РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА
И ГОСУДАРСТВЕННОЙ СЛУЖБЫ
ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ

АКАДЕМИЧЕСКИЙ УЧЕБНИК

Фумио Хайяши

ЭКОНОМЕТРИКА

Издательский дом ДЕЛО

Фумио Хайяши

ЭКОНОМЕТРИКА

Перевод с английского
под научной редакцией В. П. Носко

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 ◦ 2017

'-151

УДК 330.4
ББК 65.05
Х 12

Перевод с английского:
А.А. Скроботов (гл. 2, 5, 6, 9, 10), В.Д. Петренко (гл. 1, 8), В.Е. Зямалов (гл. 3),
Ю.К. Ачкасов (гл. 7), А.В. Зубарев, В.Д. Петренко (гл. 4)

Хайяши, Фумио

Х 12 Эконометрика / Фумио Хайяши; пер. с англ. под науч. ред. В.П. Носко. — М.: Издательский дом «Дело» РАНХиГС, 2017. — 728 с. — (Академический учебник).

ISBN 978-5-7749-1197-4

Эконометрика Хайяши обещает стать очередным большим резюме современной эконометрики. Она вводит аспирантов первого года в стандартный материал аспирантских курсов эконометрики, рассматриваемый с современной точки зрения. Исследование охватывает весь стандартный материал, необходимый для понимания основных методов эконометрики, от обычного метода наименьших квадратов до коинтеграции. Книга совершенно самобытна в разработке и анализа временных рядов, и кросс-секционного анализа, предоставляя читателю единую структуру для понимания и интегрирования результатов.

Эконометрика имеет много полезных особенностей и в сжатой форме охватывает все важные темы в эконометрике. Восемь из десяти глав включают серьезное эмпирическое применение, взятое из экономики труда, экономики отраслевых рынков, внутренних и международных финансов и макроэконометрики. Все результаты формулируются как утверждения, так что студенты понимают обсуждаемую тему, а также условия, при которых соответствующие результаты имеют силу. Большинство утверждений доказывается в тексте.

Для тех, кто собирается писать диссертацию по прикладной тематике, эмпирические применения, собранные в книге, являются хорошим способом обучиться тому, как следует проводить эмпирическое исследование. Для склонных к теории бескомпромиссное рассмотрение основных методов является хорошей подготовкой к более продвинутым теоретическим курсам.

УДК 330.4
ББК 65.05

ISBN 978-5-7749-1197-4

Содержание

Предисловие **17**

Обязательные предварительные требования 17

Структура книги 18

Построение курса на базе этой книги 19

Контрольные вопросы и аналитические упражнения 19

Эмпирические упражнения 20

Математические обозначения 21

Благодарности 21

Предисловие к русскому изданию **22**

Глава 1. Свойства OLS в конечных выборках **24**

1.1. Классическая линейная модель регрессии 25

Предположение линейности 25

Матричные обозначения 27

Предположение о строгой экзогенности 28

Следствия строгой экзогенности 29

Строгая экзогенность в моделях временных рядов 31

Остальные предположения модели 32

Классическая регрессионная модель для случайных выборок 34

«Фиксированные» регрессоры 34

1.2. Алгебра метода наименьших квадратов 36

OLS минимизирует сумму квадратов остатков 36

Нормальные уравнения 37

Два выражения для OLS-оценки 39

Еще несколько терминов из алгебры 40

Анализ влиятельных наблюдений (дополнительно) 43

Замечание относительно вычисления OLS-оценок 46

1.3. Свойства OLS в конечных выборках 48

Распределение $\boldsymbol{b}$ в конечных выборках 48

Свойства s^2 в конечных выборках 52

Оценка $\mathrm{Var}(\boldsymbol{b}|\boldsymbol{X})$ 53
1.4. Тестирование гипотез в предположении нормальности 55
Нормально распределенные ошибки 55
Тестирование гипотез об отдельных коэффициентах регрессии 57
Правило принятия решений на основании t-теста 59
Доверительный интервал 60
p-значение 61
Линейные гипотезы 61
F-тест 63
Более удобное выражение для F 65
Сравнение t и F 66
Пример тестовой статистики, распределение которой зависит от $\boldsymbol{X}$ 68
1.5. Связь с методом максимального правдоподобия 69
Метод максимального правдоподобия 70
Условная и безусловная функции правдоподобия 70
Логарифмическое правдоподобие для регрессионной модели 71
Максимизация правдоподобия с помощью концентрированной функции правдоподобия 71
Граница Крамера — Рао для классической регрессионной модели 73
F-тест как тест отношения правдоподобий 76
Метод квазимаксимального правдоподобия 76
1.6. Обобщенный метод наименьших квадратов (GLS) 78
Последствия ослабления предположения 1.4 79
Эффективное оценивание при известной матрице V 79
Частный случай: взвешенный метод наименьших квадратов (WLS) 81
Ограничивающее свойство GLS 82
1.7. Приложение: отдача от масштаба в производстве и распределении электроэнергии 84
Отрасль производства и распределения электроэнергии 84
Данные 84
Для чего нам нужна эконометрика 85
Технология Кобба — Дугласа 86
Откуда мы знаем, что действительно имеем дело с технологией Кобба — Дугласа? 88
Выполнены ли предположения OLS? 88
Метод наименьших квадратов с ограничениями 89
Тестирование однородности функции издержек 90

Лирическое отступление: об осторожности с интерпретацией R^2 91
Тестирование постоянной отдачи от масштаба 92
Важность графического анализа остатков 93
Дальнейшая разработка темы 94
Набор задач для главы 1 96
Ответы на избранные вопросы 110

Глава 2. Теория больших выборок 114

2.1. Обзор предельных теорем для последовательностей случайных величин 115
Различные виды сходимости 115
Три полезных результата 118
Взгляд на оценки как на последовательности случайных величин 121
Законы больших чисел и центральные предельные теоремы 122
2.2. Фундаментальные понятия в анализе временных рядов 124
Необходимость эргодической стационарности 124
Различные классы случайных процессов 125
Другая формулировка отсутствия сериальной зависимости 133
CLT для эргодических стационарных последовательностей мартингал-разностей 134
2.3. Распределение OLS-оценки на больших выборках 136
Модель 137
Асимптотическое распределение OLS-оценки 141
s^2 состоятельна 143
2.4. Тестирование гипотез 145
Тестирование линейных гипотез 145
Тест является состоятельным 147
Асимптотическая мощность 148
Тестирование нелинейных гипотез 150
2.5. Состоятельное оценивание $\mathrm{E}(\varepsilon_i^2 \boldsymbol{x}_i \boldsymbol{x}_i')$ 151
Использование остатков вместо ошибок 151
Представление $\boldsymbol{S}$ в терминах матриц данных 153
Рассмотрение конечных выборок 154
2.6. Последствия условной гомоскедастичности 154
Противопоставление условной и безусловной гомоскедастичности 155
Редукция к формулам для конечных выборок 155
Распределение на больших выборках t- и F-статистик 156

Варианты асимптотических тестов при условной гомоскедастичности 157
2.7. Тестирование условной гомоскедастичности 159
2.8. Оценивание параметризованной условной гетероскедастичности (дополнительно) 161
Функциональная форма 162
WLS с известным α 163
Регрессия e_i^2 на $\mathbf{z}_i$ дает состоятельную оценку α 164
WLS с оцененным α 164
Противопоставление OLS и WLS 165
2.9. Проекция методом наименьших квадратов 166
Оптимальное предсказание значения зависимой переменной 166
Наилучший линейный предиктор 167
OLS состоятельно оценивает коэффициенты проекции 169
2.10. Тестирование сериальной корреляции 170
Статистики Бокса — Пирса и Льюнга — Бокса 171
Выборочные автокорреляции, вычисленные с помощью остатков 173
Тестирование с предопределенными, но не строго экзогенными регрессорами 175
Тест, основанный на вспомогательной регрессии 176
2.11. Приложение: эконометрика рациональных ожиданий 179
Гипотеза эффективности рынка 179
Тестируемые следствия 181
Тестирование на сериальную корреляцию 183
Является ли номинальная процентная ставка оптимальным предиктором? 185
R_t не является строго экзогенной 187
Последующее развитие 188
2.12. Регрессия на время 190
Асимптотическое распределение OLS-оценки 190
Тестирование гипотез в регрессиях на время 192
Приложение 2.А. Асимптотика с фиксированными регрессорами 194
Приложение 2.В. Доказательство утверждения 2.10 194
Набор задач для главы 2 197
Ответы на избранные вопросы 213

Глава 3. Одномерный GMM 217

3.1. Смещение из-за эндогенности: пример Уоркинга 218
Система одновременных уравнений рыночного равновесия 218

Смещение из-за эндогенности 219
Наблюдаемые сдвиги предложения 220
3.2. Другие примеры 223
Простая макроэкономическая модель 224
Ошибки в переменных 224
Производственная функция 226
3.3. Общая постановка задачи 228
Регрессоры и инструменты 228
Идентификация 230
Порядковое условие идентифицируемости 232
Предположение асимптотической нормальности 232
3.4. Обобщенный метод моментов (GMM) 234
Метод моментов 235
Обобщенный метод моментов 236
Ошибка оценки 237
3.5. Асимптотические свойства GMM 238
Асимптотическое распределение GMM-оценки 238
Оценивание дисперсии ошибки 239
Тестирование гипотез 240
Оценивание $\boldsymbol{S}$ 241
Эффективная GMM-оценка 242
Асимптотическая мощность 243
Свойства в малых выборках 244
3.6. Тестирование сверхидентифицирующих ограничений 246
Тестирование подмножеств условий ортогональности 248
3.7. Тестирование гипотез по принципу отношения правдоподобий 251
LR-статистика для модели регрессии 253
Тест добавления переменной (необязательный материал) 253
3.8. Приложения условной гомоскедастичности 255
Эффективный GMM становится 2SLS 256
J становится статистикой Саргана 257
Свойства 2SLS на конечных выборках 258
Альтернативные выводы 2SLS 259
Когда регрессоры предопределены 261
Тестирование подмножества условий ортогональности 262
Тестирование условной гомоскедастичности 264
Тестирование на наличие автокорреляции 264
3.9. Применение: отдача от образования 265
Данные NLS-Y 266
Полулогарифмическое уравнение заработной платы 267
Смещение из-за невключения переменной 267

IQ как мера способностей ... 268
Ошибки в переменных ... 270
Коррекция смещения с помощью 2SLS ... 271
Дальнейшее развитие ... 272
Набор задач для главы 3 ... 273
Ответы на избранные вопросы ... 283

Глава 4. Обобщенный метод моментов для систем уравнений ... 287

4.1. Модель с несколькими уравнениями ... 288
Линейность ... 288
Стационарность и эргодичность ... 289
Условия ортогональности ... 291
Идентифицируемость ... 291
Предположение для асимптотической нормальности ... 293
Связь с «полной» системой одновременных уравнений ... 294
4.2. GMM для системы уравнений ... 294
4.3. Теория больших выборок ... 296
4.4. Сравнение оценивания одного уравнения и системы уравнений ... 299
Когда они «эквивалентны»? ... 300
Совместное оценивание может быть рискованным ... 302
4.5. Частные случаи GMM-оценивания системы уравнений: FIVE, 3SLS и SUR ... 303
Условная гомоскедастичность ... 303
Инструментальная оценка с полной информацией (FIVE) ... 304
Трехшаговый метод наименьших квадратов (3SLS) ... 305
Кажущиеся несвязанными регрессии (SUR) ... 308
SUR и OLS ... 310
4.6. Общие коэффициенты ... 314
Модель с общими коэффициентами ... 315
GMM-оценка ... 316
OLS в модели пула ... 319
Приведение формул к более красивому виду ... 320
Ограничение, которое не является ограничивающим ... 322
4.7. Приложение. Спрос на взаимосвязанные факторы ... 325
Транслогарифмическая функция издержек ... 325
Доли факторов ... 326
Эластичности замещения ... 327
Свойства функции издержек ... 328
Стохастические спецификации ... 329
Природа ограничений ... 330

Модель многомерной регрессии
с перекрестными ограничениями 331
Какое уравнение исключать 333
Результаты 334
Набор задач для главы 4 336
Ответы на избранные вопросы 349

Глава 5. Панельные данные 352

5.1. Модель компонент ошибки 353
Компоненты ошибки 354
Групповые средние 356
Репараметризация 357
5.2. Оценка фиксированных эффектов 359
Формула 359
Свойства на больших выборках 360
Отступление: когда η_i является сферической 362
Сравнение случайных эффектов с фиксированными 363
Ослабление условной гомоскедастичности 365
5.3. Несбалансированные панели (необязательный параграф) 367
Обнуление пропущенных наблюдений 367
Обнуление против сжатия 368
Отсутствие смещения вследствие отбора 369
5.4. Приложение: Международные различия в темпах роста 371
Вывод оцениваемого уравнения 371
Добавление ошибок 373
Трактовка α_i 373
Состоятельное оценивание скорости конвергенции 374
Приложение 5.А. Распределение статистики Хаусмана 376
Набор задач для главы 5 378
Ответы на избранные вопросы 392

Глава 6. Сериальная корреляция 394

6.1. Моделирование сериальной корреляции: линейные процессы 394
MA(q) 395
MA(∞) как предел в среднем квадратическом 395
Фильтры 399
Обращение лаговых полиномов 402
6.2. Процессы ARMA 405
AR(1) и его MA(∞)-представление 405
Автоковариации AR(1) 408
AR(p) и его MA(∞)-представление 408
ARMA($p.q$) 410

ARMA(p, q) с общими корнями ... 412
Обратимость ... 412
Производящая функция автоковариаций и спектр ... 413
6.3. Векторные процессы ... 416
6.4. Оценивание авторегрессий ... 421
Оценивание AR(1) ... 421
Оценивание AR(p) ... 423
Выбор глубины запаздываний ... 424
Оценивание VAR ... 427
Оценивание ARMA(p, q) ... 428
6.5. Асимптотика для выборочных средних сериально коррелированных процессов ... 430
LLN для ковариационно стационарных процессов ... 430
Две центральные предельные теоремы ... 431
Многомерное обобщение ... 434
6.6. Включение сериальной корреляции в GMM ... 436
Модель и асимптотические результаты ... 436
Оценивание $\boldsymbol{S}$, когда автоковариации обращаются в нуль после конечного количества лагов ... 437
Использование ядер для оценивания $\boldsymbol{S}$ ... 438
VARHAC ... 440
6.7. Оценивание при условной гомоскедастичности (дополнительно) ... 443
Ядерное оценивание $\boldsymbol{S}$ при условной гомоскедастичности ... 443
Представление оцененной долговременной дисперсии через матрицу данных ... 444
Связь с GLS ... 445
6.8. Приложение: Форвардные обменные курсы как оптимальные предикторы ... 448
Гипотеза эффективности рынка ... 449
Тестирование равенства нулю безусловного среднего ... 450
Регрессионные тесты ... 453
Набор задач для главы 6 ... 458
Ответы на избранные вопросы ... 471

Глава 7. Экстремальная оценка ... 475

7.1. Экстремальные оценки ... 476
«Измеримость» $\widehat{\boldsymbol{\theta}}$ ... 476
Два класса экстремальных оценок ... 477
Метод максимального правдоподобия (ML) ... 478
Метод условного максимального правдоподобия ... 481

Инвариантность оценок ML 483
Нелинейный метод наименьших квадратов (NLS) 484
Линейный и нелинейный GMM 485
7.2. Состоятельность 487
Две теоремы о состоятельности экстремальных оценок 487
Состоятельность M-оценок 490
Вогнутость после репараметризации 492
Идентификация в NLS и ML 493
NLS 493
ML 494
Состоятельность GMM 499
7.3. Асимптотическая нормальность 501
Асимптотическая нормальность M-оценок 502
Состоятельное оценивание асимптотической ковариационной матрицы 506
Асимптотическая нормальность условного ML 506
Два примера 509
Асимптотическая нормальность GMM 511
Сопоставление GMM с ML 514
Выражение для ошибки оценки в едином формате 515
7.4. Проверка гипотез 520
Нулевая гипотеза 520
Рабочие предпосылки 521
Статистика Вальда 522
Статистика множителей Лагранжа (LM-статистика) 524
Статистика отношения правдоподобий (LR-статистика) 526
Итоговые выводы для триады статистик 527
7.5. Численная оптимизация 529
Алгоритм Ньютона — Рафсона 530
Алгоритм Гаусса — Ньютона 530
Запись алгоритмов Ньютона — Рафсона и Гаусса — Ньютона в общем формате 531
Уравнения, нелинейные только по параметрам 531
Набор задач для главы 7 533
Ответы на избранные вопросы 537

Глава 8. Примеры применения метода максимального правдоподобия... 539

8.1. Модели с качественным откликом ... 539
Вклад и гессиан для наблюдения t ... 540
Состоятельность ... 541
Асимптотическая нормальность ... 542
8.2. Усеченные регрессионные модели ... 543
Модель ... 543
Усеченные распределения ... 544
Функция правдоподобия ... 546
Репараметризация функции правдоподобия ... 547
Проверка состоятельности и асимптотической нормальности ... 547
Восстановление исходных параметров ... 549
8.3. Цензурированные регрессионные (тобит) модели ... 550
Функция правдоподобия для тобит-модели ... 551
Репараметризация ... 552
8.4. Многомерные регрессии ... 554
Переформулирование модели многомерной регрессии ... 554
Функция правдоподобия ... 555
Максимизация функции правдоподобия ... 557
Состоятельность и асимптотическая нормальность ... 558
8.5. FIML ... 559
Система одновременных уравнений с общими инструментами: новые обозначения ... 559
Полная система одновременных уравнений ... 562
Взаимосвязь между $(\mathbf{\Gamma}_0, \mathbf{B}_0)$ и δ_0 ... 563
Функция правдоподобия для FIML ... 564
Концентрированная функция правдоподобия для FIML ... 565
Тестирование сверхидентифицирующих ограничений ... 566
Свойства FIML-оценки ... 567
ML-оценивание SUR-модели ... 569
8.6. LIML ... 571
Определение LIML ... 571
Вычисление LIML ... 573
Сравнение LIML и 2SLS ... 575
8.7. Сериально коррелированные наблюдения ... 576
Два вопроса ... 577
Безусловный ML для зависимых наблюдений ... 578
ML-оценка процесса AR(1) ... 579
Условное ML-оценивание AR(1)-процессов ... 580
Условное ML-оценивание процессов AR(p) и VAR(p) ... 583
Набор задач для главы 8 ... 585

Глава 9. Эконометрика единичного корня 591

9.1. Моделирование трендов 591
Интегрированные процессы 593
Почему важно знать, является ли рассматриваемый процесс I(1)-процессом 595
Что следует брать в качестве нулевой гипотезы, I(0) или I(1)? 597
Другие подходы к моделированию трендов 598
9.2. Инструментарий для эконометрики единичных корней 598
Линейные I(0)-процессы 599
Аппроксимация I(1)-процесса случайным блужданием 600
Связь с моделями ARMA 601
Винеровский процесс 602
Полезная лемма 605
9.3. Тесты Дики — Фуллера 608
AR(1)-модель 608
Получение предельного распределения при нулевой гипотезе I(1) 609
Включение постоянной составляющей 613
Включение временного тренда 616
9.4. Расширенные тесты Дики — Фуллера 621
Расширенная авторегрессия 621
Предельное распределение OLS-оценки 622
Получение тестовых статистик 625
Тестирование гипотез о ζ 627
Что делать, когда p неизвестно 627
Предложение для выбора $p_{\max}(T)$ 629
Включение в регрессию постоянной составляющей 630
Включение линейного тренда 633
Сводный обзор DF- и ADF-тестов и других тестов на единичный корень 635
9.5. Какой тест на единичный корень использовать 637
Асимптотика локальности к единице 637
Свойства в конечных выборках 638
9.6. Приложение: Паритет покупательной способности 639
Смущающая жизнеспособность модели случайного блуждания? 640
Набор задач для главы 9 641
Ответы на избранные вопросы 655

Глава 10. Коинтеграция 660

10.1. Коинтегрированные системы 661
Линейные векторные I(0)- и I(1)-процессы 662
Разложение Converse Бевериджа — Нельсона 665
Определение коинтеграции 666
10.2. Альтернативные представления коинтегрированных систем ... 671
Треугольная система Филлипса 671
VAR и коинтеграция 673
Векторная модель коррекции ошибок (VECM) 675
ML-процедура Йохансена 677
10.3. Тестирование нулевой гипотезы об отсутствии коинтеграции 680
Ложные регрессии 680
Тест на коинтеграцию, основанный на остатках 681
Тестирование нулевой гипотезы о коинтеграции 687
10.4. Статистические выводы о коинтегрирующих векторах 688
Оценка SOLS 688
Двумерный пример 690
Продолжение двумерного примера 691
Допущение сериальной корреляции 692
Общий случай 695
Другие оценки и свойства на конечных выборках 697
10.5. Приложение: Спрос на деньги в США 697
Данные 698
$(m - p, y, R)$ как коинтегрированная система 698
DOLS 700
Нестабильный спрос на деньги? 701
Набор задач для главы 10 704

Приложение А. Блочные (клеточные) матрицы и произведения Кронекера 708
Блочные (клеточные) матрицы 708
Сложение и умножение блочных (клеточных) матриц 708
Обращение блочных (клеточных) матриц 710
Произведения Кронекера 711

Предметный указатель 712

Предисловие

Эта книга призвана служить в качестве учебника по курсу эконометрики для студентов первого года аспирантуры. Она имеет две отличительные черты. Во-первых, охватывает целый спектр методов, используя в качестве организующего принципа так называемый обобщенный метод моментов. Я считаю, что такой унифицированный подход является наиболее эффективным способом изложить материал первого года в доступной, но еще строгой манере. Во-вторых, в большинстве глав имеется параграф, в котором детально рассматриваются оригинальные прикладные статьи из таких различных областей экономики, как экономика отраслевых рынков, труд, финансы, международная экономика и макроэкономика. Таким образом, читатель будет знать, как *использовать* методы, рассматриваемые в соответствующей главе, и при каких условиях они применимы.

Конспекты лекций, на основе которых базируется эта книга, использовались на протяжении нескольких последних лет в Университете Пенсильвании, Колумбийском университете, Принстонском университете, Университете Токио, Бостонском колледже, Гарвардском университете и Университете штата Огайо. Студентам, похоже, книга весьма нравится. Мой собственный опыт преподавания по ней таков, что, по мнению студентов, эта книга лучше, чем преподаватель.

Обязательные предварительные требования

Предполагается, что читатель владеет основами математического анализа, теории вероятностей и линейной алгебры. Считается само собой разумеющимся понимание следующих понятий: функции нескольких переменных, частные производные, интегралы, случайные величины, совместные распределения, независимость, безусловные и условные математические ожидания, нормальные распределения, распределения хи-квадрат, умножение матриц, обратные матрицы, ранг матрицы, детерминанты и положительно определенные матрицы. Более сложные понятия будут вводиться по ходу обсуждения. Результаты, касающиеся блочных матриц и кронекеровских произведений, собраны в приложении. Предварительного изучения бакалаврского курса эконометрики не требуется.

Структура книги

Чтобы понять организацию материала в этой книге, полезно различать *модель* и *процедуру оценивания*. Основная предпосылка эконометрики состоит в том, что экономические данные (такие как послевоенный ВНП США) являются случайными величинами. *Модель* представляет семейство распределений вероятностей, которое могло породить эти данные. *Процедура оценивания* представляет основанный на данных протокол выбора из модели некоторого конкретного распределения, которое, скорее всего, породило эти данные. Большинство процедур оценивания, используемых в эконометрике, является тем или иным частным случаем принципа GMM-оценивания. Например, когда GMM применяется к модели, известной как классическая линейная модель регрессии, результирующая процедура есть просто обычный метод наименьших квадратов (*Ordinary Least Squares — OLS*) — основная процедура оценивания в эконометрике. Такая точка зрения является организующим принципом в первых шести главах книги, где представлено большинство стандартных процедур оценивания.

GMM можно было бы изложить сразу в главе 1, но это лишило бы читателя возможности рассматривать ряд вопросов, специфических для OLS, не отвлекаясь на GMM. По этой причине я решил использовать две первые главы для представления теории OLS для конечных и больших выборок. Сам GMM вводится в главе 3 как обобщение OLS.

Главная новизна представления материала в книге — это трактовка процедур оценивания нескольких уравнений, таких как **SUR** (*Seemingly Unrelated Regression* — кажущиеся несвязанными регрессии), **3SLS** (*Three-Stage Least Squares* — трехшаговый метод наименьших квадратов), метод случайных эффектов, рассматриваемый в главе 4, и метод фиксированных эффектов, рассматриваемый в главе 5, — как частных случаев GMM для одного уравнения из главы 3. В главе 6 книги обсуждение GMM завершается указанием на то, каким образом можно инкорпорировать в GMM автокоррелированность ошибок.

Для некоторых эконометрических моделей более естественным, чем GMM, принципом оценивания является принцип максимального правдоподобия (*Maximum Likelihood* — ML), который рассматривается в главах 7 и 8. Чтобы сделать связь между GMM и ML более ясной, обсуждение ML начинается в книге в главе 7 с принципа экстремальных оценок (*Extremum Estimators*), который включает ML и GMM в качестве частных случаев. В главе 8 рассматривается применение ML к различным моделям.

Книга содержит также всестороннее рассмотрение анализа временных рядов. Основной материал, изложенный в параграфе 2.2 и в первой части главы 6, дает возможность перейти к изучению важных недавних достижений в анализе нестационарных временных рядов, которые рассматриваются в двух последних главах книги.

Построение курса на базе этой книги

На базе этой книги можно построить несколько различных курсов.

- Если курс читается два раза в неделю по полтора часа, то восьми недель будет достаточно для покрытия ядра теории, изложенного в главах 1–4 и 6 (исключая параграфы, посвященные конкретным экономическим приложениям), главе 7 (с пропуском доказательств и примеров) и главе 8.

- Двенадцатинедельный семестровый курс может покрывать, в дополнение к ядру теории, главу 5 и экономические применения, содержащиеся в главах 1–6.

- Краткий (скажем, шестинедельный) курс, посвященный GMM-оцениванию в кросс-секционных и панельных данных, мог бы покрывать главы 1–5 (исключая параграфы 2.10–2.12, но включая параграф 2.2), главы 7 и 8 (исключая параграф 8.7) могли бы быть дополнительной компонентой курса, посвященной ML-оцениванию.

- Краткий курс временных рядов, покрывающий недавние достижения с экономическими приложениями, мог бы включать главу 1 (исключая параграф 1.7), главу 2, части глав 6 и 7 (параграфы 6.1–6.5, 6.8 и 7.1), а также главы 9 и 10. Если курс фокусируется только на теории, прикладные параграфы в главах 2, 6, 9, и 10 можно опустить.

Контрольные вопросы и аналитические упражнения

В конце каждой главы книги помещено несколько коротких контрольных вопросов с большим количеством указаний (и даже ответов). Эти вопросы можно использовать для проверки того, что читатель действительно разобрался в материале параграфа. Если не при первом, то при повторном прочтении читатели должны пытаться ответить на указанные вопросы, прежде чем двигаться дальше. Ответы на некоторые контрольные вопросы доступны на сайте книги: http://pup.princeton.edu/titles/6946.html

В конце каждой главы имеется несколько аналитических упражнений, предлагающих читателю доказать результаты, оставленные недоказанными в тексте, или дополнительные результаты, полезные сами по себе. Если не указано противное, аналитические упражнения можно пропустить без потери непрерывности.

Эмпирические упражнения

Каждая глава, как правило, содержит одно большое эмпирическое упражнение. В нем предлагается воспроизвести эмпирические результаты оригинальной статьи, обсуждавшейся в прикладном параграфе этой главы, и произвести оценивание различных обобщений рассмотренной в статье модели с использованием процедур оценивания, представленных в главе. Набор данных для оценивания, который обычно тот же, что и использованный в оригинальной статье, можно скачать с упомянутого выше сайта книги.

Для применения процедуры оценивания к множеству данных читатели должны прогнать на компьютере пакет статистических программ. В эконометрике используется довольно много статистических пакетов, включая GAUSS (www.aptech.com), MATLAB (www.mathworks.com), Eviews (www.eviews.com), LIMDEP (www.limdep.com), RATS (www.estima.com), SAS (www.sas.com), STATA (www.stata.com) и TSP (www.tsp.com).

GAUSS и MATLAB отличаются от остальных пакетов тем, что они являются матричными языками, а не набором процедур. Рассмотрим, например, реализацию метода OLS посредством пакетов GAUSS или MATLAB. После загрузки данных в рабочее пространство компьютера в нескольких строках отражаются матричные операции OLS для получения OLS-оценки и сопутствующих статистик (например, R^2). В других пакетах, которые процедурно-ориентированы и иногда называются — «canned packages» («стандартные, готовые»), эти несколько строк могут заменяться однострочной командой, вызывающей OLS-процедуру. Например, в пакете TSP такой командой является OLSQ.

Стандартные пакеты имеют свои преимущества и недостатки. Ясно, что для выполнения того же самого в них требуется меньше строк, что ведет к меньшим затратам времени на программирование. С другой стороны, в этих пакетах процедуры, принимающие данные и выдающие точечные оценки и сопутствующие статистики, являются по существу, черным ящиком. Из документации пакета иногда неясно, как вычисляются те или иные статистики. Хотя новые достижения в эконометрике регулярно инкорпорируются в такие пакеты, желаемые процедуры все же могут отсутствовать в текущий момент времени, и тогда возникает необходимость в написании своих собственных процедур в GAUSS или MATLAB. Но это может быть и благом: реальное выписывание соответствующих матричных операций предоставляет вам отличную возможность понять процедуру оценивания.

За несколькими исключениями, все вычисления, необходимые для эмпирических упражнений, предлагаемых в книге, можно выполнить, используя любой из вышеперечисленных стандартных пакетов. Поэтому я рекомендую использовать GAUSS или MATLAB тем аспирантамэкономистам, которые планируют написание прикладной диссертации, используя современные процедуры оценивания, или написание теоретической диссертации, предлагающей новые процедуры. В противном случае студентам следует использовать любой из стандартных пакетов, упомянутых выше.

Математические обозначения

В эконометрике нет единых математических обозначений. Система обозначений в этой книге следует наиболее общепринятой, если не универсальной, практике. Векторы рассматриваются как векторы-столбцы и записываются полужирными строчными буквами. Для обозначения матриц используются полужирные прописные буквы. Транспонированная матрица для $\boldsymbol{A}$ обозначается как $\boldsymbol{A}'$. Для скалярных переменных используются (по большей части строчные) курсивные буквы.

Благодарности

Я весьма признателен за помощь, полученную от следующих лиц и учреждений. Марк Уотсон (Mark Watson), Дэйл Иоргенсон (Dale Jorgenson), Бо Оноре (Bo Honore), Серена Нг (Serena Ng), Масао Огаки (Masao Ogaki) и Юшан Баи (Jushan Bai) были достаточно любезны и отважны, чтобы использовать ранние варианты этого труда в качестве учебников для их курсов эконометрики. Комментарии, сделанные ими и их студентами, учтены в данной финальной версии. Юзо Хонда (Yuzo Honda) прочитал рукопись и сделал ряд полезных предложений. Наото Кунитомо (Naoto Kunitomo), Уитни Ньюи (Whitney Newey), Серена Нг (Serena Ng), Пьер Перрон (Pierre Perron), Джим Сток (Jim Stock), Кацуто Танака (Katsuto Tanaka), Марк Уотсон (Mark Watson), Хэл Уайт (Hal White) и Йошихиро Ядзима (Yoshihiro Yajima) нашли время для ответа на мои вопросы и запросы. Два студента Токийского университета — Мари Сакудо (Mari Sakudo) и Наоки Шимои (Naoki Shimoi) — вычитали всю рукопись с целью устранения опечаток. Их усилия были поддержаны грантом Zengin Foundation for Studies on Economics and Finance. Питер Дагерти (Peter Dougherty), мой редактор в Princeton University Press, обеспечивал энтузиазм и надлежащее давление. Стефани Хоуг (Stephanie Hogue) была достаточно разносторонним экспертом по LaTeX, чтобы приспособиться к моим прихотям в разметке набора. Эллен Фус (Ellen Foos) контролировала производство книги. Наконец, что не менее важно, Джессика Хелфанд (Jessica Helfand) согласилась, вероятно по дружбе, выполнить дизайн обложки.

В течение более пяти лет все мое свободное время уходило на написание этой книги. Теперь, закончив ее, я чувствую себя как человек, только что освобожденный из тюрьмы. Моя исследовательская деятельность пострадала, но, надеюсь, профессия этого не заметила.

Предисловие к русскому изданию

Я весьма польщен тем, что мой учебник «Эконометрика» предлагается российским читателям.

Прошло уже 15 лет с момента появления этой книги на английском языке. За эти годы у меня была возможность общаться со студентами из России, слушавшими мой курс или комментировавшими содержание этой книги. Я был впечатлен их технической компетентностью, причина которой, как я полагаю, кроется в математическом образовании в России. Ведь Россия — это страна, которая дала Ляпунова, Слуцкого, Колмогорова, Гордина и многие другие аналитические умы. И поэтому я со скромностью и в то же время с удовольствием воспринимаю то, что моя книга, которая не пытается идти на компромиссы в отношении строгости, наконец достигает широкого круга студентов с сильной базой знаний.

В 1980-е и 1990-е годы произошли два важных изменения в стиле и содержании эмпирических исследований. Одним из них является широкое использование GMM (обобщенного метода моментов), а другим является обобщение анализа временных рядов на процессы с единичным корнем (покрываемое двумя последними главами этой книги). Конечно, авторитетная программа аспирантуры хотела бы предложить для второго года курсы, посвященные этим новым разработкам. Однако я понял из опыта преподавания профильного курса эконометрики для студентов первого года аспирантуры, что вместо того, чтобы обобщенный метод моментов относить на второй год обучения, им можно — и нужно — заменить традиционное изложение базового материала на первом году обучения. На сегодняшний день моя книга остается единственным базовым учебником эконометрики, использующим GMM в качестве организующего принципа.

Выглядела бы моя книга иначе, если бы я начал сейчас новую? Ненамного. Разумеется, многое произошло с тех пор, как эта книга была опубликована на английском языке в 2000 году. Два наиболее важных продвижения произошли в использовании контролируемых экспериментов и в байесовской эконометрике. Первое только подчеркивает важность методологии, уже хорошо освещенной в книге. Как вы узнаете из главы 3, вам достаточно принять лишь одно предположение, которое сделает пригодным метод инструментальных переменных. Вся цель контролируемых экспериментов — сделать доступными наборы данных, для которых это идентифицирующее предположение гарантированно выполняется. В отличие от этого возрастающая популярность байесовской эконометрики, несомненно обязанная постоянному снижению стоимости вычислений, может в один прекрасный день заставить серьезно пересмотреть материал первого года. Это потому, что байесовская точка зрения радикально отличается от классического, или частотного, подхода, который принят в большинстве профильных учебников эконометрики, в том числе и в моем. Однако смена парадигмы, если она вообще состоится, — дело далекого будущего.

Поэтому я уверен, что моя книга предлагает наиболее эффективный и по сей день путь обучения эконометрике на первом году аспирантуры.

Фумио Хайяши
Июнь 2015

Глава 1. Свойства OLS в конечных выборках

Оценивание **методом наименьших квадратов** (*Ordinary Least Squares* **OLS**) является основной процедурой оценивания в эконометрике. В этой главе описываются свойства OLS-оценки[1]в **малых**, или **конечных, выборках**, то есть статистические свойства OLS-оценки, которые действительны для выборки любого заданного размера. Материалы, представленные в этой главе, совершенно стандартны. Изложение здесь отличается от большинства других книг подчеркиванием роли, которую играет предположение о «строгой экзогенности» регрессоров.

В заключительном параграфе мы применяем теорию для конечных выборок для оценивания функции издержек, используя кросс-секционные данные по отдельным фирмам. Важное практическое значение имеет вопрос, поставленный в работе [Nerlove, 1963]: имеет ли место возрастающая отдача от масштаба в отрасли электроснабжения? Если да, то микроэкономика говорит нам, что отрасль должна регулироваться государством. Кроме предоставления вам практических навыков использования методов для проверки интересующих гипотез в работе Нерлова имеется тщательное обсуждение вопроса о том, почему OLS является подходящим методом оценивания в конкретном прикладном исследовании.

---

[1]В оригинале здесь используется термин *estimator* («оцениватель»). Обычно при переводе этого термина на русский язык используют термин «оценка» и для метода (формулы), и для результата применения этого метода (формулы) к конкретной выборке данных, то есть для числового значения, которое при этом получается. Поскольку сам автор говорит далее (см. абзац, следующий за формулой (1.2.5)) о том, что в этой книге оба термина *estimator* (оцениватель) и *estimate* (оценка) используются практически как синонимы, мы будем придерживаться указанной выше практики перевода. — *Прим. науч. ред. перевода.*

1.1. Классическая линейная модель регрессии

В данном параграфе мы приводим предположения, на которых базируется классическая линейная модель регрессии.

В этой модели рассматриваемая переменная (которая называется **зависимой переменной** (*dependent variable*), **регрессируемой переменной** (*regressand*), или, по происхождению, *переменной в левой части* (*left-hand [-side] variable*), связана с несколькими другими переменными (которые называются **регрессорами** (*regressors*), **объясняющими переменными** (*explanatory variables*), или **переменными в правой части** (*right-hand [-side] variables*)). Допустим, мы наблюдаем n значений этих переменных. Пусть y_i обозначает i-е наблюдение зависимой переменной, а $(x_{i1}, x_{i2}, \ldots, x_{iK})$ — i-е наблюдение K регрессоров. **Выборка** (*sample*) или **данные** (*data*) — это набор этих n наблюдений.

Данные в экономике не могут быть получены в результате экспериментов (исключая экспериментальную экономику), так что и зависимая, и независимые[1] переменные рассматриваются как случайные величины — величины, значения которых подвержены случайности. **Модель** (*model*) — это набор ограничений на совместное распределение зависимой и независимых переменных. Иначе говоря, модель представляет собой совокупность совместных распределений, удовлетворяющих некоторому набору предположений. Классическая линейная модель регрессии — это совокупность совместных распределений, удовлетворяющих предположениям 1.1–1.4, сформулированным ниже.

Предположение линейности

Первое предположение заключается в том, что связь между зависимой переменной и регрессором линейна.

Предположение 1.1 (линейность):

$$y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + \varepsilon_i \quad (i = 1, 2, \ldots, n), \tag{1.1.1}$$

где $\beta_1, \ldots, \beta_K$ — неизвестные параметры, подлежащие оцениванию, ε_i — ненаблюдаемое значение ошибки с определенными свойствами, указанными ниже.

Выражение $\beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_K x_{iK}$ в правой части уравнения называется **регрессией** (*regression*), или **функцией регрессии** (*regression function*), а коэффициенты β_i — **коэффициентами регрессии** (*regression coefficients*). Они представляют собой предельные эффекты отдельных регрессоров. Например, β_2 представляет изменение зависимой

[1] Это переменные в правой части. — *Прим. науч. ред. перевода.*

переменной, когда второй регрессор увеличивается на единицу, в то время как остальные остаются неизменными. Математически это может быть записано как $\partial y_i / \partial x_{i2} = \beta_2$. Линейность означает, что предельные эффекты не зависят от значений регрессоров. Ошибка представляет собой часть зависимой переменной, которая остается не объясненной регрессорами.

Пример 1.1 (функция потребления): Простая функция потребления, знакомая из вводного курса экономики, — это

$$CON_i = \beta_1 + \beta_2 YD_i + \varepsilon_i, \tag{1.1.2}$$

где CON — потребление, YD — располагаемый доход. Если данные представляют собой годовые макроэкономические временные ряды, то CON_i и YD_i — агрегированное потребление и доход в году i. Если же данные получены из обследования домашних хозяйств, то CON_i — это потребление i-го домашнего хозяйства в кросс-секционной выборке из n домашних хозяйств. Функцию потребления можно записать в виде (1.1.1), обозначив $y_i = CON_i$, $x_{1i} = 1$ (константа) и $x_{i2} = YD_i$. Ошибка ε_i представляет другие переменные помимо располагаемого дохода, которые влияют на потребление. Они включают как переменные — такие как финансовые активы, — которые могут быть наблюдаемы, но исследователь решил их не включать в качестве регрессоров, так и другие переменные — такие как «настроение» потребителей, — которые трудно измерить. Когда уравнение включает только один регрессор, отличный от константы, как в этом примере, это называется **простой моделью регрессии**[1] (*simple regression model*).

Предположение линейности не является ограничительным, как может показаться на первый взгляд, так как зависимая переменная и регрессоры могут быть следствием преобразований рассматриваемых переменных. Рассмотрим

Пример 1.2 (уравнение заработной платы): Упрощенная версия уравнения заработной платы, обычно оцениваемая в экономике труда, имеет вид:

$$\log(WAGE_i) = \beta_1 + \beta_2 S_i + \beta_3 TENURE_i + \beta_4 EXPR_i + \varepsilon_i, \tag{1.1.3}$$

где $WAGE$ = уровень заработной платы человека, S — количество лет обучения, $TENURE$ — количество лет, отработанных на текущем рабочем месте, $EXPR$ — количество лет трудового стажа (то есть количество лет, проработанных человеком на текущем и предыдущих рабочих местах). Уравнение заработной платы вписывается в общий формат (1.1.1), если обозначить $y_i = \log(WAGE_i)$. Такую спецификацию уравнения называют **полулогарифмической** (*semi-log*), поскольку логарифмическому

[1] В русскоязычной литературе в этом случае говорят также о парной линейной регрессии.

преобразованию подвергается только зависимая переменная. Указанное уравнение выводится из следующей нелинейной связи между уровнем заработной платы и регрессорами:

$$WAGE_i = \exp(\beta_1)\exp(\beta_2 S_i)\times$$
$$\times\exp(\beta_3 TENURE_i)\exp(\beta_4 EXPR_i)\exp(\varepsilon_i). \quad (1.1.4)$$

Взяв логарифм от обеих частей (1.1.4) и учитывая, что $\log[\exp(x)] = x$, получаем (1.1.3). Коэффициенты в полулогарифмической форме имеют интерпретацию *процентных изменений*, а не изменений в уровнях. Например, значение 0,05 коэффициента β_2 означает, что дополнительный год обучения приводит к пятипроцентному росту уровня заработной платы[1]. Разница в интерпретации получается из-за того, что зависимая переменная — это логарифм заработной платы, а не сама заработная плата, и изменение в значении логарифма равно[2] процентному изменению уровня.

К виду (1.1.1) могут быть сведены и некоторые другие разновидности нелинейного влияния. Предположим, например, что предельный эффект образования уменьшается по мере роста уровня образования. Это может быть отражено путем включения в уравнение заработной платы дополнительного регрессора — квадрата количества лет обучения (S^2). Если обозначить коэффициент перед новым регрессором как β_5, то предельный эффект от образования будет рассчитываться по формуле:

$$\beta_2 + 2\beta_5 S (= \partial\log(WAGE)/\partial S).$$

Если коэффициент β_5 отрицателен, предельный эффект от образования снижается при росте уровня образования.

Конечно, существуют нелинейные спецификации, не сводимые к линейным. Например, уравнение регрессии (1.1.4) было бы невозможно свести к линейному, если бы ошибка входила не мультипликативно, а аддитивно:

$$WAGE_i = \exp(\beta_1)\exp(\beta_2 S_i)\exp(\beta_3 TENURE_i)\exp(\beta_4 EXPR_i) + \varepsilon_i.$$

Методы оценивания нелинейных уравнений регрессии (подобных этому) будут обсуждаться в главе 7.

Матричные обозначения

Прежде чем обсуждать прочие предположения классической линейной регрессионной модели, введем векторные и матричные обозначения. Эти обозначения будут полезны как для формулировки прочих предположений, так и для вывода OLS-оценки β. Обозначим K-мерные векторы-

[1]К приблизительно пятипроцентному росту уровня заработной платы. — *Прим. науч. ред. перевода.*

[2]Приблизительно равно. — *Прим. науч. ред. перевода.*

столбцы $\boldsymbol{x}_i$ и $\boldsymbol{\beta}$ как

$$\underset{(K\times 1)}{\boldsymbol{x}_i} = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{iK} \end{bmatrix}, \quad \underset{(K\times 1)}{\boldsymbol{\beta}} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{bmatrix}. \tag{1.1.5}$$

По определению скалярного произведения векторов, $\boldsymbol{x}_i'\boldsymbol{\beta} = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_K x_{ik}$. Поэтому уравнения в предположении 1.1 могут быть переписаны следующим образом:

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i \quad (i = 1, 2, \dots, n). \tag{1.1.1'}$$

Также определим

$$\underset{(n\times 1)}{\boldsymbol{y}} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad \underset{(n\times 1)}{\boldsymbol{\varepsilon}} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}, \quad \underset{(n\times K)}{\boldsymbol{X}} = \begin{bmatrix} \boldsymbol{x}_1' \\ \vdots \\ \boldsymbol{x}_n' \end{bmatrix} = \begin{bmatrix} x_{11} & \dots & x_{1K} \\ \vdots & \dots & \vdots \\ x_{n1} & \dots & x_{nK} \end{bmatrix}. \tag{1.1.6}$$

В (1.1.6) количество строк в векторах и матрицах равно числу наблюдений. По этой причине $\boldsymbol{y}$ и $\boldsymbol{X}$ иногда называют **вектором данных** и **матрицей данных** (*data vector, data matrix*). Поскольку число столбцов в $\boldsymbol{X}$ равняется числу строк в $\boldsymbol{\beta}$, $\boldsymbol{X}$ и $\boldsymbol{\beta}$ являются согласованными, и $\boldsymbol{X}\boldsymbol{\beta}$ — вектор размера $n \times 1$. i-й элемент этого вектора равен $\boldsymbol{x}_i'\boldsymbol{\beta}$. Поэтому предположение 1.1 можно переписать более компактно:

$$\underset{(n\times 1)}{\boldsymbol{y}} = \underbrace{\underset{(n\times K)}{\boldsymbol{X}}\underset{(K\times 1)}{\boldsymbol{\beta}}}_{(n\times 1)} + \underset{(n\times 1)}{\boldsymbol{\varepsilon}}.$$

Предположение о строгой экзогенности

Сформулируем следующее предположение классической регрессионной модели:

Предположение 1.2 (строгая экзогенность):

$$\mathrm{E}(\varepsilon_i|\boldsymbol{X}) = 0 \quad (i = 1, 2, \dots, n). \tag{1.1.7}$$

Здесь математическое ожидание (среднее) рассчитывается как условное относительно значений регрессоров для всех наблюдений. Это замечание можно сделать более наглядным, если переписать указанное предположение без использования матрицы данных:

$$\mathrm{E}(\varepsilon_i|\boldsymbol{x}_1 \dots \boldsymbol{x}_n) = 0 \quad (i = 1, 2, \dots, n).$$

Это предположение можно сформулировать иначе. Для любого наблюдения i возьмем совместное распределение $nK+1$ случайных величин, $f(\varepsilon_i, \boldsymbol{x}_1, \dots \boldsymbol{x}_n)$ и рассмотрим условное распределение $f(\varepsilon_i | \boldsymbol{x}_1, \dots \boldsymbol{x}_n)$. Условное математическое ожидание $\mathrm{E}(\varepsilon_i | \boldsymbol{x}_1, \dots \boldsymbol{x}_n)$ является, в общем случае, нелинейной функцией от $(\boldsymbol{x}_1, \dots \boldsymbol{x}_n)$. Предположение о строгой экзогенности говорит, что эта функция равна константе, а точнее нулю[1].

Предположение о том, что эта константа равна нулю, не является ограничительным, если регрессоры включают константу, поскольку тогда уравнение может быть переписано так, что условное математическое ожидание ошибки равно нулю. Чтобы продемонстрировать это, предположим, что $\mathrm{E}(\varepsilon_i | \boldsymbol{X})$ равно μ и что $x_{i1} = 1$. Тогда уравнение можно записать в виде:

$$
\begin{aligned}
y_i &= \beta_1 + \beta_2 x_{i2} + \dots + \beta_K x_{iK} + \varepsilon_i = \\
&= (\beta_1 + \mu) + \beta_2 x_{i2} + \dots + \beta_K x_{iK} + (\varepsilon_i - \mu).
\end{aligned}
$$

Если переобозначить β_1 как $\beta_1 + \mu$ и ε_i как $\varepsilon_i - \mu$, то условное математическое ожидание новой ошибки равно нулю. Почти во всех приложениях регрессоры включают постоянную составляющую.

Пример 1.3 (продолжение примера 1.1): Для простой регрессионной модели из примера 1.1 предположение строгой экзогенности можно записать как

$$\mathrm{E}(\varepsilon_i | YD_1,\ YD_2, \dots YD_n) = 0.$$

Поскольку $\boldsymbol{x}_i = (1, YD_i)'$, вы могли бы захотеть переписать предположение о строгой экзогенности в виде:

$$\mathrm{E}(\varepsilon_i | 1, YD_1, 1, YD_2, \dots, 1, YD_n) = 0.$$

Однако, поскольку константа не несет в себе информации, условное математическое ожидание при условии

$$(1, YD_1, 1, YD_2, \dots, 1, YD_n)$$

совпадает с математическим ожиданием при условии

$$(YD_1, YD_2, \dots, YD_n).$$

Следствия строгой экзогенности

Предположение о строгой экзогенности имеет ряд важных следствий.

- *Без*условное математическое ожидание ошибки равно нулю, то есть

$$\mathrm{E}(\varepsilon_i) = 0 \quad (i = 1, 2, \dots, n). \tag{1.1.8}$$

[1]Некоторые авторы определяют понятие «строгой экзогенности» немного иначе. Например, в [Koopmans and Hood, 1953] и [Engle, Hendry and Richard, 1983] регрессоры называются строго экзогенными, если $\boldsymbol{x}_i$ независим от ε_j для всех i, j. Это определение является более сильным, чем определение строгой экзогенности, используемое здесь, но не противоречит ему.

Это так, поскольку, по закону полных математических ожиданий[1], $\mathrm{E}[\mathrm{E}(\varepsilon_i|\boldsymbol{X})] = \mathrm{E}(\varepsilon_i)$.

- Если смешанный момент $\mathrm{E}(xy)$ двух случайных величин x и y равен нулю, мы будем говорить, что x **ортогональна** y (или y ортогональна x). При выполнении строгой экзогенности регрессоры ортогональны ошибке *для всех* наблюдений, то есть

$$\mathrm{E}(x_{jk}\varepsilon_i) = 0 \quad (i, j = 1, \ldots, n; k = 1, \ldots K)$$

или

$$\mathrm{E}(\boldsymbol{x}_j \cdot \varepsilon_i) = \begin{bmatrix} \mathrm{E}(x_{j1}\,\varepsilon_i) \\ \mathrm{E}(x_{j2}\,\varepsilon_i) \\ \vdots \\ \mathrm{E}(x_{jK}\,\varepsilon_i) \end{bmatrix} = \underset{(K\times 1)}{\mathbf{0}} \quad \text{(для всех } i, j). \tag{1.1.9}$$

Доказательство этого утверждения, приведенное ниже, представляет собой прекрасный пример применения свойств условных математических ожиданий.

Доказательство

Поскольку x_{jk} является элементом $\boldsymbol{X}$, строгая экзогенность влечет

$$\mathrm{E}(\varepsilon_i|x_{jk}) = \mathrm{E}[\mathrm{E}(\varepsilon_i|\boldsymbol{X})|x_{jk}] = 0 \tag{1.1.10}$$

по формуле повторного математического ожидания[2]. Из этого следует, что

$$\begin{aligned} \mathrm{E}(x_{jk}\varepsilon_i) &= \mathrm{E}[\mathrm{E}(x_{jk}\varepsilon_i|x_{jk})] \\ &\text{(по формуле повторного математического ожидания)} \\ &= \mathrm{E}[x_{jk}\,\mathrm{E}(\varepsilon_i|x_{jk})] \\ &\text{(из линейности условных математических ожиданий}^3\text{)} \\ &= 0. \end{aligned}$$

■

Следует подчеркнуть, что строгая экзогенность требует ортогональности регрессоров не только ошибке в том же наблюдении (то есть $\mathrm{E}(x_{ik}\varepsilon_i) = 0$ для всех k), но также ошибкам в других наблюдениях (то есть $\mathrm{E}(x_{jk}\varepsilon_i) = 0$ для всех k и для $j \neq i$).

[1] Закон полных математических ожиданий имеет вид: $\mathrm{E}[\mathrm{E}(y|\boldsymbol{x})] = \mathrm{E}(y)$.

[2] Формула повторного математического ожидания имеет вид: $\mathrm{E}[\mathrm{E}(y|\boldsymbol{x}, z)|\boldsymbol{x}] = \mathrm{E}(y|\boldsymbol{x})$.

[3] Линейность условных математических ожиданий утверждает, что $\mathrm{E}[f(\boldsymbol{x})y|\boldsymbol{x}] = f(\boldsymbol{x})\,\mathrm{E}(y|\boldsymbol{x})$.

- Так как математическое ожидание ошибки равно нулю, условия ортогональности (1.1.9) эквивалентны условиям нулевой корреляции. Это верно, так как

$$\begin{aligned}\mathrm{Cov}(\varepsilon_i, x_{jk}) &= \mathrm{E}(x_{jk}\varepsilon_i) - \mathrm{E}(x_{jk})\,\mathrm{E}(\varepsilon_i)\\ &\quad \text{(по определению ковариации)}\\ &= \mathrm{E}(x_{jk}\varepsilon_i) \quad (\text{так как } \mathrm{E}(\varepsilon_i) = 0, \text{ см. } (1.1.8)) =\\ &= 0 \quad (\text{из условий ортогональности } (1.1.9)).\end{aligned}$$

  В частности, для $i = j$ $\mathrm{Cov}(x_{ik}, \varepsilon_i) = 0$. Поэтому строгая экзогенность подразумевает (знакомое тем, кто ранее изучал эконометрику) требование того, что регрессоры не коррелированы с ошибкой в том же наблюдении (*contemporaneously uncorrelated*).

Строгая экзогенность в моделях временных рядов

В моделях временных рядов, где i обозначает время, следствие строгой экзогенности (1.1.9) может быть переформулировано следующим образом: регрессоры ортогональны прошлым, текущим и будущим ошибкам (это эквивалентно ортогональности ошибок прошлым, текущим и будущим значениям регрессоров). Но для большинства моделей временных рядов это условие (и тем более строгая экзогенность) не выполнено. Иными словами, теорию конечных выборок, базирующуюся на понятии строгой экзогенности, которая будет изложена в этой главе, редко можно применить в случае временных рядов. Однако в следующей главе будет показано, что для больших выборок предложенная оценка обладает рядом хороших свойств и без предположения о строгой экзогенности.

Самый очевидный пример невыполнения свойства строгой экзогенности — это модель, в которой в качестве регрессора выступает **лаг зависимой переменной**. Вот простейшая из подобных моделей:

$$y_i = \beta y_{i-1} + \varepsilon_i \quad (i = 1, 2, \ldots, n). \tag{1.1.11}$$

Она называется **моделью авторегрессии первого порядка** (*first-order autoregressive model* — **AR(1)**). (Мы рассмотрим эту модель более подробно в главе 6.) Предположим, что (в соответствии с предположением о строгой экзогенности) регрессор для наблюдения i, y_{i-1}, ортогонален ошибке для того же наблюдения i, то есть $\mathrm{E}(y_{i-1}\varepsilon_i) = 0$. Тогда

$$\begin{aligned}\mathrm{E}(y_i\varepsilon_i) &= \mathrm{E}[(\beta y_{i-1} + \varepsilon_i)\varepsilon_i] = \quad (\text{из } (1.1.11))\\ &= \beta\,\mathrm{E}(y_{i-1}\varepsilon_i) + \mathrm{E}(\varepsilon_i^2) =\\ &= \mathrm{E}(\varepsilon_i^2) \quad (\text{так как } \mathrm{E}(y_{i-1}\varepsilon_i) = 0 \text{ по предположению}).\end{aligned}$$

Поэтому (если исключить случай, когда ошибка всегда равна нулю) $\mathrm{E}(y_i\varepsilon_i)$ не равно нулю. Однако y_i является регрессором для наблюдения $i+1$. Таким образом, регрессор неортогонален прошлой ошибке, что нарушает строгую экзогенность.

Остальные предположения модели

Ниже приведены остальные предположения классической регрессионной модели.

Предположение 1.3 (отсутствие мультиколлинеарности): *Ранг матрицы данных $\boldsymbol{X}$ размера $n \times K$ равен K с вероятностью 1.*

Предположение 1.4 (сферичность ковариационной матрицы ошибок):

$$\text{(гомоскедастичность)}\quad \mathrm{E}(\varepsilon_i^2|\boldsymbol{X}) = \sigma^2 > 0 \quad (i = 1, 2, \ldots, n)^1. \tag{1.1.12}$$

(отсутствие корреляции между наблюдениями)

$$\mathrm{E}(\varepsilon_i\varepsilon_j|\boldsymbol{X}) = 0 \quad (i, j = 1, 2, \ldots, n; i \neq j). \tag{1.1.13}$$

Чтобы понять предположение 1.3, необходимо вспомнить из линейной алгебры, что ранг матрицы равен числу линейно независимых столбцов этой матрицы. Это предположение говорит, что ни один из K столбцов матрицы данных $\boldsymbol{X}$ не может быть выражен как линейная комбинация остальных столбцов матрицы $\boldsymbol{X}$. Это означает, что матрица $\boldsymbol{X}$ имеет **полный столбцовый ранг** (*full column rank*). Так как K столбцов не могут быть линейно независимыми, если их размерность меньше K, из данного предположения вытекает, что $n \geq K$, то есть наблюдений должно быть по меньшей мере столько же, сколько регрессоров. Говорят, что регрессоры **(строго) мультиколлинеарны** (*(perfectly) multicollinear*), если это предположение не выполняется. В некоторых случаях несложно увидеть, что регрессоры мультиколлинеарны и какие проблемы возникают вследствие этого.

Пример 1.4 (продолжение примера 1.2): Если в выборке ни один из индивидов никогда не менял место работы, то $TENURE_i = EXPR_i$ для всех i, что противоречит нашему предположению об отсутствии мультиколлинеарности. Очевидно, что в этом случае невозможно разделить влияние на заработную плату количества лет, проработанных

[1] Если написано, что момент распределения (в данном случае второй момент, $\mathrm{E}(\varepsilon_i^2|\boldsymbol{X})$) равен какому-либо значению (здесь σ^2), то это неявно подразумевает, что этот момент существует и конечен. Далее мы будем следовать этому соглашению.

на текущем рабочем месте, и влияние общего трудового стажа. Если подставить это равенство в уравнение заработной платы, чтобы исключить переменную $TENURE_i$, уравнение заработной платы примет вид:

$$\log(WAGE_i) = \beta_1 + \beta_2 S_i + (\beta_3 + \beta_4) EXPR_i + \varepsilon_i,$$

из чего следует, что можно оценить только сумму коэффициентов $\beta_3+\beta_4$, но не коэффициенты β_3 и β_4 по отдельности.

Предположение гомоскедастичности (1.1.12) говорит, что условный момент второго порядка (который в общем случае является нелинейной функцией от $\boldsymbol{X}$) — постоянная величина. Используя требование строгой экзогенности, это предположение можно переформулировать в более привычных терминах. Рассмотрим условную дисперсию $\mathrm{Var}(\varepsilon_i|\boldsymbol{X})$. Она равна этой же постоянной, так как

$$\begin{aligned}\mathrm{Var}(\varepsilon_i|\boldsymbol{X}) &\equiv \mathrm{E}(\varepsilon_i^2|\boldsymbol{X}) - \mathrm{E}(\varepsilon_i|\boldsymbol{X})^2 \\ &\quad \text{(по определению условной дисперсии)} \\ &= \mathrm{E}(\varepsilon_i^2|\boldsymbol{X}) \\ &\quad \text{(так как } \mathrm{E}(\varepsilon_i|\boldsymbol{X}) = 0 \text{ вследствие строгой экзогенности).}\end{aligned}$$

Аналогично (1.1.13) эквивалентно следующему требованию:

$$\mathrm{Cov}(\varepsilon_i, \varepsilon_j|\boldsymbol{X}) = 0 \quad (i, j = 1, 2, \ldots, n; i \neq j).$$

То есть в совместном распределении $(\varepsilon_i, \varepsilon_j)$ при условии $\boldsymbol{X}$ ковариация равна нулю. Применительно к моделям временных рядов (1.1.13) эквивалентно требованию отсутствия **сериальной корреляции** у ошибки.

Так как элемент (i, j) матрицы $\boldsymbol{\varepsilon\varepsilon}'$ размера $n \times n$ равен $\varepsilon_i\varepsilon_j$, предположение 1.4 можно переписать в более компактной форме:

$$\mathrm{E}(\boldsymbol{\varepsilon\varepsilon}'|\boldsymbol{X}) = \sigma^2 \boldsymbol{I}_n. \tag{1.1.14}$$

Дискуссия в предыдущем параграфе показывает, что это предположение можно представить и в следующем виде:

$$\mathrm{Var}(\boldsymbol{\varepsilon}|\boldsymbol{X}) = \sigma^2 \boldsymbol{I}_n.$$

Однако представление (1.1.14) предпочтительнее, поскольку более удобной мерой разброса являются вторые моменты (такие как $\mathrm{E}(\varepsilon_i^2|\boldsymbol{X})$), а не дисперсии. Смысл этого замечания станет понятен после изучения следующей главы, в которой речь пойдет о теории больших выборок. Предположение 1.4 иногда называют предположением **сферичности** ковариационной матрицы, поскольку матрица вторых моментов размера

$n \times n$ (элементами которой являются дисперсии и ковариации) пропорциональна единичной матрице $\boldsymbol{I}_n$. В дальнейшем это предположение будет ослаблено.

Классическая регрессионная модель для случайных выборок

Выборка $(\boldsymbol{y}, \boldsymbol{X})$ называется **случайной**, если последовательность $\{y_i, \boldsymbol{x}_i\}$ является i.i.d. (независимой и одинаково распределенной) по наблюдениям. Так как по предположению 1.1 ε_i — это функция от $(y_i, \boldsymbol{x}_i)$ и поскольку $(y_i, \boldsymbol{x}_i)$ независимы от $(y_j, \boldsymbol{x}_j)$ для $j \neq i$, $(\varepsilon_i, \boldsymbol{x}_i)$ независимы от $\boldsymbol{x}_j$ для $j \neq i$. Поэтому

$$\begin{aligned} &\mathrm{E}(\varepsilon_i|\boldsymbol{X}) = \mathrm{E}(\varepsilon_i|\boldsymbol{x}_i), \\ &\mathrm{E}(\varepsilon_i^2|\boldsymbol{X}) = \mathrm{E}(\varepsilon_i^2|\boldsymbol{x}_i), \\ \text{и } &\mathrm{E}(\varepsilon_i\varepsilon_j|\boldsymbol{X}) = \mathrm{E}(\varepsilon_i|\boldsymbol{x}_i)\,\mathrm{E}(\varepsilon_j|\boldsymbol{x}_j) \quad (\text{для } i \neq j). \end{aligned} \tag{1.1.15}$$

(Доказательство последнего равенства в (1.1.15) вынесено в контрольные вопросы.) Следовательно, предположения 1.2 и 1.4 сводятся к таким:

$$\text{предположение 1.2:}\quad \mathrm{E}(\varepsilon_i|\boldsymbol{x}_i) = 0 \quad (i = 1, 2, \ldots, n), \tag{1.1.16}$$

$$\text{предположение 1.4:}\quad \mathrm{E}(\varepsilon_i^2|\boldsymbol{x}_i) = \sigma^2 > 0 \quad (i = 1, 2, \ldots, n). \tag{1.1.17}$$

Следствием одинакового распределения случайной выборки является то, что совместное распределение $(\varepsilon_i, \boldsymbol{x}_i)$ одинаково для всех i. То есть *без*условный второй момент $\mathrm{E}(\varepsilon_i^2)$ не зависит от i (это называют **безусловной гомоскедастичностью** (unconditional homoskedasticity)) и функциональная форма условного второго момента $\mathrm{E}(\varepsilon_i^2|\boldsymbol{x}_i)$ одинакова для всех i. Однако из предположения 1.4 не следует, что *значение* условного второго момента одинаково для всех i. Поэтому предположение 1.4 остается ограничительным в ситуации случайной выборки: без этого предположения условный второй момент $\mathrm{E}(\varepsilon_i^2|\boldsymbol{x}_i)$ мог бы быть различным для разных i вследствие его потенциальной зависимости от $\boldsymbol{x}_i$. Чтобы подчеркнуть разницу, мы будем называть ограничения на условные моменты второго порядка, (1.1.12) и (1.1.17), **условной гомоскедастичностью**.

«Фиксированные» регрессоры

Классическая линейная регрессионная модель была изложена в предположении, что регрессоры являются стохастическими. Такое изложение противоречит большинству учебников, в которых регрессоры предполагаются детерминированными, или «фиксированными» (*deterministic*, «*fixed*»). Если регрессоры $\boldsymbol{X}$ фиксированы, то нет необходимости делать различие между условным, $f(\varepsilon_i.|\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n)$, и безусловным, $f(\varepsilon_i)$, распределением ошибки. Тогда предположения 1.2 и 1.4 можно переписать в виде:

$$\text{предположение 1.2:} \qquad \mathrm{E}(\varepsilon_i) = 0 \quad (i = 1,\ldots,n), \tag{1.1.18}$$

$$\text{предположение 1.4:} \qquad \mathrm{E}(\varepsilon_i^2) = \sigma^2 \quad (i = 1,\ldots,n);$$

$$\mathrm{E}(\varepsilon_i\varepsilon_j) = 0 \quad (i,j = 1,\ldots,n; i \neq j). \tag{1.1.19}$$

Очевидно, что такое предположение является нереалистичным в неэкспериментальной науке (в частности, в эконометрике). Однако предположение о детерминированности регрессоров остается популярным, так как регрессионную модель с фиксированными $\boldsymbol{X}$ можно интерпретировать как набор утверждений, условных относительно $\boldsymbol{X}$, что позволяет исключить «$|\boldsymbol{X}$» из рассуждений, в частности из предположений 1.2 и 1.4.

Однако за экономичность в обозначениях необходимо платить свою цену. При таких обозначениях очень легко упустить из виду предположение, что ошибка не коррелирована с текущими, прошлыми и будущими значениями регрессоров. Кроме того, в случае детерминированных регрессоров исчезает разница между условной и безусловной гомоскедастичностью. Поэтому в этой книге регрессоры предполагаются стохастическими, и, если не указано иное, утверждения, условные относительно $\boldsymbol{X}$, явно записываются с помощью вставки «$|\boldsymbol{X}$» (если не сказано иного).

Контрольные вопросы

1. (Смена единиц измерения в полулогарифмической модели.) Как изменится уравнение заработной платы (1.1.3) в примере 1.2, если $WAGE$ измерять не в долларах, а в центах? **Указание:** $\log(xy) = \log(x) + \log(y)$.

2. Докажите последнее равенство в (1.1.5).
Указание: $\mathrm{E}(\varepsilon_i\varepsilon_j|\boldsymbol{X}) = \mathrm{E}\left[\varepsilon_j\,\mathrm{E}(\varepsilon_i|\boldsymbol{X},\varepsilon_j)|\boldsymbol{X}\right]$. $(\varepsilon_i,\boldsymbol{X})$ *независимы от переменных* $(\varepsilon_j,\boldsymbol{x}_1,\ldots,\boldsymbol{x}_{i-1},\boldsymbol{x}_{i+1},\ldots,\boldsymbol{x}_n)$ *для* $i \neq j$.

3. (Одновременное выполнение линейности и строгой экзогенности.) Покажите, что из предположений 1.1 и 1.2 следует, что

$$\mathrm{E}(y_i|\boldsymbol{X}) = \boldsymbol{x}_i'\beta \quad (i = 1,2,\ldots,n). \tag{1.1.20}$$

Докажите в обратную сторону: из выполнения условия (1.1.20) следует, что существуют ошибки, которые удовлетворяют предположениям 1.1 и 1.2.

4. (Нормально распределенная случайная выборка.) Рассмотрим случайную выборку пар потребление-доход: (CON_i, YD_i) $(i = 1,2,\ldots,n)$. Предположим, что совместное распределение (CON_i, YD_i) (оно одинаково для всех i вследствие предположения о случайности выборки) нормально. Очевидно, что предположение 1.3 выполнено; ранг $\boldsymbol{X}$ может быть меньше K только по чистой случайности. Покажите, что прочие предположения (1.1, 1.2 и 1.4)

также выполнены. **Указание:** *Если две случайные величины, y и x, совместно нормально распределены, тогда условное математическое ожидание линейно по x, то есть*

$$\mathrm{E}(y|x) = \beta_1 + \beta_2 x,$$

и условная дисперсия, $\mathrm{Var}(y|x)$, *не зависит от x. Ключевым является тот факт, что это распределение одинаково для всех i, так как в противном случае β_1 и β_2 могут отличаться для разных i.*

5. (Мультиколлинеарность в модели парной регрессии.) Покажите, что предположение 1.3 в случае парной регрессии равносильно утверждению, что регрессор, отличный от постоянной, (x_{i2}) не является постоянной (то есть $x_{i2} \neq x_{j2}$ для некоторых пар $(i, j), i \neq j$ с вероятностью, равной единице).

6. (Упражнение на условные и безусловные математические ожидания.) Покажите, что из предположений 1.1 и 1.4 следует, что

$$\mathrm{Var}(\varepsilon_i) = \sigma^2 \quad (i = 1, \ldots, n)$$
$$\text{и} \quad \mathrm{Cov}(\varepsilon_i \varepsilon_j) = 0 \quad (i \neq j; i, j = 1, \ldots, n). \qquad (*)$$

Указание: *Из строгой экзогенности вытекает* $\mathrm{E}(\varepsilon_i) = 0$. *Поэтому* $(*)$ *эквивалентно*

$$\mathrm{E}(\varepsilon_i^2) = \sigma^2 \quad (i = 1, \ldots, n)$$
$$\textit{и} \quad \mathrm{E}(\varepsilon_i \varepsilon_j) = 0 \quad (i \neq j; i, j = 1, \ldots, n).$$

1.2. Алгебра метода наименьших квадратов

Этот параграф посвящен вычислительной процедуре получения OLS-оценки $\boldsymbol{b}$ неизвестного вектора коэффициентов $\boldsymbol{\beta}$. Также вводятся некоторые понятия, связанные с вектором $\boldsymbol{b}$.

OLS минимизирует сумму квадратов остатков

Хотя мы не наблюдаем ошибку, мы можем при некотором заданном гипотетическом значении $\widetilde{\boldsymbol{\beta}}$ вектора коэффициентов $\boldsymbol{\beta}$ вычислить значение

$$y_i - \boldsymbol{x}_i' \widetilde{\boldsymbol{\beta}}.$$

Это называется **остатком** (*residual*) для наблюдения i. Тогда **сумма квадратов остатков** (*sum of squared residuals* — **SSR**) представляет собой

$$SSR(\widetilde{\boldsymbol{\beta}}) \equiv \sum_{i=1}^{n} (y_i - \boldsymbol{x}_i' \widetilde{\boldsymbol{\beta}})^2 = (\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}).$$

Эту сумму называют также **остаточной суммой квадратов** (*residual sum of squares* — **RSS**), или **суммой квадратов ошибок** (*error sum of squares* — **ESS**). Сумма квадратов остатков является функцией от $\widetilde{\boldsymbol{\beta}}$, так как от него зависят остатки. **OLS-оценка**, b, вектора параметров β — это такой вектор $\tilde{\beta}$, который минимизирует эту функцию:

$$b \equiv \underset{\tilde{\beta}}{\operatorname{argmin}} SSR(\tilde{\beta}). \tag{1.2.1}$$

Взаимосвязь между неизвестным вектором коэффициентов β, его OLS-оценкой b и гипотетическим значением β ($\tilde{\beta}$) проиллюстрирована на рис. 1.1 для $K = 1$. Так как $SSR(\tilde{\beta})$ — квадратичная функция по $\tilde{\beta}$, ее график имеет U-образную форму. Значение $\tilde{\beta}$, соответствующее минимуму, — это OLS-оценка b. Поскольку она зависит от выборки $(\boldsymbol{y}, \boldsymbol{X})$, OLS-оценка b в общем случае отличается от «истинного» значения β. Равенство может достигаться только по чистой случайности.

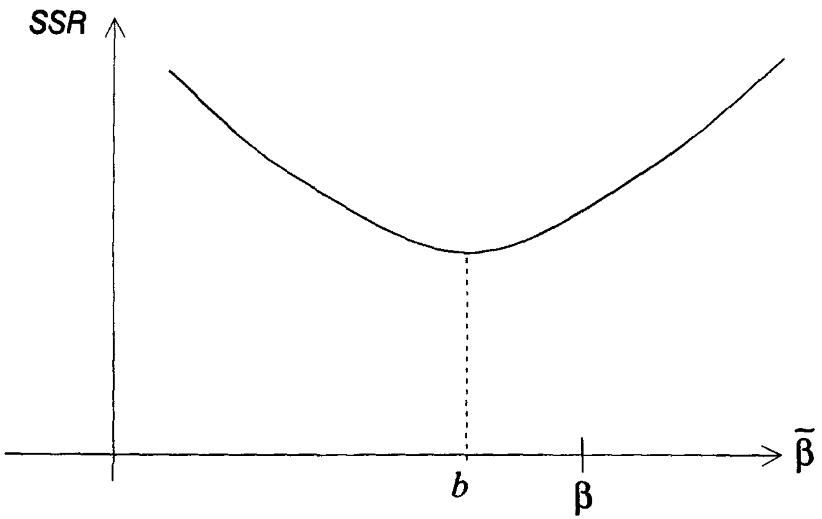

Рис. 1.1. Гипотетическое, истинное и оцененное значения

Поскольку в целевой функции используются квадраты остатков, этот метод приписывает больший «штраф» большим значениям остатков. OLS-оценка используется, чтобы предотвратить большие значения остатков для некоторых наблюдений, допуская при этом небольшие остатки для большого числа наблюдений. В следующем параграфе мы увидим, что применение именно такого критерия приводит к некоторым хорошим свойствам оценки.

Нормальные уравнения

Классический способ решения задачи минимизации состоит в получении условий первого порядка приравниванием частных производных нулю. Для этого необходимо найти K-мерный вектор частных производных

$\partial SSR(\widetilde{\boldsymbol{\beta}})/\partial\widetilde{\boldsymbol{\beta}}$[1]. Задача упрощается, если записать $SSR(\widetilde{\boldsymbol{\beta}})$ иначе:

$$
\begin{aligned}
SSR(\widetilde{\boldsymbol{\beta}}) &= (\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}) \\
&\quad (\text{так как } i\text{-й элемент } \boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}} \text{ — это } y_i - \boldsymbol{x}_i'\widetilde{\boldsymbol{\beta}}) \\
&= (\boldsymbol{y}' - \widetilde{\boldsymbol{\beta}}'\boldsymbol{X}')(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}) \quad (\text{поскольку } (\boldsymbol{X}\widetilde{\boldsymbol{\beta}})' = \widetilde{\boldsymbol{\beta}}'\boldsymbol{X}') = \\
&= \boldsymbol{y}'\boldsymbol{y} - \widetilde{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{X}\widetilde{\boldsymbol{\beta}} + \widetilde{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\widetilde{\boldsymbol{\beta}} = \\
&= \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{y}'\boldsymbol{X}\widetilde{\boldsymbol{\beta}} + \widetilde{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\widetilde{\boldsymbol{\beta}} \\
&\quad (\widetilde{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{y}'\boldsymbol{X}\widetilde{\boldsymbol{\beta}}, \text{ так как значение скаляра} \\
&\qquad \text{при транспонировании не изменяется}) \\
&\equiv \boldsymbol{y}'\boldsymbol{y} - 2\boldsymbol{a}'\widetilde{\boldsymbol{\beta}} + \widetilde{\boldsymbol{\beta}}'\boldsymbol{A}\widetilde{\boldsymbol{\beta}}, \quad \text{где } \boldsymbol{a} \equiv \boldsymbol{X}'\boldsymbol{y} \text{ и } \boldsymbol{A} \equiv \boldsymbol{X}'\boldsymbol{X}. \qquad (1.2.2)
\end{aligned}
$$

Первое слагаемое $\boldsymbol{y}'\boldsymbol{y}$ не зависит от $\widetilde{\boldsymbol{\beta}}$, поэтому им можно пренебречь при дифференцировании $SSR(\widetilde{\boldsymbol{\beta}})$. Вспомним из линейной алгебры, что

$$\frac{\partial(\boldsymbol{a}'\widetilde{\boldsymbol{\beta}})}{\partial\widetilde{\boldsymbol{\beta}}} = \boldsymbol{a} \quad \text{и} \quad \frac{\partial(\widetilde{\boldsymbol{\beta}}'\boldsymbol{A}\widetilde{\boldsymbol{\beta}})}{\partial\widetilde{\boldsymbol{\beta}}} = 2\boldsymbol{A}\widetilde{\boldsymbol{\beta}} \quad \text{для симметричной матрицы } \boldsymbol{A},$$

так что K-мерный вектор частных производных равен

$$\frac{\partial SSR(\widetilde{\boldsymbol{\beta}})}{\partial\widetilde{\boldsymbol{\beta}}} = -2\boldsymbol{a} + 2\boldsymbol{A}\widetilde{\boldsymbol{\beta}}.$$

Условия первого порядка получаются приравниванием этого выражения нулю. Применяя (1.2.2) к нашей задаче, заменяя вектор $\boldsymbol{a}$ на $\boldsymbol{X}'\boldsymbol{y}$, $\boldsymbol{A}$ — на $\boldsymbol{X}'\boldsymbol{X}$ и перегруппировывая, получаем условия первого порядка:

$$\underset{(K\times K)}{\boldsymbol{X}'\boldsymbol{X}}\ \underset{(K\times 1)}{\boldsymbol{b}} = \boldsymbol{X}'\boldsymbol{y}. \qquad (1.2.3)$$

Мы заменили $\widetilde{\boldsymbol{\beta}}$ на $\boldsymbol{b}$, поскольку OLS-оценка $\boldsymbol{b}$ — это значение $\widetilde{\boldsymbol{\beta}}$, удовлетворяющее условиям первого порядка. Эти K уравнений называются **нормальными уравнениями**.

Вектор остатков, вычисленный в точке $\widetilde{\boldsymbol{\beta}} = \boldsymbol{b}$,

$$\underset{(n\times 1)}{\boldsymbol{e}} \equiv \boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}. \qquad (1.2.4)$$

[1]Если $h\colon \mathbb{R}^K \to \mathbb{R}$ — скалярная функция от K-мерного вектора $\boldsymbol{x}$, производная h по $\boldsymbol{x}$ представляет собой K-мерный вектор, k-й элемент которого равен $\partial h(\boldsymbol{x})/\partial x_k$, где x_k — k-й элемент $\boldsymbol{x}$. (Этот K-мерный вектор называется **градиентом**.) В нашем случае $\boldsymbol{x}$ — это $\widetilde{\boldsymbol{\beta}}$, а функция $h(\boldsymbol{x})$ — $SSR(\widetilde{\boldsymbol{\beta}})$.

называется вектором **OLS остатков** (*OLS residuals*). Его i-й элемент $e_i \equiv y_i - \boldsymbol{x}_i'\boldsymbol{b}$. Перегруппировывая (1.2.3), получаем:

$$\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}) = \boldsymbol{0}, \quad \text{или} \quad \boldsymbol{X}'\boldsymbol{e} = \boldsymbol{0}, \quad \text{или}$$
$$\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \cdot e_i = \boldsymbol{0}, \quad \text{или} \quad \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \cdot (y_i - \boldsymbol{x}_i'\boldsymbol{b}) = \boldsymbol{0}. \tag{1.2.3'}$$

это уравнение показывает, что нормальные уравнения могут быть проинтерпретированы как выборочные аналоги условий ортогональности $\mathrm{E}(\boldsymbol{x}_i \cdot \varepsilon_i) = \boldsymbol{0}$. Эта идея будет более подробно обсуждаться в последующих главах.

Однако условия первого порядка являются лишь необходимыми условиями минимума. Следовательно, чтобы убедиться, что в $\boldsymbol{b}$ достигается минимум, а не максимум, нам необходимо проверить условия второго порядка. Читатели, которые знакомы с понятием гессиана[1] функции нескольких переменных[2], могут сразу понять, что условия второго порядка выполнены, поскольку (как отмечено ниже) матрица $\boldsymbol{X}'\boldsymbol{X}$ положительно определена. Однако то, что в $\boldsymbol{b}$ достигается минимум, возможно доказать и непосредственно. Для этого необходимо вычесть и прибавить к целевой функции одну и ту же величину (такая стратегия эффективна в случае квадратичной целевой функции, как здесь). Применение этой стратегии к алгебре метода наименьших квадратов оставлено читателям в качестве аналитического упражнения.

Два выражения для OLS-оценки

Итак, мы получили систему из K линейных одновременных уравнений с K неизвестными, составляющими вектор $\boldsymbol{b}$. По предположению 1.3 (об отсутствии мультиколлинеарности), матрица коэффициентов $\boldsymbol{X}'\boldsymbol{X}$ положительно определена (по поводу доказательства этого утверждения см. контрольный вопрос 1 текущего параграфа) и, следовательно, невырождена. То есть система нормальных уравнений имеет единственное решение $\boldsymbol{b}$, которое можно найти, умножив обе части (1.2.3) слева на $(\boldsymbol{X}'\boldsymbol{X})^{-1}$:

$$\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}. \tag{1.2.5}$$

Рассматривая (1.2.5) как функцию от выборки $(\boldsymbol{y}, \boldsymbol{X})$, это выражение иногда называют **OLS-оценивателем** (*OLS-estimator*). Для любой конкретной выборки $(\boldsymbol{y}, \boldsymbol{X})$ значение этой функции называется **OLS-оцен-**

[1] В оригинале *Hessian*. Этот термин автор употребляет, имея в виду именно матрицу вторых производных, которую в русскоязычной литературе часто именуют ***матрицей Гессе***. – *Прим. науч. ред. перевода.*

[2] **Гессиан** функции $h(\boldsymbol{x})$ представляет собой квадратную матрицу, в которой элемент (k, l) равен $\partial^2 h(\boldsymbol{x})/\partial x_k \partial x_l$.

кой (*OLS-estimate*). В этой книге (как и в большинстве других) эти термины будут использоваться почти как синонимы[1].

Поскольку $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = (\boldsymbol{X}'\boldsymbol{X}/n)^{-1}\boldsymbol{X}'\boldsymbol{y}/n$, OLS-оценка может быть также записана как

$$\boldsymbol{b} = \boldsymbol{S}_{\boldsymbol{xx}}^{-1}\boldsymbol{s}_{\boldsymbol{xy}}, \text{ где} \tag{1.2.5$'$}$$

$$\boldsymbol{S}_{\boldsymbol{xx}} = \frac{1}{n}\boldsymbol{X}'\boldsymbol{X} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i' \quad (\text{выборочное среднее } \boldsymbol{x}_i\boldsymbol{x}_i'), \tag{1.2.6a}$$

$$\boldsymbol{s}_{\boldsymbol{xy}} = \frac{1}{n}\boldsymbol{X}'\boldsymbol{y} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i \cdot y_i \quad (\text{выборочное среднее } \boldsymbol{x}_i \cdot y_i). \tag{1.2.6b}$$

Использовать матрицу данных из (1.2.5) более удобно для вывода свойств оценки в малых выборках, а использование выборочных средних из представления (1.2.5$'$) — для теории больших выборок.

Еще несколько терминов из алгебры

После того как мы вывели OLS-оценку, следует определить еще несколько терминов.

- **Подобранным значением** (*fitted value*) объясняющей переменной для наблюдения i называется $\widehat{y}_i \equiv \boldsymbol{x}_i'\boldsymbol{b}$. Вектор подобранных значений $\widehat{\boldsymbol{y}}$ равен $\boldsymbol{Xb}$. Таким образом, вектор OLS остатков можно записать как $\boldsymbol{e} = \boldsymbol{y} - \widehat{\boldsymbol{y}}$.

- Определим **проекционную матрицу** (*projection matrix*) $\boldsymbol{P}$ и **аннулятор** (*annihilator*) $\boldsymbol{M}$:

$$\underset{(n\times n)}{\boldsymbol{P}} \equiv \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'. \tag{1.2.7}$$

$$\underset{(n\times n)}{\boldsymbol{M}} \equiv \boldsymbol{I}_n - \boldsymbol{P}. \tag{1.2.8}$$

Они обладают рядом привлекательных свойств (доказательство которых оставлено читателям в качестве одного из контрольных вопросов):

$$\text{Матрицы } \boldsymbol{P} \text{ и } \boldsymbol{M} \text{ симметричны и идемпотентны}^2, \tag{1.2.9}$$

$$\boldsymbol{PX} = \boldsymbol{X} \quad (\text{отсюда термин «\textbf{проекционная матрица}»}). \tag{1.2.10}$$

$$\boldsymbol{MX} = \boldsymbol{0} \quad (\text{отсюда термин «\textbf{аннулятор}»}). \tag{1.2.11}$$

[1]Поскольку автор указывает на то, что упомянутые два термина используются в тексте почти как синонимы, мы употребляем при переводе термин «оценка» как для *OLS-estimator*, так и для *estimate*, что соответствует традиции, принятой в русскоязычной литературе. -- *Прим. науч. ред. перевода.*

[2]Квадратная матрица $\boldsymbol{A}$ называется **идемпотентной**, если $\boldsymbol{A} = \boldsymbol{A}^2$.

Поскольку e — вектор остатков при $\tilde{\beta} = b$, то сумма квадратов остатков OLS (SSR) равна $e'e$. Ее можно также переписать как:

$$SSR = e'e = \varepsilon' M \varepsilon. \tag{1.2.12}$$

(Доказательство этого оставлено в качестве контрольного вопроса.) Это выражение, связывающее SSR и истинные ошибки ε, будет использовано далее.

- OLS-оценка σ^2 (дисперсии ошибки), которую обозначим s^2, равна сумме квадратов остатков, деленной на $n - K$:

$$s^2 \equiv \frac{SSR}{n - K} = \frac{e'e}{n - K}. \tag{1.2.13}$$

(В этом определении предполагается, что $n > K$. В противном случае s^2 не определена.) Как будет показано в утверждении 1.2, приведенном ниже, при делении суммы квадратов остатков на $n-K$ (**число степеней свободы**), а не на n (размер выборки), оценка σ^2 оказывается несмещенной. Интуитивное объяснение состоит в том, что сначала необходимо оценить K параметров (β) и лишь потом получить вектор остатков e, необходимый для вычисления s^2. С формальной точки зрения e должен удовлетворять системе K нормальных уравнений (1.2.3′), которая накладывает ограничения на возможные значения остатков.

- Квадратный корень из s^2 (s) называется **стандартной ошибкой регрессии** (*standard error of regression* — **SER**), или **стандартной ошибкой уравнения** (*standard error of equation* — **SEE**). SER представляет собой оценку стандартного отклонения ошибок.

- **Ошибка оценки** (*sampling error*)[1] определяется как $b-\beta$. Ее также можно связать с ε:

$$\begin{aligned} b - \beta &= (X'X)^{-1}X'y - \beta = \quad \text{(из (1.2.5))} \\ &= (X'X)^{-1}X'(X\beta + \varepsilon) - \beta \\ &\quad \text{(так как } y = X\beta + \varepsilon \text{ по предположению 1.1)} \\ &= (X'X)^{-1}(X'X)\beta + (X'X)^{-1}X'\varepsilon - \beta = \\ &= \beta + (X'X)^{-1}X'\varepsilon - \beta = (X'X)^{-1}X'\varepsilon. \end{aligned} \tag{1.2.14}$$

[1]В оригинале *sampling error* (буквально: ошибка выборки). Мы используем при переводе термин «ошибка оценки», поскольку в англоязычной литературе термин *sampling error* и соответствующий ему в русскоязычной литературе термин «ошибка выборки» обычно означают нечто иное (например, систематическую ошибку выборки). — *Прим. науч. ред. перевода.*

- **Нецентрированный R^2.** Одной из мер изменчивости зависимой переменной является сумма квадратов $\sum y_i^2 = y'y$. Поскольку OLS остатки удовлетворяют системе нормальных уравнений, мы получаем следующую декомпозицию $y'y$:

$$\begin{aligned}
y'y &= (\widehat{y} + e)'(\widehat{y} + e) \quad (\text{так как } e = y - \widehat{y}) = \\
&= \widehat{y}'\widehat{y} + 2\widehat{y}'e + e'e = \\
&= \widehat{y}'\widehat{y} + 2b'X'e + e'e \quad (\text{поскольку } \widehat{y} \equiv Xb) \\
&= \widehat{y}'\widehat{y} + e'e
\end{aligned} \tag{1.2.15}$$

(так как $X'e = \mathbf{0}$ из системы нормальных уравнений; см. (1.2.3′)).

Нецентрированный R^2 (*uncentered* R^2) определяется как

$$R^2_{uc} \equiv 1 - \frac{e'e}{y'y}. \tag{1.2.16}$$

Применяя декомпозицию (1.2.15), получаем, что R^2_{uc} равен:

$$\frac{\widehat{y}'\widehat{y}}{y'y}.$$

Поскольку $\widehat{y}'\widehat{y}$ и $e'e$ неотрицательны, $0 \leq R^2_{uc} \leq 1$. Таким образом, нецентрированный R^2 может быть проинтерпретирован как доля изменчивости зависимой переменной, объясненная независимыми переменными. Чем ближе подобранные значения к значениям зависимой переменной, тем ближе нецентрированный R^2 к единице.

- **(Центрированный) R^2, коэффициент детерминации.** Если единственный регрессор в модели — константа (так что $K = 1$ и $x_{i1} = 1$), то тогда из (1.2.5) вытекает, что $b = \overline{y}$ ($\overline{y}$ — выборочное среднее зависимой переменной). Из этого следует, что $\widehat{y}_i = \overline{y}_i$ для всех i; в (1.2.15) $\widehat{y}'\widehat{y} = n\overline{y}^2$, и $e'e = \sum_i (y_i - \overline{y})^2$. Если же в модели, кроме константы, присутствуют и другие регрессоры, то можно показать (доказательство оставлено в качестве аналитического упражнения), что $\sum_i (y_i - \overline{y})^2$ может быть разложена как

$$\sum_{i=1}^{n} (y_i - \overline{y})^2 = \sum_{i=1}^{n} (\widehat{y}_i - \overline{y})^2 + \sum_{i=1}^{n} e_i^2 \quad , \text{где } \overline{y} \equiv \frac{1}{n}\sum_{i=1}^{n} y_i. \tag{1.2.17}$$

Коэффициент детерминации R^2 определяется следующим образом:

$$R^2 \equiv 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2}. \tag{1.2.18}$$

В силу декомпозиции (1.2.17), R^2 равен:

$$\frac{\sum_{i=1}^{n}(\widehat{y}_i - \overline{y})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}.$$

Поэтому, если регрессоры включают константу (так что разложение (1.2.17) выполнено), $0 \leq R^2 \leq 1$. Таким образом, R^2, определенный в (1.2.18), является мерой, объясняющей способности регрессоров, не являющихся константой.

Если среди регрессоров нет константы, то при вычислении R^2 по формуле (1.2.18) (как делают некоторые программные пакеты) R^2 может получиться отрицательным. Это происходит потому, что регрессия без константы может аппроксимировать объясняемую переменную хуже, чем выборочное среднее. С другой стороны, некоторые эконометрические пакеты (например, STATA) при исключении константы из модели автоматически используют формулу (1.2.16) для расчета коэффициента детерминации во избежание получения отрицательных значений R^2. Это палка о двух концах. Предположим, что среди регрессоров нет константы, однако линейная комбинация регрессоров равна константе. Такое может произойти, например, когда константу в модели заменяют набором сезонных дамми-переменных[1]. Такая регрессия по сути не изменится, если один из регрессоров в линейной комбинации заменить на константу. Более того, не изменится вектор подобранных значений. Однако, если R^2 рассчитывается по формуле (1.2.16) для регрессии без константы и по формуле (1.2.18) для регрессии с константой, R^2 будет меньше при добавлении константы (см. контрольный вопрос 7 ниже).

Анализ влиятельных наблюдений (дополнительно)

Поскольку метод наименьших квадратов стремится предотвратить появление нескольких больших остатков, допуская большое количество сравнительно малых остатков, только несколько наблюдений могут быть чрезвычайно влиятельными, в том смысле, что при исключении таких наблюдений из выборки некоторые элементы вектора параметров $\boldsymbol{b}$ могут значительно измениться. Существует систематический подход к поиску таких наблюдений, называемых **влиятельными наблюдениями** (*influential observations*)[2]. Обозначим через $\boldsymbol{b}^{(i)}$ OLS-оценку вектора коэффициентов $\boldsymbol{\beta}$, полученную с использованием выборки, из которой было исключено i-е наблюдение. Ключевым для дальнейшего анализа является

[1] Понятие дамми-переменной будет введено в эмпирических упражнениях к этой главе.

[2] Для более подробного ознакомления с темой см.: [Krasker, Kuh, and Welsh, 1983].

следующее уравнение:

$$b^{(i)} - b = -\left(\frac{1}{1-p_i}\right)(X'X)^{-1}x_i \cdot e_i, \tag{1.2.19}$$

где x_i', как и прежде, обозначает i-ю строку матрицы данных X, e_i — OLS остаток для наблюдения i, а p_i вычисляется по формуле:

$$p_i \equiv x_i'(X'X)^{-1}x_i, \tag{1.2.20}$$

то есть p_i является i-м диагональным элементом проекционной матрицы P. (Доказательство (1.2.19) было бы хорошим упражнением в матричной алгебре, но мы не будем приводить его здесь.) Нетрудно показать (см. контрольный вопрос 7 к параграфу 1.3), что

$$0 \le p_i \le 1 \quad \text{и} \quad \sum_{i=1}^{n} p_i = K. \tag{1.2.21}$$

Поэтому p_i в среднем равно K/n.

Для иллюстрации применения (1.2.19) на конкретном примере рассмотрим взаимосвязь между инвестициями в основной капитал и экономическим ростом в беднейших странах мира в период с 1960 по 1985 г. На рис. 1.2 изображена диаграмма рассеяния для 13 стран, ВВП которых на одного работающего в 1965 году составлял менее 10% аналогичного показателя в США[1]. По горизонтальной оси откладывается среднегодовое отношение инвестиций в основной капитал к ВВП за период с 1960 по 1985 г. По вертикальной — среднегодовой рост ВВП на работающего за тот же период. На основе визуального анализа рисунка нетрудно заметить, что положение оцененной линии регрессии в значительной мере будет зависеть от единственного выделяющегося наблюдения (Ботсваны). Действительно, при исключении Ботсваны из выборки, оцененный коэффициент наклона падает с 0,37 до 0,058. В случае парной регрессии (как здесь) выделяющиеся наблюдения легко обнаружить с помощью визуального анализа диаграммы рассеяния, подобной рис. 1.2. Однако эта стратегия не может быть использована в случае, когда имеется более одного регрессора, отличного от постоянной. Анализ, основанный на формуле (1.2.19), не ограничен парной регрессией. В табл. 1.1 для каждого наблюдения приведены OLS остатки, значения p_i и значения (1.2.19). Значение p_i для Ботсваны, равное 0,7196, намного превосходит среднее, равное 0,154 ($= K/n = 2/13$). То есть это наблюдение является весьма **влиятельным**, о чем свидетельствуют два последних столбца таблицы. Обратите внимание, что мы бы не смогли обнаружить это влиятельное

[1] Данные были взяты из Penn World Table, перепечатанной в: [DeLong and Summers, 1991]. Однако в их анализе использовалась вся выборка, включающая 61 страну.

наблюдение, рассматривая остатки. Это неудивительно, поскольку метод наименьших квадратов алгебраически устроен так, чтобы не допускать больших значений остатков (за счет большого количества малых остатков для остальных наблюдений).

