05078) | 0,07675 (0,03124) | 0,06167 (0,02168) | 0,06268 (0,02269) | 0,06645 (0,02287) |
| β_3 | 0,1186 | 0,01236 (0,03450) | 0,0786 (0,0150) | 0,07572 (0,01381) | 0,07545 (0,01397) | 0,07392 (0,01399) |
| β_4 | 0,00856 | 0,01235 (0,04062) | -0,11478 (0,01814) | -0,09672 (0,01683) | -0,1004 (0,01730) | -0,1004 (0,01698) |
| β_5 | 0,5497 | 0,01554 (0,06770) | 0,8011 (0,02976) | 0,7450 (0,02482) | 0,7542 (0,02664) | 0,7539 (0,02613) |
| β_6 | -0,00727 | 0,001384 (0,002946) | -0,005179 (0,000980) | -0,005963 (0,0008814) | -0,005809 (0,0009014) | -0,005878 (0,0009002) |
| σ_ε | 0,085422 | | 0,03676493 | 0,0367649 | 0,0366974 | 0,0366964 |
| σ_u | | | | 0,0771064 | 0,0875682 | 0,0791243 |
| σ_v | | | | | | 0,0386299 |
| $\ln L$ | 853,1372 | | 1565,501 | | 1429,075 | 1430,30576 |

^a В скобках устойчивые (кластерные) стандартные ошибки. Ковариационная матрица скорректирована на число степеней свободы.

В табл. 14.10 представлены оценки производственной функции при использовании OLS для модели пула, OLS для модели с фиксированными эффектами, а также обеих оценок FGLS и оценки максимального правдоподобия для модели со случайными эффектами. В целом оценки достаточно схожи, хотя OLS-оценки несколько различаются. Это свидетельствует, как можно было бы подозревать с самого начала, что в объединенной модели существуют упущенные эффекты. F -статистика для проверки значимости постоянных эффектов составляет 76,712 с 47 и 762 степенями свободы. Критическое значение из таблицы равно 1,379, поэтому исходя из этого надо отвергнуть гипотезу об отсутствии общих эффектов. Также обратим внимание на очень большие различия между обычными OLS стандартными ошибками и устойчивыми (кластерными) исправленными значениями. Различия в три-четыре раза убедительно показывают, что существуют скрытые эффекты, по крайней мере, на уровне штата. Остается понять, какой подход предпочтительнее: фиксированные или случайные эффекты. Тест Хаусмана для сравнения постоянных и случайных эффектов дает значение хи-квадрат, равное 18,987. Критическое значение составляет 12,592. Отсюда следует, что модель

с фиксированными эффектами является более верной спецификацией. Если повторить тест Хаусмана при использовании трехуровневых оценок (последний столбец табл. 14.10), то статистика несколько уменьшается и становится равной 15,327. Наконец, обратим внимание на сходство всех трех наборов оценок в модели со случайными эффектами. В самом деле при гипотезе о независимости средних все три являются состоятельными оценками. На этой стадии возникает желание выполнить тест отношения правдоподобий для проверки гипотезы о двухуровневой модели против более охватывающей альтернативной трехуровневой модели. Тестовая статистика будет составлять двойную разницу логарифмических правдоподобий, что равно 2,46. Для одной степени свободы критический хи-квадрат с одной степенью свободы равен 3,84, поэтому исходя из этого мы не отвергаем гипотезу о двухуровневой модели. Тем не менее мы обращаем внимание, что с этим методом тестирования связана одна проблема. Гипотеза о том, что дисперсия равна нулю, для критерия отношения правдоподобия не является корректно заданной: параметр при нулевой гипотезе находится на границе пространства допустимых параметров ($\sigma_v^2 \geq 0$). В данном случае мы не можем применять знакомую нам теорию о распределении этой статистики.

14.9.6с. Случайные эффекты в нелинейных моделях: MLE, использующая квадратуру

В подразделе 14.9.5b описывается нелинейная модель для панельных данных: геометрическая регрессионная модель

$$\begin{aligned} \text{Prob}[Y_{it} = y_{it} | \mathbf{x}_{it}] &= \theta_{it}(1 - \theta_{it})^{y_{it}}, y_{it} = 0, 1, \dots; i = 1, \dots, n, t = 1, \dots, T_i, \\ \theta_{it} &= 1/(1 + \lambda_{it}), \lambda_{it} = \exp(\mathbf{x}'_{it}\beta). \end{aligned}$$

Как отмечено, это модель панельных данных, хотя в таком виде у нее нет ни одного из тех свойств, которые мы использовали для панельных данных в линейном случае. Это регрессионная модель

$$E[y_{it} | \mathbf{x}_{it}] = \lambda_{it},$$

которая означает, что

$$y_{it} = \lambda_{it} + \varepsilon_{it}.$$

Это просто тавтология, которая определяет отклонение y_{it} от ее условного среднего. Здесь могло бы показаться естественным ввести общий фиксированный или случайный эффект, как мы делали ранее в линейном случае:

$$y_{it} = \lambda_{it} + \varepsilon_{it} + c_i.$$

Однако проблема этой спецификации состоит в том, что, в то время как ε_{it} определяется как разница между y_{it} и ее средним, c_i — это свободно варьирующая случайная переменная. Без очень сложных ограничений на то,

²⁵ Данные были загружены с веб-сайта книги Baltagi (2005) по адресу <http://www.wiley.com/legacy/wileychi/baltagi3e/>. См. табл. F10.1 приложения.

как может варьировать c_i , описанная модель не может гарантировать, что y_{it} не будет отрицательным. При определении спецификации нелинейной модели необходимо соблюдать повышенную степень осторожности, чтобы сохранить внутреннюю непротиворечивость модели. Часто встречающийся подход в **моделях индексных функций** (index function models), подобных этой, заключается в том, чтобы ввести общий эффект в функцию условного среднего. Например, геометрическая регрессионная модель со случайными эффектами будет иметь вид

$$\text{Prob}[Y_{it} = y_{it} | \mathbf{x}_{it}] = \theta_{it}(1 - \theta_{it})^{y_{it}}, y_{it} = 0, 1, \dots; i = 1, \dots, n, t = 1, \dots, T_i,$$

$$\theta_{it} = 1/(1 + \lambda_{it}), \lambda_{it} = \exp(\mathbf{x}'_{it}\boldsymbol{\beta} + u_i),$$

$f(u_i)$ = спецификация распределения случайных эффектов по индивидам.

При помощи этой спецификации модель задается следующим образом:

$$\text{Prob}[Y_{it} = y_{it} | \mathbf{x}_{it}, u_i] = \theta_{it}(1 - \theta_{it})^{y_{it}}.$$

Таким образом, наше утверждение о вероятности теперь обусловлено как данными наблюдений, так и ненаблюдаемыми случайными эффектами. Тогда случайный общий эффект может свободно варьировать и внутренние характеристики модели могут сохраняться непротиворечивыми.

Возникают два вопроса:

- Как получить оценки методом максимального правдоподобия для параметров модели? Мы будем исследовать этот вопрос сейчас.
- Если мы пренебрежем индивидуальной гетерогенностью и просто оценим объединенную модель, получим ли мы состоятельные оценки для параметров модели? Ответ на этот вопрос следующий: иногда да, но, как правило, нет. Благоприятные случаи — это простые логлинейные модели, такие как геометрическая и пуассоновская модели, которые мы рассматривали в этой главе. Неблагоприятные случаи — это большинство других моделей, часто используемых в литературе, включая в особенности модели бинарного выбора, цензурированные регрессии, модели выборочной селективности и, более обще, нелинейные модели, у которых нет простых экспоненциальных средних. [Обратим внимание, что это ключевой момент в изучении робастного оценивания ковариационной матрицы в подразделах 14.8.3 и 14.8.4. См. также работу Freedman (2006).]

Теперь мы рассмотрим оценку методом максимального правдоподобия для нелинейной модели со случайными эффектами. Чтобы установить методологию, используемую далее, мы будем рассматривать общую модель, а затем вернемся к частному случаю геометрической регрессионной модели из примера 14.10. Предположим, что модель панельных данных определяет распределение вероятностей случайной переменной y_{it} условно по вектору данных \mathbf{x}_{it} и ненаблюдаемым общим случайным эффектам u_i . Как всегда, у нас есть T_i наблюдений в группе и предполагается, что данные \mathbf{x}_{it} , а теперь и u_i определяются строго экзогенно. Тогда наша модель для одного индивида выражается как

$$p(y_{it} | \mathbf{x}_{it}, u_i) = f(y_{it} | \mathbf{x}_{it}, u_i, \boldsymbol{\theta}),$$

где $p(y_{it} | \mathbf{x}_{it}, u_i)$ показывает, что мы определяем плотность условного распределения, а $f(y_{it} | \mathbf{x}_{it}, u_i, \theta)$ определяет функциональную форму и подчеркивает, что требуется оценивать вектор параметров. Мы также предположим, что, за исключением общего u_i , наблюдения внутри группы являются независимыми: зависимость наблюдений в группе будет в любом случае ввиду общего u_i . Плотность совместного распределения T_i наблюдений y_{it} при заданном u_i при этих предположениях будет равна

$$p(y_{i1}, y_{i2}, \dots, y_{iT_i} | \mathbf{X}_i, u_i) = \prod_{t=1}^{T_i} f(y_{it} | \mathbf{x}_{it}, u_i, \theta),$$

поскольку условно по u_i наблюдения независимы. Но так как u_i являются частью наблюдения над группой, то, чтобы построить логарифмическое правдоподобие, нам потребуется

$$p(y_{i1}, y_{i2}, \dots, y_{iT_i}, u_i | \mathbf{X}_i) = \left[\prod_{t=1}^{T_i} f(y_{it} | \mathbf{x}_{it}, u_i, \theta) \right] f(u_i).$$

Функция правдоподобия есть плотность совместного распределения наблюдаемых случайных переменных. Поскольку u_i является ненаблюдаемым случайным эффектом, то для построения функции правдоподобия мы должны проинтегрировать по ней плотность совместного распределения. Таким образом,

$$p(y_{i1}, y_{i2}, \dots, y_{iT_i} | \mathbf{X}_i) = \int_{u_i} \left[\prod_{t=1}^{T_i} f(y_{it} | \mathbf{x}_{it}, u_i, \theta) \right] f(u_i) du_i.$$

Тогда вклад группы i в логарифмическую функцию правдоподобия может быть записан как

$$\ln L_i = \ln \int_{u_i} \left[\prod_{t=1}^{T_i} f(y_{it} | \mathbf{x}_{it}, u_i, \theta) \right] f(u_i) du_i.$$

Существует две практические задачи, которые надо решить, чтобы воспользоваться этой оценкой. Во-первых, редко когда интеграл можно вычислить в явном виде. (Это так, когда плотность y_{it} является нормальной с линейным условным средним и случайный эффект является нормальным, потому что, как мы видели, модель превращается в линейную модель со случайными эффектами.) Таким образом, возникающая на практике сложность состоит в том, как вычислить эти интегралы. Во-вторых, остается выяснить распределение u_i , по которому проводится интегрирование. Распределение общего эффекта представляет собой часть спецификации модели. В литературе встречается несколько подходов к этой модели. Подход, который мы используем здесь, обобщает модель со случайными эффектами с моделью с нормально распределенными эффектами, которая рассматривалась в предыдущем разделе. Эта техника носит название **метод Батлера и Моффита** (Butler, Moffitt (1982)). Первоначально она была предложена для обобщения модели со случайными эффектами на случай бинарных моделей (см. главу

17), но, как мы сейчас убедимся, она легко обобщается и на широкий диапазон других моделей. При вычислениях мы используем методику для аппроксимации интегралов, известную как **квадратура Гаусса–Эрмита** (Gauss–Hermite quadrature).