Рис. 1.2. Инвестиции в основной капитал и рост

Таблица 1.1

Анализ влияния

Страна	**Рост ВВП на одного работающего**	**Инвестиции в основной капитал/ВВП**	**Остаток**	p_i	**(1.2.19) для β_1**	**(1.2.19) для β_2**
Ботсвана	0,0676	0,1310	0,0119	0,7196	0,0104	−0,3124
Камерун	0,0458	0,0415	0,0233	0,0773	−0,0021	0,0045
Эфиопия	0,0094	0,0212	−0,0056	0,1193	0,0010	−0,0119
Индия	0,0115	0,0278	−0,0059	0,0980	0,0009	−0,0087
Индонезия	0,0345	0,0221	0,0192	0,1160	−0,0034	0,0394
Кот-д'Ивуар	0,0278	0,0243	0,0117	0,1084	−0,0019	0,0213
Кения	0,0146	0,0462	−0,0096	0,0775	0,0007	0,0023
Мадагаскар	−0,0102	0,0219	−0,0254	0,1167	0,0045	−0,0527
Малави	0,0153	0,0361	−0,0052	0,0817	0,0006	−0,0036
Мали	0,0044	0,0433	−0,0188	0,0769	0,0016	−0,0006
Пакистан	0,0295	0,0263	0,0126	0,1022	−0,0020	0,0205
Танзания	0,0184	0,0860	−0,0206	0,2281	−0,0021	0,0952
Таиланд	0,0341	0,0395	0,0123	0,0784	−0,0012	0,0047

Что делать с выделяющимися наблюдениями? Смотря по обстоятельствам. Если выделяющиеся наблюдения удовлетворяют регрессионной модели, они содержат в себе ценную информацию о выборке, которая не содержится в других наблюдениях. В таком случае их нельзя исключать. Однако более вероятным является то, что выделяющиеся наблюдения — это наблюдения, нетипичные для выборки, из-за того, что они не удовлетворяют модели. Тогда они однозначно должны быть исключены из выборки. В рассматриваемом примере наблюдался рост мирового спроса на алмазы (главную статью экспорта Ботсваны), а для их добычи необходима дорогостоящая буровая техника. Если мы предполагаем, что за эконометрической моделью стоит следующая логическая цепочка: инвестиции означают внедрение новых технологий, которые, в свою очередь, ведут к росту производительности труда и, следовательно, к экономическому росту, — тогда Ботсвана, высокие темпы роста ВВП которой были обусловлены повышенным спросом, должна быть исключена из выборки.

Замечание относительно вычисления OLS-оценок[1]

До сих пор основное внимание уделялось концептуальным аспектам алгебры OLS. Но прикладным исследователям, которые на практике вычисляют OLS-оценки с использованием компьютеров, важно знать некоторые особенности компьютерных вычислений, чтобы избежать возможности получения недостоверных результатов. Потенциальным источником проблем может быть то, что компьютеры аппроксимируют натуральные числа так называемыми **числами с плавающей запятой**. Когда в арифметической операции участвуют и очень большие, и очень маленькие числа, использование чисел с плавающей запятой может привести к неточным результатам. При вычислении OLS-оценки такая ситуация может возникнуть, когда регрессоры значительно отличаются по величине. Например, одним из регрессоров может быть процентная ставка, выраженная дробью, а другим — ВВП США в долларах. Матрица $\boldsymbol{X}'\boldsymbol{X}$ будет тогда содержать как очень маленькие, так и очень большие числа, и операция вычисления обратной матрицы может дать недостоверный результат.

Эту проблему можно решить путем выбора единиц измерения таким образом, чтобы регрессоры были близки по величине. Например, указывать ставку процента в процентах, а ВВП США — в триллионах долларов. Такая мера предосторожности в большинстве случаев помогает избежать проблем. Более систематический подход к преобразованию матрицы $\boldsymbol{X}$ состоит в вычитании выборочных средних из всех регрессоров и делении на выборочные стандартные ошибки. Стандартизация переменных проводится до формирования $\boldsymbol{X}'\boldsymbol{X}$. Полученные OLS-оценки надо скорректировать, переводя их в масштаб исходных данных. Большинство компьютерных пакетов, вычисляющих OLS-оценки (например, TSP), используют более сложное преобразование матрицы $\boldsymbol{X}$ (которое называется **QR-декомпозицией**).

[1] Более подробно эта тема изложена в Section 1.5 [Davidson and MacKinnon, 1993].

Контрольные вопросы

1. Докажите, что матрица $\boldsymbol{X}'\boldsymbol{X}$ положительно определена, если матрица $\boldsymbol{X}$ имеет полный столбцовый ранг. **Указание:** *Необходимо показать, что* $\boldsymbol{c}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{c} > 0$ *для* $\boldsymbol{c} \neq \boldsymbol{0}$. *Определим* $\boldsymbol{z} \equiv \boldsymbol{X}\boldsymbol{c}$. *Тогда* $\boldsymbol{c}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{c} = \boldsymbol{z}'\boldsymbol{z} = \sum_{i=1}^{n} z_i^2$. *Если* $\boldsymbol{X}$ — *матрица полного столбцового ранга, то* $\boldsymbol{z} \neq \boldsymbol{0}$ *для всех* $\boldsymbol{c} \neq \boldsymbol{0}$.

2. Проверьте, что $\boldsymbol{X}'\boldsymbol{X}/n = \frac{1}{n}\sum_i \boldsymbol{x}_i\boldsymbol{x}_i'$ и $\boldsymbol{X}'\boldsymbol{y}/n = \frac{1}{n}\sum_i \boldsymbol{x}_i \cdot \boldsymbol{y}_i$ как в (1.2.6). **Указание:** *Элемент* (k, l) *матрицы* $\boldsymbol{X}'\boldsymbol{X}$ *равен* $\sum_i x_{ik}x_{il}$.

3. (OLS-оценка для простой регрессионной модели.) В простой регрессионной модели $K = 2$ и $x_{i1} = 1$. Покажите, что

$$\boldsymbol{S}_{\boldsymbol{XX}} = \begin{bmatrix} 1 & \overline{x}_2 \\ \overline{x}_2 & \frac{1}{n}\sum_{i=1}^{n} x_{i2}^2 \end{bmatrix}, \quad \boldsymbol{s}_{\boldsymbol{XY}} = \begin{bmatrix} \overline{y} \\ \frac{1}{n}\sum_{i=1}^{n} x_{i2}y_i \end{bmatrix},$$

где

$$\overline{y} \equiv \frac{1}{n}\sum_{i=1}^{n} y_i \quad \text{и} \quad \overline{x}_2 \equiv \frac{1}{n}\sum_{i=1}^{n} x_{i2}.$$

Покажите, что

$$b_2 = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_{i2} - \overline{x}_2)(y_i - \overline{y})}{\frac{1}{n}\sum_{i=1}^{n}(x_{i2} - \overline{x}_2)^2} \quad \text{и} \quad b_1 = \overline{y} - \overline{x}_2 b_2.$$

(Можно заметить, что в знаменателе выражения для b_2 стоит выборочная дисперсия регрессора, отличного от константы. А в числителе находится выборочная ковариация зависимой переменной и регрессора, отличного от константы.) **Указание:**

$$\frac{1}{n}\sum_{i=1}^{n} x_{i2}^2 - (\overline{x}_2)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_{i2} - \overline{x}_2)^2$$

и

$$\frac{1}{n}\sum_{i=1}^{n} x_{i2}y_i - \overline{x}_2\overline{y} = \frac{1}{n}\sum_{i=1}^{n}(x_{i2} - \overline{x}_2)(y_i - \overline{y}).$$

Вы можете взять уравнение (1.2.5′) и просто произвести обращение матрицы. Однако есть и другой путь: выписать систему из двух нормальных уравнений. Первое уравнение системы будет иметь следующий вид: $b_1 = \overline{y} - \overline{x}_2 b_2$. *Подставьте это выражение для* b_1 *во второе уравнение и найдите* b_2.

4. Докажите (1.2.9)–(1.2.11). **Указание:** *Эти свойства выводятся напрямую из определений* $\boldsymbol{P}$ *и* $\boldsymbol{M}$.

5. (Матричная алгебра для подобранных значений и остатков.) Покажите следующее:

   (a) $\widehat{\boldsymbol{y}} = \boldsymbol{P}\boldsymbol{y}$, $\boldsymbol{e} = \boldsymbol{M}\boldsymbol{y} = \boldsymbol{M}\boldsymbol{\varepsilon}$. **Указание:** *Используйте* (1.2.5).

   (b) (1.2.12), а именно: $SSR = \boldsymbol{\varepsilon}'\boldsymbol{M}\boldsymbol{\varepsilon}$.

6. (Изменение единиц измерения и R^2.) Влияет ли изменение единиц измерения зависимой переменной на R^2? А изменение единиц измерения регрессоров? **Указание:** *Проверьте, влияют ли вышеуказанные изменения на числитель и знаменатель в определении R^2.*

7. (Взаимосвязь между R^2_{uc} и R^2.) Покажите, что

$$1 - R^2 = \left(1 + \frac{n \cdot \overline{y}^2}{\sum_{i=1}^n (y_i - \overline{y})^2}\right)(1 - R^2_{uc}).$$

Указание: *Используйте* (1.2.16), (1.2.18) *и тождество* $\sum_{i=1}^n (y_i - \overline{y})^2 = \sum_{i=1}^n y_i^2 - n \cdot \overline{y}^2$.

8. Покажите, что

$$R^2_{uc} = \frac{\boldsymbol{y}'\boldsymbol{P}\boldsymbol{y}}{\boldsymbol{y}'\boldsymbol{y}}.$$

9. (Вычисление статистик.) Проверьте, что $\boldsymbol{b}$, SSR, s^2 и R^2 можно вычислить, используя следующие выборочные средние: $\boldsymbol{S_{XX}}$, $\boldsymbol{s_{XY}}$, $\boldsymbol{y}'\boldsymbol{y}/n$ и $\overline{y}$. (Если константа включена в модель, то $\overline{y}$ является элементом $\boldsymbol{s_{XY}}$, соответствующим константе.) Поэтому для получения коэффициентов регрессии и связанных с ними статистик эти выборочные средние достаточно вычислить только один раз.

1.3. Свойства OLS в конечных выборках

Мы уже вывели OLS-оценку, теперь нам предстоит исследовать ее свойства в конечных выборках, а именно свойства распределения оценки, которые имеют место для любого объема выборки n.

Распределение b в конечных выборках

Утверждение 1.1 (свойства OLS-оценки β в конечных выборках):

(a) *(Несмещенность.) При выполнении предположений 1.1–1.3,* $\mathrm{E}(\boldsymbol{b}|\boldsymbol{X}) = \boldsymbol{\beta}$.

(b) *(Выражение для дисперсии[1].) При выполнении предположений 1.1–1.4,* $\mathrm{Var}(\boldsymbol{b}|\boldsymbol{X}) = \sigma^2 \cdot (\boldsymbol{X}'\boldsymbol{X})^{-1}$.

(c) *(Теорема Гаусса — Маркова.) При выполнении предположений 1.1–1.4, OLS-оценка является **«эффективной»** в классе линейных несмещен-*

[1]Здесь и далее у автора «variance» — «дисперсия», хотя в действительности имеется в виду ковариационная матрица случайного вектора (которую также называют дисперсионной или дисперсионно-ковариационной). — *Прим. науч. ред. перевода.*

ных оценок. То есть для любой несмещенной оценки $\widehat{\boldsymbol{\beta}}$, *линейной по* $\boldsymbol{y}$, $\mathrm{Var}(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}) \geq \mathrm{Var}(\boldsymbol{b}|\boldsymbol{X})$ *в матричном смысле*[1].

(d) *При выполнении предположений 1.1–1.4,* $\mathrm{Cov}(\boldsymbol{b}, \boldsymbol{e}|\boldsymbol{X}) = \mathbf{0}$[2], *где* $\boldsymbol{e} \equiv \boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}$.

Прежде чем перейти к доказательству, остановимся подробнее на смысле этого утверждения.

- Матричное неравенство в пункте (c) говорит, что матрица $\mathrm{Var}(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}) - \mathrm{Var}(\boldsymbol{b}|\boldsymbol{X})$ размерности $K \times K$ положительно полуопределена. Значит,
$$\boldsymbol{a}'[\mathrm{Var}(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}) - \mathrm{Var}(\boldsymbol{b}|\boldsymbol{X})]\boldsymbol{a} \geq 0$$
или
$$\boldsymbol{a}'\,\mathrm{Var}(\widehat{\boldsymbol{\beta}}|\boldsymbol{X})\boldsymbol{a} \geq \boldsymbol{a}'\,\mathrm{Var}(\boldsymbol{b}|\boldsymbol{X})\boldsymbol{a}$$
для любого K-мерного вектора $\boldsymbol{a}$. В частности, и для вектора, состоящего из нулей, кроме k-го элемента, равного единице. Для такого вектора $\boldsymbol{a}$ квадратичная форма $\boldsymbol{a}'\boldsymbol{A}\boldsymbol{a}$ выбирает элемент (k, k) из матрицы $\boldsymbol{A}$. Элемент (k, k) матрицы $\mathrm{Var}(\widehat{\boldsymbol{\beta}}|\boldsymbol{X})$, например, равен $\mathrm{Var}(\widehat{\beta}_k|\boldsymbol{X})$, где $\widehat{\beta}_k$ — k-й элемент вектора $\widehat{\boldsymbol{\beta}}$. То есть из матричного неравенства (c) следует, что
$$\mathrm{Var}(\widehat{\beta}_k|\boldsymbol{X}) \geq \mathrm{Var}(b_k|\boldsymbol{X}) \quad (k = 1, 2, \ldots, K). \tag{1.3.1}$$
То есть для любого коэффициента регрессии дисперсия его OLS-оценки не превышает дисперсии любой альтернативной линейной несмещенной оценки.

- Из выражения (1.2.5) легко видеть, что OLS-оценка линейна по $\boldsymbol{y}$. Вообще существует много других линейных и несмещенных оценок $\boldsymbol{\beta}$ (в качестве одного из контрольных вопросов вам будет предложено построить одну из таких оценок). Теорема Гаусса — Маркова утверждает, что OLS-оценка является **эффективной** в том смысле, что ее условная дисперсия $\mathrm{Var}(\boldsymbol{b}|\boldsymbol{X})$ является наименьшей среди дисперсий всех линейных несмещенных оценок. По этой причине OLS-оценку называют **наилучшей линейной несмещенной оценкой** (*Best Linear Unbiased Estimator* — **BLUE**).

---

[1] Пусть $\boldsymbol{A}$ и $\boldsymbol{B}$ — две квадратные матрицы одинакового размера. Говорят, что $\boldsymbol{A} \geq \boldsymbol{B}$, если разность $\boldsymbol{A} - \boldsymbol{B}$ положительно полуопределена. Матрица $\boldsymbol{C}$ размерности $K \times K$ называется положительно полуопределенной (или неотрицательно определенной), если $\boldsymbol{x}'\boldsymbol{C}\boldsymbol{x} \geq 0$ для любого K-мерного вектора $\boldsymbol{x}$.

[2] Здесь и далее $\mathrm{Cov}(\boldsymbol{v}, \boldsymbol{w})$ — матрица взаимных ковариаций случайных векторов $\boldsymbol{v}$ и $\boldsymbol{w}$. — *Прим. науч. ред. перевода.*

- OLS-оценка $\boldsymbol{b}$ является функцией от выборки $(\boldsymbol{y}, \boldsymbol{X})$. Поскольку выборка $(\boldsymbol{y}, \boldsymbol{X})$ случайна, это же можно сказать и о $\boldsymbol{b}$. Теперь предположим, что мы зафиксировали конкретную матрицу данных $\boldsymbol{X}$, вычислили $\boldsymbol{b}$ для всех выборок, соответствующих всем возможным реализациям $\boldsymbol{y}$, а затем взяли среднее значение $\boldsymbol{b}$ (вас попросят сделать это в упражнении на метод Монте-Карло к этой главе). Это среднее представляет собой (теоретическое) условное математическое ожидание $\mathrm{E}(\boldsymbol{b}|\boldsymbol{X})$. Пункт (a) (несмещенность) говорит о том, что это математическое ожидание равно истинному значению $\boldsymbol{\beta}$.

- Существует еще одно определение несмещенности, которое является более слабым, чем указанное в пункте (a). По формуле полного математического ожидания, $\mathrm{E}\left[\mathrm{E}(\boldsymbol{b}|\boldsymbol{X})\right] = \mathrm{E}(\boldsymbol{b})$. То есть из (a) следует
$$\mathrm{E}(\boldsymbol{b}) = \boldsymbol{\beta}. \tag{1.3.2}$$
Это уравнение утверждает, что если мы вычислим $\boldsymbol{b}$ для всех возможных выборок, различающихся не только зависимой переменной $\boldsymbol{y}$, но и $\boldsymbol{X}$, то его среднее будет равно истинному значению. Это безусловное утверждение, скорее всего, является более применимым к экономике, так как на практике выборки отличаются как $\boldsymbol{y}$, так и $\boldsymbol{X}$. То есть важность условного утверждения (a) состоит в том, что из него следует безусловный аналог (1.3.2), который, в свою очередь, ближе к практике.

- Те же рассуждения применимы к утверждению (c) относительно дисперсии. В контрольном вопросе, приведенном далее, вам будет предложено показать, что из утверждений (a) и (b) следует, что
$$\mathrm{Var}(\widehat{\boldsymbol{\beta}}) \geq \mathrm{Var}(\boldsymbol{b}), \tag{1.3.3}$$
где $\widehat{\boldsymbol{\beta}}$ — любая линейная несмещенная оценка (то есть $\mathrm{E}(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}) = \boldsymbol{\beta}$).

Теперь настало время перейти к доказательству этого важного результата. Может показаться, что оно слишком длинное, однако эта видимость создается лишь из-за его подробности. На самом деле это доказательство весьма простое. Тем не менее при первом чтении вам, возможно, захочется пропустить доказательство пункта (c). Доказательство части (d) оставлено читателям в качестве контрольных вопросов.

Доказательство

(a) (Доказательство того, что $\mathrm{E}(\boldsymbol{b}|\boldsymbol{X}) = \boldsymbol{\beta}$.) $\mathrm{E}(\boldsymbol{b}-\boldsymbol{\beta}|\boldsymbol{X}) = \boldsymbol{0}$ тогда и только тогда, когда $\mathrm{E}(\boldsymbol{b}|\boldsymbol{X}) = \boldsymbol{\beta}$. Итак, мы доказали первое из вышеупомянутых утверждений. Из выражения для ошибки оценки (1.2.14), $\boldsymbol{b} - \boldsymbol{\beta} = \boldsymbol{A}\boldsymbol{\varepsilon}$, где $\boldsymbol{A} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. То есть
$$\mathrm{E}(\boldsymbol{b} - \boldsymbol{\beta}|\boldsymbol{X}) = \mathrm{E}(\boldsymbol{A}\boldsymbol{\varepsilon}|\boldsymbol{X}) = \boldsymbol{A}\,\mathrm{E}(\boldsymbol{\varepsilon}|\boldsymbol{X}).$$

Второе равенство следует из линейности условного математического ожидания; $\boldsymbol{A}$ — функция от $\boldsymbol{X}$ и, следовательно, может рассматриваться как детерминированная матрица. Так как $\mathrm{E}(\boldsymbol{\varepsilon}|\boldsymbol{X}) = \mathbf{0}$, последнее выражение равно нулю.

(b) (Доказательство того, что $\mathrm{Var}(\boldsymbol{b}|\boldsymbol{X}) = \sigma^2 \cdot (\boldsymbol{X}'\boldsymbol{X})^{-1}$.)

$$\begin{aligned}
\mathrm{Var}(\boldsymbol{b}|\boldsymbol{X}) &= \mathrm{Var}(\boldsymbol{b} - \boldsymbol{\beta}|\boldsymbol{X}) \quad \text{(так как } \boldsymbol{\beta} \text{ детерминирован)} \\
&= \mathrm{Var}(\boldsymbol{A}\boldsymbol{\varepsilon}|\boldsymbol{X}) \quad \text{(из (1.2.14) и } \boldsymbol{A} \equiv (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}') = \\
&= \boldsymbol{A}\,\mathrm{Var}(\boldsymbol{\varepsilon}|\boldsymbol{X})\boldsymbol{A}' \quad \text{(поскольку } \boldsymbol{A} \text{ — функция от } \boldsymbol{X}) = \\
&= \boldsymbol{A}\,\mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\boldsymbol{X})\boldsymbol{A}' \quad \text{(по предположению 1.2)} = \\
&= \boldsymbol{A}(\sigma^2 \boldsymbol{I}_n)\boldsymbol{A}' \quad \text{(по предположению 1.4, см. (1.1.14))} = \\
&= \sigma^2 \boldsymbol{A}\boldsymbol{A}' = \\
&= \sigma^2 \cdot (\boldsymbol{X}'\boldsymbol{X})^{-1}
\end{aligned}$$

$$\text{(так как } \boldsymbol{A}\boldsymbol{A}' = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} = (\boldsymbol{X}'\boldsymbol{X})^{-1}).$$

(c) (Теорема Гаусса — Маркова.) Поскольку $\widehat{\boldsymbol{\beta}}$ линеен по $\boldsymbol{y}$, его можно записать как $\widehat{\boldsymbol{\beta}} = \boldsymbol{C}\boldsymbol{y}$ для некоторой матрицы $\boldsymbol{C}$, которая, возможно, является функцией от $\boldsymbol{X}$. Пусть $\boldsymbol{D} \equiv \boldsymbol{C} - \boldsymbol{A}$ или $\boldsymbol{C} = \boldsymbol{D} + \boldsymbol{A}$, где $\boldsymbol{A} \equiv (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. Тогда

$$\begin{aligned}
\widehat{\boldsymbol{\beta}} &= (\boldsymbol{D} + \boldsymbol{A})\boldsymbol{y} = \\
&= \boldsymbol{D}\boldsymbol{y} + \boldsymbol{A}\boldsymbol{y} = \\
&= \boldsymbol{D}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) + \boldsymbol{b} \\
&\text{(так как } \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \text{ и } \boldsymbol{A}\boldsymbol{y} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{b}) \\
&= \boldsymbol{D}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{D}\boldsymbol{\varepsilon} + \boldsymbol{b}.
\end{aligned}$$

После взятия условного математического ожидания от обеих частей получаем:

$$\mathrm{E}(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}) = \boldsymbol{D}\boldsymbol{X}\boldsymbol{\beta} + \mathrm{E}(\boldsymbol{D}\boldsymbol{\varepsilon}|\boldsymbol{X}) + \mathrm{E}(\boldsymbol{b}|\boldsymbol{X}).$$

Поскольку обе оценки, $\boldsymbol{b}$ и $\widehat{\boldsymbol{\beta}}$, несмещенные и поскольку $\mathrm{E}(\boldsymbol{D}\boldsymbol{\varepsilon}|\boldsymbol{X}) = \boldsymbol{D}\,\mathrm{E}(\boldsymbol{\varepsilon}|\boldsymbol{X}) = \mathbf{0}$, отсюда следует, что $\boldsymbol{D}\boldsymbol{X}\boldsymbol{\beta} = \mathbf{0}$. Чтобы это было справедливым для любого вектора $\boldsymbol{\beta}$, необходимо, чтобы $\boldsymbol{D}\boldsymbol{X} = \mathbf{0}$. Поэтому $\widehat{\boldsymbol{\beta}} = \boldsymbol{D}\boldsymbol{\varepsilon} + \boldsymbol{b}$ и

$$\begin{aligned}
\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} &= \boldsymbol{D}\boldsymbol{\varepsilon} + (\boldsymbol{b} - \boldsymbol{\beta}) = \\
&= (\boldsymbol{D} + \boldsymbol{A})\boldsymbol{\varepsilon} \quad \text{(из (1.2.14))}.
\end{aligned}$$

Следовательно,

$$\begin{aligned}
\mathrm{Var}(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}) &= \mathrm{Var}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}|\boldsymbol{X}) = \\
&= \mathrm{Var}[(\boldsymbol{D}+\boldsymbol{A})\boldsymbol{\varepsilon}|\boldsymbol{X}] = \\
&= (\boldsymbol{D}+\boldsymbol{A})\,\mathrm{Var}(\boldsymbol{\varepsilon}|\boldsymbol{X})(\boldsymbol{D}'+\boldsymbol{A}') \\
&\qquad \text{(так как } \boldsymbol{D} \text{ и } \boldsymbol{A} \text{ — функции от } \boldsymbol{X}\text{)} \\
&= \sigma^2 \cdot (\boldsymbol{D}+\boldsymbol{A})(\boldsymbol{D}'+\boldsymbol{A}') \quad \text{(поскольку } \mathrm{Var}(\boldsymbol{\varepsilon}|\boldsymbol{X}) = \sigma^2 \boldsymbol{I}_n\text{)} = \\
&= \sigma^2 \cdot (\boldsymbol{D}\boldsymbol{D}' + \boldsymbol{A}\boldsymbol{D}' + \boldsymbol{D}\boldsymbol{A}' + \boldsymbol{A}\boldsymbol{A}').
\end{aligned}$$

Однако $\boldsymbol{D}\boldsymbol{A}' = \boldsymbol{D}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} = \boldsymbol{0}$, поскольку $\boldsymbol{D}\boldsymbol{X} = \boldsymbol{0}$. Кроме того, $\boldsymbol{A}\boldsymbol{A}' = (\boldsymbol{X}'\boldsymbol{X})^{-1}$, как было показано в (b). Значит,

$$\begin{aligned}
\mathrm{Var}(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}) &= \sigma^2 \cdot [\boldsymbol{D}\boldsymbol{D}' + (\boldsymbol{X}'\boldsymbol{X})^{-1}] \geq \\
&\geq \sigma^2 \cdot (\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&\text{(так как матрица } \boldsymbol{D}\boldsymbol{D}' \text{ положительно полуопределена)} \\
&= \mathrm{Var}(\boldsymbol{b}|\boldsymbol{X}) \quad \text{(из (b)).}
\end{aligned}$$

■

Следует подчеркнуть, что предположение о строгой экзогенности (предположение 1.2) является ключевым для доказательства свойства несмещенности. Без строгой экзогенности нельзя было бы заключить этого. Например, недостаточно предположить,что $\mathrm{E}(\varepsilon_i|\boldsymbol{x}_i) = 0$ для всех i или что $\mathrm{E}(\boldsymbol{x}_i \cdot \varepsilon_i) = \boldsymbol{0}$ для всех i. Как уже отмечалось в параграфе 1.1, большинство моделей временных рядов не удовлетворяют условию строгой экзогенности, даже если выполнены более слабые требования (такие как условие ортогональности $\mathrm{E}(\boldsymbol{x}_i \cdot \varepsilon_i) = \boldsymbol{0}$). Следовательно, для подобных моделей OLS-оценка не обладает свойством несмещенности.

Свойства s^2 в конечных выборках

Мы уже определили OLS-оценку σ^2 в (1.2.13). Она также является несмещенной.

Утверждение 1.2 (несмещенность s^2): *При выполнении предположений 1.2–1.4, $\mathrm{E}(s^2|\boldsymbol{X}) = \sigma^2$ (и, следовательно, $\mathrm{E}(s^2) = \sigma^2$), при $n > K$ (то есть когда s^2 определена).*

Это утверждение можно легко доказать, используя свойства оператора следа[1].

Доказательство Поскольку $s^2 = \boldsymbol{e}'\boldsymbol{e}/(n-K)$, то доказательство заключается в том, чтобы показать, что $\mathrm{E}(\boldsymbol{e}'\boldsymbol{e}|\boldsymbol{X}) = (n-K)\sigma^2$. Как было

[1] **Следом** *trace* квадратной матрицы $\boldsymbol{A}$ называют сумму диагональных элементов $\boldsymbol{A}$: $\mathrm{trace}(\boldsymbol{A}) = \sum_i a_{ii}$.

показано в контрольном вопросе 5 (b) параграфа 1.2, $\boldsymbol{e}'\boldsymbol{e} = \boldsymbol{\varepsilon}'\boldsymbol{M}\boldsymbol{\varepsilon}$, где $\boldsymbol{M}$ — аннулятор. Таким образом, необходимо доказать два свойства: 1) $\mathrm{E}(\boldsymbol{\varepsilon}'\boldsymbol{M}\boldsymbol{\varepsilon}|\boldsymbol{X}) = \sigma^2 \cdot \operatorname{trace}(\boldsymbol{M})$ и 2) $\operatorname{trace}(\boldsymbol{M}) = n - K$.

(1) (Доказательство того, что $\mathrm{E}(\boldsymbol{\varepsilon}'\boldsymbol{M}\boldsymbol{\varepsilon}|\boldsymbol{X}) = \sigma^2 \cdot \operatorname{trace}(\boldsymbol{M})$.) Поскольку $\boldsymbol{\varepsilon}'\boldsymbol{M}\boldsymbol{\varepsilon} = \sum_{i=1}^{n}\sum_{j=1}^{n} m_{ij}\varepsilon_i\varepsilon_j$ (это видно, если расписать квадратичную форму $\boldsymbol{\varepsilon}'\boldsymbol{M}\boldsymbol{\varepsilon}$), получаем:

$$\begin{aligned}
\mathrm{E}(\boldsymbol{\varepsilon}'\boldsymbol{M}\boldsymbol{\varepsilon}|\boldsymbol{X}) &= \sum_{i=1}^{n}\sum_{j=1}^{n} m_{ij}\,\mathrm{E}(\varepsilon_i\,\varepsilon_j|\boldsymbol{X}) \text{ (так как } m_{ij} \text{ — функции от } \boldsymbol{X}, \\
&\qquad \mathrm{E}(m_{ij}\,\varepsilon_i\,\varepsilon_j|\boldsymbol{X}) = m_{ij}\,\mathrm{E}(\varepsilon_i\,\varepsilon_j|\boldsymbol{X})) \\
&= \sum_{i=1}^{n} m_{ii}\,\sigma^2 \\
&\quad \text{(поскольку } \mathrm{E}(\varepsilon_i\,\varepsilon_j|\boldsymbol{X}) = 0 \text{ для } i \neq j \text{ по предположению 1.4)} \\
&= \sigma^2 \sum_{i=1}^{n} m_{ii} = \\
&= \sigma^2 \cdot \operatorname{trace}(\boldsymbol{M}).
\end{aligned}$$

(2) (Доказательство того, что $\operatorname{trace}(\boldsymbol{M}) = n - K$.)

$$\begin{aligned}
\operatorname{trace}(\boldsymbol{M}) &= \operatorname{trace}(\boldsymbol{I}_n - \boldsymbol{P}) \\
&\quad \text{(так как } \boldsymbol{M} \equiv \boldsymbol{I}_n - \boldsymbol{P}\text{; см. (1.2.8))} \\
&= \operatorname{trace}(\boldsymbol{I}_n) - \operatorname{trace}(\boldsymbol{P}) \\
&\quad \text{(факт: оператор следа является линейным)} \\
&= n - \operatorname{trace}(\boldsymbol{P}). \\
\operatorname{trace}(\boldsymbol{P}) &= \operatorname{trace}[\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'] \\
&\quad \text{(так как } \boldsymbol{P} \equiv \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\text{; см. (1.2.7))} \\
&= \operatorname{trace}[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}] \text{ (факт: } \operatorname{trace}(\boldsymbol{AB}) = \operatorname{trace}(\boldsymbol{BA})) \\
&= \operatorname{trace}(\boldsymbol{I}_K) = K.
\end{aligned}$$

Итак, $\operatorname{trace}(\boldsymbol{M}) = n - K$. ■

Оценка $\mathrm{Var}(b|X)$

Если s^2 — оценка σ^2, то естественной оценкой $\mathrm{Var}(\boldsymbol{b}|\boldsymbol{X}) = \sigma^2 \cdot (\boldsymbol{X}'\boldsymbol{X})^{-1}$ является:

$$\widehat{\mathrm{Var}(\boldsymbol{b}|\boldsymbol{X})} \equiv s^2 \cdot (\boldsymbol{X}'\boldsymbol{X})^{-1}. \tag{1.3.4}$$

Это одна из статистик, которые входят в компьютерную выдачу после OLS-оценивания в любом компьютерном пакете.

Контрольные вопросы

1. (Роль предположения об отсутствии мультиколлинеарности.) Где мы использовали в утверждениях 1.1 и 1.2 предположение 1.3 о том, что $\operatorname{rank}(\boldsymbol{X}) = K$? **Указание:** *Условие отсутствия мультиколлинеарности необходимо нам, чтобы обеспечить обратимость матрицы $\boldsymbol{X}'\boldsymbol{X}$.*

2. (Пример линейной оценки.) Для примера 1.1 с функцией потребления предложите линейную и несмещенную оценку β_2, отличную от OLS-оценки. **Указание:** *Как насчет $\widehat{\beta}_2 = (CON_2 - CON_1)/(YD_2 - YD_1)$? Является ли она линейной по $(CON_1, \ldots, CON_n)$? Является ли она несмещенной: $\mathrm{E}(\widehat{\beta}_2 | YD_1, \ldots, YD_n) = \beta_2$?*

3. (Чего не утверждает теорема Гаусса — Маркова.) Существует ли при предположениях 1.2–1.4 линейная, но необязательно несмещенная оценка $\boldsymbol{\beta}$, дисперсия которой меньше, чем у OLS-оценки? Если да, то насколько мала может быть дисперсия? **Указание:** *Если оценка $\boldsymbol{\beta}$ — константа, то она тривиальным образом линейна по $\boldsymbol{y}$.*

4. (Теорема Гаусса — Маркова и безусловная дисперсия.)

   (a) Докажите: $\operatorname{Var}(\widehat{\boldsymbol{\beta}}) = \mathrm{E}\left[\operatorname{Var}(\widehat{\boldsymbol{\beta}}|\boldsymbol{X})\right] + \operatorname{Var}\left[\mathrm{E}(\widehat{\boldsymbol{\beta}}|\boldsymbol{X})\right]$.
   Указание: *По определению,*
   $$\operatorname{Var}(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}) \equiv \mathrm{E}\left[\left(\widehat{\boldsymbol{\beta}} - \mathrm{E}(\widehat{\boldsymbol{\beta}}|\boldsymbol{X})\right)\left(\widehat{\boldsymbol{\beta}} - \mathrm{E}(\widehat{\boldsymbol{\beta}}|\boldsymbol{X})\right)' |\boldsymbol{X}\right]$$
   и
   $$\operatorname{Var}\left[\mathrm{E}(\widehat{\boldsymbol{\beta}}|\boldsymbol{X})\right] \equiv \mathrm{E}\left\{\left[\mathrm{E}(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}) - \mathrm{E}(\widehat{\boldsymbol{\beta}})\right]\left[\mathrm{E}(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}) - \mathrm{E}(\widehat{\boldsymbol{\beta}})\right]'\right\}.$$
   Используйте стратегию прибавления-вычитания: возьмите $\widehat{\boldsymbol{\beta}} - \mathrm{E}(\widehat{\boldsymbol{\beta}}|\boldsymbol{X})$ и прибавьте и вычтите $\mathrm{E}(\widehat{\boldsymbol{\beta}})$.

   (b) Докажите (1.3.3). **Указание:** *Если $\operatorname{Var}(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}) \geq \operatorname{Var}(\boldsymbol{b}|\boldsymbol{X})$, то*
   $$\mathrm{E}\left[\operatorname{Var}(\widehat{\boldsymbol{\beta}}|\boldsymbol{X})\right] \geq \mathrm{E}\left[\operatorname{Var}(\boldsymbol{b}|\boldsymbol{X})\right].$$

5. Предложите несмещенную оценку σ^2, считая, что $\boldsymbol{\varepsilon}$ вам известен. **Указание:** *Как насчет $\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}/n$? Является ли она несмещенной?*

6. Докажите часть (d) утверждения 1.1. **Указание:** *По определению,*
   $$\operatorname{Cov}(\boldsymbol{b}, \boldsymbol{e}|\boldsymbol{X}) \equiv \mathrm{E}\left\{[\boldsymbol{b} - \mathrm{E}(\boldsymbol{b}|\boldsymbol{X})][\boldsymbol{e} - \mathrm{E}(\boldsymbol{e}|\boldsymbol{X})]'|\boldsymbol{X}\right\}.$$
   Так как $\mathrm{E}(\boldsymbol{b}|\boldsymbol{X}) = \boldsymbol{\beta}$, получаем $\boldsymbol{b} - \mathrm{E}(\boldsymbol{b}|\boldsymbol{X}) = \boldsymbol{A}\boldsymbol{\varepsilon}$, где $\boldsymbol{A} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. Используйте $\boldsymbol{M}\boldsymbol{\varepsilon} = \boldsymbol{e}$ (см. контрольный вопрос 5 параграфа 1.2), чтобы показать, что $\boldsymbol{e} - \mathrm{E}(\boldsymbol{e}|\boldsymbol{X}) = \boldsymbol{M}\boldsymbol{\varepsilon}$. $\mathrm{E}(\boldsymbol{A}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{M}|\boldsymbol{X}) = \boldsymbol{A}\,\mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\boldsymbol{X})\boldsymbol{M}$, поскольку $\boldsymbol{A}$ и $\boldsymbol{M}$ являются функциями от $\boldsymbol{X}$. Наконец, используйте $\boldsymbol{M}\boldsymbol{X} = \boldsymbol{0}$ (см. (1.2.11)).

7. Докажите (1.2.21). **Указание:** *Поскольку матрица $\boldsymbol{P}$ положительно полуопределена, ее диагональные элементы неотрицательны. Заметьте, что $\sum_{i=1}^{n} p_i = \operatorname{trace}(\boldsymbol{P})$.*

1.4. Тестирование гипотез в предположении нормальности

Очень часто экономическая теория, мотивировавшая спецификацию регрессионного уравнения, также накладывает ограничения на коэффициенты модели. Предположим, что экономическая теория накладывает ограничение, что коэффициент β_2 равен 1. Хотя предположение 1.1 гарантирует, что в среднем b_2 (OLS-оценка β_2) равна 1, если ограничение верно, b_2 может не быть в точности равной единице на конкретной выборке, которой вы располагаете. Очевидно, мы не можем заключить, что ограничение не выполнено, только на том основании, что оценка b_2 отлична от 1. Чтобы решить, является ли ошибка оценки $b_2 - 1$ «слишком большой», чтобы принять истинность ограничения, необходимо на основании ошибки оценки построить тестовую статистику, распределение которой было бы известным при условии истинности нулевой гипотезы. Может показаться, что для этого необходимо определить совместное распределение $(\boldsymbol{X}, \varepsilon)$, поскольку (как следует из (1.2.14)) ошибка оценки — это функция от $(\boldsymbol{X}, \varepsilon)$. Однако в этом параграфе будет изложен замечательный факт состоящий в том, что при нормальном условном относительно $\boldsymbol{X}$ распределении ε не нужно специфицировать распределение $\boldsymbol{X}$, чтобы найти распределение тестовой статистики.

На языке проверки гипотез тестируемое ограничение (такое, например, как «$\beta_2 = 1$») называется **нулевой гипотезой** (*null hypothesis, null*). Нулевая гипотеза является ограничением на **поддерживаемую гипотезу** (*maintained hypothesis*) — набор предположений, которые совместно с нулевой гипотезой позволяют получить тестовую статистику с известным распределением. Для рассматриваемого случая тестирования коэффициентов регрессии, для формирования такой поддерживаемой гипотезы к классической линейной регрессионной модели (предположения 1.2–1.4) необходимо добавить только предположение о нормальности условного распределения ε. (Как только что отмечалось, специфицировать совместное распределение $(\boldsymbol{X}, \varepsilon)$ нет необходимости.) Иногда поддерживаемую гипотезу достаточно вольно называют «моделью». Говорят, что модель **правильно специфицирована**, если поддерживаемая гипотеза верна. Хотя слишком большое значение тестовой статистики интерпретируется как невыполнение нулевой гипотезы, такие выводы возможно сделать лишь в случае правильной спецификации модели. Распределение тестовой статистики может отличаться от предполагаемого, если нулевая гипотеза верна, но модель неправильно специфицирована.

Нормально распределенные ошибки

Во многих приложениях ошибка включает большое количество разнообразных факторов, не охваченных регрессорами. Центральная предельная теорема наводит на мысль о том, что такая ошибка имеет нормальное распределение. В других приложениях ошибка возникает вследствие ошибок измерения зависимой переменной. Известно, что ошибки измерений очень часто распределены нормально (на самом деле нормальное распределение изначально выводилось для ошибок измерения). Поэтому заслуживает внимания следующее предположение о нормальности.

Предположение 1.5 (нормальность ошибки): *Условное распределение $\boldsymbol{\varepsilon}$ при условии $\boldsymbol{X}$ является многомерным нормальным распределением.*

Вспомним из теории вероятностей, что нормальное распределение обладает рядом удобных свойств:

- Это распределение зависит только от математического ожидания и дисперсии. Таким образом, при известных математическом ожидании и дисперсии можно записать функцию плотности. Если распределение, условное относительно $\boldsymbol{X}$, нормально, то его математическое ожидание и дисперсия могут зависеть от $\boldsymbol{X}$. Следовательно, если распределение, условное относительно $\boldsymbol{X}$, нормально и ни математическое ожидание, ни дисперсия не зависят от $\boldsymbol{X}$, то маргинальное (то есть безусловное) распределение является тем же самым нормальным распределением.

- В общем случае из независимости двух случайных величин вытекает их некоррелированность, но не наоборот. Однако если две случайные величины совместно нормально распределены, то из их некоррелированности следует их независимость. Иными словами, для совместно нормально распределенных случайных величин некоррелированность и независимость эквивалентны. Это верно и для условного распределения: если две случайные величины совместно нормальны и некоррелированы при условии $\boldsymbol{X}$, то они независимы при условии $\boldsymbol{X}$.

- Линейная функция от случайных величин с совместным нормальным распределением также распределена нормально. То же верно и для условных распределений. Если условное распределение $\boldsymbol{\varepsilon}$ при условии $\boldsymbol{X}$ нормально, то $\boldsymbol{A\varepsilon}$ (где элементы матрицы $\boldsymbol{A}$ — функции от $\boldsymbol{X}$) имеет условное нормальное распределение при условии $\boldsymbol{X}$.

Именно благодаря этим особенностям нормального распределения предположение 1.5 обеспечивает следующие свойства, которые используются для вывода тестовой статистики:

- Математическое ожидание и дисперсия распределения $\boldsymbol{\varepsilon}$, условного относительно $\boldsymbol{X}$, уже были указаны в предположениях 1.2 и 1.4 Добавление к предположениям 1.2 и 1.4 предположения 1.5

приводит к тому, что распределение ε, условное относительно $\boldsymbol{X}$, является нормальным распределением $N(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$:

$$\varepsilon | \boldsymbol{X} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I}_n). \tag{1.4.1}$$

Таким образом, распределение ε, условное относительно $\boldsymbol{X}$, не зависит от $\boldsymbol{X}$. Из этого следует, что ε и $\boldsymbol{X}$ *независимы*. Поэтому, в частности, маргинальное, или безусловное, распределение ε нормально $N(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$.

- Из (1.2.14) нам известно, что ошибка оценки $\boldsymbol{b} - \boldsymbol{\beta}$ линейна по ε при условии $\boldsymbol{X}$. Так как ε распределен нормально при условии $\boldsymbol{X}$, то это же справедливо для ошибки оценки. Его математическое ожидание и дисперсия заданы частями (a) и (b) утверждения 1.1. Таким образом, при предположениях 1.1–1.5

$$(\boldsymbol{b} - \boldsymbol{\beta}) | \boldsymbol{X} \sim N(\mathbf{0}, \sigma^2 \cdot (\boldsymbol{X}'\boldsymbol{X})^{-1}). \tag{1.4.2}$$

Тестирование гипотез об отдельных коэффициентах регрессии

Сначала рассмотрим гипотезу о k-м коэффициенте:

$$H_0\colon \beta_k = \overline{\beta}_k.$$

Здесь $\overline{\beta}_k$ — некоторое известное значение, определяемое нулевой гипотезой. Мы хотим протестировать эту гипотезу против альтернативы $\mathrm{H}_1\colon \beta_k \neq \overline{\beta}_k$ на уровне значимости α. Берем k-ю компоненту (1.4.2), накладываем ограничение, заданное нулевой гипотезой, и получаем:

$$(b_k - \overline{\beta}_k) \mid \boldsymbol{X} \sim N\Big(0, \sigma^2 \cdot \big((\boldsymbol{X}'\boldsymbol{X})^{-1}\big)_{kk}\Big),$$

где $\big((\boldsymbol{X}'\boldsymbol{X})^{-1}\big)_{kk}$ — элемент (k, k) матрицы $(\boldsymbol{X}'\boldsymbol{X})^{-1}$. Если мы определим отношение z_k как результат от деления разности $b_k - \overline{\beta}_k$ на ее стандартное отклонение,

$$z_k \equiv \frac{b_k - \overline{\beta}_k}{\sqrt{\sigma^2 \cdot \big((\boldsymbol{X}'\boldsymbol{X})^{-1}\big)_{kk}}}, \tag{1.4.3}$$

то z_k будет иметь стандартное нормальное распределение $N(0, 1)$.

Временно предположим, что σ^2 известна. Тогда статистика z_k обладает рядом желательных свойств как тестовая статистика. Во-первых, ее значение может быть вычислено по выборке. Во-вторых, ее распределение, условное относительно $\boldsymbol{X}$, не зависит от $\boldsymbol{X}$ (не путать с фактом, что *значение* z_k зависит от $\boldsymbol{X}$). То есть z_k и $\boldsymbol{X}$ независимо распределены, и вне зависимости от значения $\boldsymbol{X}$ распределение z_k совпадает с безусловным распределением. Это удобно, так как разные выборки отличаются между собой не только значениями $\boldsymbol{y}$, но и значениями $\boldsymbol{X}$. В-третьих, распределение указанной статистики известно. В частности, оно не зависит от неизвестных параметров (таких как $\boldsymbol{\beta}$). (Если распределение статистики зависит от неизвестных параметров, эти параметры называются **мешающими** (*nuisance parameters*).) Используя эту статистику, мы можем определить, является ли ошибка оценки слишком большой: она является слишком большой, если тестовая статистика принимает значение, которое нетипично для реализации из указанного распределения.

Если мы не знаем значение σ^2, то естественно заменить мешающий параметр σ^2 на его OLS-оценку s^2. Получаемая подстановкой s^2 вместо σ^2-статистика называется **t-отношением**, или **t-значением** (*t-ratio, t-value*). Знаменатель этой статистики называется **стандартной ошибкой** (*standard error*) OLS-оценки коэффициента β_k и иногда обозначается $SE(b_k)$:

$$SE(b_k) \equiv \sqrt{s^2 \cdot \left((\boldsymbol{X}'\boldsymbol{X})^{-1}\right)_{kk}} = \sqrt{(k,k)\text{-й элемент } \widehat{\operatorname{Var}(\boldsymbol{b}|\boldsymbol{X})} \text{ в } (1.3.4)}. \tag{1.4.4}$$

Поскольку s^2 — случайная величина (так как она является функцией от выборки), эта подстановка изменяет распределение статистики, но, к счастью, новое распределение также известно и не зависит ни от мешающих параметров, ни от $\boldsymbol{X}$.

Утверждение 1.3 (распределение t-статистики):
Пусть предположения 1.1–1.5 выполнены. Тогда при нулевой гипотезе $H_0\colon \beta_k = \overline{\beta}_k$, t-статистика, определенная как

$$t_k \equiv \frac{b_k - \overline{\beta}_k}{SE(b_k)} \equiv \frac{b_k - \overline{\beta}_k}{\sqrt{s^2 \cdot \left((\boldsymbol{X}'\boldsymbol{X})^{-1}\right)_{kk}}}. \tag{1.4.5}$$

распределена как $t(n-K)$ (t-распределение Стьюдента с $n-K$ степенями свободы).

Доказательство

Чтобы отразить замену s^2 на σ^2, мы можем переписать t-статистику в виде:

$$t_k = \frac{b_k - \overline{\beta}_k}{\sqrt{\sigma^2 \cdot \left((\boldsymbol{X}'\boldsymbol{X})^{-1}\right)_{kk}}} \cdot \sqrt{\frac{\sigma^2}{s^2}} = \frac{z_k}{\sqrt{s^2/\sigma^2}} =$$

$$= \frac{z_k}{\sqrt{\frac{\boldsymbol{e}'\boldsymbol{e}/(n-K)}{\sigma^2}}} = \frac{z_k}{\sqrt{\frac{q}{n-K}}}.$$

где $q \equiv \boldsymbol{e}'\boldsymbol{e}/\sigma^2$. Мы уже показали, что z_k имеет распределение $N(0, 1)$. Теперь мы покажем, что

(1) $q|\boldsymbol{X} \sim \chi^2(n-K)$,

(2) две случайные величины z_k и q независимы условно относительно $\boldsymbol{X}$.

Тогда, по определению t-распределения, отношение z_k к $\sqrt{q/(n-K)}$ распределено как t-распределение с $n-K$ степенями свободы[1], и на этом доказательство будет завершено.

(1) Поскольку $\boldsymbol{e}'\boldsymbol{e} = \boldsymbol{\varepsilon}'\boldsymbol{M}\boldsymbol{\varepsilon}$ из (1.2.12), мы имеем:

$$q = \frac{\boldsymbol{e}'\boldsymbol{e}}{\sigma^2} = \frac{\boldsymbol{\varepsilon}'}{\sigma}\boldsymbol{M}\frac{\boldsymbol{\varepsilon}}{\sigma}.$$

Матрица $\boldsymbol{M}$, стоящая в середине, идемпотентна (так как она является аннулятором). Кроме того, $\boldsymbol{\varepsilon}/\sigma|\boldsymbol{X} \sim N(\boldsymbol{0}, \boldsymbol{I}_n)$ из (1.4.1). Поэтому эта квадратичная форма распределена как χ^2 с количеством степеней свободы, равным $\operatorname{rank}(\boldsymbol{M})$[2]. Но $\operatorname{rank}(\boldsymbol{M}) = \operatorname{trace}(\boldsymbol{M})$, так как матрица $\boldsymbol{M}$ идемпотентна[3]. В процессе доказательства утверждения 1.2 мы уже показали, что $\operatorname{trace}(\boldsymbol{M}) = n - K$. То есть $q|\boldsymbol{X} \sim \chi^2(n-K)$.

(2) $\boldsymbol{b}$ и $\boldsymbol{e}$ — линейные функции от $\boldsymbol{\varepsilon}$ (из (1.2.14) и того факта, что $\boldsymbol{e} = \boldsymbol{M}\boldsymbol{\varepsilon}$), следовательно, они совместно нормально распределены при условии $\boldsymbol{X}$. Кроме того, они некоррелированы при условии $\boldsymbol{X}$ (см. часть (d) утверждения 1.1). Значит, $\boldsymbol{b}$ и $\boldsymbol{e}$ независимо распределены при условии $\boldsymbol{X}$. Однако z_k — функция от $\boldsymbol{b}$, а q — функция от $\boldsymbol{e}$. Таким образом, z_k и q распределены независимо при условии $\boldsymbol{X}$[4]. ■

Правило принятия решений на основании t-теста

Тестирование нулевой гипотезы с помощью t-статистики называется t-тестом и проводится следующим образом:

Шаг 1: При заданном предполагаемом значении, $\overline{\beta}_k$, истинного значения β_k постройте t-отношение по формуле (1.4.5). Слишком большое отклонение t_k от 0 является признаком нарушения нулевой гипотезы. Следующий шаг уточняет, что именно означает «слишком большое» отклонение.

---

[1] Факт: Если $x \sim N(0, 1)$, $y \sim \chi^2(m)$ и если x и y независимы, то отношение $x/\sqrt{y/m}$ имеет t-распределение с m степенями свободы.

[2] Факт: Если $\boldsymbol{x} \sim N(\boldsymbol{0}, \boldsymbol{I}_n)$, и матрица $\boldsymbol{A}$ симметрична и идемпотентна, то $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ имеет χ^2 распределение с количеством степеней свободы, равным $\operatorname{rank}(\boldsymbol{A})$.

[3] Факт: если $\boldsymbol{A}$ – идемпотентная и симметричная матрица, то $\operatorname{rank}(\boldsymbol{A}) = \operatorname{trace}(\boldsymbol{A})$.

[4] Факт: Если $\boldsymbol{x}$ и $\boldsymbol{y}$ независимо распределены, то же можно сказать и о $f(\boldsymbol{x})$ и $g(\boldsymbol{y})$.

Шаг 2: Откройте таблицу критических значений t-распределения (большинство учебников по эконометрике и статистике содержат эти таблицы) и посмотрите на вход этой таблицы, соответствующий $n-K$ степеням свободы. Найдите **критическое значение** (*critical value*), $t_{\alpha/2}(n-K)$, такое, что площадь под кривой t-распределения справа от $t_{\alpha/2}(n-K)$ была бы равна $\alpha/2$. Это проиллюстрировано на рис. 1.3. (Если, например, $n-K=30$, и $\alpha = 5\%$, то $t_{\alpha/2}(n-K) = 2.042$.) Так как t-распределение симметрично относительно нуля,

$$\text{Prob}\big(-t_{\alpha/2}(n-K) < t < t_{\alpha/2}(n-K)\big) = 1-\alpha.$$

Шаг 3: H_0 не отвергается, если $-t_{\alpha/2}(n-K) < t_k < t_{\alpha/2}(n-K)$ (то есть, если $|t_k| < t_{\alpha/2}(n-K)$), где t_k — t-статистика из *шага 1*. В противном случае H_0 отвергается. Поскольку $t_k \sim t(n-K)$ при H_0, то вероятность отвержения H_0, когда H_0 верна, равна α. То есть размер (уровень значимости) теста равен α.

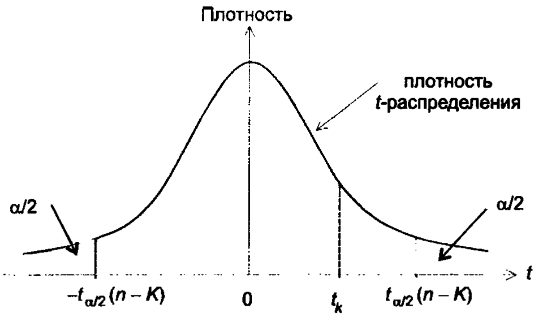

Рис. 1.3. ***t*-распределение**

Удобным свойством t-теста является то, что критические значения не зависят от $\boldsymbol{X}$: нет необходимости вычислять критические значения для каждой выборки.

Доверительный интервал

Шаг 3 может быть переформулирован в терминах b_k и $SE(b_k)$. Поскольку t_k определена соотношением (1.4.5), H_0 не отвергается, если

$$-t_{\alpha/2}(n-K) < \frac{b_k - \overline{\beta}_k}{SE(b_k)} < t_{\alpha/2}(n-K)$$

или

$$b_k - SE(b_k)\cdot t_{\alpha/2}(n-K) < \overline{\beta}_k < b_k + SE(b_k)\cdot t_{\alpha/2}(n-K).$$

Поэтому мы не отвергаем H_0 тогда и только тогда, когда гипотетическое значение $\overline{\beta}_k$ попадает в интервал:

$$[b_k - SE(b_k) \cdot t_{\alpha/2}(n-K), b_k + SE(b_k) \cdot t_{\alpha/2}(n-K)]. \qquad (1.4.6)$$

Этот интервал называется **доверительным интервалом с уровнем доверия** $(1-\alpha)$, или $(1-\alpha)$**-доверительным интервалом** (*level* $1-\alpha$ *confidence interval*). Доверительный интервал тем у́же, чем меньше стандартная ошибка. Таким образом, чем меньше стандартная ошибка, тем более точна оценка.

p-значение

Правило принятия решений t-теста можно переформулировать в терминах **p-значения** (*p-value*).

Шаг 1: Тот же.

Шаг 2: Вместо нахождения критического значения $t_{\alpha/2}(n-K)$, вычислите

$$p = \text{Prob}(t > |t_k|) \times 2.$$

Так как t-распределение симметрично относительно 0,

$$\text{Prob}(t > |t_k|) = \text{Prob}(t < -|t_k|),$$

так что

$$\text{Prob}(-|t_k| < t < |t_k|) = 1 - p. \qquad (1.4.7)$$

Шаг 3: H_0 не отвергается, если $p > \alpha$. В противном случае нулевая гипотеза отвергается.

Чтобы увидеть эквивалентность двух этих правил принятия решений (одного, основанного на критических значениях $t_{\alpha/2}(n-K)$, другого — на p-значении), обратимся к рис. 1.3. Если $\text{Prob}(t > |t_k|)$ больше, чем $\alpha/2$ (как на рисунке), то есть p-значение больше, чем α, то тогда $|t_k|$ должен быть слева от $t_{\alpha/2}(n-K)$. Это значит (из *шага 3*), что нулевая гипотеза не отвергается. Таким образом, когда p мало, t-отношение нетипично велико для случайной величины из t-распределения. Чем меньше p, тем более веским является решение об отвержении.

Примеры применения t-теста можно найти в параграфе 1.7.

Линейные гипотезы

Нулевая гипотеза, которую мы хотим протестировать, может отличаться от только что рассмотренного ограничения на индивидуальный коэффициент. Часто речь идет о линейных комбинациях коэффициентов, и гипотеза записывается в виде системы линейных уравнений:

$$H_0\colon \boldsymbol{R\beta} = \boldsymbol{r}, \tag{1.4.8}$$

где значения $\boldsymbol{R}$ и $\boldsymbol{r}$ известны и определяются проверяемой гипотезой. Обозначим число уравнений (размерность $\boldsymbol{r}$) через $\#\boldsymbol{r}$. Тогда $\boldsymbol{R}$ имеет размер $\#\boldsymbol{r} \times K$. Эти $\#\boldsymbol{r}$ уравнений — ограничения на коэффициенты в поддерживаемой гипотезе. Это называется линейной гипотезой, так как каждое уравнение линейно. Чтобы удостовериться в том, что в системе нет излишних уравнений и что уравнения не противоречат друг другу, потребуем, чтобы $\operatorname{rank}(\boldsymbol{R}) = \#\boldsymbol{r}$ (то есть чтобы матрица $\boldsymbol{R}$ имела полный *строковый* ранг, равный количеству строк). Однако не стоит придавать слишком большое значение этому ранговому условию: в конкретных случаях заметить нарушение этого рангового условия, если таковое имеет место, очень легко.

Пример 1.5 (продолжение примера 1.2): Вернемся к уравнению заработной платы из примера 1.2, где $K = 4$. Мы могли бы захотеть протестировать гипотезу о том, что образование и продолжительность работы на текущем рабочем месте имеют одинаковое влияние на уровень заработной платы, а общий стаж на заработную плату не влияет. Эта гипотеза состоит из двух уравнений (то есть $\#\boldsymbol{r} = 2$):

$$\beta_2 = \beta_3 \quad \text{и} \quad \beta_4 = 0.$$

То же самое можно переписать и в виде $\boldsymbol{R\beta} = \boldsymbol{r}$, если определить $\boldsymbol{R}$ и $\boldsymbol{r}$ следующим образом:

$$\boldsymbol{R} = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{r} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Так как две строки $\boldsymbol{R}$ линейно независимы, ранговое условие выполнено.

Но предположим, что мы дополним гипотезу еще одним ограничением:

$$\beta_2 - \beta_3 = \beta_4.$$

Это ограничение избыточно, так как оно выполнено всегда, когда выполнены первые два ограничения. Учитывая это ограничение, получаем систему из трех уравнений ($\#\boldsymbol{r} = 3$), и

$$\boldsymbol{R} = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & -1 & -1 \end{bmatrix}, \quad \boldsymbol{r} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Поскольку третья строка $\boldsymbol{R}$ представляет собой разность двух первых, $\boldsymbol{R}$ теперь не имеет полного строкового ранга. Следствием добавления избыточного ограничения является то, что матрица $\boldsymbol{R}$ больше не удовлетворяет условию полного строкового ранга.

В качестве примера несовместных уравнений добавим к первым двум ограничениям еще одно: $\beta_4 = 0.5$. Очевидно, коэффициент β_4 не может быть равен 0 и 0,5 одновременно. Гипотеза несостоятельна, так как не существует такого вектора $\boldsymbol{\beta}$, который бы одновременно удовлетворял всем условиям. Если мы все же добавим это условие, то матрица $\boldsymbol{R}$ и $\boldsymbol{r}$ будут выглядеть так:

$$\boldsymbol{R} = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \boldsymbol{r} = \begin{bmatrix} 0 \\ 0 \\ 0.5 \end{bmatrix}.$$

И опять условие полного строчного ранга не выполнено, так как ранг $\boldsymbol{R}$ равен 2, в то время как $\#\boldsymbol{r} = 3$.

F-тест

Для проверки линейных гипотез необходима тестовая статистика, которая имела бы известное распределение при нулевой гипотезе.

Утверждение 1.4 (распределение F-отношения): *Пусть предположения 1.1–1.5 выполнены. При нулевой гипотезе H_0: $\boldsymbol{R\beta} = \boldsymbol{r}$, где $\boldsymbol{R}$ — матрица размерности $\#\boldsymbol{r} \times K$, и $\operatorname{rank}(\boldsymbol{R}) = \#\boldsymbol{r}$,* **«$F$-отношение»** *($F$-ratio), определяемое как*

$$\begin{aligned} F &\equiv \frac{(\boldsymbol{Rb} - \boldsymbol{r})' \left[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\right]^{-1} (\boldsymbol{Rb} - \boldsymbol{r})/\#\boldsymbol{r}}{s^2} = \\ &= (\boldsymbol{Rb} - \boldsymbol{r})'[\boldsymbol{R}\widehat{\operatorname{Var}(\boldsymbol{b}|\boldsymbol{X})}\boldsymbol{R}']^{-1}(\boldsymbol{Rb} - \boldsymbol{r})/\#\boldsymbol{r} \quad (\text{из } (1.3.4)) \end{aligned} \tag{1.4.9}$$

распределено как $F(\#\boldsymbol{r}, n-K)$ (F-распределение Фишера с $\#\boldsymbol{r}$ и $n-K$ степенями свободы).

Как и в утверждении 1.3, достаточно показать, что распределение, условное относительно $\boldsymbol{X}$, — это $F(\#\boldsymbol{r}, n-K)$. Поскольку F-распределение не зависит от $\boldsymbol{X}$, оно совпадает с безусловным распределением.

Доказательство: Так как $s^2 = \boldsymbol{e}'\boldsymbol{e}/(n-K)$, мы можем записать:

$$F = \frac{w/\#\boldsymbol{r}}{q/(n-K)},$$

где

$$w \equiv (\boldsymbol{Rb} - \boldsymbol{r})'[\sigma^2 \cdot \boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{Rb} - \boldsymbol{r}) \quad \text{и} \quad q \equiv \frac{\boldsymbol{e}'\boldsymbol{e}}{\sigma^2}.$$

Теперь нам необходимо показать:

(1) $w|\boldsymbol{X} \sim \chi^2(\#\boldsymbol{r})$,

(2) $q|\boldsymbol{X} \sim \chi^2(n-K)$ (это часть 1) доказательства утверждения 1.3),

(3) w и q независимо распределены при условии $\boldsymbol{X}$.

Тогда, по определению F-распределения, F-отношение $\sim F(\#\boldsymbol{r}, n-K)$.

1) Пусть $\boldsymbol{v} \equiv \boldsymbol{Rb} - \boldsymbol{r}$. При H_0, $\boldsymbol{Rb} - \boldsymbol{r} = \boldsymbol{R}(\boldsymbol{b} - \boldsymbol{\beta})$. Значит, из (1.4.2), условное относительно $\boldsymbol{X}$ распределение $\boldsymbol{v}$ нормально с математическим ожиданием $\boldsymbol{0}$ и дисперсией

$$\begin{aligned}\operatorname{Var}(\boldsymbol{v}|\boldsymbol{X}) &= \operatorname{Var}(\boldsymbol{R}(\boldsymbol{b}-\boldsymbol{\beta})|\boldsymbol{X}) = \boldsymbol{R}\operatorname{Var}(\boldsymbol{b}-\boldsymbol{\beta}|\boldsymbol{X})\boldsymbol{R}' = \\ &= \sigma^2 \cdot \boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}',\end{aligned}$$

которая является не чем иным, как обратной матрицей к средней матрице в квадратичной форме для w. Значит, w можно записать как $\boldsymbol{v}'\operatorname{Var}(\boldsymbol{v}|\boldsymbol{X})^{-1}\boldsymbol{v}$. Поскольку $\boldsymbol{R}$ — матрица полного строкового ранга и $\boldsymbol{X}'\boldsymbol{X}$ невырожденная, $\sigma^2\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'$ невырождена. (Почему? Показать это потребуется в одном из контрольных вопросов.) Поэтому, по определению χ^2-распределения, $w|\boldsymbol{X} \sim \chi^2(\#\boldsymbol{r})$[1].

3) w является функцией от $\boldsymbol{b}$, а q — функция от $\boldsymbol{e}$. Но $\boldsymbol{b}$ и $\boldsymbol{e}$ независимо распределены при условии $\boldsymbol{X}$, как показано в части (2) утверждения 1.3. То есть w и q независимо распределены при условии $\boldsymbol{X}$. ■

Если нулевая гипотеза $\boldsymbol{R\beta} = \boldsymbol{r}$ верна, то мы ожидаем, что $\boldsymbol{R\beta} - \boldsymbol{r}$ будет небольшим. То есть большие значения F должны восприниматься как свидетельство неподтверждения нулевой гипотезы. Это означает, что мы смотрим только на правый хвост распределения F-статистики. Правило принятия решений на основании F-теста на уровне значимости α сформулировано ниже.

Шаг 1: Вычислите F-отношение по формуле (1.4.9).

Шаг 2: Обратитесь к таблице критических значений для F-распределения и посмотрите на вход для $\#\boldsymbol{r}$ (количество степеней свободы числителя) и $n-K$ (количество степеней свободы знаменателя). Найдите критическое значение $F_\alpha(\#\boldsymbol{r}, n-K)$, для которого вероятность попасть в правый хвост F-распределения равна α. Это проиллюстрировано на рис. 1.4. Например, когда $\#\boldsymbol{r} = 3$, $n-K = 30$ и $\alpha = 5\%$, критическое значение $F_{.05}(3.30) = 2.92$.

[1]Факт: Пусть $\boldsymbol{x}$ — m-мерный случайный вектор. Если $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, где матрица $\boldsymbol{\Sigma}$ невырождена, то $(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) \sim \chi^2(m)$.

Шаг 3: Нулевая гипотеза не отвергается, если F-отношение из *шага 1* меньше, чем $F_\alpha(\#\boldsymbol{r}, n-K)$. В противном случае нулевая гипотеза отвергается.

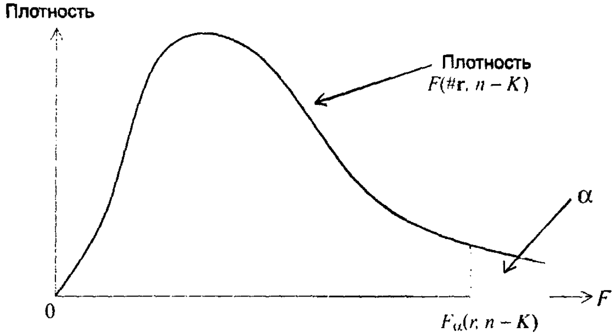

Рис. 1.4. F-распределение

Это правило принятия решений может также быть сформулировано в терминах p-значения:

Шаг 1: Тот же, что и был.

Шаг 2: Вычислите

$p =$ площадь правого хвоста F-распределения справа от F-отношения.

Шаг 3: Нулевая гипотеза не отвергается, если $p > \alpha$. В противном случае H_0 отвергается.

Таким образом, *малое* p-значение является сигналом невыполнения нулевой гипотезы.

Более удобное выражение для F

Процесс вывода F-отношения, изложенный выше, базируется на **принципе Вальда** (*Wald principle*), поскольку в его основе лежит оценка без ограничений (то есть не учитывающая ограничения, накладываемые нулевой гипотезой). Вычисление F-отношения по формуле (1.4.9) требует обращения матрицы и умножения матриц. Однако существует альтернативная формула, требующая только подсчета двух сумм квадратов остатков: минимизированной суммы квадратов остатков, полученной по формуле (1.2.1) (которую обозначим как SSR_U), и минимизированной суммы квадратов остатков, рассчитанной с учетом ограничений (ее обозначим как SSR_R), получаемой как

$$\min_{\widetilde{\boldsymbol{\beta}}} SSR(\widetilde{\boldsymbol{\beta}}) \quad \text{при ограничении} \quad \boldsymbol{R}\widetilde{\boldsymbol{\beta}} = \boldsymbol{r}. \qquad (1.4.10)$$

Отыскание такой оценки $\tilde{\beta}$, которая удовлетворяет задаче минимизации с ограничениями, называется **регрессией с ограничениями, или методом наименьших квадратов с ограничениями** (*restricted regression, restricted least squares*). В качестве аналитического упражнения читателям предлагается показать, что F-отношение можно представить в следующем виде:

$$F = \frac{(SSR_R - SSR_U)/\#\boldsymbol{r}}{SSR_U/(n-K)}, \tag{1.4.11}$$

то есть как разность двух целевых функций, деленную на оценку дисперсии ошибки. Такой способ выведения F-отношения похож на выведение статистики отношения максимальных правдоподобий при оценивании методом максимального правдоподобия (эта статистика рассчитывается как разница логарифмов правдоподобия с ограничениями, накладываемыми нулевой гипотезой, и без них). По этой причине о выводе F-отношения в такой форме говорят, что он основывается на **принципе отношения правдоподобий** (*Likelihood-Ratio principle*). Выражение для OLS-оценки с ограничениями можно найти аналитически. Это оставлено читателям в качестве аналитического упражнения. Вычисление OLS-оценки с ограничениями будет проиллюстрировано в параграфе 1.7 на эмпирическом примере.

Сравнение t и F

Поскольку гипотезы относительно отдельных коэффициентов регрессии являются линейными, t-тест $H_0\colon \beta_k = \overline{\beta}_k$ является частным случаем F-теста. Чтобы увидеть это, заметим, что такая гипотеза может быть переписана в форме $\boldsymbol{R}\beta = \boldsymbol{r}$, если определить $\boldsymbol{R}$ и $\boldsymbol{r}$ следующим образом:

$$\underset{(1\times K)}{\boldsymbol{R}} = \begin{bmatrix} 0 & \cdots & 0 & \underset{(k)}{1} & 0 & \cdots & 0 \end{bmatrix}, \quad \boldsymbol{r} = \overline{\beta}_k.$$

Поэтому, в соответствии с (1.4.9), F-отношение равно:

$$F = (b_k - \overline{\beta}_k)\left[s^2 \cdot (k,k) \text{ элемент матрицы } (\boldsymbol{X}'\boldsymbol{X})^{-1}\right]^{-1}(b_k - \overline{\beta}_k),$$

что равно квадрату t-отношения (1.4.5). Поскольку случайная величина, распределенная по $F(1, n-k)$, представляет собой квадрат случайной величины, распределенной по $t(n-K)$, то F- и t-тесты дают одинаковые результаты.

Иногда нулевая гипотеза состоит в том, что несколько коэффициентов регрессии принимают конкретные значения. Например, предположим, что $K = 2$, и рассмотрим нулевую гипотезу:

$$H_0\colon \beta_1 = 1 \quad \text{и} \quad \beta_2 = 0.$$

Это можно записать в виде линейной гипотезы $\boldsymbol{R\beta} = \boldsymbol{r}$, где $\boldsymbol{R} = \boldsymbol{I}_2$ и $\boldsymbol{r} = (1,0)'$. То есть можно использовать F-тест. Однако заманчивой является идея использования t-тестов для каждого коэффициента по отдельности. Мы могли бы не отвергать H_0, когда оба ограничения $\beta_1 = 1$ и $\beta_2 = 0$ проходят t-тест. Это сводится к использованию доверительной области

$$\big\{(\beta_1, \beta_2) | b_1 - SE(b_1) \cdot t_{\alpha/2}(n-K) < \beta_1 < b_1 + SE(b_1) \cdot t_{\alpha/2}(n-K), \\ b_2 - SE(b_2) \cdot t_{\alpha/2}(n-K) < \beta_2 < b_2 + SE(b_2) \cdot t_{\alpha/2}(n-K)\big\},$$

которая представляет собой прямоугольную область на плоскости (β_1, β_2), как показано на рис. 1.5. Если точка $(1,0)$ (точка на плоскости (β_1, β_2), специфицированная нулевой гипотезой) попадает в эту область, нулевая гипотеза не отвергается. С другой стороны, доверительной областью для F-теста является:

$$\left\{(\beta_1, \beta_2) | (b_1 - \beta_1, b_2 - \beta_2)\big(\widehat{\mathrm{Var}(\boldsymbol{b}|\boldsymbol{X})}\big)^{-1} \begin{bmatrix} b_1 - \beta_1 \\ b_2 - \beta_2 \end{bmatrix} < 2F_\alpha(\#\boldsymbol{r}, n-K)\right\}.$$

Поскольку матрица $\widehat{\mathrm{Var}(\boldsymbol{b}|\boldsymbol{X})}$ положительно определена, доверительная область для F-теста представляет собой эллипс в плоскости (β_1, β_2). Две найденные доверительные области обычно выглядят примерно так, как изображено на рис. 1.5.

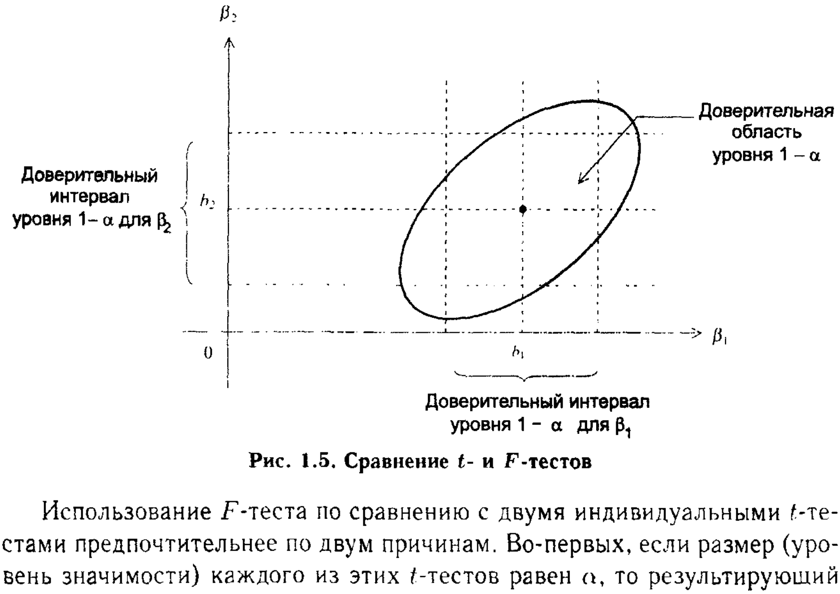

Рис. 1.5. Сравнение t- и F-тестов

Использование F-теста по сравнению с двумя индивидуальными t-тестами предпочтительнее по двум причинам. Во-первых, если размер (уровень значимости) каждого из этих t-тестов равен α, то результирующий размер (вероятность того, что точка $(1,0)$ находится за пределами прямоугольной области) не равен α. Во-вторых, как будет отмечено в следующем параграфе (см. (1.5.19)), F-тест является тестом отношения правдоподобий, а тесты отношения правдоподобий обладают определенными желательными свойствами. То есть, даже если уровень значимости каждого t-теста подобран так, чтобы результирующий уровень значимости был равен α, такой путь все равно менее предпочтителен, чем использование F-теста[1].

Пример тестовой статистики, распределение которой зависит от X

Чтобы вписать текущий параграф в более широкую картину, необходимо отметить, что существуют некоторые статистики, условное распределение которых зависит от $\boldsymbol{X}$. Рассмотрим знаменитую **статистику Дарбина — Уотсона** (*Durbin — Watson statistic*):

$$\frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}.$$

Условное распределение этой статистики и, следовательно, ее критические значения зависят от $\boldsymbol{X}$. Однако Дарбин и Уотсон показали, что эти критические значения попадают в определенный интервал (границы которого зависят от размера выборки, количества регрессоров и от того, включена ли в модель константа). Поэтому критические значения безусловного распределения также попадают в эти границы.

Эта статистика предназначена для проверки гипотезы об отсутствии сериальной корреляции в ошибках. Таким образом, нулевая гипотеза отражает предположение 1.4, в то время как поддерживаемая гипотеза представляет прочие предположения классической регрессионной модели (включая строгую экзогенность) и нормальность. Однако, как уже подчеркивалось в параграфе 1.1, предположение о строгой экзогенности не выполняется в моделях временных рядов, а сериальная корреляция может возникнуть только в таких моделях. Таким образом, статистика Дарбина — Уотсона бесполезна в эконометрике. Тесты на сериальную корреляцию, основанные на асимптотической теории, будут изложены в следующей главе.

Контрольные вопросы

1. (Условное и безусловное распределение.) Следует ли из предположений 1.1–1.5, что предельное (безусловное) распределение $\boldsymbol{b}$ нормально? [Ответ: Нет.] Распределены ли статистики z_k (см. (1.4.3)), t_k и F независимо от $\boldsymbol{X}$? [Ответ: Да, так как их условные распределения относительно $\boldsymbol{X}$ не зависят от $\boldsymbol{X}$.]

2. (Вычисление тестовых статистик.) Проверьте, что $SE(b_k)$, $\boldsymbol{b}$, SSR, s^2 и R^2 можно вычислить, на основе следующих выборочных средних: $\boldsymbol{S}_{XX}$, $\boldsymbol{s}_{XY}$, $\boldsymbol{y}'\boldsymbol{y}/n$ и $\overline{y}$.

[1] Взаимосвязь между t- и F-тестами более подробно изложена в: [Scheffe, 1959, p. 46].

3. Чтобы формула (1.4.9) для F была определена, матрица $\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'$ должна быть невырожденной. Докажите более строгий результат, что эта матрица должна быть положительно определенной. **Указание:** $\boldsymbol{X}'\boldsymbol{X}$ — *положительно определенная матрица. Матрица, обратная к положительно определенной матрице, также положительно определена. Так как матрица* $\boldsymbol{R}(\#\boldsymbol{r}\times K)$ *полного строкового ранга, то для любого ненулевого* $\#\boldsymbol{r}$*-мерного вектора* $\boldsymbol{z}$ $\boldsymbol{R}'\boldsymbol{z}\neq\boldsymbol{0}$.

4. (Односторонний t-тест.) t-тест, описанный выше, является **двусторонним t-тестом** (*two-tailed t-test*), так как уровень значимости α поровну распределен между двумя хвостами t-распределения. Предположим, что альтернатива односторонняя и записывается как $\mathrm{H}_1\colon \beta_k > \overline{\beta}_k$. Рассмотрим следующую модификацию правила принятия решений на основании t-теста.

   Шаг 1: Тот же, что и ранее.

   Шаг 2: Найдите такое критическое значение t_α, чтобы площадь под кривой плотности t-распределения справа от t_α равнялась бы α. Обратите внимание на отличие от двустороннего теста: левый хвост распределения игнорируется и площадь размера α приписывается только правому хвосту.

   Шаг 3: Нулевая гипотеза не отвергается, если $t_k < t_\alpha$. В противном случае нулевая гипотеза отвергается.

   Покажите, что размер (уровень значимости) такого **одностороннего t-теста** равен α.

5. (Взаимосвязь между $F(1, n-K)$ и $t(n-K)$.) Воспользуйтесь таблицами критических значений для F- и t-распределений и проверьте, что $F_\alpha(1, n-K) = \left(t_{\alpha/2}(n-K)\right)^2$ для количества степеней свободы и уровня значимости по вашему выбору.

6. (t в сравнении с F.) «Бессмысленно тестировать гипотезу, состоящую из большого количества ограничений в виде равенства, так как t-тест, скорее всего, отвергнет по крайней мере некоторые из этих ограничений». Укажите на ошибки в этом утверждении.

7. (Дисперсия s^2.) Покажите, что при предположениях 1.1–1.5

$$\mathrm{Var}(s^2|\boldsymbol{X}) = \frac{2\sigma^4}{n-K}.$$

   Указание: *Если случайная величина распределена как $\chi^2(m)$, то ее математическое ожидание равно m, а дисперсия равна $2m$.*

1.5. Связь с методом максимального правдоподобия

После того как мы специфицировали распределение вектора ошибок ε, для оценивания параметров модели (β, σ^2) можно использовать **метод максимального правдоподобия** (*maximum likelihood principle* — **ML**)[1].

[1]Более подробно метод максимального правдоподобия описан в главе 7.

В этом параграфе мы покажем, что OLS-оценка вектора β ($\boldsymbol{b}$) совпадает с ML-оценкой, а OLS-оценка σ^2 лишь слегка отличается от ML-оценки при предположении нормальности ошибок. Мы также покажем, что $\boldsymbol{b}$ достигает **нижней границы Крамера — Рао** (*Cramer — Rao lower bound*).

Метод максимального правдоподобия

Как вы, возможно, помните из вводного курса статистики, основная идея метода максимального правдоподобия состоит в том, что оценки параметров вычисляются таким образом, чтобы максимизировать вероятность получения имеющейся выборки. Более формально, мы предполагаем, что плотность выборки $(\boldsymbol{y}, \boldsymbol{X})$ принадлежит семейству плотностей с конечномерным вектором параметров $\tilde{\boldsymbol{\zeta}}$: $f(\boldsymbol{y}, \boldsymbol{X}, \tilde{\boldsymbol{\zeta}})$. (Это называется **параметризацией** (*parameterizing*) функции плотности.) Эта функция, если рассматривать ее как функцию от гипотетического вектора параметров $\tilde{\boldsymbol{\zeta}}$, называется **функцией правдоподобия** (*likelihood function*). При истинном значении вектора параметров $\boldsymbol{\zeta}$ функция плотности $(\boldsymbol{y}, \boldsymbol{X})$ равна $f(\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\zeta})$. ML-оценкой неизвестного вектора параметров $\boldsymbol{\zeta}$ называется вектор $\tilde{\boldsymbol{\zeta}}$, который максимизирует функцию правдоподобия с учетом имеющейся выборки $(\boldsymbol{y}, \boldsymbol{X})$.

Условная и безусловная функции правдоподобия

Поскольку (совместная) функция плотности вероятности равна произведению частной функции плотности вероятности на условную, плотность $(\boldsymbol{y}, \boldsymbol{X})$ можно записать как

$$f(\boldsymbol{y}, \boldsymbol{X}; \boldsymbol{\zeta}) = f(\boldsymbol{y}|\boldsymbol{X}; \boldsymbol{\theta}) \cdot f(\boldsymbol{X}; \boldsymbol{\psi}), \tag{1.5.1}$$

где $\boldsymbol{\theta}$ — подмножество вектора параметров $\boldsymbol{\zeta}$, которое определяет условную функцию плотности, а $\boldsymbol{\psi}$ — подмножество, определяющее частную функцию плотности. Интересующие нас параметры заключены в векторе $\boldsymbol{\theta}$. Для линейной модели регрессии с нормально распределенными ошибками $\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma^2)'$ и $f(\boldsymbol{y}|\boldsymbol{X}; \boldsymbol{\theta})$ определена в (1.5.4) ниже по тексту.

Пусть $\tilde{\boldsymbol{\zeta}} \equiv (\tilde{\boldsymbol{\theta}}', \tilde{\boldsymbol{\psi}}')'$ — гипотетическое значение $\boldsymbol{\zeta} = (\boldsymbol{\theta}', \boldsymbol{\psi}')'$. Тогда (безусловная, или совместная) функция правдоподобия имеет вид:

$$f(\boldsymbol{y}, \boldsymbol{X}; \tilde{\boldsymbol{\zeta}}) = f(\boldsymbol{y}|\boldsymbol{X}; \tilde{\boldsymbol{\theta}}) \cdot f(\boldsymbol{X}; \tilde{\boldsymbol{\psi}}). \tag{1.5.2}$$

Если бы параметрическая форма $f(\boldsymbol{X}; \tilde{\boldsymbol{\psi}})$ была известна, то мы могли бы максимизировать эту совместную функцию правдоподобия по всему гипотетическому вектору параметров $\tilde{\boldsymbol{\zeta}}$ и ML-оценка $\boldsymbol{\theta}$ была бы просто подмножеством вектора ML-оценок $\boldsymbol{\zeta}$. Мы не можем пойти по такому пути, так как модель классической линейной регрессии не специфицирует функцию плотности вероятности $f(\boldsymbol{X}; \tilde{\boldsymbol{\zeta}})$. Однако при отсутствии функциональной зависимости между $\tilde{\theta}$ и $\tilde{\psi}$ (например, возможно, что подмножество $\tilde{\theta}$ является функцией от $\tilde{\psi}$) максимум (1.5.2) по $\tilde{\zeta}$ достигается при независимой максимизации $f(\boldsymbol{y}|\boldsymbol{X}:\tilde{\boldsymbol{\theta}})$ по $\tilde{\boldsymbol{\theta}}$ и $f(\boldsymbol{X}:\tilde{\psi})$ по $\tilde{\psi}$. Таким образом, ML-оценка $\boldsymbol{\theta}$ также доставляет максимум **условной функции правдоподобия** (*conditional likelihood*) $f(\boldsymbol{y},\boldsymbol{X}:\tilde{\zeta})$.

Логарифмическое правдоподобие для регрессионной модели

Как мы уже видели, из совокупности предположений 1.5 (о нормальности ошибок), 1.2 и 1.4 следует, что распределение ε при условии $\boldsymbol{X}$ нормально $N(\boldsymbol{0},\sigma^2\boldsymbol{I}_n)$ (см. (1.4.1)). Учитывая, что $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, по предположению 1.1, мы получаем:

$$\boldsymbol{y}|\boldsymbol{X} \sim N(\boldsymbol{X}\boldsymbol{\beta},\sigma^2\boldsymbol{I}_n). \tag{1.5.3}$$

Таким образом, условная плотность $\boldsymbol{y}$ при условии $\boldsymbol{X}$ равна[1]:

$$f(\boldsymbol{y}|\boldsymbol{X}) = (2\pi\sigma^2)^{-n/2}\exp\left[-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\beta})\right]. \tag{1.5.4}$$

Заменяя истинные значения параметров $(\boldsymbol{\beta},\sigma^2)$ их гипотетическими значениями $(\tilde{\boldsymbol{\beta}},\tilde{\sigma}^2)$ и беря логарифм, получаем **логарифмическую функцию правдоподобия** (*log likelihood function*):

$$\log L(\tilde{\boldsymbol{\beta}},\tilde{\sigma}^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\tilde{\sigma}^2) - \frac{1}{2\tilde{\sigma}^2}(\boldsymbol{y}-\boldsymbol{X}\tilde{\boldsymbol{\beta}})'(\boldsymbol{y}-\boldsymbol{X}\tilde{\boldsymbol{\beta}}). \tag{1.5.5}$$

Поскольку взятие логарифма — монотонное преобразование, ML-оценкой $(\boldsymbol{\beta},\sigma^2)$ является пара $(\tilde{\boldsymbol{\beta}},\tilde{\sigma}^2)$, максимизирующая логарифм функции правдоподобия.

Максимизация правдоподобия с помощью концентрированной функции правдоподобия

Полезным упражнением является нахождение максимума логарифмической функции правдоподобия в два этапа. Сначала найдите максимум выражения по $\tilde{\boldsymbol{\beta}}$ для любого заданного значения $\tilde{\sigma}^2$. В общем случае $\tilde{\boldsymbol{\beta}}$, при котором достигается максимум, может зависеть от $\tilde{\sigma}^2$ (однако в данном случае это не так при выполнении предположений 1.1–1.5). После этого найдите максимум по $\tilde{\sigma}^2$, принимая во внимание, что $\tilde{\boldsymbol{\beta}}$ (найденный

[1] Вспомним из элементарной теории вероятностей, что функция плотности вероятности для n-мерного нормального распределения с математическим ожиданием $\boldsymbol{\mu}$ и ковариационной матрицей $\boldsymbol{\Sigma}$ равна:

$$(2\pi)^{-n/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\left[-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{y}-\boldsymbol{\mu})\right].$$

Чтобы получить (1.5.4), просто положите $\boldsymbol{\mu} = \boldsymbol{X}\boldsymbol{\beta}$ и $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}_n$.

на предыдущем шаге) может зависеть от $\widetilde{\sigma}^2$. Логарифмическая функция правдоподобия, в которой $\widetilde{\boldsymbol{\beta}}$ представляет собой выражение из первого шага процедуры, называется **концентрированной (по $\widetilde{\boldsymbol{\beta}}$) логарифмической функцией правдоподобия** (*concentrated (with respect to $\boldsymbol{\beta}$) log likelihood function*). Для нормального логарифмического правдоподобия (1.5.5) первый шаг сводится к минимизации суммы квадратов $(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}})$. Значение вектора $\widetilde{\boldsymbol{\beta}}$, при котором достигается минимум, совпадает с OLS-оценкой $\boldsymbol{b}$, а значение суммы квадратов в точке минимума равно $\boldsymbol{e}'\boldsymbol{e}$. Таким образом, концентрированное логарифмическое правдоподобие равно:

$$\text{концентрированное логарифмическое правдоподобие} = \\ = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\widetilde{\sigma}^2) - \frac{1}{2\widetilde{\sigma}^2}\,\boldsymbol{e}'\boldsymbol{e}. \quad (1.5.6)$$

Оно представляет собой функцию только от $\widetilde{\sigma}^2$, а значение $\widetilde{\sigma}^2$, максимизирующее концентрированное логарифмическое правдоподобие, — это ML-оценка σ^2. В рассматриваемом случае классической линейной регрессионной модели максимум можно найти непосредственно, так как $\boldsymbol{e}'\boldsymbol{e}$ не зависит от σ^2 и, следовательно, может рассматриваться как константа. Однако взятие производной по $\widetilde{\sigma}^2$ (а не по $\widetilde{\sigma}$) может быть затруднительным. Но проблем можно избежать, если обозначить $\widetilde{\sigma}^2$ за $\widetilde{\gamma}$. Беря производную от (1.5.6) по $\widetilde{\gamma}$ ($\equiv \widetilde{\sigma}^2$) и приравнивая ее нулю, получаем следующий результат.

Утверждение 1.5 (ML-оценка $(\boldsymbol{\beta}, \sigma^2)$): *Пусть предположения 1.1–1.5 выполнены. Тогда ML-оценка $\boldsymbol{\beta}$ совпадает с OLS-оценкой ($\boldsymbol{b}$), и*

$$\textit{ML-оценка } \sigma^2 = \frac{1}{n}\,\boldsymbol{e}'\boldsymbol{e} = \frac{SSR}{n} = \frac{n-K}{n}s^2. \quad (1.5.7)$$

Из утверждения 1.2 нам известно, что s^2 не смещена. Так как s^2 умножается на $(n-K)/n$ (множитель, отличный от единицы), то ML-оценка σ^2 является смещенной, хотя смещение стремится к нулю при количестве наблюдений n, стремящемся к бесконечности, для любого фиксированного K.

Чтобы воспользоваться этим в дальнейшем, вычислим значение функции правдоподобия в точке максимума. Подставляя (1.5.7) в (1.5.6), получаем:

$$\text{значение логарифма правдоподобия в точке максимума} = \\ = -\frac{n}{2}\log\Big(\frac{2\pi}{n}\Big) - \frac{n}{2} - \frac{n}{2}\log(SSR).$$

так что значение функции правдоподобия в точке максимума равно:

$$\max_{\tilde{\beta},\tilde{\sigma}^2} L(\tilde{\beta},\tilde{\sigma}^2) = \left(\frac{2\pi}{n}\right)^{-n/2} \cdot \exp\left(-\frac{n}{2}\right) \cdot (SSR)^{-n/2}. \qquad (1.5.8)$$

Граница Крамера — Рао для классической регрессионной модели

Чтобы освежить ваши знания статистики, временно отложим классическую модель линейной регрессии и приведем без доказательства неравенство Крамера — Рао для ковариационной матрицы произвольной несмещенной оценки. Для этого определим **вектор вклада** (*score vector*) для гипотетического значения параметра $\tilde{\boldsymbol{\theta}}$ как градиент (вектор частных производных) логарифмической функции правдоподобия:

$$\text{вклад:}\quad \boldsymbol{s}(\tilde{\boldsymbol{\theta}}) \equiv \frac{\partial \log L(\tilde{\boldsymbol{\theta}})}{\partial \tilde{\boldsymbol{\theta}}}. \qquad (1.5.9)$$

Неравенство Крамера — Рао: *Пусть $\boldsymbol{z}$ — вектор случайных величин (необязательно независимых), совместная функция плотности вероятности которых задана как $f(\boldsymbol{z};\boldsymbol{\theta})$, где $\boldsymbol{\theta}$ — m-мерный вектор параметров в некотором пространстве параметров Θ. Пусть $L(\tilde{\boldsymbol{\theta}}) \equiv f(\boldsymbol{z};\tilde{\boldsymbol{\theta}})$ — функция правдоподобия, а $\widehat{\boldsymbol{\theta}}(\boldsymbol{z})$ — несмещенная оценка $\boldsymbol{\theta}$ с конечной ковариационной матрицей. Тогда если выполнены некоторые условия регулярности $f(\boldsymbol{z};\boldsymbol{\theta})$ (неформулируемые здесь), то*

$$\operatorname{Var}[\widehat{\boldsymbol{\theta}}(\boldsymbol{z})] \geq \underset{(m\times m)}{\boldsymbol{I}(\boldsymbol{\theta})^{-1}} \quad (\equiv \textit{нижняя граница Крамера — Рао}),$$

*где $\boldsymbol{I}(\boldsymbol{\theta})$ — **информационная матрица**, определенная как*

$$\boldsymbol{I}(\boldsymbol{\theta}) \equiv \mathrm{E}[\boldsymbol{s}(\boldsymbol{\theta})\, \boldsymbol{s}(\boldsymbol{\theta})']. \qquad (1.5.10)$$

(Обратите внимание, что в формуле стоит значение вклада в точке истинного значения параметра $\boldsymbol{\theta}$.) Также при выполнении указанных условий регулярности информационная матрица равна матрице, обратной к математическому ожиданию матрицы Гессе (матрице вторых производных) логарифмической функции правдоподобия:

$$\boldsymbol{I}(\boldsymbol{\theta}) = -\,\mathrm{E}\left[\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial \tilde{\boldsymbol{\theta}}\, \partial \tilde{\boldsymbol{\theta}}'}\right]. \qquad (1.5.11)$$

*Это выражение называется **равенством для информационных матриц** (information matrix equality).*

По поводу формулирования условий регулярности и доказательства соответствующего утверждения см., например: [Amemiya, 1985, Theorem 1.3.1]. Условия регулярности обеспечивают то, что операции дифференцирования и взятия математического ожидания можно поменять местами. Поэтому, например,

$$\mathrm{E}[\partial L(\boldsymbol{\theta})/\partial\widetilde{\boldsymbol{\theta}}] = \partial\,\mathrm{E}[L(\boldsymbol{\theta})]/\partial\widetilde{\boldsymbol{\theta}}.$$

Вернемся к модели классической линейной регрессии (предположения 1.1–1.5 выполнены). Функция правдоподобия $L(\widetilde{\boldsymbol{\theta}})$ в неравенстве Крамера — Рао равна условной плотности (1.5.4), так что дисперсия в неравенстве равна условной дисперсии при условии $\boldsymbol{X}$. Можно показать, что условия регулярности выполнены для функции плотности нормального распределения (1.5.4) (см., например: [Amemiya, 1985, Sections 1.3.2 and 1.3.3]). В оставшейся части этого подпараграфа мы вычислим **информационную матрицу** (*information matrix*) для (1.5.4). Вектор параметров $\boldsymbol{\theta}$ равен $(\boldsymbol{\beta}', \sigma^2)'$. Поэтому $\widetilde{\boldsymbol{\theta}} = (\widetilde{\boldsymbol{\beta}}', \widetilde{\gamma}')'$, и матрица вторых производных равна:

$$\underset{((K+1)\times(K+1))}{\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial\widetilde{\boldsymbol{\theta}}\,\partial\widetilde{\boldsymbol{\theta}}'}} = \begin{bmatrix} \underset{(K\times K)}{\dfrac{\partial^2 \log L(\boldsymbol{\theta})}{\partial\widetilde{\boldsymbol{\beta}}\,\partial\widetilde{\boldsymbol{\beta}}'}} & \underset{(K\times 1)}{\dfrac{\partial^2 \log L(\boldsymbol{\theta})}{\partial\widetilde{\boldsymbol{\beta}}\,\partial\widetilde{\gamma}}} \\ \underset{(1\times K)}{\dfrac{\partial^2 \log L(\boldsymbol{\theta})}{\partial\widetilde{\gamma}\,\partial\widetilde{\boldsymbol{\beta}}'}} & \underset{(1\times 1)}{\dfrac{\partial^2 \log L(\boldsymbol{\theta})}{\partial\widetilde{\gamma}^2}} \end{bmatrix}. \tag{1.5.12}$$

Первая и вторая производные логарифмической функции правдоподобия (1.5.5) по $\widetilde{\boldsymbol{\theta}}$ в точке истинного значения вектора $\boldsymbol{\theta}$ равны:

$$\frac{\partial \log L(\boldsymbol{\theta})}{\partial\widetilde{\boldsymbol{\beta}}} = \frac{1}{\gamma}\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}), \tag{1.5.13a}$$

$$\frac{\partial \log L(\boldsymbol{\theta})}{\partial\widetilde{\gamma}} = -\frac{n}{2\gamma} + \frac{1}{2\gamma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}). \tag{1.5.13b}$$

$$\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial\widetilde{\boldsymbol{\beta}}\,\partial\widetilde{\boldsymbol{\beta}}'} = -\frac{1}{\gamma}\boldsymbol{X}'\boldsymbol{X}. \tag{1.5.14a}$$

$$\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial\widetilde{\gamma}^2} = \frac{n}{2\gamma^2} - \frac{1}{\gamma^3}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}), \tag{1.5.14b}$$

$$\frac{\partial^2 \log L(\boldsymbol{\theta})}{\partial\widetilde{\boldsymbol{\beta}}\,\partial\widetilde{\gamma}} = -\frac{1}{\gamma^2}\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}). \tag{1.5.14c}$$

Поскольку эти производные вычислены при истинном значении параметра, в этих выражениях $\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} = \boldsymbol{\varepsilon}$. Подставляя (1.5.14) в (1.5.12) и используя $\mathrm{E}(\boldsymbol{\varepsilon}|\boldsymbol{X}) = \boldsymbol{0}$ (предположение 1.2), $\mathrm{E}(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}|\boldsymbol{X}) = n\sigma^2$ (следствие из предположения 1.4), также вспоминая о произведенной замене $\gamma = \sigma^2$, несложно вывести

$$\boldsymbol{I}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{1}{\sigma^2}\boldsymbol{X}'\boldsymbol{X} & \boldsymbol{0} \\ \boldsymbol{0}' & \frac{n}{2\sigma^4} \end{bmatrix}. \tag{1.5.15}$$

Здесь используется условное математическое ожидание относительно $\boldsymbol{X}$, так как функция правдоподобия (1.5.4) представляет собой условную плотность вероятности относительно $\boldsymbol{X}$. Обращая эту блочно-диагональную матрицу, получаем границу Крамера — Рао:

$$\text{граница Крамера — Рао} \equiv \boldsymbol{I}(\boldsymbol{\theta})^{-1} = \begin{bmatrix} \sigma^2 \cdot (\boldsymbol{X}'\boldsymbol{X})^{-1} & \mathbf{0} \\ \mathbf{0}' & \frac{2\sigma^4}{n} \end{bmatrix}. \tag{1.5.16}$$

Поэтому несмещенная оценка $\boldsymbol{b}$ с дисперсией $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$ (по утверждению 1.1) достигает границы Крамера — Рао. Таким образом, мы доказали

Утверждение 1.6 ($\boldsymbol{b}$ — наилучшая несмещенная оценка (BUE)): *При предположениях 1.1–1.5 OLS-оценка $\boldsymbol{b}$ неизвестного вектора параметров $\boldsymbol{\beta}$ является **наилучшей несмещенной оценкой** (Best Unbiased Estimator) в том смысле, что любая другая несмещенная оценка (но необязательно линейная) обладает большей (в матричном смысле) условной дисперсией.*

Этот результат не следует путать с теоремой Гаусса — Маркова, которая утверждает, что $\boldsymbol{b}$ обладает наименьшей дисперсией в классе несмещенных *и* линейных по $\boldsymbol{y}$ оценок. Утверждение 1.6 говорит, что $\boldsymbol{b}$ обладает наименьшей дисперсией в более широком классе оценок, который включает в себя линейные несмещенные оценки. Более сильное утверждение мы получили с использованием предположения нормальности (предположение 1.5), которого нет в теореме Гаусса — Маркова. Иными словами, теорема Гаусса — Маркова не исключает возможности существования нелинейной оценки с меньшей дисперсией, чем дисперсия OLS-оценки, однако такая возможность исключается при введении предположения нормальности.

Как мы уже видели, ML-оценка σ^2 является смещенной, следовательно, неравенство Крамера — Рао неприменимо. Однако МНК-оценка дисперсии s^2 является несмещенной. Достигает ли она границы Крамера — Рао? В одном из контрольных вопросов к предыдущему параграфу мы показали, что

$$\operatorname{Var}(s^2|\boldsymbol{X}) = \frac{2\sigma^4}{n-K}$$

при том же самом наборе предположений, что и в утверждении 1.6. Поэтому s^2 не достигает границы Крамера — Рао, равной $2\sigma^4/n$. Однако можно показать, что не существует несмещенной оценки σ^2 с дисперсией, меньшей чем $2\sigma^4/(n-K)$ (см., например: [Рао, 1973, р. 319]).

F-тест как тест отношения правдоподобий

Тест отношения правдоподобий (*likelihood ratio test*) для некоторой нулевой гипотезы сравнивает значения L_U (максимальное значение функции правдоподобия, полученное без учета ограничений, накладываемых нулевой гипотезой) и L_R (максимальное значение функции правдоподобия, полученное с учетом ограничений, накладываемых нулевой гипотезой). Если значение отношения правдоподобий $\lambda \equiv L_U/L_R$ слишком велико, то это должно быть сигналом к тому, что нулевая гипотеза неверна. F-тест, проверяющий нулевую гипотезу H_0: $\boldsymbol{R\beta} = \boldsymbol{r}$, рассмотренный в предыдущем параграфе, является тестом отношения правдоподобий, поскольку F-отношение является монотонным преобразованием отношения правдоподобий λ. Для рассматриваемой сейчас модели L_U вычисляется по формуле (1.5.8), где SSR (сумма квадратов остатков, минимизированная без учета ограничений H_0) равна SSR_U в (1.4.11). Значение L_R получается путем замены SRR на сумму квадратов остатков при ограничении, SSR_R. Таким образом,

$$L_R = \max_{\widetilde{\boldsymbol{\beta}}, \widetilde{\sigma}^2 \text{ при } H_0} L(\widetilde{\boldsymbol{\beta}}, \widetilde{\sigma}^2) = \left(\frac{2\pi}{n}\right)^{-n/2} \cdot \exp\left(-\frac{n}{2}\right) \cdot (SSR_R)^{-n/2}, \quad (1.5.17)$$

и отношение правдоподобий равно:

$$\lambda \equiv \frac{L_U}{L_R} = \left(\frac{SSR_U}{SSR_R}\right)^{-n/2}. \quad (1.5.18)$$

Сравнивая эту формулу с формулой (1.4.11) для F-отношения, можно увидеть, что F-отношение является монотонным преобразованием отношения правдоподобий λ:

$$F = \frac{n-K}{\#\boldsymbol{r}}(\lambda^{2/n} - 1), \quad (1.5.19)$$

следовательно, эти два теста эквивалентны.

Метод квазимаксимального правдоподобия

Все результаты, полученные до сих пор, базируются на предположении о нормальности ошибок. Без этого предположения нет гарантии того, что ML-оценка вектора параметров $\boldsymbol{\beta}$ совпадает с OLS-оценкой (утверждение 1.5) или что OLS-оценка $\boldsymbol{b}$ достигает границы Крамера — Рао (утверждение 1.6). Однако из утверждения 1.5 следует, что $\boldsymbol{b}$ является оценкой **метода квази- (или псевдо-) максимального правдоподобия** (*quasi- (pseudo-) maximum likelihood estimator* — **quasi-ML**). Так называется оценка, максимизирующая неправильно специфицированную функцию правдоподобия. В нашем случае это функция правдоподобия для нормального распределения. Следовательно, результаты параграфа 1.3 могут быть интерпретированы как дающие свойства квази-ML-оценки на конечных выборках, когда мы некорректно предполагаем нормальность ошибок.

Контрольные вопросы

1. (Использование условий регулярности.) Предполагая, что можно менять местами операции взятия математического ожидания (то есть взятия интеграла) и дифференцирования, докажите, что математическое ожидание вклада в (1.5.9), оцененное при истинном значении вектора $\boldsymbol{\theta}$, равно нулю. **Указание:** *Необходимо показать, что*

$$\int \frac{\partial \log f(\boldsymbol{z}; \boldsymbol{\theta})}{\partial \widetilde{\boldsymbol{\theta}}} f(\boldsymbol{z}; \boldsymbol{\theta}) \mathrm{d}z = \mathbf{0}.$$

Так как $f(\boldsymbol{z}; \widetilde{\boldsymbol{\theta}})$ — *функция плотности, то* $\int f(\boldsymbol{z}; \widetilde{\boldsymbol{\theta}}) \mathrm{d}\boldsymbol{z} = 1$ *для любого* $\widetilde{\boldsymbol{\theta}}$. *Продифференцируйте обе части по* $\widetilde{\boldsymbol{\theta}}$ *и используйте условия регулярности, позволяющие менять местами операции интегрирования и дифференцирования, чтобы получить* $\int [\partial f(\boldsymbol{z}; \boldsymbol{\theta})] \, \mathrm{d}\boldsymbol{z} = \mathbf{0}$. *Также вспомните, что*

$$\frac{\partial \log f(\boldsymbol{z}; \boldsymbol{\theta})}{\partial \widetilde{\boldsymbol{\theta}}} = \frac{1}{f(\boldsymbol{z}; \boldsymbol{\theta})} \frac{\partial f(\boldsymbol{z}; \boldsymbol{\theta})}{\partial \widetilde{\boldsymbol{\theta}}}.$$

2. (Максимизация логарифма совместной функции правдоподобия.) Рассмотрим задачу максимизации логарифма совместной функции правдоподобия (1.5.2) для классической модели регрессии, где $\widetilde{\boldsymbol{\theta}} = (\widetilde{\boldsymbol{\beta}}', \widetilde{\sigma}^2)'$, и $\log f(\boldsymbol{y} | \boldsymbol{X}; \widetilde{\boldsymbol{\theta}})$ задан в (1.5.5). Вам необходимо параметризовать маргинальную функцию правдоподобия $f(\boldsymbol{X}; \boldsymbol{\psi})$ и взять логарифм от (1.5.2), чтобы получить целевую функцию, которую, в свою очередь, необходимо будет максимизировать по $\boldsymbol{\zeta} = (\boldsymbol{\theta}', \boldsymbol{\psi}')'$. Что является ML-оценкой $\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma^2)'$? [Ответ: Она будет такой же, как и оценка из утверждения 1.5.] Выведите границу Крамера — Рао для $\boldsymbol{\beta}$. **Указание:** *Из равенства для информационных матриц*

$$\boldsymbol{I}(\boldsymbol{\zeta}) = - \mathrm{E} \left[\frac{\partial^2 \log L(\boldsymbol{\zeta})}{\partial \widetilde{\boldsymbol{\zeta}} \partial \widetilde{\boldsymbol{\zeta}}'} \right].$$

Кроме того, $\partial^2 \log L(\boldsymbol{\zeta}) / (\partial \widetilde{\boldsymbol{\theta}} \partial \widetilde{\boldsymbol{\psi}}') = \mathbf{0}$.

3. (Логарифм концентрированной по $\widetilde{\sigma}^2$ функции правдоподобия.) Если обозначить $\widetilde{\sigma}^2$ как $\widetilde{\gamma}$, получаем логарифм функции правдоподобия для классической регрессионной модели в виде:

$$\log L(\widetilde{\boldsymbol{\beta}}, \widetilde{\gamma}) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\widetilde{\gamma}) - \frac{1}{2\widetilde{\gamma}} (\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}).$$

При использовании двухшаговой процедуры максимизации, описанной в тексте, мы сначала максимизировали эту функцию по $\widetilde{\boldsymbol{\beta}}$. Сейчас же на первом шаге найдем максимум по $\widetilde{\gamma}$, считая $\widetilde{\boldsymbol{\beta}}$ заданными. Покажите, что логарифм концентрированной по $\tilde{\gamma} \equiv \tilde{\sigma}^2$ функции правдоподобия равен:

$$-\frac{n}{2}[1+\log(2\pi)] - \frac{n}{2}\log\left(\frac{(\boldsymbol{y}-\boldsymbol{X}\tilde{\boldsymbol{\beta}})'(\boldsymbol{y}-\boldsymbol{X}\tilde{\boldsymbol{\beta}})}{n}\right).$$

4. (Равенство для информационных матриц в классической регрессионной модели.) Проверьте (1.5.11) для линейной регрессионной модели.

5. (Уравнения правдоподобия для классической регрессионной модели.) Мы использовали двухшаговую процедуру, чтобы вывести ML-оценку для классической регрессионной модели. Альтернативный подход поиска ML-оценки состоит в приравнивании нулю правых частей (1.5.13) и решении полученной системы (эти условия первого порядка для логарифмической функции правдоподобия называются **уравнениями правдоподобия** (*likelihood equations*)). Проверьте, что ML-оценка из утверждения 1.5 является решением этой системы уравнений.

1.6. Обобщенный метод наименьших квадратов (GLS)

Предположение 1.4 (о сферичности возмущений ошибок) утверждает, что матрица размера $n \times n$ условных вторых моментов $\mathrm{E}(\boldsymbol{\varepsilon\varepsilon}'|\boldsymbol{X}$ $(= \mathrm{Var}(\boldsymbol{\varepsilon}|\boldsymbol{X}))$ пропорциональна единичной матрице. Без этого предположения каждый элемент матрицы $n \times n$ является в общем случае нелинейной функцией от $\boldsymbol{X}$. Если ошибки (условно) негомоскедастичны, значения диагональных элементов $\mathrm{E}(\boldsymbol{\varepsilon\varepsilon}'|\boldsymbol{X})$ различны и если присутствует корреляция ошибок между наблюдениями (случай сериальной корреляции для моделей временных рядов), то значения внедиагональных элементов не равны нулю. Для любого заданного положительного скаляра σ^2 определим $\boldsymbol{V}(\boldsymbol{X}) \equiv \mathrm{E}(\boldsymbol{\varepsilon\varepsilon}'|\boldsymbol{X})/\sigma^2$ и предположим, что $\boldsymbol{V}(\boldsymbol{X})$ — невырожденная и известная матрица. То есть

$$\mathrm{E}(\boldsymbol{\varepsilon\varepsilon}'|\boldsymbol{X}) = \sigma^2 \underset{(n\times n)}{\boldsymbol{V}(\boldsymbol{X})},\quad \boldsymbol{V}(\boldsymbol{X}) \text{ — невырожденная и известная матрица.} \tag{1.6.1}$$

Причина проведенной декомпозиции $\mathrm{E}(\boldsymbol{\varepsilon\varepsilon}'|\boldsymbol{X})$ (на компоненту σ^2, общую для всех элементов матрицы $\mathrm{E}(\boldsymbol{\varepsilon\varepsilon}'|\boldsymbol{X})$, и оставшуюся часть $\boldsymbol{V}(\boldsymbol{X})$) состоит в том, что нам не нужно знать значение σ^2 для нахождения эффективной оценки. Модель, которую мы получаем при замене предположения 1.4 на (1.6.1), которое лишь предполагает невырожденность матрицы условных вторых моментов $\mathrm{E}(\boldsymbol{\varepsilon\varepsilon}'|\boldsymbol{X})$, называется **обобщенной регрессионной моделью** (*generalized regression model*).

Последствия ослабления предположения 1.4

Те из полученных в предыдущих параграфах результатов, которые используют предположение 1.4, не сохраняются для обобщенной регрессионной модели. А именно:

- Теорема Гаусса — Маркова больше не выполняется для OLS-оценки
$$b \equiv (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}.$$
BLUE теперь является другая оценка.
- t-отношение более не имеет t-распределение Стьюдента. Значит, t-тест более неприменим. Те же замечания относятся и к F-тесту.
- Однако OLS-оценка все еще *является* несмещенной, так как для выведения этого свойства (утверждение 1.1(a)) предположение 1.4 не использовалось.

Эффективное оценивание при известной матрице V

Если значение $V(\boldsymbol{X})$ известно, существует ли BLUE для обобщенной регрессионной модели? Ответ положителен, и такая оценка называется **оценкой обобщенного метода наименьших квадратов** (*generalized least squares GLS estimator*). Сейчас мы выведем эту оценку. Вывод будет построен на идее сведения обобщенной линейной модели (которая состоит из предположений 1.1–1.3 и (1.6.1)) к модели классической линейной регрессии (удовлетворяющей всем предположениям, включая предположение 1.4).

Для упрощения обозначений далее будем использовать $\boldsymbol{V}$ вместо $\boldsymbol{V}(\boldsymbol{X})$. Так как $\boldsymbol{V}$ — по построению симметричная и положительно определенная матрица, существует такая невырожденная матрица $\boldsymbol{C}$ размера $n \times n$, что
$$\boldsymbol{V}^{-1} = \boldsymbol{C}'\boldsymbol{C}. \tag{1.6.2}$$

Эта декомпозиция неединственна (с более чем одним выбором матрицы $\boldsymbol{C}$), однако, как будет показано ниже, выбор $\boldsymbol{C}$ не играет никакой роли. Создадим теперь новую регрессионную модель для переменных, преобразуя $(\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\varepsilon})$ с помощью матрицы $\boldsymbol{C}$:
$$\widetilde{\boldsymbol{y}} \equiv \boldsymbol{C}\boldsymbol{y}, \quad \widetilde{\boldsymbol{X}} \equiv \boldsymbol{C}\boldsymbol{X}, \quad \widetilde{\boldsymbol{\varepsilon}} \equiv \boldsymbol{C}\boldsymbol{\varepsilon}. \tag{1.6.3}$$

Тогда предположение 1.1 о линейности модели для $(\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\varepsilon})$ ведет к тому, что для преобразованных переменных $(\widetilde{\boldsymbol{y}}, \widetilde{\boldsymbol{X}}, \widetilde{\boldsymbol{\varepsilon}})$ свойство линейности сохраняется:
$$\widetilde{\boldsymbol{y}} = \widetilde{\boldsymbol{X}}\boldsymbol{\beta} + \widetilde{\boldsymbol{\varepsilon}}. \tag{1.6.4}$$

Модель для преобразованных переменных удовлетворяет также и остальным предположениям классической линейной регрессионной модели. Требование строгой экзогенности выполнено, так как

$$
\begin{aligned}
\mathrm{E}(\widetilde{\varepsilon}|\widetilde{\boldsymbol{X}}) &= \mathrm{E}(\widetilde{\varepsilon}|\boldsymbol{X}) \\
&\text{(поскольку матрица } \boldsymbol{C} \text{ невырождена, } \boldsymbol{X} \text{ и } \widetilde{\boldsymbol{X}} \text{ содержат одну и ту же информацию)} \\
&= \mathrm{E}(\boldsymbol{C}\varepsilon|\boldsymbol{X}) = \\
&= \boldsymbol{C}\,\mathrm{E}(\varepsilon|\boldsymbol{X}) \\
&\text{(из линейности условного математического ожидания)} \\
&= \mathbf{0} \quad \text{(поскольку } \mathrm{E}(\varepsilon|\boldsymbol{X}) = \mathbf{0} \text{ по предположению 1.2).}
\end{aligned}
$$

Так как $\boldsymbol{V}$ положительно определена, предположение об отсутствии мультиколлинеарности также выполнено (доказательство этого остается в качестве контрольного вопроса). Предположение 1.4 выполнено для преобразованной модели:

$$
\begin{aligned}
\mathrm{E}(\widetilde{\varepsilon}\widetilde{\varepsilon}'|\widetilde{\boldsymbol{X}}) &= \mathrm{E}(\widetilde{\varepsilon}\widetilde{\varepsilon}'|\boldsymbol{X}) \\
&\text{(так как } \widetilde{\boldsymbol{X}} \text{ и } \boldsymbol{X} \text{ содержат одинаковый объем информации)} \\
&= \boldsymbol{C}\,\mathrm{E}(\varepsilon\varepsilon'|\boldsymbol{X})\boldsymbol{C}' \quad \text{(поскольку } \widetilde{\varepsilon}\widetilde{\varepsilon}'\boldsymbol{C}\varepsilon\varepsilon'\boldsymbol{C}') = \\
&= \boldsymbol{C}\cdot\sigma^2\cdot\boldsymbol{V}\boldsymbol{C}' \quad \text{(из (1.6.1))} = \\
&= \sigma^2\boldsymbol{C}\boldsymbol{V}\boldsymbol{C}' = \\
&= \sigma^2\boldsymbol{I}_n \\
&\text{(так как } (\boldsymbol{C}')^{-1}\boldsymbol{V}^{-1}\boldsymbol{C}^{-1} = \boldsymbol{I}_n \text{ или } \boldsymbol{C}\boldsymbol{V}\boldsymbol{C}' = \boldsymbol{I}_n \text{ из (1.6.2)).}
\end{aligned}
$$

То есть дисперсия преобразованного вектора ошибок $\widetilde{\varepsilon}$ на самом деле сферична. Наконец, распределение $\widetilde{\varepsilon}|\widetilde{\boldsymbol{X}}$ нормально, так как распределение $\widetilde{\varepsilon}|\widetilde{\boldsymbol{X}}$ совпадает с распределением $\widetilde{\varepsilon}|\boldsymbol{X}$, а $\widetilde{\varepsilon}$ является линейным преобразованием ε. На этом проверка выполнения предположений 1.1–1.5 для модели с преобразованными переменными закончена.

Из теоремы Гаусса — Маркова для модели с преобразованными переменными следует, что BLUE для $\boldsymbol{\beta}$ в обобщенной линейной регрессионной модели — это OLS-оценка, примененная к модели (1.6.4):

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}} &= (\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{X}})^{-1}\widetilde{\boldsymbol{X}}'\widetilde{\boldsymbol{y}} = \\
&= [(\boldsymbol{C}\boldsymbol{X})'(\boldsymbol{C}\boldsymbol{X})]^{-1}(\boldsymbol{C}\boldsymbol{X})'\boldsymbol{C}\boldsymbol{y} = \\
&= (\boldsymbol{X}'\boldsymbol{C}'\boldsymbol{C}\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{C}'\boldsymbol{C}\boldsymbol{y}) = \\
&= (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y} \quad \text{(из (1.6.2)).} \qquad (1.6.5)
\end{aligned}
$$

Это GLS-оценка. Ее условная дисперсия равна:

$$\begin{aligned}
&\mathrm{Var}(\widehat{\boldsymbol\beta}_{\mathrm{GLS}}|\boldsymbol X) = \\
&= (\boldsymbol X'\boldsymbol V^{-1}\boldsymbol X)^{-1}\boldsymbol X'\boldsymbol V^{-1}\,\mathrm{Var}(\boldsymbol y|\boldsymbol X)\boldsymbol V^{-1}\boldsymbol X(\boldsymbol X'\boldsymbol V^{-1}\boldsymbol X)^{-1} = \\
&= (\boldsymbol X'\boldsymbol V^{-1}\boldsymbol X)^{-1}\boldsymbol X'\boldsymbol V^{-1}(\sigma^2\boldsymbol V)\boldsymbol V^{-1}\boldsymbol X(\boldsymbol X'\boldsymbol V^{-1}\boldsymbol X)^{-1} \\
&(\text{так как } \mathrm{Var}(\boldsymbol y|\boldsymbol X) = \mathrm{Var}(\boldsymbol\varepsilon|\boldsymbol X)) \\
&= \sigma^2\cdot(\boldsymbol X'\boldsymbol V^{-1}\boldsymbol X)^{-1}.
\end{aligned} \tag{1.6.6}$$

Поскольку замена $\boldsymbol X$ на $\sigma^2\boldsymbol X$ ($=\mathrm{Var}(\boldsymbol\varepsilon|\boldsymbol X)$) в (1.6.5) не приводит к количественному изменению оценки, то GLS-оценку можно также представить в виде:

$$\widehat{\boldsymbol\beta}_{\mathrm{GLS}} = \left[\boldsymbol X'\,\mathrm{Var}(\boldsymbol\varepsilon|\boldsymbol X)^{-1}\boldsymbol X\right]^{-1}\boldsymbol X'\,\mathrm{Var}(\boldsymbol\varepsilon|\boldsymbol X)^{-1}\boldsymbol y.$$

Как уже отмечалось выше, OLS-оценка $(\boldsymbol X'\boldsymbol X)^{-1}\boldsymbol X'\boldsymbol y$ является несмещенной при отсутствии предположения 1.4, однако несмотря на это GLS-оценка (при известной $\boldsymbol V$) предпочтительнее, так как она является более эффективной в том смысле, что ее дисперсия меньше в матричном смысле. Выигрыш в эффективности достигается за счет учета гетероскедастичности и корреляции между наблюдениями в ошибках. На практике это означает, что надо вставить матрицу, обратную к $\mathrm{Var}(\boldsymbol\varepsilon|\boldsymbol X)$ (или пропорциональную ей) в формулу для OLS-оценки, как это было сделано в (1.6.5). Эти рассуждения можно обобщить следующим образом:

Утверждение 1.7 (свойства GLS в конечных выборках):

(a) *(Несмещенность.) При предположениях 1.1–1.3,* $\mathrm{E}(\widehat{\boldsymbol\beta}_{\mathrm{GLS}}|\boldsymbol X) = \boldsymbol\beta$.

(b) *(Формула для дисперсии.) При предположениях 1.1–1.3 и предположении* (1.6.1) *о том, что условный второй момент пропорционален* $\boldsymbol V(\boldsymbol X)$,

$$\mathrm{Var}(\widehat{\boldsymbol\beta}_{\mathrm{GLS}}|\boldsymbol X) = \sigma^2\cdot(\boldsymbol X'\boldsymbol V(\boldsymbol X)^{-1}\boldsymbol X)^{-1}.$$

(c) *(Эффективность GLS.) При том же наборе предположений, что и в пункте (b), GLS-оценка является эффективной: условная дисперсия любой несмещенной оценки, линейной по* $\boldsymbol y$, *больше или равна* $\mathrm{Var}(\widehat{\boldsymbol\beta}_{\mathrm{GLS}}|\boldsymbol X)$ *в матричном смысле.*

Частный случай: взвешенный метод наименьших квадратов (WLS)

Идея корректировки на ковариационную матрицу ошибок наиболее очевидна в случае, когда между ошибками нет корреляции (то есть матрица $\boldsymbol V$ диагональна). Обозначим через $v_i(\boldsymbol X)$ i-й диагональный элемент матрицы $\boldsymbol V(\boldsymbol X)$. Тогда

$$\mathrm{E}(\varepsilon_i^2|\boldsymbol X)\ (=\mathrm{Var}(\varepsilon_i|\boldsymbol X)) = \sigma^2\cdot v_i(\boldsymbol X).$$

Несложно заметить, что матрица $\boldsymbol{C}$ также диагональна, причем ее i-й диагональный элемент равен квадратному корню из $1/v_i(\boldsymbol{X})$. Таким образом, компоненты $(\widetilde{\boldsymbol{y}}, \widetilde{\boldsymbol{X}})$ равны:

$$\widetilde{y}_i = \frac{y_i}{\sqrt{v_i(\boldsymbol{X})}}, \quad \widetilde{\boldsymbol{x}}_i = \frac{\boldsymbol{x}_i}{\sqrt{v_i(\boldsymbol{X})}} \quad (i = 1, 2, \ldots, n).$$

Поэтому эффективное оценивание при гетероскедастичности известной формы состоит в придании каждому наблюдению веса, обратного квадратному корню из дисперсии этого наблюдения $v_i(\boldsymbol{X})$, и последующем применении OLS. Это называется **взвешенной регрессией**, или **взвешенным методом наименьших квадратов** (*weighted regression, weighted least squares* — **WLS**).

Другой важный частный случай — это случайная выборка, в которой $\{y_i, \boldsymbol{x}_i\}$ i.i.d. по i. Как уже было замечено в параграфе 1.1, ошибки здесь безусловно гомоскедастичны (то есть $\mathrm{E}(\varepsilon_i^2)$ не зависит от i), однако GLS все равно можно использовать для повышения эффективности оценки, поскольку ошибки могут быть условно гетероскедастичными. Условный момент второго порядка $\mathrm{E}(\varepsilon_i^2|\boldsymbol{X})$ в ситуации случайной выборки зависит только от $\boldsymbol{x}_i$, и функциональная форма $\mathrm{E}(\varepsilon_i^2|\boldsymbol{X})$ одинакова для всех i. Таким образом,

$$v_i(\boldsymbol{X}) = v(\boldsymbol{x}_i) \quad \text{для случайных выборок.} \tag{1.6.7}$$

То есть знание матрицы $\boldsymbol{V}(\cdot)$ равносильно знанию одной функции K переменных, $v(\cdot)$.

Ограничивающее свойство GLS

Все сделанные ранее радужные выводы о свойствах GLS в конечных выборках основаны на предположении, что регрессоры в обобщенной регрессионной модели строго экзогенны ($\mathrm{E}(\widetilde{\boldsymbol{\varepsilon}}|\widetilde{\boldsymbol{X}}) = \boldsymbol{0}$). Это ограничивает сферу применения GLS. Предположим, как в случае с моделями временных рядов, что регрессоры не удовлетворяют свойству строгой экзогенности и ошибки сериально коррелированы. Тогда ни OLS-, ни GLS-оценки не обладают хорошими свойствами, такими как несмещенность. Однако, как будет показано в следующей главе, OLS-оценка, игнорирующая наличие сериальной корреляции в ошибках, будет обладать рядом хороших свойств в больших выборках (такими как «состоятельность» и «асимптотическая нормальность»), при условии, что регрессоры являются «предопределенными» (это требование слабее, чем строгая экзогенность). GLS-оценка, напротив, не обладает такими подкупающими свойствами. То есть, если регрессоры не строго экзогенны, но являются предопределенными, GLS-процедура, корректирующая на сериальную корреляцию, может дать несостоятельную оценку (см. параграф 6.7).

Процедура, позволяющая учесть сериальную корреляцию и при этом сохранить состоятельность, будет представлена в главе 6.

Несмотря на то что GLS-процедура неприменима для коррекции сериальной корреляции, она все равно может быть использована для коррекции гетероскедастичности, если ошибки сериально не коррелированы с диагональной матрицей $\boldsymbol{V}(\boldsymbol{X})$ (в форме WLS). Однако это только при условии, что матрица $\boldsymbol{V}(\boldsymbol{X})$ известна. Очень редко в нашем распоряжении имеется априорная информация относительно значений диагональных элементов $\boldsymbol{V}(\boldsymbol{X})$, которые необходимы для взвешивания наблюдений. Имея дело со случайной выборкой, в которой не может возникнуть сериальная корреляция, знание $\boldsymbol{V}(\boldsymbol{X})$ сводится, как мы только что видели, к знанию одной функции K переменных, $v(\boldsymbol{x}_i)$. Однако даже в этом случае такая функция неизвестна в большинстве приложений.

Если мы не знаем $\boldsymbol{V}(\boldsymbol{X})$, мы можем оценить ее функциональную форму по имеющейся выборке. Такой подход называется **доступным обобщенным методом наименьших квадратов** (*Feasible Generalized Least Squares* — **FGLS**). Однако если $\boldsymbol{V}(\boldsymbol{X})$ оценивается по выборке, то ее значение $\boldsymbol{V}$ становится случайной величиной, что влияет на распределение GLS-оценки. Для малых выборок о свойствах FGLS известно немного. Асимптотические свойства FGLS будут рассмотрены в следующей главе.

В заключение стоит отметить одну положительную сторону GLS: большинство линейных процедур оценивания (включая 2SLS, 3SLS и оценку в модели со случайными эффектами, которые будут введены позднее) можно представить как GLS-оценку при некотором способе формирования матрицы данных. Однако все эти оценки и OLS могут быть так же проинтерпретированы как оценки обобщенного метода моментов (GMM), а GMM-интерпретация более полезна для вывода асимптотических свойств.

Контрольные вопросы

1. (Предположение об отсутствии мультиколлинеарности в модели с преобразованными переменными) предположение 1.3 для модели с преобразованными переменными состоит в том, что $\operatorname{rank}(\boldsymbol{C}\boldsymbol{X}) = K$. Это условие выполнено, так как $\boldsymbol{C}$ — невырожденная матрица, а $\boldsymbol{X}$ — матрица полного столбцового ранга. Покажите это. **Указание:** *Так как* $\boldsymbol{X}$ *— матрица полного столбцового ранга, то для любого* K*-мерного вектора* $\boldsymbol{c} \neq \boldsymbol{0}$, $\boldsymbol{X}\boldsymbol{c} \neq \boldsymbol{0}$.

2. (Обобщенная сумма квадратов остатков.) Покажите, что при $\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}$ достигается минимум $(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}})'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}})$.

3. Выведите выражение для $\operatorname{Var}(\boldsymbol{b}|\boldsymbol{X})$ для обобщенной регрессионной модели. Как оно соотносится с $\operatorname{Var}(\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}|\boldsymbol{X})$? Проверьте, что из утвержде-

ния 1.7(c) следует, что

$$(X'X)^{-1}X'VX(X'X)^{-1} \geq (X'V^{-1}X)^{-1}.$$

4. (Ошибка GLS.) Покажите, что $\widehat{\beta}_{GLS} - \beta = (X'V^{-1}X)^{-1}X'V^{-1}\varepsilon$.

1.7. Приложение: отдача от масштаба в производстве и распределении электроэнергии

Работа Нерлова [Nerlove, 1963] представляет собой классический пример изучения отдачи от масштаба в регулируемой отрасли. Она также идеальна для применения методов, изученных в этой главе, и ознакомления читателя с некоторыми другими, еще не рассматривавшимися.

Отрасль производства и распределения электроэнергии

В то время, когда Нерлов писал свою статью, отрасль производства и распределения электроэнергии в США характеризовалась следующими особенностями:

(1) Местные частные монополии поставляли электроэнергию по заявкам.

(2) Тарифы (цена электроэнергии) устанавливались Комиссией по коммунальному обслуживанию.

(3) Цены факторов производства (например, ставка заработной платы) устанавливались для фирмы либо на совершенно конкурентном рынке, либо посредством долговременных контрактов с профсоюзами.

Перечисленные особенности будут важны для нас при обсуждении вопроса, правомерно ли использование в таких условиях OLS[1].

Данные

Нерлов использовал кросс-секционные данные по 145 фирмам в 44 штатах за 1955 год. В модели были использованы следующие переменные: общие издержки, цены факторов (зарплата, цена топлива, стоимость аренды капитала) и выпуск. Хотя фирмы обладают своим капиталом (например, электростанциями, оборудованием, сооружениями), в соответствии со стандартной теорией инвестиций [Jorgenson, 1963], фирма должна вести себя (при предположении, что издержек изменения размера

[1] Вследствие дерегулирования отрасли, произошедшего после выхода статьи Нерлова, на одном местном рынке было разрешено конкурировать нескольким фирмам, и строгий контроль за ценами был отменен во многих штатах. То есть первые две особенности в настоящее время не имеют места.

капитала нет) так, как будто она арендует капитал у себя же по рыночной ставке (она называется «стоимость использования капитала»), которая равна $(r + \delta) \cdot p_I$, где r — реальная ставка процента (далее мы будем использовать букву r для обозначения отдачи от масштаба), δ — норма амортизации, p_I — цена капитала. По этой причине к капиталу можно относиться так, как будто его количество фирма может легко изменять (так же как количество используемого труда и топлива).

Приложение B статьи [Nerlove, 1963] содержит тщательное обсуждение источников данных. Данные по выпуску, затратам на топливо и на труд (которые в сумме с затратами капитала составляют общие затраты) были получены от Федеральной комиссии по энергетике (1956). В качестве ставки заработной платы использовалась средняя по штату заработная плата работников коммунальных компаний. В идеале стоимость капитала рассчитывается как произведение восстановительной стоимости на стоимость использования. Однако из-за отсутствия соответствующих данных этот показатель рассчитывался как сумма процентных платежей и амортизационных отчислений, которые были получены из бухгалтерских отчетов.

Для чего нам нужна эконометрика

Зачем нам использовать модные эконометрические модели (например, OLS) для установления отдачи от масштаба? Почему мы не можем быть наивными и просто построить график зависимости средних издержек (которые могут быть легко вычислены из данных как отношение общей суммы затрат к выпуску) против выпуска в плоскости выпуск — издержки и проверить, является ли наклон кривой АС («средних издержек», от англ. *average cost*) отрицательным? Причина состоит в том, что фирмы могут иметь различные кривые АС. Если цены факторов производства не одинаковы для разных фирм, то для фирм с меньшими ценами факторов средние издержки будут ниже. Предположение о том, что все фирмы в выборке сталкиваются с одинаковыми ценами на факторы производства в заданный момент времени, обычно является адекватным предположением, но не в случае с электроэнергетикой США, где цены на ресурсы значительно варьируются от региона к региону. Поэтому нужно каким-то способом изолировать влияние цен факторов на кривую АС. Подход, которым воспользовался Нерлов, ставший привычным в прикладных эконометрических исследованиях, состоит в оценивании параметризованной функции издержек.

Другим фактором, сдвигающим индивидуальную кривую АС, является уровень эффективности производства. Если более эффективные фирмы выпускают больше продукции, то возможно, что индивидуальные кривые АС имеют положительный наклон, в то время как кривая, соединяющая наблюдаемые комбинации средних издержек и выпуска, имеет отрицательный наклон. В качестве иллюстрации этого рассмотрим конкурентную отрасль, изображенную на рис. 1.6, где кривые AC и MC (предельные издержки, от англ. *marginal cost*) изображены для двух фирм, работающих на одном рынке. Чтобы сконцентрироваться на связи эффективности и выпуска, предположим, что цены факторов производства для всех фирм одинаковы. То есть единственная причина различия кривых AC и MC двух фирм — это различие в эффективности производства. AC и MC имеют положительные наклоны, что отражает убывающую отдачу от масштаба. Кривые AC и MC для фирмы A лежат выше кривых фирмы B, так как фирма A менее эффективна, чем B. Поскольку отрасль совершенно конкурентна, обе фирмы продают свой выпуск по одной и той же цене p. Так как выпуск фирмы определяется точкой пересечения кривой MC и рыночной цены, то выпуски и средние издержки двух фирм соответствуют на графике точкам A и B. Кривая, которую мы получим, если соединим эти две точки, имеет отрицательный наклон, давая нам ложную иллюзию *положительной* отдачи от масштаба.

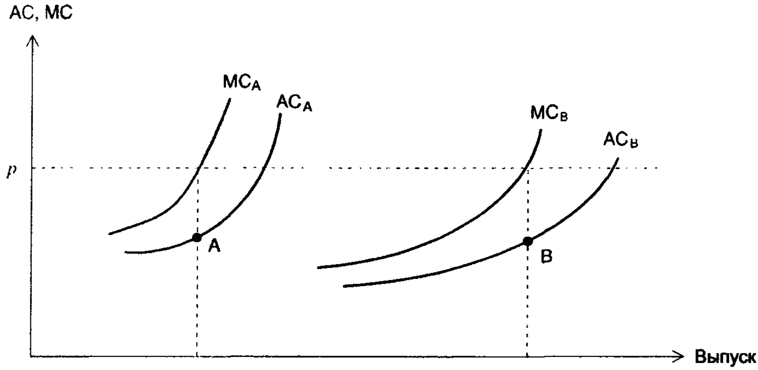

Рис. 1.6. Определение выпуска

Технология Кобба — Дугласа

Чтобы вывести параметризованную функцию издержек, начнем с производственной функции Кобба — Дугласа:

$$Q_i = A_i x_{i1}^{\alpha_1} x_{i2}^{\alpha_2} x_{i3}^{\alpha_3}, \tag{1.7.1}$$

где Q_i — выпуск фирмы i, x_{i1} — количество труда, используемое фирмой i, x_{i2} — капитал, x_{i3} — топливо. Параметр A_i отражает ненаблюдаемые различия в эффективности производства (это часто называется **гетерогенностью фирм** (*firm heterogeneity*)). Сумма коэффициентов $\alpha_1 + \alpha_2 + \alpha_3 \equiv r$ характеризует отдачу от масштаба. Таким образом, мы *априори* предполагаем, что отдача от масштаба постоянна (это не следует путать с постоянной отдачей от масштаба, что означает, что $r = 1$). Так как в выборке фирмы по производству электроэнергии являются частными, логично предположить, что каждая из них решает задачу минимизации издержек (посмотрите, однако, обсуждение в конце этого параграфа). Из микроэкономики нам известно, что функция издержек, соответствующая производственной функции Кобба — Дугласа, также имеет форму Кобба — Дугласа:

$$TC_i = r \cdot (A_i \alpha_1^{\alpha_1} \alpha_2^{\alpha_2} \alpha_3^{\alpha_3})^{-1/r} Q_i^{1/r} p_{i1}^{\alpha_1/r} p_{i2}^{\alpha_2/r} p_{i3}^{\alpha_3/r}, \qquad (1.7.2)$$

где TC_i — общие издержки фирмы i. Беря логарифм от обеих частей, получаем линейное в логарифмах соотношение:

$$\log(TC_i) = \mu_i + \frac{1}{r}\log(Q_i) + \frac{\alpha_1}{r}\log(p_{i1}) + \frac{\alpha_2}{r}\log(p_{i2}) + \frac{\alpha_3}{r}\log(p_{i3}), \quad (1.7.3)$$

где $\mu_i = \log[r \cdot (A_i \alpha_1^{\alpha_1} \alpha_2^{\alpha_2} \alpha_3^{\alpha_3})^{-1/r}]$. Это уравнение называется **логлинейным** (*log-linear*), так как и зависимая, и независимые переменные стоят в логарифмах. Коэффициенты в логлинейном уравнении интерпретируются как **эластичности**. Например, коэффициент при $\log(p_{i1})$ показывает эластичность общих издержек по ставке заработной платы, то есть на сколько процентов изменятся общие издержки при росте заработной платы на 1 процент. Отдача от масштаба, которая в уравнении (1.7.3) равна величине, обратной эластичности выпуска по общим издержкам, не зависит от уровня выпуска.

Пусть $\mu \equiv \mathrm{E}(\mu_i)$ и определим $\varepsilon_i \equiv \mu_i - \mu$, так что $\mathrm{E}(\varepsilon_i) = 0$. Этот ε_i представляет собой величину, обратную к отношению эффективности фирмы к средней эффективности отрасли. Фирмы с положительным ε_i менее эффективны. Применяя эти обозначения, перепишем (1.7.3) как

$$\begin{aligned}\log(TC_i) = \beta_1 + \beta_2\log(Q_i) + \beta_3\log(p_{i1}) + \beta_4\log(p_{i2}) + \\ + \beta_5\log(p_{i3}) + \varepsilon_i.\end{aligned} \qquad (1.7.4)$$

где

$$\beta_1 = \mu, \quad \beta_2 = \frac{1}{r}, \quad \beta_3 = \frac{\alpha_1}{r}, \quad \beta_4 = \frac{\alpha_2}{r} \quad \text{и} \quad \beta_5 = \frac{\alpha_3}{r}. \qquad (1.7.5)$$

Таким образом, мы преобразовали функцию издержек к формату регрессии из предположения 1.1 с $K = 5$. Мы уже отмечали, что наивный метод построения графика зависимости средних издержек от выпуска не может принять во внимание влияние цен факторов. Что мы только что показали, — это то, что в случае производственной функции Кобба — Дугласа влияние цен факторов принимается во внимание посредством включения в функцию издержек логарифмов цен. Так как уравнение выведено из явного предположения о функциональном виде технологии фирм, ошибка и коэффициенты регрессии имеют ясную экономическую интерпретацию.

Откуда мы знаем, что действительно имеем дело с технологией Кобба — Дугласа?

Технология Кобба — Дугласа, конечно, является удобным способом параметризации производственной функции. Но откуда мы знаем, что истинной производственной функцией является именно производственная функция Кобба — Дугласа? Форма Кобба — Дугласа удовлетворяет ряду свойств (например, дает убывающую предельную производительность), выполнения которых мы обычно требуем от производственной функции. В то же время этим свойствам также удовлетворяют многие другие производственные функции. Однако функция вида Кобба — Дугласа, несмотря на свою простоту, на удивление хорошо зарекомендовала себя. Статья Нерлова входит в то небольшое число работ, в которых логлинейная спецификация функции Кобба — Дугласа признана неадекватной. Однако этот факт лишь подчеркивает важность функции Кобба — Дугласа как ориентира, который дает возможность обдумывать подходящие обобщения.

Выполнены ли предположения OLS?

Чтобы обосновать возможность применения метода наименьших квадратов, необходимо удостовериться, что для уравнения (1.7.4) выполнены предположения 1.2–1.4. Очевидно, что предположение 1.1 (линейность) выполнено с

$$y_i = \log(TC_i)\,,\ \boldsymbol{x}_i = (1, \log(Q_i), \log(p_{i1})\,, \log(p_{i2})\,, \log(p_{i3}))'\,.$$

Нет никаких оснований предполагать, что регрессоры в (1.7.4) строго мультиколлинеарны. Кроме того, в данных Нерлова $\operatorname{rank}(\boldsymbol{X}) = 5$ и $n = 145$, следовательно, предположение 1.3 (об отсутствии мультиколлинеарности) также выполняется.

При проверке предположения о строгой экзогенности регрессоров (предположение 1.2) актуальны особенности рынка электроэнергетики, описанные выше. Логично предположить, что, как и в большинстве кросс-секционных данных, $\boldsymbol{x}_i$ не зависит от ε_j для $i \neq j$. Но остается вопрос: зависит ли $\boldsymbol{x}_i$ от ε_i? Если нет, то $\mathrm{E}(\boldsymbol{\varepsilon}|\boldsymbol{X}) = \boldsymbol{0}$. В соответствии с третьей особенностью рынка цены факторов не зависят от эффективности фирмы, поэтому резонно предположить, что цены независимы от ошибок.

А что насчет выпуска? Так как фирмы отпускают электроэнергию по заявкам (первая особенность рынка), выпуск зависит от цены электроэнергии, которая (в свою очередь) устанавливается Комиссией по коммунальному обслуживанию (вторая особенность). Если регулятор устанавливает цену без учета эффективности той или иной фирмы, то тогда $\log(Q_i)$ и ε_i независимы. С другой стороны, если цена устанавливается таким образом, чтобы она покрывала средние издержки, то тогда эффективность фирмы влияет на выпуск через влияние цены электричества на спрос. В таком случае выпуск **эндогенен** (коррелирован с ошибкой). В конце мы очень кратко вернемся к этому вопросу, а до этого будем игнорировать возможную эндогенность выпуска. Это нельзя было бы сделать, если бы мы имели дело с совершенно конкурентной отраслью. Так как фирмы с высокими издержками производят меньше, наблюдалась бы *отрицательная* корреляция между $\log(Q_i)$ и ε_i, что сделало бы OLS неприменимым.

Что касается предположения 1.4, то предположение об отсутствии корреляции между ошибками разных фирм (наблюдений) не выполнялось бы, если бы, например, фирмы перенимали технологические новшества у других фирм, расположенных неподалеку. Для рассматриваемого примера этого эффекта, скорее всего, нет.

Не существует *априорной* причины считать, что условие гомоскедастичности выполнено. Кроме того, график остатков, приведенный ниже[1], свидетельствует в пользу того, что это предположение не выполнено. Большая часть работы Нерлова посвящена рассмотрению способов преодоления этой проблемы.

Метод наименьших квадратов с ограничениями

Уравнение (1.7.4) **сверхидентифицировано** (*overidentified*), в том смысле, что пять его коэффициентов, будучи функциями от четырех параметров технологии ($\alpha_1, \alpha_2, \alpha_3$ и μ), не являются свободными параметрами. Из (1.7.5) легко видеть, что $\beta_1+\beta_2+\beta_3 = 1$ (вспомните: $r \equiv \alpha_1+\alpha_2+\alpha_3$). Это является отражением общего свойства функции издержек: линейность и однородность по факторам производства. Кроме того, умножение общих издержек TC_i и цен факторов производства (p_{i1}, p_{i2}, p_{i3}) на одну и ту же величину не приведет к изменению функции издержек (1.7.4), если и только если $\beta_3 + \beta_5 + \beta_5 = 1$.

Оценивание уравнения методом наименьших квадратов при наложении *априорных* ограничений на вектор коэффициентов называется методом наименьших квадратов с ограничениями. Оно может быть легко реализовано путем получения из исходного уравнения регрессии другого уравнения регрессии, в котором учтены ограничения. В нашем примере, чтобы ввести ограничение однородности $\beta_3 + \beta_5 + \beta_5 = 1$ на функцию издержек, нужно взять любую из цен факторов производства (например,

[1]См. рис. 1.7.

p_{i3}) и вычесть $\log(p_{i3})$ из обеих частей (1.7.4). В результате получаем:

$$\log\left(\frac{TC_i}{p_{i3}}\right) = \beta_1 + \beta_2\log(Q_i) + \beta_3\log\left(\frac{p_{i1}}{p_{i3}}\right) + \beta_4\log\left(\frac{p_{i2}}{p_{i3}}\right) + \varepsilon_i. \quad (1.7.6)$$

Теперь в регрессии осталось четыре коэффициента, на основании которых можно восстановить значения четырех параметров технологии. Оценка OLS с ограничениями коэффициентов $(\beta_1, \ldots, \beta_4)$ представляет собой обычную OLS-оценку коэффициентов уравнения (1.7.6). Оценка OLS с ограничениями коэффициента β_5 вычисляется исходя из оценок $(\beta_1, \ldots, \beta_4)$ и указанного ограничения.

Тестирование однородности функции издержек

Прежде чем переходить к оцениванию модели с ограничением (1.7.6), протестируем ограничение однородности производственной функции $\beta_3 + \beta_4 + \beta_5 = 1$. Для этого сначала оценим модель без ограничений (1.7.4). При использовании данных из статьи Нерлова, которые доступны в печатном виде, получаем следующую OLS-оценку:

$$\begin{aligned}
\log(TC_i) = \underset{(1.8)}{-3.5} &+ \underset{(0.017)}{0.72}\log(Q_i) + \underset{(0.29)}{0.44}\log(p_{i1}) \\
&- \underset{(0.34)}{0.22}\log(p_{i2}) + \underset{(0.10)}{0.43}\log(p_{i3})
\end{aligned}$$

$$R^2 = 0{,}926,\ \text{среднее зависимой переменной} = 1{,}72,$$
$$SER = 0.392,\ SSR = 21.552,\ n = 145. \quad (1.7.7)$$

В скобках приведены стандартные ошибки OLS-оценок коэффициентов. Поскольку $\beta_2 = 1/r$, оценка отдачи от масштаба, рассчитанная на основании OLS-оценки коэффициента β_2, приблизительно равна 1,4 (=1/0,72). OLS-оценка $\beta_4 = \alpha_2/r$ имеет неправильный знак. Как отметил Нерлов, это может быть следствием того, что цена использования капитала p_{i2} измерена плохо. Это может объяснить, почему коэффициент b_4 определен столь неточно (то есть его стандартная ошибка велика по отношению к оценке коэффициента), что гипотеза о его равенстве нулю $\beta_4 = 0$ не отвергается со значением t-отношения, равным $-0.65(= -0.22/0.34)$[1].

Чтобы проверить гипотезу однородности $H_0\colon \beta_3 + \beta_4 + \beta_5 = 1$, мы можем переписать эту гипотезу в форме $\boldsymbol{R\beta} = \boldsymbol{r}$, где $\boldsymbol{R} = (0, 0, 1, 1, 1)$, $\boldsymbol{r} = 1$, а затем использовать формулу (1.4.9) для вычисления F-отношения. Поддерживаемой гипотезой является в данном случае модель

[1]Следствием ошибки в измерении является не только то, что коэффициент при переменной, измеренной с ошибкой, плохо определен, но это может повлиять и на оценки коэффициентов при других переменных. Более подробно эта проблема будет обсуждаться в контексте асимптотической теории для эндогенных регрессоров в главе 3.

без ограничений (1.4.9) (то есть предположения 1.1–1.5, причем уравнение в предположении 1.1 имеет спецификацию (1.7.4)). То есть оценка коэффициентов $\boldsymbol{b}$ и оценка дисперсии $\boldsymbol{b}$ в F-отношении должны быть получены из OLS-оценивания уравнения (1.7.4). Однако можно пойти другим путем: взять формулу (1.4.11) для F-отношения. Модель без ограничений, из которой берется SSR_U, — это (1.7.4), а модель с ограничениями (которая накладывает ограничения нулевой гипотезы на модель без ограничений), из которой получаем SSR_R, — это модель (1.7.6). Ниже представлена OLS-оценка (1.7.6).

$$\begin{aligned}\log\left(\frac{TC_i}{p_{i3}}\right) = \underset{(0,88)}{-4,7} &+ \underset{(0,017)}{0,72}\log(Q_i)\\ &+ \underset{(0,20)}{0,59}\log(p_{i1}/p_{i3}) - \underset{(0.19)}{0,007}\log(p_{i2}/p_{i3})\end{aligned}$$

$$R^2 = 0{,}932,\ \text{среднее зависимой переменной} = -1{,}48,$$
$$SER = 0{,}39,\ SSR = 21{,}640,\ n = 145. \tag{1.7.8}$$

F-тест для гипотезы однородности выполняется следующим образом.

Шаг 1: Используя (1.4.11), можно вычислить F-отношение:

$$\frac{(21{,}640 - 21{,}552)/1}{21{,}552/(145 - 5)} = 0{,}57.$$

Шаг 2: Находим критическое значение. Количество ограничений (уравнений), накладываемых нулевой гипотезой, равно 1, а K (число коэффициентов) в модели без ограничений (которая является поддерживаемой гипотезой) равно 5. То есть нас интересует F-распределение со степенями свободы 1 и 140 ($= 145 - 5$). Из соответствующей таблицы находим, что критическое значение равно приблизительно 3,9.

Шаг 3: Таким образом, нулевая гипотеза однородности не отвергается. Это обнадеживающее заключение для тех, кто всерьез относится к микроэкономической теории (подобно нам).

Лирическое отступление: об осторожности с интерпретацией R^2

Значение R^2, равное 0.926, на удивление велико для кросс-секционных данных. Однако часть объясняющей силы регрессии можно объяснить эффектом масштаба: общие издержки фирмы растут с ростом размера фирмы. Чтобы оценить вклад этого эффекта в R^2, вычтем $\log(Q_i)$

из обеих частей (1.7.4), получая эквивалентную функцию издержек:

$$\log\left(\frac{TC_i}{Q_i}\right) = \beta_1 + (\beta_2 - 1)\log(Q_i) + \\ + \beta_3 \log(p_{i1}) + \beta_4 \log(p_{i2}) + \beta_5 \log(p_{i3}) + \varepsilon_i. \tag{1.7.4'}$$

Теперь в качестве зависимой переменной выступают средние издержки, а не общие, как было ранее. Применение OLS к (1.7.4′) дает для тех же данных следующие результаты:

$$\begin{aligned}\log\left(\frac{TC_i}{Q_i}\right) = \underset{(1{,}8)}{-3{,}5} - \underset{(0{,}017)}{0{,}28}\log(Q_i) + \\ + \underset{(0{,}29)}{0{,}44}\log(p_{i1}) - \underset{(0{,}34)}{0{,}22}\log(p_{i2}) + \underset{(0{,}10)}{0{,}43}\log(p_{i3})\end{aligned}$$

$$R^2 = 0{,}695, \text{ среднее зависимой переменной} = -4{,}83, \\ SER = 0{,}392,\ SSR = 21{,}552,\ n = 145. \tag{1.7.9}$$

Как вы, без сомнения, ожидали, коэффициент при выпуске теперь равен $-0{,}28$ $(= 0{,}72 - 1)$, при этом оценки прочих коэффициентов и стандартных ошибок не изменились. R^2 изменился только из-за изменения зависимой переменной. Было бы глупо утверждать, что вследствие более высокого значения коэффициента детерминации спецификация (1.7.4) предпочтительнее, чем (1.7.4′), так как эти два уравнения представляют одну и ту же модель. Мораль: при сравнении двух переменных на основании R^2 необходимо, чтобы в обоих случаях зависимая переменная была одной и той же.

Тестирование постоянной отдачи от масштаба

В качестве примера применения t-теста рассмотрим тестирование гипотезы о постоянной отдаче от масштаба ($r = 1$). Поддерживаемой гипотезой является модель (1.7.6). Поскольку β_2 (коэффициент при логарифме выпуска) равен 1, если и только если $r = 1$, то нулевая гипотеза может быть сформулирована как H_0: $\beta_2 = 1$. t-тест для проверки гипотезы о постоянстве отдачи от масштаба проводится следующим образом:

Шаг 1: Вычисляем t-отношение для гипотезы. Из оценивания модели с ограничением имеем $b_2 = 0.72$ со стандартной ошибкой, равной 0,017, так что

$$t\text{-отношение} = \frac{0.72 - 1}{0{,}017} = -16.$$

Так как поддерживаемой гипотезой здесь является модель с ограничениями (1.7.6), K (количество коэффициентов) $= 4$.

Шаг 2: Находим критическое значение для распределения Стьюдента $t(151)$. Размер теста равен 5%, критическое значение равно 1,98.

Шаг 3: Поскольку абсолютное значение t-отношения превышает критическое значение, мы отвергаем нулевую гипотезу о постоянной отдаче от масштаба.

Важность графического анализа остатков

У рассматриваемого уравнения есть проблема, о существовании которой нельзя догадаться из анализа регрессионных коэффициентов и их стандартных ошибок. На рис. 1.7 изображены остатки модели (по горизонтальной оси отложена переменная $\log(Q_i)$).

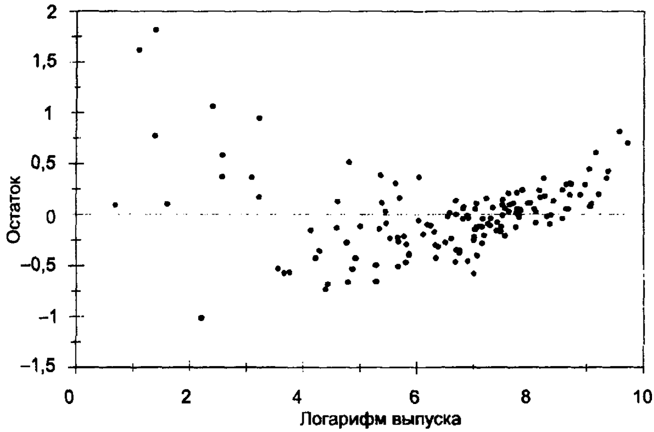

Рис. 1.7. График «логарифм выпуска — остатки»

Обратите внимание на два обстоятельства. Во-первых, с ростом объема выпуска остатки сначала имеют тенденцию быть положительными, затем отрицательными, потом опять положительными. Это, скорее всего, свидетельствует в пользу того, что отдача от масштаба (r) непостоянна, что противоречит логлинейной спецификации. Во-вторых, разброс остатков больше при небольших значениях выпуска, что является признаком нарушения предположения о гомоскедастичности ошибок (дисперсия ошибок не зависит от регрессоров). Чтобы справиться с этими проблемами, Нерлов разделил выборку из 145 фирм на 5 групп по 29 фирм (предварительно упорядочив фирмы по объему выпуска) и оценил уравнение (1.7.6) отдельно для каждой группы. Это позволило коэффициентам (включая $\beta_2 = 1/r$) и дисперсии ошибок быть разными для пяти разных групп фирм (отличающихся по размеру). Нерлов установил, что отдача от масштаба постепенно уменьшается (от значения r, значительно большего 2, до значения, немного меньшего 1) по мере роста выпуска. В эмпирическом упражнении к этой главе читателям будет предложено повторить эти результаты и провести дополнительный анализ, используя **дамми-переменные** (*dummy variables*) и взвешенный метод наименьших квадратов.

Дальнейшая разработка темы

Одним из направлений в последующих исследованиях является обобщение технологии Кобба — Дугласа при сохранении предположения о минимизации издержек. Очевидной альтернативой рассмотренной функции Кобба — Дугласа является производственная **функция с постоянной эластичностью замещения факторов** (*Constant Elasticity of Substitution* — **CES**). Однако ее использование сопряжено с двумя трудностями. Во-первых, функция издержек, получаемая при CES-спецификации производственной функции, нелинейна. Эту проблему можно решить, используя нелинейный метод наименьших квадратов, представленный в главе 7. Во-вторых, из CES-спецификации производственной функции следует постоянная отдача от масштаба. Одним из главных выводов в работе Нерлова было то, что отдача от масштаба изменяется при изменении объема выпуска. [Christensen and Greene, 1976], возможно, были первыми, кто оценивал параметры технологии в предположении изменяющейся отдачи от масштаба. Используя **транслогарифмическую функцию издержек** (*translog cost function*), предложенную [Christensen, Jorgenson and Lau, 1973], они выяснили, что значимый эффект масштаба, очевидный для данных за 1955 год, перестал проявляться к 1970 году, когда большинство фирм производили намного больший объем выпуска при практически плоских кривых AC. Мы обсудим эту работу более подробно в главе 4.

Другой вопрос: действительно ли фирмы в регулируемой отрасли минимизируют издержки? Известная статья [Averch and Johnson, 1962] утверждает, что практика регулирующих органов устанавливать тарифы коммунальных предприятий таким образом, чтобы гарантировать им «справедливую норму доходности» их капитала, нарушает выбор входных уровней. Так как справедливая норма доходности обычно выше рыночной ставки процента, то у коммунальных предприятий появляются стимулы к повышению уровня инвестиций выше оптимального. Иными словами, фирмы минимизируют издержки, однако норма доходности в определении стоимости использования капитала — это справедливая норма доходности. Следовательно, истинный параметр технологии не может быть определен из функции издержек (если, конечно, не использовать справедливую норму доходности для расчета p_{i2}). Однако такой способ назначения тарифов создает еще одну проблему: чтобы обеспечить фирмам справедливую норму доходности, цена на электроэнергию должна быть выше на рынках, где работают предприятия с большими издержками. Поэтому выпуск будет эндогенным.

Наконец, более свежее направление исследований — это изучение вопроса: действительно ли регулятор обладает достаточным объемом информации, чтобы его решение привело к минимизации издержек. Если фирма располагает большим объемом информации об издержках, у нее есть стимулы сообщать регулятору неверное значение своего параметра эффективности. Схемы, которые может использовать регулятор, чтобы предотвратить эту проблему, связанную со стимулированием, могут не приводить к минимизации издержек. Работа [Wolak, 1994], посвященная предприятиям коммунального водоснабжения в Калифорнии, указывает, что наблюдаемый уровень издержек и выпуска лучше моделируется как результат взаимодействия между регулятором и фирмой при асимметричной информации. Волак решает проблему эндогенности выпуска путем оценивания функции спроса одновременно с функцией издержек. Однако для этого необходима более изощренная техника, чем метод наименьших квадратов.

Контрольные вопросы

1. (Повторение теории двойственности.) Откройте ваш любимый учебник по микроэкономике и вспомните, как выводится функция издержек Кобба — Дугласа из производственной функции Кобба — Дугласа.

2. (Изменение единиц измерения.) В работе Нерлова выпуск измеряется в киловатт-часах. Как изменятся оценки коэффициентов в уравнении, если выпуск будет измеряться в мегаватт-часах?

3. (Восстановление параметров технологии на основе регрессионных коэффициентов.) Покажите, что параметры производственной функции $(\mu, \alpha_1, \alpha_2, \alpha_3)$ могут быть однозначно определены из первых четырех уравнений в (1.7.5) и определения $r \equiv \alpha_1 + \alpha_2 + \alpha_3$). (Не используйте пятое уравнение $\beta_5 = \alpha_3/r$.)

4. (Восстановление коэффициентов, исключенных из модели с ограничениями.) Вычислите OLS-оценку коэффициента β_5 из (1.7). Как вычислить стандартную ошибку b_5 на основании оценки модели с ограничениями? **Указание:** *Запишите $b_5 = a + c'b$ для специальным образом выбранных a и c, где $b = (b_1, \dots, b_4)'$, так что $\mathrm{Var}(b_5|X) = c'\,\mathrm{Var}(b|X)c$. Таблица результатов OLS-оценивания с ограничениями должна включать $\widehat{\mathrm{Var}(b|X)}$.*

5. Если вы возьмете p_{i2} вместо p_{i3} и вычтете из обеих частей (1.7.4) $\log(p_{i2})$, как будет выглядеть регрессия с ограничениями в таком случае? Без реального оценивания этой регрессии на данных Нерлова, используя только результат оценивания регрессии с ограничениями в тексте параграфа, можете ли вы сказать, чему будут равны оценки коэффициентов $(\beta_1, \dots, \beta_5)$, получаемые OLS с ограничениями? Чему равны их стандартные ошибки? SSR? R^2?

6. Почему R^2 в модели (1.7.7), равный 0,926, *меньше*, чем в модели (1.7) (где $R^2 = 0{,}932$)?

7. Более реалистичным предположением о цене капитала может быть существование единого рынка капитала на территории всей страны, так что p_{i2} одинакова для всех фирм. В таком случае

   (a) Можем ли мы оценить параметры технологий? **Указание:** *Ответ: да. Но почему? Когда p_{i2} постоянна, в (1.7.4) возникнет проблема полной мультиколлинеарности. Но вспомним, что $(\beta_1, \ldots, \beta_5)$ не являются свободными параметрами.*

   (b) Можем ли мы проверить гипотезу об однородности функции издержек по ценам факторов?

8. Беря логарифмы от обеих частей производственной функции (1.7.1), можно вывести линейное в логарифмах соотношение:

$$\log(Q_i) = \alpha_0 + \alpha_1 \log x_{i1} + \alpha_2 \log x_{i2} + \alpha_3 \log x_{i3} + \varepsilon_i,$$

где ε_i определяется как $\log(A_i) - \mathrm{E}[\log(A_i)]$ и $\alpha_0 = \mathrm{E}[\log(A_i)]$. Предположим, что в дополнение к данным об общих издержках, выпуске и ценах факторов производства в нашем распоряжении есть также данные о количестве использованных факторов производства. Можем ли мы оценить коэффициенты в логлинейном соотношении с помощью метода наименьших квадратов? Почему да или почему нет? **Указание:** *Зависят ли уровни факторов производства от ε_i?*

Предложите альтернативный способ оценивания коэффициентов α. **Указание:** *Посмотрите на доли факторов производства.*

Набор задач для главы 1

Аналитические упражнения

1. (Доказательство того, что b минимизирует SSR.) Обозначим через b OLS-оценку β. Докажите, что для любой гипотетической оценки вектора β, $\widetilde{\beta}$, выполнено

$$(y - X\widetilde{\beta})'(y - X\widetilde{\beta}) \geq (y - Xb)'(y - Xb).$$

При доказательстве используйте стратегию прибавления — вычитания: возьмите разность $y - X\widetilde{\beta}$, прибавьте к ней, а затем вычтете Xb. Тогда получится следующее разложение $y - X\widetilde{\beta}$:

$$y - X\widetilde{\beta} = (y - Xb) + (Xb - X\widetilde{\beta}).$$

Указание: $(y - X\widetilde{\beta})'(y - X\widetilde{\beta}) = [(y - Xb) + X(b - \widetilde{\beta})]'[(y - Xb) + X(b - \widetilde{\beta})]$. *Используя систему нормальных уравнений, покажите, что это равно:*

$$(y - Xb)'(y - Xb) + (b - \widetilde{\beta})'X'X(b - \widetilde{\beta}).$$

2. (Аннулятор для вектора, состоящего из единиц.) Обозначим через $\mathbf{1}$ n-мерный вектор-столбец, состоящий из единиц. Определим $\boldsymbol{M}_1 \equiv \boldsymbol{I}_n - \mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'$. То есть $\boldsymbol{M}_1$ — аннулятор для вектора $\mathbf{1}$. Докажите:

   (a) Матрица $\boldsymbol{M}_1$ симметричная и идемпотентная.

   (b) $\boldsymbol{M}_1\mathbf{1} = \mathbf{0}$.

   (c) $\boldsymbol{M}_1\boldsymbol{y} = \boldsymbol{y} - \overline{y}\cdot\mathbf{1}$, где
$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i.$$
   $\boldsymbol{M}_1\boldsymbol{y}$ — вектор **отклонений от среднего**.

   (d) $\boldsymbol{M}_1\boldsymbol{X} = \boldsymbol{X} - \mathbf{1}\overline{\boldsymbol{x}}'$, где $\overline{\boldsymbol{x}} = \boldsymbol{X}'\mathbf{1}/n$. k-й элемент вектора $\overline{\boldsymbol{x}}$ размерности $K \times 1$ равен $\sum_{i=1}^{n} x_{ik}$.

3. (Модель в отклонениях от среднего.) Рассмотрим модель регрессии с константой. Пусть $\boldsymbol{X}$ разбивается на блоки следующим образом:
$$\underset{(n\times K)}{\boldsymbol{X}} = \begin{bmatrix} \underset{n\times 1}{\mathbf{1}} & \vdots & \underset{n\times(K-1)}{\boldsymbol{X}_2} \end{bmatrix},$$
   то есть первый регрессор — константа. Векторы $\boldsymbol{\beta}$ и $\boldsymbol{b}$ разбиваются соответственно:
$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} \begin{matrix} \leftarrow \quad \text{скаляр} \\ \leftarrow \quad (K-1)\times 1 \end{matrix}, \boldsymbol{b} = \begin{bmatrix} b_1 \\ \boldsymbol{b}_2 \end{bmatrix}.$$
   Обозначим $\widetilde{\boldsymbol{X}}_2 \equiv \boldsymbol{M}_1\boldsymbol{X}_2$ и $\widetilde{\boldsymbol{y}} \equiv \boldsymbol{M}_1\boldsymbol{y}$. Это отклонения от среднего для всех регрессоров, кроме константы, и зависимой переменной. Докажите:

   (a) Система K нормальных уравнений имеет вид:
$$\overline{y} - b_1 - \overline{\boldsymbol{x}}_2'\boldsymbol{b}_2 = 0,$$
   где $\overline{\boldsymbol{x}}_2 = \boldsymbol{X}_2'\mathbf{1}/n$,
$$\boldsymbol{X}_2'\boldsymbol{y} - n\cdot b_1\boldsymbol{x}_2 - \boldsymbol{X}_2'\boldsymbol{X}_2\boldsymbol{b}_2 = \underset{((K-1)\times 1))}{\mathbf{0}}.$$

   (b) $\boldsymbol{b}_2 = (\widetilde{\boldsymbol{X}}_2'\widetilde{\boldsymbol{X}}_2)^{-1}\widetilde{\boldsymbol{X}}_2'\widetilde{\boldsymbol{y}}$. **Указание:** *Подставьте первое уравнение системы в остальные $K-1$, чтобы избавиться от b_1, и найдите $\boldsymbol{b}_2$. Это является обобщением результата, который вы доказали в контрольном вопросе 3 к параграфу 1.2.*

4. (Регрессия с блочной структурой матрицы наблюдений, обобщение упражнения 3.) Пусть $\boldsymbol{X}$ разбита на блоки следующим образом:

$$\underset{(n\times K)}{\boldsymbol{X}} = \left[\underset{n\times K_1}{\boldsymbol{X}_1} \;\vdots\; \underset{n\times K_2}{\boldsymbol{X}_2}\right].$$

Тогда соответствующим разбиением $\boldsymbol{\beta}$ будет:

$$\boldsymbol{\beta} = \begin{bmatrix}\boldsymbol{\beta}_1\\ \boldsymbol{\beta}_2\end{bmatrix} \begin{matrix}\leftarrow & K_1\times 1\\ \leftarrow & K_2\times 1\end{matrix}.$$

Поэтому регрессионную модель можно записать в виде:

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}.$$

Обозначим $\boldsymbol{P}_1 \equiv \boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'$, $\boldsymbol{M}_1 \equiv \boldsymbol{I} - \boldsymbol{P}_1$, $\widetilde{\boldsymbol{X}_2} \equiv \boldsymbol{M}_1\boldsymbol{X}_2$ и $\widetilde{\boldsymbol{y}} \equiv \boldsymbol{M}_1\boldsymbol{y}$. Таким образом, $\widetilde{\boldsymbol{y}}$ представляет собой вектор остатков от регрессии $\boldsymbol{y}$ на $\boldsymbol{X}$, а k-й столбец $\widetilde{\boldsymbol{X}_2}$ — вектор остатков от регрессии соответствующего k-го столбца матрицы $\boldsymbol{X}_2$ на $\boldsymbol{X}_1$. Докажите:

(a) система нормальных уравнений имеет вид:

$$\boldsymbol{X}_1'\boldsymbol{X}_1\boldsymbol{b}_1 + \boldsymbol{X}_1'\boldsymbol{X}_2\boldsymbol{b}_2 = \boldsymbol{X}_1'\boldsymbol{y}, \tag{$*$}$$

$$\boldsymbol{X}_2'\boldsymbol{X}_1\boldsymbol{b}_1 + \boldsymbol{X}_2'\boldsymbol{X}_2\boldsymbol{b}_2 = \boldsymbol{X}_2'\boldsymbol{y}. \tag{$**$}$$

(b) $\boldsymbol{b}_2 = (\widetilde{\boldsymbol{X}_2}'\widetilde{\boldsymbol{X}_2})^{-1}\widetilde{\boldsymbol{X}_2}'\widetilde{\boldsymbol{y}}$. Таким образом, оценка $\boldsymbol{b}_2$ может быть получена из регрессии остатков $\widetilde{\boldsymbol{y}}$ на матрицу остатков $\widetilde{\boldsymbol{X}_2}$. **Указание:** *Выведите* $\boldsymbol{X}_1\boldsymbol{b}_1 = -\boldsymbol{P}_1\boldsymbol{X}_2\boldsymbol{b}_2 + \boldsymbol{P}_1\boldsymbol{y}$ *из* ($*$). *Подставьте это выражение в* ($**$), *чтобы получить* $\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{b}_2 = \boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{y}$. *После этого используйте тот факт, что* $\boldsymbol{M}_1$ *— симметричная и идемпотентная матрица. Или, если пожелаете, можете просто применить формулу (A.10) приложения А к матрице коэффициентов:*

$$\boldsymbol{X}'\boldsymbol{X} = \begin{bmatrix}\boldsymbol{X}_1'\boldsymbol{X}_1 & \boldsymbol{X}_1'\boldsymbol{X}_2\\ \boldsymbol{X}_2'\boldsymbol{X}_1 & \boldsymbol{X}_2'\boldsymbol{X}_2\end{bmatrix}.$$

Покажите, что второй диагональный блок матрицы $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ *равен* $(\widetilde{\boldsymbol{X}_2}'\widetilde{\boldsymbol{X}_2})^{-1}$.

(c) Остатки от регрессии $\widetilde{\boldsymbol{y}}$ на $\widetilde{\boldsymbol{X}_2}$ численно равны $\boldsymbol{e}$, остаткам от регрессии $\boldsymbol{y}$ на $\boldsymbol{X}$ ($\equiv (\boldsymbol{X}_1 \vdots \boldsymbol{X}_2)$). **Указание:** *Если* $\boldsymbol{e}$ *— остаток от регрессии* $\boldsymbol{y}$ *на* $\boldsymbol{X}$, *то*

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{b}_1 + \boldsymbol{X}_2\boldsymbol{b}_2 + \boldsymbol{e}.$$

Умножая слева обе части на $\boldsymbol{M}_1$ *и используя* $\boldsymbol{M}_1\boldsymbol{X}_1 = \boldsymbol{0}$, *получаем*

$$\widetilde{\boldsymbol{y}} = \widetilde{\boldsymbol{X}}_2\boldsymbol{b}_2 + \boldsymbol{M}_1\boldsymbol{e}.$$

Покажите, что $\boldsymbol{M}_1\boldsymbol{e} = \boldsymbol{e}$, *и вы увидите, что* $\boldsymbol{b}_2$ *равен OLS-оценке коэффициента в регрессии* $\widetilde{\boldsymbol{y}}$ *на* $\widetilde{\boldsymbol{X}}_2$.

(d) $\boldsymbol{b}_2 = (\widetilde{\boldsymbol{X}}_2'\widetilde{\boldsymbol{X}}_2)^{-1}\widetilde{\boldsymbol{X}}_2'\boldsymbol{y}$. Обратите внимание на отличие от пункта (b). Здесь в роли зависимой переменной выступает $\boldsymbol{y}$, а не $\widetilde{\boldsymbol{y}}$. Совпадают ли численно остатки от регрессии $\boldsymbol{y}$ на $\widetilde{\boldsymbol{X}}_2$ с $\boldsymbol{e}$? [Ответ: нет.] Равна ли SSR регрессии $\boldsymbol{y}$ на $\widetilde{\boldsymbol{X}}_2$ сумме квадратов остатков регрессии $\widetilde{\boldsymbol{y}}$ на $\widetilde{\boldsymbol{X}}_2$? [Ответ: Нет.]

Результаты пунктов (b)–(d) носят название **теоремы Фриша — Во — Ловелла**.

(e) Покажите:

$$\widetilde{\boldsymbol{y}}'\widetilde{\boldsymbol{y}} - \boldsymbol{e}'\boldsymbol{e} = \widetilde{\boldsymbol{y}}'\boldsymbol{X}_2(\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{X}_2)^{-1}\boldsymbol{X}_2'\widetilde{\boldsymbol{y}}.$$

Указание: *Примените общую формулу декомпозиции* (1.2.15) *к регрессии в части (c), чтобы получить*

$$\widetilde{\boldsymbol{y}}'\widetilde{\boldsymbol{y}} = \boldsymbol{b}_2'\widetilde{\boldsymbol{X}}_2'\widetilde{\boldsymbol{X}}_2\boldsymbol{b}_2 + \boldsymbol{e}'\boldsymbol{e}.$$

Затем используйте пункт (b).

(f) Рассмотрите следующие четыре регрессии:

(1) $\widetilde{\boldsymbol{y}}$ на $\boldsymbol{X}_1$.

(2) $\widetilde{\boldsymbol{y}}$ на $\widetilde{\boldsymbol{X}}_2$.

(3) $\widetilde{\boldsymbol{y}}$ на $\boldsymbol{X}_1$ и $\boldsymbol{X}_2$.

(4) $\widetilde{\boldsymbol{y}}$ на $\boldsymbol{X}_2$.

Обозначим через SSR_j сумму квадратов остатков в регрессии j. Покажите:

(i) $SSR_1 = \widetilde{\boldsymbol{y}}'\widetilde{\boldsymbol{y}}$. **Указание:** $\widetilde{\boldsymbol{y}}$ *построен таким образом, что* $\boldsymbol{X}_1'\widetilde{\boldsymbol{y}} = \boldsymbol{0}$, *то есть* $\boldsymbol{X}_1$ *не имеет объясняющей силы.*

(ii) $SSR_2 = \boldsymbol{e}'\boldsymbol{e}$. **Указание:** *Используйте (c).*

(iii) $SSR_3 = \boldsymbol{e}'\boldsymbol{e}$. **Указание:** *Примените теорему Фриша — Во — Ловелла к регрессии (3).*

$$\boldsymbol{M}_1\widetilde{\boldsymbol{y}} = \widetilde{\boldsymbol{y}}.$$

(iv) Проверьте на численном примере, что SSR_4 необязательно равна $\boldsymbol{e}'\boldsymbol{e}$.

5. (Регрессия с ограничениями и F.) При использовании метода наименьших квадратов с ограничениями сумма квадратов остатков минимизируется при условии ограничений, накладываемых нулевой гипотезой $\boldsymbol{R\beta} = \boldsymbol{r}$. Запишем Лагранжиан

$$\mathcal{L} = \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}) + \boldsymbol{\lambda}'(\boldsymbol{R}\widetilde{\boldsymbol{\beta}} - \boldsymbol{r}).$$

где λ — $\#\mathbf{r}$-мерный вектор множителей Лагранжа (вспомните: вектор $\boldsymbol{R}$ размера $\#\mathbf{r}\times K$, $\widetilde{\boldsymbol{\beta}}$ размера $K \times 1$, $\boldsymbol{r}$ размера $\#\mathbf{r}\times 1$). Обозначим через $\widehat{\boldsymbol{\beta}}$ OLS-оценку параметра $\boldsymbol{\beta}$ в регрессии с ограничениями. Она является решением задачи минимизации с ограничениями.

(a) Пусть $\boldsymbol{b}$ — OLS-оценка без ограничений. Покажите:

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{b} - (\boldsymbol{X}')^{-1}\boldsymbol{R}'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{Rb} - \boldsymbol{r}),$$
$$\boldsymbol{\lambda} = [\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{Rb} - \boldsymbol{r}).$$

Указание: *Условия первого порядка:* $\boldsymbol{X}'\boldsymbol{y} - (\boldsymbol{X}'\boldsymbol{X})\widehat{\boldsymbol{\beta}} = \boldsymbol{R}'\boldsymbol{\lambda}$, *или* $\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) = \boldsymbol{R}'\boldsymbol{\lambda}$. *Добавьте к этому ограничение* $\boldsymbol{R}\widehat{\boldsymbol{\beta}} = \boldsymbol{r}$ *и решите относительно* $\boldsymbol{\lambda}$ *и* $\widehat{\boldsymbol{\beta}}$.

(b) Обозначим остатки от регрессии с ограничением через $\widehat{\boldsymbol{\varepsilon}} \equiv \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}$. Покажите:

$$\begin{aligned} SSR_R - SSR_U &= (\boldsymbol{b} - \widehat{\boldsymbol{\beta}})'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{b} - \widehat{\boldsymbol{\beta}}) = \\ &= (\boldsymbol{Rb} - \boldsymbol{r})'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{Rb} - \boldsymbol{r}) = \\ &= \boldsymbol{\lambda}'\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\boldsymbol{\lambda} = \\ &= \widehat{\boldsymbol{\varepsilon}}'\boldsymbol{P}\widehat{\boldsymbol{\varepsilon}}. \end{aligned}$$

где $\boldsymbol{P}$ — проекционная матрица. **Указание:** *Для получения первого равенства используйте стратегию прибавления — вычитания:*

$$\begin{aligned} SSR_R &= (\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) = \\ &= [(\boldsymbol{y} - \boldsymbol{Xb}) + \boldsymbol{X}(\boldsymbol{b} - \widehat{\boldsymbol{\beta}})]'[(\boldsymbol{y} - \boldsymbol{Xb}) + \boldsymbol{X}(\boldsymbol{b} - \widehat{\boldsymbol{\beta}})]. \end{aligned}$$

Используйте систему нормальных уравнений $\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{Xb}) = \boldsymbol{0}$. *Для второго и третьего равенств используйте (a). Для доказательства четвертого равенства самым простым путем является использование условия первого порядка, упомянутого в (a):* $\boldsymbol{R}'\boldsymbol{\lambda} = \boldsymbol{X}'\widehat{\boldsymbol{\varepsilon}}$.

(c) Проверьте, что в пункте (b) вы доказали, что (1.4.9) = (1.4.11).

6. (Доказательство декомпозиции (1.2.17).) Пусть модель без ограничений — регрессия, в которой, помимо постоянной составляющей имеются и другие регрессоры, а модель с ограничениями — регрессия, в которой постоянная составляющая является единственным регрессором.

   (a) Покажите, что (b) в предыдущем упражнении представляет собой декомпозицию (1.2.17) для этого случая. **Указание:** *Что такое $\widehat{\boldsymbol{\beta}}$ для этого случая? Покажите, что $SSR_R = \sum_i (y_i - \overline{y})^2$, и $(\boldsymbol{b} - \widehat{\boldsymbol{\beta}})'(\boldsymbol{X}'\boldsymbol{X})(\boldsymbol{b} - \widehat{\boldsymbol{\beta}}) = \sum_i (\widehat{y}_i - \overline{y})^2$.*

   (b) (R^2 и F-отношение.) Для модели, в которой константа является одним из регрессоров, докажите, что

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

7. (Принцип Хаусмана в конечных выборках.) Для обобщенной регрессионной модели докажите следующее. Здесь подразумевается, что математические ожидания, дисперсии и ковариации являются условными относительно $\boldsymbol{X}$.

   (a) $\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{GLS}, \boldsymbol{b} - \widehat{\boldsymbol{\beta}}_{GLS}) = \boldsymbol{0}$. **Указание:** *Вспомните, что для любых двух случайных векторов $\boldsymbol{x}$ и $\boldsymbol{y}$*

$$\mathrm{Cov}(\boldsymbol{x}, \boldsymbol{y}) \equiv \mathrm{E}\left[(\boldsymbol{x} - \mathrm{E}(\boldsymbol{x}))(\boldsymbol{y} - \mathrm{E}(\boldsymbol{y}))\right].$$

   Поэтому

$$\mathrm{Cov}(\boldsymbol{A}\boldsymbol{x}, \boldsymbol{B}\boldsymbol{y}) = \boldsymbol{A}\,\mathrm{Cov}(\boldsymbol{x}, \boldsymbol{y})\boldsymbol{B}'.$$

   Кроме того, так как $\boldsymbol{\beta}$ неслучайный, то

$$\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{GLS}, \boldsymbol{b} - \widehat{\boldsymbol{\beta}}_{GLS}) = \mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{GLS} - \boldsymbol{\beta}, \boldsymbol{b} - \widehat{\boldsymbol{\beta}}_{GLS}).$$

   (b) Пусть $\widetilde{\boldsymbol{\beta}}$ — линейная по $\boldsymbol{y}$ и несмещенная оценка, и определим $\boldsymbol{q} \equiv \widetilde{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}_{GLS}$. Пусть $\widetilde{\boldsymbol{\beta}}$ определена так, что матрица $\boldsymbol{V_q} \equiv \mathrm{Var}(\boldsymbol{q})$ невырождена. Докажите: $\mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{GLS}, \boldsymbol{q}) = \boldsymbol{0}$. (Если мы положим $\widetilde{\boldsymbol{\beta}} = \boldsymbol{b}$, то мы вернемся в пункт (a).) **Указание:** *Определим $\widehat{\boldsymbol{\beta}} \equiv \widehat{\boldsymbol{\beta}}_{GLS} + \boldsymbol{H}\boldsymbol{q}$ для некоторой матрицы $\boldsymbol{H}$. Покажите:*

$$\mathrm{Var}(\widehat{\boldsymbol{\beta}}) = \mathrm{Var}(\widehat{\boldsymbol{\beta}}_{GLS}) + \boldsymbol{C}\boldsymbol{H}' + \boldsymbol{H}\boldsymbol{C}' + \boldsymbol{H}\boldsymbol{V_q}\boldsymbol{H}',$$

   где $\boldsymbol{C} \equiv \mathrm{Cov}(\widehat{\boldsymbol{\beta}}_{GLS}, \boldsymbol{q})$. Покажите, что если $\boldsymbol{C} \neq \boldsymbol{0}$, то $\mathrm{Var}(\widehat{\boldsymbol{\beta}})$ может быть меньше, чем $\mathrm{Var}(\widehat{\boldsymbol{\beta}}_{GLS})$, если положить $\boldsymbol{H} = -\boldsymbol{C}\boldsymbol{V}_q^{-1}$. Наконец, укажите, что это противоречит утверждению 1.7(c).

(c) (Дополнительное, только для читателей с хорошим знанием линейной алгебры.) Докажите: если K столбцов $\boldsymbol{X}$ являются собственными векторами $\boldsymbol{V}$, где $\boldsymbol{V}$ — $n \times n$ ковариационная матрица n-мерного вектора ошибок ε, то тогда $\boldsymbol{b} = \widehat{\beta}_{GLS}$. (То есть не все несмещенные линейные оценки удовлетворяют требованию невырожденности матрицы $\mathrm{Var}(\widetilde{\beta} - \widehat{\beta}_{GLS})$, предъявленному в пункте (b).) **Указание:** *Для любой симметричной матрицы $\boldsymbol{V}$ размера $n \times n$ существует матрица $\boldsymbol{H}$ размера $n \times n$ такая, что $\boldsymbol{H}'\boldsymbol{H} = \boldsymbol{I}_n$ (то есть $\boldsymbol{H}$ — ортогональная матрица), и $\boldsymbol{H}'\boldsymbol{V}\boldsymbol{H} = \boldsymbol{\Lambda}$, где $\boldsymbol{\Lambda}$ — диагональная матрица, диагональные элементы которой являются собственными значениями $\boldsymbol{V}$ (которые являются натуральными числами, поскольку матрица $\boldsymbol{V}$ симметрична). Столбцы $\boldsymbol{H}$ называются собственными векторами $\boldsymbol{H}$. Покажите, что*

$$\boldsymbol{H}^{-1} = \boldsymbol{H}', \quad \boldsymbol{H}'\boldsymbol{V}^{-1}\boldsymbol{H} = \boldsymbol{\Lambda}^{-1}, \quad \boldsymbol{H}'\boldsymbol{V}^{-1} = \boldsymbol{\Lambda}^{-1}\boldsymbol{H}'.$$

Без потери общности, в качестве $\boldsymbol{X}$ можно взять первые K столбцов матрицы $\boldsymbol{H}$. То есть $\boldsymbol{X} = \boldsymbol{H}\boldsymbol{F}$, где

$$\underset{(n \times K)}{\boldsymbol{F}} = \begin{bmatrix} \boldsymbol{I}_K \\ \boldsymbol{0} \end{bmatrix}.$$

Эмпирические упражнения

Прежде чем приступить к этому упражнению, прочитайте статью Марка Нерлова (Marc Nerlove) «Returns to Scale in Electricity Supply» (*отдача от масштаба в производстве и распределении электроэнергии*) (кроме параграфов с уравнениями (6)–(9), части параграфа 2 со с. 184 и далее и приложений A и C). Файл NERLOVE.ASC содержит следующую информацию по 145 электроэнергетическим компаниям за 1955 год:

Столбец 1: общие издержки (TC) в миллионах долларов.

Столбец 2: выпуск (Q) в миллиардах киловатт-часов.

Столбец 3: цена труда (PL).

Столбец 4: цена топлива (PF).

Столбец 5: цена капитала (PK).

Данные взяты из приложения статьи. Данные содержат 145 наблюдений, которые упорядочены по размеру: первой идет самая маленькая компания, 145-й — самая большая. Используя возможности программного обеспечения, имеющегося на вашем компьютере, создайте для каждой из 145 компаний переменные, необходимые для оценивания модели. Чтобы оценить, например, уравнение (1.7.4), вам понадобятся следующие переменные: $\log(TC)$, константа, $\log(Q)$, $\log(PL)$, $\log(PK)$ и $\log(PF)$ для каждой из 145 фирм.

(a) (Вопрос относительно данных.) Соответствует ли цена капитала (как она использована Нерловым) определению стоимости использования капитала? **Указание:** *Прочитайте приложение B.4 статьи Нерлова.*

(b) Оцените модель без ограничений (1.7.4) методом наименьших квадратов. Совпали ли ваши результаты с оценками, приведенными в тексте?

(c) (Метод наименьших квадратов с ограничениями.) Оцените модель с ограничениями (1.7.6) методом наименьших квадратов. Чтобы сделать это, необходимо создать дополнительные переменные для всех фирм в выборке. Например, зависимой переменной теперь будет $\log(TC/PF)$, а не $\log(TC)$, как раньше. Совпали ли ваши результаты с оценками, приведенными в тексте? Можете ли вы повторить результаты Нерлова? Например, его оценка коэффициента β_2 равна 0,721 со стандартной ошибкой, равной 0,0174 (в статье написано, что стандартная ошибка равна 0,175, но это, скорее всего, опечатка). Где именно в статье Нерлова находятся эти результаты? А как насчет остальных коэффициентов? (Предупреждение: у вас не получится повторить результаты Нерлова в точности. Одной из причин является то, что он использовал не натуральные, а десятичные логарифмы, однако это влияет только на оценку константы. Другая причина — для его результатов использовалась исправленная версия данных, опубликованных в статье.)

Как уже отмечалось в тексте главы, график остатков свидетельствует в пользу существования нелинейной связи между $\log(TC)$ и $\log(Q)$. Нерлов предположил, что отдача от масштаба изменяется с размером фирмы. Следуя за ним, разбейте выборку из 145 фирм на группы по 29. (Вспомните, что данные уже отсортированы по возрастанию размера фирмы, поэтому первые 29 наблюдений представляют группу фирм с наименьшим уровнем выпуска, последние — с наибольшим.) Рассмотрите следующие три обобщения модели (1.7.6):

Модель 1: И коэффициенты (β), и дисперсии ошибок в (1.7.6) различны для разных групп.

Модель 2: Коэффициенты различны, но дисперсии ошибок одинаковы.

Модель 3: Коэффициенты β_3 и β_4 (ценовые эластичности) и дисперсия ошибок одинаковы для всех групп, а коэффициент β_2 и константа разные. Нерлов назвал модель 3 «нейтральной вариацией в отдаче от масштаба».

Для модели 1 с различными для разных групп коэффициентами и дисперсиями ошибок оценки могут быть найдены из:

$$\boldsymbol{y}^{(j)} = \boldsymbol{X}^{(j)}\boldsymbol{\beta}^{(j)} + \boldsymbol{\varepsilon}^{(j)} \quad (j = 1, \ldots, 5).$$

где $\boldsymbol{y}^{(j)}$ (29×1) — вектор значений зависимой переменной для группы j, $\boldsymbol{X}^{(j)}$ (29×4) — матрица значений четырех регрессоров для группы j, $\boldsymbol{\beta}^{(j)}$ (4×1) — вектор коэффициентов группы j, $\boldsymbol{\varepsilon}^{(j)}$ (29×1) — вектор ошибок. Например, второй столбец $\boldsymbol{X}^{(5)}$ равен $\log(Q)$ для $i=117,\ldots,145$. Модель 1 предполагает выполнение требования условной гомоскедастичности остатков: $\mathrm{E}(\varepsilon^{(j)}\varepsilon^{(j)\prime}|\boldsymbol{X}^{(j)}) = \sigma_j^2\boldsymbol{I}_{29}$ внутри каждой группы (но необязательно между группами).

(d) Оцените модель 1 методом наименьших квадратов. Совпадают ли ваши результаты с оценками Нерлова? На основе вашей оценки коэффициента β_2 рассчитайте точечные оценки отдачи от масштаба в каждой из пяти групп. Каков общий характер зависимости оцененной отдачи от масштаба от уровня выпуска? Как в общих чертах изменяется оцененная дисперсия ошибок при росте выпуска?

Модель 2 накладывает на модель 1 дополнительное предположение: $\sigma_j^2=\sigma^2$ для всех j. Это ограничение, уравнивающее дисперсии, можно инкорпорировать в модель, используя штабелирование векторов и матриц и записывая модель в следующем виде:

$$\boldsymbol{y}=\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon},$$

где

$$\underset{(145\times1)}{\boldsymbol{y}}=\begin{bmatrix}\boldsymbol{y}^{(1)}\\ \vdots\\ \boldsymbol{y}^{(5)}\end{bmatrix},\quad \underset{(145\times20)}{\boldsymbol{X}}=\begin{bmatrix}\boldsymbol{X}^{(1)} & & \\ & \ddots & \\ & & \boldsymbol{X}^{(5)}\end{bmatrix},\quad \underset{(145\times1)}{\boldsymbol{\varepsilon}}=\begin{bmatrix}\boldsymbol{\varepsilon}^{(1)}\\ \vdots\\ \boldsymbol{\varepsilon}^{(5)}\end{bmatrix}. \qquad (*)$$

В частности, матрица $\boldsymbol{X}$ теперь является блочно-диагональной. Предположение об одинаковой дисперсии можно сформулировать таким образом: $\mathrm{E}(\varepsilon\varepsilon'|\boldsymbol{X})=\sigma^2\boldsymbol{I}_{145}$. Теперь в нашей модели 20 переменных, которые были построены на основании изначальных четырех регрессоров. Например, первыми 29 элементами 145-мерного вектора, соответствующего второй переменной, являются $\log(Q_1),\ldots,\log(Q_{29})$, а на остальных местах стоят нули. На 30–58-х местах вектора, соответствующего шестой переменной (логарифм выпуска второй группы фирм), стоят $\log(Q_{30}),\ldots,\log(Q_{58})$, а на остальных местах стоят нули. И так далее.

Операция штабелирования, необходимая для формирования $\boldsymbol{y}$ и $\boldsymbol{X}$ в уравнении $(*)$, может быть легко реализована, если программный пакет, которым вы пользуетесь, поддерживает операции с матрицами. В противном случае вам придется схитрить и добиться от программы того же конечного результата с помощью дамми-переменных. Определим j-ю дамми-переменную как

$$D_{ij} = \begin{cases} 1 & \text{если } i\text{-я фирма принадлежит } j\text{-й группе,} \\ 0 & \text{в противном случае} \end{cases} \quad (i = 1, \ldots, 145).$$

Тогда второй регрессор можно построить как $D_{1i} \cdot \log(Q_i)$. Шестой — $D_{i2} \cdot \log(Q_i)$, и так далее.