Предположим, что u_i распределены нормально со средним ноль и дисперсией σ_u^2 . Таким образом,

$$f(u_i) = \frac{1}{\sqrt{2\pi\sigma_u^2}} \exp\left(-\frac{u_i^2}{2\sigma_u^2}\right).$$

При этом предположении i -й элемент в логарифмическом правдоподобии есть

$$\ln L_i = \ln \int_{-\infty}^{\infty} \left[\prod_{t=1}^{T_i} f(y_{it} | \mathbf{x}_{it}, u_i, \boldsymbol{\theta}) \right] \frac{1}{\sqrt{2\pi\sigma_u^2}} \exp\left(-\frac{u_i^2}{2\sigma_u^2}\right) du_i.$$

Для того чтобы привести эту функцию к удобному нам виду, положим $w_i = u_i/(\sigma_u\sqrt{2})$, так что $u_i = \sigma_u\sqrt{2}w_i = \phi w_i$, а якобиан преобразования из u_i к w_i есть $du_i = \phi dw_i$. Теперь мы сделаем замену переменной в интеграле и получим функцию

$$\ln L_i = \ln \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \left[\prod_{t=1}^{T_i} f(y_{it} | \mathbf{x}_{it}, \phi w_i, \boldsymbol{\theta}) \right] \exp(-w_i^2) dw_i.$$

Пусть

$$g(w_i) = \prod_{t=1}^{T_i} f(y_{it} | \mathbf{x}_{it}, \phi w_i, \boldsymbol{\theta}).$$

Тогда преобразуемая нами функция выражается как

$$\ln L_i = \ln \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} g(w_i) \exp(-w_i^2) dw_i.$$

Целью всех этих преобразований является то, что интегралы такого вида могут вычисляться очень точно при помощи квадратуры Гаусса–Эрмита. Квадратура Гаусса–Эрмита заменяет интегрирование взвешенной суммой функций, вычисленных в специальном множестве точек. Для общего случая это выглядит так:

$$\int_{-\infty}^{\infty} g(w_i) \exp(-w_i^2) dw_i \approx \sum_{h=1}^H z_h g(v_h),$$

где z_h — веса и v_h — узловые точки. Таблицы весов и узловых точек можно найти в широко распространенных источниках, таких как справочник Abramovitz, Stegun (1971). Например, узловые точки и веса для четырехточечной квадратуры составляют

$$\begin{aligned} v_h &= \pm 0,52464762327529002 \text{ и } \pm 1,6506801238857849, \\ z_h &= 0,80491409000549996 \text{ и } 0,081312835447250001. \end{aligned}$$

На практике обычно используется восемь и более точек (на практике не более 96). Таким образом, мы получаем аппроксимацию вклада в логарифмическое правдоподобие, равную

$$\ln L_i = \ln \frac{1}{\sqrt{\pi}} \sum_{h=1}^H z_h \left[\prod_{t=1}^{T_i} f(y_{it} | \mathbf{x}_{it}, \phi v_h, \boldsymbol{\theta}) \right].$$

Приближение Эрмита к логарифмической функции правдоподобия равняется

$$\ln L = \frac{1}{\sqrt{\pi}} \sum_{i=1}^n \ln \sum_{h=1}^H z_h \left[\prod_{t=1}^{T_i} f(y_{it} | \mathbf{x}_{it}, \phi v_h, \boldsymbol{\theta}) \right]. \quad (14-90)$$

Теперь требуется максимизировать эту функцию по $\boldsymbol{\theta}$ и ϕ . Максимизация представляет собой сложную проблему. Тем не менее она была автоматизирована в современном программном обеспечении для некоторых моделей, особенно упомянутых ранее моделей двоичного выбора, и фактически этот алгоритм несложно переделать для многих других моделей. Первые и вторые производные логарифмической функции правдоподобия являются соответственно сложными, но их все еще можно вычислить с помощью квадратур. Оценка σ_u и соответствующая стандартная ошибка получаются из $\hat{\phi}$, используя $\phi = \sigma_u \sqrt{2}$. Гипотеза об отсутствии корреляции между периодами может быть в принципе проверена с использованием любого из трех стандартных тестов.

Пример 14.13. Геометрическая регрессионная модель со случайными эффектами

Используем описанную теорию, чтобы построить модель со случайными эффектами для количественной переменной *DocVis*, рассматриваемой в примере 14.10. Используя (14-90), аппроксимированная логарифмическая функция правдоподобия будет равна

$$\ln L_H = \frac{1}{\sqrt{\pi}} \sum_{i=1}^n \ln \sum_{h=1}^H z_h \left[\prod_{t=1}^{T_i} \theta_{it} (1 - \theta_{it})^{y_{it}} \right],$$

$$\theta_{it} = 1 / (1 + \lambda_{it}), \lambda_{it} = \exp(\mathbf{x}'_{it} \boldsymbol{\beta} + \phi v_h).$$

Производные логарифмического правдоподобия также аппроксимированы. Следующая формула представляет собой общий результат, доказательство оставлено в качестве упражнения:

$$\frac{\partial \log L}{\partial \begin{pmatrix} \boldsymbol{\beta} \\ \phi \end{pmatrix}} = \sum_{i=1}^n \frac{1}{L_i} \frac{\partial L_i}{\partial \begin{pmatrix} \boldsymbol{\beta} \\ \phi \end{pmatrix}} \approx$$

$$\frac{\sum_{i=1}^n \left\{ \frac{1}{\sqrt{\pi}} \sum_{h=1}^H z_h \left[\prod_{t=1}^{T_i} f(y_{it} | \mathbf{x}_{it}, \phi v_h, \boldsymbol{\beta}) \right] \left[\sum_{t=1}^{T_i} \frac{\partial \log f(y_{it} | \mathbf{x}_{it}, \phi v_h, \boldsymbol{\beta})}{\partial \begin{pmatrix} \boldsymbol{\beta} \\ \phi \end{pmatrix}} \right] \right\}}{\left\{ \frac{1}{\sqrt{\pi}} \sum_{h=1}^H z_h \left[\prod_{t=1}^{T_i} f(y_{it} | \mathbf{x}_{it}, \phi v_h, \boldsymbol{\beta}) \right] \right\}}$$

Остается только применить это к нашей геометрической регрессионной модели. Для этого случая плотность была выписана ранее. Неуставленные компоненты рассматриваемых производных — это частные производные по β и ϕ , которые были получены в подразделе 14.9.5. Таким образом, необходимый нам результат имеет вид

$$\frac{\partial \ln f(y_{it} | \mathbf{x}_{it}, \phi v_h, \beta)}{\partial \begin{pmatrix} \beta \\ \phi \end{pmatrix}} = [\theta_{it}(1 + y_{it}) - 1] \begin{pmatrix} \mathbf{x}_{it} \\ v_h \end{pmatrix}.$$

Оценки методом максимального правдоподобия для параметров геометрической регрессионной модели со случайными эффектами даны в примере 14.13 вместе с оценками для модели с фиксированными эффектами.

14.9.6d. Фиксированные эффекты в нелинейных моделях: полная ML-оценка

Используя ту же схему построения модели, что и в предыдущем разделе, определим модель с постоянными эффектами как модель индексной функции со специфическим по группе постоянным членом. Как и прежде, «модель» есть предполагаемая плотность для случайной переменной

$$p(y_{it} | d_{it}, \mathbf{x}_{it}) = f(y_{it} | \alpha_i d_{it} + \mathbf{x}'_{it} \beta),$$

где d_{it} — фиктивная переменная, которая принимает значение один в каждом периоде для индивида i и ноль в противном случае. (В более сложных моделях, таких как цензурированная регрессионная модель, разбираемая в главе 19, могли бы быть другие параметры, такие как дисперсия. Сейчас нам удобно ими пренебречь: рассуждение может быть обобщено, чтобы учесть их, позднее.) Для удобства мы переопределили \mathbf{x}_{it} как непостоянные переменные в модели²⁶. К оцениваемым параметрам относятся K элементов β и n индивидуальных постоянных. Логарифмическая функция правдоподобия для модели с постоянными эффектами может быть записана как

$$\ln L = \sum_{i=1}^n \sum_{t=1}^{T_i} \ln f(y_{it} | \alpha_i + \mathbf{x}'_{it} \beta),$$

где $f(\cdot)$ — плотность вероятности наблюдаемого результата, например геометрической регрессионной модели, использованной в предыдущем примере. Нам будет удобно положить $z_{it} = \alpha_i + \mathbf{x}'_{it} \beta$, так что $p(y_{it} | d_{it}, \mathbf{x}_{it}) = f(y_{it} | z_{it})$.

²⁶При оценивании линейной регрессионной модели с фиксированными эффектами в разделе 11.4 мы получили, что невозможно анализировать модели с не зависящими от времени переменными. Такое же ограничение получается в нелинейном случае по тем же, в сущности, причинам. Не зависящие от времени эффекты поглощаются в постоянном члене. В оценивании столбцы матрицы данных с не зависящими от времени переменными будут преобразованы в столбцы нулей, когда мы будем вычислять производные логарифмической функции правдоподобия.

В случае линейной регрессии с постоянными эффектами мы получили, что оценивание параметров оказалось возможным благодаря преобразованию данных к отклонениям от внутригрупповых средних, что уничтожило специфические для каждого индивида постоянные из уравнения. (См. подраздел 11.4.1.) В нескольких случаях в нелинейных моделях можно также исключить постоянные эффекты из функции правдоподобия, хотя в общем случае не с помощью использования отклонений от средних. Примером тому может служить **экспоненциальная регрессионная модель** (exponential regression model), которая используется для сроков службы радиодеталей и электрооборудования, такого как электрические лампочки:

$$f(y_{it} | \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta}) = \theta_{it} \exp(-\theta_{it}y_{it}), \theta_{it} = \exp(\alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta}), y_{it} \geq 0.$$

Удобно будет переписать это в виде $\theta_{it} = \gamma_i \exp(\mathbf{x}'_{it}\boldsymbol{\beta}) = \gamma_i \Delta_{it}$. Мы используем свойство инвариантности MLE: оценивание $\gamma_i = \exp(\alpha_i)$ эквивалентно оцениванию α_i . Логарифмическое правдоподобие записывается как

$$\begin{aligned} \ln L &= \sum_{i=1}^n \sum_{t=1}^{T_i} \ln \theta_{it} - \theta_{it} y_{it} \\ &= \sum_{i=1}^n \sum_{t=1}^{T_i} \ln(\gamma_i \Delta_{it}) - (\gamma_i \Delta_{it}) y_{it}. \end{aligned} \quad (14-91)$$

MLE находится приравниванием $n + K$ частных производных по γ_i и $\boldsymbol{\beta}$ к нулю. Для каждого постоянного члена

$$\frac{\partial \ln L}{\partial \gamma_i} = \sum_{t=1}^{T_i} \left(\frac{1}{\gamma_i} - \Delta_{it} y_{it} \right).$$

Приравнивание этой величины к нулю дает следующее решение для γ_i в терминах данных и $\boldsymbol{\beta}$:

$$\gamma_i = \frac{T_i}{\sum_{t=1}^{T_i} \Delta_{it} y_{it}}. \quad (14-92)$$

[Обратим внимание на аналогичный результат для линейной модели в (11-16).] Подставляя это решение обратно в логарифмическую функцию правдоподобия в (14-91), мы получаем концентрированное логарифмическое правдоподобие

$$\ln L_C = \sum_{i=1}^n \sum_{t=1}^{T_i} \left[\ln \left(\frac{T_i \Delta_{it}}{\sum_{s=1}^{T_i} \Delta_{is} y_{is}} \right) - \left(\frac{T_i \Delta_{it}}{\sum_{s=1}^{T_i} \Delta_{is} y_{is}} \right) y_{it} \right],$$

которое теперь является функцией только от $\boldsymbol{\beta}$. Теперь мы можем максимизировать эту функцию только по $\boldsymbol{\beta}$. ML-оценки для α_i находятся как логарифмы формулы (14-92). Еще раз обратим внимание, что мы исключили постоянные из задачи оценивания, но не с помощью вычисления отклонений от внутригрупповых средних. Это преобразование годится лишь для линейной модели.