(e) Оцените модель 2 методом наименьших квадратов. Проверьте, что OLS-оценки, полученные в этом пункте, совпадают с оценками пункта (d). Также проверьте, что

$$\sum_{j=1}^{5} SSR_j = SSR,$$

где SSR_j — это SSR, полученная при оценивании модели 1 в (d) для j-й группы, а SSR — это SSR в модели 2. Такое равенство неслучайно (то есть не является специфической чертой имеющегося набора данных). Докажите, что упомянутые выше тождества для коэффициентов и SSR выполняются вообще (без потери общности, рассмотрите случай двух групп). **Указание:** *Сначала покажите, что коэффициенты модели 1 и модели 2 совпадают. Используйте формулы (A.4), (A.5) и (A.9) приложения A.*

(f) (Тест Чоу.) Модель 2 является более общей, чем модель (1.7.6), так как коэффициенты могут различаться для разных групп. Проверьте нулевую гипотезу о том, что коэффициенты одинаковы для всех групп. Сколько уравнений (ограничений) в этой гипотезе? Этот тест иногда называют **тестом Чоу на наличие структурного сдвига** (*Chow test for structural change*). Рассчитайте p-значение для F-отношения. **Указание:** *Это линейная гипотеза относительно коэффициентов модели 2. То есть модель 2 — поддерживаемая гипотеза, а* (1.7.6) *— модель с ограничениями. Используйте формулу* (1.4.11) *для F-отношения.*

***Совет при работе в пакете* Gauss:** Если обозначить за x F-отношение, то команда **Gauss** `cdffc(x,df1,df2)` выдает площадь под кривой плотности справа от x для F-распределения со степенями свободы $df1$ и $df2$.

Совет при работе в пакете TSP: Аналогичная TSP-команда имеет вид: `cdf(f, df1=df1, df2=df2) x`. Результат применения команды `OLSQ`, реализующей OLS, выдаваемый как `@SSR`, представляет SSR регрессии.

Совет при работе в пакете RATS: В RATS команда выглядит так: `cdf ftest x df1 df2`. RSS регрессии может быть получена после использования команды OLS, `LINREG`, с помощью `%RSS`.

Ограничение в модели 3, предполагающее, что ценовые эластичности одинаковы внутри групп фирм, можно ввести в модель 2, применяя преобразование, использующее дамми-переменные только для константы и логарифма выпуска. Таким образом, в новую модель включаются $12(= 2\times 5+2)$ переменных, причем матрица $\boldsymbol{X}$ выглядит так:

$$\boldsymbol{X} = \begin{bmatrix} 1\ \log(Q_1) & 0 & 0 & \log(PL_1/PF_1) & \log(PK_1/PF_1) \\ \vdots\quad \vdots & \vdots & \vdots & \vdots & \vdots \\ 1\ \log(Q_{29}) & 0 & 0 & \log(PL_{29}/PF_{29}) & \log(PK_{29}/PF_{29}) \\ & \ddots & & \vdots & \vdots \\ 0 \quad 0 & 1 & \log(Q_{117}) & \log(PL_{117}/PF_{117}) & \log(PK_{117}/PF_{117}) \\ \vdots\quad \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 \quad 0 & 1 & \log(Q_{145}) & \log(PL_{145}/PF_{145}) & \log(PK_{145}/PF_{145}) \end{bmatrix}. \tag{**}$$

(g) Оцените модель 3. Эта модель является частным случаем модели 2, соответствующим гипотезе о том, что две ценовые эластичности одинаковы для всех пяти групп. Протестируйте эту гипотезу на 5%-ном уровне значимости, предполагая нормальность ошибок. (Замечание: F-отношение на с. 183 статьи Нерлова неверно.)

Как мы уже установили из графического анализа остатков на рис. 1.7, условный момент второго порядка $\mathrm{E}(\varepsilon_i^2|\boldsymbol{X})$, скорее всего, зависит от логарифма выпуска, что является нарушением предположения об условной гомоскедастичности остатков. Сейчас мы не будем пытаться тестировать условную гомоскедастичность, так как для этого необходима асимптотическая теория, которая будет рассмотрена в следующей главе. Вместо этого мы будем делать вид, что нам известна функциональная форма зависимости условного момента второго порядка от логарифма выпуска. Эта функция, спецификация которой приведена ниже, такова, что условный момент второго порядка непрерывно зависит от логарифма выпуска, в отличие от трех моделей, рассмотренных ранее. Еще одно отличие от предыдущих моделей состоит в том, что теперь мы предполагаем, что отдача от масштаба изменяется непрерывно с изменением выпуска. Чтобы добиться этого, добавим в модель квадрат логарифма выпуска[1]. Тогда модель 4 имеет следующий вид:

Модель 4:

$$\log\left(\frac{TC_i}{p_{i3}}\right) = \beta_1 + \beta_2\log(Q_i) + \beta_3[\log(Q_i)]^2 + \\ + \beta_4\log\left(\frac{p_{i1}}{p_{i3}}\right) + \beta_5\log\left(\frac{p_{i2}}{p_{i3}}\right) + \varepsilon_i.$$

$$\mathrm{E}(\varepsilon_i^2|\boldsymbol{X}) = \sigma^2\cdot\left(0{,}0565 + \frac{2{,}1377}{Q_i}\right) \quad (i = 1, 2, \ldots, 145)$$

для некоторого неизвестного значения σ^2.

(h) Оцените модель 4 с помощью взвешенного метода наименьших квадратов для всей выборки из 145 фирм. (Будьте внимательны в отношении константы: после проведения процедуры взвешивания ни один из регрессоров не будет постоянным.) Нарисуйте график остатков. Свидетельствует ли он об условной гомоскедастичности ошибок или, как и прежде, о наличии нелинейности?

Упражнения на метод Монте-Карло

Метод **Монте-Карло** (*Monte Carlo*) состоит в генерации большого количества выборок на основе некоторой модели, чтобы изучить распределение оценок в конечных выборках. Здесь мы будем использовать этот метод для того, чтобы подтвердить два результата для малых выборок, упоминавшихся в тексте главы: несмещенность OLS-оценки коэффициентов и распределение t-отношения. Для этого рассмотрим модель простой линейной регрессии, удовлетворяющую предположениям 1.1–1.5, с количеством наблюдений $n = 32$. Регрессионные уравнения выглядят следующим образом:

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i \quad (i = 1, 2, \ldots, n)$$
$$\text{или } \boldsymbol{y} = \mathbf{1}\cdot\beta_1 + \boldsymbol{x}\cdot\beta_2 + \varepsilon = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}. \qquad (*)$$

где $\boldsymbol{X} = (\mathbf{1}\,\vdots\,\boldsymbol{x})$ и $\boldsymbol{\beta} = (\beta_1, \beta_2)'$. Параметрами модели являются $(\beta_1, \beta_2, \sigma^2)$.

[1]Мы вывели линейную в логарифмах функцию издержек из производственной функции Кобба – Дугласа. Существует ли такая производственная функция, из которой можно было бы вывести обобщенную функцию издержек, содержащую квадрат логарифма выпуска? Это вопрос об «интегрируемости» («*integrability*») функций издержек, подробное обсуждение которого можно найти в: [Christensen et al., 1973].

Как уже упоминалось в тексте, модель представляет собой набор совместных распределений $(\boldsymbol{y}, \boldsymbol{X})$. Из этого набора мы выберем конкретное совместное распределение, определяя регрессионную модель следующим образом. Пусть $\beta_1 = 1$, $\beta_2 = 0{,}5$ и $\sigma^2 = 1$. Распределение $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)'$ зададим следующим AR(1)-процессом:

$$x_i = c + \phi x_{i-1} + \eta_i \quad (i = 1, 2, \ldots, n), \tag{**}$$

где $\{\eta_i\} \sim i.i.d. N(0, 1)$, и

$$x_0 \sim N\left(\frac{c}{1-\phi}, \frac{1}{1-\phi^2}\right), \quad c = 2, \quad \phi = 0{,}6.$$

Таким образом, задается совместное распределение $(\boldsymbol{y}, \boldsymbol{X})$. Именно из этого распределения будет извлекаться большое количество выборок.

Для осуществления симуляций с помощью компьютера будет полезным представление $\boldsymbol{x}$ в следующем виде. Решая разностное уравнение первого порядка (**), получаем:

$$\begin{aligned} x_i = &\phi^i x_0 + (1 + \phi + \phi^2 + \ldots + \phi^{i-1})c + \\ &+ (\eta_i + \phi\eta_{i-1} + \phi^2\eta_{i-2} + \ldots + \phi^{i-1}\eta_1), \end{aligned}$$

или, используя матричные обозначения,

$$\underset{(n\times 1)}{\boldsymbol{x}} = \underset{(n\times 1)}{\boldsymbol{r}} \cdot x_0 + \underset{(n\times 1)}{\boldsymbol{d}} + \underset{(n\times n)}{\boldsymbol{A}}\underset{(n\times 1)}{\boldsymbol{\eta}}, \tag{***}$$

где $\boldsymbol{d} = (d_1, d_2, \ldots, d_n)')$ и

$$d_1 = c, \quad d_2 = (1+\phi)c, \ldots, \quad d_i = (1 + \phi + \phi^2 + \ldots + \phi^{i-1})c, \ldots,$$

$$\boldsymbol{r} = \begin{bmatrix} \phi \\ \phi^2 \\ \vdots \\ \phi^n \end{bmatrix}, \quad \boldsymbol{A} = \begin{bmatrix} 1 & 0 & \ldots & \ldots & 0 \\ \phi & 1 & 0 & \ldots & 0 \\ \phi^2 & \phi & 1 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \phi^{n-1} & \phi^{n-2} & \ldots & \phi & 1 \end{bmatrix}, \quad \boldsymbol{\eta} = \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{bmatrix}.$$

***Совет при работе с пакетом* Gauss:** Для построения матрицы $\boldsymbol{r}$ используйте команду `seqm`. Для получения матрицы $\boldsymbol{A}$ используйте `toeplitz` и `lowmat`.

(a) Проведите две симуляции методом Монте-Карло. Первая симуляция служит для вычисления $\mathrm{E}(\boldsymbol{b}|\boldsymbol{x})$ и распределения t-отношения как распределения, условного относительно $\boldsymbol{x}$. Компьютерная программа для этой первой симуляции должна состоять из следующих шагов.

(1) (Сгенерируйте $\boldsymbol{x}$ только один раз.) Используя генератор случайных чисел, создайте вектор $\boldsymbol{\eta}$, состоящий из реализаций n i.i.d. $N(0, 1)$ случайных величин, и выберите x_0 из $N(c/(1-\phi), 1/(1-\phi^2))$, рассчитайте $\boldsymbol{x}$ по формуле (***). ($\boldsymbol{x}$ также может быть рассчитан рекурсивно из (**) с использованием цикла, однако операции с векторами, подобные (***), требуют меньшего процессорного времени, чем циклы. Это замечание будет актуальным во второй симуляции, где $\boldsymbol{x}$ надо будет генерировать при каждом повторении.)

(2) Создайте переменную-счетчик и положите ее равной нулю. В этой переменной будет записываться количество случаев, когда $|t| > t_{0.025}(n-2)$. Кроме того, создайте двумерный вектор, первоначально состоящий из нулей; этот вектор будет использоваться для вычисления математического ожидания OLS-оценки $\boldsymbol{b}$ для $(\beta_1, \beta_2)'$.

(3) Начните цикл для большого числа повторений (например, для миллиона). В каждом повторении выполните следующие действия:

  (i) (Создайте $\boldsymbol{y}$.) Сгенерируйте n-мерный вектор $\boldsymbol{\varepsilon}$, состоящий из реализаций n i.i.d. случайных величин $N(0, 1)$, вычислите $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$ по формуле (*). Теперь этот $\boldsymbol{y}$ вместе с $\boldsymbol{x}$ из шага (1) составляют нашу выборку $(\boldsymbol{y}, \boldsymbol{x})$.

  (ii) По имеющейся выборке вычислите OLS-оценку $\boldsymbol{b}$ и t-отношение для H_0: $\beta_2 = 0{,}5$.

  (iii) Увеличьте значение счетчика на 1, если $|t| > t_{0.025}(n-2)$. Прибавьте $\boldsymbol{b}$ к двумерному вектору.

(4) После окончания цикла разделите значение переменной-счетчика на количество повторений, чтобы вычислить частоту отвержения нулевой гипотезы. Также разделите двумерный вектор, равный сумме OLS-оценок $\boldsymbol{b}$ для всех повторений, на число повторений. Результат деления должен стать равным $\mathrm{E}(\boldsymbol{b}|\boldsymbol{x})$ при бесконечном количестве повторений.

Заметьте, что $\boldsymbol{x}$ в этой первой симуляции генерируется только один раз и остается *неизменным* в процессе получения всех значений $\boldsymbol{y}$. Во второй симуляции вычисляется *без*условное распределение t-отношения. Это выполняется в несколько этапов:

(1) Создайте переменную-счетчик, положите ее равной нулю.

(2) Начните цикл для большого числа повторений (например, для миллиона). В каждом повторении выполните следующие действия:

  (i) (Создайте $\boldsymbol{x}$.) Создайте вектор $\boldsymbol{\eta}$, состоящий из реализаций n i.i.d. $N(0, 1)$ случайных величин, и выберите x_0 из $N(c/(1-\phi), 1/(1-\phi^2))$, рассчитайте $\boldsymbol{x}$ по формуле (***).

(ii) (Создайте $\boldsymbol{y}$.) Сгенерируйте n-мерный вектор ε, состоящий из реализаций n i.i.d. случайных величин $N(0,1)$, вычислите $\boldsymbol{y} = (y_1, \dots, y_n)'$ по формуле $(*)$.

(iii) По имеющейся выборке $(\boldsymbol{y}, \boldsymbol{x})$ вычислите t-отношение для $H_0\colon \beta_2 = 0{,}5$.

(iv) Увеличьте значение счетчика на 1, если $|t| > t_{0{,}025}(n-2)$.

(3) После окончания цикла разделите значение переменной-счетчика на количество повторений.

Для этих двух симуляций проверьте, что для достаточно большого количества наблюдений:

1. Среднее OLS-оценок $\boldsymbol{b}$ из первого эксперимента неограниченно приближается к истинному значению $(1,\ 0{,}5)$;
2. Частота отвержения нулевой гипотезы H_0 (ошибка первого рода) неограниченно приближается к 5% в обоих экспериментах.

(b) Является ли регрессор (не являющийся константой) строго экзогенным в этих двух симуляциях? Является ли ошибка условно гомоскедастичной?

Ответы на избранные вопросы

Аналитические упражнения

1. $(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}) =$

$= [(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}) + \boldsymbol{X}(\boldsymbol{b} - \widetilde{\boldsymbol{\beta}})]'[(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}) + \boldsymbol{X}(\boldsymbol{b} - \widetilde{\boldsymbol{\beta}})]$

(прибавление и вычитание одинаковой величины)

$= [(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})' + (\boldsymbol{b} - \widetilde{\boldsymbol{\beta}})'\boldsymbol{X}'][(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}) + \boldsymbol{X}(\boldsymbol{b} - \widetilde{\boldsymbol{\beta}})] =$

$= (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}) + (\boldsymbol{b} - \widetilde{\boldsymbol{\beta}})'\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}) +$

$+ (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})'\boldsymbol{X}(\boldsymbol{b} - \widetilde{\boldsymbol{\beta}}) + (\boldsymbol{b} - \widetilde{\boldsymbol{\beta}})'\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{b} - \widetilde{\boldsymbol{\beta}}) =$

$= (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}) + 2(\boldsymbol{b} - \widetilde{\boldsymbol{\beta}})\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}) +$

$+ (\boldsymbol{b} - \widetilde{\boldsymbol{\beta}})'\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{b} - \widetilde{\boldsymbol{\beta}})$

(так как $(\boldsymbol{b} - \widetilde{\boldsymbol{\beta}})'\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}) = (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})'\boldsymbol{X}(\boldsymbol{b} - \widetilde{\boldsymbol{\beta}})$)

$= (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}) + (\boldsymbol{b} - \widetilde{\boldsymbol{\beta}})'\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{b} - \widetilde{\boldsymbol{\beta}})$

(поскольку $\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b}) = \boldsymbol{0}$ из системы нормальных уравнений)

$\geq (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{b})$

(так как $(\boldsymbol{b} - \widetilde{\boldsymbol{\beta}})'\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{b} - \widetilde{\boldsymbol{\beta}}) = \boldsymbol{z}'\boldsymbol{z} = \sum\limits_{i=1}^{n} z_i^2 \geq 0$, где $\boldsymbol{z} \equiv \boldsymbol{X}(\boldsymbol{b} - \widetilde{\boldsymbol{\beta}})$).

7a. $\widehat{\beta}_{GLS} - \beta = A\varepsilon$, где $A \equiv (X'V^{-1}X)^{-1}X'V^{-1}$ и $b - \widehat{\beta}_{GLS} = B\varepsilon$, где $B \equiv (X'X)^{-1}X' - (X'V^{-1}X)^{-1}X'V^{-1}$. Поэтому

$$\begin{aligned}&\mathrm{Cov}(\widehat{\beta}_{GLS} - \beta, b - \widehat{\beta}_{GLS}) =\\ &= \mathrm{Cov}(A\varepsilon, B\varepsilon) =\\ &= A\,\mathrm{Var}(\varepsilon)B' =\\ &= \sigma^2 AVB'.\end{aligned}$$

Равенство $AVB' = 0$ доказывается напрямую.

7b. Для матрицы H, определенной в указании к этому пункту,

$$\mathrm{Var}(\widehat{\beta}) - \mathrm{Var}(\widehat{\beta}_{GLS}) = -CV_q^{-1}C'.$$

Если $C \neq 0$, то тогда существует ненулевой вектор z, такой, что $C'z \equiv v \neq 0$. Для такого z

$$z'\left[\mathrm{Var}(\widehat{\beta}) - \mathrm{Var}(\widehat{\beta}_{GLS})\right]z = -v'V_q^{-1}v < 0$$

(поскольку матрица V_q положительно определена), получаем противоречие, так как $\widehat{\beta}_{GLS}$ — эффективная оценка.

Эмпирические упражнения

(a) Из описания Нерлова в приложении B.4 следует, что он не использовал норму амортизации δ для расчета цены капитала.

(b) Ваши оценки должны совпасть с (1.7.7).

(c) Наши оценки немного отличаются от оценок Нерлова. Они отличались бы, даже если бы данные, которыми пользовался Нерлов, совпадали с теми, которые есть у вас. Причина состоит в том, что компьютеры в то время производили не такие точные вычисления и допускали больше ошибок округления.

(d) Насколько точно вы можете повторить приведенные у Нерлова результаты? Достаточно точно. Точечные оценки отдачи от масштаба в каждой из пяти групп равны 2,5, 1,5, 1,1, 1,1 и 0,96. При увеличении выпуска отдача от масштаба снижается.

(e) Модель 2 может быть записана в виде $y = X\beta + \varepsilon$, где y, X и ε определены так же, как и в $(*)$. Поэтому (полагая $j = 2$)

$$X'X = \begin{bmatrix} X^{(1)\prime}X^{(1)} & 0 \\ 0 & X^{(2)\prime}X^{(2)} \end{bmatrix}.$$

откуда следует

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{bmatrix} (\boldsymbol{X}^{(1)\prime}\boldsymbol{X}^{(1)})^{-1} & \mathbf{0} \\ \mathbf{0} & (\boldsymbol{X}^{(2)\prime}\boldsymbol{X}^{(2)})^{-1} \end{bmatrix}.$$

Кроме того

$$\boldsymbol{X}'\boldsymbol{y} = \begin{bmatrix} \boldsymbol{X}^{(1)\prime}\boldsymbol{y}^{(1)} \\ \boldsymbol{X}^{(2)\prime}\boldsymbol{y}^{(2)} \end{bmatrix}.$$

Поэтому

$$(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = \begin{bmatrix} (\boldsymbol{X}^{(1)\prime}\boldsymbol{X}^{(1)})^{-1}\boldsymbol{X}^{(1)\prime}\boldsymbol{y}^{(1)} \\ (\boldsymbol{X}^{(2)\prime}\boldsymbol{X}^{(2)})^{-1}\boldsymbol{X}^{(2)\prime}\boldsymbol{y}^{(2)} \end{bmatrix}.$$

Таким образом, вектор OLS-оценок коэффициентов модели 2 такой же, как и в модели 1. Суммы квадратов остатков также совпадают, вследствие равенства оценок коэффициентов.

(f) Количество ограничений равно 16. K = количество коэффициентов в модели 2 = 20. То есть число степеней свободы должно равняться (16, 125). SSR_U = 12,262 и SSR_R = 21,640. F-отношение = 5,97, p-значение = = 0,0000. Поэтому нулевая гипотеза отвергается на любом разумном уровне значимости.

(g) SSR_U = 12,262 и SSR_R = 12,577. Тогда F = 0,40 с 8 и 125 степенями свободы. p-значение равно 0,92. Нулевая гипотеза не отвергается на любом разумном уровне значимости. F-отношение в работе Нерлова (см. с. 183, 8-я строка снизу) равно 1,576.

(h) График остатков все равно свидетельствует о том, что условный момент второго порядка больше для меньших фирм, однако теперь нет намека на его нелинейность.

Литература

Amemiya, T., 1985, *Advanced Econometrics*, Cambridge: Harvard University Press.

Averch, H., and L. Johnson, 1962, “Behavior of the Firm under Regulatory Constraint,” *American Economic Review*, 52, 1052–1069.

Christensen, L., and W. Greene, 1976, "Economies of Scale in US Electric Power Generation," *Journal of Political Economy*, 84, 655–676.

Christensen, L., D. Jorgenson, and L. Lau, 1973, "Transcendental Logarithmic Production Frontiers," *Review of Economics and Statistics*, 55, 28–45.

Davidson, R., and J. MacKinnon, 1993, *Estimation and Inference in Econometrics*, Oxford: Oxford University Press.

DeLong, B., and L. Summers, 1991, “Equipment Investment and Growth,” *Quarterly Journal of Economics*, 99, 28–45.

Engle, R., D. Hendry, and J.-F. Richard, 1983, "Exogeneity," *Econometrica*, 51, 277–304.

Federal Power Commission, 1956, *Statistics of Electric Utilities in the United States, 1955, Class A and B Privately Owned Companies*, Washington, D.C.

Jorgenson, D., 1963, "Capital Theory and Investment Behavior," *American Economic Review*, 53, 247–259.

Koopmans, T., and W. Hood, 1953, "The Estimation of Simultaneous Linear Economic Relationships," in W. Hood, and T. Koopmans (eds.), *Studies in Econometric Method*, New Haven: Yale University Press.

Krasker, W., E. Kuh, and R. Welsch, 1983, "Estimation for Dirty Data and Flawed Models," Chapter 11 in Z. Griliches, and M. Intriligator (eds.), *Handbook of Econometrics*, Volume 1, Amsterdam: North-Holland.

Nerlove, M., 1963, "Returns to Scale in Electricity Supply," in C. Christ (ed.), *Measurement in Economics: Studies in Mathematical Economics and Econometrics in Memory of Yehuda Grunfeld*, Stanford: Stanford University Press.

Rao, C. R., 1973, *Linear Statistical Inference and Its Applications* (2d ed.), New York: Wiley.

Scheffe, H., 1959, *The Analysis of Variance*, New York: Wiley.

Wolak, F., 1994, "An Econometric Analysis of the Asymmetric Information, Regulator-Utility Interaction," *Annales D'Economie et de Statistique*, 34, 13–69.

Глава 2. Теория больших выборок

В предыдущей главе мы получили точные распределения, или распределения на конечных выборках, OLS-оценки и соответствующих ей тестовых статистик. Однако в экономике предположения о точном распределении выполняются не очень часто. Теория конечных выборок теряет силу, если нарушается одно из трех следующих предположений: 1) экзогенность регрессоров, 2) нормальность ошибок и 3) линейность уравнения регрессии. В настоящей главе разрабатывается альтернативный подход, в котором сохраняется только третье предположение. При таком подходе, называемом **асимптотической теорией**, или **теорией больших выборок**, аппроксимацию распределения оценки и соответствующих ей статистик получают в предположении достаточно большого объема выборки.

Вместо предположений о выборке заданного объема в теории больших выборок делаются предположения о случайном процессе, порождающем эту выборку. В первых двух параграфах этой главы вводится язык, необходимый для описания случайных процессов.

Понятия, введенные в данной главе, имеют важное значение в эконометрике рациональных ожиданий, что было проиллюстрировано в классической работе Фамы о гипотезе Фишера, согласно которой реальная процентная ставка постоянна. Эта гипотеза имеет очень серьезные последствия для политики: монетарная и фискальная политика не могут влиять на агрегированный спрос посредством реальной процентной ставки. Весьма удивительно, что никто не может отвергнуть данную гипотезу для Соединенных Штатов (по крайней мере, если период выборки заканчивается в начале 1970-х).

2.1. Обзор предельных теорем для последовательностей случайных величин

Материал этого параграфа касается предельного поведения последовательности случайных величин, $(z_1, z_2, \ldots)$. Поскольку этот материал может быть уже известен вам, мы представляем его достаточно формально, рядом определений и теорем. Авторитетным источником является книга Рао [Rao, 1973, Chapter 2c], в которой приведены доказательства всех теорем, включенных в данный параграф. В этом параграфе и в оставшейся части книги последовательность $(z_1, z_2, \ldots)$ будет обозначаться как $\{z_n\}$.

Различные виды сходимости

Сходимость по вероятности

Последовательность случайных скаляров $\{z_n\}$ **сходится по вероятности** (*converges in probability*) к константе (неслучайной) α, если для любого $\varepsilon > 0$

$$\lim_{n\to\infty} \mathrm{Prob}(|z_n - \alpha| > \varepsilon) = 0. \tag{2.1.1}$$

Константа α называется **пределом по вероятности** (*probability limit*) z_n, и это записывается как «$\mathrm{plim}_{n\to\infty} z_n = \alpha$» или «$z_n \underset{p}{\to} \alpha$». Очевидно, что

$$«z_n \underset{p}{\to} \alpha» \quad \text{то же самое, что и} \quad «z_n - \alpha \underset{p}{\to} 0».$$

Это определение сходимости по вероятности обобщается на последовательность случайных векторов или случайных матриц (матрица рассматривается как вектор, элементы которого переставлены) требованием поэлементной сходимости по вероятности. То есть последовательность K-мерных случайных векторов $\{\mathbf{z}_n\}$ сходится по вероятности к K-мерному вектору констант $\boldsymbol{\alpha}$, если для любого $\varepsilon > 0$

$$\lim_{n\to\infty} \mathrm{Prob}(|z_{nk} - \alpha_k| > \varepsilon) = 0 \ \text{ для всех } k \ (= 1, 2, \ldots, K). \tag{2.1.2}$$

где z_{nk} есть k-й элемент $\mathbf{z}_n$ и α_k есть k-й элемент $\boldsymbol{\alpha}$.

Сходимость почти наверное

Последовательность случайных скаляров $\{z_n\}$ **сходится почти наверное** (*converges almost surely*) к константе α, если

$$\mathrm{Prob}\Big(\lim_{n\to\infty} z_n = \alpha\Big) = 1. \tag{2.1.3}$$

Мы записываем это как «$z_n \to_{\text{a.s.}} \alpha$». Обобщение на случайные векторы аналогично обобщению сходимости по вероятности. Как будет упомянуто ниже, это понятие сходимости более строгое, чем сходимость по вероятности; то есть если последовательность сходится почти наверное, то она сходится и по вероятности. Понятие, включенное в (2.1.3), осознать сложнее, поскольку указанная вероятность относится к событию, касающемуся бесконечной последовательности $(z_1, z_2, \ldots)$. Для наших целей, однако все, что имеет значение, — это то, что сходимость почти наверное более строгая, чем сходимость по вероятности. Если мы можем показать, что последовательность сходится почти наверное, то это один из способов доказать, что последовательность сходится по вероятности.

Сходимость в среднеквадратичном

Последовательность случайных скаляров $\{z_n\}$ **сходится в среднеквадратичном** (*converges in mean square* или *in quadratic mean*) к α (записывается как «$z_n \to_{\text{m.s.}} \alpha$»), если

$$\lim_{n\to\infty} \mathrm{E}[(z_n - \alpha)^2] = 0. \tag{2.1.4}$$

Обобщение на случайные векторы аналогично обобщению сходимости по вероятности: $\boldsymbol{z}_n \to_{\text{m.s.}} \boldsymbol{\alpha}$, если каждый элемент $\boldsymbol{z}_n$ сходится в среднеквадратичном к соответствующей компоненте $\boldsymbol{\alpha}$.

Сходимость к случайной величине

В приведенных определениях сходимости предел является константой (то есть действительным числом). Предел может быть случайной величиной. Мы говорим, что последовательность K-мерных случайных величин $\{\boldsymbol{z}_n\}$ сходится к K-мерной случайной величине $\boldsymbol{z}$, и пишем $\boldsymbol{z}_n \to_{\text{p}} \boldsymbol{z}$, если $\{\boldsymbol{z}_n - \boldsymbol{z}\}$ сходится к $\mathbf{0}$:

$$«\boldsymbol{z}_n \underset{\text{p}}{\to} \boldsymbol{z}» \quad \text{то же самое, что и} \quad «\boldsymbol{z}_n - \boldsymbol{z} \underset{\text{p}}{\to} \mathbf{0}». \tag{2.1.5a}$$

Аналогично,

$$«\boldsymbol{z}_n \underset{\text{a.s.}}{\to} \boldsymbol{z}» \quad \text{то же самое, что и} \quad «\boldsymbol{z}_n - \boldsymbol{z} \underset{\text{a.s.}}{\to} \mathbf{0}». \tag{2.1.5b}$$

$$«\boldsymbol{z}_n \underset{\text{m.s.}}{\to} \boldsymbol{z}» \quad \text{то же самое, что и} \quad «\boldsymbol{z}_n - \boldsymbol{z} \underset{\text{m.s.}}{\to} \mathbf{0}». \tag{2.1.5c}$$

Сходимость по распределению

Пусть $\{z_n\}$ — последовательность случайных скаляров, а F_n — кумулятивная функция распределения (c.d.f.) z_n. Мы говорим, что $\{z_n\}$ **сходится по распределению** (*converges in distribution*) к случайному скаляру z, если c.d.f. F_n от z_n сходится к c.d.f. F от z в каждой точке непрерывности F[1]. Мы пишем «$z_n \to_{\mathrm{d}} z$» или «$z_n \to_{\mathrm{L}} z$» и называем F **асимптотическим** (*asymptotic*) (или **предельным** (*limit* или *limiting*)) **распределением** z_n. Иногда мы пишем «$z_n \to_{\mathrm{d}} F$», когда распределение F хорошо известно. Например, «$z_n \to_{\mathrm{d}} N(0,1)$» следует читать так: «$z_n \to_{\mathrm{d}} z$ и распределением z является $N(0,1)$ (нормальное распределение со средним 0 и дисперсией 1)». Пользуясь определением, можно показать, что сходимость по вероятности является более строгой, чем сходимость по распределению, то есть

$$«z_n \underset{\mathrm{p}}{\to} z» \Rightarrow «z_n \underset{\mathrm{d}}{\to} z». \tag{2.1.6}$$

В частном случае сходимости по распределению z является константой (тривиальная случайная величина).

Обобщение на последовательность случайных векторов прямое: $\mathbf{z}_n \to_{\mathrm{d}} \mathbf{z}$, если совместная c.d.f. F_n случайного вектора $\mathbf{z}_n$ сходится к совместной c.d.f. F случайного вектора $\mathbf{z}$ в каждой точке непрерывности F. Заметим, однако, что, в отличие от других понятий сходимости, в случае сходимости по распределению поэлементная сходимость необязательно означает сходимость для последовательности векторов. То есть из того, что «каждый элемент $\mathbf{z}_n \to_{\mathrm{d}}$ к соответствующему элементу $\mathbf{z}$» не обязательно вытекает, что «$\mathbf{z}_n \to_{\mathrm{d}} \mathbf{z}$». Распространенным способом установить связь между сходимостью скаляров и сходимостью векторов по распределению является

Теорема о многомерной сходимости по распределению: *(изложенная в [Rao, 1973, p. 128]). Пусть $\{\mathbf{z}_n\}$ — последовательность K-мерных случайных векторов. Тогда:*

$$«\mathbf{z}_n \underset{d}{\to} \mathbf{z}» \Leftrightarrow «\boldsymbol{\lambda}'\mathbf{z}_n \underset{d}{\to} \boldsymbol{\lambda}'\mathbf{z}$$

для любого K-мерного вектора действительных чисел $\boldsymbol{\lambda}'$».

Соотношение между сходимостью по распределению и сходимостью моментов

Стоит подчеркнуть, что моменты предельного распределения z_n необязательно равны пределам моментов z_n. Например, из «$z_n \to_{\mathrm{d}} z$» необязательно вытекает «$\lim_{n\to\infty} \mathrm{E}(z_n) = \mathrm{E}(z)$». Однако

[1]Не обращайте внимания на определение «в каждой точке непрерывности F». По большей части, за исключением, возможно, глав о дискретной или ограниченной зависимой переменной, соответствующее распределение является непрерывным, и функция распределения непрерывна во всех точках.

Лемма 2.1 (сходимость по распределению и сходимость моментов): *Пусть α_{sn} является s-м моментом z_n и $\lim_{n\to\infty}\alpha_{sn} = \alpha_s$, где α_s конечное (то есть действительное число). Предположим, что для некоторого $\delta > 0$ $\mathrm{E}(|z_n|^{s+\delta}) < M < \infty$ для всех n. Тогда:*

$$«z_n \underset{d}{\to} z» \Rightarrow «\alpha_s \text{ есть } s\text{-й момент } z».$$

Таким образом, например, если дисперсия последовательности случайных величин, сходящихся по распределению, сходится к некоторому *конечному* числу, то тогда это число является дисперсией предельного распределения.

Соотношение между видами сходимости

Некоторые виды сходимости являются более слабыми, чем другие. Следующая теорема устанавливает соотношение между четырьмя видами сходимости.

Лемма 2.2 (соотношение между четырьмя видами сходимости):

(a) $«\boldsymbol{z}_n \to_{m.s.} \boldsymbol{\alpha}» \Rightarrow «\boldsymbol{z}_n \to_p \boldsymbol{\alpha}»$. Аналогично $«\boldsymbol{z}_n \to_{m.s.} \boldsymbol{z}» \Rightarrow «\boldsymbol{z}_n \to_p \boldsymbol{z}»$.

(b) $«\boldsymbol{z}_n \to_{a.s.} \boldsymbol{\alpha}» \Rightarrow «\boldsymbol{z}_n \to_p \boldsymbol{\alpha}»$. Аналогично $«\boldsymbol{z}_n \to_{a.s.} \boldsymbol{z}» \Rightarrow «\boldsymbol{z}_n \to_p \boldsymbol{z}»$.

(c) $«\boldsymbol{z}_n \to_p \boldsymbol{\alpha}» \Leftrightarrow «\boldsymbol{z}_n \to_d \boldsymbol{\alpha}»$. (То есть, если предельная случайная величина является константой [тривиальной случайной величиной], сходимость по распределению равносильна сходимости по вероятности.)

Три полезных результата

Определив виды сходимости, мы можем установить три результата, важных для разработки теории больших выборок.

Лемма 2.3 (сохранение сходимости для непрерывных преобразований): *Предположим, что $\boldsymbol{a}(\cdot)$ — векторнозначная непрерывная функция[1], которая не зависит от n.*

(a) $«\boldsymbol{z}_n \to_p \boldsymbol{\alpha}» \Rightarrow «\boldsymbol{a}(\boldsymbol{z}_n) \to_p \boldsymbol{a}(\boldsymbol{\alpha})»$. Иными словами,

$$\operatorname*{plim}_{n\to\infty} \boldsymbol{a}(\boldsymbol{z}_n) = \boldsymbol{a}(\operatorname*{plim}_{n\to\infty} \boldsymbol{z}_n)$$

при условии существования plim.

[1]Часть (a) требует непрерывности только в $\boldsymbol{\alpha}$, в то время как часть (b) требует непрерывности всюду. — *Прим. пер.*

(b) «$\boldsymbol{z}_n \to_d \boldsymbol{z}$» $\Rightarrow$ «$\boldsymbol{a}(\boldsymbol{z}_n) \to_d \boldsymbol{a}(\boldsymbol{z})$».

Немедленным следствием Леммы 2.3(a) является то, что обычные арифметические операции сохраняют сходимость по вероятности. Например:

«$x_n \underset{\mathrm{p}}{\to} \beta$, $y_n \underset{\mathrm{p}}{\to} \gamma$» $\Rightarrow$ «$x_n + y_n \underset{\mathrm{p}}{\to} \beta + \gamma$».

«$x_n \underset{\mathrm{p}}{\to} \beta$, $y_n \underset{\mathrm{p}}{\to} \gamma$» $\Rightarrow$ «$x_n y_n \underset{\mathrm{p}}{\to} \beta\gamma$».

«$x_n \underset{\mathrm{p}}{\to} \beta\, y_n \underset{\mathrm{p}}{\to} \gamma$» $\Rightarrow$ «$x_n / y_n \underset{\mathrm{p}}{\to} \beta/\gamma$» при условии, что $\gamma \neq 0$.

«$\boldsymbol{Y}_n \underset{\mathrm{p}}{\to} \boldsymbol{\Gamma}$» $\Rightarrow$ «$\boldsymbol{Y}_n^{-1} \underset{\mathrm{p}}{\to} \boldsymbol{\Gamma}^{-1}$»

при условии, что $\boldsymbol{\Gamma}$ является обратимой.

Следующий результат о комбинациях сходимости по вероятности и сходимости по распределению будет неоднократно использоваться для получения асимптотического распределения оценки.

Лемма 2.4:

(a) «$\boldsymbol{x}_n \to_d \boldsymbol{x}$, $\boldsymbol{y}_n \to_p \boldsymbol{\alpha}$» $\Rightarrow$ «$\boldsymbol{x}_n + \boldsymbol{y}_n \to_d \boldsymbol{x} + \boldsymbol{\alpha}$».

(b) «$\boldsymbol{x}_n \to_d \boldsymbol{x}$, $\boldsymbol{y}_n \to_p \mathbf{0}$» $\Rightarrow$ «$\boldsymbol{y}_n' \boldsymbol{x}_n \to_p 0$».

(c) «$\boldsymbol{x}_n \to_d \boldsymbol{x}$, $\boldsymbol{A}_n \to_p \boldsymbol{A}$» $\Rightarrow$ «$\boldsymbol{A}_n \boldsymbol{x}_n \to_d \boldsymbol{A}\boldsymbol{x}$» *при условии, что размеры* $\boldsymbol{A}_n$ *и* $\boldsymbol{x}_n$ *согласованны. В частности, если* $\boldsymbol{x} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$, *то* $\boldsymbol{A}_n \boldsymbol{x}_n \to_d N(\mathbf{0}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}')$.

(d) «$\boldsymbol{x}_n \to_d \boldsymbol{x}$, $\boldsymbol{A}_n \to_p \boldsymbol{A}$» $\Rightarrow$ «$\boldsymbol{x}_n' \boldsymbol{A}_n^{-1} \boldsymbol{x}_n \to_d \boldsymbol{x}' \boldsymbol{A}^{-1} \boldsymbol{x}$» *при условии, что размеры* $\boldsymbol{A}_n$ *и* $\boldsymbol{x}_n$ *согласованны, и* $\boldsymbol{A}$ *невырождена.*

Части (a) и (c) иногда называют **теоремой Слуцкого** (*Slutzky's Theorem*). Если положить $\boldsymbol{\alpha} = 0$, то из части (a) вытекает:

$$«\boldsymbol{x}_n \underset{\mathrm{d}}{\to} \boldsymbol{x}, \boldsymbol{y}_n \underset{\mathrm{p}}{\to} \mathbf{0}». \Rightarrow «\boldsymbol{x}_n + \boldsymbol{y}_n \underset{\mathrm{d}}{\to} \boldsymbol{x}». \tag{2.1.7}$$

То есть если $\boldsymbol{z}_n = \boldsymbol{x}_n + \boldsymbol{y}_n$ и $\boldsymbol{y}_n \to_{\mathrm{p}} \mathbf{0}$ (то есть если $\boldsymbol{z}_n - \boldsymbol{x}_n \to_{\mathrm{d}} \mathbf{0}$), то тогда асимптотическое распределение $\boldsymbol{z}_n$ то же самое, что и у $\boldsymbol{x}_n$. Если $\boldsymbol{z}_n - \boldsymbol{x}_n \to_{\mathrm{d}} \mathbf{0}$, мы иногда (но не всегда) говорим, что эти две последовательности **асимптотически эквивалентны** (*asymptotically equivalent*) и записываем это как

$$«\boldsymbol{z}_n \underset{\mathrm{a}}{\sim} \boldsymbol{x}_n» \text{ или } «\boldsymbol{z}_n = \boldsymbol{x}_n + o_p».$$

где o_p — некоторая подходящая случайная величина (здесь $\boldsymbol{y}_n$), которая сходится к нулю по вероятности.

Стандартный трюк получения асимптотического распределения последовательности случайных величин заключается в нахождении асимптотически эквивалентной последовательности, асимптотическое распределение которой получить легче. В частности, заменяя в части (b) леммы $\boldsymbol{y}_n$ на $\boldsymbol{y}_n - \boldsymbol{\alpha}$, мы получаем:

$$«\boldsymbol{x}_n \underset{\mathrm{d}}{\rightarrow} \boldsymbol{x},\ \boldsymbol{y}_n \underset{\mathrm{p}}{\rightarrow} \boldsymbol{\alpha}» \Rightarrow «\boldsymbol{y}_n'\boldsymbol{x}_n \underset{\mathrm{a}}{\sim} \boldsymbol{\alpha}'\boldsymbol{x}_n» \text{ или } «\boldsymbol{y}_n'\boldsymbol{x}_n = \boldsymbol{\alpha}'\boldsymbol{x}_n + o_p» \quad (2.1.8)$$

Здесь o_p есть $(\boldsymbol{y}_n - \boldsymbol{\alpha})'\boldsymbol{x}_n$. Следовательно, замена $\boldsymbol{y}_n$ на его предел по вероятности не изменяет асимптотическое распределение $\boldsymbol{y}_n'\boldsymbol{x}_n$ при условии, что $\boldsymbol{x}_n$ сходится по распределению к некоторой случайной величине.

Третий результат позволит нам тестировать нелинейные гипотезы при заданном асимптотическом распределении оценки.

Лемма 2.5 («дельта-метод»): *Предположим, что $\{\boldsymbol{x}_n\}$ — последовательность K-мерных случайных векторов, таких, что $\boldsymbol{x}_n \rightarrow_p \boldsymbol{\beta}$ и*

$$\sqrt{n}(\boldsymbol{x}_n - \boldsymbol{\beta}) \underset{d}{\rightarrow} \boldsymbol{z},$$

и предположим, что $\boldsymbol{a}(\cdot)\colon \mathbb{R}^K \rightarrow \mathbb{R}^r$ имеет непрерывные первые производные и $\boldsymbol{A}(\boldsymbol{\beta})$ обозначает $r \times K$ матрицу первых производных, вычисленных в $\boldsymbol{\beta}$:

$$\underset{(r\times K)}{\boldsymbol{A}(\boldsymbol{\beta})} \equiv \frac{\partial \boldsymbol{a}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}.$$

Тогда

$$\sqrt{n}[\boldsymbol{a}(\boldsymbol{x}_n) - \boldsymbol{a}(\boldsymbol{\beta})] \underset{d}{\rightarrow} \boldsymbol{A}(\boldsymbol{\beta})\boldsymbol{z}.$$

В частности:

$$«\sqrt{n}(\boldsymbol{x}_n - \boldsymbol{\beta}) \underset{d}{\rightarrow} N(\boldsymbol{0}, \boldsymbol{\Sigma})» \Rightarrow «\sqrt{n}[\boldsymbol{a}(\boldsymbol{x}_n) - \boldsymbol{a}(\boldsymbol{\beta})] \underset{d}{\rightarrow} N\big(\boldsymbol{0}, \boldsymbol{A}(\boldsymbol{\beta})\boldsymbol{\Sigma}\boldsymbol{A}(\boldsymbol{\beta})'\big)».$$

Доказательство этого является хорошим способом научиться тому, как использовать результаты, которые изучались до сих пор.

Доказательство. По теореме о среднем из математического анализа (см. параграф 7.3 по поводу формулировки теоремы) существует K-мерный вектор $\boldsymbol{y}_n$ между $\boldsymbol{x}_n$ и $\boldsymbol{\beta}$, такой, что

$$\boldsymbol{a}(\boldsymbol{x}_n) - \boldsymbol{a}(\boldsymbol{\beta}) = \underset{(r\times K)}{\boldsymbol{A}(\boldsymbol{y}_n)}\underset{(K\times 1)}{(\boldsymbol{x}_n - \boldsymbol{\beta})}.$$

Умножая обе части на $\sqrt{n}$, мы получаем:

$$\sqrt{n}[\boldsymbol{a}(\boldsymbol{x}_n) - \boldsymbol{a}(\boldsymbol{\beta})] = \boldsymbol{A}(\boldsymbol{y}_n)\sqrt{n}(\boldsymbol{x}_n - \boldsymbol{\beta}).$$

Поскольку $\boldsymbol{y}_n$ находится между $\boldsymbol{x}_n$ и $\boldsymbol{\beta}$ и поскольку $\boldsymbol{x}_n \to_{\mathrm{p}} \boldsymbol{\beta}$, мы знаем, что $\boldsymbol{y}_n \to_{\mathrm{p}} \boldsymbol{\beta}$. Кроме того, первая производная $\boldsymbol{A}(\cdot)$ непрерывна по предположению. Таким образом, по лемме 2.3(а):

$$\boldsymbol{A}(\boldsymbol{y}_n) \underset{\mathrm{p}}{\to} \boldsymbol{A}(\boldsymbol{\beta}).$$

По лемме 2.4(с), этот факт и гипотеза о том, что $\sqrt{n}(\boldsymbol{x}_n - \boldsymbol{\beta}) \to_{\mathrm{d}} \boldsymbol{z}$, влекут за собой:

$$\boldsymbol{A}(\boldsymbol{y}_n)\sqrt{n}(\boldsymbol{x}_n - \boldsymbol{\beta}) \left(= \sqrt{n}[\boldsymbol{a}(\boldsymbol{x}_n) - \boldsymbol{a}(\boldsymbol{\beta})]\right) \underset{\mathrm{d}}{\to} \boldsymbol{A}(\boldsymbol{\beta})\boldsymbol{z}. \qquad \blacksquare$$

Взгляд на оценки как на последовательности случайных величин

Пусть $\widehat{\boldsymbol{\theta}}_n$ — оценка вектора параметров $\boldsymbol{\theta}$, основанная на выборке размера n. Последовательность $\{\widehat{\boldsymbol{\theta}}_n\}$ является примером последовательности случайных величин, так что понятия, введенные в этом параграфе для последовательностей случайных величин, применимы к $\{\widehat{\boldsymbol{\theta}}_n\}$. Мы говорим, что оценка $\widehat{\boldsymbol{\theta}}_n$ является **состоятельной для $\boldsymbol{\theta}$** (*consistent for* $\boldsymbol{\theta}$), если

$$\operatorname*{plim}_{n\to\infty} \widehat{\boldsymbol{\theta}}_n = \boldsymbol{\theta} \quad \text{или} \quad \widehat{\boldsymbol{\theta}}_n \underset{\mathrm{p}}{\to} \boldsymbol{\theta}.$$

Асимптотическое смещение (*asymptotic bias*) $\widehat{\boldsymbol{\theta}}_n$ определяется как[1] $\operatorname{plim}_{n\to\infty} \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}$. Таким образом, если оценка состоятельна, ее асимптотическое смещение равно нулю. Состоятельная оценка $\widehat{\boldsymbol{\theta}}_n$ является **асимптотически нормальной** (*asymptotically normal*), если

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \underset{\mathrm{d}}{\to} N(\mathbf{0}, \boldsymbol{\Sigma}).$$

Такая оценка называется **$\sqrt{n}$-состоятельной** (*$\sqrt{n}$-consistent*). Акроним, иногда используемый для «состоятельности и асимптотической нормальности», — **CAN**. Дисперсионная матрица $\boldsymbol{\Sigma}$ называется **асимптотической дисперсией** (*asymptotic variance*) и обозначается как $\operatorname{Avar}(\widehat{\boldsymbol{\theta}}_n)$. Некоторые авторы используют обозначение $\operatorname{Avar}(\widehat{\boldsymbol{\theta}}_n)$ для среднего $\boldsymbol{\Sigma}/n$ (которое равно нулю в пределе). В этой книге $\operatorname{Avar}(\widehat{\boldsymbol{\theta}}_n)$ — это дисперсия предельного распределения $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$.

[1]Некоторые авторы используют термин «асимптотическое смещение» иным образом. [Amemiya, 1985], например, определяет его для обозначения $\lim_{n\to\infty} \mathrm{E}(\widehat{\boldsymbol{\theta}}_n) - \boldsymbol{\theta}$.

Законы больших чисел и центральные предельные теоремы

Для последовательности случайных скаляров $\{z_n\}$ **выборочное среднее** (*sample mean*) $\overline{z}_n$ определяется как

$$\overline{z}_n \equiv \frac{1}{n}\sum_{i=1}^{n} z_i.$$

Рассмотрим последовательность $\{\overline{z}_n\}$. Закон больших чисел (*Law of large numbers* — **LLN**) относится к условиям, при которых $\{\overline{z}_n\}$ сходится или по вероятности, или почти наверное. LLN называется усиленным, если имеет место сходимость почти наверное, и слабым, если имеет место сходимость по вероятности. Из части (а) леммы 2.2 мы можем легко получить следующий слабый LLN.

Версия слабого LLN Чебышева:

$$«\lim_{n\to\infty} \mathrm{E}(\overline{z}_n) = \mu,\ \lim_{n\to\infty} \mathrm{Var}(\overline{z}_n) = 0» \Rightarrow «\overline{z}_n \underset{p}{\to} \mu».$$

Это выполняется потому, что при данном условии легко доказать (см. аналитический вопрос), что $\overline{z}_n \to_{\text{m.s.}} \mu$. Следующий усиленный LLN предполагает, что $\{z_i\}$ является i.i.d.-последовательностью (i.i.d. — независимые и одинаково распределенные), но дисперсия необязательно должна быть конечной.

Второй усиленный закон больших чисел Колмогорова: *Пусть последовательность $\{z_i\}$ есть i.i.d. с $\mathrm{E}(z_i) = \mu$[1]. Тогда $\overline{z}_n \to_{\text{a.s.}} \mu$.*

Эти LLNs легко обобщаются на случайные векторы требованием поэлементной сходимости.

Центральные предельные теоремы (*Central Limit Theorems* — **CLT**) говорят о предельном поведении разности между $\overline{z}_n$ и $\mathrm{E}(\overline{z}_n)$ (которое равно $\mathrm{E}(z_i)$, если $\{z_i\}$ является i.i.d.), увеличенной в $\sqrt{n}$ раз. Единственная центральная предельная теорема, которая нам нужна в случае последовательности i.i.d., такова:

CLT Линдеберга — Леви: *Пусть $\{\mathbf{z}_i\}$ есть i.i.d. с $\mathrm{E}\,\mathbf{z}_i = \boldsymbol{\mu}$ и $\mathrm{Var}\,\mathbf{z}_i = \boldsymbol{\Sigma}$. Тогда*

$$\sqrt{n}(\overline{\mathbf{z}}_n - \boldsymbol{\mu}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\mathbf{z}_n - \boldsymbol{\mu}) \underset{d}{\to} N(\mathbf{0}, \boldsymbol{\Sigma}).$$

[1] Таким образом, среднее существует и конечно (действительное число). Когда указывается на некоторый момент (например, среднее), как здесь, то косвенно предполагается, что этот момент существует и конечен.

Это читается так: последовательность случайных векторов $\{\sqrt{n}(\bar{\boldsymbol{z}}_n - \boldsymbol{\mu})\}$ сходится по распределению к случайному вектору, имеющему распределение $N(\mathbf{0}, \boldsymbol{\Sigma})$. (Обычно CLT Линдеберга — Леви используется для последовательности *скалярных* случайных величин. Векторная версия, показанная выше, выводится из скалярной версии следующим образом. Пусть $\{\boldsymbol{z}_i\}$ есть i.i.d. с $\mathrm{E}(\boldsymbol{z}_i) = \boldsymbol{\mu}$ и $\mathrm{Var}\, \boldsymbol{z}_i = \boldsymbol{\Sigma}$ и пусть $\boldsymbol{\lambda}$ есть любой вектор действительных чисел той же самой размерности. Тогда $\{\boldsymbol{\lambda}' \boldsymbol{z}_n\}$ является последовательностью скалярных случайных переменных с $\mathrm{E}(\boldsymbol{\lambda}' \boldsymbol{z}_n) = \boldsymbol{\lambda}' \boldsymbol{\mu}$ и $\mathrm{Var}\, \boldsymbol{\lambda}' \boldsymbol{z}_n = \boldsymbol{\lambda}' \boldsymbol{\Sigma} \boldsymbol{\lambda}$. Из скалярной версии Линдеберга — Леви вытекает тогда, что

$$\sqrt{n}(\boldsymbol{\lambda}' \bar{\boldsymbol{z}}_n - \boldsymbol{\lambda}' \boldsymbol{\mu}) = \boldsymbol{\lambda}' \sqrt{n}(\bar{\boldsymbol{z}}_n - \boldsymbol{\mu}) \underset{\mathrm{d}}{\to} N(0, \boldsymbol{\lambda}' \boldsymbol{\Sigma} \boldsymbol{\lambda}).$$

Но это предельное распределение есть распределение $\boldsymbol{\lambda}' \boldsymbol{x}$, где $\boldsymbol{x} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$. Таким образом, по теореме о многомерной сходимости по распределению, сформулированной выше, $\{\sqrt{n}(\bar{\boldsymbol{z}}_n - \boldsymbol{\mu})\} \to_{\mathrm{d}} \boldsymbol{x}$, что и утверждается в векторной версии Линдеберга — Леви.)

Контрольные вопросы

1. (Обычная сходимость против сходимости по вероятности.) Последовательность действительных чисел является тривиальным примером последовательности случайных величин. Верно ли, что «$\lim_{n \to \infty} z_n = \alpha$» $\Rightarrow$ «$\operatorname{plim}_{n \to \infty} z_n = \alpha$»? **Указание:** *Посмотрите на определение plim. Поскольку $\lim_{n \to \infty} z_n = \alpha$, $|z_n - \alpha| < \varepsilon$ для достаточно больших n.*

2. (Альтернативное определение сходимости для последовательностей векторов.) Проверьте, что данное в тексте определение «$\boldsymbol{z}_n \to_{\mathrm{m.s.}} \boldsymbol{z}$» эквивалентно

$$\lim_{n \to \infty} \mathrm{E}[(\boldsymbol{z}_n - \boldsymbol{z})'(\boldsymbol{z}_n - \boldsymbol{z})] = 0.$$

Указание: $\mathrm{E}[(\boldsymbol{z}_n - \boldsymbol{z})'(\boldsymbol{z}_n - \boldsymbol{z})] = \mathrm{E}[(z_{n1} - z_1)^2] + \cdots + \mathrm{E}[(z_{nK} - z_K)^2]$, *где K — размерность $\boldsymbol{z}$.*

Аналогично проверьте, что данное в тексте определение «$\boldsymbol{z}_n \to_{\mathrm{p}} \boldsymbol{\alpha}$» эквивалентно

$$\lim_{n \to \infty} \mathrm{Prob}\,((\boldsymbol{z}_n - \boldsymbol{z})'(\boldsymbol{z}_n - \boldsymbol{z}) > \varepsilon) = 0 \text{ для всех } \varepsilon > 0.$$

3. Докажите лемму 2.4(c), используя пункты (a) и (b) этой леммы. **Указание:** $\boldsymbol{A}_n \boldsymbol{x}_n = (\boldsymbol{A}_n - \boldsymbol{A})\boldsymbol{x}_n + \boldsymbol{A}\boldsymbol{x}_n$. *Из (b)* $(\boldsymbol{A}_n - \boldsymbol{A})\boldsymbol{x}_n \to_p \mathbf{0}$.

4. Предположим, что $\sqrt{n}(\widehat{\theta}_n - \theta) \to_{\mathrm{d}} N(0, \sigma^2)$. Следует ли из этого, что $\widehat{\theta}_n \to_{\mathrm{p}} \theta$? **Указание:**

$$\widehat{\theta}_n - \theta = \frac{1}{\sqrt{n}} \cdot \sqrt{n}(\widehat{\theta}_n - \theta), \quad \operatorname*{plim}_{n \to \infty} \frac{1}{\sqrt{n}} = 0.$$

5. (Объедините дельта-метод с Линдебергом — Леви.) Пусть $\{z_i\}$ — последовательность i.i.d. (независимых и одинаково распределенных) случайных величин с $\mathrm{E}(z_i) = \mu \neq 0$ и $\mathrm{Var}(z_i) = \sigma^2$, и пусть $\bar{z}_n$ — выборочное среднее. Покажите, что

$$\sqrt{n}\left(\frac{1}{\bar{z}_n} - \frac{1}{\mu}\right) \underset{\mathrm{d}}{\to} N\left(0, \frac{\sigma^2}{\mu^4}\right).$$

Указание: *В лемме 2.5 положите* $\boldsymbol{\beta} = \mu$, $\boldsymbol{a}(\boldsymbol{\beta}) = 1/\mu$, $\boldsymbol{x}_n = \bar{z}_n$.

2.2. Фундаментальные понятия в анализе временных рядов

В этом параграфе мы определяем самые основные понятия в анализе временных рядов, которые будут неотъемлемой частью нашего языка. Основным понятием является **случайный процесс** (*stochastic process*), который является только красивым названием последовательности случайных величин. Если индекс случайных величин интерпретируется как представляющий время, случайный процесс называется **временным рядом** (*time series*). Если $\{z_i\}$ ($i = 1, 2, \dots$) является случайным процессом, его **реализация** (*realization*) или **выборочная траектория** (*sample path*) приписывает каждому i некоторое допустимое значение z_i. Таким образом, реализация $\{z_i\}$ есть последовательность действительных чисел. Мы будем часто использовать термин «временной ряд» и для обозначения реализации процесса, и для обозначения процесса, реализацией которого временной ряд является.

Необходимость эргодической стационарности

Фундаментальной проблемой в анализе временных рядов является то, что мы можем наблюдать реализацию процесса только один раз. Например, выборка годового темпа инфляции США за период с 1946 по 1995 год является строкой из 50 определенных чисел, являющейся только одним из возможных результатов лежащего в основе случайного процесса для темпа инфляции; если бы история пошла другим путем, мы бы получили другую выборку. Если бы мы могли наблюдать историю много раз, мы могли бы собрать много выборок, каждая из которых содержит, возможно, другую строку из 50 чисел. Среднее темпа инфляции для, скажем, 1995 года можно тогда оценить, беря среднее темпа инфляции 1995 года (50-й элемент строки) по этим выборкам. Популяционное среднее, оцененное таким образом, иногда называется **средним по ансамблю** (*ensemble mean*). На языке общей теории равновесия в экономике среднее по ансамблю является средним по всем возможным состояниям природы для любого заданного календарного времени.

Конечно, невозможно наблюдать множество различных альтернативных историй. Но если распределение темпа инфляции остается неизменным (на это свойство будем ссылаться как на **стационарность** (*stationarity*)), определенную строку из 50 чисел, которую мы наблюдаем, можно рассматривать как 50 различных значений из *одного и того же* распределения. Кроме того, если процесс не слишком инерционен (такое свойство имеет то, что называется **эргодичностью** (*ergodicity*)), каждый элемент строки будет содержать некоторую информацию, не доступную из других элементов, и, как показано ниже, среднее по времени по всем элементам единственной строки будет соответствовать среднему по ансамблю).

Различные классы случайных процессов[1]

Стационарные процессы

Случайный процесс $\{z_i\}$ $(i = 1, 2, \ldots)$ является (строго) стационарным, если для любого заданного конечного целого r и для любого множества индексов $i_1, i_2, \ldots, i_r$ совместное распределение $(z_i, z_{i_1}, \ldots, z_{i_r})$ зависит только от $i_1 - i, i_2 - i, \ldots, i_r - i$, но не от i. Например, совместное распределение (z_1, z_5) то же самое, что и для (z_{12}, z_{16}). Что является важным для распределения — это относительное местоположение в последовательности. В частности, распределение z_i не зависит от абсолютного местоположения z_i, то есть i. Таким образом, среднее, дисперсия и другие более высокие моменты, если они существуют, остаются одинаковыми i. Из этого определения также следует, что любое преобразование (функция) стационарного процесса и само стационарно, то есть если $\{z_i\}$ является стационарным, то и $\{f(z_i)\}$ также[2]. Например, последовательность $z_i z_i'$ стационарна, если стационарна последовательность z_i.

Пример 2.1 (i.i.d.-последовательности): Последовательность независимых и одинаково распределенных случайных величин является стационарным процессом, который не проявляет сериальной зависимости.

Пример 2.2 (ряд из одинаковых констант): Извлеките z_1 из некоторого распределения, а затем положите $z_i = z_1$ $(i = 2, 3, \ldots)$. Таким образом, значение процесса заморожено начальной датой. Так, построенный процесс $\{z_i\}$ является стационарным процессом, проявляющим максимальную сериальную зависимость.

[1]Многие из понятий, собранных в этом параграфе, также можно найти в Section 4.7 книги [Davidson and MacKinnon, 1993].

[2]Функция $f(\cdot)$ должна быть «измерима», так чтобы $f(z_i)$ была корректно определенной случайной величиной. Любая непрерывная функция является измеримой. В дальнейшем мы не утруждаем себя в добавлении определения «измеримой», когда функция f от случайной величины понимается как случайная величина.

Очевидно, что если векторный процесс $\{z_i\}$, построенный таким образом, является стационарным процессом, то тогда каждый элемент этого вектора образует одномерный стационарный процесс. Однако обратное утверждение неверно.

Пример 2.3 (противопоставление поэлементной и совместной стационарности): Пусть $\{\varepsilon_i\}$ $(i = 1, 2, \dots)$ — скалярный i.i.d.-процесс. Построим на его основе двумерный процесс $\{z_i\}$, определяя $z_{i1} = \varepsilon_i$ и $z_{i2} = \varepsilon_1$. Скалярный процесс $\{z_{i1}\}$ стационарный (это процесс из примера 2.1). Скалярный процесс $\{z_{i2}\}$ также стационарный (процесс из примера 2.2). Векторный процесс $\{z_i\}$, однако, не является (совместно) стационарным, поскольку (совместное) распределение $z_1 (= (\varepsilon_1, \varepsilon_1)')$ отличается от распределения $z_2 (= (\varepsilon_2, \varepsilon_1)')$.

Многие агрегированные временные ряды, такие как ВВП, не являются стационарными, поскольку они проявляют временной тренд. Менее очевидным примером нестационарности являются международные обменные курсы, которые, как утверждается, имеют возрастающую дисперсию. Но многие временные ряды с трендом могут быть редуцированы к стационарным процессам. Процесс называется **стационарным относительно тренда** (*trend stationary*), если он является стационарным после удаления из него (обычно линейной) функции времени (которое представлено индексом i). Если процесс не является стационарным, но его первая разность, $z_i - z_{i-1}$, стационарна, то процесс $\{z_i\}$ называется **стационарным в разностях** (*difference stationary*). Стационарные относительно тренда и стационарные в разностях процессы будут рассмотрены в главе 9.

Стационарные в ковариациях
(ковариационно-стационарные) процессы

Случайный процесс $\{z_i\}$ является **слабо** (или **в ковариациях, ковариационно**) **стационарным** (*weakly* (или *covariance*) *stationary*), если

(i) $\mathrm{E}(z_i)$ не зависит от i и

(ii) $\mathrm{Cov}(z_i, z_{i-j})$ существует, конечна и зависит только от j, но не от i (например, $\mathrm{Cov}(z_1, z_5)$ равна $\mathrm{Cov}(z_{12}, z_{16})$).

Для среднего и ковариации ковариационно-стационарного процесса имеет значение не абсолютное, а относительное местоположение в последовательности[1]. Очевидно, что если последовательность является (строго) стационарной и если дисперсия и ковариации конечны, то тогда эта последовательность является слабо стационарной (отсюда и термин «стро-

[1] Для среднего такого процесса местоположение не имеет никакого значения. — *Прим. науч. ред. перевода.*

гая»). Пример стационарного в ковариациях, но не строго стационарного процесса будет дан ниже, в примере 2.4.

Автоковариация j-го порядка (*j-th order autocovariance*), обозначаемая $\mathbf{\Gamma}_j$, определяется как

$$\mathbf{\Gamma}_j \equiv \mathrm{Cov}(\mathbf{z}_i, \mathbf{z}_{i-j}) \quad (j = 0, 1, 2, \ldots).$$

Термин «авто» объясняется тем, что входящие в определение две случайные величины взяты из одного и того же процесса. $\mathbf{\Gamma}_j$ не зависит от i из-за стационарности в ковариациях. Также из стационарности в ковариациях $\mathbf{\Gamma}_j$ удовлетворяет

$$\mathbf{\Gamma}_j = \mathbf{\Gamma}'_{-j}. \tag{2.2.1}$$

(Обоснование этого является контрольным вопросом, приведенным ниже.) Ковариация 0-го порядка является дисперсией $\mathbf{\Gamma}_0 = \mathrm{Var}(\mathbf{z}_i)$. Процессы в примерах 2.1 и 2.2 являются стационарными в ковариациях, если эта дисперсия существует и конечна. Для процесса из примера 2.1 $\mathbf{\Gamma}_0$ является дисперсией распределения, и $\mathbf{\Gamma}_j = \mathbf{0}$ для $j \geq 1$. Для процесса из примера 2.2 $\mathbf{\Gamma}_j = \mathbf{\Gamma}_0$.

Для скалярного стационарного в ковариациях процесса $\{z_i\}$ автоковариация j-го порядка является скаляром. Если γ_j является этой ковариацией, она удовлетворяет

$$\gamma_j = \gamma_{-j}. \tag{2.2.2}$$

Возьмем цепочку из n последовательных значений, $(z_i, z_{i+1}, \ldots, z_{i+n-1})$, скалярного процесса. Вследствие стационарности в ковариациях его $n \times n$ ковариационная матрица является той же самой, что и для $(z_1, z_2, \ldots, z_n)$, и является матрицей линейного спектра (*band spectrum matrix*):

$$\mathrm{Var}(z_i, z_{i+1}, \ldots, z_{i+n-1}) = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{n-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \gamma_{n-2} & \cdots & \gamma_1 & \gamma_0 & \gamma_1 \\ \gamma_{n-1} & \cdots & \gamma_2 & \gamma_1 & \gamma_0 \end{bmatrix}.$$

Она называется **автоковариационной матрицей** (*autocovariance matrix*) процесса. **Коэффициент автокорреляции j-го порядка** (*j-th order autocorrelation coefficient*), ρ_j, определяется как

$$\text{коэффициент автокорреляции } j\text{-го порядка} \equiv$$
$$\equiv \rho_j \equiv \frac{\gamma_j}{\gamma_0} = \frac{\mathrm{Cov}(z_i, z_{i-j})}{\mathrm{Var}(z_i)} \quad (j = 1, 2, \ldots). \tag{2.2.3}$$

Для $j = 0$ $\rho_j = 1$. График зависимости $\{\rho_j\}$ от $j = 0, 1, 2, \ldots$ называется **коррелограммой** (*correlogram*).

Процессы белого шума

Очень важным классом слабо стационарных процессов является **процесс белого шума** (*white noise process*), процесс с нулевым средним и отсутствием сериальной корреляции:

> стационарный в ковариациях процесс $\{z_i\}$ является **белым шумом** (*white noise*), если $\mathrm{E}(z_i) = 0$ и $\mathrm{Cov}(z_i, z_{i-j}) = 0$ для $j \neq 0$.

Ясно, что последовательность независимых и одинаково распределенных случайных величин (i.i.d.) с нулевым средним и конечной дисперсией является частным случаем процесса белого шума. По этой причине она называется **независимым процессом белого шума** (independent white noise process).

Пример 2.4 (процесс белого шума, не являющийся строго стационарным[1]): Пусть w — случайная величина, равномерно распределенная в интервале $(0, \pi)$, определим

$$z_i = \cos(iw) \quad (i = 1, 2, \ldots).$$

Можно показать, что $\mathrm{E}(z_i) = 0$, $\mathrm{Var}(z_i) = 1/2$ и $\mathrm{Cov}(z_i, z_j) = 0$ для $i \neq j$. Поэтому $\{z_i\}$ является белым шумом. Однако ясно, что это не независимый процесс белого шума. Он даже не строго стационарный.

Эргодичность

Говорят, что стационарный процесс $\{z_i\}$ является **эргодическим** (*ergodic*), если для любых двух ограниченных функций $f : \mathbb{R}^{k+1} \to \mathbb{R}$ и $g : \mathbb{R}^{l+1} \to \mathbb{R}$

$$\begin{aligned}\lim_{n\to\infty} \left|\mathrm{E}[f(z_i, \ldots, z_{i+k})g(z_{i+n}, \ldots, z_{i+n+\ell})]\right| &= \\ = \left|\mathrm{E}[f(z_i, \ldots, z_{i+k})]\right| &\left|\mathrm{E}[g(z_i, \ldots, z_{i+\ell})]\right|.\end{aligned}$$

Эвристически стационарный процесс является эргодическим, если он асимптотически независим, то есть если две случайные величины, расположенные в последовательности далеко друг от друга, почти независимо распределены. Стационарный процесс, который является эргодическим, будет называться **эргодически стационарным** (*ergodic stationary*). Эргодическая стационарность будет неотъемлемой частью разработки теории больших выборок из-за следующего свойства.

[1] Взят из Example 7.8 в [Anderson, 1971, p. 379].

Эргодическая теорема *(см., например, Theorem 9.5.5 в [Karlin and Taylor, 1975]): Пусть $\{z_i\}$ является стационарным и эргодическим процессом с* $\mathrm{E}(z_i) = \mu$[1]. *Тогда*

$$\bar{z}_n \equiv \frac{1}{n}\sum_{i=1}^{n} z_i \underset{a.s.}{\to} \mu.$$

Таким образом, эргодическая теорема является существенным обобщением LLN Колмогорова. Сериальная зависимость, которая исключена предположением i.i.d. в LLN Колмогорова, в эргодической теореме допускается при условии, что она исчезает в долговременной перспективе. Поскольку для любой (измеримой) функции $f(\cdot)$ процесс $\{f(z_i\}$ является эргодически стационарным всякий раз, когда эргодически стационарен z_i, из этой теоремы вытекает, что любой момент стационарного и эргодического процесса (если он существует и конечен) состоятельно оценивается выборочным средним. Предположим, например, что процесс z_i является стационарным и эргодическим, и $\mathrm{E}(z_i z_i')$ существует и конечно. Тогда среднее $\frac{1}{n}\sum_i z_i z_i'$ состоятельно для $\mathrm{E}(z_i z_i')$.

Простейшим примером эргодически стационарного процесса является процесс независимого белого шума. (Процесс белого шума, в котором независимость ослаблена до отсутствия сериальной корреляции, не обязательно эргодичен; см. пример 2.4 выше.) Другой важный пример — AR(1) — процесс, удовлетворяющий

$$z_i = c + \rho z_{i-1} + \varepsilon_i, \quad |\rho| < 1,$$

где $\{\varepsilon_i\}$ является независимым белым шумом.

Мартингалы

Пусть x_i является элементом z_i. Скалярный процесс $\{x_i\}$ называется **мартингалом по отношению к** $\{z_i\}$ (*martingale with respect to* $\{z_i\}$), если

$$\mathrm{E}(x_i | z_{i-1}, z_{i-2}, \dots, z_1) = x_{i-1} \text{ для } i \geq 2^2. \tag{2.2.4}$$

Множество $(z_{i-1}, z_{i-2}, \dots, z_1)$, стоящее в условии, часто называется **информационным множеством** (*information set*) в точке (на дату) $i-1$.

---

[1]Так что среднее предполагается существующим и конечным.

[2]Если процесс начинается в бесконечном прошлом, так что i пробегает значения от $-\infty$ до ∞, определением является $\mathrm{E}(x_i | z_{i-1}, z_{i-2}, \dots) = x_{i-1}$, и уточнение «$i > 2$» не нужно. Начинается ли процесс с бесконечного прошлого или с даты $i = 1$, для теории больших выборок, которая будет излагаться ниже, не имеет значения. Что будет иметь значение, так это то, что процесс начинается ранее периода выборки.

Процесс $\{x_i\}$ называется просто **мартингалом** (*martingale*), если информационное множество образовано его собственными прошлыми значениями $(x_{i-1}, x_{i-2}, \ldots, x_1)$. Если $\boldsymbol{z}_i$ включает x_i, то $\{x_i\}$ является мартингалом, поскольку

$$\begin{aligned}
&\mathrm{E}(x_i|x_{i-1}, x_{i-2}, \ldots, x_1) = \\
&= \mathrm{E}[\mathrm{E}(x_i|\boldsymbol{z}_{i-1}, \boldsymbol{z}_{i-2}, \ldots, \boldsymbol{z}_1)|x_{i-1}, x_{i-2} \ldots, x_1] \\
&\qquad\qquad \text{(закон повторных математических ожиданий)} \\
&= \mathrm{E}(x_{i-1}|x_{i-1}, x_{i-2}, \ldots, x_1) = x_{i-1}.
\end{aligned}$$

Векторный процесс $\{\boldsymbol{z}_i\}$ называется **мартингалом**, если

$$\mathrm{E}(\boldsymbol{z}_i|\boldsymbol{z}_{i-1}, \ldots, \boldsymbol{z}_1) = \boldsymbol{z}_{i-1} \text{ для } i \geq 2. \tag{2.2.5}$$

Пример 2.5 (мартингальная гипотеза Холла (*Hall's Martingale Hypothesis*)): Пусть $\boldsymbol{z}_i$ — вектор, содержащий совокупность макроэкономических переменных (таких как предложение денег или ВВП), включая агрегированное потребление c_i за период i. Мартингальная гипотеза в [Hall, 1978] состоит в том, что потребление является мартингалом относительно $\boldsymbol{z}_i$:

$$\mathrm{E}(c_i|\boldsymbol{z}_{i-1}, \boldsymbol{z}_{i-2}, \ldots, \boldsymbol{z}_1) = c_{i-1}.$$

Это формализует понятие теории потребления, называемое «сглаживанием потребления»: потребитель, желающий избежать колебаний в уровне жизни, корректирует потребление на дату $i-1$ до такого уровня, чтобы не ожидать никаких изменений в последующем уровне потребления.

Случайные блуждания

Важным примером мартингалов является **случайное блуждание** (*random walk*). Пусть $\{\boldsymbol{g}_i\}$ — векторный процесс независимого белого шума (таким образом, он является i.i.d. со средним $\mathbf{0}$ и конечной дисперсионной матрицей). Случайное блуждание, $\{\boldsymbol{z}_i\}$, является последовательностью накопленных сумм:

$$\boldsymbol{z}_1 = \boldsymbol{g}_1, \ \boldsymbol{z}_2 = \boldsymbol{g}_1 + \boldsymbol{g}_2, \ldots, \ \boldsymbol{z}_i = \boldsymbol{g}_1 + \boldsymbol{g}_2 + \cdots + \boldsymbol{g}_i, \cdots. \tag{2.2.6}$$

Если задана последовательность $\{\boldsymbol{z}_i\}$, лежащую в ее основе последовательность независимого белого шума, $\{\boldsymbol{g}_i\}$, можно восстановить, беря первые разности:

$$\boldsymbol{g}_1 = \boldsymbol{z}_1, \ \boldsymbol{g}_2 = \boldsymbol{z}_2 - \boldsymbol{z}_1, \ldots, \ \boldsymbol{g}_i = \boldsymbol{z}_i - \boldsymbol{z}_{i-1}, \ldots. \tag{2.2.7}$$

Таким образом, первая разность случайного блуждания является независимым белым шумом. Случайное блуждание является мартингалом, поскольку

$$
\begin{aligned}
&\mathrm{E}(\boldsymbol{z}_i|\boldsymbol{z}_{i-1},\dots,\boldsymbol{z}_1) = \mathrm{E}(\boldsymbol{z}_i|\boldsymbol{g}_{i-1},\dots,\boldsymbol{g}_1)\\
&\text{(так как } (\boldsymbol{z}_{i-1},\dots,\boldsymbol{z}_1) \text{ и } (\boldsymbol{g}_{i-1},\dots,\boldsymbol{g}_1) \text{ содержат одинаковую}\\
&\text{информацию, как мы только что видели)}\\
&= \mathrm{E}(\boldsymbol{g}_1 + \boldsymbol{g}_2 + \dots + \boldsymbol{g}_i|\boldsymbol{g}_{i-1},\dots \boldsymbol{g}_1)\\
&= \mathrm{E}(\boldsymbol{g}_i|\boldsymbol{g}_{i-1},\dots,\boldsymbol{g}_1) + (\boldsymbol{g}_1 + \dots + \boldsymbol{g}_{i-1}) =\\
&= \boldsymbol{g}_1 + \dots + \boldsymbol{g}_{i-1} \; (\mathrm{E}(\boldsymbol{g}_i|\boldsymbol{g}_{i-1},\dots,\boldsymbol{g}_1) = \boldsymbol{0}\\
&\text{так как } \{\boldsymbol{g}_i\} \text{ является независимым белым шумом}) =\\
&= \boldsymbol{z}_{i-1} \quad (\text{по определению } \boldsymbol{z}_{i-1}).
\end{aligned} \tag{2.2.8}
$$

Последовательность мартингал-разностей

Векторный процесс $\boldsymbol{g}_i$ с $\mathrm{E}(\boldsymbol{g}_i) = \boldsymbol{0}$ называется **последовательностью мартингал-разностей** (*martingale difference sequence*, **m.d.s.**) или **мартингал-разностью** (*martingale differences*), если математическое ожидание, условное относительно его прошлых значений, также равно нулю:

$$\mathrm{E}(\boldsymbol{g}_i|\boldsymbol{g}_{i-1}, \boldsymbol{g}_{i-2},\dots,\boldsymbol{g}_1) = \boldsymbol{0} \quad \text{для} \quad i \ge 2. \tag{2.2.9}$$

Процесс называется так потому, что накопленная сумма $\boldsymbol{z}_i$, созданная на основе последовательности мартингал-разностей $\boldsymbol{g}_i$, является мартингалом; доказательство этого аналогично (2.2.8). Обратно, если $\boldsymbol{z}_i$ — мартингал, то его первая разность, построенная, как в (2.2.7), является последовательностью мартингал-разностей.

Последовательность мартингал-разностей не имеет сериальной корреляции (то есть $\mathrm{Cov}(\boldsymbol{g}_i, \boldsymbol{g}_{i-j}) = \boldsymbol{0}$ для всех i и $j \neq 0$). Доказательство этого утверждения проводится следующим образом.

Доказательство. Прежде всего заметим, что мы можем, без потери общности, предположить, что $j \ge 1$. Так как среднее нулевое, достаточно показать, что $\mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_{i-j}') = \boldsymbol{0}$. Поэтому рассмотрим это выражение, переписав его следующим образом:

$$
\begin{aligned}
&\mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_{i-j}') =\\
&= \mathrm{E}[\mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_{i-j}'|\boldsymbol{g}_{i-j})] \quad \text{(по закону полных математических ожиданий)}\\
&= \mathrm{E}[\mathrm{E}(\boldsymbol{g}_i|\boldsymbol{g}_{i-j})\boldsymbol{g}_{i-j}']
\end{aligned}
$$

(вследствие линейности условных математических ожиданий).

Далее, так как $j \geq 1$, $(\boldsymbol{g}_{i-1}, \ldots, \boldsymbol{g}_{i-j}, \ldots, \boldsymbol{g}_1)$ включает $\boldsymbol{g}_{i-j}$. Следовательно,

$$\begin{aligned}
&\mathrm{E}(\boldsymbol{g}_i | \boldsymbol{g}_{i-j}) = \\
&= \mathrm{E}[\mathrm{E}(\boldsymbol{g}_i | \boldsymbol{g}_{i-1}, \ldots, \boldsymbol{g}_{i-j}, \ldots, \boldsymbol{g}_1) | \boldsymbol{g}_{i-j}] \\
&\text{(по закону повторных математических ожиданий)} \\
&= \mathbf{0}.
\end{aligned}$$

Последнее равенство выполняется потому, что

$$\mathrm{E}(\boldsymbol{g}_i | \boldsymbol{g}_{i-1}, \ldots, \boldsymbol{g}_{i-j}, \ldots, \boldsymbol{g}_1) = \mathbf{0}. \qquad \blacksquare$$

ARCH-процессы

Примером мартингал-разностей, часто используемым в анализе доходностей активов, является процесс с **авторегрессионной условной гетероскедастичностью** (*autoregressive conditional heteroskedastic* — **ARCH**), введенный в [Engle, 1982]. Процесс $\{g_i\}$ называется **ARCH-процессом порядка 1** (ARCH process of order 1) (ARCH(1)), если он может быть записан в виде:

$$g_i = \sqrt{\zeta + \alpha g_{i-1}^2} \cdot \varepsilon_i, \tag{2.2.10}$$

где $\{\varepsilon_i\}$ есть i.i.d. с нулевым средним и единичной дисперсией. Если g_1 является начальным значением процесса, мы можем использовать (2.2.10) для вычисления значений g_i. Например,

$$g_2 = \sqrt{\zeta + \alpha g_1^2} \cdot \varepsilon_2.$$

В более общем случае $g_i (i \geq 2)$ есть функция от g_1 и $\{\varepsilon_2, \varepsilon_3, \ldots, \varepsilon_i\}$. Следовательно, ε_i не зависит от $(g_1, g_2, \ldots, g_{i-1})$. Тогда легко показать, что $\{g_i\}$ есть m.d.s., поскольку

$$\begin{aligned}
&\mathrm{E}(g_i | g_{i-1}, g_{i-2}, \ldots, g_1) = && (2.2.11)\\
&= \mathrm{E}\left(\sqrt{\zeta + \alpha g_{i-1}^2} \cdot \varepsilon_i | g_{i-1}, g_{i-2}, \ldots, g_1\right) = \\
&= \sqrt{\zeta + \alpha g_{i-1}^2}\, \mathrm{E}(\varepsilon_i | g_{i-1}, g_{i-2}, \ldots, g_1) = \\
&= \sqrt{\zeta + \alpha g_{i-1}^2}\, \mathrm{E}(\varepsilon_i) \quad (\text{так как } \varepsilon_i \text{ не зависит от } (g_1, g_2, \ldots, g_{i-1})) = \\
&= 0 \quad (\text{так как } \mathrm{E}(\varepsilon_i) = 0). && (2.2.12)
\end{aligned}$$

Аналогичная аргументация приводит к

$$\mathrm{E}(g_i^2 | g_{i-1}, g_{i-2}, \ldots, g_1) = \zeta + \alpha g_{i-1}^2. \tag{2.2.13}$$

Поэтому условный второй момент (который равен условной дисперсии, так как $\mathrm{E}(g_i | g_{i-1}, g_{i-2}, \ldots, g_1) = 0$)) является функцией от собственной истории процесса. В этом смысле указанный процесс выражает **собственную условную гетероскедастичность** (*own conditional heteroskedasticity*). Можно показать (см., например [Engle, 1982]), что такой процесс является строго стационарным и эргодическим, если $|\alpha| < 1$, при условии, что g_1 извлекается из соответствующего распределения, или при условии, что процесс начинается в бесконечном прошлом. Если процесс g_i стационарный, легко найти его *без*условный второй момент. Беря безусловное математическое ожидание от обеих сторон (2.2.13) и замечая, что

$$\mathrm{E}[\mathrm{E}(g_i^2 | g_{i-1}, g_{i-2}, \ldots, g_1)] = \mathrm{E}(g_i^2) \quad \text{и} \quad \mathrm{E}(g_i^2) = \mathrm{E}(g_{i-1}^2),$$

если g_i стационарный, мы получаем:

$$\mathrm{E}(g_i^2) = \zeta + \alpha\, \mathrm{E}(g_i^2) \quad \text{или} \quad \mathrm{E}(g_i^2) = \frac{\zeta}{1-\alpha}. \tag{2.2.14}$$

Если $\alpha > 0$, эта модель отражает характерное свойство, найденное для доходностей активов: за большими значениями, как правило, следуют также большие значения. (Для более подробной информации об ARCH-процессах см., например [Hamilton, 1994, Section 21.1].)

Другая формулировка отсутствия сериальной зависимости

Очевидно, что процесс независимого белого шума является стационарной последовательностью мартингал-разностей с конечной дисперсией. И, как уже было видно, последовательность мартингал-разностей не имеет сериальной корреляции. Таким образом, у нас есть три формулировки отсутствия сериальной зависимости для стационарного в ковариациях процесса с нулевым средним. Они, в порядке ослабления, таковы:

$$\begin{aligned} &(1)\ «\{g_i\}\ \text{независимый белый шум»} \Rightarrow \\ &\Rightarrow (2)\ «\{g_i\}\ \text{стационарная m.d.s. с конечной дисперсией»} \Rightarrow \\ &\Rightarrow (3)\ «\{g_i\}\ \text{белый шум»}. \end{aligned} \tag{2.2.15}$$

Условие (1) более сильное, чем (2), поскольку существуют процессы, удовлетворяющие (2), но не (1). Примером является ARCH(1)-процесс (2.2.10) с $|\alpha| < 1$. Рис. 2.1 показывает, как реализация процесса, удовлетворяющего (1), обычно отличается от реализации процесса, удовлетворяющего (2). Верхний график (a) изображает реализацию последовательности независимых и нормально распределенных случайных величин со средним 0 и единичной дисперсией. Нижний график (b) изображает ARCH(1)-процесс (2.2.10) с $\zeta = 0{,}2$ и $\alpha = 0{,}8$ (так что безусловная дисперсия, $\zeta/(1-\alpha)$, равна единице, как и на графике (a)), где i.i.d.-последовательность ε_i в (2.2.10) та же, что и использованная в части (a) рис. 2.1, так что знаки в (a) и (b) одинаковы в каждой точке. Ряд в (b) в целом менее волатилен, чем в (a), но в некоторых точках он намного более волатилен. Тем не менее этот ряд стационарный.

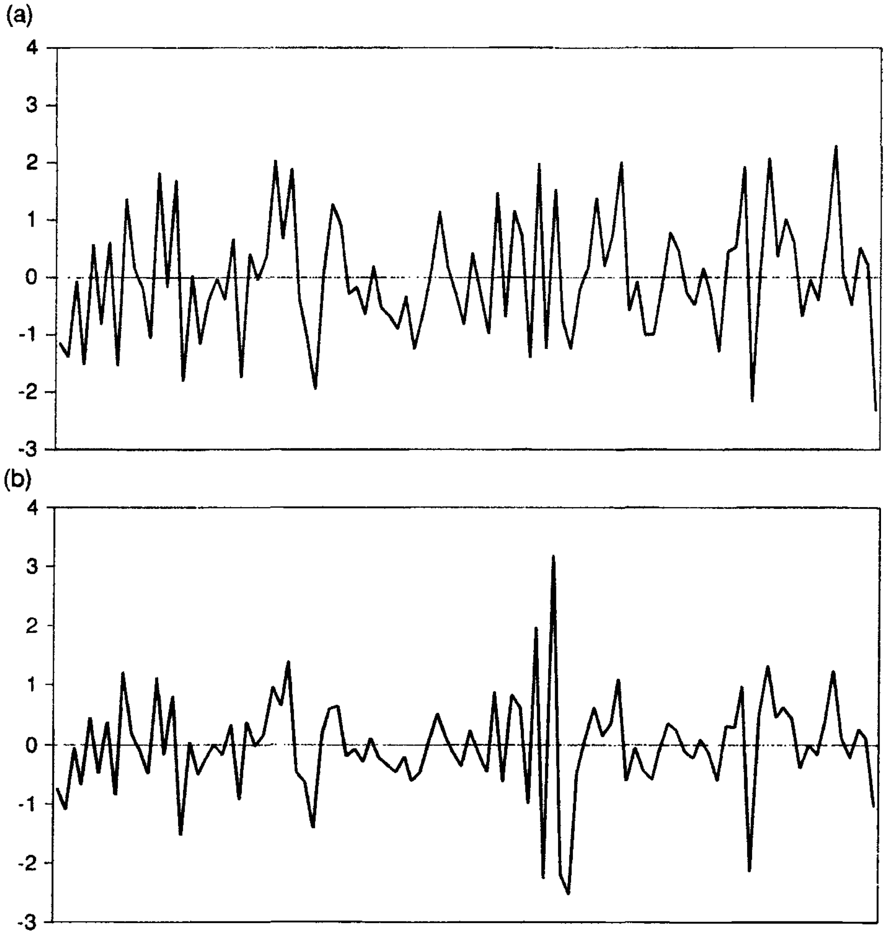

Рис. 2.1. Графики сериально некоррелированных временных рядов: (a) последовательность i.i.d. $N(0,1)$, (b) ARCH(1) с шоками, взятыми из (a)

Условие (2) более сильное, чем (3): процесс в примере 2.4. является белым шумом, но (как вы покажете в контрольном вопросе) он не удовлетворяет (2).

CLT для эргодических стационарных последовательностей мартингал-разностей

Следующая CLT обобщает CLT Линдеберга—Леви на стационарную и эргодическую m.d.s.

CLT для эргодических стационарных мартингал-разностей [Billingsley, 1961]: *Пусть* $\{g_i\}$ *— векторная последовательность мартингал-разностей, которая является стационарной и эргодической с* $\mathrm{E}(g_i g_i') = \mathbf{\Sigma}$[1], *и пусть* $\overline{g} \equiv \frac{1}{n}\sum_{i=1}^{n} g_i$. *Тогда*

$$\sqrt{n}\,\overline{g} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_i \underset{d}{\to} N(\mathbf{0}, \mathbf{\Sigma}).$$

В отличие от CLT Линдеберга — Леви, здесь нет необходимости вычитать среднее из g_i, поскольку безусловное среднее m.d.s. по определению ноль. По той же причине $\mathbf{\Sigma}$ также равна $\mathrm{Var}(g_i)$. Эта CLT, будучи применимой не только к последовательности i.i.d., но и к стационарным мартингал-разностям, таким как ARCH(1)-процессы, является более общей, чем CLT Линдеберга — Леви.

Мы представили LLN для сериально коррелированных процессов в форме эргодической теоремы. Центральная предельная теорема для сериально коррелированных процессов будет представлена в главе 6.

Контрольные вопросы

1. Докажите, что $\mathbf{\Gamma}_j = \mathbf{\Gamma}'_{-j}$. **Указание:**

$$\mathrm{Cov}(z_i, z_{i-j}) = \mathrm{E}[(z_i - \boldsymbol{\mu})(z_{i-j} - \boldsymbol{\mu})'],$$

где $\boldsymbol{\mu} = \mathrm{E}(z_i)$. *В силу стационарности в ковариациях,*

$$\mathrm{Cov}(z_i, z_{i-j}) = \mathrm{Cov}(z_{i+j}, z_i).$$

2. (Прогнозирование белого шума.) Для процесса белого шума в примере 2.4 $\mathrm{E}(z_i) = 0$. Каково значение $\mathrm{E}(z_i z_1)$ для $i \geq 2$? **Указание:** *Вы должны быть в состоянии прогнозировать будущее точно, если знаете значение* z_1.

   Является ли процесс m.d.s.? [Ответ: Нет.]

3. (Отсутствие ожидаемых изменений в мартингалах.) Предположим, что $\{x_i\}$ — мартингал относительно $\{z_i\}$. Покажите, что

$$\mathrm{E}(x_{i+j}|z_{i-1}, z_{i-2}, \ldots, z_1) = x_{i-1} \text{ и } \mathrm{E}(x_{i+j+1} - x_{i+j}|z_{i-1}, z_{i-2}, \ldots, z_1) = 0$$

для $j = 0, 1, \ldots$. **Указание:** *Используйте закон повторных математических ожиданий.*

4. Пусть $\{x_i\}$ — последовательность действительных чисел, которые изменяются по i, и $\{\varepsilon_i\}$ — последовательность i.i.d. случайных величин со средним 0 и конечной дисперсией. Является ли процесс $\{x_i \cdot \varepsilon_i\}$ i.i.d.? [Ответ: Нет.] Является ли он сериально независимым? [Ответ: Да.] А m.d.s.? [Ответ: Да.] Стационарным? [Ответ: Нет.]

[1] Так как последовательность $\{g_i\}$ стационарна, эта матрица перекрестных моментов не зависит от i. Также поскольку матрица перекрестных моментов указана, то подразумевается, что все перекрестные моменты существуют и конечны.

5. Покажите, что случайное блуждание нестационарно. **Указание:** *Проверьте дисперсию.*

6. (Первая разность мартингала является мартингал-разностью.) Пусть $\{z_i\}$ — мартингал. Покажите, что процесс $\{g_i\}$, построенный по (2.2.7), является m.d.s. **Указание:** *$(g_1, \ldots, g_i)$ и $(z_1, \ldots, z_i)$ используют одну и ту же информацию.*

7. (m.d.s., не являющаяся независимым белым шумом.) Пусть $g_t = \varepsilon_i \cdot \varepsilon_{i-1}$, где $\{\varepsilon_i\}$ — процесс независимого белого шума. Очевидно, что $\{g_i\}$ не является i.i.d. Убедитесь, что $\{g_i\}$ $(i = 2, 3, \ldots)$ есть m.d.s. **Указание:**

$$\mathrm{E}(g_i | g_{i-1}, \ldots, g_2) = \mathrm{E}[\mathrm{E}(\varepsilon_i \cdot \varepsilon_{i-1} | \varepsilon_{i-1}, \ldots, \varepsilon_1) | \varepsilon_{i-1} \cdot \varepsilon_{i-2}, \varepsilon_{i-2} \cdot \varepsilon_{i-3}, \ldots, \varepsilon_2 \cdot \varepsilon_1].$$

8. (Пересмотр ожиданий является m.d.s.) Пусть $\{y_i\}$ — процесс такой, что $\mathrm{E}(y_i | y_{i-1}, y_{i-2}, \ldots, y_1)$ существует и конечно, и определим

$$r_{i1} = \mathrm{E}(y_i | y_{i-1}, y_{i-2}, \ldots, y_1) - \mathrm{E}(y_i | y_{i-2}, y_{i-3}, \ldots, y_1).$$

Таким образом, r_{i1} есть изменение ожиданий при добавлении еще одного наблюдения к информационному множеству. Покажите, что $\{r_{i1}\}$ $(i \geq 2)$ есть m.d.s.

9. (CLT Биллингсли сильнее CLT Линдеберга—Леви.) Пусть $\{\boldsymbol{z}_i\}$ — последовательность i.i.d. с $\mathrm{E}(\boldsymbol{z}_i) = \boldsymbol{\mu}$ и $\mathrm{Var}(\boldsymbol{z}_i) = \boldsymbol{\Sigma}$, как в CLT Линдеберга—Леви. Используйте CLT для мартингал-разностей, чтобы доказать утверждение CLT Линдеберга—Леви, а именно что $\sqrt{n}(\overline{\boldsymbol{z}}_n - \boldsymbol{\mu}) \to_{\mathrm{d}} N(\mathbf{0}, \boldsymbol{\Sigma})$. **Указание:** *$\{\overline{\boldsymbol{z}}_n - \boldsymbol{\mu}\}$ является процессом независимого белого шума и, следовательно, эргодически стационарной m.d.s.*

2.3. Распределение OLS-оценки на больших выборках

Важность в эконометрике процедуры OLS, первоначально разработанной для классической регрессионной модели главы 1, заключается в том, что она имеет хорошие асимптотические свойства для класса моделей, которые полезны в экономике, но отличны от классической. Среди этих моделей модель, представленная в этом параграфе, имеет, вероятно, наиболее широкий диапазон экономических приложений. Для получения асимптотического распределения OLS-оценки не требуется никаких специфических предположений о распределении (таких как нормальность ошибок). Требование теории конечных выборок, чтобы регрессоры были строго экзогенными или «фиксированными», заменяется на более слабое требование того, чтобы они были «предопределенными». (Для полноты изложения, в приложении развивается параллельная асимптотическая теория для модели с «фиксированными» регрессорами.)

Модель

Мы используем термин **«процесс порождения данных»** (*data generating process* — **DGP**) для случайного процесса, порождающего конечную выборку $(\boldsymbol{y}, \boldsymbol{X})$. Следовательно, если мы специфицируем DGP, то можно определить совместное распределение конечной выборки $(\boldsymbol{y}, \boldsymbol{X})$. В теории конечных выборок, где размер выборки фиксированный и конечный, мы определяем модель как множество совместных распределений $(\boldsymbol{y}, \boldsymbol{X})$. В теории больших выборок модель формулируется как множество DGP. Модель, которую мы исследуем, — это множество DGP, удовлетворяющих следующей совокупности предположений.