Концентрированное логарифмическое правдоподобие может быть получено всего в нескольких случаях, включая линейную модель, экспоненциальную модель (как только что было показано), модель пуассоновской регрессии и некоторые другие. В работе Lancaster (2000) перечислены некоторые из них и исследованы основные методологии. В большинстве случаев при желании оценить параметры модели с постоянными эффектами будет необходимо честно вычислять возможно большее число постоянных членов α_i одновременно с основными параметрами β . Это часто интерпретируется как практическое препятствие для оценивания этой модели, связанное с необходимостью обращаться большую матрицу вторых производных, но это заблуждение. [См., например, Maddala (1987), p. 317.] Уравнения правдоподобия для модели с постоянными эффектами имеют вид

$$\frac{\partial \ln L}{\partial \alpha_i} = \sum_{t=1}^{T_i} \frac{\partial \ln f(y_{it} | z_{it})}{\partial z_{it}} \frac{\partial z_{it}}{\partial \alpha_i} = \sum_{t=1}^{T_i} g_{it} = g_i = 0$$

и

$$\frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^n \sum_{t=1}^{T_i} \frac{\partial \ln f(y_{it} | z_{it})}{\partial z_{it}} \frac{\partial z_{it}}{\partial \beta} = \sum_{i=1}^n \sum_{t=1}^{T_i} g_{it} \mathbf{x}_{it} = \mathbf{0}.$$

Матрица вторых производных есть

$$\frac{\partial^2 \ln L}{\partial \alpha_i^2} = \sum_{t=1}^{T_i} \frac{\partial^2 \ln f(y_{it} | z_{it})}{\partial z_{it}^2} = \sum_{t=1}^{T_i} h_{it} = h_i < 0,$$

$$\frac{\partial^2 \ln L}{\partial \beta \partial \alpha_i} = \sum_{t=1}^{T_i} h_{it} \mathbf{x}_{it},$$

$$\frac{\partial^2 \ln L}{\partial \beta \partial \beta'} = \sum_{i=1}^n \sum_{t=1}^{T_i} h_{it} \mathbf{x}_{it} \mathbf{x}_{it}' = \mathbf{H}_{\beta\beta'},$$

где $\mathbf{H}_{\beta\beta'}$ — отрицательно определенная матрица. Уравнения правдоподобия представляют собой большую систему, но решение оказывается на удивление несложным. [См. Greene (2001).]

Используя формулу для блочной обратной матрицы, мы получаем, что $K \times K$, подматрица обратного гессиана, которая соответствует β и дает асимптотическую ковариационную матрицу для MLE, есть

$$\begin{aligned} \mathbf{H}^{\beta\beta'} &= \left\{ \sum_{i=1}^n \left[\sum_{t=1}^{T_i} h_{it} \mathbf{x}_{it} \mathbf{x}_{it}' - \frac{1}{h_i} \left(\sum_{t=1}^{T_i} h_{it} \mathbf{x}_{it} \right) \left(\sum_{t=1}^{T_i} h_{it} \mathbf{x}_{it}' \right) \right] \right\}^{-1}, \\ &= \left\{ \sum_{i=1}^n \left[\sum_{t=1}^{T_i} h_{it} (\mathbf{x}_{it} - \bar{\mathbf{x}}_i) (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' \right] \right\}^{-1}, \quad \text{где } \bar{\mathbf{x}}_i = \frac{\sum_{t=1}^{T_i} h_{it} \mathbf{x}_{it}}{h_i}. \end{aligned}$$

Обратим внимание на поразительное сходство с результатом, который у нас получился в (11-20) для модели с фиксированными эффектами в линейном случае. [Схожий результат вкратце отмечен в работе Chamberlain (1984).] Собирая гессиан как блок для β и блок для полного вектора постоянных членов

с помощью (A-66b) и предыдущих определений, относящихся к выделению одного диагонального элемента, мы получаем

$$\mathbf{H}^{\alpha_i \alpha_i} = \frac{1}{h_i} + \bar{\mathbf{x}}_i' \mathbf{H}^{\beta \beta'} \bar{\mathbf{x}}_i.$$

Вновь результат имеет такой же вид, как и его аналог в линейной модели. [См. (11.19).] В принципе эти матрицы, взятые со знаком минус, будут оценками асимптотических дисперсий оценок максимального правдоподобия. (Как мы скоро увидим, асимптотические свойства в этой модели наталкиваются на затруднения.)

Все они могут быть довольно легко рассчитаны сразу же, как только станут известны оценки параметров. Поэтому в сущности оценивание модели на практике не наталкивается на сложность. [Однако тут требуется кое-что уточнить. Рассмотрим уравнение правдоподобия для одной из постоянных в геометрической регрессионной модели. Оно будет иметь вид

$$\sum_{t=1}^{T_i} [\theta_{it}(1 + y_{it}) - 1] = 0.$$

Допустим, y_{it} равно нулю в каждый момент для индивида i . Тогда решение имеет место там, где выполняется $\sum_i (\theta_{it} - 1) = 0$. Но θ_{it} находится между нулем и единицей, поэтому сумма должна быть отрицательной и не может равняться нулю. Уравнение правдоподобия не имеет решения с конечными коэффициентами. Такие группы должны были быть удалены из выборки для оценивания этой модели.]

В работе Greene (2001) показано, что при потенциально большом числе параметров в модели можно использовать метод Ньютона со следующей итерацией, для которой требуется только вычисленная ранее $K \times K$ матрица и несколько $K \times 1$ векторов:

$$\hat{\beta}^{(s+1)} = \hat{\beta}^{(s)} - \left\{ \sum_{i=1}^n \left[\sum_{t=1}^{T_i} h_{it} (\mathbf{x}_{it} - \bar{\mathbf{x}}_i) (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' \right] \right\}^{-1} \times \\ \times \left\{ \sum_{i=1}^n \left[\sum_{t=1}^{T_i} g_{it} (\mathbf{x}_{it} - \bar{\mathbf{x}}_i) \right] \right\} = \hat{\beta}^{(s)} + \Delta_{\beta}^{(s)}$$

и

$$\hat{\alpha}_l^{(s+1)} = \hat{\alpha}_l^{(s)} - [(g_l/h_l) + \bar{\mathbf{x}}_l' \Delta_{\beta}^{(s)}]^{27}.$$

Этот алгоритм требует большого объема вычислений, включая много суммирований, однако он линеен по числу параметров и не включает никаких $n \times n$ матриц.

В дополнение к теоретическим преимуществам и недостаткам этой модели нужно отметить практический аспект оценивания потенциально большого числа параметров $n + K$. В случае фиксированных эффектов n не ограничено и зачастую на практике составляет тысячи. [В примере 14.14 n равно 7,293. На момент написания книги крупнейшее применение описанного

²⁷ Похожие результаты встречаются в работе Prentice, Gloeckler (1978), которые ссылаются на работы Rao (1973) и Chamberlain (1980, 1984).

здесь метода, о котором нам известно, есть исследование Кингдона и Кассена (Kingdon, Cassen (2007)), в котором они подбирают пробит-модель с фиксированными эффектами, включающую более чем 140,000 коэффициентов при фиктивных переменных.] Проблемы с оценкой фиксированных эффектов носят статистический, а не практический характер²⁸. Оценка основывается на T_i , которое должно увеличиваться, чтобы постоянные члены были состоятельными: по сути, каждое α_i оценивается с помощью T_i наблюдений. В этой постановке задачи T_i является не только постоянным, но, скорее всего, и совсем небольшим. Таким образом, оценки постоянных членов не являются состоятельными (не потому, что они сходятся к чему-то другому, чем параметр, который они пытаются оценить, а потому, что они вообще не сходятся). В связи с этим возникает смещение из-за выборки небольшого объема (небольшого T_i) в оценках коэффициентов наклона. Это носит название **проблема несущественных параметров** (incidental parameters problem). [Neyman, Scott (1948) и Lancaster (2000).] Мы изучим проблему несущественных параметров несколько подробнее при рассмотрении исследования с помощью метода Монте-Карло в подразделе 15.5.2.

Пример 14.14. Геометрическая регрессия с фиксированными и случайными эффектами

В примере 14.10 были представлены оценки для геометрической регрессионной модели пула

$$f(y_{it} | \mathbf{x}_{it}) = \theta_{it}(1 - \theta_{it})^{y_{it}}, \theta_{it} = 1/(1 + \lambda_{it}), \lambda_{it} = \exp(c_i + \mathbf{x}'_{it}\beta), y_{it} = 0, 1, \dots$$

Сейчас мы переоценим модель, предполагая наличие случайных и фиксированных эффектов. Мы будем непосредственно применять методы двух предыдущих разделов: их модификация не требуется. В табл. 14.11 приведены три набора оценок максимального правдоподобия. Оценки значительно различаются. Средний размер группы составляет около пяти. Это означает, что оценка постоянных эффектов может вполне подлежать смещению малых выборок. Если не считать коэффициент при переменной *Kids*, оценки фиксированных и случайных эффектов очень похожи. С другой стороны, две панельные модели дают результаты, похожие на модель пула, за исключением коэффициента при переменной *Income*. Основываясь только на результатах, трудно предпочесть одну из этих моделей. Заметим также, что модель является нелинейной, поэтому модель пула, которую иначе можно было бы предпочесть исходя из легкости расчетов, сейчас не имеет таких преимуществ. Никакая из трех моделей не будет робастной (устойчивой) к неправильной спецификации. В отличие от линейной модели в этой и других нелинейных моделях оценка фиксированных эффектов будет несостоятельной, когда T является небольшим в случае модели как со случайными, так и с постоянными эффектами. Оценка случайных эффектов становится состоятельной

²⁸См. работы Vytlačil, Aakvik, Heckman (2005), Chamberlain (1980, 1984), Newey (1994), Bover, Arellano (1997) и Chen (1998) для изучения некоторых обобщений параметрических и полупараметрических видов моделей двоичного выбора с фиксированными эффектами.

в модели со случайными эффектами, но, как всегда, не в модели с фиксированными эффектами. Оценка пула считается несостоятельной при наличии как случайных, так и фиксированных эффектов (что сводит на нет преимущество робастной ковариационной матрицы). Могло бы возникнуть желание использовать критерий Хаусмана (см. подраздел 11.5.5), однако условия этого критерия не соблюдаются: в отличие от линейной модели, где оценка постоянных эффектов состоятельна в обоих случаях, здесь она несостоятельна в обоих случаях. Так или иначе исследователю придется выбирать модель, основываясь на теоретических предположениях.

Таблица 14.11. Оценки модели панельных данных геометрической регрессии для *DOCVIS*

| Переменная | Суммарная | | Случайных эффектов ^a | | Постоянных эффектов | |
|----------------------------|-----------|------------|---------------------------------|------------|---------------------|------------|
| | оценка | станд. ош. | оценка | станд. ош. | оценка | станд. ош. |
| Постоянная | 1,0918 | 0,1112 | 0,3998 | 0,09531 | | |
| Age (Возраст) | 0,0180 | 0,0013 | 0,02208 | 0,001220 | 0,04845 | 0,003511 |
| Education (Образование) | -0,0473 | 0,0069 | -0,04507 | 0,006262 | -0,05437 | 0,03721 |
| Income (Доход) | -0,0468 | 0,0075 | -0,1959 | 0,06103 | -0,1892 | 0,09127 |
| Kids (Дети) | -0,1569 | 0,0319 | -0,1242 | 0,02336 | -0,002543 | 0,03687 |

14.10. Модели скрытых классов и модели смеси распределений с конечным числом значений параметра

В этом завершающем главу приложении метода максимального правдоподобия мы будем исследовать, скорее, методику, используемую во многих различных случаях, чем конкретную модель. Постановка задачи моделей скрытых классов предполагает, что распределение наблюдений есть смесь конечного числа некоторых распределений. Мотивация такой модели следует из разных оснований:

- В классическом применении этой методики данные наблюдений берутся из смеси различных генеральных совокупностей. Рассмотрим, например, исторический или иным образом фиксированный набор данных из двух генеральных совокупностей. Этот набор состоит из измерений некоторой переменной, которая отличается не полностью, но существенно от генеральной совокупности. Тем не менее у исследователя нет отличительного признака, из какой именно совокупности берутся наблюде-

ния. При заданной выборке наблюдений исследователя интересуют две статистические задачи: (1) оценить параметры этих генеральных совокупностей и (2) классифицировать имеющиеся наблюдения по генеральным совокупностям. У этой методики есть ряд современных приложений, связанных с исследованием здоровья. Например, в исследовании проблемы ожирения в работе Greene, Harris, Hollingsworth, Maitra (2008) авторы предположили, что их упорядоченная модель выбора (см. главу 19) могла изменяться в выборке, которая включала (по предположению) некоторых индивидов, имеющих наследственную предрасположенность к ожирению, и большинство тех, у кого ее нет. В другом недавнем исследовании, Lambert (1992), автор изучал количество брака при производстве продукции. Когда наблюдается условие «отсутствия дефектов», оно может означать как режим 1 «процесс под контролем», так и режим 2 «процесс не под контролем, но просто брак не наблюдается».