Предположение 2.1 (линейность):

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i \quad (i = 1, 2, \ldots, n),$$

где $\boldsymbol{x}_i$ есть K-мерный вектор объясняющих переменных (регрессоров), $\boldsymbol{\beta}$ — K-мерный вектор коэффициентов и ε_i — ненаблюдаемая ошибка.

Предположение 2.2 (эргодическая стационарность): *$(K+1)$-мерный векторный случайный процесс $\{y_i, \boldsymbol{x}_i\}$ является совместно стационарным и эргодическим.*

Предположение 2.3 (предопределенные регрессоры): *Все регрессоры являются **предопределенными** (predetermined) в том смысле, что они ортогональны* текущей *ошибке: $\mathrm{E}(x_{ik}\varepsilon_i) = 0$ для всех i и k $(= 1, 2, \ldots, K)$*[1]*. Это может быть записано как*

$$\mathrm{E}[\boldsymbol{x}_i \cdot (y_i - \boldsymbol{x}_i'\boldsymbol{\beta})] = \mathbf{0} \quad \text{или, эквивалентно,} \quad \mathrm{E}(\boldsymbol{g}_i) = \mathbf{0},$$

где $\boldsymbol{g}_i \equiv \boldsymbol{x}_i \cdot \varepsilon_i$.

Предположение 2.4 (ранговое условие): *$K \times K$ матрица $\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')$ является невырожденной (и, следовательно, конечной). Мы обозначаем эту матрицу $\boldsymbol{\Sigma}_{\boldsymbol{xx}}$.*

Предположение 2.5 ($\boldsymbol{g}_i$ является последовательностью мартингал-разностей с конечными вторыми моментами): *$\{\boldsymbol{g}_i\}$ есть последовательность мартингал-разностей (поэтому заведомо $\mathrm{E}(\boldsymbol{g}_i) = \mathbf{0}$). Матрица*

[1] Наше определение термина «*предопределенные*» неуниверсально. Некоторые авторы говорят, что регрессоры предопределенные, если $\mathrm{E}(\boldsymbol{x}_{i-j} \cdot \varepsilon_i) = \mathbf{0}$ для всех $j \geq 0$, а не только для $j = 0$. То есть ошибка ортогональна не только текущим, но и прошлым значениям регрессоров. В [Koopmans and Hood, 1953] регрессоры называются предопределенными, если ε_i не зависит от $\boldsymbol{x}_{i-j}$ для всех $j \geq 0$. Наше определение такое же, как в [Hamilton, 1994].

размера $K \times K$ перекрестных моментов, $\mathrm{E}(\boldsymbol{g}_i \boldsymbol{g}_i')$, невырождена. Мы используем $\boldsymbol{S}$ для $\mathrm{Avar}(\overline{\boldsymbol{g}})$ (дисперсия асимптотического распределения $\sqrt{n}\,\overline{\boldsymbol{g}}$, где $\overline{\boldsymbol{g}} \equiv \frac{1}{n}\sum_i \boldsymbol{g}_i$). По предположению 2.2 и CLT для эргодической стационарной мартингал-разности $\boldsymbol{S} = \mathrm{E}(\boldsymbol{g}_i \boldsymbol{g}_i')$.

Первое предположение — это просто воспроизведение предположения 1.1. Оставшиеся предположения требуют длительных комментариев.

- (Эргодическая стационарность.) Тривиальный, но важный специальный случай эргодической стационарности состоит в том, что $\{y_i, \boldsymbol{x}_i\}$ есть i.i.d., то есть выборка является случайной[1]. В большинстве существующие микроданные о домохозяйствах являются случайными выборками, с наблюдениями, случайно выбранными из генеральной совокупности национальных домохозяйств. Таким образом, мы ни в коем случае не исключаем модели, которые используют кросс-секционные данные.

- (Модель адаптирована к условной гетероскедастичности.) Если $\{y_i, \boldsymbol{x}_i\}$ стационарный, то ошибка $\varepsilon_i = y_i - \boldsymbol{x}_i'\boldsymbol{\beta}$ также стационарна. Поэтому из предположения 2.2 вытекает, что безусловный второй момент $\mathrm{E}(\varepsilon_i^2)$ — если он существует и конечен — постоянен по i. То есть ошибка безусловно гомоскедастична. Тем не менее ошибка может быть условно гетероскедастичной — условный второй момент, $\mathrm{E}(\varepsilon_i^2 | \boldsymbol{x}_i)$, может зависеть от $\boldsymbol{x}_i$. Пример, в котором ошибка гомоскедастична безусловно, но не условно, включен в параграф 2.6, где будут изучены последствия наложения на модель условной гомоскедастичности (что $\mathrm{E}(\varepsilon_i^2 | \boldsymbol{x}_i) = \sigma^2$).

- ($\mathrm{E}(\boldsymbol{x}_i \cdot \boldsymbol{\varepsilon}_i) = \mathbf{0}$ против $\mathrm{E}(\boldsymbol{\varepsilon}_i | \boldsymbol{x}_i) = 0$.) Иногда вместо условия ортогональности $\mathrm{E}(\boldsymbol{x}_i \cdot \varepsilon_i) = \mathbf{0}$ предполагается, что ошибки не связаны в том смысле, что $\mathrm{E}(\boldsymbol{\varepsilon}_i | \boldsymbol{x}_i) = 0$. Последнее условие более строгое, чем условие ортогональности, поскольку предполагает, что для любой (измеримой) функции f от $\boldsymbol{x}_i$, $f(\boldsymbol{x}_i)$ ортогональна ε_i:

$$\mathrm{E}[f(\boldsymbol{x}_i)\varepsilon_i] = \mathrm{E}[\mathrm{E}(f(\boldsymbol{x}_i)\varepsilon_i | \boldsymbol{x}_i)] = \mathrm{E}[f(\boldsymbol{x}_i)\,\mathrm{E}(\varepsilon_i | \boldsymbol{x}_i)] = 0.$$

  Это более строгое условие удовлетворяется в моделях рациональных ожиданий, но для целей разработки асимптотической теории нам нужно только более слабое предположение условия ортогональности.

[1]На самом деле, как только сделано предположение о независимости, те же самые результаты для больших выборок могут быть доказаны для более общего случая, в котором $\{y_i, \boldsymbol{x}_i\}$ независимы, но не одинаково распределены (i.n.i.d.), при условии, что удовлетворяются некоторые условия на более высокие моменты совместного распределения $(\varepsilon_i, \boldsymbol{x}_i)$. Мы не будем рассматривать это обобщение, потому что предположение i.i.d. выполняется в большинстве наборов микроданных.

- (Предопределенные против строго экзогенных регрессоров.) От регрессоров не требуется быть строго экзогенными. Как мы отмечали в параграфе 1.1, из предположения экзогенности (предположение 1.2) вытекает, что для k-го регрессора $\mathrm{E}(x_{jk}\varepsilon_i) = 0$ для всех i и j, а не только для $i = j$, что исключает возможность того, что текущая ошибка, ε_i, коррелирована с будущими регрессорами, $\boldsymbol{x}_{i+j}$ для $j \geq 1$. Предположение 2.3, ограничивающее только *одновременную* связь между ошибкой и регрессорами, не исключает этой возможности. Например, AR(1)-процесс, который не удовлетворяет предположению экзогенности классической регрессионной модели, может быть вмещен в модель этой главы. Это более слабое предположение о предопределенных регрессорах будет далее ослаблено в следующей главе.

- (Ранговое условие как отсутствие мультиколлинеарности в пределе.) Так как $\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')$ конечно по предположению 2.4, $\lim_{n\to\infty} \boldsymbol{S_{xx}} = \boldsymbol{\Sigma_{xx}}$ (где $\boldsymbol{S_{xx}} \equiv \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i'$) с вероятностью единица по эргодической теореме. Поэтому для достаточно больших n выборочный перекрестный момент регрессоров $\boldsymbol{S_{xx}}$, который может быть записан как $\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}$, невырожден, по предположениям 2.2 и 2.4. Так как $\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}$ невырождена, если и только если $\operatorname{rank}(\boldsymbol{X}) = K$, предположение 1.3 (отсутствие мультиколлинеарности) удовлетворяется с вероятностью единица для достаточно больших n. В OLS-формуле $\boldsymbol{b} = \boldsymbol{S_{xx}^{-1}}\boldsymbol{s_{xy}}$ (где $\boldsymbol{s_{xy}} \equiv \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \cdot y_i$) $\boldsymbol{S_{xx}}$ необходимо обращать. Если $\boldsymbol{S_{xx}}$ вырождена на конечных выборках (так что она не может быть обращена), мы просто присваиваем $\boldsymbol{b}$ произвольное значение так, чтобы OLS-оценка была определена для любой выборки.

- (Достаточное условие для того, чтобы $\{\boldsymbol{g}_i\}$ была m.d.s.) Так как m.d.s. (последовательность мартингал-разностей) имеет нулевое среднее по определению, предположение 2.5 более строгое, чем предположение 2.3. Предположение 2.5 будет нужно нам, чтобы доказать асимптотическую нормальность OLS-оценки. Предположение о произведении регрессоров и ошибки может быть сложно интерпретировать. Достаточное условие, которое интерпретировать легче, таково:

$$\mathrm{E}(\varepsilon_i|\varepsilon_{i-1}, \varepsilon_{i-2}, \ldots, \varepsilon_1, \boldsymbol{x}_i, \boldsymbol{x}_{i-1}, \ldots, \boldsymbol{x}_1) = 0. \tag{2.3.1}$$

Заметим, что в информационное множество включены как *текущие*, так и лагированные регрессоры. Из этого условия вытекает, что ошибка сериально некоррелирована, а также не коррелирована с текущими и прошлыми регрессорами (доказательство очень похоже на доказательство в предыдущем параграфе того, что m.d.s.

сериально некоррелирована). То, что (2.3.1) достаточно для того, чтобы $\{g_i\}$ был m.d.s., можно увидеть следующим образом. Мы имеем:

$$\begin{aligned}&\mathrm{E}(\boldsymbol{g}_i|\boldsymbol{g}_{i-1},\dots,\boldsymbol{g}_1) =\\ &= \mathrm{E}[\mathrm{E}(\boldsymbol{g}_i|\varepsilon_{i-1},\varepsilon_{i-2},\dots,\varepsilon_1,\boldsymbol{x}_i,\boldsymbol{x}_{i-1},\dots,\boldsymbol{x}_1)|\boldsymbol{g}_{i-1},\dots,\boldsymbol{g}_1].\end{aligned}$$

Это выполняется по закону повторных математических ожиданий, поскольку во «внутреннем» информационном множестве $(\varepsilon_{i-1},\varepsilon_{i-2},\dots,\varepsilon_1,\boldsymbol{x}_i,\boldsymbol{x}_{i-1},\dots,\boldsymbol{x}_1)$, содержится больше информации чем во «внешнем» информационном множестве $(\boldsymbol{g}_{i-1},\dots,\boldsymbol{g}_1)$. Следовательно,

$$\begin{aligned}&\mathrm{E}(\boldsymbol{g}_i|\boldsymbol{g}_{i-1},\dots,\boldsymbol{g}_1) =\\ &= \mathrm{E}[\boldsymbol{x}_i\,\mathrm{E}(\varepsilon_i|\varepsilon_{i-1},\varepsilon_{i-2},\dots,\varepsilon_1,\boldsymbol{x}_i,\boldsymbol{x}_{i-1},\dots,\boldsymbol{x}_1)|\boldsymbol{g}_{i-1},\dots,\boldsymbol{g}_1]\\ &\qquad \text{(в силу линейности условных математических ожиданий)}\\ &= 0 \quad \text{(из (2.3.1))}. \qquad (2.3.2)\end{aligned}$$

- (Когда регрессоры включают константу.) Практически во всех приложениях регрессоры включают константу. Если регрессоры включают константу, так что $x_{i1} = 1$ для всех i, то предположение 2.3 о предопределенных регрессорах может быть сформулировано более привычным образом: среднее ошибки нулевое (что вытекает из условия $\mathrm{E}(x_{ik}\varepsilon_i) = 0$ для $k = 1$), одновременная корреляция между компонентой ошибок и регрессорами нулевая (что вытекает из условий $\mathrm{E}(x_{ik}\varepsilon_i) = 0$ для $k \neq 1$ и $\mathrm{E}(\varepsilon_i) = 0$). Также, поскольку первый элемент K-мерного вектора $\boldsymbol{g}_i(\equiv \boldsymbol{x}_i \cdot \varepsilon_i)$ есть ε_i, из предположения 2.5 вытекает, что

$$\mathrm{E}(\varepsilon_i|\boldsymbol{g}_{i-1},\boldsymbol{g}_{i-2},\dots,\boldsymbol{g}_1) = 0.$$

Тогда, по закону повторных математических ожиданий, ε_i является скалярной m.d.s.:

$$\mathrm{E}(\varepsilon_i|\varepsilon_{i-1},\varepsilon_{i-2},\dots,\varepsilon_1) = 0. \qquad (2.3.3)$$

Следовательно, из предположения 2.5 вытекает, что ошибка сама является m.d.s. и, следовательно, сериально некоррелирована.

- ($\boldsymbol{S}$ является матрицей четвертых моментов.) Поскольку $\boldsymbol{g}_i \equiv \boldsymbol{x}_i \cdot \varepsilon_i$, матрица $\boldsymbol{S}$ в предположении 2.5 может быть записана как $\mathrm{E}(\varepsilon_i^2\boldsymbol{x}_i\boldsymbol{x}_i')$. Ее (k,j) элементом является $\mathrm{E}(\varepsilon_i^2 x_{ik}x_{ij})$. Поэтому $\boldsymbol{S}$ является матрицей четвертых моментов (математических ожиданий произведений четырех различных переменных). Состоятельное оценивание $\boldsymbol{S}$ будет требовать дополнительного предположения, которое будет специфицировано в параграфе 2.5.

- (Отсутствие предположения 2.5 приводит к другому выражению для матрицы $\boldsymbol{S}$.) Благодаря предположению о том, что $\{\boldsymbol{g}_i\}$ есть m.d.s., $\boldsymbol{S}(\equiv \mathrm{Avar}(\overline{\boldsymbol{g}}))$ равна $\mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_i')$. Без этого предположения, как мы увидим в главе 6, выражение для $\boldsymbol{S}$ будет более сложным и включает автоковариации $\boldsymbol{g}_i$.

Асимптотическое распределение OLS-оценки

Мы теперь докажем, что OLS-оценка является состоятельной и асимптотически нормальной. В остальной части этой главы следует иметь в виду, что OLS-оценка $\boldsymbol{b}$ зависит от размера выборки n (хотя эта зависимость не делается явной из-за нашего выбора не снабжать $\boldsymbol{b}$ индексом n) и что число регрессоров K остается фиксированным, когда мы отслеживаем последовательность OLS-оценок, снабженным индексом n. На данный момент мы предполагаем, что для $\boldsymbol{S}(\equiv \mathrm{Avar}(\overline{\boldsymbol{g}}) = \mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_i') = \mathrm{E}(\varepsilon_i^2\boldsymbol{x}_i\boldsymbol{x}_i'))$ доступна некоторая состоятельная оценка, обозначаемая как $\widehat{\boldsymbol{S}}$. Задача состоятельного оценивания $\boldsymbol{S}$ будет рассмотрена позже.

Утверждение 2.1 (асимптотическое распределение OLS-оценки):

(a) *(Состоятельность $\boldsymbol{b}$ для $\boldsymbol{\beta}$) При предположениях 2.1–2.4, имеет место равенство $\mathrm{plim}_{n\to\infty}\boldsymbol{b} = \boldsymbol{\beta}$. (Поэтому предположение 2.5 не является необходимым для состоятельности.)*

(b) *(Асимптотическая нормальность $\boldsymbol{b}$.) Если предположение 2.3 усилить до предположения 2.5, то тогда*

$$\sqrt{n}(\boldsymbol{b}-\boldsymbol{\beta}) \underset{d}{\to} N(\mathbf{0}, \mathrm{Avar}(\boldsymbol{b})) \quad \textit{при} \quad n \to \infty,$$

где

$$\mathrm{Avar}(\boldsymbol{b}) = \boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1}\,\boldsymbol{S}\,\boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1}. \tag{2.3.4}$$

(Напомним: $\boldsymbol{\Sigma}_{\boldsymbol{xx}} \equiv \mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')$, $\boldsymbol{S} = \mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_i')$, $\boldsymbol{g}_i \equiv \boldsymbol{x}_i \cdot \varepsilon_i$.)

(c) *(Состоятельная оценка $\mathrm{Avar}(\boldsymbol{b})$.) Предположим, что доступна состоятельная оценка $\widehat{\boldsymbol{S}}$ для $\boldsymbol{S}$ $(K \times K)$. Тогда при предположении 2.2, $\mathrm{Avar}(\boldsymbol{b})$ состоятельно оценивается посредством*

$$\widehat{\mathrm{Avar}(\boldsymbol{b})} = \boldsymbol{S}_{\boldsymbol{xx}}^{-1}\,\widehat{\boldsymbol{S}}\,\boldsymbol{S}_{\boldsymbol{xx}}^{-1}. \tag{2.3.5}$$

где $\boldsymbol{S}_{\boldsymbol{xx}}$ есть выборочное среднее от $\boldsymbol{x}_i\boldsymbol{x}_i'$:

$$\underset{(K\times K)}{\boldsymbol{S}_{\boldsymbol{xx}}} \equiv \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i' = \frac{1}{n}\boldsymbol{X}'\boldsymbol{X}. \tag{2.3.6}$$

Доказательство является ярким примером всех стандартных трюков в асимптотике. Для доказательства (а) и (b) будут использоваться три трюка: 1) записать рассматриваемый объект в терминах выборочных средних, 2) применить соответствующие LLN (Эргодическая теорема в данном контексте) и CLT (CLT для эргодической стационарной мартингал-разности) к выборочным средним и 3) использовать лемму 2.4(с) для получения асимптотического распределения. Доказательство части (с) не приводим, поскольку оно непосредственно следует из эргодической стационарности.

Доказательство (части (а) и (b)).

(1) Сначала мы запишем выборочную ошибку $\boldsymbol{b}-\boldsymbol{\beta}$ в терминах выборочных средних:

$$\begin{aligned}\boldsymbol{b}-\boldsymbol{\beta} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon} = \\ &= \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{\varepsilon}\right) = \\ &= \left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\cdot\varepsilon_i\right) \equiv \\ &\equiv \boldsymbol{S}_{\boldsymbol{xx}}^{-1}\overline{\boldsymbol{g}},\end{aligned} \tag{2.3.7}$$

где $$\overline{\boldsymbol{g}} \equiv \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{g}_i,\quad \boldsymbol{g}_i \equiv \boldsymbol{x}_i\cdot\varepsilon_i.$$

Выборочные средние $\boldsymbol{S}_{\boldsymbol{xx}}$ и $\overline{\boldsymbol{g}}$ зависят от размера выборки n, хотя обозначения не делают это явным.

(2) (Состоятельность.) Так как, по предположению 2.2, процесс $\{\boldsymbol{x}_i\boldsymbol{x}_i'\}$ является эргодически стационарным, $\boldsymbol{S}_{\boldsymbol{xx}} \to_{\mathrm{p}} \boldsymbol{\Sigma}_{\boldsymbol{xx}}$. (Сходимость здесь фактически почти наверное, но из сходимости почти наверное вытекает сходимость по вероятности.) Поскольку матрица $\boldsymbol{\Sigma}_{\boldsymbol{xx}}$ обратима, по предположению 2.4, $\boldsymbol{S}_{\boldsymbol{xx}}^{-1} \to_{\mathrm{p}} \boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1}$ по лемме 2.3(а). Аналогично $\overline{\boldsymbol{g}} \to_{\mathrm{p}} \mathrm{E}(\boldsymbol{g}_i)$, что, по предположению 2.3, есть $\boldsymbol{0}$. Поэтому, по лемме 2.3(а), $\boldsymbol{S}_{\boldsymbol{xx}}^{-1}\overline{\boldsymbol{g}} \to_{\mathrm{p}} \boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1}\boldsymbol{0} = \boldsymbol{0}$. Следовательно, $\operatorname{plim}_{n\to\infty}(\boldsymbol{b}-\boldsymbol{\beta}) = \boldsymbol{0}$, откуда вытекает, что $\operatorname{plim}_{n\to\infty}\boldsymbol{b} = \boldsymbol{\beta}$.

(3) (Асимптотическая нормальность.) Перепишем (2.3.7) в виде:

$$\sqrt{n}(\boldsymbol{b}-\boldsymbol{\beta}) = \boldsymbol{S}_{\boldsymbol{xx}}^{-1}(\sqrt{n}\,\overline{\boldsymbol{g}}). \tag{2.3.8}$$

Как упомянуто в формулировке предположения 2.5, $\sqrt{n}\overline{\boldsymbol{g}} \to_{\mathrm{d}} N(\boldsymbol{0}, \boldsymbol{S})$. Поэтому, по лемме 2.4(с), $\sqrt{n}(\boldsymbol{b}-\boldsymbol{\beta})$ сходится к нормальному распределению со средним $\boldsymbol{0}$ и дисперсией $\boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1}\boldsymbol{S}(\boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1})'$. Но так как матрица $\boldsymbol{\Sigma}_{\boldsymbol{xx}}$ является симметричной, это выражение равно (2.3.4). ■

Этот результат говорит о том, что распределение умноженной на $\sqrt{n}$ выборочной ошибки сколь угодно точно аппроксимируется нормальным распределением, когда размер выборки достаточно большой. Естественный вопрос заключается в том, насколько большой «большой»: насколько большим должен быть размер выборки, для того чтобы асимптотическая аппроксимация была обоснованной? Асимптотический результат, который мы только что получили, справедлив для всех DGP, удовлетворяющих предположениям модели. Однако размер выборки, необходимый для достижения той или иной меры близости к асимптотическому распределению, зависит от DGP. Мы затронем эту проблему в этой главе в эксперименте Монте-Карло.

s^2 состоятельна

Обратимся теперь к OLS-оценке дисперсии ошибок, s^2.

Утверждение 2.2 (состоятельная оценка дисперсии ошибок): *Пусть $e_i \equiv y_i - \boldsymbol{x}_i'\boldsymbol{b}$ есть OLS-остаток для наблюдения i. При предположениях 2.1–2.4,*

$$s^2 \equiv \frac{1}{n-K}\sum_{i=1}^{n} e_i^2 \underset{p}{\to} \mathrm{E}(\varepsilon_i^2),$$

при условии, что $\mathrm{E}(\varepsilon_i^2)$ существует и конечно.

Если бы мы могли наблюдать ошибку, ε_i, то тогда очевидной оценкой было бы выборочное среднее от ε_i^2. Она является состоятельной в силу эргодической стационарности. Основная идея утверждения 2.2 заключается в том, что замена настоящей ошибки, ε_i, OLS-остатком, e_i, не нарушает состоятельности. Давайте тщательно разберем пункт за пунктом схему доказательства, поскольку знание того, как справиться с расхождением между ε_i и его оценкой e_i, будет полезным и в других контекстах. Поскольку

$$s^2 = \frac{n}{n-K}\left(\frac{1}{n}\sum_{i=1}^{n} e_i^2\right),$$

достаточно показать, что выборочное среднее от e_i^2, $\frac{1}{n}\sum_{i=1}^{n} e_i^2$, сходится по вероятности к $\mathrm{E}(\varepsilon_i^2)$. Связь между e_i и ε_i задается как

$$\begin{aligned} e_i &\equiv y_i - \boldsymbol{x}_i'\boldsymbol{b} = \\ &= y_i - \boldsymbol{x}_i'\boldsymbol{\beta} - \boldsymbol{x}_i'(\boldsymbol{b}-\boldsymbol{\beta}) \quad \text{(добавляя и вычитая } \boldsymbol{x}_i'\boldsymbol{\beta}) \\ &= \varepsilon_i - \boldsymbol{x}_i'(\boldsymbol{b}-\boldsymbol{\beta}). \end{aligned} \tag{2.3.9}$$

так что

$$e_i^2 = \varepsilon_i^2 - 2(\boldsymbol{b}-\boldsymbol{\beta})'\boldsymbol{x}_i \cdot \varepsilon_i + (\boldsymbol{b}-\boldsymbol{\beta})'\boldsymbol{x}_i\boldsymbol{x}_i'(\boldsymbol{b}-\boldsymbol{\beta}). \tag{2.3.10}$$

Суммируя по i, мы получаем:

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} e_i^2 = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 - 2(\boldsymbol{b}-\boldsymbol{\beta})'\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\cdot\varepsilon_i &+ \\ + (\boldsymbol{b}-\boldsymbol{\beta})'\Big(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\Big)(\boldsymbol{b}-\boldsymbol{\beta}) &= \\ = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 - 2(\boldsymbol{b}-\boldsymbol{\beta})'\overline{\boldsymbol{g}} + (\boldsymbol{b}-\boldsymbol{\beta})'\boldsymbol{S}_{\boldsymbol{xx}}(\boldsymbol{b}-\boldsymbol{\beta}).& \qquad (2.3.11)\end{aligned}$$

Остаток доказательства, который должен показать, что plim двух последних слагаемых равны нулю, так что $\operatorname{plim}\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 = \operatorname{plim}\frac{1}{n}\sum_{i=1}^{n}e_i^2$, остается в качестве контрольного вопроса. Если вы будете рассматривать этот вопрос, вам будет ясно, что все, что требуется от оценки коэффициента, — это ее состоятельность; если для формирования остатков используется другая состоятельная оценка, а не OLS-оценка $\boldsymbol{b}$, указанная оценка дисперсии ошибок остается состоятельной для $\mathrm{E}(\varepsilon_i^2)$.

Контрольные вопросы

1. Предположим, что $\mathrm{E}(y_i|\boldsymbol{x}_i) = \boldsymbol{x}_i'\boldsymbol{\beta}$, то есть предположим, что регрессия y_i на $\boldsymbol{x}_i$ является линейной функцией по $\boldsymbol{x}_i$. Определим $\varepsilon_i \equiv y_i - \boldsymbol{x}_i'\boldsymbol{\beta}$. Покажите, что $\boldsymbol{x}_i$ ортогонален ε_i. **Указание:** *Сначала покажите, что* $\mathrm{E}(\varepsilon_i|\boldsymbol{x}_i) = 0$.

2. (Предполагается ли $\mathrm{E}(\varepsilon_i^2)$ конечным?)

   (a) Вытекает ли из предположений 2.1–2.5, что $\mathrm{E}(\varepsilon_i^2)$ существует и конечно? **Указание:** *Вторые моменты строго стационарного процесса могут не быть конечными.*

   (b) Если один из регрессоров в нашей модели является константой, то дисперсия ошибки конечна. Докажите это. **Указание:** *Если* $x_{i1} = 1$, *то элемент* $(1,1)$ *матрицы* $\boldsymbol{g}_i\boldsymbol{g}_i'$ *равен* ε_i^2.

3. (Альтернативное выражение для S.) Пусть $f(\boldsymbol{x}_i) \equiv \mathrm{E}(\varepsilon_i^2|\boldsymbol{x}_i)$. Покажите, что $S(\equiv \mathrm{E}(\varepsilon_i^2\boldsymbol{x}_i\boldsymbol{x}_i'))$ можно записать как

$$\boldsymbol{S} = \mathrm{E}[f(\boldsymbol{x}_i)\boldsymbol{x}_i\boldsymbol{x}_i'].$$

   Указание: *закон полных математических ожиданий.*

4. Завершите доказательство утверждения 2.2. **Указание:** *Мы уже доказали для утверждения 2.1, что* $\operatorname{plim}\overline{\boldsymbol{g}} = \boldsymbol{0}$, $\operatorname{plim}\boldsymbol{S}_{\boldsymbol{xx}} = \boldsymbol{\Sigma}_{\boldsymbol{xx}}$ *и* $\operatorname{plim}(\boldsymbol{b}-\boldsymbol{\beta}) = \boldsymbol{0}$ *при предположениях 2.1–2.4. Используйте лемму 2.3(a), чтобы показать, что* $\operatorname{plim}(\boldsymbol{b}-\boldsymbol{\beta})'\overline{\boldsymbol{g}} = 0$ *и* $\operatorname{plim}(\boldsymbol{b}-\boldsymbol{\beta})'\boldsymbol{S}_{\boldsymbol{xx}}(\boldsymbol{b}-\boldsymbol{\beta}) = 0$.

5. (Утверждение 2.2 с состоятельной $\widehat{\boldsymbol{\beta}}$.) Докажите следующее обобщение утверждения 2.2:

Пусть $\widehat{\varepsilon}_i \equiv y_i - \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}$, где $\widehat{\boldsymbol{\beta}}$ — любая состоятельная оценка $\boldsymbol{\beta}$. При предположениях 2.1, 2.2 и предположении того, что $\mathrm{E}(\boldsymbol{x}_i \cdot \varepsilon_i)$ и $\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')$ конечны, $\frac{1}{n}\sum_i \widehat{\varepsilon}_i^2 \to_{\mathrm{p}} \mathrm{E}(\varepsilon_i^2)$.

Поэтому регрессоры не обязаны быть ортогональными ошибке.

2.4. Тестирование гипотез

Статистические выводы в теории больших выборок основаны на тестовых статистиках, асимптотическое распределение которых известно в случае правильности нулевой гипотезы. Получить распределения тестовых статистик легче, чем в теории конечных выборок, поскольку нас интересует только *аппроксимация* точного распределения на больших выборках. В этом параграфе мы получаем тестовые статистики, предполагая всюду, что для $\boldsymbol{S}(\equiv \mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_i'))$ доступна состоятельная оценка, $\widehat{\boldsymbol{S}}$. Проблема состоятельного оценивания $\boldsymbol{S}$ будет рассмотрена в следующем параграфе.

Тестирование линейных гипотез

Рассмотрим тестирование гипотезы о значении k-го коэффициента β_k. Из утверждения 2.1 вытекает, что при $H_0 : \beta_k = \overline{\beta}_k$

$$\sqrt{n}(b_k - \overline{\beta}_k) \underset{\mathrm{d}}{\to} N\left(0, \mathrm{Avar}(b_k)\right) \quad \text{и} \quad \widehat{\mathrm{Avar}(b_k)} \underset{\mathrm{p}}{\to} \mathrm{Avar}(b_k).$$

где b_k есть k-й элемент $\boldsymbol{b}$ и $\mathrm{Avar}(b_k)$ есть (k, k)-й элемент $K \times K$ матрицы $\mathrm{Avar}(\boldsymbol{b})$. Поэтому лемма 2.4(c) гарантирует, что

$$t_k \equiv \frac{\sqrt{n}(b_k - \overline{\beta}_k)}{\sqrt{\widehat{\mathrm{Avar}(b_k)}}} = \frac{b_k - \overline{\beta}_k}{SE^*(b_k)} \underset{\mathrm{d}}{\to} N(0, 1), \tag{2.4.1}$$

где

$$SE^*(b_k) \equiv \sqrt{\frac{1}{n} \cdot \widehat{\mathrm{Avar}(b_k)}} \equiv \sqrt{\frac{1}{n} \cdot \left(\boldsymbol{S}_{\boldsymbol{xx}}^{-1}\,\widehat{\boldsymbol{S}}\,\boldsymbol{S}_{\boldsymbol{xx}}^{-1}\right)_{kk}}.$$

Знаменатель этого t-отношения, $SE^*(b_k)$, называется **состоятельной при гетероскедастичности стандартной ошибкой** (*Heteroskedasticity-consistent standard error*), **робастной (к гетероскедастичности) стандартной ошибкой** ((*heteroskedasticity-*)*robust standard error*), или **стандартной ошибкой в форме Уайта** (*White's standard error*). Причиной для такой терминологии является то, что ошибка может быть условно гетероскедастичной; вспомните, что мы не предполагали условную гомоскедастичность (то есть что $\mathrm{E}(\varepsilon_i^2|\boldsymbol{x}_i)$ не зависит от $\boldsymbol{x}_i$) для получения асимптотического распределения t_k. Это t-отношение называется **робастным** (*robust*) t-отношением, чтобы отличать его от t-отношения из главы 1. Связь между этими двумя видами t-отношения будет обсуждаться в параграфе 2.6. Тестирование нулевой гипотезы $H_0 : \beta_k = \overline{\beta}_k$ на уровне значимости α, опирающееся на это t-отношение, осуществляется следующим образом:

Шаг 1: Вычисляем t_k по формуле (2.4.1).

Шаг 2: Смотрим в таблицу $N(0, 1)$, чтобы найти критическое значение $t_{\alpha/2}$, которое оставляет $\alpha/2$ верхнему хвосту стандартного нормального распределения[1]. (Пример: если $\alpha = 5\%$, $t_{\alpha/2} = 1.96$.)

Шаг 3: Нулевую гипотезу принимаем, если $|t_k| < t_{\alpha/2}$; в противном случае отвергаем ее.

Отличия от t-теста на конечных выборках: 1) другой способ вычисления стандартной ошибки, 2) мы используем таблицу $N(0, 1)$, а не $t(n - K)$ и 3) **фактический размер** (*actual size*), или **точный размер** (*exact size*), теста (вероятность ошибки I рода при заданном объеме выборки) равен **номинальному размеру** (*nominal size*) (то есть желаемому уровню значимости α) только приблизительно, хотя аппроксимация становится сколь угодно хорошей с увеличением объема выборки. Разница между точным размером и номинальным размером теста называется **искажением размера** (*size distortion*). Поскольку t_k асимптотически стандартная нормальная, искажение размера t-теста сходится к нулю при стремлении объема выборки n к бесконечности.

Таким образом, мы доказали первую половину следующего утверждения:

Утверждение 2.3 (робастное t-отношение и статистика Вальда): *Предположим, что выполняются предположения 2.1–2.5, и предположим, что существует доступная состоятельная оценка $\widehat{\boldsymbol{S}}$ для $\boldsymbol{S}$ ($= \mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_i')$). Как и прежде, пусть*

$$\widehat{\mathrm{Avar}(\boldsymbol{b})} \equiv \boldsymbol{S}_{\boldsymbol{xx}}^{-1}\, \widehat{\boldsymbol{S}}\, \boldsymbol{S}_{\boldsymbol{xx}}^{-1}.$$

Тогда

(a) *При нулевой гипотезе H_0: $\beta_k = \overline{\beta}_k$, t_k, определенная в формуле (2.4.1), $\to_d N(0, 1)$.*

(b) *При нулевой гипотезе H_0: $\boldsymbol{R\beta} = \boldsymbol{r}$, где $\boldsymbol{R}$ есть $\#\boldsymbol{r} \times K$ матрица ($\#\boldsymbol{r}$ — размерность $\boldsymbol{r}$, число ограничений) полного строкового ранга,*

$$W \equiv n \cdot (\boldsymbol{Rb} - \boldsymbol{r})'\{\boldsymbol{R}[\widehat{\mathrm{Avar}(\boldsymbol{b})}]\boldsymbol{R}'\}^{-1}(\boldsymbol{Rb} - \boldsymbol{r}) \underset{d}{\to} \chi^2(\#\boldsymbol{r}). \qquad (2.4.2)$$

[1]Обратите внимание на то, что здесь $t_{\alpha/2}$ — это не квантиль уровня $\alpha/2$, а квантиль уровня $1 - \alpha/2$. — *Прим. науч. ред. перевода.*

Что остается показать — так это то, что $W \to_{\mathrm{d}} \chi^2(\#\boldsymbol{r})$, а это есть непосредственное применение леммы 2.4(d).

Доказательство (продолжение). Запишем W как

$$W = \boldsymbol{c}_n' \boldsymbol{Q}_n^{-1} \boldsymbol{c}_n, \text{ где } \quad \boldsymbol{c}_n \equiv \sqrt{n}(\boldsymbol{R}\boldsymbol{b} - \boldsymbol{r}) \text{ и } \boldsymbol{Q}_n \equiv \boldsymbol{R}\widehat{\mathrm{Avar}(\boldsymbol{b})}\boldsymbol{R}'.$$

При H_0 имеем: $\boldsymbol{c}_n = \boldsymbol{R}\sqrt{n}(\boldsymbol{b} - \boldsymbol{\beta})$. Поэтому, по предположению 2.1,

$$\boldsymbol{c}_n \underset{\mathrm{d}}{\to} \boldsymbol{c}, \text{ где } \quad \boldsymbol{c} \sim N(\boldsymbol{0}, \boldsymbol{R}\,\mathrm{Avar}(\boldsymbol{b})\boldsymbol{R}').$$

Также по предположению 2.1

$$\boldsymbol{Q}_n \underset{\mathrm{d}}{\to} \boldsymbol{Q}, \text{ где } \quad \boldsymbol{Q} \equiv \boldsymbol{R}\,\mathrm{Avar}(\boldsymbol{b})\boldsymbol{R}'.$$

Поскольку $\boldsymbol{R}$ имеет полный строковый ранг и $\mathrm{Avar}(\boldsymbol{b})$ положительно определена, $\boldsymbol{Q}$ обратима. Следовательно, по лемме 2.4(d),

$$W \underset{\mathrm{d}}{\to} \boldsymbol{c}'\boldsymbol{Q}^{-1}\boldsymbol{c}.$$

Так как $\#\boldsymbol{r}$-мерный случайный вектор $\boldsymbol{c}$ распределен нормально со средним $\boldsymbol{0}$ и так как $\boldsymbol{Q}$ равна $\mathrm{Var}(\boldsymbol{c})$, $\boldsymbol{c}'\boldsymbol{Q}^{-1}\boldsymbol{c} \sim \chi^2(\#\boldsymbol{r})$. ■

Эта статистика хи-квадрат W является статистикой Вальда, поскольку она основана на неограниченных оценках (здесь это $\boldsymbol{b}$ и $\widehat{\mathrm{Avar}(\boldsymbol{b})}$), не ограниченных нулевой гипотезой H_0. Тестирование H_0 на уровне значимости α осуществляется следующим образом.

Шаг 1: Вычисляем статистику W по формуле (2.4.2).

Шаг 2: Смотрим в таблицу распределения $\chi^2(\#\boldsymbol{r})$, чтобы найти критические значения $\chi^2_\alpha(\#\boldsymbol{r})$, которые дают α верхнему хвосту распределения $\chi^2(\#\boldsymbol{r})$.

Шаг 3: Если $W < \chi^2(\#\boldsymbol{r})$, то H_0 принимаем; в противном случае отвергаем ее.

Вероятность ошибки I рода приближается к α при увеличении выборки. Как будет ясно из параграфа 2.6, эта статистика Вальда тесно связана с известным F-тестом при условной гомоскедастичности.

Тест является состоятельным

Вспомним из основ статистики, что **мощность** (*power*) теста на конечных выборках есть вероятность отвергнуть нулевую гипотезу, когда она ложна, при заданной выборке конечного объема (то есть мощность есть 1 минус вероятность ошибки II рода). Мощность будет, очевидно, зависеть от DGP (то есть фактически сгенерированных данных), рассмотренного как альтернатива, так же как и от размера (уровня значимости) теста. Для примера рассмотрим DGP $\{y_i, \boldsymbol{x}_i\}$, удовлетворяющий предположениям 2.1–2.5, но *не* нулевой гипотезе $H_0 : \beta_k = \overline{\beta}_k$. Мощность t-теста с размером α против этой альтернативы есть

$$\text{мощность} = \Pr(|t_k| > t_{\alpha/2}),$$

которая зависит от DGP в вопросе, поскольку DGP контролирует распределение t_k. Мы говорим, что тест является **состоятельным** (*consistent*) против множества DGP, ни один из которых не удовлетворяет нулевой гипотезе, если его мощность против любого конкретного элемента этого множества приближается к единице при $n \to \infty$ для любого предполагаемого уровня значимости.

То, что t-тест является состоятельным против множества альтернатив (DGP), удовлетворяющих предположениям 2.1–2.5, можно увидеть из следующего. Посмотрим на воспроизводимое здесь выражение (2.4.1) для t-отношения:

$$t_k \equiv \frac{\sqrt{n}(b_k - \overline{\beta}_k)}{\sqrt{\widehat{\text{Avar}(b_k)}}}.$$

Знаменатель, с одной стороны, сходится к $\sqrt{\text{Avar}(b_k)}$ вопреки тому, что DGP не удовлетворяет нулевой гипотезе (вспомним, что все части утверждения 2.1 выполняются независимо от того, верна ли нулевая гипотеза, при условии выполнения предположений 2.1–2.5). С другой стороны, числитель стремится к $+\infty$ или $-\infty$, поскольку b_k сходится по вероятности к β_k из DGP, отличающемуся от $\overline{\beta}_k$. Поэтому мощность стремится к единице при стремлении объема выборки n к бесконечности, откуда вытекает, что t-тест в утверждении 2.3 состоятелен против тех альтернатив, DGP, которые не удовлетворяют нулевой гипотезе. То же самое верно для теста Вальда.

Асимптотическая мощность

Для последующего использования в следующей главе мы определяем здесь **асимптотическую мощность состоятельного теста** (*asymptotic power of a consistent test*). Как замечено выше, мощность t-теста приближается к единице при увеличении объема выборки, в то время как DGP, взятый в качестве альтернативы, остается фиксированным. Но если при увеличении объема выборки DGP становится все ближе и ближе к нулевой гипотезе, мощность может не сходиться к единице. Последовательность таких DGPs называется **последовательностью локальных альтернатив** (*sequence of local alternatives*). Для регрессионной модели и для нулевой гипотезы $H_0 : \beta_k = \overline{\beta}_k$ это последовательность DGP, такая, что (i) n-й DGP $\{y_i^{(n)}, \boldsymbol{x}_i^{(n)}\}$ $(i = 1, 2, \ldots)$ удовлетворяет предположениям 2.1–2.5 и сходится в обычном смысле к фиксированному DGP $\{y_i, \boldsymbol{x}_i\}$[1] и (ii) значение β_k n-го DGP, $\beta_k^{(n)}$, сходится к $\overline{\beta}_k$. Предположим далее, что $\beta_k^{(n)}$ удовлетворяет

$$\beta_k^{(n)} = \overline{\beta}_k + \frac{\gamma}{\sqrt{n}} \tag{2.4.3}$$

для некоторого заданного $\gamma \neq 0$. Таким образом, $\beta_k^{(n)}$ приближается к $\overline{\beta}_k$ со скоростью, пропорциональной $1/\sqrt{n}$. Этот частный случай последовательности локальных альтернатив называется **Питменовским сносом** (*Pitman drift*) или **Питменовской последовательностью** (*Pitman sequence*). Подставляя (2.4.3) в (2.4.1), t-отношение выше можно записать как

$$t_k = \frac{\sqrt{n}(b_k - \beta_k^{(n)})}{\sqrt{\widehat{\mathrm{Avar}(b_k)}}} + \frac{\gamma}{\sqrt{\widehat{\mathrm{Avar}(b_k)}}}. \tag{2.4.4}$$

Если выборка объема n порождается n-м DGP Питменовского сноса, сходится ли t_k к нетривиальному распределению? Поскольку n-й DGP удовлетворяет предположениям 2.1–2.5, первая компонента в правой части (2.2.4) сходится по распределению к $N(0, 1)$ согласно частям (b) и (c) утверждения 2.1. В силу части (c) утверждения 2.1 и того факта, что $\{y_i^{(n)}, \boldsymbol{x}_i^{(n)}\}$ «сходится» к фиксированному DGP, вторая компонента сходится по вероятности к

$$\mu \equiv \frac{\gamma}{\sqrt{\mathrm{Avar}(b_k)}}, \tag{2.4.5}$$

где $\mathrm{Avar}(b_k)$ вычисляется при фиксированном DGP. Следовательно, $t_k \to_{\mathrm{d}} N(\mu, 1)$ вдоль этой последовательности локальных альтернатив. Если уровень значимости равен α, мощность сходится к

$$\mathrm{Prob}(|x| > t_{\alpha/2}). \tag{2.4.6}$$

где $x \sim N(\mu, 1)$ и $t_{\alpha/2}$ — критическое значение для уровня значимости α. Эта вероятность называется **асимптотической мощностью** (*asymptotic power*). Она является мерой способности теста обнаруживать малые отклонения модели от нулевой гипотезы. Очевидно, что чем больше $|\mu|$, тем выше асимптотическая мощность для заданного размера α. Исходя из аналогичных аргументов, легко показать, что статистика Вальда сходится к распределению, называемому **нецентральным хи-квадрат-распределением** (*noncentral chi-squared*).

[1] См., например, предположение 1 в [Newey, 1985] для точной формулировки.

Тестирование нелинейных гипотез

Статистика Вальда может быть обобщена для тестирования множества нелинейных ограничений на $\boldsymbol{\beta}$. Рассмотрим нулевую гипотезу в форме:

$$H_0 : \boldsymbol{a}(\boldsymbol{\beta}) = \mathbf{0}.$$

Здесь $\boldsymbol{a}$ — векторозначная функция с непрерывными первыми производными. Пусть $\#\boldsymbol{a}$ — размерность $\boldsymbol{a}(\boldsymbol{\beta})$ (так что нулевая гипотеза имеет $\#\boldsymbol{a}$ ограничений), и $\boldsymbol{A}(\boldsymbol{\beta})$ является $\#\boldsymbol{a} \times K$-матрицей первых производных, вычисленных в $\boldsymbol{\beta}$: $\boldsymbol{A}(\boldsymbol{\beta}) = \partial\boldsymbol{a}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}'$. Для того чтобы указанная гипотеза была определена, мы предполагаем, что $\boldsymbol{A}(\boldsymbol{\beta})$ имеет полный строковый ранг (это является обобщением требования линейной гипотезы $\boldsymbol{R\beta} = \boldsymbol{r}$ о том, что $\boldsymbol{R}$ имеет полный строковый ранг). Из леммы 2.5 параграфа 2.1 и утверждения 2.1(b) вытекает, что

$$\sqrt{n}[\boldsymbol{a}(\boldsymbol{b}) - \boldsymbol{a}(\boldsymbol{\beta})] \underset{\mathrm{d}}{\to} \boldsymbol{c}, \ \boldsymbol{c} \sim N\big(\mathbf{0}, \boldsymbol{A}(\boldsymbol{\beta}) \operatorname{Avar}(\boldsymbol{b}) \boldsymbol{A}(\boldsymbol{\beta})'\big). \tag{2.4.7}$$

Так как $\boldsymbol{a}(\boldsymbol{\beta}) = \mathbf{0}$ при H_0, (2.4.7) принимает вид:

$$\sqrt{n}\boldsymbol{a}(\boldsymbol{b}) \underset{\mathrm{d}}{\to} \boldsymbol{c}, \ \boldsymbol{c} \sim N\big(\mathbf{0}, \boldsymbol{A}(\boldsymbol{\beta}) \operatorname{Avar}(\boldsymbol{b}) \boldsymbol{A}(\boldsymbol{\beta})'\big). \tag{2.4.8}$$

Поскольку $\boldsymbol{b} \to_{\mathrm{p}} \boldsymbol{\beta}$ по утверждению 2.1(a), лемма 2.3(a) влечет $\boldsymbol{A}(\boldsymbol{b}) \to_{\mathrm{p}} \boldsymbol{A}(\boldsymbol{\beta})$. По предположению 2.1(c), $\widehat{\operatorname{Avar}(\boldsymbol{b})} \to_{\mathrm{p}} \operatorname{Avar}(\boldsymbol{b})$. Поэтому, по лемме 2.3(a),

$$\boldsymbol{A}(\boldsymbol{b})\widehat{\operatorname{Avar}(\boldsymbol{b})}\, \boldsymbol{A}(\boldsymbol{b})' \underset{\mathrm{p}}{\to} \boldsymbol{A}(\boldsymbol{\beta}) \operatorname{Avar}(\boldsymbol{b})\, \boldsymbol{A}(\boldsymbol{\beta})' = \operatorname{Var}(\boldsymbol{c}). \tag{2.4.9}$$

Поскольку $\boldsymbol{A}(\boldsymbol{\beta})$ имеет полный строковый ранг и $\operatorname{Avar}(\boldsymbol{b})$ положительно определена, $\operatorname{Var}(\boldsymbol{c})$ обратима. Тогда лемма 2.4(d), (2.4.8) и (2.4.9) влекут

$$\sqrt{n}\, \boldsymbol{a}(\boldsymbol{b})'\{\boldsymbol{A}(\boldsymbol{b})\widehat{\operatorname{Avar}(\boldsymbol{b})}\boldsymbol{A}(\boldsymbol{b})'\}^{-1}\sqrt{n}\, \boldsymbol{a}(\boldsymbol{b}) \underset{\mathrm{d}}{\to} \boldsymbol{c}' \operatorname{Var}(\boldsymbol{c})^{-1}\boldsymbol{c} \sim \chi^2(\#\boldsymbol{a}). \tag{2.4.10}$$

Объединяя два $\sqrt{n}$ в (2.4.10) в единственный n, мы доказали

Утверждение 2.3 (продолжение):

(c) *При нулевой гипотезе с $\#\boldsymbol{a}$ ограничениями $H_0 : \boldsymbol{a}(\boldsymbol{\beta}) = \mathbf{0}$, такой, что $\boldsymbol{A}(\boldsymbol{\beta})$, матрица размера $\#\boldsymbol{a} \times K$ непрерывных первых производных $\boldsymbol{a}(\boldsymbol{\beta})$, имеет полный строковый ранг, мы имеем:*

$$W \equiv n \cdot \boldsymbol{a}(\boldsymbol{b})'\{\boldsymbol{A}(\boldsymbol{b})\widehat{\operatorname{Avar}(\boldsymbol{b})}\, \boldsymbol{A}(\boldsymbol{b})'\}^{-1}\boldsymbol{a}(\boldsymbol{b}) \underset{d}{\to} \chi^2(\#\boldsymbol{a}). \tag{2.4.11}$$

Часть (c) является обобщением (b): при $\boldsymbol{a}(\boldsymbol{b}) = \boldsymbol{R\beta} - \boldsymbol{r}$, (2.4.11) сводится к (2.4.2), статистике Вальда для линейных ограничений.

Выбор $\boldsymbol{a}(\cdot)$ для представления заданного множества ограничений не является единственным. Например, равенство $\beta_1\beta_2 = 1$ можно записать как $\alpha(\boldsymbol{\beta}) = 0$ с $\alpha(\boldsymbol{\beta}) = \beta_1\beta_2 - 1$ или с $\alpha(\boldsymbol{\beta}) = \beta_1 - 1/\beta_2$. В то время как часть (c) утверждения гарантирует, что на больших выборках результат теста Вальда одинаковый независимо от выбора функции $\boldsymbol{a}$, численное значение статистики Вальда W в действительности зависит от выбранного представления, и результат теста может быть различным на конечных выборках. В вышеприведенном примере второе представление, $\alpha(\boldsymbol{\beta}) = \beta_1 - 1/\beta_2$, не удовлетворяет требованию непрерывности производных при $\beta_2 = 0$. Действительно, исследование Монте-Карло [Gregory and Veall, 1985] показывает, что, когда β_2 близко к нулю, тест Вальда, основанный на втором представлении, отвергает нулевую гипотезу на малых выборках слишком часто.

Контрольные вопросы

1. Верно ли, что $SE^*(b_k) \to_{\mathrm{p}} 0$ при $n \to \infty$?
2. (Стандартная ошибка нелинейной функции.) Для простоты пусть $K = 1$ и пусть b — OLS-оценка для β. Стандартной ошибкой для b является $\sqrt{\widehat{\mathrm{Avar}(b)}/n}$. Предположим, что $\lambda = -\log(\beta)$. Оценка λ, порождаемая OLS-оценкой для β, имеет вид: $\widehat{\lambda} = -\log(b)$. Проверьте, что стандартная ошибка $\widehat{\lambda}$ равна $(1/b) \cdot \sqrt{\widehat{\mathrm{Avar}(b)}/n}$.
3. (Инвариантность [или ее отсутствие] статистики Вальда.) Не существует однозначного способа записать линейную гипотезу $\boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r}$, поскольку для любой невырожденной матрицы $\boldsymbol{F}$ размера $\#\boldsymbol{r} \times \#\boldsymbol{r}$ то же самое множество ограничений может быть представлено как $\widetilde{\boldsymbol{R}}\boldsymbol{\beta} = \widetilde{\boldsymbol{r}}$ с $\widetilde{\boldsymbol{R}} \equiv \boldsymbol{F}\boldsymbol{R}$ и $\widetilde{\boldsymbol{r}} \equiv \boldsymbol{F}\boldsymbol{r}$. Влияют ли различные выборы $\boldsymbol{R}$ и r на асимптотическое распределение W? На распределение на конечных выборках? На численное значение?

2.5. Состоятельное оценивание $\mathrm{E}(\varepsilon_i^2 \boldsymbol{x}_i \boldsymbol{x}_i')$

Теория, разработанная ранее, предполагает, что существует доступная состоятельная оценка, $\widehat{\boldsymbol{S}}$, для $\boldsymbol{S}(= \mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_i') = \mathrm{E}(\varepsilon_i^2 \boldsymbol{x}_i \boldsymbol{x}_i'))$, которая будет использоваться для вычисления оцененной асимптотической дисперсии, $\widehat{\mathrm{Avar}}\,\boldsymbol{b}$. Этот параграф объясняет, как получить $\widehat{\boldsymbol{S}}$ по выборке $(\boldsymbol{y}, \boldsymbol{X})$.

Использование остатков вместо ошибок

Если бы ошибка была наблюдаема, то тогда выборочное среднее $\varepsilon_i^2 \boldsymbol{x}_i \boldsymbol{x}_i'$, очевидно, было бы состоятельно в силу эргодической стационарности. Но мы не наблюдаем ошибку, и подстановка вместо нее некоторой ее состоятельной оценки дает:

$$\widehat{\boldsymbol{S}} \equiv \frac{1}{n}\sum_{i=1}^{n} \widehat{\varepsilon}_i^2 \, \boldsymbol{x}_i \boldsymbol{x}_i', \tag{2.5.1}$$

где $\widehat{\varepsilon}_i \equiv y_i - \boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}$, и $\widehat{\boldsymbol{\beta}}$ — некоторая состоятельная оценка $\boldsymbol{\beta}$. (Хотя очевидным кандидатом для состоятельной оценки $\widehat{\boldsymbol{\beta}}$ является OLS-оценка $\boldsymbol{b}$, мы используем здесь $\widehat{\boldsymbol{\beta}}$, а не $\boldsymbol{b}$, чтобы отметить, что результаты этого параграфа справедливы для любой состоятельной оценки.) Чтобы эта оценка была состоятельной для $\boldsymbol{S}$, нам нужно сделать предположение относительно четвертых моментов регрессоров.

Предположение 2.6 (конечные четвертые моменты регрессоров): $\mathrm{E}[(x_{ik}x_{ij})^2]$ *существует и конечно для всех* k, j $(= 1, 2, \ldots, K)$.

Утверждение 2.4 (состоятельная оценка для S): *Предположим, что оценка коэффициента* $\widehat{\boldsymbol{\beta}}$*, использованная для вычисления остатков* $\widehat{\varepsilon}_i$ *для* $\widehat{\boldsymbol{S}}$ *в* (2.5.1)*, является состоятельной, и предположим, что* $\boldsymbol{S} = \mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_i')$ *существует и конечно. Тогда при предположениях 2.1, 2.2 и 2.6* $\widehat{\boldsymbol{S}}$*, заданная в* (2.5.1)*, является состоятельной оценкой для* $\boldsymbol{S}$.

Чтобы указать, почему необходимо предположение о четвертом моменте для регрессоров, мы даем набросок доказательства для частного случая $K = 1$ (только один регрессор). Так что $\boldsymbol{x}_i$ сейчас скаляр x_i, $\boldsymbol{g}_i$ — скаляр $g_i = x_i\varepsilon_i$ и (2.3.10) (с $\boldsymbol{b} = \widehat{\boldsymbol{\beta}}$ и $e_i = \widehat{\varepsilon}_i$) упрощается до

$$\widehat{\varepsilon}_i^2 = \varepsilon_i^2 - 2(\widehat{\beta} - \beta)x_i\varepsilon_i + (\widehat{\beta} - \beta)^2 x_i^2. \tag{2.5.2}$$

Домножая обе стороны на x_i^2 и суммируя по i,

$$\frac{1}{n}\sum_{i=1}^{n} \widehat{\varepsilon}_i^2 x_i^2 - \frac{1}{n}\sum_{i=1}^{n} \varepsilon_i^2 x_i^2 = -2(\widehat{\beta} - \beta)\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i x_i^3 + (\widehat{\beta} - \beta)^2 \frac{1}{n}\sum_{i=1}^{n} x_i^4. \tag{2.5.3}$$

Сейчас мы можем видеть, почему требуется предположение о конечном четвертом моменте для x_i: если четвертый момент $\mathrm{E}(x_i^4)$ конечен, то тогда по эргодической стационарности выборочное среднее x_i^4 сходится по вероятности к некоторому конечному числу, поэтому последний член в (2.5.3) исчезает (сходится к 0 по вероятности), если $\widehat{\beta}$ состоятельна для β. Можно также показать (см. аналитическое упражнение 4 для доказательства), комбинируя то же предположение о четвертых моментах регрессоров и предположение о конечности $\mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_i')(= \mathrm{E}(\varepsilon_i^2\boldsymbol{x}_i\boldsymbol{x}_i'))$, что выборочное среднее $x_i^3\varepsilon_i$ сходится по вероятности к некоторому конечному числу, так что другая составляющая в правой части (2.5.3) также исчезает.

Согласно утверждению 2.1(а), предположения, сделанные в утверждении 2.3, являются достаточными, чтобы гарантировать состоятельность $\boldsymbol{b}$, поэтому мы можем положить $\widehat{\boldsymbol{\beta}} = \boldsymbol{b}$ в (2.5.1) и использовать OLS-остатки для вычисления $\widehat{\boldsymbol{S}}$. К тому же предположение, сделанное в утверждении 2.4, о том, что $\mathrm{E}(\boldsymbol{g}_i \boldsymbol{g}_i')$ конечно, является частью предположения 2.5, которое предполагается выполненным в утверждении 2.3. Следовательно, подразумеваемый смысл утверждения 2.4:

Если к гипотезе в утверждении 2.3 добавить предположение 2.6, то тогда $\widehat{\boldsymbol{S}}$, заданная в (2.5.1) *с* $\boldsymbol{b} = \widehat{\boldsymbol{\beta}}$ *(так что $\widehat{\varepsilon}_i$ — OLS-остатки e_i) может быть использована в* (2.3.5) *для вычисления оцененной асимптотической дисперсии:*

$$\widehat{\mathrm{Avar}(\boldsymbol{b})} = \boldsymbol{S}_{\boldsymbol{xx}}^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} e_i^2 \boldsymbol{x}_i \boldsymbol{x}_i' \right) \boldsymbol{S}_{\boldsymbol{xx}}^{-1}, \tag{2.5.4}$$

которая является состоятельной для $\mathrm{Avar}(\boldsymbol{b})$.

Представление S в терминах матриц данных

Если $\boldsymbol{B}$ — диагональная матрица размера $n \times n$, i-й диагональный элемент которой есть $\widehat{\varepsilon}_i^2$, то тогда $\widehat{\boldsymbol{S}}$ в (2.5.1) можно представить в терминах матриц данных как

$$\widehat{\boldsymbol{S}} = \frac{\boldsymbol{X}'\boldsymbol{B}\boldsymbol{X}}{n} \tag{2.5.1$'$}$$

с

$$\boldsymbol{B} = \begin{bmatrix} \widehat{\varepsilon}_1^2 & & \\ & \ddots & \\ & & \widehat{\varepsilon}_n^2 \end{bmatrix}.$$

Поэтому (2.5.4) может быть записано (с $\widehat{\varepsilon}_i$ в $\boldsymbol{B}$ равным e_i) как

$$\widehat{\mathrm{Avar}(\boldsymbol{b})} = n \cdot (\boldsymbol{X}'\boldsymbol{X})^{-1} (\boldsymbol{X}'\boldsymbol{B}\boldsymbol{X}) (\boldsymbol{X}'\boldsymbol{X})^{-1}. \tag{2.5.4$'$}$$

Эти выражения, хотя они и полезны для некоторых целей, не следует использовать для вычислений, поскольку $n \times n$ матрица $\boldsymbol{B}$ будет занимать слишком много оперативной памяти компьютера, особенно когда размер выборки велик. Для вычисления $\widehat{\boldsymbol{S}}$ по выборке формула (2.5.1) более полезна, чем (2.5.1$'$).

Рассмотрение конечных выборок

Разумеется, на конечных выборках мощность вполне может быть намного ниже единицы против конкретных альтернатив. Кроме того, вероятность отвергнуть нулевую гипотезу, когда DGP действительно удовлетворяет нулевой гипотезе (ошибка I рода) может очень отличаться от предполагаемого уровня значимости. Дэвидсон и Мак-Киннон [Davidson and MacKinnon, 1993, Section 16.3] сообщают, что, по крайней мере для наблюдавшихся ими симуляций Монте-Карло, робастное t-отношение, основанное на (2.5.1), отвергает нулевую гипотезу слишком часто и что простая замена знаменателя n в (2.5.1) на количество степеней свободы $n - K$ или, эквивалентно, домножение (2.5.1) на $n/(n-K)$ (это будет **коррекцией на число степеней свободы** (*degrees of freedom correction*) уменьшает проблему слишком частого отвержения. Они также сообщают, что робастное t-отношение, основанное на следующей корректировке $\widehat{\boldsymbol{S}}$, работает даже лучше:

$$\widehat{\boldsymbol{S}} = \frac{1}{n}\sum_{i=1}^{n} \frac{e_i^2}{(1-p_i)^d}\boldsymbol{x}_i\boldsymbol{x}_i', \quad d = 1 \text{ или } 2, \tag{2.5.5}$$

где p_i есть p_i, определенное в контексте анализа влияния в главе 1: это есть i-й диагональный элемент проекционной матрицы $\boldsymbol{P}$, то есть

$$p_i \equiv \boldsymbol{x}_i'(\boldsymbol{X}'\boldsymbol{X})^{-1}x_i = \frac{\boldsymbol{x}_i'\boldsymbol{S}_{\boldsymbol{xx}}^{-1}\boldsymbol{x}_i}{n}.$$

Контрольные вопросы

1. (Вычисление робастных стандартных ошибок.) В контрольном вопросе 2 параграфа 1.4 мы наблюдали, что стандартные ошибки OLS-оценок коэффициентов могут быть получены из $\boldsymbol{S}_{\boldsymbol{xx}}$, $\boldsymbol{s}_{\boldsymbol{xy}}$ (выборочное среднее $\boldsymbol{x}_i \cdot y_i$), $\boldsymbol{y}'\boldsymbol{y}/n$ и $\overline{y}$, поэтому выборочные моменты должны быть вычислены только один раз. Верно ли то же самое для робастных стандартных ошибок, где $\widehat{\boldsymbol{S}}$ вычисляется по формуле (2.5.1) с $\widehat{\varepsilon}_i = e_i$?

2. Дисперсия на конечных выборках OLS-оценки в обобщенной регрессионной модели главы 1 равна $\operatorname{Var}(\boldsymbol{b}|\boldsymbol{X}) = (\boldsymbol{X}\boldsymbol{X})^{-1}\boldsymbol{X}'(\sigma^2\boldsymbol{V})\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$. Сравните ее с (2.5.4′). Каковы различия?

2.6. Последствия условной гомоскедастичности

Тестовые статистики, разработанные в параграфах 2.4 и 2.5. отличаются от их аналогов на конечных выборках из параграфа 1.4, предназначенных для тестирования той же самой нулевой гипотезы. Как они соотносятся? Каково асимптотическое распределение t- и F-статистик из главы 1? Данный параграф отвечает на эти вопросы.

Оказывается, что робастное t-отношение численно равно t-отношению из параграфа 1.4 для конкретного выбора $\widehat{S}$. Следовательно, асимптотическое распределение t-отношения из параграфа 1.4 то же самое, что и у робастного t-отношения, если этот конкретный выбор является состоятельным для $\boldsymbol{S}$. Аналогичное соотношение выполняется между F-отношением из параграфа 1.4 и статистикой Вальда W данной главы. При предположении об условной гомоскедастичности, сформулированном ниже, этот конкретный выбор действительно является состоятельным.

Противопоставление условной и безусловной гомоскедастичности

Предположение об условной гомоскедастичности таково:

Предположение 2.7 (условная гомоскедастичность):

$$\mathrm{E}(\varepsilon_i^2|\boldsymbol{x}_i) = \sigma^2 > 0. \tag{2.6.1}$$

Из этого предположения вытекает, что безусловный второй момент $\mathrm{E}(\varepsilon_i^2)$ равен σ^2 по закону полных математических ожиданий. Чтобы прояснить различие между безусловной и условной гомоскедастичностью, рассмотрим следующий пример.

Пример 2.6 (безусловно гомоскедастичные, но условно гетероскедастичные ошибки): Как уже отмечалось, если $\{y_i, \boldsymbol{x}_i\}$ стационарен, то стационарен и $\{\varepsilon_i\}$, и ошибка *без*условно гомоскедастична по той причине, что $\mathrm{E}(\varepsilon_i^2)$ не зависит от i. Для иллюстрации того, что ошибка может быть тем не менее условно гетероскедастична, предположим, что ε_i записывается как $\varepsilon_i \equiv \eta_i f(\boldsymbol{x}_i)$, где $\{\eta_i\}$ имеет нулевое среднее $\mathrm{E}(\eta_i) = 0$ и не зависит от $\boldsymbol{x}_i$. Условный второй момент ε_i зависит от $\boldsymbol{x}_i$, поскольку

$$\begin{aligned}
&\mathrm{E}(\varepsilon_i^2|\boldsymbol{x}_i) = \mathrm{E}(\eta_i^2 f(\boldsymbol{x}_i)^2|\boldsymbol{x}_i) \qquad (\text{так как } \varepsilon_i \equiv \eta_i f(\boldsymbol{x}_i)) = \\
&= f(\boldsymbol{x}_i)^2\,\mathrm{E}(\eta_i^2|\boldsymbol{x}_i) \\
&(\text{по линейности условных математических ожиданий}) \\
&= f(\boldsymbol{x}_i)^2\,\mathrm{E}(\eta_i^2) \qquad (\text{так как } \eta_i \text{ не зависит от } \boldsymbol{x}_i \text{ по предположению}),
\end{aligned}$$

где правая часть изменяется по i из-за изменчивости $f(\boldsymbol{x}_i)$ в зависимости от i.

Редукция к формулам для конечных выборок

Для изучения распределений на больших выборках t- и F-статистик из главы 1 при дополнительном предположении 2.7, мы сначала изучим *алгебраическую* связь с их робастными аналогами. Рассмотрим следующий выбор для оценки $\boldsymbol{S}$:

$$\widehat{\boldsymbol{S}} = s^2 \boldsymbol{S}_{\boldsymbol{xx}}.$$

где s^2 — OLS-оценка для σ^2. (Мы очень скоро покажем, что эта оценка состоятельна при условной гомоскедастичности.) Тогда выражение (2.3.5) для $\widehat{\mathrm{Avar}(\boldsymbol{b})}$ принимает вид:

$$\widehat{\mathrm{Avar}(\boldsymbol{b})} = s^2 \boldsymbol{S}_{\boldsymbol{xx}}^{-1} = n \cdot s^2 \cdot (\boldsymbol{X}'\boldsymbol{X})^{-1}. \tag{2.6.2}$$

Подставляя это выражение в (2.4.1), мы видим, что робастная стандартная ошибка становится равной

$$\sqrt{s^2, \text{ умноженная на } (k,k) \text{ элемент матрицы } (\boldsymbol{X}'\boldsymbol{X})^{-1}}, \tag{2.6.3}$$

а это есть обычная стандартная ошибка в теории конечных выборок. Поэтому робастное t-отношение численно идентично обычному t-отношению для конечных выборок, когда мы полагаем $\widehat{\boldsymbol{S}} = s^2 \boldsymbol{S}_{\boldsymbol{xx}}$. Аналогично, подставляя (2.6.2) в выражение для статистики Вальда (2.4.2), мы получаем:

$$\begin{aligned}
W &= n \cdot (\boldsymbol{Rb} - \boldsymbol{r})' \{\boldsymbol{R}[n \cdot s^2 \cdot (\boldsymbol{X}'\boldsymbol{X})^{-1}]\boldsymbol{R}'\}^{-1} (\boldsymbol{Rb} - \boldsymbol{r}) = \\
&= (\boldsymbol{Rb} - \boldsymbol{r})' \{\boldsymbol{R}[s^2 \cdot (\boldsymbol{X}'\boldsymbol{X})^{-1}]\boldsymbol{R}'\}^{-1} (\boldsymbol{Rb} - \boldsymbol{r}) \quad \text{(два } n \text{ сокращаются)} \\
&= (\boldsymbol{Rb} - \boldsymbol{r})' \left\{\boldsymbol{R}[(\boldsymbol{X}'\boldsymbol{X})^{-1}]\boldsymbol{R}'\right\}^{-1} (\boldsymbol{Rb} - \boldsymbol{r})/s^2 = \\
&= \#\boldsymbol{r} \cdot F \quad \text{(по определению (1.4.9) } F\text{-отношения)} \\
&= (SSR_R - SSR_U)/s^2 \quad \text{(по (1.4.11))}.
\end{aligned}$$

Таким образом, когда мы полагаем $\widehat{\boldsymbol{S}} = s^2 \boldsymbol{S}_{\boldsymbol{xx}}$, статистика Вальда W численно идентична $\#\boldsymbol{r} \cdot F$ (где $\#\boldsymbol{r}$ есть число ограничений в нулевой гипотезе).

Распределение на больших выборках t- и F-статистик

Из утверждения 2.3 тогда следует, что t-отношение (2.4.1) асимптотически $N(0,1)$ и $\#\boldsymbol{r} \cdot F$ асимптотически $\chi^2(\#\boldsymbol{r})$, если $s^2 \boldsymbol{S}_{\boldsymbol{xx}}$ состоятельна для $\boldsymbol{S}$. То, что $s^2 \boldsymbol{S}_{\boldsymbol{xx}}$ состоятельна для $\boldsymbol{S}$, можно видеть из следующего. При условной гомоскедастичности матрица четвертых моментов $\boldsymbol{S}$ может быть выражена как произведение вторых моментов:

$$\begin{aligned}
\boldsymbol{S} &= \mathrm{E}(\boldsymbol{g}_i \boldsymbol{g}_i') = \mathrm{E}(\boldsymbol{x}_i \boldsymbol{x}_i' \varepsilon_i^2) \quad \text{(так как } \boldsymbol{g}_i \equiv \boldsymbol{x}_i \cdot \varepsilon_i) \\
&= \mathrm{E}[\mathrm{E}(\boldsymbol{x}_i \boldsymbol{x}_i' \varepsilon_i^2 | \boldsymbol{x}_i)] \quad \text{(по закону полных математических ожиданий)} \\
&= \mathrm{E}[\boldsymbol{x}_i \boldsymbol{x}_i' \, \mathrm{E}(\varepsilon_i^2 | \boldsymbol{x}_i)] \\
&\quad \text{(по линейности условных математических ожиданий)} \\
&= \mathrm{E}(\boldsymbol{x}_i \boldsymbol{x}_i' \sigma^2) \quad \text{(по предположению 2.7)} \\
&= \sigma^2 \, \mathrm{E}(\boldsymbol{x}_i \boldsymbol{x}_i') = \sigma^2 \boldsymbol{\Sigma}_{\boldsymbol{xx}}.
\end{aligned} \tag{2.6.4}$$

Это представление имеет несколько важных следствий.

- ($\mathbf{\Sigma}_{\boldsymbol{xx}}$ является невырожденной.) Так как, по предположению 2.5, $\boldsymbol{S}$ является невырожденной, это разложение $\boldsymbol{S}$ влечет $\sigma^2 > 0$, и $\mathbf{\Sigma}_{\boldsymbol{xx}}$ является невырожденной. Следовательно, отсюда вытекает предположение 2.4 (ранговое условие).

- (Нет необходимости в предположении о четвертом моменте.) В силу эргодической стационарности, $\boldsymbol{S}_{\boldsymbol{xx}} \to_{\mathrm{p}} \mathbf{\Sigma}_{\boldsymbol{xx}}$. Из утверждения 2.2 вытекает, что s^2 состоятельна для σ^2 при предположениях 2.1–2.4. Таким образом, $s^2\boldsymbol{S}_{\boldsymbol{xx}} \to_{\mathrm{p}} \sigma^2\mathbf{\Sigma}_{\boldsymbol{xx}} = \boldsymbol{S}$. Для нас нет необходимости требовать предположение о четвертом моменте (предположение 2.6) для состоятельности.

В качестве другого следствия из (2.6.4) заметим, что выражение для $\mathrm{Avar}(\boldsymbol{b})$ можно упростить: подставляя (2.6.4) в (b) из предположения 2.1, приводим выражение для $\mathrm{Avar}(\boldsymbol{b})$ к виду:

$$\mathrm{Avar}(\boldsymbol{b}) = \sigma^2\mathbf{\Sigma}_{\boldsymbol{xx}}^{-1}. \tag{2.6.5}$$

Таким образом, мы доказали

Утверждение 2.5 (свойства на больших выборках $\boldsymbol{b}$, $\boldsymbol{t}$ и $\boldsymbol{F}$ при условной гомоскедастичности): *Предположим, что выполняются предположения 2.1–2.5 и 2.7. Тогда*

(a) (Асимптотическое распределение $\boldsymbol{b}$.*) OLS-оценка* $\boldsymbol{b}$ *для* $\boldsymbol{\beta}$ *является состоятельной и асимптотически нормальной с* $\mathrm{Avar}(\boldsymbol{b}) = \sigma^2\mathbf{\Sigma}_{\boldsymbol{xx}}^{-1}$.

(b) (Состоятельное оценивание асимптотической дисперсии.) При том же множестве предположений $\mathrm{Avar}(\boldsymbol{b})$ *состоятельно оценивается посредством* $\widehat{\mathrm{Avar}(\boldsymbol{b})} = s^2\,\boldsymbol{S}_{\boldsymbol{xx}}^{-1} = n \cdot s^2 \cdot (\boldsymbol{X}'\boldsymbol{X})^{-1}$.

(c) (Асимптотическое распределение t- *и* F*-статистик теории конечных выборок.) При* H_0: $\beta_k = \overline{\beta}_k$ *обычное* t*-отношение (1.4.5) асимптотически нормально распределено как* $N(0,1)$. *При* H_0: $\boldsymbol{R\beta} = \boldsymbol{r}$, $\#\boldsymbol{r} \cdot F$ *имеет асимптотическое распределение* $\chi^2(\#\boldsymbol{r})$, *где* F *есть* F*-статистика из (1.4.9), и* $\#\boldsymbol{r}$ *равно числу ограничений в* H_0.

Варианты асимптотических тестов при условной гомоскедастичности

Согласно данному результату, вы должны посмотреть таблицу $N(0,1)$, чтобы найти критическое значение для сравнения с t-отношением (1.4.5) и таблицу χ^2 для статистики $\#\boldsymbol{r} \cdot F$, полученной из (1.4.9). Некоторые исследователи заменяют s^2 в (1.4.5) и (1.4.9) на $\frac{1}{n}\sum_i e_i^2$. То есть количество степеней свободы $n - K$ заменяется на n, или коррекция степеней свободы, скрытая в s^2, устраняется. Отличие, возникающее при этой замене, исчезает на больших выборках, поскольку $\lim_{n\to\infty}[n/(n-K)] = 1$. Следовательно, независимо от того, какой тест использовать, результат теста будет одинаковым, если объем выборки достаточно большой.

Другой вариант заключается в том, чтобы оставить количество степеней свободы $n - K$, но использовать таблицу $t(n - K)$ для t-отношения и таблицу $F(\#\boldsymbol{r}, n - K)$ для F, что *в точности* предписано в теории конечных выборок. Это также асимптотически обоснованно, поскольку при $n - K$, стремящемся к бесконечности (что и происходит, когда $n \to \infty$ с фиксированным K), распределение $t(n - K)$ сходится к $N(0, 1)$ (просто сравните таблицу t-распределения для большого количества степеней свободы с таблицей стандартного нормального распределения), а $F(\#\boldsymbol{r}, n - K)$ сходится к $\chi^2(\#\boldsymbol{r})/\#\boldsymbol{r}$. Иными словами, даже если ошибка не распределена нормально и регрессоры всего лишь предопределенные (ортогональны ошибке) и не строго экзогенные, распределение t-отношения (1.4.5) хорошо аппроксимируется распределением $t(n-K)$, а распределение F-отношения — распределением $F(\#\boldsymbol{r}, n - K)$.

Все эти варианты асимптотически эквивалентны в том, что разница в их значениях исчезает на больших выборках и, следовательно (по лемме 2.4(a)), их асимптотическое распределение одинаково. Однако, когда объем выборки только умеренно большой, аппроксимация распределения на конечных выборках, или точного распределения, тестовых статистик может быть лучше при использовании $t(n-K)$ и $F(\#\boldsymbol{r}, n-K)$, а не $N(0, 1)$ и $\chi^2(\#\boldsymbol{r})$. Поскольку точное распределение зависит от DGP, нет простого указания на то, какой вариант работает лучше на конечных выборках. Вопрос о том, какую таблицу — $N(0, 1)$ или $t(n - K)$ — нужно использовать для умеренных объемов выборки, будет рассмотрен в упражнении на метод Монте-Карло в этой главе.

Контрольные вопросы

1. (Несостоятельность формул для конечных выборок без условной гомоскедастичности.) Без предположения 2.7 $\text{Avar}(\boldsymbol{b})$ определяется посредством (b) в утверждении 2.1. Оценивается ли она состоятельно посредством (2.6.2) без предположения 2.7? [Ответ: Нет. Почему?] Является ли t-отношение (1.4.5) асимптотически стандартным нормальным без такого предположения? [Ответ: Нет.]

2. (Преимущество формул для конечных выборок при условной гомоскедастичности.) Наоборот, при предположении 2.7 $\text{Avar}(\boldsymbol{b})$ определяется в (2.6.5). Оценивается ли она состоятельно посредством (2.3.5) при предположении 2.7? Если предположение 2.7 выполняется, то как вы думаете, есть ли преимущество использования (2.6.2) над (2.3.5) для оценивания асимптотической дисперсии? [**Замечание:** Свойства оценки на конечных выборках, в общем случае, тем лучше, чем меньше число популяционных параметров, оценива-

емых для формирования этой оценки. Как много популяционных параметров необходимо оценить для формирования (2.3.5)? (2.6.2)?]

3. (Отношение F к χ^2.) Найдите 5%-ное критическое значение для $F(10, \infty)$ и сравните его с 5%-ным критическим значением для $\chi^2(10)$. Как они соотносятся между собой?

4. Без условной гомоскедастичности имеет ли $(SSR_R - SSR_U)/s^2$ асимптотически распределение $\chi^2(\#\boldsymbol{r})$? [Ответ: Нет.]

5. (nR^2.) Для регрессии с константой рассмотрим нулевую гипотезу о том, что коэффициенты при $K-1$ регрессорах, не являющихся константой, все равны нулю. Покажите, что $nR^2 \to_{\mathrm{d}} \chi^2(K-1)$ при гипотезе в утверждении 2.5. **Указание:** *Вы доказали в аналитическом упражнении главы 1, что алгебраическое соотношение между F-отношением для нулевой гипотезы и R^2 есть*

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

Можете ли вы использовать статистику nR^2, когда ошибки не являются условно гомоскедастичными? [Ответ: Нет.]

2.7. Тестирование условной гомоскедастичности

С появлением робастных стандартных ошибок, позволяющих нам делать выводы без спецификации условного второго момента $\mathrm{E}(\varepsilon_i^2|\boldsymbol{x}_i)$, тестирование условной гомоскедастичности стало не столь важным, как раньше. Этот параграф представляет только самый популярный тест, предложенный в работе [White, 1980] для случая случайных выборок[1].

Напомним, что $\widehat{\boldsymbol{S}}$, заданная в (2.5.1) (с $\widehat{\varepsilon}_i = e_i$), является состоятельной для $\boldsymbol{S}$ (утверждение 2.4), и $s^2\boldsymbol{S}_{\boldsymbol{xx}}$ является состоятельной для $\sigma^2\boldsymbol{\Sigma}_{\boldsymbol{xx}}$ (как следствие утверждения 2.2). Но при условной гомоскедастичности $\boldsymbol{S} = \sigma^2\boldsymbol{\Sigma}_{\boldsymbol{xx}}$ (см. (2.6.4)), поэтому разница между двумя состоятельными оценками должна исчезать:

$$\begin{aligned}\widehat{\boldsymbol{S}} - s^2\,\boldsymbol{S}_{\boldsymbol{xx}} &= \frac{1}{n}\sum_{i=1}^{n} e_i^2\boldsymbol{x}_i\boldsymbol{x}_i' - s^2\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i' = \\ &= \frac{1}{n}\sum_{i=1}^{n}(e_i^2 - s^2)\boldsymbol{x}_i\boldsymbol{x}_i' \underset{\mathrm{p}}{\to} \boldsymbol{0}. \qquad (2.7.1)\end{aligned}$$

Пусть $\boldsymbol{\psi}_i$ будет вектором, собирающим уникальные и непостоянные элементы симметричной матрицы $\boldsymbol{x}_i\boldsymbol{x}_i'$ размерности $K \times K$. (Построение

[1] См., например, [Judge et al. 1985, Section 11.3] или [Greene, 1997, Section 12.3] относительно других тестов.

ψ_i из $x_i x_i'$ будет проиллюстрировано в примере 2.7 ниже.) Тогда (2.7.1) влечет:

$$c_n \equiv \frac{1}{n}\sum_{i=1}^{n}(e_i^2 - s^2)\psi_i \xrightarrow[p]{} \mathbf{0} \tag{2.7.2}$$

Это c_n является выборочным средним, сходящимся к нулю. При некоторых условиях, соответствующих применимости центральной предельной теоремы, мы можем ожидать, что $\sqrt{n}c_n$ сходится по распределению к нормальному распределению с нулевым средним и некоторой асимптотической дисперсией $\boldsymbol{B}$, поэтому для любой состоятельной оценки $\widehat{\boldsymbol{B}}$ для $\boldsymbol{B}$:

$$n \cdot c_n' \widehat{\boldsymbol{B}}^{-1} c_n \xrightarrow[d]{} \chi^2(m), \tag{2.7.3}$$

где m есть размерность c_n. Для конкретного выбора $\widehat{\boldsymbol{B}}$ эта статистика может быть вычислена как nR^2 из следующей вспомогательной регрессии:

$$\text{регрессия } e_i^2 \text{ на константу и } \psi_i. \tag{2.7.4}$$

В [White, 1980] эта аргументация реализована строго[1] для доказательства следующего утверждения.

Утверждение 2.6 (тест Уайта для условной гетероскедастичности): *В дополнение к предположениям 2.1 и 2.4 предположим, что а) $\{y_i, x_i\}$ есть i.i.d. с конечным $\mathrm{E}(\varepsilon_i^2 x_i x_i')$ (таким образом, усиливая предположения 2.2 и 2.5), b) ε_i независима от x_i (таким образом, усиливая предположение 2.3 и условную гомоскедастичность) и c) выполняются определенные условия на моменты ε_i и x_i. Тогда*

$$nR^2 \xrightarrow[d]{} \chi^2(m),$$

где R^2 есть R^2 вспомогательной регрессии (2.7.4), а m есть размерность ψ_i.

Пример 2.7 (регрессоры в тесте Уайта nR^2): Рассмотрим функцию издержек Кобба — Дугласа из параграфа 1.7:

$$\log\left(\frac{TC_i}{p_{i3}}\right) = \beta_1 + \beta_2 \log(Q_i) + \beta_3 \log\left(\frac{p_{i1}}{p_{i3}}\right) + \beta_4 \log\left(\frac{p_{i2}}{p_{i3}}\right) + \varepsilon_i.$$

[1] Оригинальная формулировка теоремы Уайта более общая, так как охватывает случай, где $\{y_i, x_i\}$ независимы, но не одинаково распределены (i.n.i.d.).

Здесь $\boldsymbol{x}_i' = (1, \log(Q_i), \log(p_{i1}/p_{i3}), \log(p_{i1}/p_{i3}))$ — четырехмерный вектор. Матрица $\boldsymbol{x}_i\boldsymbol{x}_i'$ содержит 10 ($= 4 \cdot 5/2$) уникальных элементов:

$$1, \quad \log(Q_i), \quad \log\left(\frac{p_{i1}}{p_{i3}}\right), \quad \log\left(\frac{p_{i2}}{p_{i3}}\right),$$

$$[\log(Q_i)]^2, \quad \log(Q_i)\cdot\log\left(\frac{p_{i1}}{p_{i3}}\right), \quad \log(Q_i)\cdot\log\left(\frac{p_{i2}}{p_{i3}}\right),$$

$$\left[\log\left(\frac{p_{i1}}{p_{i3}}\right)\right]^2, \quad \log\left(\frac{p_{i1}}{p_{i3}}\right)\cdot\log\left(\frac{p_{i2}}{p_{i3}}\right), \quad \left[\log\left(\frac{p_{i2}}{p_{i3}}\right)\right]^2.$$

ψ_i есть девятимерный вектор, исключающий константу из этого списка. Поэтому m в утверждении 2.6 равно 9.

Если тест Уайта принимает нулевую гипотезу об условной гомоскедастичности, то тогда применяются результаты параграфа 2.6, и статистические выводы могут быть основаны на t- и F-отношениях из главы 1. В противном случае статистические выводы должны основываться на робастных статистиках t и Вальда из утверждения 2.3.

Поскольку регрессоры ψ_i во вспомогательной регрессии имеют много элементов, состоящих из квадратов и перекрестных произведений элементов $\boldsymbol{x}_i$, тест должен быть состоятельным (то есть его мощность приближается к единице при $n \to \infty$) против большинства альтернатив гетероскедастичности, но может потребоваться достаточно большая выборка, чтобы получить мощность, близкую к единице. Если исследователь знает, что некоторые элементы ψ_i не влияют на условный второй момент, то их можно исключить из вспомогательной регрессии, и мощность может возрасти на конечных выборках. Недостаток этого, конечно, заключается в том, что, если такое знание ложно, тест не будет иметь никакой мощности против альтернатив о гетероскедастичности, которые соотносят условный второй момент к тем элементам, которые исключены из вспомогательной регрессии.

Контрольные вопросы

1. (Размерность ψ_i.) Предположим, что $x_i = (1, q_i, q_i^2, p_i)'$, четырехмерный вектор. Как много непостоянных и уникальных элементов в $\boldsymbol{x}_i\boldsymbol{x}_i'$? [Ответ: 8.]

2.8. Оценивание параметризованной условной гетероскедастичности (дополнительно)

Даже когда обнаружено, что ошибка является условно гетероскедастичной, OLS-оценка все еще состоятельна и асимптотически нормальна, и обоснованные статистические выводы могут проводиться с робастными стандартными ошибками и робастными статистиками Вальда. Однако в (довольно маловероятном) случае *априорного* знания функциональной формы условного второго момента $\mathrm{E}(\varepsilon_i^2|\boldsymbol{x}_i)$ возможно получить более точные оценки с меньшей асимптотической дисперсией. Действительно, в теории конечных выборок WLS (взвешенный метод наименьших квадратов, *weighted least squares*) может включать в себя такие знания для повышения эффективности в смысле меньшей дисперсии на конечных выборках. Переносится ли этот результат для конечных выборок на теорию больших выборок? Этот параграф посвящен свойствам WLS-оценки на больших выборках. Чтобы упростить обсуждение, в этом параграфе мы усиливаем предположения 2.2 и 2.5, предполагая, что $\{y_i, \boldsymbol{x}_i\}$ является i.i.d. Такое предположение является естественным, поскольку оно обычно в контексте кросс-секционного анализа, где привлекается WLS.

Функциональная форма

Параметрическая функциональная форма для условного второго момента, которую мы рассмотрим, имеет вид:

$$\mathrm{E}(\varepsilon_i^2|\boldsymbol{x}_i) = \boldsymbol{z}_i'\boldsymbol{\alpha}. \tag{2.8.1}$$

где $\boldsymbol{z}_i$ — некоторая функция от $\boldsymbol{x}_i$.

Пример 2.8 (параметрическая форма условного второго момента): Функциональная форма, использованная в WLS-оценивании в эмпирическом упражнении к главе 1, имела вид:

$$\mathrm{E}\left[\varepsilon_i^2|\log(Q_i), \log(Q_i)^2, \log\left(\frac{p_{i1}}{p_{i3}}\right), \log\left(\frac{p_{i2}}{p_{i3}}\right)\right] = \alpha_1 + \alpha_2 \cdot \left(\frac{1}{Q_i}\right).$$

Поскольку элементы $\boldsymbol{z}_i$ могут быть нелинейными функциями от $\boldsymbol{x}_i$, эта спецификация является более гибкой, чем могло бы показаться на первый взгляд, но все еще исключает некоторые нелинейности. Например, функциональная форма

$$\mathrm{E}(\varepsilon_i^2|\boldsymbol{x}_i) = \exp(\boldsymbol{z}_i'\boldsymbol{\alpha}) \tag{2.8.2}$$

может быть более привлекательной, поскольку ее значение гарантированно положительно. Мы рассмотрим линейную спецификацию (2.8.1) только потому, что оценивание параметров в нелинейных моделях, таких как (2.8.2), требует использования нелинейного метода наименьших квадратов.

WLS с известным α

Чтобы исключить осложнения, возникающие из-за того, что приходится оценивать неизвестный вектор параметров $\boldsymbol{\alpha}$, мы сначала изучим распределение на больших выборках WLS-оценки с известным $\boldsymbol{\alpha}$. Если $\boldsymbol{\alpha}$ известен, условный второй момент может быть вычислен по данным как $\mathbf{z}_i'\boldsymbol{\alpha}$, и WLS применяется точно так же, как и в параграфе 1.6: делением обеих сторон оцениваемого уравнения $y_i = \mathbf{x}_i'\boldsymbol{\beta}+\varepsilon_i$ на квадратный корень из $\mathbf{z}_i'\boldsymbol{\alpha}$, чтобы получить

$$\widetilde{y}_i = \widetilde{\mathbf{x}}_i'\boldsymbol{\beta} + \widetilde{\varepsilon}_i, \tag{2.8.3}$$

где

$$\widetilde{y}_i \equiv \frac{y_i}{\sqrt{\mathbf{z}_i'\boldsymbol{\alpha}}},\quad \widetilde{\mathbf{x}}_i \equiv \frac{\mathbf{x}_i}{\sqrt{\mathbf{z}_i'\boldsymbol{\alpha}}},\quad \widetilde{\varepsilon}_i \equiv \frac{\varepsilon_i}{\sqrt{\mathbf{z}_i'\boldsymbol{\alpha}}},$$

а затем применить OLS. Для дальнейшего использования запишем результирующую WLS-оценку как $\widehat{\boldsymbol{\beta}}(\mathbf{V})$. Она задается как:

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}(\mathbf{V}) &\equiv \left(\frac{1}{n}\sum_{i=1}^{n}\widetilde{\mathbf{x}}_i\widetilde{\mathbf{x}}_i'\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\widetilde{\mathbf{x}}_i\cdot\widetilde{y}_i = \\
&= \left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\mathbf{z}_i'\boldsymbol{\alpha}}\mathbf{x}_i\mathbf{x}_i'\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\mathbf{z}_i'\boldsymbol{\alpha}}\mathbf{x}_i\cdot y_i = \\
&= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y},
\end{aligned} \tag{2.8.4}$$

с

$$\mathbf{V} \equiv \begin{bmatrix} \mathbf{z}_1'\boldsymbol{\alpha} & & \\ & \ddots & \\ & & \mathbf{z}_n'\boldsymbol{\alpha} \end{bmatrix}.$$

Если предположение 2.3 усилено условием, что

$$\mathrm{E}(\varepsilon_i|\mathbf{x}_i) = 0, \tag{2.8.5}$$

то $\mathrm{E}((\widetilde{\varepsilon}_i|(\widetilde{\mathbf{x}}_i) = 0$. Чтобы увидеть это, заметим, во-первых, что, поскольку $\mathbf{z}_i$ есть функция от $\mathbf{x}_i$,

$$\mathrm{E}(\widetilde{\varepsilon}_i|\mathbf{x}_i) = \mathrm{E}\left(\frac{\varepsilon_i}{\sqrt{\mathbf{z}_i'\boldsymbol{\alpha}}}\;\middle|\;\mathbf{x}_i\right) = \frac{1}{\sqrt{\mathbf{z}_i'\boldsymbol{\alpha}}}\,\mathrm{E}\,(\varepsilon_i|\mathbf{x}_i) = 0 \tag{2.8.6}$$

Во-вторых, поскольку $\widetilde{\mathbf{x}}_i$ есть функция от $\mathbf{x}_i$, в $\widetilde{\mathbf{x}}_i$ информации не больше, чем в $\mathbf{x}_i$. Поэтому, по закону повторных математических ожиданий мы имеем:

$$\mathrm{E}(\widetilde{\varepsilon}_i|\widetilde{\mathbf{x}}_i) = \mathrm{E}[\mathrm{E}(\widetilde{\varepsilon}_i|\mathbf{x}_i)|\widetilde{\mathbf{x}}_i] = 0.$$

Следовательно, при условии, что $\mathrm{E}(\widetilde{\boldsymbol{x}}_i\widetilde{\boldsymbol{x}}_i')$ невырождена, предположения 2.1–2.5 удовлетворяются для уравнения (2.8.3). Кроме того, по построению, ошибка $\widetilde{\varepsilon}_i$ условно гомоскедастична: $\mathrm{E}(\widetilde{\varepsilon}_i^2|\widetilde{\boldsymbol{x}}_i) = 1$. Поэтому утверждение 2.5 применяется с $\sigma^2 = 1$: WLS-оценка является состоятельной и асимптотически нормальной, а асимптотическая дисперсия равна

$$\begin{aligned}
\mathrm{Avar}(\widehat{\boldsymbol{\beta}}(\boldsymbol{V})) &= \mathrm{E}(\widetilde{\boldsymbol{x}}_i\widetilde{\boldsymbol{x}}_i')^{-1} \quad \text{(так как дисперсия ошибок есть 1)} = \\
&= \operatorname{plim}\Big(\frac{1}{n}\sum_{i=1}^{n}\widetilde{\boldsymbol{x}}_i\widetilde{\boldsymbol{x}}_i'\Big)^{-1} \quad \text{(по эргодической стационарности)} = \\
&= \operatorname{plim}\Big(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\boldsymbol{z}_i'\boldsymbol{\alpha}}\boldsymbol{x}_i\boldsymbol{x}_i'\Big)^{-1} \quad \text{(так как } \widetilde{\boldsymbol{x}}_i = \boldsymbol{x}_i/\sqrt{\boldsymbol{z}_i'\boldsymbol{\alpha}}) = \\
&= \operatorname{plim}\Big(\frac{1}{n}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\Big)^{-1}. \qquad (2.8.7)
\end{aligned}$$

Поэтому $(\frac{1}{n}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}$ является состоятельной оценкой $\mathrm{Avar}(\widehat{\boldsymbol{\beta}}(\boldsymbol{V}))$.

Регрессия e_i^2 на z_i дает состоятельную оценку α

Если $\boldsymbol{\alpha}$ неизвестен, его можно оценить, прогоняя отдельную регрессию. Соотношение (2.8.1) говорит, что $\boldsymbol{z}_i'\boldsymbol{\alpha}$ есть регрессия ε_i^2. Если мы определим $\eta_i \equiv \varepsilon_i^2 - \mathrm{E}(\varepsilon_i^2|\boldsymbol{x}_i)$, то (2.8.1) может быть записано как уравнение регрессии:

$$\varepsilon_i^2 = \boldsymbol{z}_i'\boldsymbol{\alpha} + \eta_i. \qquad (2.8.8)$$

По построению $\mathrm{E}(\eta_i|\boldsymbol{x}_i) = 0$, что вместе с тем фактом, что $\boldsymbol{z}_i$ есть функция от $\boldsymbol{x}_i$, влечет ортогональность регрессоров $\boldsymbol{z}_i$ ошибке η_i. Следовательно, при условии, что $\mathrm{E}(\boldsymbol{z}_i\boldsymbol{z}_i')$ невырождена, утверждение 2.1 применимо к этой вспомогательной регрессии (2.8.8): OLS-оценка для $\boldsymbol{\alpha}$ состоятельна и асимптотически нормальна.

Конечно, мы не можем сделать это, потому что мы не наблюдаем ошибку. Однако, так как OLS-оценка $\boldsymbol{b}$ для первоначальной регрессии $y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i$ состоятельна, несмотря на наличие условной гетероскедастичности, OLS-остатки e_i дают состоятельные оценки для ε_i. Вам остается в качестве аналитического упражнения показать, что, когда ε_i заменяется на e_i в регрессии (2.8.8), OLS-оценка, назовем ее $\widehat{\boldsymbol{\alpha}}$, состоятельна для $\boldsymbol{\alpha}$.

WLS с оцененным α

WLS-оценка первоначального уравнения $y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i$ с оцененным $\boldsymbol{\alpha}$ есть $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{V}})$, где $\widehat{\boldsymbol{V}}$ есть диагональная матрица $n \times n$, i-й диагональный элемент которой равен $\boldsymbol{z}_i'\widehat{\boldsymbol{\alpha}}$. При соответствующих дополнительных условиях (см., например: [Amemiya, 1997] относительно точной формулировки таких условий) можно показать, что

(a) $\sqrt{n}(\widehat{\boldsymbol{\beta}}(\boldsymbol{V})-\boldsymbol{\beta})$ и $\sqrt{n}(\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{V}})-\boldsymbol{\beta})$ асимптотически эквивалентны в том смысле, что их разность сходится по вероятности к нулю при $n \to \infty$. Следовательно, по лемме 2.4(a), асимптотическое распределение $\sqrt{n}(\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{V}})-\boldsymbol{\beta})$ то же самое, что и у $\sqrt{n}(\widehat{\boldsymbol{\beta}}(\boldsymbol{V})-\boldsymbol{\beta})$. Поэтому $\mathrm{Avar}(\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{V}}))$ равна $\mathrm{Avar}(\widehat{\boldsymbol{\beta}}(\boldsymbol{V}))$, которая, в свою очередь, задается (2.8.7);

(b) $\operatorname{plim} \frac{1}{n}\boldsymbol{X}'\widehat{\boldsymbol{V}}^{-1}\boldsymbol{X} = \operatorname{plim} \frac{1}{n}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}$. Поэтому $(\frac{1}{n}\boldsymbol{X}'\widehat{\boldsymbol{V}}^{-1}\boldsymbol{X})^{-1}$ состоятельна для $\mathrm{Avar}(\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{V}}))$.

Все это может показаться сложным, но смысл его для WLS-оценивания уравнения $y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i$ очень ясен:

Шаг 1: Оцениваем уравнение $y_i = \boldsymbol{x}_i'\boldsymbol{\beta}+\varepsilon_i$ OLS и вычисляем OLS-остатки e_i.

Шаг 2: Регрессируем e_i^2 на $\boldsymbol{z}_i$, чтобы получить OLS-оценку коэффициента $\boldsymbol{\alpha}$.

Шаг 3: Повторно оцениваем уравнение $y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i$, используя WLS с $1/\sqrt{\boldsymbol{z}_i'\widehat{\boldsymbol{\alpha}}}$ в качестве веса для наблюдения i.

Так как правильная оценка асимптотической дисперсии, $(\frac{1}{n}\boldsymbol{X}'\widehat{\boldsymbol{V}}^{-1}\boldsymbol{X})^{-1}$ в (b), есть умноженная на n дисперсионная матрица, обычно распечатываемая в стандартных регрессионных пакетах для *шага 3*, вычислять соответствующие тестовые статистики довольно просто: для статистических выводов можно использовать стандартные t- и F-отношения из *шага 3* регрессии.

Противопоставление OLS и WLS

Таким образом, у нас есть две состоятельные и асимптотически нормальные оценки, OLS- и WLS-оценки. Мы говорим, что состоятельная и асимптотически нормальная оценка **асимптотически более эффективна** (*asymptotically more efficient*), чем другая состоятельная и асимптотически нормальная оценка того же самого параметра, если асимптотическая дисперсия первой не больше, чем последней. Вам остается в качестве аналитического упражнения показать, что WLS-оценка асимптотически более эффективна, чем OLS-оценка.

Превосходство WLS над OLS, однако, опирается на предпосылку о том, что объем выборки достаточно велик *и* функциональная форма условного второго момента правильно специфицирована. Если же функциональная форма специфицирована неправильно, WLS-оценка будет все еще состоятельной, но ее асимптотическая дисперсия может быть, а может и не быть меньше, чем $\text{Avar}(\boldsymbol{b})$. На конечных выборках, даже если функциональная форма специфицирована верно, аппроксимация для больших выборок будет, вероятно, работать для WLS-оценки хуже, чем для OLS-оценки, поскольку в WLS-процедуру включено оценивание дополнительных параметров ($\boldsymbol{\alpha}$).

Контрольные вопросы

1. Докажите: «$\text{E}(\eta_i|\boldsymbol{x}_i) = 0$, $\boldsymbol{z}_i$ является функцией от $\boldsymbol{x}_i$» $\Rightarrow$ «$\text{E}(\boldsymbol{z}_i \cdot \eta_i) = \boldsymbol{0}$.» **Указание:** *закон полных математических ожиданий.*

2. Являются ли ошибки условно гомоскедастичными во вспомогательной регрессии (2.8.8)? Если да, имеют ли они влияние на асимптотическое распределение WLS-оценки?

2.9. Проекция методом наименьших квадратов

Что делать, если предположения, обосновывающие свойства OLS-оценки для больших выборок (за исключением эргодической стационарности), не выполняются, но мы тем не менее действуем по своему усмотрению и применяем OLS к выборке? Что тогда мы оцениваем? Ответ на этот вопрос находится в данном параграфе. OLS дает оценку лучшего способа линейно объединить объясняющие переменные для предсказания зависимой переменной. Такая линейная комбинация называется **проекцией методом наименьших квадратов** (*least squares projection*).

Оптимальное предсказание значения зависимой переменной

Мы были обеспокоены оцениванием неизвестных параметров по выборке. Давайте временно приостановим роль эконометристов и поставим себя в следующую ситуацию. Существуют случайный скаляр y и случайный вектор $\boldsymbol{x}$. Мы знаем совместное распределение $(y, \boldsymbol{x})$ и значение $\boldsymbol{x}$. На основе этого знания мы хотим предсказать y. Поэтому предиктор есть функция $f(\boldsymbol{x})$ от $\boldsymbol{x}$ с функциональной формой $f(\cdot)$, определенной совместным распределением $(y, \boldsymbol{x})$. Естественно, мы выбираем функцию $f(\cdot)$ таким образом, чтобы минимизировать некоторый показатель, который является функцией от **ошибки прогноза** (*forecast error*) $y - f(\boldsymbol{x})$. Мы берем в качестве функции потерь **среднеквадратичную ошибку** (*mean squared error*) $\text{E}[(y - f(\boldsymbol{x}))^2]$, поскольку она кажется столь же разумной функцией потерь, как и какая-нибудь другая, и, что более важно, поскольку она приводит к следующему удобному результату:

Утверждение 2.7: $\text{E}(y|\boldsymbol{x})$ *есть наилучший предиктор для* y *в том смысле, что он минимизирует среднеквадратичную ошибку.*

Мы видели в главе 1, что стратегия прибавления — вычитания эффективна для того, чтобы показать, что вариант решения минимизирует квадратичную функцию. Давайте применим здесь эту стратегию к квадратам ошибок. Пусть $f(\boldsymbol{x})$ будет любым прогнозом. Прибавим $\mathrm{E}(y|\boldsymbol{x})$ к ошибке прогноза $y - f(\boldsymbol{x})$ и затем вычтем его для получения разложения

$$y - f(\boldsymbol{x}) = (y - \mathrm{E}(y|\boldsymbol{x})) + (\mathrm{E}(y|\boldsymbol{x}) - f(\boldsymbol{x})). \tag{2.9.1}$$

Поэтому квадрат ошибки прогноза равен

$$(y - f(\boldsymbol{x}))^2 = (y - \mathrm{E}(y|\boldsymbol{x}))^2 + 2(y - \mathrm{E}(y|\boldsymbol{x}))(\mathrm{E}(y|\boldsymbol{x}) - f(\boldsymbol{x})) + \\ + (\mathrm{E}(y|\boldsymbol{x}) - f(\boldsymbol{x}))^2. \tag{2.9.2}$$

Берем математическое ожидание от обеих сторон и получаем:

$$\text{среднеквадратичная ошибка} \equiv \mathrm{E}\big[(y - f(\boldsymbol{x}))^2\big] = \\ = \mathrm{E}\big[(y - \mathrm{E}(y|\boldsymbol{x}))^2\big] + 2\,\mathrm{E}\big[(y - \mathrm{E}(y|\boldsymbol{x}))(\mathrm{E}(y|\boldsymbol{x}) - f(\boldsymbol{x}))\big] + \\ + \mathrm{E}\big[(\mathrm{E}(y|\boldsymbol{x}) - f(\boldsymbol{x}))^2\big]. \tag{2.9.3}$$

Непосредственным применением закона полных математических ожиданий можно показать, что средняя компонента, являющаяся ковариацией между ошибкой оптимального прогноза и разностью между прогнозами, равна нулю (контрольный вопрос). Следовательно,

$$\text{среднеквадратичная ошибка} = \mathrm{E}\big[(y - \mathrm{E}(y|\boldsymbol{x}))^2\big] + \\ + \mathrm{E}\big[(\mathrm{E}(y|\boldsymbol{x}) - f(\boldsymbol{x}))^2\big] \geq \mathrm{E}\big[(y - \mathrm{E}(y|\boldsymbol{x}))^2\big]. \tag{2.9.4}$$

что показывает, что среднеквадратичная ошибка ограничена снизу величиной $\mathrm{E}[(y - \mathrm{E}(y|\boldsymbol{x})^2]$, и эта нижняя граница достигается условным математическим ожиданием.

Наилучший линейный предиктор

Для вычисления $\mathrm{E}(y|\boldsymbol{x})$, которое может быть сильно нелинейным, требуется знание совместного распределения $(y, \boldsymbol{x})$. Мы теперь ограничимся предикторами, которые являются линейными функциями от $\boldsymbol{x}$ и зададимся вопросом: какой предиктор является наилучшим (в смысле минимизации среднеквадратичной ошибки) линейным предиктором для y, основанным на $\boldsymbol{x}$? Для этой цели рассмотрим $\boldsymbol{\beta}^*$, который удовлетворяет условию ортогональности

$$\mathrm{E}[\boldsymbol{x} \cdot (y - \boldsymbol{x}'\boldsymbol{\beta}^*)] = \mathbf{0} \quad \text{или} \quad \mathrm{E}(\boldsymbol{x}\boldsymbol{x}')\boldsymbol{\beta}^* = \mathrm{E}(\boldsymbol{x} \cdot y). \tag{2.9.5}$$

Идея состоит в том, чтобы выбрать $\boldsymbol{\beta}^*$ таким образом, чтобы ошибка прогноза $y - \boldsymbol{x}'\boldsymbol{\beta}^*$ была ортогональна $\boldsymbol{x}$. Если $\mathrm{E}(\boldsymbol{x}\boldsymbol{x}')$ невырождена, условие ортогональности может быть разрешено относительно $\boldsymbol{\beta}^*$:

$$\boldsymbol{\beta}^* = [\mathrm{E}(\boldsymbol{x}\boldsymbol{x}')]^{-1}\,\mathrm{E}(\boldsymbol{x}\cdot y). \tag{2.9.6}$$

Проекция методом наименьших квадратов (или **линейная**) (*least squares (or linear) projection*) y на $\boldsymbol{x}$, обозначаемая $\widehat{\mathrm{E}}^*(y|\boldsymbol{x})$, определяется как $\boldsymbol{x}'\boldsymbol{\beta}^*$, где $\boldsymbol{\beta}^*$ удовлетворяет (2.9.5) и называется **коэффициентом проекции методом наименьших квадратов** (*least squares projection coefficients*).

Утверждение 2.8: *Проекция методом наименьших квадратов $\widehat{\mathrm{E}}^*(y|\boldsymbol{x})$ есть наилучший линейный предиктор y в том смысле, что он минимизирует среднеквадратичную ошибку.*

Стратегия прибавления — вычитания работает и здесь.

Доказательство
Для любого линейного предиктора $\boldsymbol{x}'\widetilde{\boldsymbol{\beta}}$

среднеквадратичная ошибка $\equiv \mathrm{E}\big[(y - \boldsymbol{x}'\widetilde{\boldsymbol{\beta}})^2\big] =$

$= \mathrm{E}\big\{\big[(y - \boldsymbol{x}'\boldsymbol{\beta}^*) + \boldsymbol{x}'(\boldsymbol{\beta}^* - \widetilde{\boldsymbol{\beta}})\big]^2\big\}$

(по стратегии прибавления — вычитания)

$= \mathrm{E}\big[(y - \boldsymbol{x}'\boldsymbol{\beta}^*)^2\big] + 2(\boldsymbol{\beta}^* - \widetilde{\boldsymbol{\beta}})'\,\mathrm{E}\big[\boldsymbol{x}\cdot(y - \boldsymbol{x}'\boldsymbol{\beta}^*)\big] + \mathrm{E}\big[(\boldsymbol{x}'(\boldsymbol{\beta}^* - \widetilde{\boldsymbol{\beta}}))^2\big] =$

$= \mathrm{E}\big[(y - \boldsymbol{x}'\boldsymbol{\beta}^*)^2\big] + \mathrm{E}\big[(\boldsymbol{x}'(\boldsymbol{\beta}^* - \widetilde{\boldsymbol{\beta}}))^2\big]$

(по условию ортогональности (2.9.5))

$\geq \mathrm{E}\big[(y - \boldsymbol{x}'\boldsymbol{\beta}^*)^2\big]$. ■

В отличие от наилучшего предиктора, который является условным математическим ожиданием, наилучший линейный предиктор требует для своего вычисления только знания второго момента совместного распределения $(y, \boldsymbol{x})$ для вычисления (см. (2.9.6)).

Если один из регрессоров $\boldsymbol{x}$ является константой, коэффициенты проекции метода наименьших квадратов могут быть записаны в терминах дисперсий и ковариаций. Пусть $\widetilde{\boldsymbol{x}}$ будет вектором непостоянных регрессоров, так что

$$\boldsymbol{x} = \begin{bmatrix} 1 \\ \widetilde{\boldsymbol{x}} \end{bmatrix}.$$

В аналитическом упражнении к этой главе вас просят доказать, что

$$\widehat{\mathrm{E}}^*(y|\boldsymbol{x}) \equiv \widehat{\mathrm{E}}^*(y|1, \widetilde{\boldsymbol{x}}) = \mu + \boldsymbol{\gamma}'\widetilde{\boldsymbol{x}}, \tag{2.9.7}$$

где

$$\boldsymbol{\gamma} = \mathrm{Var}(\widetilde{\boldsymbol{x}})^{-1}\,\mathrm{Cov}(\widetilde{\boldsymbol{x}}, y),\ \ \mu = \mathrm{E}(y) - \boldsymbol{\gamma}'\,\mathrm{E}(\widetilde{\boldsymbol{x}}).$$

Эта формула является популяционным аналогом формулы «регрессии в отклонениях от среднего» (*deviations from the mean*) из главы 1 (аналитическое упражнение 3).

OLS состоятельно оценивает коэффициенты проекции

Давайте возвратимся теперь к роли эконометриста и рассмотрим оценивание $\boldsymbol{\beta}^*$. Предположим, что у нас есть выборка объема n, извлеченная из эргодически стационарного случайного процесса $\{y_i, \boldsymbol{x}_i\}$ с совместным распределением $(y_i, \boldsymbol{x}_i)$ (не зависящим от i из-за стационарности), которое совпадает с распределением $(y, \boldsymbol{x})$ выше. Поэтому, например, $\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i') = \mathrm{E}(\boldsymbol{x}\boldsymbol{x}')$. По эргодической теореме вторые моменты в (2.9.6) могут быть состоятельно оценены соответствующими выборочными вторыми моментами. Таким образом, состоятельной оценкой вектора коэффициентов проекции $\boldsymbol{\beta}^*$ является:

$$\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\cdot y_i\right) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

которая есть не что иное, как OLS-оценка $\boldsymbol{b}$. То есть при предположении 2.2 (эргодическая стационарность) и предположении 2.4, гарантирующем невырожденность $\mathrm{E}(\boldsymbol{x}\boldsymbol{x}')$, OLS-оценка всегда состоятельна для вектора коэффициентов проекции $\boldsymbol{\beta}^*$, который удовлетворяет условию ортогональности (2.9.5).

Контрольные вопросы

1. (Непрогнозируемость ошибок прогноза.) Для прогноза с минимальным квадратом ошибки прогноза $\mathrm{E}(y|\boldsymbol{x})$ ошибка прогноза ортогональна любой функции $\phi(\boldsymbol{x})$ от $\boldsymbol{x}$. То есть $\mathrm{E}[\eta\phi(\boldsymbol{x})] = 0$, где $\eta = y - \mathrm{E}(y|\boldsymbol{x})$. Докажите это. **Указание:** *Закон полных математических ожиданий.*

   Покажите, что средний член в правой части (2.9.3) равен нулю, полагая $\phi(\boldsymbol{x}) = \mathrm{E}(y|\boldsymbol{x}) - f(\boldsymbol{x})$.

2. (Прогнозирование белого шума.) Предположим, что $\{\varepsilon_i\}$ — белый шум. Что есть $\widehat{\mathrm{E}}^*(\varepsilon_i|\varepsilon_{i-1}, \varepsilon_{i-2}, \ldots, \varepsilon_{i-m})$? Что есть $\widehat{\mathrm{E}}^*(\varepsilon_i|1, \varepsilon_{i-1}, \ldots, \varepsilon_{i-m})$? Верно ли, что $\mathrm{E}(\varepsilon_i|\varepsilon_{i-1}, \ldots, \varepsilon_{i-m}) = 0$? **Указание:** *Является ли оно нулем для процесса в примере 2.4?*

3. (Условные математические ожидания, которые линейны.) Предположим, что $\mathrm{E}(y|\tilde{\boldsymbol{x}}) = \mu + \boldsymbol{\gamma}'\tilde{\boldsymbol{x}}$. Покажите, что $\widehat{\mathrm{E}}^*(y|1, \tilde{\boldsymbol{x}}) = \mathrm{E}(y|\tilde{\boldsymbol{x}})$.

4. (Блочная проекция.) Рассмотрим модель $y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \boldsymbol{z}_i'\boldsymbol{\delta} + \varepsilon_i$ с $\mathrm{E}(\boldsymbol{x}_i \cdot \varepsilon_i) = 0$, $\mathrm{E}(\boldsymbol{z}_i \cdot \varepsilon_i) \neq 0$ и $\mathrm{E}(\boldsymbol{z}_i\boldsymbol{x}_i') = \mathbf{0}$. Таким образом, $\boldsymbol{z}_i$ не предопределен (то есть не ортогонален ошибке), но не связан с предопределенным регрессором $\boldsymbol{x}_i$ в том смысле, что перекрестные моменты равны нулю.

   (a) Покажите, что коэффициент проекции метода наименьших квадратов для $\boldsymbol{x}_i$ в проекции y_i на $\boldsymbol{x}_i$ и $\boldsymbol{z}_i$ равен $\boldsymbol{\beta}$. **Указание:** *Вычислите* $\widehat{\mathrm{E}}^*(\varepsilon_i|\boldsymbol{x}_i, \boldsymbol{z}_i)$.

   (b) Чему равен коэффициент проекции метода наименьших квадратов для $\boldsymbol{x}_i$ в проекции y_i на $\boldsymbol{x}_i$? **Указание:** *Трактуйте* $\boldsymbol{z}_i'\boldsymbol{\delta} + \varepsilon_i$ *как ошибку.*

   (c) Какую проекцию вы бы использовали для оценивания $\boldsymbol{\beta}$? **Указание:** *Вы хотите, чтобы дисперсия ошибки была поменьше.*

2.10. Тестирование сериальной корреляции

Как было отмечено в параграфе 2.3 (см. (2.3.3)), если регрессоры включают константу (это так практически во всех известных приложениях), из предположения 2.5 вытекает, что ошибка является скалярной последовательностью мартингал-разностей (m.d.s.), поэтому если обнаружена сериальная корреляция ошибок, это является указанием на нарушение предположения 2.5. Сериальная корреляция традиционно была важной темой в эконометрике, и имеется необходимое количество тестов на сериальную корреляцию (то есть проверка нулевой гипотезы об отсутствии сериальной корреляции у ошибок). Некоторые из них, однако, требуют, чтобы регрессоры были строго экзогенными. Тест, который будет представлен в этом параграфе, не требует экзогенности. Поскольку проблема сериальной корреляции возникает только в моделях временных рядов, мы используем в данном параграфе (и в следующих) индекс «t» вместо «i». Всюду в этом параграфе мы предполагаем, что регрессоры включают константу.

Было бы хорошо обобщить эти тесты для учета сериальной корреляции в $\boldsymbol{g}_i (\equiv \boldsymbol{x}_i \cdot \varepsilon_i)$, но таких тестов, которые получили бы признание, не было предложено. Это пробел в литературе, но не серьезный, поскольку в настоящее время исследователи знают, как жить с сериальной корреляцией в $\boldsymbol{g}_i$. То есть, как будет показано в главе 6, имеется способ сделать вывод в присутствии сериальной корреляции в $\boldsymbol{g}_i$. Ранее в этой главе мы изучали, как вычислять стандартные ошибки, робастные к условной гетероскедастичности. В главе 6 эти стандартные ошибки будут сделаны также робастными к сериальной корреляции.

Статистики Бокса — Пирса и Льюнга — Бокса

Прежде чем перейти к тестам на сериальную корреляцию ошибок, мы временно выходим за рамки регрессионной основы и рассмотрим сериальную корреляцию одномерных временных рядов. Предположим, что у нас есть выборка объема n, $\{z_1, \dots, z_n\}$, извлеченная из скалярного стационарного в ковариациях процесса. В параграфе 2.2 мы определили (популяционную) автоковариацию j-го порядка как γ_j. **Выборочная автоковариация j-го порядка** (*sample j-th order autocovariance*) определяется соотношением:

$$\widehat{\gamma}_j \equiv \frac{1}{n} \sum_{t=j+1}^{n} (z_t - \overline{z}_n)(z_{t-j} - \overline{z}_n) \quad (j = 0, 1, \dots), \tag{2.10.1}$$

где

$$\overline{z}_n \equiv \frac{1}{n} \sum_{t=1}^{n} z_t.$$

(Если популяционное среднее $\mathrm{E}(z_t)$ известно, можно заменить им выборочное среднее; делая это, можно улучшить свойства на конечных выборках.) Здесь, даже если в сумме используется только $n - j$ слагаемых, знаменателем является n, а не $n - j$. Делится ли сумма на $n - j$ или на n — это не влияет на распределение на больших выборках. Однако для выборок умеренного объема численное различие может быть существенным, и вы должны всегда быть точны в том, что используется в знаменателе. **Выборочный коэффициент автокорреляции j-го порядка** (*sample j-th order autocorrelation coefficient*), $\widehat{\rho}_j$, определяется как

$$\widehat{\rho}_j \equiv \frac{\widehat{\gamma}_j}{\widehat{\gamma}_0} \quad (j = 1, 2, \dots). \tag{2.10.2}$$

Если $\{z_t\}$ эргодически стационарный, то легко показать (см. контрольный вопрос), что $\widehat{\gamma}_j$ является состоятельной для γ_j $(j = 0, 1, 2, \dots)$. Следовательно, по лемме 2.3(a), $\widehat{\rho}_j$ является состоятельной для ρ_j $(j = 1, 2, \dots)$. В частности если $\{z_t\}$ сериально некоррелирована, то тогда все выборочные коэффициенты автокорреляции сходятся к 0 по вероятности. Для тестирования сериальной корреляции нам, однако, надо знать асимптотическое распределение $\sqrt{n}\widehat{\rho}_j$. Это обеспечивает

Утверждение 2.9 (частный случай теоремы 6.7 [Hall and Heyde, 1980]): *Предположим,что $\{z_t\}$ может быть записан как $\mu + \varepsilon_t$, где ε_t — стационарная последовательность мартингал-разностей с «собственной» условной гомоскедастичностью:*

(собственная условная гомоскедастичность)

$$\mathrm{E}(\varepsilon_t^2 | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) = \sigma^2, \ \sigma^2 > 0.$$

Пусть выборочная автокорреляция $\widehat{\rho}_j$ определяется, как в (2.10.1) *и* (2.10.2). *Тогда*

$$\sqrt{n}\widehat{\gamma} \underset{d}{\to} N(\mathbf{0}, \sigma^4 \boldsymbol{I}_p) \quad \text{и} \quad \sqrt{n}\widehat{\boldsymbol{\rho}} \underset{d}{\to} N(\mathbf{0}, \boldsymbol{I}_p),$$

где $\widehat{\boldsymbol{\gamma}} = (\widehat{\gamma}_1, \widehat{\gamma}_2, \ldots, \widehat{\gamma}_p)'$ *и* $\widehat{\boldsymbol{\rho}} = (\widehat{\rho}_1, \widehat{\rho}_2, \ldots \widehat{\rho}_p)'$.

Здесь от процесса ε_t не требуется быть эргодическим. Доказательство этого при дополнительном условии эргодичности остается в качестве аналитического упражнения. Таким образом, умноженные на $\sqrt{n}$ автокорреляции асимптотически являются i.i.d. и имеют распределение $N(0, 1)$. Процесс $\{\varepsilon_t\}$ предполагается здесь более общим, чем независимый процесс белого шума, но условный второй момент должен быть константой, поэтому этот результат, например, не охватывает ARCH-процессы.

Один из способов тестирования сериальной корреляции в ряде — проверить, равна ли 0 автокорреляция первого порядка, ρ_1. Утверждение 2.9 влечет

$$\sqrt{n}\widehat{\rho}_1 = \frac{\widehat{\rho}_1}{1/\sqrt{n}} \underset{\mathrm{d}}{\to} N(0, 1). \tag{2.10.3}$$

Поэтому t-статистика, сформированная как отношение $\widehat{\rho}_1$ к «стандартной ошибке» $1/\sqrt{n}$, является асимптотически стандартной нормальной.

Мы можем также тестировать на одновременное равенство нулю целой группы автокорреляций. Пусть $\widehat{\boldsymbol{\rho}} = (\widehat{\rho}_1, \ldots, \widehat{\rho}_p)'$ является p-мерным вектором, содержащим первые p выборочных автокорреляций. Так как компоненты $\sqrt{n}\widehat{\boldsymbol{\rho}}$ асимптотически независимы и одинаково распределены как стандартные нормальные, их сумма квадратов, называемая $\boldsymbol{Q}$ **Бокса — Пирса** (*Box — Pierce Q*), поскольку она была впервые предложена в [Box and Pierce, 1970], имеет асимптотическое распределение хи-квадрат:

$$\text{статистика } Q \text{ Бокса — Пирса} \equiv n\sum_{j=1}^{p} \widehat{\rho}_j^2 = \sum_{j=1}^{p} \left(\sqrt{n}\widehat{\rho}_j\right)^2 \underset{\mathrm{d}}{\to} \chi^2(p). \tag{2.10.4}$$

Легко показать (см. контрольный вопрос), что следующая модификация, называемая $\boldsymbol{Q}$ **Льюнга — Бокса** (Ljung — Box Q), является асимптотически эквивалентной в том смысле, что ее отличие от Q Бокса — Пирса исчезает на больших выборках. Поэтому, по лемме 2.4(a), она также имеет асимптотическое распределение хи-квадрат:

$$\text{Статистика } Q \text{ Льюнга — Бокса} \equiv$$

$$\equiv n \cdot (n+2) \sum_{j=1}^{p} \frac{\widehat{\rho}_j^2}{n-j} = \sum_{j=1}^{p} \frac{n+2}{n-j} \left(\sqrt{n}\widehat{\rho}_j\right)^2 \underset{\mathrm{d}}{\to} \chi^2(p). \tag{2.10.5}$$

Эта модификация часто дает лучшую аппроксимацию распределения хи-квадрат для умеренного объема выборки (вам будет предложено проверить это в упражнении Монте-Карло). Для каждой из статистик нет четкого руководства по выбору p. Если p слишком мало, то есть опасность пропуска существующей автокорреляции высокого порядка, но если p слишком велико относительно объема выборки, распределение статистики на конечных выборках, вероятно, ухудшится, значительно удаляясь от распределения хи-квадрат.

Выборочные автокорреляции, вычисленные с помощью остатков

Вернемся теперь к регрессионной модели, описанной предположениями 2.1–2.5. Если бы ошибка ε_t была наблюдаема, мы могли бы вычислить выборочные автокорреляции как

$$\widetilde{\rho}_j \equiv \frac{\widetilde{\gamma}_j}{\widetilde{\gamma}_0} \quad (j = 1, 2, \ldots), \tag{2.10.6}$$

где

$$\widetilde{\gamma}_j \equiv \frac{1}{n} \sum_{t=j+1}^{n} \varepsilon_t \varepsilon_{t-j} \quad (j = 0, 1, 2, \ldots). \tag{2.10.7}$$

(Нет необходимости вычитать выборочное среднее, поскольку популяционное среднее равно нулю, как результат включения константы в число регрессоров.) Так как $\{\varepsilon_t \varepsilon_{t-j}\}$ эргодически стационарен, по предположению 2.2, $\widetilde{\gamma}_j$ сходится по вероятности к соответствующему популяционному среднему, $\mathrm{E}(\varepsilon_t \varepsilon_{t-j})$, для всех j, и $\widetilde{\rho}_j$ является состоятельной для популяционного j-го коэффициента автокорреляции для ε_t.

Далее рассмотрим более реалистичный случай, когда мы не наблюдаем ошибку. Мы можем заменить ε_t в формуле выше на его OLS-оценку e_t и вычислить выборочную автокорреляцию как

$$\widehat{\rho}_j \equiv \frac{\widehat{\gamma}_j}{\widehat{\gamma}_0} \quad (j = 1, 2, \ldots), \tag{2.10.8}$$

где

$$\widehat{\gamma}_j \equiv \frac{1}{n} \sum_{t=j+1}^{n} e_t e_{t-j} \quad (j = 0, 1, 2, \ldots). \tag{2.10.9}$$

(Поскольку регрессоры включают константу, нормальное уравнение, соответствующее константе, гарантирует, что выборочное среднее e_t нулевое. Поэтому нет необходимости вычитать выборочное среднее.) Правильно ли использовать $\widehat{\rho}_j$ (вычисленное с помощью остатков) вместо $\widetilde{\rho}_j$

и основанную на остатках статистику Q, полученную из $\{\widehat{\rho}_j\}$, для тестирования сериальной корреляции? [Ответ: Да. Но только если регрессоры строго экзогенны.]

Вспомним выражение, связывающее остаток e_t и действительную ошибку ε_t, которое было дано в (2.3.9). Используя его, можно записать:

$$\begin{aligned}\widehat{\gamma}_j &\equiv \frac{1}{n}\sum_{t=j+1}^{n} e_t\, e_{t-j} = \\ &= \frac{1}{n}\sum_{t=j+1}^{n}[\varepsilon_t - \boldsymbol{x}_t'(\boldsymbol{b}-\boldsymbol{\beta})][\varepsilon_{t-j} - \boldsymbol{x}_{t-j}'(\boldsymbol{b}-\boldsymbol{\beta})] = \\ &= \widetilde{\gamma}_j - \frac{1}{n}\sum_{t=j+1}^{n}(\boldsymbol{x}_{t-j}\cdot\varepsilon_t + \boldsymbol{x}_t\cdot\varepsilon_{t-j})'(\boldsymbol{b}-\boldsymbol{\beta}) + \\ &\quad + (\boldsymbol{b}-\boldsymbol{\beta})'\Big(\frac{1}{n}\sum_{t=j+1}^{n}\boldsymbol{x}_t\,\boldsymbol{x}_{t-j}'\Big)(\boldsymbol{b}-\boldsymbol{\beta}).\end{aligned} \tag{2.10.10}$$

Если $\mathrm{E}(\boldsymbol{x}_t\cdot\varepsilon_{t-j})$, $\mathrm{E}(\boldsymbol{x}_{t-j}\cdot\varepsilon_t)$ и $\mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}_{t-j}')$ все конечны, то вторая и третья компоненты исчезают (сходятся к нулю по вероятности), поскольку $\boldsymbol{b}-\boldsymbol{\beta} \to_{\mathrm{p}} \mathbf{0}$. Следовательно,

$$\widehat{\gamma}_j - \widetilde{\gamma}_j \underset{\mathrm{p}}{\to} 0 \quad (j = 0, 1, 2, \ldots),$$

и, таким образом, различие между $\widetilde{\rho}_j$ и $\widehat{\rho}_j$ также исчезает на больших выборках.

Однако это будет не так, если обе эти величины умножить на $\sqrt{n}$. $\sqrt{n}\widetilde{\rho}_j$ и $\sqrt{n}\widehat{\rho}_j$ могут быть записаны как

$$\sqrt{n}\widetilde{\rho}_j = \frac{\sqrt{n}\widetilde{\gamma}_j}{\widetilde{\gamma}_0} \quad \text{и} \quad \sqrt{n}\widehat{\rho}_j = \frac{\sqrt{n}\widehat{\gamma}_j}{\widehat{\gamma}_0}. \tag{2.10.11}$$

Мы уже показали, что для обоих $\widetilde{\rho}_j$ и $\widehat{\rho}_j$ знаменатель $\to_{\mathrm{p}} \sigma^2$. Поэтому различие между $\widetilde{\rho}_j$ и $\widehat{\rho}_j$ будет исчезать, если будет исчезать различие между $\widetilde{\gamma}_j$ и $\widehat{\gamma}_j$ (если вы не уверены, см. контрольный вопрос 3 ниже). Теперь, умножая обе стороны (2.10.10) на $\sqrt{n}$, мы получаем:

$$\begin{aligned}\sqrt{n}\widehat{\gamma}_j &= \sqrt{n}\widetilde{\gamma}_j - \frac{1}{n}\sum_{t=j+1}^{n}(\boldsymbol{x}_{t-j}\cdot\varepsilon_t + \boldsymbol{x}_t\cdot\varepsilon_{t-j})'\sqrt{n}\,(\boldsymbol{b}-\boldsymbol{\beta}) + \\ &\quad + \sqrt{n}\,(\boldsymbol{b}-\boldsymbol{\beta})'\Big(\frac{1}{n}\sum_{t=j+1}^{n}\boldsymbol{x}_t\,\boldsymbol{x}_{t-j}'\Big)(\boldsymbol{b}-\boldsymbol{\beta}).\end{aligned} \tag{2.10.12}$$

Поскольку $\sqrt{n}(\boldsymbol{b}-\boldsymbol{\beta})$ сходится к случайной величине (распределение которой нормальное), третья компонента в правой части исчезает, по лемме 2.4(b). Что касается второй компоненты, мы имеем:

$$\frac{1}{n}\sum_{t=j+1}^{n}(\boldsymbol{x}_{t-j}\cdot\varepsilon_t+\boldsymbol{x}_t\cdot\varepsilon_{t-j})\xrightarrow[\mathrm{p}]{}\mathrm{E}(\boldsymbol{x}_{t-j}\cdot\varepsilon_t)+\mathrm{E}(\boldsymbol{x}_t\cdot\varepsilon_{t-j}). \tag{2.10.13}$$

Если регрессоры строго экзогенны в том смысле, что $\mathrm{E}(\boldsymbol{x}_t\cdot\varepsilon_s)=\mathbf{0}$ для всех t,s, то

$$\mathrm{E}(\boldsymbol{x}_{t-j}\cdot\varepsilon_t)+\mathrm{E}(\boldsymbol{x}_t\cdot\varepsilon_{t-j})=\mathbf{0}, \tag{2.10.14}$$

и, по лемме 2.4(b), вторая компонента сходится к нулю по вероятности. Поэтому различие между $\sqrt{n}\widetilde{\rho}_j$ и $\sqrt{n}\widehat{\rho}_j$ исчезает, и это означает, что статистика Q, вычисленная с помощью остатков $\{e_t\}$, также имеет асимптотическое распределение хи-квадрат, и мы можем использовать эту основанную на остатках статистику Q для тестирования сериальной корреляции. Если, с другой стороны, регрессоры не являются строго экзогенными, то тогда нет гарантии того, что (2.10.14) выполняется. Следовательно, основанная на остатках статистика Q может не иметь асимптотическое распределение хи-квадрат.

Тестирование с предопределенными, но не строго экзогенными регрессорами

Таким образом, если регрессоры не являются строго экзогенными, нам необходимо модифицировать статистику Q, чтобы восстановить ее асимптотическое распределение. Для этой цели рассмотрим два ограничения:

(более сильная форма предопределенности)

$$\mathrm{E}(\varepsilon_t|\varepsilon_{t-1},\varepsilon_{t-2},\ldots,\boldsymbol{x}_t,\boldsymbol{x}_{t-1},\ldots)=0. \tag{2.10.15}$$

(более сильная форма условной гомоскедастичности)

$$\mathrm{E}(\varepsilon_t^2|\varepsilon_{t-1},\varepsilon_{t-2},\ldots,\boldsymbol{x}_t,\boldsymbol{x}_{t-1},\ldots)=\sigma^2>0. \tag{2.10.16}$$

Первое условие просто воспроизводит (2.3.1) (только теперь с «t» в качестве индекса). Как было показано в параграфе 2.3, оно более сильное, чем предположение 2.3, и из него вытекает, что $\boldsymbol{g}_t(\equiv\boldsymbol{x}_t\cdot\varepsilon_t)$ является m.d.s. Условие (2.10.16) с условным множеством, включающим $\boldsymbol{x}_t$, очевидно более сильное, чем предположение 2.7 (условная гомоскедастичность). Оно также более сильное, чем предположение о собственной условной гомоскедастичности в утверждении 2.9, поскольку условное множество включает текущие и прошлые $\boldsymbol{x}$, так же как и прошлые ε. Следующий результат показывает, что при этих дополнительных условиях существует соответствующая модификация статистики Q.

Утверждение 2.10 (тестирование сериальной корреляции с предопределенными регрессорами): *Предположим, что удовлетворяются предположения 2.1, 2.2, 2.4,* (2.10.15) *и* (2.10.16). *Пусть выборочная автокорреляция OLS-остатков, $\widehat{\rho}_j$, определяется, как в* (2.10.8). *Тогда*

$$\sqrt{n}\widehat{\boldsymbol{\gamma}} \underset{d}{\to} N(\mathbf{0}, \sigma^4 \cdot (\boldsymbol{I}_p - \boldsymbol{\Phi})) \text{ и } \sqrt{n}\widehat{\boldsymbol{\rho}} \underset{d}{\to} N(\mathbf{0}, \boldsymbol{I}_p - \boldsymbol{\Phi}), \tag{2.10.17}$$

где $\widehat{\boldsymbol{\gamma}} = (\widehat{\gamma}_1, \widehat{\gamma}_2, \ldots, \widehat{\gamma}_p)'$, $\widehat{\boldsymbol{\rho}} = (\widehat{\rho}_1, \widehat{\rho}_2, \ldots, \widehat{\rho}_p)'$ и ϕ_{jk} (элемент $\equiv (j, k)$ матрицы $\boldsymbol{\Phi}$ размерности $p \times p$) задается как

$$\phi_{jk} = \mathrm{E}(\boldsymbol{x}_t \cdot \varepsilon_{t-j})' \, \mathrm{E}(\boldsymbol{x}_t \boldsymbol{x}_t')^{-1} \, \mathrm{E}(\boldsymbol{x}_t \cdot \varepsilon_{t-k})/\sigma^2. \tag{2.10.18}$$

Доказательство, хотя и не очень сложное, перенесено в приложение 2.B. По эргодической теореме, матрица $\boldsymbol{\Phi}$ состоятельно оценивается своим выборочным аналогом:

$$\widehat{\boldsymbol{\Phi}} \equiv (\widehat{\phi}_{jk}), \quad \widehat{\phi}_{jk} \equiv \overline{\boldsymbol{\mu}}_j' \boldsymbol{S}_{\boldsymbol{xx}}^{-1} \overline{\boldsymbol{\mu}}_k / s^2 \quad (j, k = 1, 2, \ldots, p), \tag{2.10.19}$$

где

$$s^2 \equiv \frac{1}{n-K} \sum_{t=1}^{n} e_t^2, \quad \overline{\boldsymbol{\mu}}_j \equiv \frac{1}{n} \sum_{t=j+1}^{n} \boldsymbol{x}_t \cdot e_{t-j}.$$

Из этого и из предположения 2.10 следует, что

$$\text{модифицированная } Q \text{ Бокса — Пирса} \equiv n \cdot \widehat{\boldsymbol{\rho}}' (\boldsymbol{I}_p - \widehat{\boldsymbol{\Phi}})^{-1} \widehat{\boldsymbol{\rho}} \underset{\mathrm{d}}{\to} \chi^2(p). \tag{2.10.20}$$

Пока регрессоры предопределены и ошибки условно гомоскедастичны в смысле (2.10.15) и (2.10.16), эта модифицированная статистика Q может быть использована, когда регрессоры не являются строго экзогенными.

Тест, основанный на вспомогательной регрессии

Хотя вычисление этой модифицированной статистики Q выполняется непосредственно с помощью матричного программного обеспечения, полезно найти асимптотически эквивалентную статистику, которая может быть вычислена в регрессионных пакетах. Для этой цели рассмотрим вспомогательную регрессию:

$$\text{регрессия } e_t \text{ на } \boldsymbol{x}_t, e_{t-1}, e_{t-2}, \ldots, e_{t-p}. \tag{2.10.21}$$

Чтобы прогнать эту вспомогательную регрессию для $t = 1, 2, \ldots, n$, нам нужны данные о $(e_0, e_{-1}, \ldots, e_{-p+1})$. Асимптотически, не имеет значения, какое конкретное значение им присвоить, но кажется разумным положить их равными 0, их ожидаемому значению[1]. Из этой вспомогательной регрессии мы можем вычислить F-статистику для гипотезы о том, что все p коэффициентов при $e_{t-1}, e_{t-2}, \ldots, e_{t-p}$ равны нулю. Учитывая утверждение 2.5(c), вполне естественно задаться вопросом: имеет ли $p \cdot F$ асимптотическое распределение $\chi^2(p)$? Эта догадка действительно верна: при гипотезе в утверждении 2.10 модифицированная статистика Q (2.10.20) асимптотически эквивалентна $p \cdot F$ (то есть разность между ними сходится к нулю по вероятности при $n \to \infty$), поэтому $p \cdot F$ также имеет асимптотическое распределение хи-квадрат (демонстрации этого посвящено аналитическое упражнение).

Эта $p \cdot F$-статистика оказывается асимптотически эквивалентной nR^2 из вспомогательной регрессии, что может быть показано следующим образом. Вспомним алгебраический результат из главы 1 о том, что F-отношение может быть вычислено на основе разности сумм квадратов остатков от неограниченной и ограниченной регрессий. Неограниченная регрессия в текущем контексте есть (2.10.21), в то время как ограниченная регрессия есть

$$\text{регрессия } e_t \text{ на } \boldsymbol{x}_t. \tag{2.10.22}$$

Следовательно, если SSR_U и SSR_R являются SSR из (2.10.21) и (2.10.22), соответственно, мы имеем:

$$p \cdot F = \frac{SSR_R - SSR_U}{SSR_U/(n - \#\boldsymbol{x}_t - p)} = (n - \#\boldsymbol{x}_t - p)\frac{SSR_R - SSR_U}{SSR_U}, \tag{2.10.23}$$

где $\#\boldsymbol{x}_t$ — число переменных в $\boldsymbol{x}_t$. Однако поскольку e_t является остатком от первоначальной регрессии (регрессии y_t на $\boldsymbol{x}_t$), регрессоры $\boldsymbol{x}_t$ в (2.10.22) не имеют объясняющей силы. Поэтому $SSR_R = \boldsymbol{e}'\boldsymbol{e}$, где $\boldsymbol{e}$ — n-мерный вектор остатков из первоначальной регрессии, а (2.10.23) численно идентична

$$(n - \#\boldsymbol{x}_t - p)\frac{R^2_{uc}}{1 - R^2_{uc}},$$

где R^2_{uc} — нецентрированный R^2 для вспомогательной регрессии (2.10.21)[2]. Но поскольку, по построению, выборочное среднее e_t равно нулю (это потому, что $\boldsymbol{x}_t$ включает константу), R^2_{uc} численно идентичен R^2 для вспо-

[1] Другой асимптотически эквивалентный выбор заключается в проведении вспомогательной регрессии для $t = p+1, p+2, \ldots, n$.

[2] Вспомним из (1.2.16), что нецентрированный R^2 для регрессии $\boldsymbol{y}$ на $\boldsymbol{x}$ определяется как

$$R^2_{uc} \equiv \frac{\boldsymbol{y}'\boldsymbol{y} - \boldsymbol{e}'\boldsymbol{e}}{\boldsymbol{y}'\boldsymbol{y}}.$$

В текущем контексте $\boldsymbol{y}'\boldsymbol{y}$ в этой формуле есть $\boldsymbol{e}'\boldsymbol{e}$, в то время как $\boldsymbol{e}'\boldsymbol{e}$ в формуле есть SSR_U.

могательной регрессии. Таким образом, мы доказали следующий алгебраический результат:

$$p \cdot F = (n - \#\boldsymbol{x}_t - p)\frac{R^2}{1 - R^2},$$

где R^2 равен R^2 во вспомогательной регрессии (2.10.21). Решая это уравнение относительно R^2 и умножая обе стороны на n, получаем:

$$nR^2 = \frac{n}{n - \#\boldsymbol{x}_t - p} \cdot \frac{1}{1 + \frac{p \cdot F}{n - \#\boldsymbol{x}_t - p}} \cdot p \cdot F.$$

Поскольку $\frac{p \cdot F}{n - \#\boldsymbol{x}_t - p} \to_0$ при $n \to \infty$ (как следствие леммы 2.4(b)), это уравнение показывает, что $p \cdot F - nR^2 \to_{\mathrm{p}} 0$. Следовательно, nR^2 из вспомогательной регрессии (2.10.21) имеет асимптотическое распределение $\chi^2(p)$. Тест, основанный на nR^2, называется **тестом Бройша — Годфри на сериальную корреляцию** (*Breusch — Godfrey test for serial correlation*). Когда $p = 1$, эта тестовая статистика асимптотически эквивалентна квадрату статистики, известной как **статистика h Дарбина** (*Durbin's h*)[1].

Контрольные вопросы

1. (Состоятельность выборочных автоковариаций.) Покажите: если $\{z_t\}$ эргодически стационарный, то тогда выборочная автоковариация $\widehat{\gamma}_j$, заданная в (2.10.1), является состоятельной для γ_j, популяционной автоковариации. **Указание:** $\gamma_j = \mathrm{E}(z_t z_{t-j}) - \mathrm{E}(z_t)\,\mathrm{E}(z_{t-j})$. *Перепишите $\widehat{\gamma}_j$ как*

$$\widehat{\gamma}_j = \frac{1}{n}\sum_{t=j+1}^{n} z_t z_{t-j} - \overline{z}_n \frac{1}{n}\sum_{t=j+1}^{n} z_{t-j} - \overline{z}_n \frac{1}{n}\sum_{t=j+1}^{n} z_t + \frac{n-j}{n}(\overline{z}_n)^2.$$

2. (Асимптотическая эквивалентность двух Q.) Докажите, что разность между Q Бокса — Пирса и Q Льюнга — Бокса сходится к 0 по вероятности при $n \to \infty$ (поэтому их асимптотические распределения одинаковые). **Указание:** *Пусть*

$$\boldsymbol{x}_n \equiv \left(\left(\sqrt{n}\widehat{\rho}_1\right)^2, \ldots, \left(\sqrt{n}\widehat{\rho}_p\right)^2\right)'.$$

Найдите p-мерный вектор $\boldsymbol{a}_n$, такой, чтобы разность между двумя Q была $\boldsymbol{a}_n'\boldsymbol{x}_n$. Покажите, что $\boldsymbol{a}_n \to_p \boldsymbol{0}$. Асимптотическое свойство, которое нужно использовать, — лемма 2.4(b).

[1]См.: [Breusch, 1978] и [Godfrey, 1978]. Тест Бройша — Годфри был первоначально разработан как тест множителей Лагранжа для случая нормально распределенных ошибок, состоящих из фиксированных регрессоров x и лагированной зависимой переменной. Обсуждение в тексте показывает, что тест применим к более общему случаю, рассмотренному в утверждении 2.10.

3. Рассмотрим $\sqrt{n}\widetilde{\rho}_j$ и $\sqrt{n}\widehat{\rho}_j$ в (2.10.11). Мы показали, что $\widehat{\gamma}_0 \to_p \sigma^2$ и $\widetilde{\gamma}_0 \to_p \sigma^2$. По утверждению 2.9 $\sqrt{n}\widetilde{\gamma}_j \to_d N(0, \sigma^4)$. Принимая эти утверждения как доказанные, покажите, что $\sqrt{n}\widetilde{\rho}_j - \sqrt{n}\widehat{\rho}_j \to_p 0$, если $\sqrt{n}\widetilde{\gamma}_j - \sqrt{n}\widehat{\gamma}_j \to_p 0$. **Указание:** *Если* $\sqrt{n}\widetilde{\gamma}_j - \sqrt{n}\widehat{\gamma}_j \to_p 0$, *то, по лемме 2.4(a),* $\sqrt{n}\widehat{\gamma}_j \to_d N(0, \sigma^4)$.

$$\frac{\widetilde{\gamma}_j}{\widetilde{\gamma}_0} - \frac{\widehat{\gamma}_j}{\widehat{\gamma}_0} = \widetilde{\gamma}_j \left(\frac{1}{\widetilde{\gamma}_0} - \frac{1}{\sigma^2} \right) - \widehat{\gamma}_j \left(\frac{1}{\widehat{\gamma}_0} - \frac{1}{\sigma^2} \right) + \frac{1}{\sigma^2}(\widetilde{\gamma}_j - \widehat{\gamma}_j).$$

Умножьте обе стороны на $\sqrt{n}$ *и примените лемму 2.4(b).*

2.11. Приложение: эконометрика рациональных ожиданий

В динамических экономических моделях и в других контекстах ожидания относительно будущего, естественно, играют важную роль. **Эконометрика рациональных ожиданий** (*rational expectations econometrics*) относится к уравнениям, включающим ожидания. Ожидания обычно ненаблюдаемы, но, если мы предполагаем, что экономические агенты формируют ожидания рационально, мы можем преодолеть проблему ненаблюдаемых ожиданий, используя технику этой главы. Приложением, которое мы рассмотрим, является **гипотеза эффективности рынка** (*efficient market hypothesis*) Фамы. По его словам, «эффективным рынком капитала является рынок, который эффективен при обработке информации.<...> В эффективном рынке цены «в полной мере отражают доступную информацию» [Fama, 1976, p. 133]. Слова «в полной мере отражают» могут быть точно формализованы. Мы будем делать это для конкретного рынка капитала и тестировать следствия из этой гипотезы.

Гипотеза эффективности рынка

Рынок капитала, изученный в [Fama, 1975], является рынком казначейских векселей США. В этом параграфе мы сосредоточимся на ставках по одномесячным казначейским векселям, наблюдаемых на основе месячного базиса. (Позднее в этой книге мы также рассмотрим казначейские векселя с различными сроками погашения.) Для формализации гипотезы эффективности рынка для рынка казначейских векселей нам нужно ввести некоторые понятия из макроэкономики и финансов. Поскольку в данном параграфе мы имеем дело исключительно с временными рядами, в качестве индекса используем «t», а не «i». Определим:

$v_t \equiv$ цена одномесячных казначейских векселей[1] на начало месяца t,
R_t = одномесячная номинальная процентная ставка за месяц t,
то есть номинальная доходность по векселю за месяц, равная
$(1 - v_t)/v_t$, так что $v_t = 1/(1 + R_t)$,

[1]Из приведенных ниже формул следует, что имеются в виду векселя номинальной стоимостью 1.

$P_t \equiv$ значение индекса потребительских цен (ИПЦ), который является нашей мерой уровня цен на начало месяца t,

$\pi_{t+1} \equiv$ темп инфляции за месяц t (то есть с начала месяца t по начало месяца $t+1) = (P_{t+1} - P_t)/P_t$,

${}_t\pi_{t+1} \equiv$ ожидаемый темп инфляции за месяц t, ожидание формируется в начале месяца t,

$\eta_{t+1} \equiv$ ошибка прогноза инфляции $= \pi_{t+1} - {}_t\pi_{t+1}$,

$r_{t+1} \equiv$ ex post реальная ставка процента за месяц

$$t = \frac{1/P_{t+1} - v_t/P_t}{v_t/P_t} = \frac{1+R_t}{1+\pi_{t+1}} - 1 \approx R_t - \pi_{t+1},$$

$${}_tr_{t+1} \equiv \text{ex ante реальная ставка процента } t = \frac{1+R_t}{1+{}_t\pi_{t+1}} - 1 \approx$$

$$\approx R_t - {}_t\pi_{t+1} = \frac{1+R_t}{1+\pi_{t+1}} - 1 \approx R_t - \pi_{t+1}.$$

В эконометрике рациональных ожиданий очень важна синхронизация — указание того, значение какой переменной наблюдается в какой момент времени. Наше правило для датированных переменных (которое отлично от Фамы) заключается в том, что переменная имеет индекс t, если ее значение является первым наблюдаемым на начало периода (месяца[1]) t. Например, поскольку π_{t+1}, темп инфляции за месяц t зависит от P_{t+1}, его индексом является $t+1$, а не t. По той же причине r_{t+1}, ex post реальная процентная ставка за месяц t, имеет индекс $t+1$. В противоположность этому номинальная процентная ставка R_t за месяц t имеет индекс t, а не $t+1$, поскольку она определена ценой казначейского векселя на начало месяца t, v_t. Рис. 2.2 показывает, значение какой переменной наблюдается в какой момент времени.

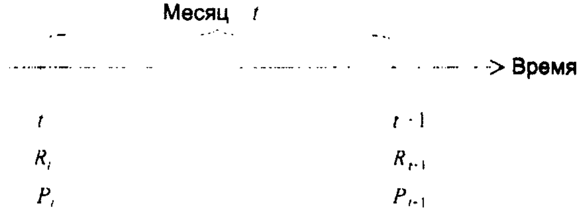

Рис. 2.2. Что когда наблюдается

Мы берем в качестве периода месяц только потому, что наши данные ежемесячные. Заметим, что срок погашения ценной бумаги (один месяц) совпадает с интервалом выборки (месяц). Если интервал выборки меньше, чем срок погашения, что имеет место, например, когда у нас

[1] В [Fama, 1975] наше π_{t+1} есть $-\widetilde{\Delta}_t$, r_{t+1} есть $\widetilde{r}_{t+1}$ и I_t есть ϕ_t.

есть данные по ставке трехмесячных казначейских векселей, то тогда сроки погашения перекрываются и нам нужна более сложная техника, вводимая в главе 6.

Гипотеза эффективности рынка есть совместная гипотеза, объединяющая:

Рациональные ожидания. Инфляционные ожидания рациональны: ${}_t\pi_{t+1} = \mathrm{E}(\pi_{t+1}|I_t)$, где I_t — информация, доступная на начало месяца t и включающая $\{R_t, R_{t-1}, \ldots, \pi_t, \pi_{t-1}, \ldots\}$. К тому же $I_t \supseteq I_{t-1} \supseteq \supseteq I_{t-2} \ldots$. То есть агенты, участвующие в рынке, не забывают прошлое.

Постоянные реальные ставки. Ex ante реальная процентная ставка постоянна: ${}_tr_{t+1} = r$.

Тестируемые следствия

Мы получим два тестируемых следствия гипотезы эффективности рынка, используя следующие ключевые наблюдения относительно ошибки прогноза инфляции при рациональных ожиданиях:

(a) $\mathrm{E}(\eta_{t+1}|I_t) = 0$, то есть ошибка прогноза инфляции является последовательностью мартингал-разностей относительно информационного множества.

(b) $\{\eta_t, \eta_{t+1}, \ldots\}$, так же как и $\{R_t, R_{t-1}, \ldots, \pi_t, \pi_{t-1}, \ldots\}$, включены в I_t (известные на начало месяца t). То есть агенты помнят все свои прошлые ошибки.

(a) Имеет интуитивный смысл; если для прогнозирования темпа инфляции люди используют всю доступную информацию, ошибка прогноза, которая известна только постфактум, не имеет какой-либо систематической связи с тем, что люди знали, когда они формировали ожидания. Это можно легко доказать:

$$
\begin{aligned}
&\mathrm{E}(\eta_{t+1}|I_t) = \\
&= \mathrm{E}(\pi_{t+1} - {}_t\pi_{t+1}|I_t) \quad (\text{поскольку } \eta_{t+1} \equiv \pi_{t+1} - {}_t\pi_{t+1}) = \\
&= \mathrm{E}(\pi_{t+1}|I_t) - \mathrm{E}({}_t\pi_{t+1}|I_t) = \\
&= \mathrm{E}(\pi_{t+1}|I_t) - \mathrm{E}[\mathrm{E}(\pi_{t+1}|I_t)|I_t] \\
&\qquad (\text{поскольку } {}_t\pi_{t+1} = \mathrm{E}(\pi_{t+1}|I_t) \text{ в силу рациональности ожиданий}) \\
&= \mathrm{E}(\pi_{t+1}|I_t) - \mathrm{E}(\pi_{t+1}|I_t) = 0. \qquad (2.11.1)
\end{aligned}
$$

(b) Вытекает из того, что агенты помнят прошлые прогнозы инфляции. Поскольку ${}_{t-j-1}\pi_{t-j} = \mathrm{E}(\pi_{t-j}|I_{t-j-1})$ является функцией от I_{t-j-1}, она включена в I_{t-j-1} (то есть известна на начало месяца $t-j-1$) и, следовательно, в I_t (известна на начало месяца t) для $j \geq 0$.

Таким образом, $\eta_{t-j} \equiv \pi_{t-j} - {}_{t-j-1}\pi_{t-j}$ включена в I_t для всех $j \geq 0$. Наблюдения (a) и (b) вместе с законом повторных математических означают, что

(c) $\{\eta_t\}$ является m.d.s.

Поэтому он является процессом с нулевым средним и сериально некоррелированным.

Это непроверяемо, поскольку ожидания ненаблюдаемы, но, когда они объединены с предположением о постоянной реальной ставке, они влекут два тестируемых следствия относительно темпа инфляции и процентной ставки, которые уже наблюдаемы.

Следствие 1: Ex post реальная процентная ставка имеет постоянное среднее и сериально некоррелирована.

Это следует из того, что

$$\begin{aligned} r_{t+1} &\equiv R_t - \pi_{t+1} = \\ &= (R_t - {}_t\pi_{t+1}) - (\pi_{t+1} - {}_t\pi_{t+1}) = \\ &= {}_t r_{t+1} - \eta_{t+1} \quad \text{(по определению)} = \\ &= r - \eta_{t+1} \quad \text{(так как } {}_t r_{t+1} = r \text{ по предположению).} \end{aligned} \tag{2.11.2}$$

В силу (c), $\{r_t\}$ имеет среднее r и сериально некоррелирована.

Следствие 2: $\mathrm{E}(\pi_{t+1}|I_t) = -r + R_t$.

То есть наилучшим (в смысле минимальной среднеквадратичной ошибки) предиктором инфляции на основе всей доступной информации является номинальная процентная ставка: номинальная ставка R_t, которая определена ценой казначейского векселя v_t, объединяет всю текущую доступную информацию, имеющую отношение к прогнозированию будущей инфляции. Это является формализацией того, что цены на активы «полностью отражают» доступную информацию. Указанное утверждение также может быть легко получено. Решая (2.11.2) относительно π_{t+1}, получаем: $\pi_{t+1} = -r + R_t + \eta_{t+1}$. Поэтому

$$\begin{aligned} &\mathrm{E}(\pi_{t+1}|I_t) = \mathrm{E}(-r + R_t + \eta_{t+1}|I_t) = \\ &= -r + R_t + \mathrm{E}(\eta_{t+1}|I_t) \\ &\text{(поскольку } r \text{ есть константа, а } R_t \text{ включена в } I_t\text{)} \\ &= -r + R_t \quad \text{(из (a)).} \end{aligned} \tag{2.11.3}$$

В оставшейся части этого параграфа мы тестируем эти два следствия эффективности рынка.

Тестирование на сериальную корреляцию

ссмотрим сначала следствие, указывающее, что ex post реальная прс нтная ставка не имеет сериальной корреляции, которое Фама тестиру , используя результат из утверждения 2.9. Мы используем рассмот энный в [Mishkin, 1992] набор месячных данных по одномесячным азначейским векселям и месячному темпу инфляции ИПЦ, указанно- іу в процентах годового темпа[1]. Рис. 2.3 изображает данные. Ex post)еальная процентная ставка, определенная как разность между став- кой по казначейским векселям и темпом инфляции ИПЦ, изображена на рис. 2.4.

Рис. 2.3. Инфляция и процентная ставка

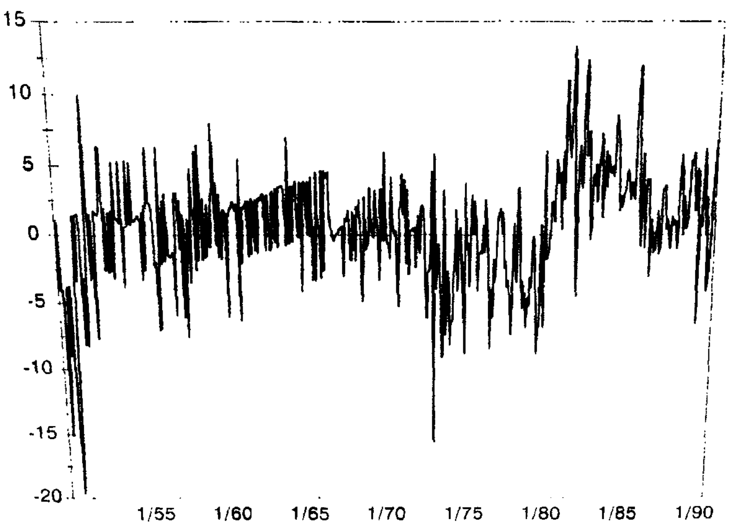

Рис. 2.4. Реальная процентная ставка

[1]Этот набор данных описан подробнее в эмпирическом упражнении. Следуя Фама, мера инфляции, использованная в тексте, вычислена на основе строки чисел CPI. Мера инфляции, рассмотренная в [Mishkin, 1992], использует компоненту CPI, относящуюся к строительству домов. См. вопрос (k) эмпирического упражнения 1.

Для повторения результатов Фама мы берем период выборки тем же, что и в [Fama, 1975], с января 1953 по июль 1971 г. Объем выборки, таким образом, равен 223. Свойства временного ряда реальной процентной ставки суммированы в табл. 2.1.

Таблица 2.1

Реальные процентные ставки, январь 1953 г. – июль 1971 г.

	среднее = 0,82%, стандартное отклонение = 2,837%, объем выборки = 223											
	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$j=7$	$j=8$	$j=9$	$j=10$	$j=11$	$j=12$
$\widehat{\rho}_j$	−0,101	0,172	−0,019	−0,004	−0,064	−0,021	−0,092	0,095	0,094	0,019	0,004	0,207
стд. ошибка	0,067	0,067	0,067	0,067	0,067	0,067	0,067	0,067	0,067	0,067	0,067	0,067
Q Льюнга Бокса	2,3	9,1	9,1	9,1	10,1	10,2	12,1	14,2	16,3	16,4	16,4	26,5
p-значение, %	12,8	1,1	2,8	5,8	7,3	11,7	9,6	7,6	6,1	8,9	12,8	0,9

При вычислении выборочных автокорреляций $\widehat{\rho}_j = \widehat{\gamma}_j/\widehat{\gamma}_0$ мы используем формулу (2.10.1) для $\widehat{\gamma}_j$, где знаменатель равен объему выборки, а не $n - j$. Аналогично в качестве формулы для стандартной ошибки $\widehat{\rho}_j$ берется $1/\sqrt{n}$, а не $1/\sqrt{n-j}$ (так что такая стандартная ошибка не зависит от j). Q-статистика Льюнга — Бокса в таблице для, скажем, $j = 2$ есть (2.10.5) с $p = 2$. Результаты в таблице показывают, что ни одна из автокорреляций не является сильно значимой индивидуально или как группа, в соответствии с первым следствием гипотезы эффективного рынка.

Является ли номинальная процентная ставка оптимальным предиктором?

Является ли номинальная процентная ставка оптимальным предиктором?

Мы можем тестировать следствие 2, оценивая соответствующую регрессию и проводя t-тест утверждения 2.3. Примечателен тот факт о гипотезе эффективности рынка, что, несмотря на свою простоту, она специфична, так что из нее вытекают важные части предположений, обосновывающих t-тест[1]. Чтобы увидеть это, положим $y_t \equiv \pi_{t+1}$, $\boldsymbol{x}_t \equiv (1, R_t)'$, $\varepsilon_t \equiv \eta_{t+1}$ (ошибка прогноза инфляции) и перепишем (2.11.2) как

$$\pi_{t+1} = \beta_1 + \beta_2 R_t + \eta_{t+1} \quad \text{или} \quad y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + \varepsilon_t, \tag{2.11.4}$$

так что предположение 2.1 очевидно удовлетворяется с

$$\boldsymbol{\beta} = (-r, 1)'. \tag{2.11.5}$$

Другие предположения модели проверяются следующим образом:

- *Предположение 2.3 (предопределенные регрессоры)*
  $\boldsymbol{g}_t (\equiv \boldsymbol{x}_t \cdot \varepsilon_t)$ в текущем контексте есть $(\eta_{t+1}, R_t\eta_{t+1})'$. Что нам нужно показать, так это то, что $\mathrm{E}(\eta_{t+1}) = 0$ и $\mathrm{E}(R_t\eta_{t+1}) = 0$. Первое вытекает из (а) и закона полных математических ожиданий:

  $$\mathrm{E}(\eta_{t+1}) = \mathrm{E}[\mathrm{E}(\eta_{t+1}|I_t)] = 0. \tag{2.11.6}$$

  Второе выполняется потому, что

  $\mathrm{E}(R_t\eta_{t+1}) =$
  $= \mathrm{E}[\mathrm{E}(R_t\eta_{t+1}|I_t)]$
  (по закону полных математических ожиданий)

[1]Техническая часть — то, что матрица четвертых моментов $\mathrm{E}(\boldsymbol{g}_t\boldsymbol{g}_t')$ существует и конечна и что регрессоры имеют конечные четвертые моменты, не предполагается гипотезой эффективности рынка и должно быть предположено.

$$= \mathrm{E}[R_t \, \mathrm{E}(\eta_{t+1}|I_t)]$$

(из-за линейности условных математических ожиданий и $R_t \in I_t$)

$$= 0 \quad (\text{по (a)}). \tag{2.11.7}$$

- *Предположение 2.5.* Из пункта (b) следует, что I_t включает в себя $(\varepsilon_{t-1}(= \eta_t), \varepsilon_{t-2}, \ldots)$, так же как и $(\boldsymbol{x}_t, \boldsymbol{x}_{t-1}, \ldots)$. (a) тогда влечет (2.3.1), что является достаточным условием для того, чтобы $\{\boldsymbol{g}_t\}$ была m.d.s.

- *Предположение 2.4.* Матрица $\mathrm{E}(\boldsymbol{x}_t \boldsymbol{x}_t')$ здесь имеет вид:

$$\mathrm{E}(\boldsymbol{x}_t \boldsymbol{x}_t') = \begin{bmatrix} 1 & \mathrm{E}(R_t) \\ \mathrm{E}(R_t) & \mathrm{E}(R_t^2) \end{bmatrix}. \tag{2.11.8}$$

Ее определитель равен $\mathrm{E}(R_t^2) - [\mathrm{E}(R_t)]^2 = \mathrm{Var}(R_t)$ и должен быть положительным; если бы дисперсия была нулевой, мы бы не наблюдали флуктуации процентной ставки R_t.

- *Предположение 2.2 (эргодическая стационарность).* Оно требует, чтобы $\{\pi_{t+1}, R_t\}$ был эргодически стационарным. График на рис. 2.3 показывает участок восходящего движения, особенно для номинальной процентной ставки, хотя за этим участком следует возвращение на более низкий уровень. Тестирование стационарности ряда будет рассмотрено в примере 9.2 главы 9. В настоящее время, несмотря на это достаточно случайное свидетельство об обратном, мы исходим из предположения стационарности (но последствие наличия тренда в ряде будет затронуто в заключительном подпараграфе). Что касается эргодичности, мы просто должны предположить ее.

Что мы показали, так это то, что при выполнении гипотезы эффективности рынка $\{y_t, \boldsymbol{x}_t\}$ (где $y_t = \pi_{t+1}$ и $\boldsymbol{x}_t = (1, R_t)'$) принадлежит множеству DGP, удовлетворяющих предположениям 2.1–2.5.

Проверив требуемые предположения, мы теперь перейдем к оцениванию. Используя наши данные, получаем оцененную регрессию: для $t = 1, \ldots, 223$ ($t = 1$ соответствует январю 1953 г.)

$$\begin{gathered} \pi_{t+1} = \underset{(0{,}431)}{-0.868} \quad \underset{(0{,}112)}{1.015} \; R_t, \\ R^2 = 0.24, \text{ среднее зависимой переменной} = 2{,}35, \\ SER = 2{,}84\%, \; n = 223. \end{gathered} \tag{2.11.9}$$

Робастные к гетероскедастичности стандартные ошибки, вычисленные как в (2.4.1) (где $\widehat{\boldsymbol{S}}$ задается (2.5.1) с $\widehat{\varepsilon}_i = e_i$), указаны в скобках. Как показано в (2.11.5), эффективность рынка влечет равенство 1 коэффициента при R_t[1]. Робастное t-отношение для нулевой гипотезы о том, что коэффициент при R_t равен 1, равно $(1{,}015 - 1)/0{,}112 = 0{,}13$. Соответствующее ему p-значение равно примерно 90%. Таким образом, мы можем легко принять нулевую гипотезу.

Существуют и другие способы тестировать следствие о том, что номинальная процентная ставка объединяет всю текущую доступную информацию. Мы можем привлечь в регрессию больше объясняющих переменных и посмотреть, имеют ли они дополнительную объяснительную мощность сверх номинальной ставки. Однако дополнительные переменные, которые будут введены, должны быть частью текущего информационного множества I_t, поскольку очевидно, что гипотеза эффективности рынка согласуется с существованием некоторой переменной, *в настоящее время не доступной*, предсказывающей будущую инфляцию лучше, чем текущая номинальная ставка. Формально, если $\boldsymbol{x}_t$ включает переменные, не включенные в I_t, доказательство, которое мы использовали для обоснования предположения 2.3, больше недействительно.

R_t не является строго экзогенной

Как было подчеркнуто, одно из важных преимуществ теории больших выборок над **теорией конечных выборок** (*finite-sample theory*) из главы 1 заключается в том, что регрессоры необязательно должны быть строго экзогенными, пока они ортогональны компоненте ошибок. Мы сейчас покажем на примере, что R_t не является строго экзогенной, так что теория конечных выборок неприменима к регрессии Фамы (2.11.4). Предположим, что темп инфляции описывается AR(1)-процессом:

$$\pi_t = c + \rho\pi_{t-1} + \eta_t, \quad \{\eta_t\} \text{ — независимый белый шум.}$$

Если η_{t+1} не зависит от любого элемента I_t, то

$$\begin{aligned}
&\mathrm{E}(\pi_{t+1}|I_t) = \\
&= c + \rho\pi_t + \mathrm{E}(\eta_{t+1}|I_t) \quad \text{(поскольку } \pi_t \text{ содержится в } I_t\text{)} \\
&= c + \rho\pi_t \\
&\text{(поскольку } \mathrm{E}(\pi_{t+1}|I_t) = \mathrm{E}(\pi_{t+1}) \text{ и } \mathrm{E}(\pi_{t+1}) = 0 \text{ по предположению).}
\end{aligned}$$

Поэтому η_{t+1} действительно является ошибкой прогноза инфляции для π_{t+1}, и

$$R_t = r + c + \rho\pi_t$$

(поскольку $R_t = r + \mathrm{E}(\pi_{t+1}|I_t)$) при эффективности рынка). Тогда легко видеть, что компонента ошибок ε_t в регрессии Фамы (π_{t+1} на константу

[1] Известно, что константа равна $-r$, но это не производит тестируемого ограничения, поскольку гипотеза эффективности рынка не специфицирует значение r.

и R_t), которая есть ошибка прогноза инфляции η_{t+1}, может быть коррелирована с будущими регрессорами. Например,

$$\begin{aligned}
&\mathrm{Cov}(\varepsilon_t, R_{t+1}) = \\
&= \mathrm{Cov}(\eta_{t+1}, r + c + \rho\pi_{t+1}) = \\
&= \rho\,\mathrm{Cov}(\eta_{t+1}, \pi_{t+1}) = \\
&= \rho\,\mathrm{Var}(\eta_{t+1}) \quad (\text{поскольку } \pi_{t+1} = c + \rho\pi_t + \eta_{t+1} \text{ и } \mathrm{Cov}(\eta_{t+1}, \pi_t) = 0),
\end{aligned}$$

что не является нулем.

Хотя теория конечных выборок и неприменима, ее рецепты статистических выводов близки к правильным при условии, что ошибки условно гомоскедастичны и размер выборки достаточно велик, о чем говорит параграф 2.6.

Последующее развитие

Хотя работа Фамы является, пожалуй, лучшим примером эконометрики рациональных ожиданий, события после ее публикации и большое количество эмпирических исследований, на которые он вдохновил, доказали, что близкое к взаимно однозначному соотношение между инфляцией и процентными ставками ограничивается послевоенным периодом до 1979 года. Такая тесная связь не может быть найдена для предвоенного периода или для многих других стран. Кроме того, увеличение реальных процентных ставок после смены операционной процедуры ФРС, которое состоялось в октябре 1979 года, противоречит предпосылке регрессии Фамы, что ожидаемая реальная процентная ставка не меняется с течением времени. Когда выборка ограничена периодом после октября 1979 года, коэффициент при процентной ставке в регрессии Фамы намного ниже единицы[1].

Эти результаты ставят вопрос о том, почему сильная связь происходит только в определенные периоды времени, но не в другие. Объяснение Мишкина [Mishkin, 1992] заключается в том, что инфляционная премия включается в процентные ставки постепенно. В периоды, когда темп инфляции показывает только краткосрочные колебания, процентные ставки не реагируют на инфляцию. Однако устойчивое движение инфляции отражается на процентных ставках. Период, когда наблюдается сильная связь (которая включает в себя период выборки Фамы, см. рис. 2.3) является именно тем, на котором инфляция показывала устойчивое восходящее движение.

29 января 1997 года Министерство финансов США продало с аукциона 7 миллиардов долларов десятилетних индексированных на инфляцию облигаций, дав возможность исследователям наблюдать ex ante

[1]См., например: [Mishkin, 1992, Table 1]. Мы обоснуем это в части (1) эмпирического упражнения 1.

реальную процентную ставку как доходность по индексированным облигациям. Данные по Великобритании, где индексированные облигации были доступны с начала 1980-х, показывают, что доходности по индексированным облигациям хотя и гораздо менее волатильны, чем доходности по обычным облигациям, но не являются постоянными во времени. Предположение о постоянной реальной ставке, которое мы сделали, было вспомогательным предположением, посредством которого мы можем проверить, отражают ли цены на облигации в полной мере доступную информацию. Теперь мы можем сделать это без вспомогательного предположения, регрессируя фактический темп инфляции на разность доходностей между обычными и индексированными облигациями. Мы также можем ослабить предположение, неявно делавшееся до сих пор, о том, что разность доходностей равна ожидаемому темпу инфляции. Инвесторы могут не быть нейтральными к риску, и инфляционная премия за риск сверх ожидаемого уровня темпа инфляции может быть необходима для того, чтобы побудить их держать обычные облигации. Если это так, то ожидаемый темп инфляции снова становится ненаблюдаемым. Эконометрика рациональных ожиданий будет оставаться полезным инструментом для работы с ненаблюдаемыми ожиданиями.

Контрольные вопросы

1. Пусть $R_t = 5\%$ и $\pi_{t+1} = 2\%$. Вычислите r_{t+1}, используя две формулы, $(1+R_t)/(1+\pi_{t+1})-1$ и $R_t-\pi_{t+1}$. Мало ли отличие? Что будет, если $R_t = 20\%$ и $\pi_{t+1} = 17\%$?

2. Покажите, что при гипотезе эффективности рынка $\widehat{\mathrm{E}}^*(\pi_{t+1}|1, R_t) = -r + R_t$. **Указание:** *Используйте* (2.9.7).

3. Если бы ошибка прогноза инфляции η_{t+1} наблюдалась, как бы вы тестировали эффективность рынка? Нужно ли нам предположение о постоянной ex ante реальной ставке?

4. Если темп инфляции и процентные ставки измерять в долях, а не в процентах (то есть 0,08 вместо 8%), то как изменится регрессионный результат (2.11.9)? А если темп инфляции измерять в процентах, но за месяц, а процентную ставку в процентах за год? **Указание:** *Если x — темп инфляции за месяц, а y — ее годовой темп, то $1 + y = (1 + x)^{12}$, так что $y \approx 12x$.*

5. Предположим в (2.11.4), что x_t включает третью переменную, которая не входит в I_t. Какая часть аргументации, использованной при выводе предположения 2.3, терпит неудачу? **Указание:** *Третий элемент g_t является третьей переменной, умноженной на η_{t+1}.*

6. Предложите пример *не*эффективности рынка такой, что следствие 1 выполняется, но следствие 2 — нет. **Указание:** *Предположите, что π_t и R_t сериально независимы и являются взаимно независимыми процессами.*

2.12. Регрессия на время

На протяжении всей этой главы мы предполагали, что $\{y_t, \boldsymbol{x}_t\}$ является стационарным. Это предположение, однако, не всегда выполняется в регрессиях временных рядов. В этой книге мы изучаем два случая, в которых регрессоры не являются стационарными и все же применим OLS. Первый рассматривается здесь, в то время как другой случай, называемый «коинтегрированными регрессиями», рассматривается в главе 10.

Регрессия, которую мы рассмотрим, записывается как

$$y_t = \alpha + \delta \cdot t + \varepsilon_t, \tag{2.12.1}$$

где $\{\varepsilon_t\}$ — независимый белый шум. Эта регрессия может быть записана как $y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + \varepsilon_t$ с

$$\boldsymbol{x}_t = (1, t)', \ \boldsymbol{\beta} = (\alpha, \delta)'. \tag{2.12.2}$$

Ясно, что $\boldsymbol{x}_t$ не является стационарным, поскольку математическое ожидание второго элемента равно t, и оно возрастает с течением времени. Аналогично нестационарным является y_t. Линейная функция $\alpha + \delta \cdot t$ называется **временным трендом** (*time trend*) y_t. Мы говорим, что процесс **стационарный относительно тренда** (*trend stationary*), если он может быть записан как сумма временного тренда и стационарного процесса. Процесс $\{y_t\}$ здесь является частным случаем процесса, стационарного относительно тренда, стационарная компонента которого является независимым белым шумом.

Асимптотическое распределение OLS-оценки

Пусть $\boldsymbol{b}$ является OLS-оценкой для $\boldsymbol{\beta}$, основанной на выборке размера n:

$$\boldsymbol{b} \equiv \begin{bmatrix} \widehat{\alpha} \\ \widehat{\delta} \end{bmatrix} = \Big(\sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{x}_t'\Big)^{-1} \Big(\sum_{t=1}^{n} \boldsymbol{x}_t \cdot y_t\Big). \tag{2.12.3}$$

Ошибка этой оценки может быть записана как

$$\boldsymbol{b} - \boldsymbol{\beta} = \begin{bmatrix} \widehat{\alpha} - \alpha \\ \widehat{\delta} - \delta \end{bmatrix} = \Big(\sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{x}_t'\Big)^{-1} \Big(\sum_{t=1}^{n} \boldsymbol{x}_t \cdot \varepsilon_t\Big). \tag{2.12.4}$$

Используя алгебраические результаты $\sum_{t=1}^{n} t = n \cdot (n+1)/2$ и $\sum_{t=1}^{n} t^2 = n \cdot (n+1)(2n+1)/6$, легко показать, что

$$\sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{x}_t' = \begin{bmatrix} n & n \cdot (n+1)/2 \\ n \cdot (n+1)/2 & n \cdot (n+1)(2n+1)/6 \end{bmatrix}. \tag{2.12.5}$$

Поэтому, в отличие от стационарного случая, $\sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{x}_t'/n$ не сходится по вероятности к невырожденной матрице; она фактически расходится.

Оказывается, что OLS-оценки $\widehat{\alpha}$ и $\widehat{\delta}$ являются состоятельными, но имеют различные скорости сходимости. Как и в стационарном случае, скорость сходимости для $\widehat{\alpha}$ есть $\sqrt{n}$. В противоположность этому скорость для $\widehat{\delta}$ есть $n^{3/2}$. То есть $n^{3/2}(\widehat{\delta}-\delta)$ сходится к невырожденному распределению (распределению непостоянной случайной величины). Чтобы присвоить эти различные скорости сходимости элементам $\boldsymbol{b}-\boldsymbol{\beta}$, умножим обе части (2.12.4) на матрицу:

$$\boldsymbol{\Upsilon}_n = \begin{bmatrix} \sqrt{n} & 0 \\ 0 & n^{3/2} \end{bmatrix}. \tag{2.12.6}$$

Это приводит к

$$\begin{aligned} \boldsymbol{\Upsilon}_n(\boldsymbol{b}-\boldsymbol{\beta}) &= \begin{bmatrix} \sqrt{n}(\widehat{\alpha}-\alpha) \\ n^{3/2}(\widehat{\delta}-\delta) \end{bmatrix} = \boldsymbol{\Upsilon}_n\Big(\sum_{t=1}^{n}\boldsymbol{x}_t\boldsymbol{x}_t'\Big)^{-1}\Big(\sum_{t=1}^{n}\boldsymbol{x}_t\cdot\varepsilon_t\Big) = \\ &= \boldsymbol{\Upsilon}_n\Big(\sum_{t=1}^{n}\boldsymbol{x}_t\boldsymbol{x}_t'\Big)^{-1}\boldsymbol{\Upsilon}_n\boldsymbol{\Upsilon}_n^{-1}\Big(\sum_{t=1}^{n}\boldsymbol{x}_t\cdot\varepsilon_t\Big) = \\ &= \Big[\boldsymbol{\Upsilon}_n^{-1}\Big(\sum_{t=1}^{n}\boldsymbol{x}_t\boldsymbol{x}_t'\Big)\boldsymbol{\Upsilon}_n^{-1}\Big]^{-1}\Big(\boldsymbol{\Upsilon}_n^{-1}\sum_{t=1}^{n}\boldsymbol{x}_t\cdot\varepsilon_t\Big) \equiv \\ &\equiv \boldsymbol{Q}_n^{-1}\boldsymbol{v}_n, \end{aligned} \tag{2.12.7}$$

где

$$\boldsymbol{Q}_n \equiv \boldsymbol{\Upsilon}_n^{-1}\Big(\sum_{t=1}^{n}\boldsymbol{x}_t\boldsymbol{x}_t'\Big)\boldsymbol{\Upsilon}_n^{-1} \quad \text{и} \quad \boldsymbol{v}_n \equiv \boldsymbol{\Upsilon}_n^{-1}\sum_{t=1}^{n}\boldsymbol{x}_t\cdot\varepsilon_t. \tag{2.12.8}$$

Подставляя (2.12.5) и (2.12.6) в эти выражения для $\boldsymbol{Q}_n$ и $\boldsymbol{v}_n$, получаем:

$$\begin{aligned} \boldsymbol{Q}_n &= \begin{bmatrix} 1 & (n+1)/(2n) \\ (n+1)/(2n) & (n+1)(2n+1)/(6n^2) \end{bmatrix}, \\ \boldsymbol{v}_n &= \begin{bmatrix} \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varepsilon_t \\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(t/n)\varepsilon_t \end{bmatrix}. \end{aligned} \tag{2.12.9}$$

Ясно, что мы имеем:

$$\boldsymbol{Q}_n \to \boldsymbol{Q} \equiv \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix}. \tag{2.12.10}$$

Для $\boldsymbol{v}_n$ можно показать (см., например: [Hamilton, 1994, pp. 458–460]), что

$$\boldsymbol{v}_n \underset{\mathrm{d}}{\to} N(\boldsymbol{0}, \sigma^2\boldsymbol{Q}), \tag{2.12.11}$$

если $\{\varepsilon_t\}$ является независимым белым шумом с $\mathrm{E}(\varepsilon_t^2)=\sigma^2$ и $\mathrm{E}(\varepsilon_t^4)<\infty$. Таким образом, асимптотическое распределение (2.12.7) является нормальным со средним $\boldsymbol{0}$ и дисперсией $\boldsymbol{Q}^{-1}(\sigma^2\boldsymbol{Q})\cdot\boldsymbol{Q}^{-1}=\sigma^2\boldsymbol{Q}^{-1}$. Подытоживая проведенное обсуждение, получаем

Утверждение 2.11 (OLS-оценивание регрессии на время): *Рассмотрим регрессию на время (2.12.1), где ε_t — независимый белый шум с $\mathrm{E}(\varepsilon_t^2) = \sigma^2$ и $\mathrm{E}(\varepsilon_t^4) < \infty$, и пусть $\widehat{\alpha}$ и $\widehat{\delta}$ являются OLS-оценками для α и δ. Тогда*

$$\begin{bmatrix} \sqrt{n}(\widehat{\alpha} - \alpha) \\ n^{3/2}(\widehat{\delta} - \delta) \end{bmatrix} \underset{d}{\to} N\left(\mathbf{0}, \sigma^2 \cdot \begin{bmatrix} 1 & 1/2 \\ 1/2 & 1/3 \end{bmatrix}^{-1}\right).$$

Как и в стационарном случае, $\widehat{\alpha}$ является $\sqrt{n}$-состоятельной, поскольку $\sqrt{n}(\widehat{\alpha} - \alpha)$ сходится к (нормальной) случайной величине. OLS-оценка коэффициента при времени, $\widehat{\delta}$, также является состоятельной, но скорость сходимости выше: она $n^{3/2}$-состоятельна в том смысле, что $n^{3/2}(\widehat{\delta} - \delta)$ сходится к случайной величине. В этом смысле $\widehat{\delta}$ является **гиперсостоятельной** (*hyperconsistent*)[1].

Тестирование гипотез в регрессиях на время

Таким образом, OLS-коэффициенты в регрессии на время асимптотически нормальны при условии, что ошибка оценки правильно шкалирована. Мы сейчас покажем, что деление ошибки оценки на стандартную ошибку приводит к шкалированию, которое делает результирующее отношение — t-значение — асимптотически нормальным.

Сначала рассмотрим t-значение для нулевой гипотезы $\alpha = \alpha_0$. Замечая, что

$$\text{элемент } (1,1) \text{ матрицы } \left(\sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{x}_t'\right)^{-1} = \begin{bmatrix} 1 & 0 \end{bmatrix} \left(\sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{x}_t'\right)^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \tag{2.12.12}$$

t-значение можно записать как

$$\begin{aligned} t &= \frac{\widehat{\alpha} - \alpha_0}{\sqrt{s^2 \cdot \begin{bmatrix} 1 & 0 \end{bmatrix} \left(\sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{x}_t'\right)^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix}}} = \\ &= \frac{\sqrt{n}(\widehat{\alpha} - \alpha_0)}{\sqrt{s^2 \cdot \begin{bmatrix} \sqrt{n} & 0 \end{bmatrix} \left(\sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{x}_t'\right)^{-1} \begin{bmatrix} \sqrt{n} \\ 0 \end{bmatrix}}} = \\ &= \frac{\sqrt{n}(\widehat{\alpha} - \alpha_0)}{\sqrt{s^2 \cdot \begin{bmatrix} 1 & 0 \end{bmatrix} \boldsymbol{\Upsilon}_n \left(\sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{x}_t'\right)^{-1} \boldsymbol{\Upsilon}_n \begin{bmatrix} 1 \\ 0 \end{bmatrix}}} \end{aligned}$$

$$\left(\text{поскольку } \begin{bmatrix} \sqrt{n} & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \end{bmatrix} \boldsymbol{\Upsilon}_n\right)$$

[1]Оценка, которая n-состоятельна, называется **суперсостоятельной** (superconsistent). Известный пример суперсостоятельных оценок возникает при оценивании процессов с единичным корнем; см. главу 9. Некоторые авторы используют термин «суперсостоятельный» более широко, для обозначения оценок, которые n^γ-состоятельны с $\gamma > \frac{1}{2}$. На их языке $\widehat{\delta}$ в тексте является суперсостоятельной.

$$= \frac{\sqrt{n}(\widehat{\alpha} - \alpha_0)}{\sqrt{s^2 \cdot [1 \quad 0]\left\{\boldsymbol{\Upsilon}_n^{-1}\left(\sum_{t=1}^n \boldsymbol{x}_t\boldsymbol{x}_t'\right)\boldsymbol{\Upsilon}_n^{-1}\right\}^{-1}\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]}} =$$

$$= \frac{\sqrt{n}(\widehat{\alpha} - \alpha_0)}{\sqrt{s^2 \cdot [1 \quad 0]\boldsymbol{Q}_n^{-1}\left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]}} \quad \text{(из (2.12.8)).} \qquad (2.12.13)$$

Можно непосредственно показать (см. контрольный вопрос 1), что s^2 состоятельна для σ^2. И из (2.12.10) $\boldsymbol{Q}_n \to_{\text{p}} \boldsymbol{Q}$. Таким образом, по лемме 2.4(с) и утверждению 2.11,

t-значение для нулевой гипотезы $\alpha = \alpha_0$

$$\underset{\text{d}}{\to} \frac{\text{1-й элемент } \boldsymbol{z}}{\sqrt{\sigma^2 \times \text{элемент } (1,1) \text{ матрицы } \boldsymbol{Q}^{-1}}}, \qquad (2.12.14)$$

где $\boldsymbol{z} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{Q}^{-1})$. Поэтому t-значение для α асимптотически $N(0,1)$, как и в стационарном случае.

Используя тот же самый трюк и замечая, что $[0 \quad n^{3/2}] = [0 \quad 1]\boldsymbol{\Upsilon}_n$, мы можем записать t-значение для нулевой гипотезы $\delta = \delta_0$ как

$$t = \frac{n^{3/2}(\widehat{\delta} - \delta_0)}{\sqrt{s^2 \cdot [0 \quad 1]\boldsymbol{Q}_n^{-1}\left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]}}. \qquad (2.12.15)$$

Поэтому t-значение для коэффициента при времени также является асимптотически $N(0,1)$! Статистические выводы относительно α или δ могут быть получены аналогично стационарному случаю.

Контрольные вопросы

1. (s^2 состоятельна для σ^2.) Покажите, что $s^2 \to_{\text{p}} \sigma^2$. **Указание:** *Исходя из формулы* (2.3.10) *получите*

$$e_t^2 = \varepsilon_t^2 - 2(\boldsymbol{b} - \boldsymbol{\beta})'\boldsymbol{\Upsilon}_n\boldsymbol{\Upsilon}_n^{-1}\boldsymbol{x}_t \cdot \varepsilon_t + (\boldsymbol{b} - \boldsymbol{\beta})'\boldsymbol{\Upsilon}_n\boldsymbol{\Upsilon}_n^{-1}\boldsymbol{x}_t\boldsymbol{x}_t'\boldsymbol{\Upsilon}_n^{-1}\boldsymbol{\Upsilon}_n(\boldsymbol{b} - \boldsymbol{\beta}).$$

Суммируйте по t, чтобы получить

$$\frac{1}{n}\sum_{t=1}^{n} e_t^2 = \frac{1}{n}\sum_{t=1}^{n} \varepsilon_t^2 - \frac{2}{n}(\boldsymbol{b} - \boldsymbol{\beta})'\boldsymbol{\Upsilon}_n\boldsymbol{v}_n + \frac{1}{n}(\boldsymbol{b} - \boldsymbol{\beta})'\boldsymbol{\Upsilon}_n\boldsymbol{Q}\boldsymbol{\Upsilon}\boldsymbol{\Upsilon}_n(\boldsymbol{b} - \boldsymbol{\beta}).$$

2. (Сериально некоррелированные ошибки.) Предположим, что $\{\varepsilon_t\}$ — общий стационарный процесс, а не процесс независимого белого шума, но предположим, что $\boldsymbol{v}_n \to_{\text{d}} N(\boldsymbol{0}, \boldsymbol{V})$, где $\boldsymbol{V} \neq \sigma^2\boldsymbol{Q}$. Являются ли OLS-оценки $\widehat{\alpha}$ и $\widehat{\delta}$ состоятельными? Сходится ли $\boldsymbol{\Upsilon}(\boldsymbol{b} - \boldsymbol{\beta})$ к нормальному случайному вектору с нулевым средним? Если да, то какова дисперсия нормального распределения? [Ответ: $\boldsymbol{Q}^{-1}\boldsymbol{V}\boldsymbol{Q}^{-1}$.] Являются ли t-значения асимптотически нормальными? [Ответ: Да. Но дисперсия не будет единичной.]

Приложение 2.А. Асимптотика с фиксированными регрессорами

Чтобы доказать результаты, аналогичные выводам утверждения 2.5, мы добавляем к предположению 2.1 следующее.

Предположение 2.А.1: *$\boldsymbol{X}$ является детерминированной матрицей. При $n \to \infty$ $\boldsymbol{S_{xx}} \to \boldsymbol{\Sigma_{xx}}$, невырожденной матрице (здесь обычная сходимость, поскольку $\boldsymbol{X}$ не является случайной).*

Предположение 2.А.2: *$\{\varepsilon_i\}$ является i.i.d. с $\mathrm{E}(\varepsilon_i) = 0$ и $\mathrm{Var}(\varepsilon_i) = \sigma^2$.*

Утверждение 2.А.1: *При предположениях 2.1, 2.А.1 и 2.А.2 плюс предположении на $\{\boldsymbol{g}_i\}$, называемом* ***условием Линдеберга*** *(где $\boldsymbol{g}_i \equiv \boldsymbol{x}_i \cdot \varepsilon_i$), сохраняются те же самые выводы, что и в утверждении 2.5.*

Вместо того чтобы сформулировать условие Линдеберга, мы объясняем, почему такое условие необходимо. Для доказательства асимптотической нормальности $\boldsymbol{b}$ мы могли бы взять (2.3.8) и показать, что $\sqrt{n}\overline{\boldsymbol{g}}$ сходится к нормальному распределению. Техническая сложность заключается в том, что последовательность $\{\boldsymbol{g}_i\}$ не является стационарной из-за предположения о том, что $\{\boldsymbol{x}_i\}$ является последовательностью неслучайных векторов. Например,

$$\mathrm{Var}(\boldsymbol{g}_i) = \mathrm{E}(\varepsilon_i^2 \boldsymbol{x}_i \boldsymbol{x}_i') = \mathrm{E}(\varepsilon_i^2) \boldsymbol{x}_i \boldsymbol{x}_i' = \sigma^2 \boldsymbol{x}_i \boldsymbol{x}_i'$$

не остается постоянной при переходе от наблюдения к наблюдению. Следовательно, ни CLT Линдеберга — Леви, ни CLT для мартингал-разностей неприменимы. Однако, по счастью, существует обобщение CLT Линдеберга — Леви на нестационарные процессы с непостоянной дисперсией, так называемая **CLT Линдеберга — Феллера** (*Lindeberg — Feller CLT*), которая устанавливает техническое ограничение на хвост распределения $\boldsymbol{g}_i$, чтобы наблюдения с большими $\boldsymbol{x}_i$ не могли доминировать в распределении $\sqrt{n}\overline{\boldsymbol{g}}$.