- В узком смысле на неоднородность параметров в генеральной совокупности можно смотреть как на вид перемешивания дискретных распределений. Мы моделировали гетерогенность параметров с помощью непрерывных распределений в разделе 11.11. Подход «конечной смеси» предполагает, что распределение параметров по индивидам является дискретным. (Конечно, это другой способ интерпретации первого пункта.)
- Подход конечного перемешивания представляет собой способ, с помощью которого распределение (модель) может быть построено из смеси базовых распределений. Смесь нормальных распределений модели Квандта и Рамсея в примере 13.4 представляет собой случай, в котором распределение, отличное от нормального, создается с помощью смеси двух нормальных распределений с различными параметрами.

14.10.1. Модель смеси распределений

Для того чтобы заложить базу для более полной модели, которая описывается ниже, вернемся сначала к смеси нормальных моделей из примера 13.4. Рассмотрим генеральную совокупность, которая состоит из скрытой смеси двух базовых нормальных распределений. Пренебрегая тем фактом, что мы не знаем, которое из распределений верно для данного индивида, для индивида i выполняется следующее:

$$\begin{aligned} f(y_i | class_i = 1) &= N[\mu_1, \sigma_1^2] = \frac{\exp[-\frac{1}{2}(y_i - \mu_1)^2 / \sigma_1^2]}{\sigma_1 \sqrt{2\pi}}, \\ &\text{и} \\ f(y_i | class_i = 2) &= N[\mu_2, \sigma_2^2] = \frac{\exp[-\frac{1}{2}(y_i - \mu_2)^2 / \sigma_2^2]}{\sigma_2 \sqrt{2\pi}}. \end{aligned} \quad (14-93)$$

Вклад в функцию правдоподобия составляет $f(y_i | class_i = 1)$ для индивида из класса 1 и $f(y_i | class = 2)$ для индивида из класса 2. Допустим, что существует истинная пропорция $\lambda = \text{Prob}(class_i = 1)$ индивидов из генеральной совокупности, которые принадлежат классу 1, и $(1 - \lambda)$ для класса 2. Тогда безусловная (одномерная) плотность для индивида i есть

$$\begin{aligned} f(y_i) &= \lambda f(y_i | class_i = 1) + (1 - \lambda) f(y_i | class_i = 2) \\ &= E_{classes} f(y_i | class_i). \end{aligned} \quad (14-94)$$

Параметрами, которые необходимо оценить, являются λ , μ_1 , μ_2 , σ_1 и σ_2 . Логарифмическое правдоподобие для выборки из n отдельных наблюдений после группировки членов имеет вид

$$\ln L = \sum_{i=1}^n \ln \left(\frac{\lambda \exp \left[-\frac{1}{2}(y_i - \mu_1)^2 / \sigma_1^2 \right]}{\sigma_1 \sqrt{2\pi}} + \frac{(1 - \lambda) \exp \left[-\frac{1}{2}(y_i - \mu_2)^2 / \sigma_2^2 \right]}{\sigma_2 \sqrt{2\pi}} \right). \quad (14-95)$$

Это та же самая смесь плотностей, с которой мы уже сталкивались в примере 13.4. Тогда мы предложили метод моментов для оценивания пяти параметров. Тем не менее оказывается, что эта задача может быть решена непосредственно методом максимального правдоподобия.

Пример 14.15. Модель скрытого класса для средних баллов учащихся

Таблица F14.1 приложения содержит набор данных из 32 наблюдений, использованных в работе Спектора и Маццео (Spector, Mazzeo (1980)), для исследования влияния нового метода преподавания экономики — индивидуальной системы обучения (ИСО, PSI, Personalized System of Instruction) на качество учебы на более продвинутых курсах экономики. Переменные в наборе данных включают

- GPA_i = средний балл учащегося,
- $GRADE_i$ = фиктивную переменную для определения того, был ли балл студента выше на промежуточном курсе макроэкономики по сравнению с базовым курсом,
- PSI_i = фиктивную переменную для определения того, участвовал ли учащийся в ИСО,
- $TUCE_i$ = балл учащегося на предварительном тесте по экономике.

Мы будем использовать эти данные, чтобы разработать модель смеси конечного числа нормальных распределений для моделирования распределения средних баллов.

Мы начнем с того, что вычислим оценки максимального правдоподобия параметров для (14-95). Чтобы оценить параметры, используя итерационные методы, необходимо определиться с набором начальных значений. Может показаться естественным использовать просто значения из модели одного класса: \bar{y} и s_y , для λ взять, например, $1/2$. Однако алгоритм сразу остановится на этих значениях, поскольку производные на этом шаге будут нулевыми. Вместо этого часто используют близкие значения, незначительно отклоняя их (на несколько процентов), просто чтобы иметь возможность начать итерации алгоритма. Таблица 14.12 содержит оценки для этой модели смеси с двумя классами. Оценками для модели с одним классом являются выборочное среднее значение и стандартное отклонение GPA . [Поскольку они являются ML-оценками, $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (GPA_i - \overline{GPA})^2$.]

Таблица 14.12. Оценка нормальной модели смеси

| Параметр | Один класс | | Скрытый класс 1 | | Скрытый класс 2 | |
|-------------|------------|------------|-----------------|------------|-----------------|------------|
| | Оценка | Станд. ош. | Оценка | Станд. ош. | Оценка | Станд. ош. |
| μ | 3,1172 | 0,08251 | 3,64187 | 0,3452 | 2,8894 | 0,2514 |
| σ | 0,4594 | 0,04070 | 0,2524 | 0,2625 | 0,3218 | 0,1095 |
| Вероятность | 1,0000 | 0,0000 | 0,3028 | 0,3497 | 0,6972 | 0,3497 |
| $\ln L$ | -20,51274 | | -19,63654 | | | |

Средние значения и стандартные отклонения двух классов заметно различаются: по-видимому, модель выявляет четкое разделение данных на два класса. (Вопрос о том, является ли два подходящим числом классов, рассматривается в подразделе 14.10.5.) На данном этапе возникает желание связать идентификацию класса с некоторой переменной, имеющейся или нет в начальном наборе данных, такой как *PSI*. Однако здесь нет оснований для этого: рассматриваемые классы являются «скрытыми». Тем не менее при дальнейшем анализе у нас будет желание исследовать, помогают ли какие-либо данные наблюдений предсказывать принадлежность к классу.

14.10.2. Измеренная и неизмеренная гетерогенность

До сих пор мы предполагали, что у исследователя нет никакой информации о принадлежности индивида к классу. Оценивание «априорных» вероятностей (λ в предыдущем примере) является частью задачи оценивания. Однако у исследователя может быть некоторая, возможно неполная, информация о разделении выборки на разные классы. В предыдущем примере, связанном со средними баллами, у нас была информация об оценке индивида на предварительном тестировании (*TUCE*). Использование этой информации могло бы уточнить оценки вероятностей принадлежности к различным классам. Задачу о смеси нормальных распределений можно, например, сформулировать следующим образом:

$$f(y_i | \mathbf{z}_i) = \left(\frac{\text{Prob}(\text{class} = 1 | \mathbf{z}_i) \exp \left[-\frac{1}{2}(y_i - \mu_1)^2 / \sigma_1^2 \right]}{\sigma_1 \sqrt{2\pi}} + \frac{[1 - \text{Prob}(\text{class} = 1 | \mathbf{z}_i)] \exp \left[-\frac{1}{2}(y_i - \mu_2)^2 / \sigma_2^2 \right]}{\sigma_2 \sqrt{2\pi}} \right),$$

где \mathbf{z}_i — вектор переменных, призванный объяснить вероятности классов. Чтобы иметь возможность оценить модель смеси, необходимо параметризовать вероятности. Для этого часто используется вероятностная логит-модель. (См. раздел 17.2. Применение этой модели можно найти в работе Грина (Greene (2007d, подраздел 2.3.3)); см. также приводимые там ссылки.) Для

случая двух классов это будет иметь вид

$$\begin{aligned} \text{Prob}(class = 1 | \mathbf{z}_i) &= \frac{\exp(\mathbf{z}'_i \boldsymbol{\theta})}{1 + \exp(\mathbf{z}'_i \boldsymbol{\theta})}, \\ \text{Prob}(class = 2 | \mathbf{z}_i) &= 1 - \text{Prob}(class = 1 | \mathbf{z}_i). \end{aligned} \quad (14-96)$$

(Более общий случай J -классов рассмотрен в подразделе 14.10.6.) Логарифмическое правдоподобие для нашей смеси двух нормальных распределений становится равным

$$\begin{aligned} \ln L &= \sum_{i=1}^n \ln L_i \\ &= \sum_{i=1}^n \ln \left(\left(\frac{\exp(\mathbf{z}'_i \boldsymbol{\theta})}{1 + \exp(\mathbf{z}'_i \boldsymbol{\theta})} \right) \frac{\exp \left[-\frac{1}{2} (y_i - \mu_1)^2 / \sigma_1^2 \right]}{\sigma_1 \sqrt{2\pi}} \right. \\ &\quad \left. + \left(\frac{1}{1 + \exp(\mathbf{z}'_i \boldsymbol{\theta})} \right) \frac{\exp \left[-\frac{1}{2} (y_i - \mu_2)^2 / \sigma_2^2 \right]}{\sigma_2 \sqrt{2\pi}} \right). \end{aligned} \quad (14-97)$$

Логарифмическое правдоподобие теперь максимизируется по $\mu_1, \sigma_1, \mu_2, \sigma_2$ и $\boldsymbol{\theta}$. Если \mathbf{z}_i содержит постоянный член и некоторые другие наблюдаемые переменные, то мы вернемся к первоначальной модели, когда все коэффициенты этих переменных будут равны нулю. В этом случае $\lambda = \ln[\theta/(1-\theta)]$. (Этот способ обычно используется, чтобы убедиться, что $0 < \lambda < 1$ в первоначальной модели.)