Приложение 2.В. Доказательство утверждения 2.10

В этом приложении приводится доказательство утверждения 2.10. Мы начинаем с выражения для $\widehat{\gamma}_j$, данного в (2.10.12). Первая часть доказательства заключается в нахождении выражения, которое является асимптотически эквивалентным.

$$\sqrt{n}\widehat{\gamma}_j = \sqrt{n}\widetilde{\gamma}_j - \frac{1}{n}\sum_{t=j+1}^{n}(\boldsymbol{x}_{t-j}\cdot\varepsilon_t + \boldsymbol{x}_t\cdot\varepsilon_{t-j})'\sqrt{n}(\boldsymbol{b}-\boldsymbol{\beta}) +$$

$$+\sqrt{n}(\boldsymbol{b}-\boldsymbol{\beta})'\left(\frac{1}{n}\sum_{t=j+1}^{n}\boldsymbol{x}_t\boldsymbol{x}'_{t-j}\right)(\boldsymbol{b}-\boldsymbol{\beta}) \underset{\mathrm{a}}{\sim}$$

$$\underset{\mathrm{a}}{\sim} \sqrt{n}\widetilde{\gamma}_j - \frac{1}{n}\sum_{t=j+1}^{n}(\boldsymbol{x}_{t-j}\cdot\varepsilon_t + \boldsymbol{x}_t\cdot\varepsilon_{t-j})'\sqrt{n}(\boldsymbol{b}-\boldsymbol{\beta})$$

(поскольку последнее слагаемое выше исчезает)

$$\underset{\mathrm{a}}{\sim} \sqrt{n}\widetilde{\gamma}_j - \mathrm{E}(\boldsymbol{x}_{t-j}\cdot\varepsilon_t + \boldsymbol{x}_t\cdot\varepsilon_{t-j})'\sqrt{n}(\boldsymbol{b}-\boldsymbol{\beta})$$

(поскольку $\frac{1}{n}\sum_{t=j+1}^{n}(\boldsymbol{x}_{t-j}\cdot\varepsilon_t + \boldsymbol{x}_t\cdot\varepsilon_{t-j}) \underset{\mathrm{p}}{\to} \mathrm{E}(\boldsymbol{x}_{t-j}\cdot\varepsilon_t + \boldsymbol{x}_t\cdot\varepsilon_{t-j})$)

$$= \sqrt{n}\widetilde{\gamma}_j - \boldsymbol{\mu}'_j\sqrt{n}(\boldsymbol{b}-\boldsymbol{\beta}) =$$

где $\boldsymbol{\mu}_j \equiv \mathrm{E}(\boldsymbol{x}_t\cdot\varepsilon_{t-j})$ (поскольку $\mathrm{E}(\boldsymbol{x}_{t-j}\cdot\varepsilon_t) = \boldsymbol{0}$ по (2.10.15))

$$= \sqrt{n}\widetilde{\gamma}_j - \boldsymbol{\mu}'_j\sqrt{n}\boldsymbol{S}_{\boldsymbol{xx}}^{-1}\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{x}_t\cdot\varepsilon_t =$$

$$= \sqrt{n}\frac{1}{n}\sum_{t=j+1}^{n}\varepsilon_t\varepsilon_{t-j} - \boldsymbol{\mu}'_j\sqrt{n}\boldsymbol{S}_{\boldsymbol{xx}}^{-1}\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{x}_t\cdot\varepsilon_t \quad \text{по (2.10.7)}$$

$$\underset{\mathrm{a}}{\sim} \sqrt{n}\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t\varepsilon_{t-j} - \boldsymbol{\mu}'_j\sqrt{n}\boldsymbol{S}_{\boldsymbol{xx}}^{-1}\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{x}_t\cdot\varepsilon_t$$

(поскольку $\sqrt{n}\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t\varepsilon_{t-j} - \sqrt{n}\frac{1}{n}\sum_{t=j+1}^{n}\varepsilon_t\varepsilon_{t-j}$

$$= \frac{1}{\sqrt{n}}\sum_{t=1}^{j}\varepsilon_t\varepsilon_{t-j} \underset{\mathrm{p}}{\to} 0 \text{ для каждого } j)$$

$$\underset{\mathrm{a}}{\sim} \sqrt{n}\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t\varepsilon_{t-j} - \boldsymbol{\mu}'_j\sqrt{n}\boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1}\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{x}_t\cdot\varepsilon_t$$

(поскольку $\boldsymbol{S}_{\boldsymbol{xx}} \underset{\mathrm{p}}{\to} \boldsymbol{\Sigma}_{\boldsymbol{xx}}$)

$$= \boldsymbol{c}'_j\sqrt{n}\overline{\boldsymbol{g}}_j, \quad \text{где}$$

$$\underset{((K+1)\times 1)}{\boldsymbol{c}_j} \equiv \begin{bmatrix} 1 \\ -\boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1}\boldsymbol{\mu}_j \end{bmatrix}, \quad \underset{((K+1)\times 1)}{\overline{\boldsymbol{g}}_j} \equiv$$

$$\equiv \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{g}_{jt}, \quad \underset{((K+1)\times 1)}{\boldsymbol{g}_{jt}} \equiv \begin{bmatrix} \varepsilon_{t-j}\varepsilon_t \\ \boldsymbol{x}_t\cdot\varepsilon_t \end{bmatrix}. \qquad (2.\mathrm{B}.1)$$

Таким образом, полагая

$$\underset{(p\times 1)}{\boldsymbol{\gamma}} \equiv \begin{bmatrix} \widehat{\gamma}_1 \\ \vdots \\ \widehat{\gamma}_p \end{bmatrix},$$

мы доказали, что

$$\sqrt{n}\widehat{\boldsymbol{\gamma}} \underset{\mathrm{a}}{\sim} \boldsymbol{C}'\sqrt{n}\overline{\boldsymbol{g}},$$

где

$$\underset{(p(K+1)\times p)}{\boldsymbol{C}} \equiv \begin{bmatrix} \boldsymbol{c}_1 & & \\ & \ddots & \\ & & \boldsymbol{c}_p \end{bmatrix},$$

$$\underset{(p(K+1)\times 1)}{\overline{\boldsymbol{g}}} \equiv \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{g}_t,$$

$$\underset{(p(K+1)\times 1)}{\boldsymbol{g}_t} \equiv \begin{bmatrix} \boldsymbol{g}_{1t} \\ \vdots \\ \boldsymbol{g}_{pt} \end{bmatrix}.$$

Вторая часть доказательства состоит в том, чтобы показать, что $\boldsymbol{g}_t$ является мартингал-разностью, а именно что

$$\mathrm{E}(\boldsymbol{g}_{jt}|\boldsymbol{g}_{t-1}, \boldsymbol{g}_{t-2}, \dots) = \mathbf{0} \quad (\text{для } j = 1, 2, \dots, p).$$

Но это легко следует из закона повторных математических ожиданий, того факта, что $(\boldsymbol{g}_{t-1}, \boldsymbol{g}_{t-2}, \dots)$ содержит меньше информации, чем $(\boldsymbol{x}_t, \boldsymbol{x}_{t-1}, \dots, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots)$, и (2.10.15).

Ясно, что $\boldsymbol{g}_t$ эргодически стационарна. Следовательно, по CLT для эргодически стационарных мартингал-разностей, $\sqrt{n}\overline{\boldsymbol{g}} \to_{\mathrm{d}} N(\mathbf{0}, \mathrm{E}(\boldsymbol{g}_t\boldsymbol{g}_t'))$. Следующий шаг состоит в вычислении математического ожидания $\mathrm{E}(\boldsymbol{g}_t\boldsymbol{g}_t')$, (j, k)-й блок которого имеет вид:

$$\underset{(((K+1)\times(K+1)))}{\mathrm{E}(\boldsymbol{g}_{jt}\boldsymbol{g}_{kt}')} = \begin{bmatrix} \mathrm{E}(\varepsilon_t^2\varepsilon_{t-j}\varepsilon_{t-k}) & \mathrm{E}(\boldsymbol{x}_t'\varepsilon_t^2\varepsilon_{t-j}) \\ \mathrm{E}(\boldsymbol{x}_t\varepsilon_t^2\varepsilon_{t-k}) & \mathrm{E}(\varepsilon_t^2\boldsymbol{x}_t\boldsymbol{x}_t') \end{bmatrix}.$$

Используя закон полных математических ожиданий и (2.10.16), легко показать, что

$$\underset{(((K+1)\times(K+1)))}{\mathrm{E}(\boldsymbol{g}_{jt}\boldsymbol{g}_{kt}')} = \begin{bmatrix} \sigma^4\delta_{jk} & \sigma^2\boldsymbol{\mu}_j' \\ \sigma^2\boldsymbol{\mu}_k & \sigma^2\boldsymbol{\Sigma}_{\boldsymbol{xx}} \end{bmatrix}, \tag{2.B.2}$$

где δ_{jk} — «дельта-символ Кронекера», принимающий значение 1, если $j = k$, и 0 в противном случае.

Поскольку, как показано выше,

$$\sqrt{n}\widehat{\gamma} \underset{\mathrm{a}}{\sim} \boldsymbol{C}'\sqrt{n}\overline{\boldsymbol{g}} \text{ и } \sqrt{n}\overline{\boldsymbol{g}} \underset{\mathrm{d}}{\to} N(\mathbf{0}, \mathrm{E}(\boldsymbol{g}_t\boldsymbol{g}_t')),$$

мы имеем:

$$\mathrm{Avar}(\widehat{\gamma})(\equiv \text{дисперсия предельного распределения } \sqrt{n}\widehat{\gamma}) = \boldsymbol{C}'\,\mathrm{E}(\boldsymbol{g}_t\boldsymbol{g}_t')\boldsymbol{C}.$$

Ее (j,k)-й элемент есть

$$(j,k)\text{-й элемент } \mathrm{Avar}(\widehat{\gamma}) = \boldsymbol{c}_j'\,\mathrm{E}(\boldsymbol{g}_{jt}\boldsymbol{g}_{kt}')\boldsymbol{c}_k.$$

Подставляя сюда (2.B.2) и используя определение $\boldsymbol{c}_j$ в (2.B.1), после некоторых простых вычислений получаем, что

$$(j,k)\text{-й элемент } \mathrm{Avar}(\widehat{\gamma}) = \sigma^4 \cdot [\delta_{jk} - \boldsymbol{\mu}_j'\boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1}\boldsymbol{\mu}_k/\sigma^2].$$

Поэтому

$$\sqrt{n}\widehat{\gamma} \underset{\mathrm{d}}{\to} N(\mathbf{0}, \sigma^4\boldsymbol{I}_p - \sigma^4\boldsymbol{\Phi}),$$

где $\boldsymbol{\Phi}$ определена в утверждении 2.10. Поскольку $\sqrt{n}\widehat{\boldsymbol{\rho}} \underset{\mathrm{a}}{\sim} n\widehat{\gamma}/\sigma^2$, предельное распределение для $\sqrt{n}\widehat{\boldsymbol{\rho}}$ то же самое, что и предельное распределение для $\sqrt{n}\widehat{\gamma}/\sigma^2$, которое является $N(\mathbf{0}, \boldsymbol{I}_p - \boldsymbol{\Phi})$. ■

Набор задач для главы 2

Аналитические упражнения

1. (Взято из примера 3.4.2 в [Amemiya, 1985].) Пусть z_n определяется как

$$z_n = \begin{cases} 0 & \text{с вероятностью } (n-1)/n, \\ n^2 & \text{с вероятностью } 1/n. \end{cases}$$

Покажите, что $\mathrm{plim}_{n\to\infty} z_n = 0$, но $\lim_{n\to\infty} \mathrm{E}(z_n) = \infty$.

2. Докажите слабый LLN Чебышева, показывая, что $\bar{z}_n \to_{\mathrm{m.s.}} \mu$. **Указание:**

$$\bar{z}_n - \mu = (\bar{z}_n - \mathrm{E}(\bar{z}_n)) + (\mathrm{E}(\bar{z}_n) - \mu).$$

Поэтому

$$(\bar{z}_n - \mu)^2 = (\bar{z}_n - \mathrm{E}(\bar{z}_n))^2 + 2(\bar{z}_n - \mathrm{E}(\bar{z}_n))(\mathrm{E}(\bar{z}_n) - \mu) + (\mathrm{E}(\bar{z}_n) - \mu)^2.$$

Покажите, что $\mathrm{E}[(\bar{z}_n - \mathrm{E}(\bar{z}_n))(\mathrm{E}(\bar{z}_n) - \mu)] = 0$.

3. (Состоятельность и асимптотическая нормальность OLS для случайных выборок.) Рассмотрите замену предположения 2.2 на

Предположение 2.2′: $\{y_i, \boldsymbol{x}_i\}$ *является случайной выборкой.*

и предположения 2.5 на

Предположение 2.5′: $\boldsymbol{S} \equiv \mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_i')$ *существует и конечно.*

Докажите следующую упрощенную версию утверждения 2.1:

Утверждение 2.1 для случайных выборок: *При предположениях 2.1, 2.3, 2.4 и 2.2′ OLS-оценка* $\boldsymbol{b}$ *является состоятельной. При дополнительном предположении 2.5′ OLS-оценка* $\boldsymbol{b}$ *является асимптотически нормальной с* $\mathrm{Avar}(\boldsymbol{b})$, *заданной в* (b).

(a) Покажите, что это является частным случаем утверждения 2.1 в тексте.

(b) Дайте прямое доказательство этого утверждения.
Указание: *Используйте результаты Колмогорова и Линдеберга — Леви.*

4. (Доказательство утверждения 2.4.) Мы хотим доказать утверждение 2.4. Чтобы избежать несущественных усложнений, мы предположим, как и в тексте, что $K = 1$ (только один регрессор). Что остается показать — это то, что выборочное среднее первой компоненты правой части (2.5.3) сходится по вероятности к конечному числу. Докажите это. **Указание:** *Пусть* $f \equiv x_i\varepsilon_i$ *и* $h \equiv x_i^2$. ***Неравенство Коши — Шварца*** *(Cauchy — Schwartz inequality) утверждает:*

$$\mathrm{E}(|f \cdot h|) \le \sqrt{\mathrm{E}(f^2)\,\mathrm{E}(h^2)}.$$

Поэтому

$$\mathrm{E}(|x_i^3\varepsilon_i|) \le \sqrt{\mathrm{E}(x_i^2\varepsilon_i^2)\,\mathrm{E}(x_i^4)}.$$

По предположению, $\mathrm{E}(x_i^2\varepsilon_i^2)$ *конечно и, по предположению 2.6,* $\mathrm{E}(x_i^4)$ *также конечно. Поэтому* $\mathrm{E}(\varepsilon_i x_i^3)$ $(= \mathrm{E}(g_i^2))$ *конечно, и* $\frac{1}{n}\sum_i \varepsilon_i x_i^3 \to_{\mathrm{p}}$ *к некоторому конечному числу в силу эргодической стационарности.*

5. (Прямое доказательство того, что изменение в SSR, деленное на s^2, имеет асимптотическое распределение χ^2.) В параграфе 2.6 мы доказали, что

$$(SSR_R - SSR_U)/s^2 \underset{\mathrm{d}}{\to} \chi^2(\#\boldsymbol{r})$$

как следствие утверждения 2.3. Дайте прямое доказательство, сначала показывая, что

$$SSR_R - SSR_U = (\sqrt{n}\bar{\boldsymbol{g}})'\boldsymbol{S}_{\boldsymbol{xx}}^{-1}\boldsymbol{R}'(\boldsymbol{R}\boldsymbol{S}_{\boldsymbol{xx}}^{-1}\boldsymbol{R}')^{-1}\boldsymbol{R}\boldsymbol{S}_{\boldsymbol{xx}}^{-1}(\sqrt{n}\bar{\boldsymbol{g}}),$$

$$\bar{g} \equiv \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \cdot \varepsilon_i.$$

6. (Дополнительно, состоятельность $\widehat{\alpha}$.) В этом упражнении мы докажем утверждение, сделанное в параграфе 2.8, о том, что $\widehat{\alpha}$, OLS-оценка для регрессии e_i^2 на z_i, состоятельна. Пусть $\widetilde{\alpha}$ будет OLS-оценкой из

$$\varepsilon_i^2 = \boldsymbol{z}_i'\boldsymbol{\alpha} + \eta_i, \quad \eta_i \equiv \varepsilon_i^2 - \mathrm{E}(\varepsilon_i^2|\boldsymbol{x}_i).$$

   (a) Сделайте соответствующие предположения в дополнение к предположениям 2.1 и 2.2, чтобы показать, что $\widetilde{\alpha}$ является состоятельной. **Указание:** *Вам нужно ранговое условие для $\boldsymbol{z}_i$.*

   (b) Получите:

$$\widehat{\boldsymbol{\alpha}} - \widetilde{\boldsymbol{\alpha}} = \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_i\boldsymbol{z}_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_i \cdot v_i \qquad (*)$$

   с $v_i \equiv -2(\boldsymbol{b} - \boldsymbol{\beta})'\boldsymbol{x}_i \cdot \varepsilon_i + (\boldsymbol{b} - \boldsymbol{\beta})'\boldsymbol{x}_i\boldsymbol{x}_i'(\boldsymbol{b} - \boldsymbol{\beta})$. **Указание:** *Расхождение между ε_i^2 и квадратом OLS-остатка e_i^2 из OLS-оценивания исходного уравнения $y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i$ задано в* (2.3.10)*. Подстановка в* (2.8.8) *дает:* $e_i^2 = \boldsymbol{z}_i'\boldsymbol{\alpha} + (\eta_i + v_i)$.

   (c) Чтобы избежать несущественных усложнений, предположим, что $\boldsymbol{x}_i$ является скаляром x_i. Тогда $\frac{1}{n}\sum_i \boldsymbol{z}_i \cdot v_i$ в $(*)$ принимает вид:

$$-2(b - \beta)\frac{1}{n}\sum_{i=1}^{n} x_i\varepsilon_i\boldsymbol{z}_i + (b - \beta)^2\frac{1}{n}\sum_{i=1}^{n} x_i^2\boldsymbol{z}_i. \qquad (**)$$

   Покажите, что plim первой компоненты равен нулю. **Указание:**

$$\mathrm{E}(x_i\varepsilon_i\boldsymbol{z}_i) = \mathrm{E}[\boldsymbol{z}_i \cdot x_i \cdot \mathrm{E}(\varepsilon_i|x_i)].$$

   Покажите, что plim второй компоненты равен нулю, если $\mathrm{E}(x_i^2\boldsymbol{z}_i)$ существует и конечно.

   (d) Таким образом, мы доказали, что $\widehat{\alpha} - \widetilde{\alpha}$ стремится к нулю. Докажите более сильный результат о том, что $\sqrt{n}(\widehat{\alpha} - \widetilde{\alpha})$ стремится к нулю. **Указание:** *Используйте лемму 2.4(b).*

7. (Дополнительно, докажите, что WLS асимптотически эффективнее, чем OLS.) В параграфе 2.8 WLS-оценка обозначена как $\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{V}})$. Мы хотим доказать, что она асимптотически эффективнее OLS, а именно

$$\mathrm{Avar}(\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{V}})) \leq \mathrm{Avar}(\boldsymbol{b}).$$

Но, поскольку $\mathrm{Avar}(\widehat{\boldsymbol{\beta}}(\widehat{\boldsymbol{V}})) = \mathrm{Avar}(\widehat{\boldsymbol{\beta}}(\boldsymbol{V}))$, достаточно доказать, что

$$\mathrm{Avar}(\widehat{\boldsymbol{\beta}}(\boldsymbol{V})) \leq \mathrm{Avar}(\boldsymbol{b}).$$

Докажите последнее матричное неравенство. **Указание:** *В главе 1 мы доказали алгебраический результат о том, что*

$$(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X})(\boldsymbol{X}'\boldsymbol{X})^{-1} \geq (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1} \qquad (*)$$

для любой положительно определенной матрицы $\boldsymbol{V}$. (Это следует из того факта, что GLS эффективнее OLS; левая часть есть (умноженная на σ^2) дисперсия OLS-оценки для обобщенной регрессионной модели, в то время как правая часть есть (умноженная на σ^2) дисперсия GLS-оценки.) Умножение обеих сторон $(*)$ на n и взятие предела по вероятности приводит к:

$$\left(\operatorname{plim}\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\operatorname{plim}\frac{1}{n}\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X}\right)\left(\operatorname{plim}\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \geq$$
$$\geq \operatorname{plim}\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X}\right)^{-1}. \qquad (**)$$

Правая часть есть $\operatorname{Avar}(\widehat{\boldsymbol{\beta}}(\boldsymbol{V}))$ (см. (2.8.7)). Покажите, что левая часть равна $\operatorname{Avar}(\boldsymbol{b}) = \boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1}\boldsymbol{S}\boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1}$.

8. Докажите (2.9.7). **Указание:** *Если $(\mu, \boldsymbol{\gamma})$ является вектором коэффициентов проекции наименьших квадратов, то*

$$\operatorname{E}(\boldsymbol{x}\boldsymbol{x}')\begin{bmatrix}\mu\\ \boldsymbol{\gamma}\end{bmatrix} = \operatorname{E}(\boldsymbol{x}\cdot y). \qquad (*)$$

или

$$\begin{bmatrix}\mu\\ \boldsymbol{\gamma}\end{bmatrix} = [\operatorname{E}(\boldsymbol{x}\boldsymbol{x}')]^{-1}\operatorname{E}(\boldsymbol{x}\cdot y). \qquad (**)$$

где

$$\operatorname{E}(\boldsymbol{x}\boldsymbol{x}') = \begin{bmatrix}1 & \operatorname{E}(\widetilde{\boldsymbol{x}})'\\ \operatorname{E}(\widetilde{\boldsymbol{x}}) & \operatorname{E}(\widetilde{\boldsymbol{x}}\widetilde{\boldsymbol{x}}')\end{bmatrix}, \operatorname{E}(\boldsymbol{x}\cdot y) = \begin{bmatrix}\operatorname{E}(y)\\ \operatorname{E}(\widetilde{\boldsymbol{x}}\cdot y)\end{bmatrix}.$$

Первым уравнением в $(*)$ является $\mu + \operatorname{E}(\widetilde{\boldsymbol{x}})'\boldsymbol{\gamma} = \operatorname{E}(y)$. Используйте его для удаления μ из остальных уравнений $(*)$, а затем найдите решение относительно $\boldsymbol{\gamma}$. При этом используйте $\operatorname{Var}(\widetilde{\boldsymbol{x}}) = \operatorname{E}(\widetilde{\boldsymbol{x}}\widetilde{\boldsymbol{x}}') - \operatorname{E}(\widetilde{\boldsymbol{x}})\operatorname{E}(\widetilde{\boldsymbol{x}}')$, $\operatorname{Cov}(\widetilde{\boldsymbol{x}}, y) = \operatorname{E}(\widetilde{\boldsymbol{x}}\cdot y) - \operatorname{E}(\widetilde{\boldsymbol{x}})\operatorname{E}(y)$. Альтернативно, используйте формулу для обращения блочных матриц (см. (A.10) приложения A), чтобы получить

$$\operatorname{E}(\boldsymbol{x}\boldsymbol{x}')^{-1} = \begin{bmatrix}1 + \operatorname{E}(\widetilde{\boldsymbol{x}})'\operatorname{Var}(\widetilde{\boldsymbol{x}})^{-1}\operatorname{E}(\widetilde{\boldsymbol{x}}) & -\operatorname{E}(\widetilde{\boldsymbol{x}})'\operatorname{Var}(\widetilde{\boldsymbol{x}})^{-1}\\ -\operatorname{Var}(\widetilde{\boldsymbol{x}})^{-1}\operatorname{E}(\widetilde{\boldsymbol{x}}) & \operatorname{Var}(\widetilde{\boldsymbol{x}})^{-1}\end{bmatrix}.$$

Затем подставьте это в $(**)$.

9. (Доказательство утверждения 2.9 для $p = 1$.) Предположим, что $\{\varepsilon_t\}$ — эргодически стационарная последовательность мартингал-разностей и что $\mathrm{E}(\varepsilon_t^2|\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots, \varepsilon_1) = \sigma^2 < \infty$. Пусть $g_t = \varepsilon_t \cdot \varepsilon_{t-1}$ и

$$\widehat{\gamma}_j = \frac{1}{n} \sum_{t=j+1}^{n} \varepsilon_t \cdot \varepsilon_{t-j} \quad (j = 0, 1), \; \widehat{\rho}_1 = \frac{\widehat{\gamma}_1}{\widehat{\gamma}_0}.$$

   (a) Покажите, что $\{g_t\}$ $(t = 2, 3, \ldots)$ является последовательностью мартингал-разностей.

   (b) Покажите, что $\mathrm{E}(g_t^2) = \sigma^4$.

   (c) Покажите, что $\sqrt{n}\widehat{\gamma}_1 \to_{\mathrm{d}} N(0, \sigma^4)$ при $n \to \infty$.

   (d) Покажите, что $\sqrt{n}\widehat{\rho}_1 \to_{\mathrm{d}} N(0, 1)$. **Указание:** *Сначала покажите, что* $\sqrt{n}\widehat{\gamma}_1/\sigma^2 \to_d N(0, 1)$. *Затем используйте лемму 2.4(c).*

10. (Асимптотика выборочного среднего для MA(2).) Пусть $(\varepsilon_{-1}, \varepsilon_0, \varepsilon_1, \ldots)$ является i.i.d. с нулевым средним и дисперсией σ_ε^2. Рассмотрим процесс $(y_{-1}, y_0, y_1, y_2, \ldots)$, порожденный как

$$y_{-1} = \varepsilon_{-1}, \; y_0 = \varepsilon_0 + \theta_1 \varepsilon_{-1}, \; y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} \quad (t = 1, 2, \ldots).$$

(Этот процесс называется процессом скользящего среднего второго порядка (*second-order moving average process*, MA(2)); см. главу 6.

   (a) Покажите, что $(y_1, y_2, \ldots)$ является стационарным в ковариациях. Получите выражение для автоковариаций γ_j $(j = 0, 1, 2, \ldots)$.

   (b) Пусть

$$r_{tj} \equiv \mathrm{E}(y_t|y_{t-j}, y_{t-j-1}, \ldots, y_0, y_{-1}) - \\ - \mathrm{E}(y_t|y_{t-j-1}, y_{t-j-2}, \ldots, y_0, y_{-1}) \\ (t = j, j+1, \ldots; j = 0, 1, 2, \ldots).$$

   Покажите, что

$$r_{t0} = \varepsilon_t, \; r_{t1} = \theta_1 \varepsilon_{t-1}, \; r_{t2} = \theta_2 \varepsilon_{t-2}, \; r_{t3} = 0, \; r_{t4} = 0, \; \ldots$$

   Указание: *Существует взаимно однозначное отображение между* $(\varepsilon_{-1}, \varepsilon_0, \ldots, \varepsilon_t)$ *и* $(y_{-1}, y_0, \ldots, y_t)$, *поэтому* $\mathrm{E}(y_t|y_{t-j}, \ldots, y_{-1}) = \mathrm{E}(y_t|\varepsilon_{t-j}, \ldots, \varepsilon_{-1})$.

   (c) Пусть

$$\overline{y}_n \equiv \frac{1}{n}(y_1 + y_2 + \cdots + y_n).$$

   Покажите, что

$$\mathrm{Var}(\sqrt{n}\overline{y}_n) = \gamma_0 + 2\left[\left(1 - \frac{1}{n}\right)\gamma_1 + \left(1 - \frac{2}{n}\right)\gamma_2\right].$$

(d) Поскольку $\{y_t\}$ не является i.i.d., CLT Линдеберга — Леви неприменима. Однако в главе 6 будет показано, что $\sqrt{n}\overline{y}_n$ сходится по распределению к нормальной случайной величине. Что является средним предельного распределения? Что является дисперсией предельного распределения (то есть $\mathrm{Avar}(\sqrt{n}\overline{y}_n)$)? Предположите, что лемма 2.1 применима к $\sqrt{n}\overline{y}_n$. **Указание:** *В лемме 2.1 положите* $z_n = \sqrt{n}\overline{y}_n$, $\lim_{n\to\infty}(1-q/n) = 1$ *для любого фиксированного* q.

11. (Дополнительно, тест Бройша — Годфри на сериальную корреляцию.) В этом упражнении мы доказываем, что модифицированная Q Бокса — Пирса асимптотически эквивалентна pF из вспомогательной регрессии (2.10.21), где p является числом лагов, а F является F-отношением для гипотезы о том, что все коэффициенты при $e_{t-1}, e_{t-2}, \ldots, e_{t-p}$ равны нулю.

Сначала установим обозначения. Пусть

$$\underset{(n\times K)}{\boldsymbol{X}} = \begin{bmatrix} \boldsymbol{x}'_1 \\ \vdots \\ \boldsymbol{x}'_n \end{bmatrix}, \quad \underset{(n\times p)}{\boldsymbol{E}} = \begin{bmatrix} 0 & 0 & \ldots & 0 \\ e_1 & 0 & \ldots & \vdots \\ e_2 & e_1 & & 0 \\ e_3 & e_2 & & e_1 \\ e_3 & e_2 & & e_1 \\ \vdots & \vdots & & \vdots \\ e_{n-1} & e_{n-2} & \ldots & e_{n-p} \end{bmatrix}, \quad \underset{(n\times 1)}{\boldsymbol{e}} = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix},$$

$$\underset{(K+p)\times(K+p)}{\boldsymbol{B}} = \begin{bmatrix} \frac{1}{n}\boldsymbol{X}'\boldsymbol{X} & \frac{1}{n}\boldsymbol{X}'\boldsymbol{E} \\ \frac{1}{n}\boldsymbol{E}'\boldsymbol{X} & \frac{1}{n}\boldsymbol{E}'\boldsymbol{E} \end{bmatrix}, \quad \widehat{\boldsymbol{B}}^{-1} = \begin{bmatrix} \underset{(K\times K)}{\widehat{\boldsymbol{B}}^{11}} & \underset{(K\times p)}{\widehat{\boldsymbol{B}}^{12}} \\ \underset{(p\times K)}{\widehat{\boldsymbol{B}}^{21}} & \underset{(p\times p)}{\widehat{\boldsymbol{B}}^{22}} \end{bmatrix}.$$

$$\underset{(p\times 1)}{\widehat{\gamma}} = \begin{bmatrix} \widehat{\gamma}_1 \\ \vdots \\ \widehat{\gamma}_p \end{bmatrix},$$

где e_t является остатком от исходной регрессии $y_t = \boldsymbol{x}'_t\boldsymbol{\beta}+\varepsilon_t$ и $\widehat{\gamma}_j$ является выборочной автоковариацией j-го порядка, определенной в (2.10.8), вычисленной по остаткам.

(a) Вспомогательная регрессия (2.10.21) имеет $K+p$ регрессоров. Пусть $\widehat{\alpha}$ является вектором $K+p$ коэффициентов. Покажите, что

$$\widehat{\alpha} = \widehat{\boldsymbol{B}}^{-1} \begin{bmatrix} \underset{(K\times 1)}{\boldsymbol{0}} \\ \underset{(p\times K)}{\widehat{\gamma}} \end{bmatrix}.$$

Указание: *Поскольку* e *является вектором остатков из исходной регрессии с* $\boldsymbol{X}$ *в качестве регрессора,* $\boldsymbol{X}'\boldsymbol{e} = \boldsymbol{0}$.

(b) Покажите:

$$\widehat{\boldsymbol{B}} \underset{\mathrm{p}}{\to} \boldsymbol{B} \equiv \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{xx}} & \boldsymbol{H} \\ \boldsymbol{H}' & \sigma^2 \boldsymbol{I}_p \end{bmatrix},$$

где

$$\underset{(K\times p)}{\boldsymbol{H}} \equiv \left[\mathrm{E}(\boldsymbol{x}_t \cdot \varepsilon_{t-1}) \quad \dots \quad \mathrm{E}(\boldsymbol{x}_t \cdot \varepsilon_{t-p}) \right].$$

Указание: j*-м столбцом* $\frac{1}{n}\boldsymbol{X}'\boldsymbol{E}$ *является* $\frac{1}{n}\sum_{t=j+1}^{n} \boldsymbol{x}_t \cdot e_{t-j}$. *Используйте* (2.3.9), *чтобы показать, что он сходится по вероятности к* $\mathrm{E}(\boldsymbol{x}_t \cdot \varepsilon_{t-j})$.

(c) (Очень легко.) Покажите, что $\widehat{\boldsymbol{\alpha}} \to_{\mathrm{p}} \boldsymbol{0}$.

(d) Пусть SSR здесь является суммой квадратов остатков в искусственной, а *не* исходной регрессии. Покажите:

$$\frac{SSR}{n-K-p} \underset{\mathrm{p}}{\to} \sigma^2,$$

где σ^2 является дисперсией ε_t, ошибки из оригинальной регрессии.

Указание: $\frac{1}{n}\boldsymbol{E}'\boldsymbol{e} = \widehat{\boldsymbol{\gamma}}$. *Покажите, что* $\frac{SSR}{n} = \frac{1}{n}\boldsymbol{e}'\boldsymbol{e} - \widehat{\boldsymbol{\alpha}}' \begin{bmatrix} \boldsymbol{0} \\ \widehat{\boldsymbol{\gamma}} \end{bmatrix}$.

(e) Покажите:

$$pF = \frac{n\widehat{\boldsymbol{\gamma}}'\widehat{\boldsymbol{B}}^{22}\widehat{\boldsymbol{\gamma}}}{SSR/(n-K-p)}.$$

Указание: *Примените формулу для* F*-отношения в* (1.4.9) *к вспомогательной регрессии.*

(f) Используйте формулу для обращения блочных матриц (см. (A.10) приложения A), чтобы показать

$$\widehat{\boldsymbol{B}}^{22} = \left[\frac{1}{n}\boldsymbol{E}'\boldsymbol{E} - \left(\frac{1}{n}\boldsymbol{E}'\boldsymbol{X}\right) \boldsymbol{S}_{\boldsymbol{xx}}^{-1} \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{E}\right) \right]^{-1}.$$

(g) Пусть $\widehat{\boldsymbol{\rho}}$ и $\widehat{\boldsymbol{\Phi}}$ будут такими же, как в (2.10.8) и (2.10.19). Покажите, что модифицированная Q Бокса — Пирса, определенная в (2.10.20), асимптотически эквивалентна pF (то есть разность между ними сходится к нулю по вероятности). **Указание:** *Покажите, что и* pF, *и модифицированная* Q *Бокса — Пирса асимптотически эквивалентны*

$$n\widehat{\boldsymbol{\gamma}}'(\boldsymbol{I}_p - \boldsymbol{\Phi})^{-1}\widehat{\boldsymbol{\gamma}}/\sigma^4,$$

где $\boldsymbol{\Phi}$ *определена в (2.10.18).*

12. (Дополнительно, тест Чоу на структурный сдвиг при больших выборках.) Рассмотрим применение теста Чоу на структурный сдвиг к регрессионной модели с условной гомоскедастичностью, описанной в утверждении 2.5. Поскольку проблема структурного сдвига возникает главным образом в моделях временных рядов, мы используем «t» в качестве индекса. Мы обобщаем предположение 2.1, допуская изменение вектора коэффициентов в момент сдвига r:

$$y_t = \begin{cases} \boldsymbol{x}_t'\boldsymbol{\beta}_1 + \varepsilon_t & \text{для } t = 1, 2, \ldots, r, \\ \boldsymbol{x}_t'\boldsymbol{\beta}_2 + \varepsilon_t & \text{для } t = r+1, r+2, \ldots, n. \end{cases}$$

Мы предполагаем, что дата сдвига r известна. (Когда дата сдвига неизвестна, ситуация более сложная и является объектом недавних исследований. См.: [Stock, 1994, Section 5] в качестве обзора.) Тестируемая нулевая гипотеза: $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$. Это множество из K ограничений. Пусть SSR_1 является суммой квадратов остатков для первого периода ($t = 1, 2, \ldots, r$), SSR_2 является SSR для второго периода ($t = r+1$, $r+2, \ldots, n$), а SSR_R является SSR для всей выборки при ограничении $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$. Статистика Чоу была определена в главе 1 как

$$F = \frac{[SSR_R - (SSR_1 + SSR_2)]/K}{(SSR_1 + SSR_2)/(n - 2K)}.$$

Пусть $\boldsymbol{b}_1$ является OLS-оценкой для $\boldsymbol{\beta}_1$, полученной по первому периоду, а $\boldsymbol{b}_2$ является OLS-оценкой для $\boldsymbol{\beta}_2$ по второму периоду.

(a) Покажите, что KF численно равно:

$$\sqrt{n}(\boldsymbol{b}_1 - \boldsymbol{b}_2)' \left[\left(\frac{\frac{1}{n}\sum_{t=1}^{r} \boldsymbol{x}_t\boldsymbol{x}_t'}{s^2} \right)^{-1} + \right.$$
$$\left. + \left(\frac{\frac{1}{n}\sum_{t=r+1}^{n} \boldsymbol{x}_t\boldsymbol{x}_t'}{s^2} \right)^{-1} \right]^{-1} \sqrt{n}(\boldsymbol{b}_1 - \boldsymbol{b}_2),$$

где $s^2 = (SSR_1 + SSR_2)/(n - 2K)$. **Указание:** *Приведенные выше n уравнений могут быть записаны в матричном виде как $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, где*

$$\underset{(n\times 2K)}{\boldsymbol{X}} \begin{bmatrix} \boldsymbol{X}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{X}_2 \end{bmatrix}, \quad \underset{(r\times K)}{\boldsymbol{X}_1} \begin{bmatrix} \boldsymbol{x}_1' \\ \vdots \\ \boldsymbol{x}_r' \end{bmatrix},$$

$$\underset{((n-r)\times K)}{\boldsymbol{X}_2} \begin{bmatrix} \boldsymbol{x}_{r+1}' \\ \vdots \\ \boldsymbol{x}_n' \end{bmatrix}, \quad \underset{(2K\times 1)}{\boldsymbol{\beta}} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix}.$$

Нулевая гипотеза о том, что $\beta_1 = \beta_2$, может быть записана как $\boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r}$, где $\boldsymbol{R} = [\boldsymbol{I}_K \ \vdots - \boldsymbol{I}_K]$ и $\boldsymbol{r} = \boldsymbol{0}$. Используйте утверждение 1.4.

(b) Пусть $\lambda \equiv r/n$. Покажите, что при $n \to \infty$ с фиксированным λ

$$\frac{1}{n}\sum_{t=1}^{r} \boldsymbol{x}_t\boldsymbol{x}_t' \underset{\mathrm{p}}{\to} \lambda\boldsymbol{\Sigma}_{\boldsymbol{xx}}, \quad \frac{1}{n}\sum_{t=r+1}^{n} \boldsymbol{x}_t\boldsymbol{x}_t' \underset{\mathrm{p}}{\to} (1-\lambda)\boldsymbol{\Sigma}_{\boldsymbol{xx}},$$

где $\boldsymbol{\Sigma}_{\boldsymbol{xx}} \equiv \mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}_t')$.

(c) Покажите, что при $n \to \infty$ с фиксированным λ

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{r} \boldsymbol{x}_t \cdot \varepsilon_t \underset{\mathrm{d}}{\to} N(\boldsymbol{0}, \lambda\sigma^2\boldsymbol{\Sigma}_{\boldsymbol{xx}}),$$

$$\frac{1}{\sqrt{n}}\sum_{t=r+1}^{n} \boldsymbol{x}_t \cdot \varepsilon_t \underset{\mathrm{d}}{\to} N(\boldsymbol{0}, (1-\lambda)\sigma^2\boldsymbol{\Sigma}_{\boldsymbol{xx}}).$$

Можно показать (см.: [Stock, 1994, Section 5]), что $\frac{1}{\sqrt{n}}\sum_{t=1}^{r}\boldsymbol{x}_t\varepsilon_t$ и $\frac{1}{\sqrt{n}}\sum_{t=r+1}^{n}\boldsymbol{x}_t\varepsilon_t$ асимптотически некоррелированы. Поэтому

$$\begin{bmatrix} \frac{1}{\sqrt{n}}\sum_{t=1}^{r} \boldsymbol{x}_t \cdot \varepsilon_t \\ \frac{1}{\sqrt{n}}\sum_{t=r+1}^{n} \boldsymbol{x}_t \cdot \varepsilon_t \end{bmatrix} \underset{\mathrm{d}}{\to} N\left(\boldsymbol{0}, \begin{bmatrix} \lambda\sigma^2\boldsymbol{\Sigma}_{\boldsymbol{xx}} & \boldsymbol{0} \\ \boldsymbol{0} & (1-\lambda)\sigma^2\boldsymbol{\Sigma}_{\boldsymbol{xx}} \end{bmatrix}\right).$$

(d) Покажите, что $\mathrm{Avar}(\boldsymbol{b}_1 - \boldsymbol{b}_2) = \lambda^{-1}\sigma^2\boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1} + (1-\lambda)^{-1}\sigma^2\boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1}$. **Указание:** *Пусть*

$$\boldsymbol{b} \equiv \begin{bmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \end{bmatrix}.$$

Покажите, что

$$\mathrm{Avar}(\boldsymbol{b}) = \begin{bmatrix} \lambda^{-1}\sigma^2\boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & (1-\lambda)^{-1}\sigma^2\boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1} \end{bmatrix}.$$

$\boldsymbol{b}_1 - \boldsymbol{b}_2 = [\boldsymbol{I}_K \ \vdots - \boldsymbol{I}_K]\boldsymbol{b}$. *Используйте лемму 2.4(c).*

(e) В заключение покажите, что $KF \to_{\mathrm{d}} \chi^2(K)$ при $n \to \infty$ с фиксированным $\lambda = r/n$.

Эмпирические упражнения

1. Перед выполнением этого упражнения прочитайте введение и параграфы I–IV в [Fama, 1975]. В файле данных MISHKIN.ASC месячные данные представлены в виде:

Столбец 1: год.

Столбец 2: месяц.

Столбец 3: одномесячный темп инфляции (в процентах, годовой темп; назовите его *PAI1*).

Столбец 4: трехмесячный темп инфляции (в процентах, годовой темп; назовите его *PAI3*).

Столбец 5: одномесячная ставка по казначейским векселям (в процентах, годовой темп; назовите его *TB1*).

Столбец 6: трехмесячная ставка по казначейским векселям (в процентах, годовой темп; назовите его *TB3*).

Столбец 7: CPI для городских потребителей, все составляющие (среднее за 1982–1984 годы приравнивается к 100; назовите его *CPI*).

Период выборки составляют наблюдения с февраля 1950 г. по декабрь 1990 г. (491 наблюдение). Данные по *PAI1*, *PAI3*, *TB1* и *TB3* те же самые, которые использовались в [Mishkin, 1992]) и стали доступными для нас благодаря ему. Данные по казначейским векселям получены из Center for Research in Security Prices (CRSP) Чикагского университета. Ставкой по казначейским векселям за конкретный месяц является ставка на последний рабочий день предыдущего месяца (и поэтому она может быть взята как процентная ставка на начало месяца). Построение *PAI1* и *PAI3* будет описано в конце этого упражнения; в течение некоторого времени мы будем использовать инфляцию, полученную из *CPI*.

(a) (Библиотечная/интернет-работа.) Чтобы проверить правильность данных в MISHKIN.ASC, найдите соответствующие таблицы из прошлых номеров *Treasury Bulletin* или *Federal Reserve Bulletin* с распечатанными данными о ставках по казначейским векселям. Или посетите веб-сайты Board of Governors (www.bog.frb.fed.us) или Treasury Department (www.ustreas.gov) для достижения той же цели. Можете ли вы найти одномесячные ставки по казначейским векселям? [Ответ: Вероятно, нет.] Являются ли ставки в MISHKIN.ASC близкими к соответствующим таблицам? Похожи ли они на ставки на начало месяца?

(b) (Библиотечная/интернет-работа.) Найдите соответствующие таблицы из прошлых номеров *Monthly Labor Review* (или посетите веб-сайт Бюро статистики труда, www.bls.gov) для проверки корректности графиков CPI в MISHKIN.ASC. Проверьте, что датировка переменных такова, что январское значение CPI является CPI за этот месяц. Относительно определения CPI проверьте следующее. 1) CPI является CPI для городских потребителей, для всех составляющих, включающих продукты питания и жилье, и без сезонной коррекции. 2) Измерение цен всех составляющих индекса производится в течение месяца. Когда объявляется CPI за месяц?

(c) CPI является индексом с фиксированными весами или переменными весами? **Указание:** *Разыщите ваш старый учебник по макроэкономике промежуточного уровня; учебник по макроэкономике для магистрантов не понадобится.*

Одномесячная ставка по казначейским векселям за месяц t в наборе данных соответствует периоду с начала месяца t до конца этого месяца (как вы только что проверили). В идеале, если бы у нас были данные по уровню цен на начало каждого периода, мы могли бы вычислить темп инфляции за тот же самый период как $(P_{t+1} - P_t)/P_t$, где P_t — уровень цен на начало периода. Мы используем CPI за месяц $t - 1$ для P_t (то есть полагаем $P_t = CPI_{t-1}$). Поскольку компоненты CPI собраны в различные моменты в течение месяца, возникает неизбежное смещение меры инфляции и данных по процентным ставкам. Другая проблема заключается в сроках опубликования данных о CPI. Из теории эффективности рынка следует, что P_t известна рынку на начало месяца t, когда установлены ставки на казначейские векселя на месяц. Однако CPI за месяц $t - 1$, который мы берем в качестве P_t, не объявляется до некоторого момента следующего месяца (месяца t). Таким образом, мы предполагаем, что люди знают CPI за месяц $t - 1$ на начало месяца t.

Совет при работе с пакетом TSP: При загрузке данных и вычислении темпа инфляции вы должны использовать способность TSP обрабатывать календарные даты. Начальная часть вашей программы в TSP могла бы выглядеть так:

```
? The data are monthly, 1950:2 thru 1990:12
freq m;smpl 50:2 90:12;
? Read in the ASCII data
read(file=?mishkin.asc?) year month pai1 pai3
tb1 tb3 cpi;
? Calculate inflation rate and the real rate
smpl 50:3 90:12;
pai=((cpi/cpi(-1))**12-1)*100;
r = tb1-pai;
```

Совет при работе с пакетом RATS: Аналогично начальная часть вашей программы RATS может выглядеть так:

```
* The data are monthly, 1950:2 thru 1990:12
cal 50 2 12;all 0 90:12
* Read in the ASCII data
open data mishkin.asc
data(org=obs)/year month pai1 pai3 tb1 tb3 cpi
* Calculate inflation rate and real rate
set pai = ((cpi(t)/cpi(t-1))**12-1)*100
set r = tb1(t)-pai(t)
```

(d) Воспроизведите результаты в табл. 2.1. Поскольку ставка по казначейским векселям указывается в процентах в расчете на год, темп инфляции должен измеряться в тех же единицах. Вычислите π_{t+1}, которая должна быть согласована с $TB1_t$ (одномесячная ставка по казначейским векселям за месяц t), то есть вычисляться по формуле сложных процентов:

$$\left[\left(\frac{P_{t+1}}{P_t}\right)^{12} - 1\right] \times 100.$$

Совет при работе с пакетом GAUSS: Один из способов вычисления выборочных автоковариаций по формуле (2.10.1) и выборочных автокорреляций по (2.10.2) таков. Пусть z является n-мерным вектором, i-я компонента которого — z_i, а nobs является размером выборки, n.

```
z=z-meanc(z); /* de-mean the series */
nlag=12; /* lag length */
rho=zeros(nlag,1);
j=1;do until j>nlag;
rho[j]=z'*shiftr(z',j,0)'/nobs;
 /* sample autocovariances */
j=j+1;endo;
rho=rho./(z'*z/nobs);
 /* sample autocorrelations */
```

Совет при работе с пакетом TSP: Используйте команду BJIDENT. Она не вычисляет p-значения, поэтому (если вы не готовы учиться, как использовать матричные команды TSP) откажитесь от идеи получения p-значений.

Совет при работе с пакетом TSP: Используйте CORRELATE или BOXJENK.

(e) Можете ли вы воспроизвести (2.11.9), которое использует робастные стандартные ошибки? Какова интерпретация константы?

Совет при работе с пакетом GAUSS: Для робастных к гетероскедастичности стандартных ошибок мы должны вычислить $\widehat{S}$ по формуле (2.5.1) с $\widehat{\varepsilon}_i = e_i$ (OLS-остаток). Пример **Gauss**-кода заключается в следующем. Пусть `x` — $n \times K$ матрица данных и `ehat` — n-мерный вектор OLS-остатков. Если `g` является $n \times K$-матрицей, i-я строка которой — $\boldsymbol{x}_i' \cdot e_i$, **Gauss**-код для порождения `g` — `g = x.*ehat`. **Gauss**-код для вычисления $\widehat{S}$ — `g'g/nobs`, где `nobs` является объемом выборки. Это, однако, не использует тот факт, что $\widehat{S}$ симметрична. Более вычислительно эффективной командой является `moment(g,0)/nobs`.

Совет при работе с пакетом TSP: Воспользуйтесь опцией `HCTYPE = 0` от OLSQ.

Совет при работе с пакетом GAUSS: Используйте опцию ROBUSTERRORS от `LINREG`.

(f) (Коррекция Дэвидсона — Мак-Киннона для оценки Уайта.) Свойства на конечных выборках робастного t-отношения могут быть улучшены коррекцией Дэвидсона — Мак-Киннона, которая рассматривалась в основном тексте. Вычислите робастные стандартные ошибки, используя три формулы, которые обсуждались в основном тексте. Первая — это коррекция на степени свободы умножением $\widehat{S}$ на $n/(n-K)$. Вторая — вычисление $\widehat{S}$ по формуле (2.5.5) с $d = 1$. Третья — (2.5.5) с $d = 2$. Они соответствуют опциям `HCTYPE = 1, 2,` 3 программы OLSQ для TSP.

Совет при работе с пакетом GAUSS: Пусть `p` — n-мерный вектор для сохранения p_i в (2.5.5). `sxxinv` = $S_{\boldsymbol{xx}}^{-1}$, `x` = матрица данных $\boldsymbol{X}$. Один из способов вычислить `p` таков:

```
p=zeros(nobs,1);
i=1;do until i>nobs;
p[i]=x[i,.]*sxxinv*x[i,.]'/nobs;
i=i+1;endo;
```

(g) (Оценивание при условной гомоскедастичности.) Проверьте эффективность рынка, регрессируя π_{t+1} на константу и $TB1_t$ *при условной гомоскедастичности* (2.6.1.) Сравните ваши результаты с полученными в (e) и (f). В какой части имеются отличия?

(h) (Тест Бройша — Годфри.) Для спецификации в (g) проведите тест Бройша — Годфри на сериальную корреляцию с $p = 12$. (Значение статистики nR^2 должно быть около 27,0). Пусть e_t $(t = 1, 2, \ldots, n)$ является OLS-остатком из регрессии в (g). Для выполнения теста Бройша — Годфри, как описано в тексте, нам нужно положить e_t $(t = 0, -1, \ldots, -11)$ равными нулю при проведении вспомогательной регрессии для $t = 1, \ldots, n$.

Совет при работе с пакетом TSP: Код TSP для вычисления статистики nR^2 таков. `resid` является OLS-остатком из исходной регрессии.

```
? Part (h): Breusch-Godfrey
resid=@res;
smpl 52:1 52:12;
resid=0;
smpl 53:1 71:7;
olsq resid c tbl resid(-1) resid(-2) resid(-3)
                 resid(-4) resid(-5) resid(-6)
                 resid(-7) resid(-8) resid(-9)
                 resid(-10) resid(-11)
                 resid(-12);
set w=@nob*@rsq;cdf(chisq,df=12) w;
```

Здесь `w` является статистикой nR^2.

Совет при работе с пакетом RATS: Код RATS для вычисления статистики nR^2 следующий. `resid` является OLS-остатком из исходной регрессии.

```
* Part (h): Breusch-Godfrey
set resid 53:1 71:7 = pai-%beta(1)-%beta(2)*tbl
set resid 52:1 52:12 = 0
linreg resid 53:1 71:7;
# constant tbl resid1 to 12
compute w = %nobs*%rsquared
cdf chisquared w 12
```

Здесь `w` является статистикой nR^2.

Совет при работе с пакетом GAUSS: Как и в кодах **Gauss** для (d), полезной является команда `shiftr`. Пусть `y` (223×1) и `x` (223×2), соответственно, являются вектором значений зависимой переменной и матрицей регрессоров из предыдущей регрессии. Пусть `b` (2×1) является вектором оцененных коэффициентов из предыдущей регрессии. Ваша программа **Gauss** для создания `y` и `x` для вспомогательной регрессии может выглядеть так:

```
"@ Part (h): Breusch-Godfrey @";
y=y-x*b; /* residual vector from the previous
regression */
j=1;do until j>12;
x=x shiftr(y',j,0)';
j=j+1;endo;
```

(i) (Реконструкция Фамы.) Каковы собственные точечные оценки и стандартные ошибки коэффициента при номинальной процентной ставке у Фамы? Идентичны ли они вашим результатам? (Фама использует другие обозначения, поэтому вы должны перевести его результаты в наши обозначения.) Почему его оценка константы отличается от вашей? (Одна очевидная причина заключается в том, что наши данные другие, но существует и другая причина.) Дополнительно: Каковы различия между его данными и нашими?

(j) (Сезонные дамми.) CPI, использованный для вычисления темпа инфляции, не являлся сезонно скорректированным. Чтобы принять во внимание сезонные факторы при продолжении использования сезонно нескорректированных данных в регрессии, определим двенадцать месячных дамми, $M1, M2, \ldots, M12$, и используем их вместо константы в регрессии в (g). Приводит ли это к изменению ваших результатов? Что случится, если вы включите константу вместе с двенадцатью месячными дамми в регрессию? Дополнительно: Есть ли какая-нибудь причина для предпочтения сезонно нескорректированных данных сезонно скорректированным?

Совет при работе с пакетом TSP: Используйте команду `DUMMY` для порождения месячных дамми.

Совет при работе с пакетом RATS: Используйте команду `SEASONAL` для порождения месячных дамми.

Хорошо известная проблема с рядом CPI заключается в том, что его компонента, связанная с жилищным строительством, представляет расходы на проценты по ипотеке до 1983 года и более подходящую меру — «эквивалент арендной платы» — с 1983 года. Переменные инфляции *PAI1* и *PAI3* в MISHKIN.ASC вычислены по индексу цен, который использует «эквивалент арендной платы» для всех периодов. Для получения дополнительной информации об этом см.: Section II [Huizinga and Mishkin, 1984]. Датировка переменных такова, что наблюдение за январь для *PAI1* вычислено по декабрьским и январским данным индекса цен, а *PAI3* — по декабрьским и мартовским данным индекса цен. Поэтому при предположении о том, что ценовой индекс соответствует концу месяца, январское наблюдение для *PAI1* является темпом инфляции в течение этого месяца, а январское наблюдение для *PAI3* является темпом инфляции с начала января по конец марта.

(k) Оцените регрессию Фамы для 1/53–7/71, используя лучшую меру одномесячного темпа инфляции, и сравните результаты с теми, которые получены в (e).

(l) Оцените регрессию Фамы за период после октября 1979 года. На много ли уменьшилась номинальная процентная ставка? (Коэффициент должен снизиться до 0,564.)

2. (Продолжение упражнения Нерлова в главе 1, с. 102.)

(a) Для модели 4 выполните тест nR^2 Уайта на условную гетероскедастичность. (Значение nR^2 должно быть равным 66,546).

(b) (Дополнительно.) Оцените Модель 4 посредством OLS. Это *шаг 1* процедуры параграфа 2.8. Выполните *шаг 2*, оценивая следующую вспомогательную регрессию: регрессируйте e_i^2 на константу и $1/Q_i$. Проверьте, что *шаг 3* является тем, что вы делали в части (h).

(c) Вычислите стандартные ошибки Уайта для модели 4.

Упражнения на метод Монте-Карло

1. (Коррекция на степени свободы.) В модели упражнения на метод Монте-Карло в главе 1 мы предполагали, что ошибки были нормальными. Предположим вместо этого, что ε_i распределены равномерно между $-0{,}5$ и $0{,}5$.

(a) Проверьте, что DGP второй симуляции (где $\boldsymbol{X}$ различна для симулируемых выборок) удовлетворяет предположениям 2.1–2.5 и 2.7. (Можно показать, что предположение 2.2 [эргодическая стационарность] также удовлетворяется. Это является следствием утверждения 6.1(d).)

(b) Проведите вторую симуляцию из упражнения на метод Монте-Карло в главе 1 с равномерно распределенными ошибками. В каждом повторении вычислите обычное t-отношение, как ранее, и сравните его с двумя критическими значениями. Первое является тем же самым критическим значением (2,042), которое указывалось распределением $t(30)$, а второе является критическим значением (1,96) для $N(0,1)$. Вычислите частоту отвержения для каждого критического значения. Какое из них ближе к номинальному размеру 5%?

2. (Противопоставление статистик Бокса — Пирса и Льюнга — Бокса.) Мы хотим проверить утверждение о том, что свойства на конечных выборках Q-статистики Льюнга — Бокса превосходят свойства на конечных выборках Q-статистики Бокса — Пирса. Создайте строку из 50 случайных чисел i.i.d. со средним 0. (Выберите распределение, которое вам нравится.) Беря эту строку в качестве данных, сделайте следующее:

(1) Вычислите статистики Q Бокса — Пирса и Q Льюнга — Бокса (см. (2.10.4) и (2.10.5)) с $p = 4$. (Среднее по ансамблю равно 0 по построению. Однако возьмите выборочное среднее ряда, когда вычисляете автокорреляции.)

(2) Для каждой статистики примите или отвергните нулевую гипотезу об отсутствии сериальной корреляции на уровне значимости 5%, предполагая, что статистика распределена как $\chi^2(4)$.

Проделайте это большое количество раз, каждый раз используя различные строки из 50 i.i.d. случайных чисел, порожденных заново. Для каждой Q-статистики вычислите частоту отвержения нулевой гипотезы. Если распределение на конечных выборках хорошо аппроксимируется $\chi^2(4)$, эта частота должна быть близка к 5%. Какая статистика дает вам частоту, более близкую к номинальному размеру 5%? Не слишком ли часто мы не отвергаем нулевую гипотезу, если используем Q Бокса — Пирса?

Ответы на избранные вопросы

Аналитические упражнения

3. Из (2.3.7) в тексте $\boldsymbol{b} - \boldsymbol{\beta} = \boldsymbol{S}_{\boldsymbol{xx}}^{-1}\overline{\boldsymbol{g}}$. По Колмогорову, $\boldsymbol{S}_{\boldsymbol{xx}} \to_{\mathrm{p}} \mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')$ и $\overline{\boldsymbol{g}} \to_{\mathrm{p}} \mathbf{0}$. Поэтому $\boldsymbol{b}$ состоятельна. По CLT Линдеберга — Леви, $\sqrt{n}\overline{\boldsymbol{g}} \to_{\mathrm{d}} N(0, \boldsymbol{S})$. Оставшаяся часть доказательства теперь должна быть рутинной.

7. Что остается показать — это то, что левая часть (**) равна:

$$\mathrm{Avar}(\boldsymbol{b}) = \boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1}\boldsymbol{S}\boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1}. \quad \boldsymbol{\Sigma}_{\boldsymbol{xx}} = \operatorname{plim}\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}$$

и

$$\begin{aligned}
\boldsymbol{S} &= \mathrm{E}(\varepsilon_i^2\boldsymbol{x}_i\boldsymbol{x}_i') = \\
&= \mathrm{E}[\mathrm{E}[\varepsilon_i^2|\boldsymbol{x}_i)\boldsymbol{x}_i\boldsymbol{x}_i'] = \\
&= \mathrm{E}[(\boldsymbol{z}_i'\boldsymbol{\alpha})\boldsymbol{x}_i\boldsymbol{x}_i'] \quad \text{(по (2.8.1))} = \\
&= \operatorname{plim}\frac{1}{n}\sum_{i=1}^{2}(\boldsymbol{z}_i'\boldsymbol{\alpha})\boldsymbol{x}_i\boldsymbol{x}_i' \quad \text{(по эргодической стационарности)} = \\
&= \operatorname{plim}\frac{1}{n}\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X} \quad \text{(по определению (2.8.4) для } \boldsymbol{V}\text{)}.
\end{aligned}$$

Эмпирические упражнения

1c. CPI, вероятно, наиболее широко использующийся индекс цен с фиксированными весами.

1e. Постоянная составляющая, взятая со знаком «минус», равна r (постоянная ex ante процентная ставка).

1g. Точечные оценки те же самые. Отличаются только стандартные ошибки.

1i. Посмотрите на таблицу 3 Фамы, его коэффициент при R_t, когда период выборки с 1/53 по 7/71, равен 0,98 с s.e. 0,10. Очень близко, но не точно то же самое. Есть два возможных объяснения этой разницы между нашими оценками и оценками Фамы. Во-первых, в вычислении темпа инфляции за месяц t, $\pi_{t+1} = (P_{t+1} - P_t)/P_t$, из Фамы неясно, что для P_t используется CPI_{t-1}, а не CPI_t. Во-вторых, вес для CPI во время его написания мог быть взят за 1958 год. Вес же для CPI в наших данных взят за 1982–1984 годы. Наша оценка константы (–0,868) отличается от оценки Фамы (0,00070), поскольку, во-первых, у него зависимой переменной является темп инфляции, взятый со знаком «минус», и, во-вторых, темп инфляции и ставка по казначейским векселям берутся у него в расчете на месяц.

1j Сезонно скорректированные ряды, выпускаемые BLS, являются своего рода двусторонними скользящими средними. Таким образом, например, сезонно скорректированный CPI_t зависит от сезонно нескорректированных значений *будущего* CPI, которые не содержатся в I_t. Значит, если нет причин принимать во внимание сезонные факторы в сотношении между темпом инфляции и номинальной процентной ставкой, нужно использовать сезонно нескорректированные данные. Если есть необходимость принять во внимание сезонные факторы, лучше включить сезонные дамми в регрессию, а не использовать сезонно скорректированные данные. Включение сезонных дамми не изменяет результаты сколько-нибудь существенным образом.

1k. Коэффициент R_t снижается с 1,014 до 0,807.

Литература

Amemiya, T., 1977, "A Note on a Heteroscedastic Model," *Journal of Economics*, 6, 365–370.

———,1985, *Advanced Econometrics*, Cambridge: Harvard University Press.

Anderson, T. W., 1971, *The Statistical Analysis of Time Series*, New York: Wiley.

Billingsley, P., 1961, "The Lindeberg-Levy Theorem for Martingales," in *Proceedings of the American Mathematical Society*, Volume 12, 788–792.

Box, G., and D. Pierce, 1970, "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time SeriesModels," *Journal of the American Statistical Association*, 65, 1509–1526.

Breusch, T., 1978, "Testing for Autocorrelation in Dynamic Linear Models," *Australian Economic Papers*, 17, 334–355.

Davidson, R., and J. MacKinnon, 1993, *Estimation and Inference in Econometrics*, Oxford: Oxford University Press.

Durbin, J., 1970, "Testing for Serial Correlation in Least-Squares Regression When Some of the Regressors are Lagged Dependent Variables," *Econometrica*, 38, 410–421.

Engle, R., 1982, "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007.

Fama, E., 1975, "Short-Term Interest Rates as Predictors of Inflation," *American Economic Review*, 65, 269–282.

———,1976, *Foundations of Finance*, Basic Books.

Godfrey, L., 1978, "Testing against General Autoregressive and Moving Average Error Models When the Regressors Include Lagged Dependent Variables," *Econometrica*, 46, 1293–1301.

Greene, W., 1997, *Econometric Analysis* (3d ed.), New York: Prentice Hall.

Gregory, A., and M. Veall, 1985, "FormulatingWald Tests of Nonlinear Restrictions," *Econometrica*, 53, 1465–1468.

Hall, P., and C. Heyde, 1980, *Martingale Limit Theory and Its Application*, New York: Academic.

Hall, R., 1978, "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence," *Journal of Political Economy*, 86, 971–987.

Hamilton, J., 1994, *Time Series Analysis*, Princeton: Princeton University Press.

Huizinga, J., and F. Mishkin, 1984, "Inflation and Real Interest Rates on Assets with Different Risk Characteristics," *Journal of Finance*, 39, 699–714.

Judge, G., W. Griffiths, R. Hill, H. Lütkepohl, and T. Lee, 1985, *The Theory and Practice of Econometrics* (2nd ed.), New York: Wiley.

Karlin, S., and H. M. Taylor, 1975, *A First Course in Stochastic Processes* (2d ed.), New York: Academic.

Koopmans, T., and W. Hood, 1953, "The Estimation of Simultaneous Linear Economic Relationships," in W. Hood and T. Koopmans (eds.), *Studies in Econometric Method*, New Haven: Yale University Press.

Mishkin, F., 1992, "Is the Fisher Effect for Real?" *Journal of Monetary Economics*, 30, 195–215.

Newey, W., 1985, "Generalized Method of Moments Specification Testing," *Journal of Economics*, 29, 229–256.

Rao, C. R., 1973, *Linear Statistical Inference and Its Applications* (2nd ed.), New York: Wiley.

Stock, J., 1994, "Unit Roots, Structural Breaks and Trends," in R. Engle and D. McFadden (eds.), *Handbook of Econometrics*, Volume IV, New York: North Holland.

White, H., 1980, "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 48, 817–838.

Глава 3. Одномерный GMM

Главным допущением метода наименьших квадратов является ортогональность случайных ошибок и регрессоров. Без этого допущения МНК-оценки не будут даже состоятельными. Поскольку во многих важных приложениях условие ортогональности не выполняется, необходимо уметь работать с эндогенными регрессорами. Решение дает метод оценивания, называемый **обобщенным методом моментов (GMM)** (*Generalized Method of Moments*), который включает в себя МНК как частный случай. В этой главе представлен GMM для моделей с одним уравнением, а в следующей главе представлено обобщение этого метода на несколько уравнений. Эти главы касаются только линейных уравнений. Оценка обобщенным методом моментов нелинейных уравнений будет описана в главе 7. В главе 8 будет представлена основная альтернатива GMM, метод максимального правдоподобия (ML).

Отражая преобладание в экономике эндогенных регрессоров, данная глава начинается с нескольких примеров. Общее определение эндогенных регрессоров приведено в параграфе 3.3. В параграфе 3.4 представлена GMM-процедура для модели из параграфа 3.3. Параграфы 3.5–3.7 представляют асимптотические свойства оценок GMM и соответствующих тестовых статистик. При наличии условной гомоскедастичности выведенные в данных параграфах формулы можно упростить. Упрощенные формулы собраны в параграфе 3.8. В частности, двухшаговый МНК, метод, разработанный изначально для оценивания систем одновременных уравнений, является частным случаем GMM. ML-альтернатива двухшагового МНК, называемая максимальным правдоподобием с ограниченной информацией (LIML), будет рассмотрена в параграфе 8.6.

Эмпирические упражнения в этой главе связаны с наиболее часто оцениваемым в экономике уравнением — уравнением заработной платы. Это уравнение связывает ставку заработной платы с различными характеристиками индивида, такими как образование и способности. Поскольку образование — это выбор индивида и поскольку способности оцениваются плохо, регрессоры, по-видимому, являются эндогенными. Применяя к этому уравнению описанные в данной главе методики, мы убеждаемся в том, что оценки параметров зависят от того, производим ли мы коррекцию на эндогенность.

3.1. Смещение из-за эндогенности: пример Уоркинга

Система одновременных уравнений рыночного равновесия

Классическим примером смещений, создаваемых эндогенными регрессорами, является пример Уоркинга (1927). Рассмотрим следующую простую модель спроса и предложения:

$$q_i^d = \alpha_0 + \alpha_1 p_i + u_i \quad \text{(уравнение спроса)}, \tag{3.1.1a}$$

$$q_i^s = \beta_0 + \beta_1 p_i + v_i \quad \text{(уравнение предложения)}, \tag{3.1.1b}$$

$$q_i^d = q_i^s \quad \text{(рыночное равновесие)}. \tag{3.1.1c}$$

где q_i^d — объем спроса на рассматриваемый товар (допустим, кофе) в период i, q_i^s — объем предложения, p_i — цена. Ошибка u_i отражает иные факторы помимо цены, влияющие на спрос на кофе, например, отношение к кофе. В зависимости от значения u_i кривая спроса в координатах «цена — объем» сдвигается вверх или вниз. Аналогично v_i отражает иные факторы предложения помимо цены. Предположим, что $E(u_i) = 0$ и $E(v_i) = 0$ (в противном случае свободные члены α_0 и β_0 изменятся на ненулевую величину). Для того чтобы избежать непринципиальных усложнений, предположим также, что $\mathrm{Cov}(u_i, v_i) = 0$. Если предположить, что $q_i = q_i^d = q_i^s$, систему из трех уравнений (3.1.1) можно свести к системе из двух уравнений:

$$q_i = \alpha_0 + \alpha_1 p_i + u_i \quad \text{(уравнение спроса)}, \tag{3.1.2a}$$

$$q_i = \beta_0 + \beta_1 p_i + v_i \quad \text{(уравнение предложения)}. \tag{3.1.2b}$$

Регрессор называется **эндогенным** (*endogenous*), если он не является предопределенным (то есть не ортогонален вектору ошибок) и, таким образом, не удовлетворяет условию ортогональности. При наличии в уравнении постоянной составляющей условие ортогональности нарушается, и, соответственно, регрессор является эндогенным тогда и только тогда, когда регрессор коррелирован с ошибкой. В данном примере регрессор p_i с необходимостью является эндогенным в обоих уравнениях. Чтобы понять почему, рассмотрим (3.1.2) как систему одновременных уравнений и решим ее относительно (p_i, q_i):

$$p_i = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{v_i - u_i}{\alpha_1 - \beta_1}, \tag{3.1.3a}$$

$$q_i = \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1} + \frac{\alpha_1 v_i - \beta_1 u_i}{\alpha_1 - \beta_1}. \tag{3.1.3b}$$

Таким образом, цена является функцией двух указанных ошибок. Используя (3.1.3а), можно вычислить ковариации регрессора p_i со сдвигом спроса u_i и сдвигом предложения v_i,

$$\mathrm{Cov}(p_i, u_i) = -\frac{\mathrm{Var}(u_i)}{\alpha_1 - \beta_1}, \ \mathrm{Cov}(p_i, v_i) = \frac{\mathrm{Var}(v_i)}{\alpha_1 - \beta_1}, \tag{3.1.4}$$

не равные нулю (кроме случая $\mathrm{Var}(u_i) = 0$ и $\mathrm{Var}(v_i) = 0$). Таким образом, если кривая спроса имеет отрицательный наклон ($\alpha_1 < 0$), а кривая предложения — положительный ($\beta_1 > 0$), цена положительно коррелирована со сдвигом спроса и отрицательно со сдвигом предложения. В данном примере эндогенность является следствием рыночного равновесия.

Смещение из-за эндогенности

Что получается в результате оценивания регрессии объема на константу и цену: кривая спроса или предложения? Ответ: ни то ни другое, так как цена эндогенна и в уравнении спроса, и в уравнении предложения. Из параграфа 2.9 следует, что оценки МНК соответствуют коэффициентам проекции наименьших квадратов. В проекции q_i на константу и p_i коэффициент при p_i равен $\mathrm{Cov}(p_i, q_i)/\mathrm{VAR}(p_i)$[1], таким образом:

$$\text{plim МНК-оценки коэффициента цены} = \frac{\mathrm{Cov}(p_i, q_i)}{\mathrm{Var}(p_i)}. \tag{3.1.5}$$

Чтобы разрешить это соотношение относительно коэффициента цены для кривой спроса (α_i), воспользуемся уравнением спроса (3.1.2а) для расчета $\mathrm{Cov}(p_i, q_i)$:

$$\mathrm{Cov}(p_i, q_i) = \alpha_1 \mathrm{Var}(p_i) + \mathrm{Cov}(p_i, u_i). \tag{3.1.6}$$

Подставив (3.1.6) в (3.1.5), получим выражение для асимптотического смещения α_1:

$$\text{plim МНК-оценки коэффициента цены} - \alpha_1 = \frac{\mathrm{Cov}(p_i, u_i)}{\mathrm{Var}(p_i)}. \tag{3.1.7}$$

Аналогично асимптотическое смещение для коэффициента цены для кривой предложения β_1 равно:

$$\text{plim МНК-оценки коэффициента цены} - \beta_1 = \frac{\mathrm{Cov}(p_i, v_i)}{\mathrm{Var}(p_i)}. \tag{3.1.8}$$

[1] Пусть $\boldsymbol{\gamma}$ — вектор коэффициентов МНК в уравнении $\widehat{\mathrm{E}}^*(y|1, \boldsymbol{x}) = \alpha + \boldsymbol{x}'\boldsymbol{\gamma}$. Тогда $\boldsymbol{\gamma} = \mathrm{Var}(\boldsymbol{x})^{-1}\,\mathrm{Cov}(\boldsymbol{x}, y)$. Доказательство этого факта было аналитическим упражнением в главе 2.

Но, как показано в (3.1.4), $\mathrm{Cov}(p_i, u_i) \neq 0$ и $\mathrm{Cov}(p_i, v_i) \neq 0$, поэтому МНК-оценки не соответствуют ни α_1, ни β_1. Это явление известно как **смещение из-за эндогенности** (*endogeneity bias*). Оно также известно как **смещение из-за одновременности уравнений** (*simultaneous equations bias*), или **смещение из-за одновременности** (*simultaneity bias*), так как регрессор и случайная ошибка часто связаны друг с другом посредством системы одновременных уравнений, как в данном примере.

В случае отсутствия сдвигов спроса ($u_i = 0$ для всех i) $\mathrm{Cov}(p_i, u_i) = 0$ формула (3.1.7) показывает, что МНК-оценка соответствует параметру α_1. В этом случае кривая спроса не сдвигается, и, как показано на рис. 3.1(a), все наблюдаемые при сдвиге кривой предложения комбинации цены и объема попадают на кривую спроса. В случае отсутствия сдвигов предложения все наблюдаемые при сдвиге кривой спроса комбинации цены и объема попадают на кривую предложения (см. рис. 3.1(b)). В общем случае сдвига обеих кривых МНК-оценка соответствует взвешенному среднему α_1 и β_1. Это можно показать аналитически выводом другого выражения для предела по вероятности МНК-оценки:

$$\text{plim МНК-оценки коэффициента цены} = \frac{\alpha_1 \mathrm{Var}(v_i) + \beta_1 \mathrm{Var}(u_i)}{\mathrm{Var}(v_i) + \mathrm{Var}(u_i)}. \tag{3.1.9}$$

Доказательство этого предлагается как контрольный вопрос.

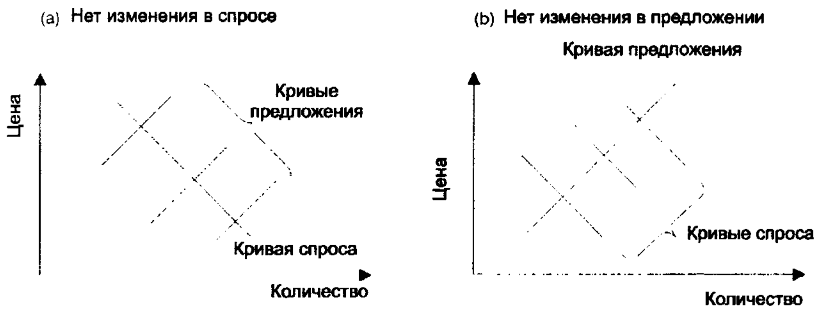

Рис. 3.1. Кривые спроса и предложения

Наблюдаемые сдвиги предложения

Причина, по которой кривые спроса и предложения не оцениваются состоятельно в общем случае, заключается в том, что мы не можем определить из данных, чем вызвано изменение цены и объема: сдвигом спроса или сдвигом предложения. Можно предположить, что оценить кривую спроса (предложения) возможно, если наблюдаются некоторые из факторов, сдвигающих кривую предложения (спроса). Предположим, что сдвиг предложения v_i можно разложить на наблюдаемый фактор x_i и на ненаблюдаемый фактор ζ_i, некоррелированный с x_i[1].

$$q_i = \beta_0 + \beta_1 p_i + \beta_2 x_i + \zeta_i, \quad \text{где } \beta_2 \neq 0 \text{ (предложение).} \tag{3.1.2b'}$$

Предположим теперь, что наблюдаемый сдвиг предложения x_i является предопределенным в уравнении спроса (то есть не коррелирует с ошибкой); x_i можно считать, например, температурой в регионах выращивания кофе. Если температура x_i не коррелирует с пристрастием к кофе u_i, то должна существовать возможность вычленить из движения цен компоненту, связанную с температурой, но не связанную со сдвигом спроса. Тогда возможно оценить кривую спроса как связь между потреблением кофе и этой компонентой цены.

Для рассматриваемого уравнения предопределенная переменная, коррелированная с эндогенным регрессором, называется **инструментальной переменной** или **инструментом** (*instrumental variable or instrument*). Ее часто называют **релевантным инструментом**[2] (*valid instrument*), чтобы подчеркнуть ее ненулевую корреляцию с эндогенным регрессором. В данном примере наблюдаемый сдвиг предложения x_i может служить инструментом для уравнения спроса. Это можно легко увидеть, решив систему одновременных уравнений (3.1.2a) и (3.1.2b') относительно (p_i, q_i):

$$p_i = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{\beta_2}{\alpha_1 - \beta_1} x_i + \frac{\zeta_i - u_i}{\alpha_1 - \beta_1}, \tag{3.1.10a}$$

$$q_i = \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1} + \frac{\alpha_1\beta_2}{\alpha_1 - \beta_1} x_i + \frac{\alpha_1\zeta_i - \beta_1 u_i}{\alpha_1 - \beta_1}. \tag{3.1.10b}$$

Поскольку $\mathrm{Cov}(x_i, \zeta_i) = 0$ по построению и $\mathrm{Cov}(x_i, u_i) = 0$ по предположению, из (3.1.10a) следует, что

$$\mathrm{Cov}(x_i, p_i) = \frac{\beta_2}{\alpha_1 - \beta_1}\,\mathrm{Var}(x_i) \neq 0.$$

Поэтому x_i — релевантный инструмент.

Имея релевантный инструмент, можно оценить коэффициент α_1 и кривую предложения. Воспользуемся уравнением (3.1.2a) для оценивания $\mathrm{Cov}(x_i, q_i)$ (а не $\mathrm{Cov}(p_i, q_i)$, как в (3.1.6)):

$$\begin{aligned}\mathrm{Cov}(x_i, q_i) &= \alpha_1 \mathrm{Cov}(x_i, p_i) + \mathrm{Cov}(x_i, u_i) = \\ &= \alpha_1 \mathrm{Cov}(x_i, p_i) \quad (\mathrm{Cov}(x_i, u_i) = 0 \text{ по предположению}).\end{aligned}$$

[1] Это разложение всегда возможно. Если проекция v_i на константу и x_i равна $\gamma_0 + \beta_2 x_i$, то определим $\zeta_i = v_i - \gamma_0 - \beta_2 x_i$, тем самым $v_i = \zeta_i + \gamma_0 + \beta_2 x_i$. ζ_i, по определению, не коррелирует с x_i. Подставив это уравнение в (3.1.2b) и включив γ_0 в константу, получим (3.1.2b$'$).

[2] В оригинале используется термин «valid instrument», однако в литературе в этом значении чаще используется термин «relevant» («релевантный»). Термин же «valid» («валидный») употребляется чаще как указание на некоррелированность инструмента с ошибкой. — *Прим. науч. ред. перевода.*

Как мы только что убедились, $\mathrm{Cov}(x_i, p_i) \neq 0$. Поэтому можно разделить обе части равенства на $\mathrm{Cov}(x_i, p_i)$ и получить

$$\alpha_1 = \frac{\mathrm{Cov}(x_i, q_i)}{\mathrm{Cov}(x_i, p_i)}. \tag{3.1.11}$$

Естественной оценкой, которая напрашивается сама собой, является:

$$\widehat{\alpha}_{1,\mathrm{IV}} = \frac{\text{выборочная ковариация } x_i \text{ и } q_i}{\text{выборочная ковариация } x_i \text{ и } p_i}. \tag{3.1.12}$$

Эта оценка называется **оценкой по методу инструментальных переменных (IV)** (*instrumental variables (IV) estimator*) с x_i в качестве инструмента. Мы иногда говорим, что «эндогенный регрессор p_i инструментируется посредством x_i».

Другой популярной процедурой состоятельного оценивания α_1 является **двухшаговый МНК (2SLS)** (*two-stage least squares (2SLS).*). Он называется так потому, что эта процедура состоит из оценивания двух регрессий. На первом шаге оценивается регрессия эндогенного регрессора p_i на предопределенную переменную x_i для получения подобранных значений $\widehat{p}_i$. На втором шаге оценивается регрессия зависимой переменной q_i на константу и $\widehat{p}_i$. Использование $\widehat{p}_i$ вместо p_i отличает 2SLS от бесхитростного применения МНК для оценивания уравнения спроса. 2SLS-оценка коэффициента α_1 — это МНК-оценка коэффициента при $\widehat{p}_i$ на втором шаге. Поэтому она равна[1]:

$$\widehat{\alpha}_{1,2\mathrm{SLS}} = \frac{\text{выборочная ковариация } \widehat{p}_i \text{ и } q_i}{\text{выборочная дисперсия } \widehat{p}_i}. \tag{3.1.13}$$

Чтобы связать регрессию на втором шаге с уравнением спроса, перепишем (3.1.2а) как

$$q_i = \alpha_0 + \alpha_1 \widehat{p}_i + [u_i + \alpha_1(p_i - \widehat{p}_i)]. \tag{3.1.14}$$

Регрессия на втором шаге оценивает это уравнение, трактуя выражение в скобках как ошибку. Получаемая при этом МНК-оценка α_1 является состоятельной по следующей причине. Если бы подобранное значение $\widehat{p}_i$ было в точности равно МНК-проекции $\widehat{\mathrm{E}}^*(p_i|1, x_i)$, то ни u_i, ни $(p_i - \widehat{p}_i)$ не были бы коррелированными с $\widehat{p}_i$: u_i — потому, что она не коррелирована с x_i и $\widehat{p}_i$ — линейная функция x_i, а $(p_i - \widehat{p}_i)$ — потому, что она является ошибкой проекции. Подобранное значение $\widehat{p}_i$ не равно точно $\widehat{\mathrm{E}}^*(p_i|1, x_i)$, но эта разница исчезает по мере увеличения объема

[1] В регрессии y_i на константу и x_i МНК-оценка коэффициента при x_i равна отношению выборочной ковариации между x_i и y_i к выборочной дисперсии x_i. Доказательство этого было контрольным вопросом к параграфу 1.2.

выборки. Таким образом, $\widehat{p}_i$ асимптотически не коррелирует с ошибкой в уравнении (3.1.14), что делает 2SLS-оценку состоятельной.

В данном примере IV- и 2SLS-оценки численно совпадают (это будет доказано позднее в более общем случае). В общем, 2SLS-оценка может быть записана как IV-оценка с надлежащим образом выбранными инструментами; а IV-оценка, в свою очередь, является частным случаем GMM-оценки.

Контрольные вопросы

1. В системе одновременных уравнений (3.1.2) предположим, что $\mathrm{Cov}(u_i, v_i)$ не обязательно равна 0. Обязательно ли, что цена и сдвиг спроса u_i положительно коррелированы, если $\alpha_1 < 0$ и $\beta_1 > 0$? [Ответ: Нет.] Почему?

2. (3.1.7) показывает, что МНК-оценка коэффициента при цене в регрессии объема на константу и цену смещена. Смещена ли МНК-оценка постоянной составляющей α_0? **Указание:** *Предел по вероятности МНК-оценки постоянной составляющей равен:*
$$E(q_i) - \frac{\mathrm{Cov}(p_i, q_i)}{\mathrm{Var}(p_i)} E(p_i).$$
Но из (3.1.2a) *следует* $E(q_i) = \alpha_0 + \alpha_1 E(p_i)$.

3. Выведите (3.1.9). **Указание:** *Покажите, что*
$$\mathrm{Var}(p_i) = \frac{\mathrm{Var}(v_i) + \mathrm{Var}(u_i)}{(\alpha_1 - \beta_1)^2},$$
$$\mathrm{Cov}(p_i, q_i) = \frac{\alpha_1 \mathrm{Var}(v_i) + \beta_1 \mathrm{Var}(u_i)}{(\alpha_1 - \beta_1)^2}.$$

4. Для модели рыночного равновесия (3.1.2a), (3.1.2b′), в которой
$$\mathrm{Cov}(u_i, \zeta_i) = 0, \quad \mathrm{Cov}(x_i, u_i) = 0, \quad \mathrm{Cov}(x_i, \zeta_i) = 0,$$
проверьте, что в уравнении спроса цена эндогенна. А в уравнении предложения? **Указание:** *Посмотрите на* (3.1.10a).

   Необходимо ли предположение о том, что сдвиги спроса и предложения некоррелированы ($\mathrm{Cov}(u_i, v_i) = 0$) для состоятельности оценок $\widehat{\alpha}_{1,IV}$ и $\widehat{\alpha}_{1,2SLS}$? **Указание:** *Является ли x_i релевантным инструментом без этого предположения?*

3.2. Другие примеры

Смещение из-за эндогенности возникает в самых различных ситуациях. Мы рассмотрим еще несколько примеров.

Простая макроэкономическая модель

Предельно простой макроэкономической моделью является иллюстративная модель Хаавелмо (1943):

$$C_i = \alpha_0 + \alpha_1 Y_i + u_i, \quad 0 < \alpha_1 < 1 \quad \text{(функция потребления)},$$
$$Y_i = C_i + I_i \quad \text{(уравнение ВВП)}.$$

где C_i — совокупное потребление в году i, Y_i — ВВП, I_i — инвестиции, α_1 — предельная склонность к потреблению (**МРС** — *Marginal Propensity to Consume*). Как известно из начального курса микроэкономики, равновесное значение ВВП:

$$Y_i = \frac{\alpha_0}{1-\alpha_1} + \frac{I_i}{1-\alpha_1} + \frac{u_i}{1-\alpha_1}. \tag{3.2.1}$$

Если инвестиции предопределены, так что $\mathrm{Cov}(I_i, u_i) = 0$, то из (3.2.1) следует, что

$$\mathrm{Cov}(Y_i, u_i) = \frac{\mathrm{Var}(u_i)}{1-\alpha_1} > 0,$$
$$\mathrm{Cov}(I_i, Y_i) = \frac{\mathrm{Var}(I_i)}{1-\alpha_1} > 0.$$

Таким образом, доход эндогенен в функции потребления, но инвестиции являются релевантным инструментом для этого эндогенного регрессора. Простые выкладки, аналогичные используемым для вывода (3.1.7), показывают, что МНК-оценка для МРС, полученная из регрессии потребления на константу и доход, асимптотически смещена:

$$\mathrm{plim}\, \widehat{\alpha}_{1,\mathrm{OLS}} - \alpha_1 = \frac{\mathrm{Cov}(Y_i, u_i)}{\mathrm{Var}(Y_i)} = \frac{1-\alpha_1}{1 + \frac{\mathrm{Var}(I_i)}{\mathrm{Var}(u_i)}} > 0. \tag{3.2.2}$$

Это, возможно, самый очевидный пример смещения из-за одновременности. Отличие от примера Уоркинга заключается в том, что здесь второе уравнение является тождеством, помогающим увидеть эндогенность регрессора. Асимптотическое смещение может быть скорректировано использованием инвестиций в качестве инструмента для дохода в функции потребления. Инвестиции здесь играют роль наблюдаемого сдвига предложения в примере Уоркинга.

Ошибки в переменных

Термин **«ошибки в переменных»** (*errors-in-variables*) относится к случаю, когда предопределенная в иных обстоятельствах переменная с необходимостью становится эндогенной, когда ее измерения проводятся с ошибкой.

Эта проблема встречается весьма часто, особенно в микроэкономических данных по домохозяйствам. Например, в **панельном исследовании динамики доходов** (**PSID** — *Panel Study of Income Dynamics*) информация о переменных, таких как потребление продуктов питания и доход, собиралась по телефону. Было бы слишком оптимистичным предполагать, что опрашиваемый сможет с ходу назвать сумму, потраченную на продукты питания в определенный период времени.

Кросс-секционная версия **гипотезы постоянного дохода** (*Permanent Income Hypothesis*) Фридмана (1957) может быть аккуратно сформулирована как задача с ошибками в переменных. Эта гипотеза утверждает, что «постоянное потребление» C_i^* домохозяйства i пропорционально «постоянному доходу» Y_I^*:

$$C_i^* = kY_i^* \quad \text{с } 0 < k < 1. \tag{3.2.3}$$

Предполагается, что измеренное потребление C_i и измеренный доход Y_i являются значениями постоянного дохода и потребления, измеренными с ошибкой:

$$C_i = C_i^* + c_i \quad \text{и} \quad Y_i = Y_i^* + y_i \tag{3.2.4}$$

Предполагается, что ошибки измерения c_i и y_i имеют нулевое среднее и не коррелированы с постоянным потреблением и доходом:

$$\mathrm{E}(c_i) = 0, \quad \mathrm{E}(y_i) = 0, \quad \mathrm{E}(c_i y_i) = 0, \tag{3.2.5}$$

$$\mathrm{E}(C_i^* c_i) = 0, \quad \mathrm{E}(Y_i^* y_i) = 0, \quad \mathrm{E}(C_i^* y_i) = 0, \quad \mathrm{E}(Y_i^* c_i) = 0. \tag{3.2.6}$$

Подставляя (3.2.4) в (3.2.3), указанное соотношение можно выразить в терминах измеренных потребления и дохода:

$$C_i = kY_i + u_i \quad \text{с} \quad u_i \equiv c_i - ky_i. \tag{3.2.7}$$

Данный пример отличается от предыдущих тем, что уравнение не содержит постоянной составляющей. Поэтому для определения того, является ли переменная предопределенной, следует использовать смешанный момент $\mathrm{E}(Y_i u_i)$, а не ковариацию $\mathrm{Cov}(Y_i, u_i)$. Из (3.2.4)–(3.2.7) нетрудно вывести, что

$$\mathrm{E}(Y_i u_i) = -k\,\mathrm{E}(y_i^2) < 0.$$

Таким образом, измеренный доход эндогенен в функции потребления (3.2.7). В отличие от предыдущих примеров, эта эндогенность является следствием ошибок измерения. С учетом того факта, что оценка МНК коэффициента при Y_i в (3.2.7) соответствует коэффициенту проекции наименьших квадратов $\mathrm{E}(C_i Y_i)/\mathrm{E}(Y_i^2)$, из (3.2.4)–(3.2.7) можно также получить, что

$$\operatorname{plim} \widehat{k}_{\mathrm{OLS}} = \frac{k\,\mathrm{E}[(Y_i^*)^2]}{\mathrm{E}[(Y_i^*)^2] + \mathrm{E}(y_i^2)} < k. \tag{3.2.8}$$

То есть оценка регрессии измеренного потребления на измеренный доход (без постоянной составляющей) дает заниженную оценку k. Фридман использовал эти результаты для объяснения того, почему оценка MPC, полученная с помощью кросс-секционной регрессии потребления на доход, меньше аналогичной оценки MPC, полученной на основе наблюдений агрегированных временных рядов.

Предположим на мгновение, что существует релевантный инструмент x_i, так что $\mathrm{E}(x_i u_i) = 0$ и $\mathrm{E}(x_i Y_i) \neq 0$. Аналогично выводу (3.1.11) можно получить, что

$$k = \frac{\mathrm{E}(x_i C_i)}{\mathrm{E}(x_i Y_i)}. \tag{3.2.9}$$

Поэтому состоятельной IV-оценкой будет

$$\widehat{k}_{\mathrm{IV}} = \frac{\text{выборочное среднее } x_i C_i}{\text{выборочное среднее } x_i Y_i}. \tag{3.2.10}$$

Но существует ли релевантный инструмент? Да, и это постоянная величина. Подставляя $x_i = 1$ в (3.2.10), мы получим состоятельную оценку для k, являющуюся отношением выборочного среднего измеренного потребления к выборочному среднему измеренного дохода. Именно так Фридман и оценивал k.

Производственная функция

Во многих ситуациях, случайная компонента модели включает в себя факторы, которые известны рассматриваемому экономическому агенту, но не наблюдаются эконометристом. Эндогенность возникает, когда регрессоры являются решениями, принимаемыми агентом на основе этих факторов. Рассмотрим кросс-секционную выборку компаний, выбирающих объем потребляемого труда для максимизации прибыли. Производственная функция фирмы i

$$Q_i = A_i \cdot (L_i)^{\phi_1} \cdot \exp(v_i), \quad 0 < \phi_1 < 1, \tag{3.2.11}$$

где Q_i — выпуск фирмы, L_i — потребляемый труд, A_i — известный фирме уровень ее эффективности, v_i — технологический шок. В отличие от A_i v_i не наблюдаем для фирмы в момент выбора объема труда L_i. Ни A_i, ни v_i не наблюдаемы для эконометриста и независимы друг от друга.

Предположим, что для каждой фирмы v_i сериально независим. Таким образом, $B \equiv \mathrm{E}[\exp(v_i)]$ одинаково для всех фирм[1], и ожидаемый фирмой

[1] Если v_i для фирмы i коррелированы во времени, фирма может использовать прошлые значения v_i для оценки текущего значения и, таким образом, B будет различным для разных фирм. Также поскольку математическое ожидание значения нелинейной функции от случайной величины с нулевым математическим ожиданием необязательно равно нулю, то B не обязано быть равным нулю, даже если $\mathrm{E}(v_i) = 0$.

уровень выпуска равен:

$$A_i \cdot (L_i)^{\phi_1} \cdot B.$$

Пусть p — цена продукции, w — уровень заработной платы. Для упрощения предположим, что все фирмы из одной и той же конкурентной отрасли, то есть p и w постоянны для всех фирм. Цель фирмы i — выбрать объем L_i, максимизирующий ожидаемую прибыль:

$$p \cdot A_i \cdot (L_i)^{\phi_1} \cdot B - w \cdot L_i.$$

Взяв производную этого выражения по L_i, приравнивая ее к нулю и решая полученное уравнение относительно L_i, получим оптимальный объем затрат труда L_i:

$$L_i = \left(\frac{w}{p}\right)^{\frac{1}{\phi_1 - 1}} (A_i B \phi_1)^{\frac{1}{1-\phi_1}}. \tag{3.2.12}$$

Пусть u_i — отклонение фирмы i от среднего логарифма эффективности: $u_i \equiv \log(A_i) - \mathrm{E}[\log(A_i)]$, а $\phi_0 \equiv \mathrm{E}[\log(A_i)]$ (так что $\mathrm{E}(u_i) = 0$ и $A_i = \exp(\phi_0 + u_i)$). Тогда производственную функцию (3.2.11) и функцию спроса на труд (3.2.12) можно записать в логарифмах как

$$\log(Q_i) = \phi_0 + \phi_1 \log(L_i) + (v_i + u_i), \tag{3.2.13}$$

$$\log(L_i) = \beta_0 + \frac{1}{1 - \phi_1} u_i, \tag{3.2.14}$$

где

$$\beta_0 = \frac{1}{\phi_1 - 1} [\log(w/p) - \phi_0 - \log(\phi_1 B)].$$

Поскольку для всех фирм цены одинаковы, β_0 постоянна для всех фирм. Формула (3.2.14) показывает, что в логлинейной производственной функции (3.2.12) $\log(L_i)$ — эндогенный регрессор, положительно коррелированный с ошибкой $(v_i + u_i)$ через u_i. Таким образом, МНК-оценка ϕ_1 при оценивании логлинейной производственной функции смешивает вклад в выпуск u_i со вкладом труда. Этот пример показывает еще один источник эндогенности — учет факторов, которые не наблюдаются эконометристом, но наблюдаются агентом, принимающим решения.

Контрольные вопросы

1. Рассмотрите регрессию C_i на константу и Y_i в рамках гипотезы постоянного дохода Фридмана. Выведите предел по вероятности МНК-оценки коэффициента при Y_i в этой модели. **Указание:** *Предел по вероятности равен отношению* $\mathrm{Cov}(C_i, Y_i)$ *к* $\mathrm{Var}(Y_i)$. *Покажите, что оно равно*

$$\frac{k \operatorname{Var}(Y_i^*)}{\operatorname{Var}(Y_i^*) + \operatorname{Var}(y_i)}.$$

2. В примере с производственной функцией покажите, что $\operatorname{plim}_{n\to\infty} \widehat{\phi}_{1.OLS} = 1$, где $\widehat{\phi}_{1.OLS}$ — МНК-оценка ϕ_1 из (3.2.13). **Указание:** *Исключите u_i из уравнения* (3.2.13) *с помощью функции спроса на труд* (3.2.14).

3. В примере с производственной функцией предположим, что перед принятием решения о затратах труда фирмы могут наблюдать и v_i, и u_i. Как изменится функция спроса на труд (3.2.12)? Покажите, что $\log(Q_i)$ и $\log(L_i)$ абсолютно коррелированы. **Указание:** $\log(Q_i)$ *будет точной линейной функцией от* $\log(L_i)$ *без случайных ошибок.*

3.3. Общая постановка задачи

Дадим теперь общую постановку задачи. Модель описывается следующим набором предположений и является обобщением модели из главы 2.

Регрессоры и инструменты

Предположение 3.1 (линейность): *Оцениваемое уравнение линейно:*

$$y_i = \boldsymbol{z}_i'\boldsymbol{\delta} + \varepsilon_i \quad (i = 1, 2, \ldots, n).$$

где $\boldsymbol{z}_i$ — L-мерный вектор регрессоров, $\boldsymbol{\delta}$ — L-мерный вектор коэффициентов, а ε_i — ненаблюдаемая ошибка.