14.10.3. Предсказание принадлежности к классу

Модель (14-97) описывает две случайные переменные: y_i — интересующую нас переменную исходов и $class_i$ — показатель того, к какому классу принадлежит индивид. У нас есть совместное распределение $f(y_i, class_i)$, которое мы моделируем в терминах плотности условного распределения $f(y_i | class_i)$ (14-93) и плотности одномерного безусловного распределения $class_i$ (14-96). Изначально мы предположили, что последнее является простой бернуллиевской величиной с $\text{Prob}(class_i = 1) = \lambda$, но в предыдущем разделе мы преобразовали это к $\text{Prob}(class_i = 1 | \mathbf{z}_i) = \Lambda(\mathbf{z}'_i \boldsymbol{\theta})$. Это можно рассматривать как априорные вероятности в байесовском смысле. Если мы хотим построить предсказание, к какому классу принадлежит данный индивид, используя всю доступную нам по этому индивиду информацию, то априорная вероятность будет использоваться не полностью. Апостериорная, или условная (по остальным данным), вероятность

$$\text{Prob}(class_i = 1 | \mathbf{z}_i y_i) = \frac{f(y_i, class = 1 | \mathbf{z}_i)}{f(y_i)} \quad (14-98)$$

будет основываться скорее на полученной информации, чем на исходных безусловных вероятностях. У нас есть все элементы, которые нам нужны для вычисления этой условной вероятности. Используя **теорему Байеса** (Bayes's

theorem), запишем это в виде

$$\begin{aligned} & \text{Prob}(class_i = 1 | \mathbf{z}_i, y_i) \\ &= \frac{f(y_i | class_i = 1, \mathbf{z}_i) \text{Prob}(class_i = 1 | \mathbf{z}_i)}{f(y_i | class_i = 1, \mathbf{z}_i) \text{Prob}(class_i = 1 | \mathbf{z}_i) + f(y_i | class_i = 2, \mathbf{z}_i) \text{Prob}(class_i = 2 | \mathbf{z}_i)}. \end{aligned} \quad (14-99)$$

Знаменателем дроби является L_i (не $\ln L_i$) из (14-97). Числитель — это первый элемент в L_i . В случае нашего примера смеси двух нормальных распределений условная (апостериорная) вероятность записывается как

$$\text{Prob}(class_i = 1 | \mathbf{z}_i, y_i) = \frac{\left(\frac{\exp(\mathbf{z}_i' \boldsymbol{\theta})}{1 + \exp(\mathbf{z}_i' \boldsymbol{\theta})} \right) \frac{\exp\left[-\frac{1}{2}(y_i - \mu_1)^2 / \sigma_1^2\right]}{\sigma_1 \sqrt{2\pi}}}{L_i}, \quad (14-100)$$

тогда как безусловная вероятность представлена формулой (14-96). Условная вероятность для второго класса рассчитывается с помощью двух других плотностей безусловного распределения в числителе (или путем вычитания из единицы). Обратим внимание, что условные вероятности представляют собой функции от данных, даже если это не так для безусловных вероятностей. Возвращаясь к задаче, сформулированной в начале параграфа, естественное предсказание для $class_i$ есть класс с наибольшей оценкой апостериорной вероятности.

14.10.4. Условная модель скрытого класса

Чтобы завершить рассмотрение модели скрытого класса, отметим, что средние (и, вообще говоря, дисперсии) в первоначальной модели могли быть также условными по данным наблюдениям. Для наших моделей смеси нормальных распределений мы могли бы заменить безусловное среднее μ_j на условное

$$\mu_{ij} = \mathbf{x}_i' \boldsymbol{\beta}_j.$$

В данных примера 14.15 присутствует также индикатор того, принимал ли индивид участие в специальной программе, разработанной в целях улучшения курса экономики (PSI_i). Мы могли бы модифицировать модель следующим образом:

$$f(y_i | class_i = 1, PSI_i) = N[\mu_{i1}, \sigma_1^2] = \frac{\exp\left[-\frac{1}{2}(y_i - \beta_{1,1} - \beta_{2,1} PSI_i)^2 / \sigma_1^2\right]}{\sigma_1 \sqrt{2\pi}}$$

и аналогичным образом для $f(y_i | class_i = 2, PSI_i)$. Теперь она является **линейной регрессионной моделью скрытого класса** (latent class linear regression model).

В более общем смысле, как мы в скором времени увидим, модель скрытого класса, или **модель конечной смеси** (finite mixture model), для переменной y_i может быть описана следующим образом:

$$f(y_i | class_i = j, \mathbf{x}_i) = h_j(y_i, \mathbf{x}_i, \boldsymbol{\gamma}_j),$$

где h_j обозначает плотность, условную по классу j ; индекс j здесь указывает, например, j -й вектор параметров $\gamma_j = (\beta_j, \sigma_j)$ и т.д. Безусловные одномерные вероятности классов записываются как

$$\text{Prob}(\text{class}_i = j | \mathbf{z}_i) = p_j(j, \mathbf{z}_i, \theta).$$

Подобная техника может быть применена к любой модели для y_i . В примере, приведенном в подразделе 14.10.6, мы смоделируем бинарную зависимую переменную с пробит-моделью. Методология нашла применение и во многих других моделях, таких как модели стохастической границы [Orea, Kumbhakar (2004), Greene (2004)], модели пуассоновской регрессии [Wedel et al. (1993)], а также в обширном числе моделей количественных переменных, дискретного выбора, ограниченных зависимых переменных [McLachlan, Peel (2000), Greene (2007b)].

Пример 14.16. Модель регрессии скрытого класса для среднего балла
Объединяя 14.10.2 и 14.10.4, мы получаем модель скрытого класса для средних баллов:

$$f(\text{GPA}_i | \text{class}_i = j, \text{PSI}_i) = \frac{\exp[-\frac{1}{2}(y_i - \beta_{1j} - \beta_{2j}\text{PSI}_i)^2/\sigma_j^2]}{\sigma_j\sqrt{2\pi}}, j = 1, 2,$$

$$\text{Prob}(\text{class}_i = 1 | \text{TUCE}_i) = \frac{\exp(\theta_1 + \theta_2 \text{TUCE}_i)}{1 + \exp(\theta_1 + \theta_2 \text{TUCE}_i)},$$

$$\text{Prob}(\text{class}_i = 2 | \text{TUCE}_i) = 1 - \text{Prob}(\text{class} = 1 | \text{TUCE}_i).$$

Теперь логарифмическое правдоподобие выглядит как $\ln L =$

$$= \sum_{i=1}^n \ln \left(\left(\frac{\exp(\theta_1 + \theta_2 \text{TUCE}_i)}{1 + \exp(\theta_1 + \theta_2 \text{TUCE}_i)} \right) \frac{\exp[-\frac{1}{2}(y_i - \beta_{1,1} - \beta_{2,1}\text{PSI}_i)^2/\sigma_1^2]}{\sigma_1\sqrt{2\pi}} + \left(\frac{1}{1 + \exp(\theta_1 + \theta_2 \text{TUCE}_i)} \right) \frac{\exp[-\frac{1}{2}(y_i - \beta_{1,2} - \beta_{2,2}\text{PSI}_i)^2/\sigma_2^2]}{\sigma_2\sqrt{2\pi}} \right).$$

Оценки параметров методом максимального правдоподобия представлены в табл. 14.13.

Таблица 14.13. Оцененная модель линейной регрессии скрытого класса для GPA

| Параметр | Один класс | | Скрытый класс 1 | | Скрытый класс 2 | |
|------------------------------------|------------|------------|-----------------|------------|-----------------|------------|
| | Оценка | Станд. ош. | Оценка | Станд. ош. | Оценка | Станд. ош. |
| β_1 | 3,1011 | 0,1117 | 3,3928 | 0,1733 | 2,7926 | 0,04988 |
| β_2 | 0,03675 | 0,1689 | -0,1074 | 0,2006 | -0,5703 | 0,07553 |
| $\sigma = \mathbf{e}'\mathbf{e}/n$ | 0,4443 | 0,0003086 | 0,3812 | 0,09337 | 0,1119 | 0,04487 |
| θ_1 | 0,0000 | 0,0000 | -6,8392 | 3,07867 | 0,0000 | 0,0000 |
| θ_2 | 0,0000 | 0,0000 | 0,3518 | 0,1601 | 0,0000 | 0,0000 |
| Prob TUCE | 1,0000 | | 0,7063 | | 0,2937 | |
| $\ln L$ | -20,48752 | | | | -13,39966 | |

Таблица 14.14. Оцененная вероятность принадлежности к скрытому классу

| GPA | TUCE | PSI | КЛАСС | P1 | P1* | P2 | P2* |
|------|------|-----|-------|--------|--------|--------|--------|
| 2,06 | 22 | 1 | 2 | 0,7109 | 0,0116 | 0,2891 | 0,9884 |
| 2,39 | 19 | 1 | 2 | 0,4612 | 0,0467 | 0,5388 | 0,9533 |
| 2,63 | 20 | 0 | 2 | 0,5489 | 0,1217 | 0,4511 | 0,8783 |
| 2,66 | 20 | 0 | 2 | 0,5489 | 0,1020 | 0,4511 | 0,8980 |
| 2,67 | 24 | 1 | 1 | 0,8325 | 0,9992 | 0,1675 | 0,0008 |
| 2,74 | 19 | 0 | 2 | 0,4612 | 0,0608 | 0,5388 | 0,9392 |
| 2,75 | 25 | 0 | 2 | 0,8760 | 0,3499 | 0,1240 | 0,6501 |
| 2,76 | 17 | 0 | 2 | 0,2975 | 0,0317 | 0,7025 | 0,9683 |
| 2,83 | 19 | 0 | 2 | 0,4612 | 0,0821 | 0,5388 | 0,9179 |
| 2,83 | 27 | 1 | 1 | 0,9345 | 1,0000 | 0,0655 | 0,0000 |
| 2,86 | 17 | 0 | 2 | 0,2975 | 0,0532 | 0,7025 | 0,9468 |
| 2,87 | 21 | 0 | 2 | 0,6336 | 0,2013 | 0,3664 | 0,7987 |
| 2,89 | 14 | 1 | 1 | 0,1285 | 1,0000 | 0,8715 | 0,0000 |
| 2,89 | 22 | 0 | 2 | 0,7109 | 0,3065 | 0,2891 | 0,6935 |
| 2,92 | 12 | 0 | 2 | 0,0680 | 0,0186 | 0,9320 | 0,9814 |
| 3,03 | 25 | 0 | 1 | 0,8760 | 0,9260 | 0,1240 | 0,0740 |
| 3,10 | 21 | 1 | 1 | 0,6336 | 1,0000 | 0,3664 | 0,0000 |
| 3,12 | 23 | 1 | 1 | 0,7775 | 1,0000 | 0,2225 | 0,0000 |
| 3,16 | 25 | 1 | 1 | 0,8760 | 1,0000 | 0,1240 | 0,0000 |
| 3,26 | 25 | 0 | 1 | 0,8760 | 0,9999 | 0,1240 | 0,0001 |
| 3,28 | 24 | 0 | 1 | 0,8325 | 0,9999 | 0,1675 | 0,0001 |
| 3,32 | 23 | 0 | 1 | 0,7775 | 1,0000 | 0,2225 | 0,0000 |
| 3,39 | 17 | 1 | 1 | 0,2975 | 1,0000 | 0,7025 | 0,0000 |
| 3,51 | 26 | 1 | 1 | 0,9094 | 1,0000 | 0,0906 | 0,0000 |
| 3,53 | 26 | 0 | 1 | 0,9094 | 1,0000 | 0,0906 | 0,0000 |
| 3,54 | 24 | 1 | 1 | 0,8325 | 1,0000 | 0,1675 | 0,0000 |
| 3,57 | 23 | 0 | 1 | 0,7775 | 1,0000 | 0,2225 | 0,0000 |
| 3,62 | 28 | 1 | 1 | 0,9530 | 1,0000 | 0,0470 | 0,0000 |
| 3,65 | 21 | 1 | 1 | 0,6336 | 1,0000 | 0,3664 | 0,0000 |
| 3,92 | 29 | 0 | 1 | 0,9665 | 1,0000 | 0,0335 | 0,0000 |
| 4,00 | 21 | 0 | 1 | 0,6336 | 1,0000 | 0,3664 | 0,0000 |
| 4,00 | 23 | 1 | 1 | 0,7775 | 1,0000 | 0,2225 | 0,0000 |

Таблица 14.14 содержит наблюдения, отсортированные с помощью GPA. Предсказание принадлежности к тому или иному классу отражает то, что можно сказать по поводу коэффициентов из соответствующей таблицы. Члены, принадлежащие классу 2, в среднем имеют более низкие GPA, чем члены класса 1. Перечисление данных в табл. 14.14 выявляет эту кластерность. Также эта таблица позволяет предположить, как в модели скрытого класса используется информация о выборке. Если использовать результаты табл. 14.12 (только оценки средних, постоянных вероятностей классов) для получения такой же таблицы, то после сортировки данных

наивысшие 10 GPA находятся в классе 1, а оставшаяся часть — в классе 2. Более тщательно разработанная модель учитывает в вычислениях информацию о *TUCE*. Невысокое значение *TUCE* может переместить учащегося с высоким GPA в класс 2. (Конечно, это в общем то же самое, что делает множественная линейная регрессия.)

14.10.5. Определение числа классов

В спецификации модели остается один нерешенный вопрос, связанный с ее оцениванием. Число классов было взято как известный параметр: два — в нашем основном примере и три — в следующей модели. В идеале нам бы хотелось статистически определить соответствующее число классов. Однако J не является параметром в модели. Так, например, критерий отношения правдоподобий не приведет к верному результату. Рассмотрим исходную модель из примера 14.15. Эта модель имеет два класса и пять параметров в совокупности. Казалось бы, естественно проверить с помощью LR-теста, нельзя ли свести модель к модели с одним классом, которая содержит только среднее значение и дисперсию. Однако здесь неясно число ограничений. Если $\mu_1 = \mu_2$ и $\sigma_1 = \sigma_2$, то использование модели смеси неуместно: плотности в модели с двумя классами совпадают и, таким образом, это есть модель с одним классом. Таким образом, число ограничений, необходимых для перехода от модели с двумя классами к модели с одним классом, является неопределенным. Это и ни два, и ни три. Один возможный путь заключается в том, чтобы двигаться, наоборот, к повышению числа классов, добавляя их до тех пор, пока добавление предельного класса будет незначительно изменять логарифмическое правдоподобие или один из информационных критериев, таких как AIC или BIC (см. подраздел 14.6.5). К сожалению, данный подход также является плохим, поскольку оценки, взятые из любой слишком короткой спецификации, являются несостоятельными. Альтернативный вариант состоит в том, чтобы идти вниз от заведомо длинной спецификации. Этот метод рассматривается в работе Heckman, Singer (1984b), в которой отмечается, что, когда число классов становится больше, чем нужно, оценка становится непригодной. В нашем примере 14.15, если мы используем четыре класса, алгоритм оптимизации перестает работать и далее становится невозможно вычислить оценки. Модель с пятью классами дает оценки, но некоторые из них совершенно бессмысленны. Но это, по крайней мере, дает пути для поиска жизнеспособной стратегии. Известная монография по конечным моделям смеси McLachlan, Peel (2000, Chapter 6) содержит подробное обсуждение этой темы.

14.10.6. Модель панельных данных

Модель скрытого класса является полезной базой для моделей панельных данных. Как мы сейчас увидим, вероятности классов частично играют роль общих случайных эффектов. Модель скрытого класса может быть интерпретирована как модель со случайными параметрами, как это делается в разделе 11.11, где распределение параметров дискретно.

Предположим, что β_j сгенерирована из дискретного распределения с J исходами, или классами, так что β_j распределено по этим классам. Таким образом, в модели предполагается, что физическое лицо принадлежит одному из J скрытых классов, обозначенных с помощью вектора параметров, однако из данных невозможно определить, какой именно это класс. Мы используем выборку для оценки векторов параметров, параметров первоначального распределения вероятностей и вероятностей принадлежности к классу. Соответствующая спецификация модели тогда выглядит следующим образом:

$$f(y_{it} | \mathbf{x}_{it}, \mathbf{z}_i, \Delta, \beta_1, \beta_2, \dots, \beta_J) = \sum_{j=1}^J p_{ij}(\mathbf{z}_i, \Delta) f(y_{it} | \text{class} = j, \mathbf{x}_{it}, \beta_j),$$

где остается параметризовать вероятности классов p_{ij} , а также структурную модель $f(y_{it} | \text{class} = j, \mathbf{x}_{it}, \beta_j)$. Матрица параметров Δ содержит параметры дискретного распределения вероятностей. Она имеет J строк, по одной для каждого класса, и M столбцов, по числу M переменных в \mathbf{z}_i . Как минимум $M = 1$, а \mathbf{z}_i содержит постоянный член, если вероятности классов представляют собой фиксированные параметры, как в примере 14.15. Наконец, чтобы адаптировать характер панельных данных к алгоритму формирования выборки, мы предполагаем, что условно по β_j , т. е. по отношению к классу j , который остается постоянным все время, наблюдения y_{it} являются независимыми. Тогда для группы из T_i наблюдений совместная плотность может быть записана как

$$f(y_{i1}, y_{i2}, \dots, y_{iT_i} | \text{class} = j, \mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iT_i}, \beta_j) = \prod_{t=1}^{T_i} f(y_{it} | \text{class} = j, \mathbf{x}_{it}, \beta_j).$$

Логарифмическая функция правдоподобия для панельных данных есть

$$\ln L = \sum_{i=1}^n \ln \left[\sum_{j=1}^J p_{ij}(\Delta, \mathbf{z}_i) \prod_{t=1}^{T_i} f(y_{it} | \text{class} = j, \mathbf{x}_{it}, \beta_j) \right].$$

Сумма вероятностей классов должна быть в точности равна 1. Обычно используемый подход состоит в том, чтобы перепараметризовать их в набор логит-вероятностей, как мы это делали в предыдущих примерах. Тогда

$$p_{ij}(\mathbf{z}_i, \Delta) = \frac{\exp(\theta_{ij})}{\sum_{j=1}^J \exp(\theta_{ij})}, \quad J = 1, \dots, J, \theta_{ij} = \mathbf{z}_i' \delta_j, \theta_{iJ} = 0 \quad (\delta_J = 0). \quad (14-101)$$

(См. подраздел 18.2.2, в котором нами рассмотрена модель для этого набора вероятностей.) Обратим внимание на ограничение, наложенное на θ_{ij} . Это идентификационное ограничение. Без него тот же набор вероятностей будет получаться при добавлении произвольного вектора к каждой δ_j . Получившееся логарифмическое правдоподобие есть непрерывная функция от параметров β_1, \dots, β_J и $\delta_1, \dots, \delta_J$. Несмотря на кажущуюся сложность, оценивание этой модели непосредственной максимизацией логарифмического правдоподобия не является особо трудным. [См. раздел E.3 и Greene (2001,

2007b). EM-алгоритм, рассматриваемый в разделе E.3.7, особенно подходит для оценивания параметров моделей скрытого класса. См. McLachlan, Peel (2000).] Число классов, которые могут быть идентифицированы, по всей видимости, является сравнительно небольшим (порядка 5, 10 максимум), что часто рассматривается как недостаток этого подхода. Вообще говоря, чем более сложной является модель для y_{it} , тем более сложно увеличивать в этой модели число классов. Кроме того, как можно ожидать, чем данные «беднее» в смысле вариации между группами, тем труднее оценить модели скрытого класса.

Оценивание дает значения для структурных параметров (β_j, δ_j) , $j = 1, \dots, J$. Имея эти оценки, мы можем вычислить априорные вероятности классов p_{ij} с помощью (14-101). Для целей предсказания нас также интересуют апостериорные (по имеющимся данным) вероятности классов, которые мы можем вычислить, используя теорему Байеса [см. (14-99)]. Условная вероятность выражается как

$$\begin{aligned} & \text{Prob}(\text{class} = j \mid \text{observation } i) \\ &= \frac{f(\text{observation } i \mid \text{class} = j) \text{Prob}(\text{class } j)}{\sum_{j=1}^J f(\text{observation } i \mid \text{class} = j) \text{Prob}(\text{class } j)} \\ &= \frac{f(y_{i1}, y_{i2}, \dots, y_{iT_i} \mid \mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iT_i}, \beta_j) p_{ij}(\mathbf{z}_j, \Delta)}{\sum_{j=1}^J f(y_{i1}, y_{i2}, \dots, y_{iT_i} \mid \mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iT_i}, \beta_j) p_{ij}(\mathbf{z}_j, \Delta)} \quad (14-102) \\ &= w_{ij}. \end{aligned}$$

Набор вероятностей $\mathbf{w}_i = (w_{i1}, w_{i2}, \dots, w_{iJ})$ дает апостериорную плотность для распределения значений β , т.е. $[\beta_1, \beta_2, \dots, \beta_J]$.

Таблица 14.15. Оценки панельных данных для геометрической регрессии для DocVis

| Переменная | Объединенная MLE (нелинейный метод наименьших квадратов) | | Случайные эффекты ^a | | Фиксированные эффекты | |
|----------------------------|---|---------------------|-----------------------------------|---------------|--------------------------|---------------|
| | Оценка | Станд. ош. | Оценка | Станд. ош. | Оценка | Станд. ош. |
| Constant (Постоянная) | 1,0918 (0,9801) | 0,1082 (0,1813) | 0,3998 | 0,09531 | | |
| Age (Возраст) | 0,0180 (0,01873) | 0,0013 (0,00198) | 0,02208 | 0,001220 | 0,04845 | 0,003511 |
| Education (Образование) | -0,0473 (-0,03613) | 0,0067 (0,01228) | -0,04507 | 0,006262 | -0,05437 | 0,03721 |
| Доход (Income) | -0,4687 (-0,5911) | 0,0726 (0,1282) | -0,1959 | 0,06103 | -0,1982 | 0,09127 |
| Дети (Kids) | -0,1569 (-0,1692) | 0,0306 (0,04882) | -0,1242 | 0,02336 | -0,002543 | 0,03687 |

^a Оценка $\sigma_u = 0,9542921$.

Пример 14.17. Модель скрытого класса для использования медицинского обслуживания

В примерах 7.6 и 11.16 мы рассматривали экспоненциальную модель регрессии

$$y_{it} = \text{DocVis}_{it} = \exp(\mathbf{x}'_{it}\beta) + \varepsilon_{it}$$

для переменной DocVis , обозначающей число посещений врача, по данным немецкой системы здравоохранения. (См. пример 11.13, в котором эта модель рассмотрена более подробно.) Результаты регрессии для спецификации

$$\mathbf{x}_{it} = (1, \text{Age}_{it}, \text{Education}_{it}, \text{Income}_{it}, \text{Kids}_{it})$$

для удобства еще раз представлены в табл. 14.15 (в скобках). Оценка нелинейным методом наименьших квадратов является полупараметрической: для нее не делается никаких предположений о распределении DocVis_{it} или ε_{it} . Мы наблюдаем сильное возрастание стандартных ошибок при использовании «кластерной устойчивой» асимптотической ковариационной матрицы. (Оценки представлены в примере 11.16.) На данном этапе мы предполагаем, что оценка нелинейным методом наименьших квадратов остается состоятельной при наличии корреляции между наблюдениями по индивидам. Учитывая способ спецификации модели, т.е. только в терминах функции условного среднего, это, возможно, является разумным. Для обобщения требуется рассматривать нелинейную обобщенную модель регрессии, а не обычную нелинейную регрессию.

В примере 14.10 мы сузили эту модель, предположив, что наблюдения посещений врача порождались геометрическим распределением

$$f(y_i | \mathbf{x}_i) = \theta_i(1 - \theta_i)^{y_i}, \theta_i = 1/(1 + \lambda_i), \lambda_i = \exp(\mathbf{x}'_i\beta), y_i = 0, 1, \dots$$

Условное среднее все еще равняется $\exp(\mathbf{x}'_{it}\beta)$, но эта спецификация накладывает структуру конкретного распределения для исходов. Объединенная модель пула была оценена в примере 14.10. В примере 14.14 в модель были добавлены предположения о случайных, а затем и о фиксированных эффектах. Тогда модель выглядит следующими образом:

$$f(y_{it} | \mathbf{x}_{it}) = \theta_{it}(1 - \theta_{it})^{y_{it}}, \theta_{it} = 1/(1 + \lambda_{it}), \lambda_{it} = \exp(c_i + \mathbf{x}'_{it}\beta), y_{it} = 0, 1, \dots$$

Оценки пула, модели со случайными и модели с постоянными эффектами показаны в табл. 14.15. Оценки пула, в которых стандартные ошибки скорректированы с учетом группировки панельных данных, сравнимы с оценками нелинейным методом наименьших квадратов с устойчивыми стандартными ошибками. Оценки параметров очень похожи: обе оценки состоятельны, а мы имеем дело с очень большой выборкой. Меньшие стандартные ошибки, наблюдаемые для ML-оценки, представляют собой следствие более детальной спецификации.

Сейчас мы ослабим спецификацию, предположив в конечной модели смеси с двумя классами. Мы также предположим, что вероятности классов есть функции от пола и семейного положения. Для спецификации модели скрытого класса

$$\text{Prob}(\text{class}_i = 1 | \mathbf{z}_i) = \Lambda(\theta_1 + \theta_2 \text{Female}_i + \theta_3 \text{Married}_i).$$

Структура модели, как и ранее, представляет собой геометрическую регрессию. Оценки параметров этой модели показаны в табл. 14.16. См. раздел E3.7, в котором обсуждаются способы оценивания.

Таблица 14.16. Оцененная модель линейной регрессии скрытого класса для GPA

| Параметр | Один класс | | Скрытый класс 1 | | Скрытый класс 2 | |
|-------------------------|------------|------------|-----------------|------------|-----------------|------------|
| | Оценка | Станд. ош. | Оценка | Станд. ош. | Оценка | Станд. ош. |
| β_1 | 1,0918 | 0,1082 | 1,6423 | 0,05351 | -0,3344 | 0,09288 |
| β_2 | 0,0180 | 0,0013 | 0,01691 | 0,0007324 | 0,02649 | 0,001248 |
| β_3 | -0,0473 | 0,0067 | -0,04473 | 0,003451 | -0,06502 | 0,005739 |
| β_4 | -0,4687 | 0,0726 | -0,4567 | 0,04688 | 0,01395 | 0,06964 |
| β_5 | -0,1569 | 0,0306 | -0,1177 | 0,01611 | -0,1388 | 0,02738 |
| θ_1 | 0,0000 | 0,0000 | -0,4280 | 0,06938 | 0,0000 | 0,0000 |
| θ_2 | 0,0000 | 0,0000 | 0,8255 | 0,06322 | 0,0000 | 0,0000 |
| θ_3 | 0,0000 | 0,0000 | -0,07829 | 0,07143 | 0,0000 | 0,0000 |
| $\text{Prob} \bar{z}$ | 1,0000 | | 0,47697 | | 0,52303 | |
| $\ln L$ | -61917,97 | | | -58708,63 | | |

В работе Деба и Триведи (Deb, Trivedi (2002)) авторы предположили, что значимое различие между группами индивидов, использующих систему здравоохранения, будет между «нечастыми» и «частыми» пользователями. Чтобы исследовать, учитывает ли наша модель скрытого класса это различие в данных, мы использовали (14-102) для предсказания отношения индивида к тому или иному классу (класс 1 или 2). Затем мы линейно регрессируем DocVis_{it} на константу и дамми-переменную для класса 2. Результаты получились следующие:

$$\text{DocVis}_{it} = 5,8034 (0,0465) - 4,7801 (0,06282) \text{Class2}_i + e_{it},$$

где в скобках указаны оцененные стандартные ошибки. Из этой линейной регрессии можно предположить, что фиктивная переменная, обозначающая отношение к классу, достаточно строго разделяет наблюдения на частых и нечастых пользователей. Результаты этой регрессии резюмируются путем описательной статистики в табл. 14.17.

Таблица 14.17. Описательная статистика для числа посещений врача

| Класс | Среднее значение | Среднеквадратическое отклонение |
|-----------------------|------------------|---------------------------------|
| Все, $n = 27,326$ | 3,18352 | 7,47579 |
| Класс 1, $n = 12,349$ | 5,80347 | 1,63076 |
| Класс 2, $n = 14,977$ | 1,02330 | 3,18352 |

14.11. Заключение и выводы

В данной главе представлены теория оценивания методом максимального правдоподобия, который является наиболее используемым в эконометрике после метода наименьших квадратов, а также ряд примеров. Оценки максимального правдоподобия состоятельны, асимптотически нормальны и являются наиболее эффективными среди всех оценок, обладающих этими свойствами. Недостаток данного способа оценивания состоит в том, что его использование требует полной параметрической, детальной спецификации процесса порождения данных. Таким образом, он чувствителен к проблемам, связанным с неверной спецификацией. В предыдущей главе мы рассматривали методы GMM-оценивания, которые являются менее параметрическими, но зато более устойчивыми к изменениям в лежащем в основе модели процессе порождения данных. Вместе ML- и GMM-оценивание преобладает в эмпирическом оценивании в эконометрике.

Основные термины и понятия

| | |
|--|---|
| AIC — Информационный критерий Акаике | Finite mixture model — Модель конечной смеси |
| Asymptotic efficiency — Асимптотическая эффективность | Fixed effects — Фиксированные эффекты |
| Asymptotic normality — Асимптотическая нормальность | Full information maximum likelihood (FIML) — Метод максимального правдоподобия при полной информации |
| Asymptotic variance — Асимптотическая дисперсия | Gauss-Hermite quadrature — Квадратура Гаусса-Эрмита |
| Autocorrelation — Автокорреляция | Generalized sum of squares — Обобщенная сумма квадратов |
| Bayes's theorem — Теорема Байеса | Geometric regression — Геометрическая регрессия |
| ВННН estimator — ВННН-оценка | GMM estimator — GMM-оценка |
| BIC — Байесовский (Шварца) информационный критерий | Identification — Идентификация |
| Butler and Moffitt's method — Метод Батлера и Моффитта | Incidental parameters problem — Проблема мешающих параметров |
| Cluster estimator — Кластерная оценка | Index function model — Модель индексной функции |
| Concentrated log-likelihood — Концентрированное логарифмическое правдоподобие | Information matrix equality — Информационное равенство |
| Conditional likelihood — Концентрированное правдоподобие | Invariance — Инвариантность |
| Consistency — Состоятельность | Jacobian — Якобиан |
| Cramér-Rao lower bound — Нижняя граница Крамера-Рао | Kullback-Leibler information criterion — Информационный критерий Кульбака-Лейблера |
| Efficient score — Эффективный скор | Latent regression — Скрытая регрессия |
| Exclusion restriction — Исключающее ограничение | Lagrange multiplier statistic — Статистика множителей Лагранжа |
| Exponential regression model — Экспоненциальная модель регрессии | Lagrange multiplier (LM) test — Тест множителей Лагранжа |

- Latent class model** — Модель скрытых классов
- Latent class linear regression model** — Модель линейной регрессии со скрытыми классами
- Likelihood equation** — Уравнение правдоподобия
- Likelihood function** — Функция правдоподобия
- Likelihood inequality** — Неравенство правдоподобия
- Likelihood ratio** — Отношение правдоподобий
- Likelihood ratio index** — Индекс отношения правдоподобия
- Likelihood ratio statistic** — Статистика отношения правдоподобий
- Likelihood ratio (LR) test** — Тест отношения правдоподобий
- Logistic probability model** — Логистическая вероятностная модель
- Loglinear conditional mean** — Логлинейная функция условного среднего
- Maximum likelihood** — Максимальное правдоподобие
- Maximum likelihood estimate** — Оценка методом максимального правдоподобия
- Maximum likelihood estimator** — Оценка максимального правдоподобия
- M-estimator** — *M*-оценка
- Method of scoring** — Метод сора
- Murphy and Topel estimator** — Оценка Мерфи–Топела
- Newton's method** — Метод Ньютона
- Noncentral chi-squared distribution** — Нецентральное распределение хи-квадрат
- Nonlinear least squares** — Нелинейный метод наименьших квадратов
- Nonnested models** — Невложенные модели
- Normalization** — Нормализация
- Oberhofer-Kmenta estimator** — Оценка Оберхофера–Кменты
- Outer product of gradients estimator (OPG)** — Оценка внешнего произведения градиентов
- Precision parameter** — Параметр точности
- Pseudo-log-likelihood function** — Функция логарифмического квазиправдоподобия
- Pseudo MLE** — КвазиML-оценка
- Pseudo R squared** — Псевдо*R*-квадрат
- Quadrature** — Квадратура
- Random effects** — Случайные эффекты
- Regularity conditions** — Условия регулярности
- Sandwich estimator** — Сэндвич-оценка
- Score test** — Скор-тест
- Score vector** — Скор-вектор, градиент
- Two-step maximum likelihood estimation** — Двухшаговый метод максимального правдоподобия
- Wald statistic** — Статистика Вальда
- Wald test** — Тест Вальда
- Vuong test** — Тест Вуонга

Упражнения

1. Пусть распределение x выражается как $f(x) = 1/\theta, 0 \leq x \leq \theta$. При условии случайности выборки из этого распределения докажите, что выборочный максимум является состоятельной оценкой θ . *Обратите внимание*, что вы можете доказать то, что максимум — это оценка максимального правдоподобия для θ . Но обычные свойства здесь неприменимы. Почему? (*Подсказка*: попробуйте проверить, что математическое ожидание первой производной логарифмического правдоподобия по θ равно нулю.)
2. При условии случайности выборки из экспоненциального распределения $f(x) = (1/\theta)e^{-x/\theta}, x \geq 0, \theta > 0$ найдите оценку максимального правдоподобия для параметра θ и выведите асимптотическое распределение этой оценки.
3. *Смесь распределений*. Пусть совместное распределение двух случайных

величин x и y выражается как

$$f(x, y) = \frac{\theta e^{-(\beta+\theta)y} (\beta y)^x}{x!}, \quad \beta, \theta > 0, y \geq 0, x = 0, 1, 2, \dots$$

- Найдите оценки максимального правдоподобия β и θ , а также их совместное асимптотическое распределение.
- Найдите оценку максимального правдоподобия $\theta/(\beta+\theta)$ и ее асимптотическое распределение.
- Докажите, что $f(x)$ есть функция вида

$$f(x) = \gamma(1 - \gamma)^x, \quad x = 0, 1, 2, \dots,$$

и найдите оценку максимального правдоподобия для γ и ее асимптотическое распределение.

- Докажите, что $f(y|x)$ есть функция вида

$$f(y|x) = \frac{\lambda e^{-\lambda y} (\lambda y)^x}{x!}, \quad y \geq 0, \lambda > 0.$$

Докажите, что интеграл от $f(y|x)$ равен 1. Найдите оценку максимального правдоподобия λ и ее асимптотическое распределение. (Подсказка: в условном распределении просто рассматривайте x как постоянные.)

- Докажите, что

$$f(y) = \theta e^{-\theta y}, \quad y \geq 0, \quad \theta > 0.$$

Найдите оценку максимального правдоподобия θ и ее асимптотическую дисперсию.

- Докажите, что

$$f(x|y) = \frac{e^{-\beta y} (\beta y)^x}{x!}, \quad x = 0, 1, 2, \dots, \beta > 0.$$

Согласно этому распределению, что является оценкой максимального правдоподобия β ?

- Пусть x имеет распределение Вейбулла

$$f(x) = \alpha \beta x^{\beta-1} e^{-\alpha x^\beta}, \quad x \geq 0, \alpha, \beta > 0.$$

- Выведите логарифмическую функцию правдоподобия для случайной выборки из n наблюдений.
- Выведите уравнения правдоподобия для оценки максимального правдоподобия для α и β . Обратите внимание, что первое из них дает решение в явном виде для α при известных данных и β . Но после подстановки во второе уравнение мы получаем решение для β лишь в неявном виде. Как бы вы получили оценки максимального правдоподобия?

- c. Выведите матрицу вторых производных логарифмического правдоподобия по α и β . Точные математические ожидания элементов, включающих β , включают производные гамма-функции, и их сложно вычислить аналитически. Конечно, ваш точный результат даст эмпирическую оценку. Как бы вы оценили асимптотическую ковариационную матрицу для ваших оценок из пункта b?
- d. Докажите, что $\alpha\beta \text{Cov}[\ln x, x^\beta] = 1$. (Подсказка: математические ожидания первых производных логарифмической функции правдоподобия равны нулю.)

5. Следующие данные получены из распределения Вейбулла из упражнения 4:

| | | | | | | |
|---------|---------|---------|--------|---------|---------|---------|
| 1,3043 | 0,49254 | 1,2742 | 1,4019 | 0,32556 | 0,29965 | 0,26423 |
| 1,0878 | 1,9461 | 0,47615 | 3,6454 | 0,15344 | 1,2357 | 0,96381 |
| 0,33453 | 1,1227 | 2,0296 | 1,2797 | 0,96080 | 2,0070 | |

- a. Получите оценки максимального правдоподобия для α и β и оцените асимптотическую ковариационную матрицу для оценок.
- b. Примените критерий Вальда для проверки гипотезы $\beta = 1$.
- c. Получите оценку максимального правдоподобия α в условиях гипотезы о том, что $\beta = 1$.
- d. Используя результаты частей a и b, примените критерий отношения правдоподобий для проверки гипотезы $\beta = 1$.
- e. Примените критерий множителей Лагранжа для проверки гипотезы $\beta = 1$.

6. **Оценивание методом максимального правдоподобия при ограниченной информации** (Limited Information Maximum Likelihood Estimation). Рассмотрим двумерное распределение для x и y , которое является функцией двух параметров α и β . Совместная плотность может быть записана как $f(x, y | \alpha, \beta)$. Нас интересует оценка двух параметров методом максимального правдоподобия. Оценка максимального правдоподобия с полной информацией теперь представляет собой хорошо известную нам оценку максимального правдоподобия для двух параметров. Предположим, что мы можем разложить на множители совместное распределение, как это сделано в упражнении 3, и получить $f(x, y | \alpha, \beta) = f(y | x, \alpha, \beta)f(x | \alpha)$. Иными словами, условная плотность y является функцией обоих параметров, но одномерное безусловное распределение x включает только α .

- a. Выпишите общий вид логарифмической функции правдоподобия, используя совместную плотность.
- b. Поскольку совместная плотность равна произведению условной и одномерной плотностей, логарифмическая функция правдоподобия может быть аналогично записана в терминах разложенной плотности. Запишите это в общем виде.

- с. Параметр α может быть оценен с использованием лишь данных x и логарифмического правдоподобия, выписанного лишь через одномерную (безусловную) плотность x . Он также может быть оценен с помощью β , если использовать полную логарифмическую функцию правдоподобия и данные как y , так x . Покажите это.
- d. Покажите, что первая оценка в части с имеет бóльшую асимптотическую дисперсию по сравнению со второй. В этом заключается разница между оценкой максимального правдоподобия с ограниченной информацией и оценкой максимального правдоподобия с полной информацией.
- е. Покажите, что если $\partial^2 \ln f(y | x, \alpha, \beta) / \partial \alpha \partial \beta = 0$, то результат, полученный в пункте d, более не является корректным.
7. Покажите, что неравенство правдоподобия в теореме 14.3 выполняется для распределения Пуассона, которое мы использовали в разделе 14.3, показав, что $E[(1/n) \ln L(\theta | y)]$ имеет единственный максимум в точке $\theta = \theta_0$. (Подсказка: сначала покажите, что математическое ожидание равно $-\theta + \theta_0 \ln \theta - E_0[\ln y_i!]$.)
8. Покажите, что неравенство правдоподобия в теореме 14.3 выполняется для нормального распределения.
9. При условии случайности выборки из классической регрессионной модели, задаваемой (14-3), перепараметризируйте функцию правдоподобия в терминах $\eta = 1/\sigma$ и $\delta = (1/\sigma)\beta$. Найдите оценки максимального правдоподобия η и δ и выведите асимптотическую ковариационную матрицу оценок этих параметров.
10. Рассмотрим выборку из многомерного нормального распределения с вектором математического ожидания $\mu = (\mu_1, \mu_2, \dots, \mu_M)$ и ковариационной матрицей $\sigma^2 \mathbf{I}$. Логарифмическая функция правдоподобия имеет вид

$$\ln L = \frac{-nM}{2} \ln(2\pi) - \frac{nM}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (\mathbf{y}_i - \mu)' (\mathbf{y}_i - \mu).$$

Покажите, что оценки максимального правдоподобия параметров равны $\hat{\mu} = \bar{y}_m$ и

$$\hat{\sigma}_{\text{ML}}^2 = \frac{\sum_{i=1}^n \sum_{m=1}^M (y_{im} - \bar{y}_m)^2}{nM} = \frac{1}{M} \sum_{m=1}^M \frac{1}{n} \sum_{i=1}^n (y_{im} - \bar{y}_m)^2 = \frac{1}{M} \sum_{m=1}^M \hat{\sigma}_m^2.$$

Выведите матрицу вторых производных и покажите, что асимптотическая ковариационная матрица для оценок максимального правдоподобия есть

$$\left\{ -E \left[\frac{\partial^2 \ln L}{\partial \theta \partial \theta'} \right] \right\}^{-1} = \begin{bmatrix} \sigma^2 \mathbf{I}/n & \mathbf{0} \\ \mathbf{0} & 2\sigma^4/(nM) \end{bmatrix}.$$

Предположим, что мы хотим проверить гипотезу, что средние значения всех M -распределений равны конкретному значению μ^0 . Покажите, что статистика Вальда будет иметь вид

$$\mathbf{W} = (\bar{\mathbf{y}} - \mu^0 \mathbf{i})' \left(\frac{\hat{\sigma}^2}{n} \mathbf{I} \right)^{-1} (\bar{\mathbf{y}} - \mu^0 \mathbf{i}) = \left(\frac{n}{s^2} \right) (\bar{\mathbf{y}} - \mu^0 \mathbf{i})' (\bar{\mathbf{y}} - \mu^0 \mathbf{i}),$$

где \bar{y} — вектор выборочных средних.

11. Докажите результат, сформулированный в примере 4.7.

Эмпирические приложения

1. **Бинарный выбор** (Binary Choice). Эта модель будет основываться на данных о здравоохранении, исследуемых в примере 14.17 и некоторых других примерах. Подробности получения данных рассматривались в примере 11.16. Нас интересует анализ зависимой переменной y_{it} , которая принимает значения 1 и 0 с вероятностями $F(\mathbf{x}'_{it}\beta)$ и $1 - F(\mathbf{x}'_{it}\beta)$, где F есть функция, определяющая вероятность. Зависимая переменная y_{it} получается из количественной переменной $DocVis$, которая равна числу посещений врача в данном году. Бинарная переменная строится как

$$y_{it} = 1, \text{ {если } } DocVis_{it} > 0, 0 \text{ в противном случае.}$$

Мы будем строить модель для вероятности того, что y_{it} равно единице. Интересующими нас независимыми переменными будут

$$\mathbf{x}_{it} = (1, age_{it}, educ_{it}, female_{it}, married_{it}, hsat_{it}).$$

а. В соответствии с моделью теоретическая плотность для y_{it} равна

$$f(y_{it} | \mathbf{x}_{it}) = F(\mathbf{x}'_{it}\beta) \text{ для } y_{it} = 1 \text{ и } 1 - F(\mathbf{x}'_{it}\beta) \text{ для } y_{it} = 0.$$

Мы предположим, что нам подходит логит-модель (см. раздел 17.2), т. е.

$$F(\mathbf{x}'_{it}\beta) = \Lambda(\mathbf{x}'_{it}\beta) = \frac{\exp(\mathbf{x}'_{it}\beta)}{1 + \exp(\mathbf{x}'_{it}\beta)}.$$

Покажите, что для двух исходов вероятности могут быть объединены в функцию плотности

$$f(y_{it} | \mathbf{x}_{it}) = g(y_{it}, \mathbf{x}_{it}, \beta) = \Lambda[(2y_{it} - 1)\mathbf{x}'_{it}\beta].$$

Теперь используйте этот результат, чтобы построить логарифмическую функцию правдоподобия для имеющейся выборки данных $(y_{it}, \mathbf{x}_{it})$. (Замечание: мы будем игнорировать панельный аспект имеющейся выборки. Постройте модель, как если бы это были межобъектные данные.)

- Выведите уравнения правдоподобия для оценки β .
- Выведите матрицу вторых производных для логарифмической функции правдоподобия. (Подсказка: при дифференцировании пригодится выражение: $d\Lambda(t)/dt = \Lambda(t)[1 - \Lambda(t)]$.)
- Покажите, как использовать метод Ньютона для оценки параметров модели.
- Отличается ли метод скоры от метода Ньютона? Выведите отрицательное математическое ожидание матрицы вторых производных.
- Получите оценки максимального правдоподобия параметров для указанных данных и переменных. Оформите ваши результаты: оценки, стандартные ошибки и т.д., а также значение логарифмического правдоподобия.

- g. Проверьте гипотезу о том, что коэффициенты при переменной *female* и переменной *married* являются нулевыми. Покажите, как тестировать эту гипотезу, используя критерий Вальда, LM-критерий и LR-критерий, а затем проведите эти тесты.
- h. Проверьте гипотезу о том, что все коэффициенты в модели, кроме константы, равны нулю.

Учебное издание

Серия «Академический учебник»

Уильям Г. Грин

Эконометрический анализ

Книга 1

Главный редактор *В.В. Анашвили*
Заведующая редакцией *Ю.В. Бандурина*
Выпускающий редактор *Е. В. Попова*
Редактор *Ф. Н. Морозова*
Художник *Е. Н. Спасская*
Оригинал-макет *О. З. Эллов*
Верстка *Н. А. Кильдишева*

Подписано в печать 15.09.2016. Формат 70x108¹/₁₆
Гарнитура PT Serif Pro. Усл. печ. л. 66,4.
Тираж 1000 экз. Изд. № 166. Заказ № 6633

Издательский дом «Дело» РАНХиГС
119571, Москва, пр-т Вернадского, 82
Коммерческий центр — тел. (495) 433-2510, (495) 433-2502
www.ranepa.ru
delo@ranepa.ru

Отпечатано в АО «Первая Образцовая типография»
Филиал «Чеховский Печатный Двор»
142300, Московская область, г. Чехов, ул. Полиграфистов, д. 1
Сайт: www.chpd.ru, E-mail: sales@chpd.ru, тел. 8(499)270-73-59

ИЗДАТЕЛЬСКИЙ ДОМ

«ДЕЛО» РАНХиГС

Москва, проспект Вернадского, д. 82

Тел.: (495) 433-25-02

delo@ranepa.ru

В серии «Академический учебник» вышли в свет:

Андреу Мас-Колелл, Майкл Д. Уинстон, Джерри Р. Грин. Микроэкономическая теория: в 2 кн. / пер. с англ.; под науч. ред. М. Левина и Е. Покатович.

Оливье Бланшар, Стэнли Фишер. Лекции по макроэкономике / пер. с англ. под науч. ред. Е. Андреевой, Н. Ранневой.

Морис Обсфельд, Кеннет Рогофф. Основы международной макроэкономики / пер. с англ.; под науч. ред. С. Дробышевского, П. Трунина.

Джеймс Сток, Марк Уотсон. Введение в эконометрику / пер. с англ.; под науч. ред. М. Турунцевой.

Майкл Уикенс. Макроэкономическая теория: подход динамического общего равновесия / пер. с англ. под науч. ред. Е. Синельниковой.

Карл Уолш. Монетарная теория и монетарная политика / пер. с англ.; под науч. ред. М. Карева.

Э. Колин Кэмерон, Правин К. Триведи. Микроэконометрика: методы и их применения: в 2 кн. / пер. с англ. под науч. ред. Б. Демешева.