Предположение 3.2 (эргодическая стационарность):
Пусть $\boldsymbol{x}_i$ — K-мерный вектор, рассматриваемый как вектор инструментов, и пусть $\boldsymbol{w}_i$ — уникальные (различные) и непостоянные элементы $(y_i, \boldsymbol{z}_i, \boldsymbol{x}_i)$[1]. *$\{\boldsymbol{w}_i\}$ совместно стационарен и эргодичен.*

Предположение 3.3 (условия ортогональности): *Все K переменных в $\boldsymbol{x}_i$* ***«предопределены»*** *(predetermined) в том смысле, что они ортогональны текущей ошибке: $\mathrm{E}(x_{ik}\varepsilon_i) = 0$ для всех i и k $(k = 1, 2, \ldots, K)$*[2].

Это может быть записано как

$$\mathrm{E}[\boldsymbol{x}_i \cdot (y_i - \boldsymbol{z}_i'\boldsymbol{\delta})] = \mathbf{0} \quad \text{или} \quad \mathrm{E}(\boldsymbol{g}_i) = \mathbf{0},$$

где $\boldsymbol{g}_i \equiv \boldsymbol{x}_i \cdot \varepsilon_i$.

---

[1] Далее будут приведены примеры $\boldsymbol{w}_i$.

[2] Как было отмечено в параграфе 2.3, наше использование термина *«предопределенность»* может быть не совсем обычным для части профессионалов. Все, чего мы требуем, — это ортогональности текущего значения ошибки и текущих значений регрессоров. Мы не требуем ортогональности текущего значения ошибки и прошлых значений регрессоров.

Мы отмечали в параграфе 2.3, что если в состав регрессоров входит константа, то условие ортогональности эквивалентно условию $\mathrm{E}(\varepsilon_i) = 0$ и что регрессоры не коррелированы с ошибкой. Теперь условия ортогональности касаются инструментальных переменных. То есть если один из инструментов — константа, то предположение 3.3 эквивалентно тому, что $\mathrm{E}(\varepsilon_i) = 0$ и что непостоянные инструменты не коррелированы с ошибкой.

Рассмотренные в двух предыдущих параграфах примеры могут быть переписаны в этом общем виде. Например,

Пример 3.1 (пример Уоркинга с наблюдаемым сдвигом предложения): Рассмотрим модель рыночного равновесия (3.1.2a) и (3.1.2b'). Предположим, что оцениваемое уравнение — уравнение спроса. Пример можно привести к общему виду, полагая

$$y_i = q,\ L = 2,\ \boldsymbol{\delta} = \begin{bmatrix}\alpha_0\\ \alpha_1\end{bmatrix},\ \boldsymbol{z}_i = \begin{bmatrix}1\\ p_i\end{bmatrix},\ \varepsilon_i = u_i,\ K = 2,\ \boldsymbol{x}_i = \begin{bmatrix}1\\ x_i\end{bmatrix}$$

и $\boldsymbol{w}_i = (q, p_i, x_i)'$. Поскольку математическое ожидание ошибки равно нулю, константа ортогональна ошибке и может быть включена в $\boldsymbol{x}_i$. Также получаем, что x_i, которое, по предположению, не коррелирует с ошибкой, удовлетворяет условию ортогональности $\mathrm{E}(x_i\varepsilon_i) = 0$. Поэтому оно тоже может быть включено в $\boldsymbol{x}_i$.

Другие рассмотренные примеры также могут быть записаны как частные случаи общей модели.

Два отдельных символа, $\boldsymbol{x}_i$ и $\boldsymbol{z}_i$, могут создать впечатление, что регрессоры и инструменты не содержат одних и тех же переменных, но это не всегда так. Более того, в приведенном примере регрессоры и инструменты содержат общую переменную — константу. Инструменты, являющиеся также и регрессорами, называют **предопределенными регрессорами** (*predetermined regressors*), а остальные регрессоры, которые не включены в x_i, — **эндогенными регрессорами** (*endogenous regressors*). Хорошим примером вышесказанного является

Пример 3.2 (уравнение заработной платы): Упрощенная версия уравнения заработной платы, которое будет оцениваться позднее в этой главе,

$$LW_i = \delta_i + \delta_2 S_i + \delta_3 EXPR_i + \delta_4 IQ_i + \varepsilon_i,$$

где LW_i — логарифм заработной платы индивида i, S_i — количество законченных лет обучения, $EXPR_i$ — количество лет стажа, IQ_i — уровень IQ, ε_i — ненаблюдаемые характеристики индивида, влияющие на уровень заработной платы. Мы предполагаем, что $\mathrm{E}(\varepsilon_i) = 0$ (в противном случае включим математическое ожидание ε_i в δ_0). В одной из спецификаций, которую мы оценим позднее, предположим, что S_i является предопределенной переменной, но IQ_i является эндогенной переменной из-за ошибок измерения. Мы также предполагаем, что $EXPR_i$, AGE_i (возраст индивида) и MED_i (продолжительность образования матери в годах) предопределены. Возраст исключен из уравнения заработной платы, отражая лежащее в основе предположение о том, что при контроле стажа возраст не будет влиять на заработную плату. В терминах общей модели,

$$y_i = LW_i,\ L = 4,\ \boldsymbol{z}_i = \begin{bmatrix} 1 \\ S_i \\ EXPR_i \\ IQ_i \end{bmatrix},\ K = 5,\ \boldsymbol{x}_i = \begin{bmatrix} 1 \\ S_i \\ EXPR_i \\ AGE_i \\ MED_i \end{bmatrix}$$

и $\boldsymbol{w}_i = (LW_i, S_i, EXPR_i, IQ_i, AGE_i, MED_i)'$. Как и в примере 3.1, мы можем включить константу в $\boldsymbol{x}_i$, так как $\mathrm{E}(\varepsilon_i) = 0$. В этом примере $\boldsymbol{x}_i$ и $\boldsymbol{z}_i$ включают три общие переменные $(1, S_i, EXPR_i)$. Если бы, например, S_i не была бы включена в $\boldsymbol{x}_i$, ее можно было бы рассматривать как эндогенную.

Если, как в этом примере, некоторые регрессоры $\boldsymbol{z}_i$ являются предопределенными, их следует включать в число инструментов $\boldsymbol{x}_i$. GMM-оценивание вектора параметров использует информацию, предоставленную условиями ортогональности. Невключение предопределенных регрессоров в число инструментов отбрасывает условия ортогональности, которые могли бы быть использованы.

Идентификация

Как видно из примеров предыдущих двух параграфов, инструмент должен быть не только предопределенным (ортогональным ошибке), но также коррелированным с регрессорами. В противном случае IV-оценка не может быть определена (см., например, (3.1.11)). Обобщением на случай более одного регрессора и более одной предопределенной переменной является

Предположение 3.4 (ранговое условие идентификации): $\mathrm{E}(\boldsymbol{x}_i\boldsymbol{z}_i')$, *матрица размера $K \times L$, имеет полный столбцовый ранг (то есть ее ранг равен L, числу столбцов). Мы обозначаем эту матрицу как $\boldsymbol{\Sigma}_{\boldsymbol{xz}}$*[1].

Чтобы показать, что это действительно обобщение, рассмотрим пример 3.1. Если $\boldsymbol{z}_i = (1, p_i)'$, $\boldsymbol{x}_i = (1, x_i)'$, то матрица $\boldsymbol{\Sigma}_{\boldsymbol{xz}}$ равна:

$$\boldsymbol{\Sigma}_{\boldsymbol{xz}} = \begin{bmatrix} 1 & \mathrm{E}(p_i) \\ \mathrm{E}(x_i) & \mathrm{E}(x_i p_i). \end{bmatrix}$$

[1] То есть существует и конечен смешанный момент $\mathrm{E}(\boldsymbol{x}_i\boldsymbol{z}_i')$. Если момент указывается, как здесь, то косвенно предполагается, что этот момент существует и конечен.

Определитель этой матрицы отличен от нуля (матрица имеет полный столбцовый ранг) тогда и только тогда, когда $\mathrm{Cov}(x_i, p_i) = \mathrm{E}(x_i p_i) - E(x_i)E(p_i) \neq 0$.

Предположение 3.4 называется **ранговым условием идентификации** (*rank condition for identification*) по следующей причине. Чтобы подчеркнуть зависимость $\boldsymbol{g}_i$ ($\equiv \boldsymbol{x}_i \varepsilon_i$) от данных и вектора параметров, перепишем его как

$$\boldsymbol{g}_i = \boldsymbol{g}(\boldsymbol{w}_i; \boldsymbol{\delta}) \equiv \boldsymbol{x}_i \cdot (y_i - \boldsymbol{z}_i' \boldsymbol{\delta}). \tag{3.3.1}$$

Тогда условие ортогональности можно переписать в виде:

$$\mathrm{E}[\boldsymbol{g}(\boldsymbol{w}_i; \boldsymbol{\delta})] = \underset{(K \times 1)}{\mathbf{0}}. \tag{3.3.2}$$

Пусть $\widetilde{\boldsymbol{\delta}}$ $(L \times 1)$ — гипотетическое значение $\boldsymbol{\delta}$. Рассмотрим систему K одновременных уравнений от L неизвестных (L элементов $\widetilde{\boldsymbol{\delta}}$):

$$\mathrm{E}[\boldsymbol{g}(\boldsymbol{w}_i; \widetilde{\boldsymbol{\delta}})] = \underset{(K \times 1)}{\mathbf{0}}. \tag{3.3.3}$$

Условие ортогональности (3.3.2) означает, что истинное значение вектора коэффициентов $\boldsymbol{\delta}$ является решением этой системы K одновременных уравнений (3.3.3). Поэтому сделанные нами предположения гарантируют, что решение системы одновременных уравнений существует. Мы говорим, что вектор коэффициентов (или уравнение) **идентифицируем** (*identified*), если $\widetilde{\boldsymbol{\delta}} = \boldsymbol{\delta}$ — единственное решение.

Поскольку оцениваемое уравнение в нашей модели линейно, то функция $\boldsymbol{g}(\boldsymbol{w}_i; \widetilde{\boldsymbol{\delta}})$ линейна по $\widetilde{\boldsymbol{\delta}}$, так как она может быть записана в виде $\boldsymbol{x}_i \cdot y_i - \boldsymbol{x}_i \boldsymbol{z}_i' \widetilde{\boldsymbol{\delta}}$. Поэтому (3.3.3) — это система из K линейных уравнений:

$$\mathrm{E}(\boldsymbol{x}_i \cdot y_i) - \mathrm{E}(\boldsymbol{x}_i \boldsymbol{z}_i')\widetilde{\boldsymbol{\delta}} = \mathbf{0} \quad \text{или} \quad \underset{(K \times L)}{\boldsymbol{\Sigma}_{\boldsymbol{xz}}} \underset{(L \times 1)}{\widetilde{\boldsymbol{\delta}}} = \underset{(K \times 1)}{\boldsymbol{\sigma}_{\boldsymbol{xy}}}, \tag{3.3.4}$$

где

$$\boldsymbol{\sigma}_{\boldsymbol{xy}} \equiv \mathrm{E}(\boldsymbol{x}_i \cdot y_i), \quad \boldsymbol{\Sigma}_{\boldsymbol{xz}} \equiv \mathrm{E}(\boldsymbol{x}_i \boldsymbol{z}_i').$$

Необходимое и достаточное условие того, что $\widetilde{\boldsymbol{\delta}} = \boldsymbol{\delta}$ является единственным решением этой системы, можно вывести из следующего результата матричной алгебры[1].

Предположим, что существует решение системы линейных одновременных уравнений относительно $\boldsymbol{x}$: $\boldsymbol{Ax} = \boldsymbol{b}$. Необходимым и достаточным условием того, что это решение единственно, является полный столбцовый ранг матрицы $\boldsymbol{A}$.

[1] См., например: [Searle, 1982, pp. 233–234].

Таким образом, $\tilde{\delta} = \delta$ — единственное решение (3.3.4) тогда и только тогда, когда матрица $\boldsymbol{\Sigma}_{\boldsymbol{xz}}$ имеет полный столбцовый ранг, о чем и говорит предположение 3.4.

Порядковое условие идентифицируемости

Поскольку $\mathrm{rank}(\boldsymbol{\Sigma}_{\boldsymbol{xz}}) < L$, если $K < L$, то *необходимым* условием идентифицируемости будет

$$K \ (= \#\text{предопределенных переменных}) \geq L \ (= \#\text{регрессоров}). \tag{3.3.5a}$$

Это называется **порядковым условием идентифицируемости** (*order condition for identification*). Его можно записать по-разному. Поскольку K является также числом условий ортогональности, а L — числом параметров, то порядковое условие можно записать эквивалентно как

$$\#\text{условий ортогональности} \geq \#\text{параметров}. \tag{3.3.5b}$$

Вычитая количество предопределенных регрессоров из обеих сторон неравенства, получим другое эквивалентное утверждение:

$$\begin{aligned}&\#\text{предопределенных переменных, не включенных в уравнение} \geq \\ &\qquad \geq \#\text{эндогенных регрессоров}.\end{aligned} \tag{3.3.5c}$$

В зависимости от выполнения порядкового условия мы говорим, что уравнение:

- **сверхидентифицируемо** (*overidentified*), если ранговое условие выполнено и $K > L$,
- **точно идентифицируемо** (*just identified*), если ранговое условие выполнено и $K = L$,
- **неидентифицируемо** (*underidentified*), если порядковое условие не выполнено (то есть $K < L$).

Поскольку порядковое условие является необходимым, его невыполнение означает, что уравнение неидентифицируемо.

Предположение асимптотической нормальности

Как и в главе 2, нам нужно усилить условия ортогональности, чтобы оценки были асимптотически нормальными.

Предположение 3.5 (g_i — мартингал-разность с конечными моментами второго порядка): *Пусть* $g_i \equiv \boldsymbol{x}_i \cdot \varepsilon_i$. $\{g_i\}$ *— мартингал-разность (так что* $\mathrm{E}(\boldsymbol{g}_i) = 0$*). Матрица смешанных моментов размера* $K \times K$, $\mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_i')$, *невырождена. Используем обозначение* $\boldsymbol{S}$ *для* $\mathrm{Avar}(\overline{\boldsymbol{g}})$ *(то есть дисперсии предельного распределения* $\sqrt{n}\,\overline{\boldsymbol{g}}$, *где* $\overline{\boldsymbol{g}} \equiv \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{g}_i$*). По предположению 3.2 и ЦПТ для стационарных эргодических мартингал-разностей,* $\boldsymbol{S} = \mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_i')$.

Это аналог предположения 2.5, к которому применимы те же примечания:

- Если инструменты включают в себя константу, то из этого предположения вытекает, что ошибка является мартингал-разностью (и *тем самым*, сериально некоррелирована).

- Достаточным и, вероятно, более простым для понимания условием для предположения 3.5 является

$$\mathrm{E}(\varepsilon_i|\varepsilon_{i-1},\ldots,\varepsilon_1,\boldsymbol{x}_i,\boldsymbol{x}_{i-1},\ldots,\boldsymbol{x}_1) = 0, \tag{3.3.6}$$

  означающее, что ошибка не просто является мартингал-разностью, но также ортогональна не только текущему, но и прошлым значениям инструментов.

- Поскольку $\boldsymbol{g}_i\boldsymbol{g}_i' = \varepsilon_i^2\boldsymbol{x}_i\boldsymbol{x}_i'$, $\boldsymbol{S}$ является матрицей четвертых моментов. Состоятельное оценивание $\boldsymbol{S}$ потребует условия на моменты четвертого порядка, которое сформулировано ниже в предположении 3.6.

- Если $\{\boldsymbol{g}_i\}$ сериально коррелирована, то $\boldsymbol{S}$ (определенная как $\mathrm{Avar}(\overline{\boldsymbol{g}})$) не равна $\mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_i')$ и имеет более сложную форму, что будет показано в главе 6.

Контрольные вопросы

1. Является ли уравнение спроса в примере Уоркинга (пример 3.1) идентифицируемым? Сверхидентифицируемым? Идентифицируемо ли уравнение предложения?

2. Пусть ранговое условие удовлетворяется для уравнения заработной платы в примере 3.2. Является ли это уравнение сверхидентифицируемым?

3. В примере с производственной функцией не определено никаких инструментов помимо константы. Поэтому $K = 1$ и $L = 2$. Порядковое условие говорит, что уравнение неидентифицируемо. Запишите условие ортогональности и удостоверьтесь, что существует бесконечно много пар (ϕ_0, ϕ_1), удовлетворяющих условию ортогональности.

4. Проверьте, что примеры из параграфа 3.2 являются частными случаями общей модели этого параграфа, определив $(y_i, \boldsymbol{x}_i, \boldsymbol{z}_i)$ и записав ранговое условие для каждого примера.

5. Покажите, что из указанной модели следует, что

$$\operatorname{rank}(\underset{(K\times L)}{\boldsymbol{\Sigma}_{\boldsymbol{xz}}}) = \operatorname{rank}(\underset{(K\times L)}{\boldsymbol{\Sigma}_{\boldsymbol{xz}}} \vdots \underset{(K\times 1)}{\sigma_{\boldsymbol{xy}}}).$$

Указание: $\widetilde{\delta} = \delta$ — *решение* (3.3.4). *(Некоторые учебники добавляют это уравнение в ранговое условие, однако оно вытекает из других предположений модели.)*

6. (Нерелевантные инструменты.) Добавим к x_i еще одну переменную, назовем ее ξ_i. Хотя она и предопределена, она не связана с регрессорами в том смысле, что $\mathrm{E}(\xi_i z_{il}) = 0$ для всех $l (= 1, 2, \ldots, L)$. Продолжает ли выполняться ранговое условие? **Указание:** *Если матрица размером $K \times L$ имеет полный столбцовый ранг, то добавление любой L-мерной строки к строкам матрицы не приводит к изменению ранга.*

7. (Линейно зависимые инструменты.) В примере 3.2 предположим, что $AGE_i = EXPR_i + S_i$ для всех индивидов. Означает ли это обязательно невыполнение рангового условия? [Ответ: Нет.] Выполняется ли условие полного ранга в предположении 3.5 (то, что матрица $\mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_i')$ невырождена)? **Указание:** *Существует K-мерный вектор $\boldsymbol{\alpha} \neq \mathbf{0}$ такой, что $\boldsymbol{\alpha}'\boldsymbol{x}_i = 0$. Покажите, что $\boldsymbol{\alpha}'\,\mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_i') = \mathbf{0}$.*

8. (Линейная комбинация инструментов.) Пусть $\boldsymbol{A}$ — матрица размера $q \times K$, имеющая полный строковый ранг (то есть $q \leq K$), такая, что $\boldsymbol{A}\boldsymbol{\Sigma}_{\boldsymbol{xz}}$ имеет полный столбцовый ранг. Пусть $\widehat{\boldsymbol{x}}_i \equiv \boldsymbol{A}\boldsymbol{x}_i$ (то есть $\widehat{\boldsymbol{x}}_i$ — это вектор из q трансформированных инструментов). Проверьте, что предположения 3.3–3.5 выполняются для $\widehat{\boldsymbol{x}}_i$, если они выполняются для $\boldsymbol{x}_i$.

9. Проверьте, что модель, построенная на предположениях 3.1–3.5, редуцируется к модели из главы 2, если $\boldsymbol{z}_i = \boldsymbol{x}_i$.

3.4. Обобщенный метод моментов (GMM)

Условия ортогональности утверждают, что все моменты из некоторого набора моментов генеральной совокупности равны нулю. Основной принцип **«метода моментов»** (*method of moments*) заключается в том, чтобы найти такую оценку параметра, при которой соответствующие выборочные моменты также были бы равны нулю. Моменты генеральной совокупности в условиях ортогональности — это $\mathrm{E}[\boldsymbol{g}(\boldsymbol{w}_i; \boldsymbol{\delta})]$. Их выборочные аналоги — это выборочные средние $\boldsymbol{g}(\boldsymbol{w}_i; \boldsymbol{\delta})$ (где $\boldsymbol{g}(\boldsymbol{w}_i; \boldsymbol{\delta})$ определено в (3.3.1)), вычисленные при некотором гипотетическом значении $\widetilde{\boldsymbol{\delta}}$:

$$\underset{(K\times 1)}{\boldsymbol{g}_n(\widetilde{\boldsymbol{\delta}})} \equiv \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{g}(\boldsymbol{w}_i; \widetilde{\boldsymbol{\delta}}). \tag{3.4.1}$$

Применение метода моментов к нашей модели сводится к выбору $\widetilde{\boldsymbol{\delta}}$, являющейся решением системы из K одновременных уравнений с L неизвестными, $\boldsymbol{g}_n(\widetilde{\boldsymbol{\delta}}) = \mathbf{0}$, которая является выборочным аналогом (3.3.3). Поскольку оцениваемое уравнение линейно, $\boldsymbol{g}_n(\widetilde{\boldsymbol{\delta}})$ можно переписать как

$$\begin{aligned}\boldsymbol{g}_n(\widetilde{\boldsymbol{\delta}}) &= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\cdot(y_i - \boldsymbol{z}_i'\widetilde{\boldsymbol{\delta}}) = \quad \text{(из выражения (3.3.1) для } \boldsymbol{g}(\boldsymbol{w}_i;\boldsymbol{\delta})\text{)}\\ &= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\cdot y_i - \Big(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{z}_i'\Big)\widetilde{\boldsymbol{\delta}} \equiv \boldsymbol{s}_{\boldsymbol{xy}} - \boldsymbol{S}_{\boldsymbol{xz}}\widetilde{\boldsymbol{\delta}},\end{aligned} \tag{3.4.2}$$

где $\boldsymbol{s}_{\boldsymbol{xy}}$ и $\boldsymbol{S}_{\boldsymbol{xz}}$ являются выборочными аналогами $\boldsymbol{\sigma}_{\boldsymbol{xy}}$ и $\boldsymbol{\Sigma}_{\boldsymbol{xz}}$:

$$\underset{(K\times 1)}{\boldsymbol{s}_{\boldsymbol{xy}}} \equiv \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\cdot y_i \quad \text{и} \quad \underset{(K\times L)}{\boldsymbol{S}_{\boldsymbol{xz}}} \equiv \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{z}_i'.$$

Таким образом, выборочный аналог $\boldsymbol{g}_n(\widetilde{\boldsymbol{\delta}}) = \mathbf{0}$ — это система K уравнений с L неизвестными:

$$\boldsymbol{S}_{\boldsymbol{xz}}\,\widetilde{\boldsymbol{\delta}} = \boldsymbol{s}_{\boldsymbol{xy}}. \tag{3.4.3}$$

Это выборочный аналог (3.3.4). Если условий ортогональности больше, чем количество параметров, система может не иметь решения. Обобщение метода моментов на такие случаи называется **обобщенным методом моментов (GMM)** (*generalized method of moments*).

Метод моментов

Если уравнение идентифицируемо точно, то тогда $K = L$ и $\boldsymbol{\Sigma}_{\boldsymbol{xz}}$ является квадратной и обратимой. Поскольку, по предположению 3.2, $\boldsymbol{S}_{\boldsymbol{xz}}$ сходится к $\boldsymbol{\Sigma}_{\boldsymbol{xz}}$ почти наверное, $\boldsymbol{S}_{\boldsymbol{xz}}$ обратима для достаточно большого объема выборки с вероятностью 1. Таким образом, при достаточно большом объеме выборки система одновременных уравнений (3.4.3) имеет единственное решение:

$$\widehat{\boldsymbol{\delta}}_{\mathrm{IV}} = \boldsymbol{S}_{\boldsymbol{xz}}^{-1}\,\boldsymbol{s}_{\boldsymbol{xy}} = \Big(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{z}_i'\Big)^{-1}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\cdot y_i. \tag{3.4.4}$$

Эта оценка также называется **оценкой с инструментальными переменными** (*instrumental variables estimator*) с $\boldsymbol{x}_t$ в качестве инструментов. Поскольку она определена в точно идентифицируемом случае, формула предполагает, что количество инструментов равно количеству регрессоров. Более того, если $\boldsymbol{z}_t = \boldsymbol{x}_t$ (все регрессоры предопределены или ортогональны ошибке регрессии), то $\widehat{\boldsymbol{\delta}}_{\mathrm{IV}}$ сводится к МНК-оценке. Таким образом, МНК-оценка является оценкой по методу моментов.

Обобщенный метод моментов

Если уравнение сверхидентифицируемо, то есть $K > L$, мы не можем в общем случае подобрать L-мерный вектор $\widetilde{\boldsymbol{\delta}}$, удовлетворяющий K уравнениям (3.4.3). Если мы не можем сделать $\boldsymbol{g}_n(\widetilde{\boldsymbol{\delta}})$ в точности равным $\mathbf{0}$, то по крайней мере мы можем подобрать $\widetilde{\delta}$ так, что $\boldsymbol{g}_n(\widetilde{\boldsymbol{\delta}})$ будет максимально близким к $\mathbf{0}$. Чтобы уточнить, что мы понимаем здесь под «близостью», определим расстояние между любыми двумя K-мерными векторами ξ и η в виде квадратичной формы $(\xi - \eta)'\widehat{\boldsymbol{W}}(\xi - \eta)$, где $\widehat{\boldsymbol{W}}$, иногда называемая **взвешивающей матрицей** (*weighting matrix*), симметричная и положительно определенная матрица, определяющая расстояние[1].

Определение 3.1 (GMM-оценка): Пусть $\widehat{\boldsymbol{W}}$ — симметричная положительно определенная матрица размера $K \times K$, возможно, зависящая от выборки, такая, что $\widehat{\boldsymbol{W}} \to_{\mathrm{p}} \boldsymbol{W}$ при стремлении объема выборки n к бесконечности, где $\boldsymbol{W}$ — симметричная и положительно определенная матрица. **«GMM-оценка»** (*GMM estimators*) $\boldsymbol{\delta}$, обозначаемая $\widehat{\delta}(\widehat{\boldsymbol{W}})$, — это

$$\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}) \equiv \underset{\widetilde{\delta}}{\operatorname{argmin}}\, J(\widetilde{\boldsymbol{\delta}}, \widehat{\boldsymbol{W}}), \tag{3.4.5}$$

где

$$J(\widetilde{\boldsymbol{\delta}}, \widehat{\boldsymbol{W}}) \equiv n \cdot \boldsymbol{g}_n(\widetilde{\boldsymbol{\delta}})'\widehat{\boldsymbol{W}}\boldsymbol{g}_n(\widetilde{\boldsymbol{\delta}}).$$

(Причина, по которой расстояние $\boldsymbol{g}_n(\widetilde{\boldsymbol{\delta}})'\widehat{\boldsymbol{W}}\boldsymbol{g}_n(\widetilde{\boldsymbol{\delta}})$ умножается на объем выборки n, будет объяснена в параграфе 3.6.) Взвешивающая матрица может быть случайной и зависящей от выборки, чтобы учесть возможность оценивания этой матрицы по выборке. Определение показывает, что GMM является частным случаем **«метода минимального расстояния»** (*minimum distance estimation*). В методе оценки минимального расстояния $\operatorname{plim} \boldsymbol{g}_n(\boldsymbol{\delta}) = 0$, как и в GMM, но функция $\boldsymbol{g}_n(\cdot)$ не обязана быть выборочным средним.

Поскольку $\boldsymbol{g}_n(\widetilde{\boldsymbol{\delta}})$ линейна по $\widetilde{\delta}$ (см. (3.4.2)), целевая функция квадратична по $\widetilde{\delta}$, когда уравнение линейно:

$$J(\widetilde{\boldsymbol{\delta}}, \widehat{\boldsymbol{W}}) = n \cdot (\boldsymbol{s}_{\boldsymbol{xy}} - \boldsymbol{S}_{\boldsymbol{xz}}\widetilde{\boldsymbol{\delta}})'\widehat{\boldsymbol{W}}(\boldsymbol{s}_{\boldsymbol{xy}} - \boldsymbol{S}_{\boldsymbol{xz}}\widetilde{\boldsymbol{\delta}}). \tag{3.4.6}$$

Вы можете сами показать, что условием первого порядка для минимизации этой целевой функции относительно $\widetilde{\delta}$ является

$$\underset{(L\times K)}{\boldsymbol{S}'_{\boldsymbol{xz}}}\ \underset{(K\times K)}{\widehat{\boldsymbol{W}}}\ \underset{(K\times 1)}{\boldsymbol{s}_{\boldsymbol{xy}}} = \underset{(L\times K)}{\boldsymbol{S}'_{\boldsymbol{xz}}}\ \underset{(K\times K)}{\widehat{\boldsymbol{W}}}\ \underset{(K\times L)}{\boldsymbol{S}_{\boldsymbol{xz}}}\ \underset{(L\times 1)}{\widetilde{\boldsymbol{\delta}}}. \tag{3.4.7}$$

[1] Не следует путать «взвешивание» в GMM и во взвешенном методе наименьших квадратов (WLS). В GMM взвешивание применяется к выборочному среднему $\widetilde{\boldsymbol{g}}$, в то время как в WLS оно применяется к каждому наблюдению.

При выполнении предположений 3.2 и 3.4 матрица $\boldsymbol{S}_{\boldsymbol{xz}}$ имеет полный столбцовый ранг для достаточно больших n с вероятностью 1. Поскольку $\widehat{\boldsymbol{W}}$ положительно определена, матрица $\boldsymbol{S}'_{\boldsymbol{xz}}\widehat{\boldsymbol{W}}\boldsymbol{S}_{\boldsymbol{xz}}$ размера $L \times L$ невырождена. Поэтому единственное решение можно получить, умножив обе части на матрицу, обратную к $\boldsymbol{S}'_{\boldsymbol{xz}}\widehat{\boldsymbol{W}}\boldsymbol{S}_{\boldsymbol{xz}}$. Это единственное решение является GMM-оценкой:

$$\text{GMM-оценка:} \quad \widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}) = (\boldsymbol{S}'_{\boldsymbol{xz}}\widehat{\boldsymbol{W}}\boldsymbol{S}_{\boldsymbol{xz}})^{-1}\boldsymbol{S}'_{\boldsymbol{xz}}\widehat{\boldsymbol{W}}\boldsymbol{s}_{\boldsymbol{xy}}. \tag{3.4.8}$$

Если $K = L$, то $\boldsymbol{S}_{\boldsymbol{xz}}$ является квадратной матрицей и (3.4.8) сводится к IV-оценке (3.4.4). Тем самым GMM является обобщением метода моментов.

Ошибка оценки

Для дальнейшего использования выведем выражение для ошибки оценки. Умножив обе стороны уравнения $y_i = \boldsymbol{z}_i'\boldsymbol{\delta} + \varepsilon_i$ слева на $\boldsymbol{x}_i$ и взяв средние, получим:

$$\boldsymbol{s}_{\boldsymbol{xy}} = \boldsymbol{S}_{\boldsymbol{xz}}\boldsymbol{\delta} + \overline{\boldsymbol{g}}, \tag{3.4.9}$$

где

$$\overline{\boldsymbol{g}} \equiv \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\cdot\varepsilon_i = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{g}(\boldsymbol{w}_i;\boldsymbol{\delta}) = \boldsymbol{g}_n(\boldsymbol{\delta}). \tag{3.4.10}$$

Подставляя (3.4.9) в (3.4.8), получаем:

$$\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}) - \boldsymbol{\delta} = (\boldsymbol{S}'_{\boldsymbol{xz}}\widehat{\boldsymbol{W}}\boldsymbol{S}_{\boldsymbol{xz}})^{-1}\boldsymbol{S}'_{\boldsymbol{xz}}\widehat{\boldsymbol{W}}\overline{\boldsymbol{g}}. \tag{3.4.11}$$

Контрольные вопросы

1. Проверьте, что (3.1.12) это IV-оценка α_1 при применении метода моментов к уравнению спроса (3.1.2а) с $(1, x_i)$ в качестве инструментов.
2. Если уравнение идентифицируемо точно, то каково минимальное значение $J(\widetilde{\boldsymbol{\delta}}, \widehat{\boldsymbol{W}})$?
3. Даже если уравнение сверхидентифицируемо, теоретический вариант системы K уравнений (3.3.4) имеет решение. Почему не имеет решения выборочный вариант системы (3.4.3)? **Указание:** *Теоретический вариант имеет решение, так как матрица $[\boldsymbol{\Sigma}_{\boldsymbol{xz}} \vdots \boldsymbol{\sigma}_{\boldsymbol{xy}}]$ имеет ранг L. Ранговое условие представляет собой набор равенств относительно элементов $\boldsymbol{\Sigma}_{\boldsymbol{xz}}$ и $\boldsymbol{\sigma}_{\boldsymbol{xy}}$. Даже если $\boldsymbol{S}_{\boldsymbol{xz}}$ и $\boldsymbol{s}_{\boldsymbol{xy}}$ сходятся к $\boldsymbol{\Sigma}_{\boldsymbol{xz}}$ и $\boldsymbol{\sigma}_{\boldsymbol{xy}}$, это не означает, что ранг $[\boldsymbol{S}_{\boldsymbol{xz}} \vdots \boldsymbol{s}_{\boldsymbol{xy}}]$ равен L для достаточно большого n. Напротив, $\boldsymbol{S}_{\boldsymbol{xz}}$ имеет полный столбцовый ранг для достаточно большого n, если ее предел, $\boldsymbol{\Sigma}_{\boldsymbol{xz}}$, имеет полный столбцовый ранг. Это связано с тем, что ранговое условие для матрицы является набором неравенств для элементов этой матрицы.*

4. Что не так со следующим утверждением?

   Даже если уравнение сверхидентифицируемо, найти решение (3.4.3) можно без проблем. Следует умножить обе стороны слева на $\boldsymbol{S}'_{\boldsymbol{xz}}$ и получить

$$\boldsymbol{S}'_{\boldsymbol{xz}}\boldsymbol{S}_{\boldsymbol{xz}}\,\tilde{\boldsymbol{\delta}} = \boldsymbol{S}'_{\boldsymbol{xz}}\boldsymbol{s}_{\boldsymbol{xy}}. \tag{3.4.12}$$

   Поскольку $\boldsymbol{S}_{\boldsymbol{xz}}$ имеет полный столбцовый ранг, $\boldsymbol{S}'_{\boldsymbol{xz}}\boldsymbol{S}_{\boldsymbol{xz}}$ обратима. Поэтому решением является

$$\tilde{\boldsymbol{\delta}} = (\boldsymbol{S}'_{\boldsymbol{xz}}\boldsymbol{S}_{\boldsymbol{xz}})^{-1}\boldsymbol{S}'_{\boldsymbol{xz}}\boldsymbol{s}_{\boldsymbol{xy}}.$$

   Указание: *Эта* $\tilde{\boldsymbol{\delta}}$ *определенно является решением* (3.4.12). *Но является ли она решением* (3.4.3)?

5. (Вырожденная $\boldsymbol{W}$.) Проверьте, что GMM-оценка (3.4.8) определена для достаточно большого n, даже если $\boldsymbol{W}$ ($\equiv \operatorname{plim}\widehat{\boldsymbol{W}}$) вырождена, если не вырождена $\boldsymbol{\Sigma}'_{\boldsymbol{xz}}\boldsymbol{W}\boldsymbol{\Sigma}_{\boldsymbol{xz}}$.

3.5. Асимптотические свойства GMM

Формула (3.4.8) определяет GMM-оценки, которые образуют множество оценок, индексируемых взвешивающей матрицей $\widehat{\boldsymbol{W}}$. Следует отметить, что все оценки, представленные в следующих главах, являются GMM-оценками для некоторых $\widehat{\boldsymbol{W}}$. Задачей данного параграфа является разработка асимптотической теории для GMM-оценок для любого заданного выбора $\widehat{\boldsymbol{W}}$, которая может быть проведена достаточно легко с использованием методов предыдущей главы. Первая половина данного параграфа распространяет утверждения 2.1–2.4 из главы 2 на GMM-оценки, которые предпочтительнее других GMM-оценок. Вопрос, который не возникал в этих утверждениях, — это вопрос выбора подходящей GMM-оценки, которая предпочтительнее других GMM-оценок. Это вопрос оптимального выбора $\widehat{\boldsymbol{W}}$, и он будет рассмотрен во второй половине данного параграфа.

Асимптотическое распределение GMM-оценки

Асимптотическая теория для $\widehat{\delta}(\widehat{\boldsymbol{W}})$, справедливая для любого выбора $\widehat{\boldsymbol{W}}$, такова:

Утверждение 3.1 (асимптотическое распределение оценок GMM):

(a) *(Состоятельность.) При предположениях 3.1–3.4,* $\operatorname*{plim}_{n\to\infty}\widehat{\delta}(\widehat{\boldsymbol{W}}) = \boldsymbol{\delta}$.

(b) *(Асимптотическая нормальность.) Если предположение 3.3 усилено, как утверждение 3.5, то*

$$\sqrt{n}(\widehat{\delta}(\widehat{\boldsymbol{W}}) - \boldsymbol{\delta}) \underset{d}{\to} N\big(\mathbf{0}, \operatorname{Avar}(\widehat{\delta}(\widehat{\boldsymbol{W}}))\big) \quad \textit{при } n \to \infty,$$

где

$$\mathrm{Avar}(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}})) = (\boldsymbol{\Sigma}'_{\boldsymbol{xz}}\boldsymbol{W}\boldsymbol{\Sigma}_{\boldsymbol{xz}})^{-1}\boldsymbol{\Sigma}'_{\boldsymbol{xz}}\boldsymbol{W}\boldsymbol{S}\boldsymbol{W}\boldsymbol{\Sigma}_{\boldsymbol{xz}}(\boldsymbol{\Sigma}'_{\boldsymbol{xz}}\boldsymbol{W}\boldsymbol{\Sigma}_{\boldsymbol{xz}})^{-1}. \tag{3.5.1}$$

(Напомним: $\boldsymbol{\Sigma}_{\boldsymbol{xz}} \equiv \mathrm{E}(\boldsymbol{x}_i\boldsymbol{z}_i')$, $\boldsymbol{S} = \mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_i') = \mathrm{E}(\varepsilon_i^2\boldsymbol{x}_i\boldsymbol{x}_i')$, $\boldsymbol{W} \equiv \mathrm{plim}\,\widehat{\boldsymbol{W}}$.*)*

(c) *(Состоятельная оценка* $\mathrm{Avar}(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}))$.*) Предположим, что доступна состоятельная оценка* $\boldsymbol{S}$ $(K \times K)$, $\widehat{\boldsymbol{S}}$. *Тогда, при предположении 3.2, состоятельной оценкой* $\mathrm{Avar}(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}))$ *является*

$$\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}))} \equiv (\boldsymbol{S}'_{\boldsymbol{xz}}\widehat{\boldsymbol{W}}\boldsymbol{S}_{\boldsymbol{xz}})^{-1}\boldsymbol{S}'_{\boldsymbol{xz}}\widehat{\boldsymbol{W}}\widehat{\boldsymbol{S}}\widehat{\boldsymbol{W}}\boldsymbol{S}_{\boldsymbol{xz}}(\boldsymbol{S}'_{\boldsymbol{xz}}\widehat{\boldsymbol{W}}\boldsymbol{S}_{\boldsymbol{xz}})^{-1}, \tag{3.5.2}$$

где $\boldsymbol{S}_{\boldsymbol{xz}}$ *— выборочное среднее* $\boldsymbol{x}_i\boldsymbol{z}_i'$:

$$\underset{(K\times L)}{\boldsymbol{S}_{\boldsymbol{xz}}} \equiv \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{z}_i'.$$

Не самое красивое выражение для асимптотической дисперсии станет гораздо приятнее при оптимальном выборе взвешивающей матрицы. Доказательство утверждения 2.1 поможет доказать утверждение 3.1. Ключевыми моментами являются:

1. $\boldsymbol{S}_{\boldsymbol{xz}} \to_p \boldsymbol{\Sigma}_{\boldsymbol{xz}}$ (согласно свойству эргодической стационарности).

2. $\overline{\boldsymbol{g}}(\equiv \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{g}_i) \to_p \boldsymbol{0}$ (согласно свойству эргодической стационарности и условиям ортогональности).

3. $\sqrt{n}\overline{\boldsymbol{g}} \to_p N(\boldsymbol{0}, \boldsymbol{S})$ (по предположению 3.5).

Состоятельность сразу получается в результате применения пунктов 1 и 2 и леммы 2.3(а) к выражению для ошибки оценки (3.4.11). Для доказательства асимптотической нормальности следует умножить обе части (3.4.11) на $\sqrt{n}$:

$$\sqrt{n}(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}) - \boldsymbol{\delta}) = (\boldsymbol{S}'_{\boldsymbol{xz}}\widehat{\boldsymbol{W}}\boldsymbol{S}_{\boldsymbol{xz}})^{-1}\boldsymbol{S}'_{\boldsymbol{xz}}\widehat{\boldsymbol{W}}\sqrt{n}\,\overline{\boldsymbol{g}} \tag{3.5.3}$$

и использовать пункт 3, леммы 2.3(а) и 2.4(с). Часть (с) утверждения 3.1 непосредственно вытекает из леммы 2.3(а).

Оценивание дисперсии ошибки

В параграфе 2.3 была доказана состоятельность МНК-оценки дисперсии ошибки. Было отмечено, что этот результат сохраняется, если остатки получены на основании состоятельной оценки вектора коэффициентов. То же самое справедливо и здесь.

Утверждение 3.2 (состоятельная оценка дисперсии ошибки): *Для любой состоятельной оценки $\boldsymbol{\delta}$, $\widehat{\boldsymbol{\delta}}$, определим $\widehat{\varepsilon}_i \equiv y_i - \boldsymbol{z}_i'\widehat{\boldsymbol{\delta}}$. При предположениях 3.1, 3.2 и дополнительном предположении, что $\mathrm{E}(\boldsymbol{z}_i\boldsymbol{z}_i')$ (вторые моменты регрессоров) существуют и конечны,*

$$\frac{1}{n}\sum_{i=1}^{n}\widehat{\varepsilon}_i^2 \underset{p}{\to} \mathrm{E}(\varepsilon_i^2),$$

если $\mathrm{E}(\varepsilon_i^2)$ существует и конечен.

Доказательство очень похоже на доказательство утверждения 2.2. Связь между $\widehat{\varepsilon}_i$ и ε_i дается формулой

$$\widehat{\varepsilon}_i \equiv y_i - \boldsymbol{z}_i'\widehat{\boldsymbol{\delta}} = \varepsilon_i - \boldsymbol{z}_i'(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}), \tag{3.5.4}$$

так что

$$\widehat{\varepsilon}_i^2 = \varepsilon_i^2 - 2(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta})'\boldsymbol{z}_i{\cdot}\varepsilon_i + (\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta})'\boldsymbol{z}_i\boldsymbol{z}_i'(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}). \tag{3.5.5}$$

Суммируя по i, получаем:

$$\frac{1}{n}\sum_{i=1}^{n}\widehat{\varepsilon}_i^2 = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2 - 2(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta})'\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i{\cdot}\varepsilon_i + (\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta})'\Big(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\boldsymbol{z}_i'\Big)(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}). \tag{3.5.6}$$

Как обычно, $\frac{1}{n}\sum_i\varepsilon_i^2 \to_p \mathrm{E}(\varepsilon_i^2)$. Поскольку $\widehat{\boldsymbol{\delta}}$ является состоятельной и $\frac{1}{n}\sum_i \boldsymbol{z}_i\boldsymbol{z}_i'$ сходится по вероятности к некоторой конечной матрице по предположению, последнее слагаемое стремится к нулю. Легко показать, что $\mathrm{E}(\boldsymbol{z}_i \cdot \varepsilon_i)$ существует и конечно[1]. Тогда, из свойства эргодической стационарности,

$$\frac{1}{n}\sum_i \boldsymbol{z}_i{\cdot}\varepsilon_i \to_p \text{некоторый конечный вектор.}$$

Таким образом, второе слагаемое в правой части (3.5.6) также стремится к нулю.

Тестирование гипотез

Из утверждений 3.1b и 3.1c очевидным образом выводится

Утверждение 3.3 (робастная t-статистика и статистика Вальда): *Пусть выполнены предположения 3.1–3.5, и пусть доступна состоятельная оценка $\widehat{\boldsymbol{S}}$ матрицы $\boldsymbol{S}$ $(\equiv \mathrm{Avar}(\overline{\boldsymbol{g}}) = \mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_i'))$. Пусть*

$$\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}))} \equiv (\boldsymbol{S}_{\boldsymbol{xz}}'\widehat{\boldsymbol{W}}\boldsymbol{S}_{\boldsymbol{xz}})^{-1}\boldsymbol{S}_{\boldsymbol{xz}}'\widehat{\boldsymbol{W}}\,\widehat{\boldsymbol{S}}\,\widehat{\boldsymbol{W}}\boldsymbol{S}_{\boldsymbol{xz}}(\boldsymbol{S}_{\boldsymbol{xz}}'\widehat{\boldsymbol{W}}\boldsymbol{S}_{\boldsymbol{xz}})^{-1}.$$

Тогда:

[1] Согласно неравенству Коши – Шварца, $\mathrm{E}(|z_{il}\varepsilon_i|) \leq \sqrt{\mathrm{E}(z_{il}^2)\,\mathrm{E}(\varepsilon_i^2)}$, где z_{il} — l-й элемент $\boldsymbol{z}_i$. $\mathrm{E}(z_{il}^2)$ и $\mathrm{E}(\varepsilon_i^2)$ конечны по предположению.

(a) *При нулевой гипотезе* $H_0 : \delta_\ell = \overline{\delta}_\ell$,

$$t_\ell \equiv \frac{\sqrt{n}(\widehat{\delta}_\ell(\widehat{\boldsymbol{W}}) - \overline{\delta}_\ell)}{\sqrt{\left(\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}))}\right)_{\ell\ell}}} = \frac{\widehat{\delta}_\ell(\widehat{\boldsymbol{W}}) - \overline{\delta}_\ell}{SE^*_\ell} \underset{d}{\rightarrow} N(0,1),$$

где $\left(\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}))}\right)_{\ell\ell}$ *является* (ℓ,ℓ) *элементом* $\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}))}$ *[являющимся* $\widehat{\mathrm{Avar}(\widehat{\delta}_\ell(\widehat{\boldsymbol{W}}))}$*] и*

$$SE^*_\ell \text{ (робастная стандартная ошибка)} \equiv \sqrt{\frac{1}{n}\cdot\left(\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}))}\right)_{\ell\ell}}. \tag{3.5.7}$$

(b) *При нулевой гипотезе* $H_0 : \boldsymbol{R\delta} = \boldsymbol{r}$, *где* $\#\boldsymbol{r}$ *— количество ограничений (размерность* $\boldsymbol{r}$*), и* $\boldsymbol{R}$ $(\#\boldsymbol{r} \times L)$ *имеет полный строковый ранг,*

$$W \equiv n\cdot(\boldsymbol{R}\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}) - \boldsymbol{r})'\left\{\boldsymbol{R}[\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}))}]\boldsymbol{R}'\right\}^{-1}(\boldsymbol{R}\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}) - \boldsymbol{r}) \underset{d}{\rightarrow} \chi^2(\#\boldsymbol{r}). \tag{3.5.8}$$

(c) *При нулевой гипотезе* $H_0 : \boldsymbol{a}(\boldsymbol{\delta}) = \boldsymbol{0}$ *такой, что* $\boldsymbol{A}(\boldsymbol{\delta})$ *— матрица первых производных* $\boldsymbol{a}(\boldsymbol{\delta})$ *размера* $\#\boldsymbol{a} \times L$ *(где* $\#\boldsymbol{a}$ *размерность* $\boldsymbol{a}$*) — непрерывна и имеет полный строковый ранг,*

$$W \equiv n\cdot\boldsymbol{a}(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}))'\left\{\boldsymbol{A}(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}))[\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}))}]\boldsymbol{A}(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}))'\right\}^{-1}\boldsymbol{a}(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}})) \underset{d}{\rightarrow}$$
$$\underset{d}{\rightarrow} \chi^2(\#\boldsymbol{a}). \tag{3.5.9}$$

Для статистики Вальда для нелинейных гипотез справедливы те же комментарии, что и для утверждения 2.6: численное значение статистики Вальда неинвариантно к представлению нелинейных ограничений.

Оценивание S

Мы уже рассматривали оценивание $\boldsymbol{S}$ $(= \mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_i') = \mathrm{E}(\varepsilon_i^2\boldsymbol{x}_i\boldsymbol{x}_i'))$ в параграфе 2.5. Предложенная оценка, с учетом оцениваемого теперь уравнения $y_i = \boldsymbol{z}_i'\boldsymbol{\delta} + \varepsilon_i$, принимает вид:

$$\widehat{\boldsymbol{S}} \equiv \frac{1}{n}\sum_{i=1}^{n}\widehat{\varepsilon}_i^2\,\boldsymbol{x}_i\boldsymbol{x}_i', \tag{3.5.10}$$

где $\widehat{\varepsilon}_i \equiv y_i - \boldsymbol{z}_i'\widehat{\boldsymbol{\delta}}$ и $\widehat{\boldsymbol{\delta}}$ — состоятельная оценка δ. Для состоятельности этой оценки нужно предположение о четвертых моментах, являющееся обобщением предположения 2.6.

Предположение 3.6 (конечные четвертые моменты): $\mathrm{E}[(x_{ik}z_{i\ell})^2]$ *существует и конечен для всех* k $(= 1, \dots, K)$ *и* ℓ $(= 1, \dots, L)$.

В качестве аналитического упражнения остается доказательство следующего результата:

Утверждение 3.4 (состоятельное оценивание S): *Предположим, что оценка* $\widehat{\delta}$, *используемая для вычисления остатка* $\widehat{\varepsilon}_i$ *для* $\widehat{S}$ *в* (3.5.10), *состоятельна, и предположим, что* $S = \mathrm{E}(g_i g_i')$ *существует и конечна. Тогда, с учетом предположений 3.1, 3.2 и 3.6,* $\widehat{S}$ *из выражения* (3.5.10) *является состоятельной оценкой для* S.

Эффективная GMM-оценка

Естественно, хотелось бы выбрать среди GMM-оценок, индексированных посредством $\widehat{W}$, такую оценку, которая имела бы наименьшую асимптотическую дисперсию. Следующее утверждение обеспечивает выбор W, минимизирующей асимптотическую дисперсию.

Утверждение 3.5 (оптимальный выбор взвешивающей матрицы): *Нижняя граница асимптотической дисперсии GMM-оценки* (3.4.8), *проиндексированной посредством* $\widehat{W}$, *задается выражением* $(\Sigma_{xz}' S^{-1} \Sigma_{xz})^{-1}$. *Нижняя граница достигается, если* $\widehat{W}$ *такая, что*[1] W $(\equiv \operatorname{plim} \widehat{W}) = S^{-1}$.

Поскольку асимптотическая дисперсия для любой заданной $\widehat{W}$ дается выражением (3.5.1), для доказательства этого утверждения нужно показать, что

$$(\Sigma_{xz}' W \Sigma_{xz})^{-1} \Sigma_{xz}' W S W \Sigma_{xz} (\Sigma_{xz}' W \Sigma_{xz})^{-1} \geq (\Sigma_{xz}' S^{-1} \Sigma_{xz})^{-1} \tag{3.5.11}$$

для любой симметричной положительно определенной матрицы W. Доказательство этого алгебраического результата остается в качестве упражнения.

GMM-оценка, удовлетворяющая условию эффективности $\operatorname{plim} \widehat{W} = S^{-1}$, будет называться **эффективной** (или **оптимальной**) **GMM-оценкой** (*efficient (or optimal) GMM estimator*). Простой подстановкой $\widehat{S}^{-1}$, являющейся состоятельной оценкой S^{-1}, вместо $\widehat{W}$ в формулах утвер-

[1] Условие $W = S^{-1}$ достаточное, но не необходимое для эффективности оценок. Необходимым и достаточным условием является существование матрицы C такой, что $\Sigma_{xz}' W = C \Sigma_{xz}' S^{-1}$. См.: [Newey and McFadden, 1994, p. 2165].

ждения 3.1 получаем:

Эффективная GMM-оценка: $\widehat{\delta}(\widehat{S}^{-1}) = (S'_{xz}\,\widehat{S}^{-1}S_{xz})^{-1}S'_{xz}\,\widehat{S}^{-1}s_{xy}$.
$$\tag{3.5.12}$$

$$\mathrm{Avar}(\widehat{\delta}(\widehat{S}^{-1})) = (\Sigma'_{xz}\,S^{-1}\Sigma_{xz})^{-1}, \tag{3.5.13}$$

$$\widehat{\mathrm{Avar}(\widehat{\delta}(\widehat{S}^{-1}))} = (S'_{xz}\,\widehat{S}^{-1}S_{xz})^{-1}. \tag{3.5.14}$$

Аналогичная подстановка следующим образом преобразует формулы из утверждения 3.3:

$$t_\ell = \frac{\widehat{\delta}_\ell(\widehat{S}^{-1}) - \overline{\delta}_\ell}{SE^*_\ell}, \tag{3.5.15}$$

где SE^*_ℓ — робастная стандартная ошибка, заданная выражением:

$$SE^*_\ell = \sqrt{\frac{1}{n}((S'_{xz}\widehat{S}^{-1}S_{xz})^{-1})_{\ell\ell}}$$

и

$$W = n\cdot a(\widehat{\delta}(\widehat{S}^{-1}))'\left\{A(\widehat{\delta}(\widehat{S}^{-1}))(S'_{xz}\,\widehat{S}^{-1}S_{xz})^{-1}A(\widehat{\delta}(\widehat{S}^{-1}))'\right\}^{-1}a(\widehat{\delta}(\widehat{S}^{-1})). \tag{3.5.16}$$

Для получения эффективной GMM-оценки необходима состоятельная оценка $\widehat{S}$. Но утверждение 3.4 гарантирует, что $\widehat{S}$, основанная на любой состоятельной оценке δ, состоятельна. Это приводит к следующей **двухступенчатой процедуре получения эффективных GMM-оценок** (*two-step efficient GMM procedure*).

Шаг 1: Выберите матрицу $\widehat{W}$, сходящуюся по вероятности к симметричной положительно определенной матрице, и минимизируйте $J(\widetilde{\delta}, \widehat{W})$ по $\widetilde{\delta}$ для получения $\widehat{\delta}(\widehat{W})$. В таких матрицах нет недостатка (например $\widehat{W} = I$), но обычно полагают $\widehat{W} = S_{xx}^{-1}$. Итоговая оценка $\widehat{\delta}(S_{xx}^{-1})$ является знаменитой оценкой двухшагового МНК (как мы увидим в параграфе 3.8). Используйте полученную оценку для вычисления остатков $\widehat{\varepsilon}_i \equiv y_i - z_i'\widehat{\delta}(\widehat{W})$ и получения состоятельной оценки $\widehat{S}$ для S по формуле (3.5.10).

Шаг 2: Минимизируйте $J(\widetilde{\delta}, \widehat{S}^{-1})$ по $\widetilde{\delta}$. Значение, доставляющее минимум, является эффективной GMM-оценкой.

Асимптотическая мощность

Подобно оценкам коэффициентов, t-статистика и статистика Вальда зависят от выбора W. Кажется интуитивно очевидным, что статистики, связанные с эффективной GMM-оценкой, предпочтительны на больших выборках. Это можно показать формально в терминах асимптотической мощности, введенной в параграфе 2.4. Возьмем для примера t-статистику для тестирования H_0: $\delta_\ell = \overline{\delta}_\ell$. Эта t-статистика записывается (см. (3.5.7)) как

$$t_\ell \equiv \frac{\sqrt{n}(\widehat{\delta}_\ell(\widehat{\boldsymbol{W}}) - \overline{\delta}_\ell)}{\sqrt{\left(\widehat{\operatorname{Avar}(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}))}\right)_{\ell\ell}}}. \tag{3.5.17}$$

Знаменатель сходится к некоторому конечному числу, даже когда эта нулевая гипотеза неверна. Напротив, предел по вероятности числителя, в случае, когда нулевая гипотеза неверна, бесконечен. Поэтому мощность при любой фиксированной альтернативе приближается к единице. Таким образом, тест состоятелен. Это справедливо для любого выбора $\boldsymbol{W}$, поэтому состоятельность теста не может служить критерием выбора $\boldsymbol{W}$.

Рассмотрим последовательность локальных альтернатив со сдвигом Питмена:

$$\delta_\ell^{(n)} = \overline{\delta}_\ell + \frac{\gamma}{\sqrt{n}} \tag{3.5.18}$$

для некоторого $\gamma \neq 0$. Подставляя (3.5.18) в (3.5.11), можно переписать t-статистику как

$$t_\ell = \frac{\sqrt{n}(\widehat{\delta}_\ell(\widehat{\boldsymbol{W}}) - \delta_\ell^{(n)})}{\sqrt{\left(\widehat{\operatorname{Avar}(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}))}\right)_{\ell\ell}}} + \frac{\gamma}{\sqrt{\left(\widehat{\operatorname{Avar}(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}))}\right)_{\ell\ell}}}. \tag{3.5.19}$$

Применяя те же соображения, что и для получения асимптотического распределения t-статистики (2.4.4), можно показать, что $t_\ell \to_{\mathrm{d}} N(\mu, 1)$, где

$$\mu \equiv \frac{\gamma}{\sqrt{\left(\widehat{\operatorname{Avar}(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}))}\right)_{\ell\ell}}}. \tag{3.5.20}$$

Если уровень значимости равен α, асимптотическая мощность задается выражением $\operatorname{Prob}(|x| > t_{\alpha/2})$, где $x \sim N(\mu, 1)$ и $t_{\alpha/2}$ — критическое значение уровня α. Очевидно, чем больше значение $|\mu|$, тем выше асимптотическая мощность. А $|\mu|$ уменьшается с ростом асимптотической дисперсии. Таким образом, асимптотическая мощность максимальна при эффективной GMM-оценке.

Свойства в малых выборках

Сохраняются ли эти желательные асимптотические свойства эффективных GMM-оценок и связанных с ними тестовых статистик в случае конечных выборок? Эффективная GMM-оценка использует $\widehat{\boldsymbol{S}}^{-1}$, функцию от оцененных четвертых моментов, для выбора $\widehat{\boldsymbol{W}}$. Обычно для надежного оценивания четвертых моментов требуется гораздо больший объем выборки, чем для оценивания первых и вторых моментов. Поэтому можно было бы ожидать от эффективных GMM-оценок худших результатов в конечных выборках по сравнению с оценками, не использующими четвертые моменты. Июльский выпуск 1996 года журнала *Journal of Business and Economic Statistics* содержит ряд статей, исследующих распределения GMM-оценок и соответствующих тестовых статистик в конечных выборках для различных процессов порождения данных. Общий их вывод заключается в том, что равновзвешенная GMM-оценка ($\widehat{\boldsymbol{W}} = \boldsymbol{I}$) обычно превосходит эффективную GMM-оценку по смещению и дисперсии в конечных выборках. Также было обнаружено, что размер статистики Вальда значительно превосходит в конечных выборках заявленный уровень значимости. То есть если α — заявленный уровень значимости, а c_α — соответствующее ему критическое значение, так что $\text{Prob}(\chi^2 > c_\alpha) = \alpha$, то вероятность того, что статистика Вальда в конечных выборках превысит c_α, намного превышает α. Другими словами, нулевая гипотеза будет отвергаться слишком часто. Однако, наряду с другими исследованиями, посвященными конечным выборкам, данные исследования не дают четких количественных указаний, которыми может руководствоваться практик.

Контрольные вопросы

1. Покажите, что все результаты параграфов 2.3–2.5 являются частными случаями результатов этой главы. В частности, покажите, что (3.5.1) сводится к (b). **Указание:** *$\boldsymbol{\Sigma}_{\boldsymbol{xz}}$ является квадратной, если $\boldsymbol{z}_i = \boldsymbol{x}_i$.*

2. (Вырожденная $\boldsymbol{W}$.) Пусть $\boldsymbol{W}$ вырождена, но $\boldsymbol{\Sigma}'_{\boldsymbol{xz}}\boldsymbol{W}\boldsymbol{\Sigma}_{\boldsymbol{xz}}$ невырождена. Покажите, что все результаты данной главы сохраняются (за исключением утверждения 3.5).

3. (Асимптотически эквивалентный выбор $\widehat{\boldsymbol{W}}$.) Предположим, что $\widehat{\boldsymbol{W}}_1 - \widehat{\boldsymbol{W}}_2 \to_{\text{p}} \boldsymbol{0}$. Покажите, что
$$\sqrt{n}\hat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}_1) - \sqrt{n}\hat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}_2) \underset{\text{p}}{\to} \boldsymbol{0}.$$
Указание:
$$\sqrt{n}\hat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}_1) - \sqrt{n}\hat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}_2) = [(\boldsymbol{S}'_{\boldsymbol{xz}}\widehat{\boldsymbol{W}}_1\boldsymbol{S}_{\boldsymbol{xz}})^{-1}\boldsymbol{S}'_{\boldsymbol{xz}}\widehat{\boldsymbol{W}}_1 - (\boldsymbol{S}'_{\boldsymbol{xz}}\widehat{\boldsymbol{W}}_2\boldsymbol{S}_{\boldsymbol{xz}})^{-1}\boldsymbol{S}'_{\boldsymbol{xz}}\widehat{\boldsymbol{W}}_2]\sqrt{n}\bar{\boldsymbol{g}}.$$
$\sqrt{n}\bar{\boldsymbol{g}}$ сходится по распределению к случайной величине. Используйте лемму 2.4(b).

4. (Трехшаговый GMM.) Предположим, что к эффективной двухшаговой процедуре GMM-оценки добавили следующий шаг: пересчитать $\widehat{\boldsymbol{S}}$ по формуле (3.5.10), но на этот раз с использованием остатков из второго шага. Вычислите GMM-оценку с использованием этого шага. Состоятельна ли она? Асимптотически нормальна? Эффективна? **Указание:** *Проверьте, что, согласно утверждению 3.4, пересчитанная $\widehat{\boldsymbol{S}}$ является состоятельной оценкой для $\boldsymbol{S}$. Зависит ли асимптотическое распределение GMM-оценки от выбора $\widehat{\boldsymbol{W}}$, если $\text{plim}_{n\to\infty}\widehat{\boldsymbol{W}}$ одинаков?*

5. (Когда z_i — строгое подмножество x_i.) Предположим, что z_i — строгое подмножество из x_i. То есть, x_i содержит, в дополнение к регрессорам (которые все предопределены), и некоторые другие предопределенные переменные. Совпадают ли МНК и GMM-оценки численно? [Ответ: Нет.]

6. (Представление эффективной GMM-оценки через матрицу данных.) Пусть $\boldsymbol{B}$ — диагональная матрица размера $n \times n$, i элемент которой равен $\widehat{\varepsilon}_i^2$, где $\widehat{\varepsilon}_i$ являются остатками первого шага оценивания. То есть

$$\boldsymbol{B} \equiv \begin{bmatrix} \widehat{\varepsilon}_1^2 & & \\ & \ddots & \\ & & \widehat{\varepsilon}_n^2 \end{bmatrix}.$$

Проверьте, что

$$\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}) = [\boldsymbol{Z}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{B}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}]^{-1}\boldsymbol{Z}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{B}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y},$$

где $\boldsymbol{X}$, $\boldsymbol{y}$ и $\boldsymbol{Z}$ — матрицы данных для инструментов, зависимой переменной и регрессоров (они определены ниже в параграфе 3.8).

7. (GLS-представление эффективного GMM.) Пусть $\boldsymbol{X}$, $\boldsymbol{y}$ и $\boldsymbol{Z}$ такие же, как в предыдущем вопросе. Тогда оцениваемое уравнение можно записать в виде $\boldsymbol{y} = \boldsymbol{Z}\boldsymbol{\delta} + \boldsymbol{\varepsilon}$. Домножьте обе стороны слева на $\boldsymbol{X}'$, чтобы получить

$$\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{X}'\boldsymbol{Z}\boldsymbol{\delta} + \boldsymbol{X}'\boldsymbol{\varepsilon}.$$

Считая $\boldsymbol{S}$ ковариационной матрицей $\boldsymbol{X}'\boldsymbol{\varepsilon}$, $\widehat{\boldsymbol{S}}$ — ее состоятельной оценкой, примените FGLS. Проверьте, что FGLS-оценка является эффективной GMM-оценкой (FGLS-оценка была определена в параграфе 1.6).

8. (Линейные комбинации условий ортогональности.) Выведите эффективную GMM-оценку, использующую линейную комбинацию условий ортогональности

$$\boldsymbol{A}\,\mathrm{E}(\boldsymbol{g}_i) = \boldsymbol{0}$$

где $\boldsymbol{A}$ — матрица размера $q \times K$ с полным строковым рангом ($q \leq K$). [Ответ: Замените $\boldsymbol{S}_{\boldsymbol{xz}}$ на $\boldsymbol{A}\boldsymbol{S}_{\boldsymbol{xz}}$, $\boldsymbol{s}_{\boldsymbol{xy}}$ на $\boldsymbol{A}\boldsymbol{s}_{\boldsymbol{xy}}$ и $\boldsymbol{S}$ на $\boldsymbol{A}\boldsymbol{S}\boldsymbol{A}'$. Формально эту оценку можно записать как (3.4.8) с $\widehat{\boldsymbol{W}} = \boldsymbol{A}'(\boldsymbol{A}\widehat{\boldsymbol{S}}\boldsymbol{A}')^{-1}\boldsymbol{A}$.] Проверьте, что при $q = K$ (так, что $\boldsymbol{A}$ невырождена) эффективная GMM-оценка численно совпадает с эффективной GMM-оценкой, соответствующей условиям ортогональности $\mathrm{E}(\boldsymbol{g}_i) = \boldsymbol{0}$.

3.6. Тестирование сверхидентифицирующих ограничений

Если уравнение идентифицируемо точно, то возможно выбрать $\widetilde{\boldsymbol{\delta}}$ таким, что все выборочные моменты $\boldsymbol{g}_n(\widetilde{\boldsymbol{\delta}})$ будут равными нулю и расстояние

$$J(\widetilde{\boldsymbol{\delta}}, \widehat{\boldsymbol{W}}) \equiv n \cdot \boldsymbol{g}_n(\widetilde{\boldsymbol{\delta}})'\,\widehat{\boldsymbol{W}}\boldsymbol{g}_n(\widetilde{\boldsymbol{\delta}})$$

также будет нулевым ($\widetilde{\boldsymbol{\delta}}$ является IV-оценкой). Если уравнение сверхидентифицируемо, то тогда это расстояние нельзя сделать точно равным нулю, но можно ожидать, что минимальное расстояние будет близким к нулю. Оказывается, что если взвешивающая матрица $\widehat{\boldsymbol{W}}$ выбирается оптимально, так что $\operatorname{plim} \widehat{\boldsymbol{W}} = \boldsymbol{S}^{-1}$, то минимальное расстояние имеет асимптотическое χ^2-распределение.

Пусть $\widehat{\boldsymbol{S}}$ является состоятельной оценкой $\boldsymbol{S}$. Рассмотрим сначала случай, когда расстояние вычисляется при истинном значении параметра $\boldsymbol{\delta}$. Поскольку, по определению, $\boldsymbol{g}_n(\widetilde{\boldsymbol{\delta}}) = \overline{\boldsymbol{g}}(\equiv \frac{1}{n}\sum_i \boldsymbol{g}_i)$ для $\widetilde{\boldsymbol{\delta}} = \boldsymbol{\delta}$, это расстояние равно:

$$J(\boldsymbol{\delta}, \widehat{\boldsymbol{S}}^{-1}) = n \cdot \overline{\boldsymbol{g}}' \, \widehat{\boldsymbol{S}}^{-1} \overline{\boldsymbol{g}} = (\sqrt{n}\,\overline{\boldsymbol{g}})' \, \widehat{\boldsymbol{S}}^{-1} (\sqrt{n}\,\overline{\boldsymbol{g}}). \tag{3.6.1}$$

Поскольку $\sqrt{n}\overline{\boldsymbol{g}} \to_{\mathrm{d}} N(\mathbf{0}, \boldsymbol{S})$ и $\widehat{\boldsymbol{S}} \to_{\mathrm{p}} \boldsymbol{S}$, оно имеет асимптотическое распределение $\chi^2(K)$ по лемме 2.4d. Если заменить $\boldsymbol{\delta}$ на $\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}})$, то число степеней свободы изменится от K до $K - L$. Очевидной причиной является то, что перед оцениванием выборочного среднего $\boldsymbol{g}_i$ необходимо оценить L параметров $\boldsymbol{\delta}$ (мы сталкивались с аналогичной ситуацией в главе 1, в контексте несмещенного оценивания σ^2). Обобщим это как

Утверждение 3.6 (тест Хансена на сверхидентифицирующие ограничения [Hansen, 1982]): *Пусть имеется состоятельная оценка, $\widehat{\boldsymbol{S}}$, для $\boldsymbol{S}$* $(= \mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_i'))$. *При предположениях 3.1–3.5,*

$$J\big(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}), \widehat{\boldsymbol{S}}^{-1}\big) \; \big(= n \cdot \boldsymbol{g}_n(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}))' \, \widehat{\boldsymbol{S}}^{-1} \boldsymbol{g}_n(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}))\big) \underset{d}{\to} \chi^2(K - L).$$

Формальное доказательство оставляем как дополнительное упражнение. Поскольку $\widehat{\boldsymbol{S}}$, заданное (3.5.10), является состоятельной оценкой (при дополнительном предположении 3.6), минимальное расстояние, вычисляемое на втором шаге эффективной GMM-оценки, имеет асимптотическое распределение $\chi^2(K - L)$.

Следует отметить следующие три момента, касающиеся использования и интерпретации указанного J-теста:

- Это **тест спецификации** (*specification test*), проверяющий выполнение всех ограничений модели (которые являются предположениями, принятыми в утверждении 3.6). Если J-статистика из утверждения 3.6 необычно велика, то либо условия ортогональности (предположение 3.3), либо иные условия, либо и те и другие, скорее, не выполняются. Только при полной уверенности в этих иных условиях можно интерпретировать большое значение J как признак эндогенности некоторых из K инструментов, включенных в $\boldsymbol{x}_i$.

- В отличие от рассмотренных ранее тестов, данный тест несостоятелен по отношению к некоторым нарушениям условий ортогональности. Основная причина этого — снижение числа степеней свободы с K до $K - L$. Легко показать, что $\bar{\boldsymbol{g}}$ связан со своим выборочным аналогом $\boldsymbol{g}_n(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}))$ как

$$\sqrt{n}\,\boldsymbol{g}_n(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1})) = \widehat{\boldsymbol{B}}\sqrt{n}\,\bar{\boldsymbol{g}}, \quad \widehat{\boldsymbol{B}} \equiv \boldsymbol{I}_K - \boldsymbol{S}_{\boldsymbol{xz}}(\boldsymbol{S}'_{\boldsymbol{xz}}\,\widehat{\boldsymbol{S}}^{-1}\boldsymbol{S}_{\boldsymbol{xz}})^{-1}\boldsymbol{S}'_{\boldsymbol{xz}}\,\widehat{\boldsymbol{S}}^{-1}. \tag{3.6.2}$$

  Проблема в том, что, так как $\widehat{\boldsymbol{B}}\boldsymbol{S}_{\boldsymbol{xz}} = \boldsymbol{0}$, матрица $\widehat{\boldsymbol{B}}$ не имеет полного столбцового ранга. Если условия ортогональности не выполняются и $\mathrm{E}(\boldsymbol{g}_i) \neq \boldsymbol{0}$, то элементы $\sqrt{n}\bar{\boldsymbol{g}}$ будут стремиться к $-\infty$ и $+\infty$. Но, так как $\widehat{\boldsymbol{B}}$ не имеет полного столбцового ранга, $\widehat{\boldsymbol{B}}\sqrt{n}\bar{\boldsymbol{g}}$ и поэтому $J\big(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}), \widehat{\boldsymbol{S}}^{-1}\big)$ могут оставаться конечными при некоторых типах нарушений ортогональности[1].

- Свойства тестов в конечных выборках стали объектом пристального внимания лишь недавно. Некоторые статьи из июльского выпуска 1996 года *Journal of Business and Economic Statistics* показали, что реальный размер J-теста в конечных выборках значительно превышает номинальный размер (нулевая гипотеза отвергается слишком часто).

Тестирование подмножеств условий ортогональности

Предположим, что мы можем разделить K инструментов на две группы: вектор $\boldsymbol{x}_{i1}$ из K_1 переменных, удовлетворяющих условиям ортогональности, и вектор $\boldsymbol{x}_{i2}$ из оставшихся $K - K_1$ переменных, находящихся под подозрением. Поскольку порядок переменных не меняет численных свойств оценок и тестовых статистик, можно предположить без потери общности, что именно последние $K - K_1$ переменных $\boldsymbol{x}_i$ являются подозрительными инструментами.

$$\boldsymbol{x}_i = \begin{bmatrix} \boldsymbol{x}_{i1} \\ \boldsymbol{x}_{i2} \end{bmatrix} \begin{matrix} \} & K_1 \text{ строк}, \\ \} & K - K_1 \text{ строк}. \end{matrix} \tag{3.6.3}$$

Часть модели, которую мы хотели бы протестировать, — это

$$\mathrm{E}(\boldsymbol{x}_{i2}\cdot\varepsilon_i) = \boldsymbol{0}. \tag{3.6.4}$$

Это ограничение тестируемо, если число надежных инструментов не меньше числа коэффициентов, то есть $K_1 \geq L$. Основная идея заключается в сравнении двух J-статистик двух отдельных GMM-оценок одного

[1] См.: [Newey, 1985, Section 3] для детального рассмотрения этого вопроса.

и того же вектора $\boldsymbol{\delta}$: с использованием только вектора $\boldsymbol{x}_{i1}$ и с использованием обоих векторов, $\boldsymbol{x}_{i1}$ и $\boldsymbol{x}_{i2}$. Если включение дополнительных инструментов увеличивает значение J, то есть основания сомневаться в их предопределенности.

В соответствии с указанным разбиением $\boldsymbol{x}_i$, выборочные условия ортогональности $\boldsymbol{g}_n(\widetilde{\boldsymbol{\delta}})$ и $\boldsymbol{S}$ могут быть записаны как

$$\underset{(K\times 1)}{\boldsymbol{g}_n(\widetilde{\boldsymbol{\delta}})} \equiv \begin{bmatrix} \underset{(K_1\times 1)}{\boldsymbol{g}_{1n}(\widetilde{\boldsymbol{\delta}})} \\ \underset{((K-K_1)\times 1)}{\boldsymbol{g}_{2n}(\widetilde{\boldsymbol{\delta}})} \end{bmatrix}, \quad \underset{(K\times K)}{\boldsymbol{S}} \equiv \begin{bmatrix} \boldsymbol{S}_{11} & \boldsymbol{S}_{12} \\ \boldsymbol{S}_{21} & \boldsymbol{S}_{22} \end{bmatrix}, \tag{3.6.5}$$

где

$$\begin{aligned} \boldsymbol{S}_{11} &= \mathrm{E}(\varepsilon_i^2\, \boldsymbol{x}_{i1}\boldsymbol{x}_{i1}'), & \boldsymbol{S}_{12} &= \mathrm{E}(\varepsilon_i^2\, \boldsymbol{x}_{i1}\boldsymbol{x}_{i2}'), \\ \boldsymbol{S}_{21} &= \mathrm{E}(\varepsilon_i^2\, \boldsymbol{x}_{i2}\boldsymbol{x}_{i1}'), & \boldsymbol{S}_{22} &= \mathrm{E}(\varepsilon_i^2\, \boldsymbol{x}_{i2}\boldsymbol{x}_{i2}'). \end{aligned}$$

В частности, $\boldsymbol{g}_{1n}(\widetilde{\boldsymbol{\delta}})$ можно записать как

$$\boldsymbol{g}_{1n}(\widetilde{\boldsymbol{\delta}}) = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_{i1}\cdot(y_i - \boldsymbol{z}_i'\widetilde{\boldsymbol{\delta}}) \equiv \boldsymbol{s}_{\boldsymbol{x}_1 y} - \boldsymbol{S}_{\boldsymbol{x}_1\boldsymbol{z}}\widetilde{\boldsymbol{\delta}},$$

где

$$\boldsymbol{s}_{\boldsymbol{x}_1 y} \equiv \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_{i1}\cdot y_i, \quad \boldsymbol{S}_{\boldsymbol{x}_1\boldsymbol{z}} \equiv \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_{i1}\boldsymbol{z}_i'. \tag{3.6.6}$$

Для состоятельной оценки $\widehat{\boldsymbol{S}}$ эффективная GMM-оценка, использующая все K инструментов, и соответствующая J-статистика уже были выведены в этом и предыдущем параграфах. Воспроизведем их:

$$\widehat{\boldsymbol{\delta}} = (\boldsymbol{S}_{\boldsymbol{xz}}'\,\widehat{\boldsymbol{S}}^{-1}\boldsymbol{S}_{\boldsymbol{xz}})^{-1}\boldsymbol{S}_{\boldsymbol{xz}}'\,\widehat{\boldsymbol{S}}^{-1}\boldsymbol{s}_{\boldsymbol{x}y}, \tag{3.6.7}$$

$$J = n\cdot\boldsymbol{g}_n(\widehat{\boldsymbol{\delta}})'\,\widehat{\boldsymbol{S}}^{-1}\boldsymbol{g}_n(\widehat{\boldsymbol{\delta}}). \tag{3.6.8}$$

Эффективная GMM-оценка *того же* вектора коэффициентов, но с использованием только первых K_1 инструментов и соответствующая J-статистика получаются заменой $\boldsymbol{x}_i$ на $\boldsymbol{x}_{i1}$.

$$\overline{\boldsymbol{\delta}} = [\boldsymbol{S}_{\boldsymbol{x}_1\boldsymbol{z}}'(\widehat{\boldsymbol{S}}_{11})^{-1}\boldsymbol{S}_{\boldsymbol{x}_1\boldsymbol{z}}]^{-1}\boldsymbol{S}_{\boldsymbol{x}_1\boldsymbol{z}}'(\widehat{\boldsymbol{S}}_{11})^{-1}\boldsymbol{s}_{\boldsymbol{x}_1 y}, \tag{3.6.9}$$

$$J_1 = n\cdot\boldsymbol{g}_{1n}(\overline{\boldsymbol{\delta}})'(\widehat{\boldsymbol{S}}_{11})^{-1}\boldsymbol{g}_{1n}(\overline{\boldsymbol{\delta}}). \tag{3.6.10}$$

где $\widehat{\boldsymbol{S}}_{11}$ — состоятельная оценка $\boldsymbol{S}_{11}$. Тест основан на следующем утверждении, определяющем асимптотическое распределение $J - J_1$ (доказательство остается в качестве дополнительного упражнения).

Утверждение 3.7 (тестирование подмножества условий ортогональности[1]): *Пусть выполнены предположения 3.1–3.5 и* $\boldsymbol{x}_{i1}$ *является подвектором* $\boldsymbol{x}_i$. *Усилим предположение 3.4 требованием выполнения рангового условия идентификации для* $\boldsymbol{x}_{i1}$ *(то есть* $\mathrm{E}(\boldsymbol{x}_{i1}\boldsymbol{z}_i')$ *имеет полный столбцовый ранг). Тогда для любых состоятельных оценок* $\widehat{\boldsymbol{S}}$ *для* $\boldsymbol{S}$ *и* $\widehat{\boldsymbol{S}}_{11}$ *для* $\boldsymbol{S}_{11}$,

$$C \equiv J - J_1 \underset{d}{\to} \chi^2(K - K_1),$$

где $K = \#\boldsymbol{x}_i$ *(размерность* $\boldsymbol{x}_i$*),* $K_1 = \#\boldsymbol{x}_{i1}$ *(размерность* $\boldsymbol{x}_{i1}$*), и* J *и* J_1 *определены в* (3.6.8) *и* (3.6.10).

Выбор $\widehat{S}$ и $\widehat{S}_{11}$ асимптотически неважен, пока они состоятельны. Но в конечных выборках статистика C может быть отрицательной. Эта проблема может быть решена, и C может быть сделана неотрицательной, если везде используется одна и та же матрица $\widehat{S}$, то есть если матрица $\widehat{\boldsymbol{S}}_{11}$ в (3.6.9) и (3.6.10) будет подматрицей $\widehat{S}$ из (3.6.7) и (3.6.8). Это достигается следующими шагами:

1. Примените эффективный двухшаговый GMM с полным набором инструментов $\boldsymbol{x}_i$ для получения $\widehat{\boldsymbol{S}}$ на первом шаге и J и $\widehat{\boldsymbol{\delta}}$ на втором.

2. Извлеките подматрицу $\widehat{S}_{11}$ из матрицы $\widehat{\boldsymbol{S}}$, вычислите $\overline{\boldsymbol{\delta}}$ по (3.6.9) с использованием этой матрицы и J_1 по (3.6.10). Затем возьмите разность J.

Доказательство неотрицательности полученной оценки C в конечных выборках остается как дополнительное аналитическое упражнение.

Утверждение 3.7 можно использовать для тестирования эндогенности подмножества регрессоров, как в следующем примере.

Пример 3.3 (тестирование предопределенности образования в уравнении заработной платы): В уравнении заработной платы из примера 3.2 предположим, что уровень образования S_i может быть эндогенным. Для тестирования S_i на эндогенность разобьем $\boldsymbol{x}_i$ как

$$\boldsymbol{x}_{i1} = \begin{bmatrix} 1 \\ EXPR_i \\ AGE_i \\ MED_i \end{bmatrix}, \qquad \boldsymbol{x}_{i2} = S_i.$$

Вектор регрессоров $\boldsymbol{z}_i$ тот же, что и в примере 3.2. Первый шаг — провести эффективное двухшаговое GMM-оценивание $\boldsymbol{\delta}$ с $\boldsymbol{x}_i \equiv (1, EXPR_i,$

---

[1] Тест разработан в [Newey, 1985, Section 4] и [Eichenbaum, Hansen, Singleton, 1988, Appendix C].

$AGE_i, MED_i, S_i)'$ в качестве инструментов. Получим J и матрицу $\widehat{S}$ размера 5×5. На втором шаге извлечем подматрицу $\widehat{S}_{11}$, соответствующую $\boldsymbol{x}_{i1}$, и оценим тот же вектор $\boldsymbol{\delta}$ с меньшим количеством инструментов и с использованием этой $\widehat{S}_{11}$. Разность J-статистик, полученных двумя различными вариантами GMM-оценивания $\boldsymbol{\delta}$, должна иметь асимптотическое распределение $\chi^2(1)$.

Контрольные вопросы

1. Верно ли, что $J(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}), \overline{\boldsymbol{S}}^{-1}) \to_{\mathrm{d}} \chi^2(K-L)$, если $\widehat{\boldsymbol{S}}$ и $\overline{\boldsymbol{S}}$ — две разные состоятельные оценки $\boldsymbol{S}$?
2. Покажите, что $J(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}), \widehat{\boldsymbol{S}}^{-1}) = n\boldsymbol{s}'_{\boldsymbol{xy}}\widehat{\boldsymbol{S}}^{-1}(\boldsymbol{s}_{\boldsymbol{xy}} - \boldsymbol{S}_{\boldsymbol{xz}}\widehat{\delta}(\widehat{\boldsymbol{S}}^{-1}))$.
3. Может ли число степеней свободы C-статистики быть больше, чем $K-L$? [Ответ: Нет.]
4. Пусть $K_1 = L$. Зависит ли численное значение C от разбиения $\boldsymbol{x}_i$ на $\boldsymbol{x}_{i1}$ и $\boldsymbol{x}_{i2}$?

3.7. Тестирование гипотез по принципу отношения правдоподобий

В параграфе 3.5 нами были выведены статистики χ^2 для $H_0 : \boldsymbol{a}(\boldsymbol{\delta}) = \boldsymbol{0}$ по принципу Вальда. В данном параграфе то же самое будет сделано по принципу отношения правдоподобий (LR), заключающемуся в сравнении значений целевой функции с учетом нулевой гипотезы и без. Вывод тестовых статистик по принципу множителей Лагранжа для GMM и обобщение на случай нелинейных уравнений даны в параграфе 7.4.

В эффективном GMM-оценивании целевой функцией является $J(\widetilde{\boldsymbol{\delta}}, \widehat{\boldsymbol{S}}^{-1})$ для некоторой состоятельной оценки $\widehat{S}$ для $\boldsymbol{S}$. **Эффективная GMM-оценка с ограничением** (*restricted efficient GMM estimator*) определяется как

$$\text{эффективная GMM-оценка с ограничением: } \overline{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}) \equiv$$
$$\equiv \operatorname*{argmin}_{\widetilde{\boldsymbol{\delta}}} J(\widetilde{\boldsymbol{\delta}}, \widehat{\boldsymbol{S}}^{-1}) \quad \text{с учетом } H_0. \tag{3.7.1}$$

LR-принцип подразумевает, что разность

$$LR \equiv J(\overline{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}), \widehat{\boldsymbol{S}}^{-1}) - J(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}), \widehat{\boldsymbol{S}}^{-1}) \tag{3.7.2}$$

должна иметь асимптотическое распределение χ^2. Это действительно так.

Утверждение 3.8 (тестовая статистика по принципу LR): *Пусть выполнены предположения 3.1–3.5 и пусть доступна состоятельная оценка $\widehat{\boldsymbol{S}}$ матрицы $\boldsymbol{S}$ $(= \mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_i'))$. Рассмотрим нулевую гипотезу из $\#\boldsymbol{a}$ ограничений $H_0 : \boldsymbol{a}(\boldsymbol{\delta}) = \boldsymbol{0}$ таких, что $\boldsymbol{A}(\boldsymbol{\delta})$ — матрица первых производных размера $\#\boldsymbol{a} \times L$, непрерывна и имеет полный строковый ранг. Определим две статистики, W и LR, согласно* (3.5.16) *и* (3.7.2) *соответственно. Тогда, при нулевой гипотезе, верно следующее:*

(a) Обе указанные статистики асимптотически эквивалентны, так как имеют одинаковые асимптотические распределения (а именно $\chi^2(\#\boldsymbol{a})$).

(b) Обе статистики асимптотически эквивалентны в более строгом смысле, так что их разность сходится к нулю по вероятности: $LR - W \to_p 0$. (По лемме 2.4(a), этот результат сильнее, чем (a).)

(c) Более того, если указанная гипотеза линейна, так что ограничения могут быть записаны в виде $\boldsymbol{R}\boldsymbol{\delta} = \boldsymbol{r}$, то тогда две указанные статистики численно равны.

Доказательство утверждения для линейного случая является алгебраическим упражнением, которое оставим в качестве аналитического упражнения. Относительно полного доказательства смотрите параграф 7.4.

Несколько комментариев к утверждению 3.8:

- С одной стороны, преимуществом LR перед W является инвариантность: численное значение LR не зависит от формы представления нелинейных ограничений посредством $\boldsymbol{a}(\cdot)$. С другой стороны, при нелинейной гипотезе вам необходимо написать компьютерную программу для поиска эффективной GMM-оценки с ограничениями.

- Утверждение 3.8 требует, чтобы матрица расстояний $\widehat{\boldsymbol{W}}$ удовлетворяла условию эффективности $\operatorname{plim} \widehat{\boldsymbol{W}} = \boldsymbol{S}^{-1}$. В противном случае LR не будет иметь асимптотического χ^2-распределения. Напротив, статистика Вальда имеет асимптотическое распределение χ^2 и без выполнения этого условия.

- При вычислении LR следует использовать одну и ту же оценку $\boldsymbol{S}$. Пусть $\widetilde{\boldsymbol{S}}$ и $\overline{\boldsymbol{S}}$ — две различные состоятельные оценки $\boldsymbol{S}$. Рассмотрим статистику

$$J(\overline{\boldsymbol{\delta}}(\overline{\boldsymbol{S}}^{-1}), \overline{\boldsymbol{S}}^{-1}) - J(\widehat{\boldsymbol{\delta}}(\widetilde{\boldsymbol{S}}^{-1}), \widetilde{\boldsymbol{S}}^{-1}),$$

полученную при применении двух отдельных двухшаговых эффективных GMM-процедур. $\overline{\boldsymbol{S}}$ получена на первом шаге с учетом ограничений, $\widetilde{\boldsymbol{S}}$ получена на первом шаге без учета ограничений. Эта статистика асимптотически эквивалентна LR (их разность сходится по вероятности к нулю), но в конечных выборках она может быть отрицательной. Использование одинаковых оценок S гарантирует неотрицательность статистики в конечных выборках. Исследователи обычно используют оценку, полученную при оценивании без учета ограничений, хотя оценка, полученная при оценивании с учетом ограничений, также допустима, поскольку она состоятельна при нулевой гипотезе.

- Часть b утверждения (асимптотическая эквивалентность в более строгом смысле) говорит, что при достаточно большом объеме выборки и справедливости гипотезы результаты LR-теста (а не только вероятность отвержения или принятия) будут такими же, как и у теста W, потому что вероятность того, что статистики отличаются даже на малую величину, равна нулю на достаточно больших выборках.

- Для численного совпадения $LR = W$ в линейном случае, одна и та же оценка $\widehat{S}$ должна быть использована для расчета не только LR, но и W. В противном случае статистики будут только асимптотически эквивалентными.

LR-статистика для модели регрессии

Поскольку модель регрессии из главы 2 является частным случаем GMM-модели из этой главы, может быть полезно узнать, как LR будет выглядеть в этом частном случае. Поскольку в этой модели регрессии $x_i = z_i$ эффективной GMM-оценкой (не учитывающей ограничения) является МНК и $J(\overline{\delta}(\widehat{S}^{-1}), \widehat{S}^{-1}) = 0$. Таким образом,

$$LR = J(\overline{\delta}(\widehat{S}^{-1}), \widehat{S}^{-1}), \tag{3.7.3}$$

где $\overline{\delta}(\widehat{S}^{-1})$ — ограниченная эффективная GMM-оценка. По предположению 3.8, эта статистика имеет асимптотическое распределение χ^2 и численно равна статистике Вальда при линейной нулевой гипотезе. Как будет показано ниже, при условной гомоскедастичности, эта статистика может быть записана как разность сумм квадратов остатков, нормированных на дисперсию ошибки.

Тест добавления переменной (необязательный материал)

В предыдущем параграфе мы рассмотрели тест спецификации, основанный на C-статистике, для проверки эндогенности подмножества x_{i2} инструментов x_i при условии предопределенности остальных инструментов $\boldsymbol{x}_{i1}$. Иногда мы сталкиваемся с частным случаем:

$$y_i = \boldsymbol{x}_{i1}'\boldsymbol{\delta} + \varepsilon_i. \tag{3.7.4}$$

Популярным способом проверить предопределенность инструментов $\boldsymbol{x}_{i2}$ является оценивание расширенной модели:

$$y_i = \boldsymbol{x}_{i1}'\boldsymbol{\delta} + \boldsymbol{x}_{i2}'\boldsymbol{\alpha} + \varepsilon_i = \boldsymbol{x}_i'\boldsymbol{\gamma} + \varepsilon_i \quad \text{с} \quad \boldsymbol{\gamma} = \begin{bmatrix} \boldsymbol{\delta} \\ \boldsymbol{\alpha} \end{bmatrix} \tag{3.7.5}$$

и тестирование нулевой гипотезы H_0: $\boldsymbol{\alpha} = \mathbf{0}$. Тестирование проводится либо с использованием статистики Вальда, либо численно равной ей LR-статистики. Этот тест часто называют **тестом добавления переменных** (*variable addition test*). Как он связан с C-тестом из утверждения 3.7?

Для расчета LR необходимо получить две эффективные GMM-оценки $\boldsymbol{\gamma}$ с одним и тем же набором инструментов $\boldsymbol{x}_i$: с ограничением $\boldsymbol{\alpha} = \mathbf{0}$ и без него. Эффективной GMM-оценкой без ограничений будет МНК-оценка уравнения (3.7.5), не учитывающего ограничений. Соответствующая J-статистика равна нулю. Пусть

$$\widehat{\boldsymbol{S}} \equiv \frac{1}{n}\sum_{i=1}^{n} e_i^2 \boldsymbol{x}_i\boldsymbol{x}_i', \tag{3.7.6}$$

где e_i — остатки МНК-оценивания уравнения (3.7.5) без ограничений. При использовании этой оценки $\boldsymbol{S}$ эффективная GMM-оценка $\boldsymbol{\gamma}$, учитывающая ограничения, минимизирует функцию

$$J(\widetilde{\boldsymbol{\gamma}}, \widehat{\boldsymbol{S}}^{-1}) = n\cdot(\boldsymbol{s}_{\boldsymbol{xy}} - \boldsymbol{S}_{\boldsymbol{xx}}\widetilde{\boldsymbol{\gamma}})'\,\widehat{\boldsymbol{S}}^{-1}(\boldsymbol{s}_{\boldsymbol{xy}} - \boldsymbol{S}_{\boldsymbol{xx}}\widetilde{\boldsymbol{\gamma}})$$

при $\boldsymbol{\alpha} = \mathbf{0}$. Очевидно, что оценку, учитывающую ограничения, можно записать следующим образом:

$$\overline{\boldsymbol{\gamma}} = \begin{bmatrix} \widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}) \\ \mathbf{0} \end{bmatrix},$$

где $\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1})$ — эффективная GMM-оценка уравнения (3.7.4), учитывающего ограничения, с инструментами $\boldsymbol{x}_i$. Итак,

$$\begin{aligned} LR &= n\cdot(\boldsymbol{s}_{\boldsymbol{xy}} - \boldsymbol{S}_{\boldsymbol{xx}}\overline{\boldsymbol{\gamma}})'\,\widehat{\boldsymbol{S}}^{-1}(\boldsymbol{s}_{\boldsymbol{xy}} - \boldsymbol{S}_{\boldsymbol{xx}}\overline{\boldsymbol{\gamma}}) \\ &\quad \text{(так как } J = 0 \text{ для GMM без ограничений)} \\ &= n\cdot(\boldsymbol{s}_{\boldsymbol{xy}} - \boldsymbol{S}_{\boldsymbol{xx}_1}\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}))'\,\widehat{\boldsymbol{S}}^{-1}(\boldsymbol{s}_{\boldsymbol{xy}} - \boldsymbol{S}_{\boldsymbol{xx}_1}\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1})) \\ &\quad \text{(так как } \boldsymbol{S}_{\boldsymbol{xx}}\overline{\boldsymbol{\gamma}} = \boldsymbol{S}_{\boldsymbol{xx}_1}\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1})). \end{aligned}$$

Это не что иное, как J-статистика Хансена для уравнения (3.7.4) без ограничений с инструментами $\boldsymbol{x}_i$. Эта статистика, в свою очередь, равна C-статистике, так как J_1 из утверждения 3.7 в настоящем случае равно нулю. Таким образом, все статистики — W из регрессии без ограничений, LR, C — численно равны J Хансена, при условии использования одной и той же $\widehat{\boldsymbol{S}}$. То есть тест добавления переменных численно эквивалентен тесту Хансена на сверхидентифицирующие ограничения.

Контрольные вопросы

1. (LR для модели регрессии.) Проверьте, что в модели регрессии с $\boldsymbol{z}_i = \boldsymbol{x}_i$

$$LR = \boldsymbol{y}'\boldsymbol{X}(n\widehat{\boldsymbol{S}})^{-1}\boldsymbol{X}'\boldsymbol{y} - 2\boldsymbol{y}'\boldsymbol{X}(n\widehat{\boldsymbol{S}})^{-1}(\boldsymbol{X}'\boldsymbol{X})\overline{\boldsymbol{\delta}} + \overline{\boldsymbol{\delta}}'(\boldsymbol{X}'\boldsymbol{X})(n\widehat{\boldsymbol{S}})^{-1}(\boldsymbol{X}'\boldsymbol{X})\overline{\boldsymbol{\delta}}.$$

2. (Выбор $\widehat{\boldsymbol{S}}$ в тесте добавления переменных.) Предположим, что $\widehat{\boldsymbol{S}}$ формируется из остатков, полученных при оценивании уравнения с ограничениями (3.7.4), и используется для получения W, LR и C. Равны ли они численно? Имеют ли они асимптотическое χ^2 распределение? **Указание:** *Состоятельна ли данная $\widehat{\boldsymbol{S}}$ при нулевой гипотезе?*

3.8. Приложения условной гомоскедастичности

До сих пор при выводе асимптотик для GMM-оценок мы не предполагали наличия условной гомоскедастичности. В этом параграфе рассматриваются последствия введения следующего предположения.

Предположение 3.7 (условная гомоскедастичность):

$$\mathrm{E}(\varepsilon_i^2|\boldsymbol{x}_i) = \sigma^2.$$

При условной гомоскедастичности матрица четвертых моментов $\boldsymbol{S}$ (равная $\mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_i') = \mathrm{E}(\varepsilon_i^2\boldsymbol{x}_i\boldsymbol{x}_i')$) может быть записана как произведение вторых моментов:

$$\boldsymbol{S} = \sigma^2\boldsymbol{\Sigma}_{\boldsymbol{xx}}. \tag{3.8.1}$$

где $\boldsymbol{\Sigma}_{\boldsymbol{xx}} = \mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')$. Как и в главе 2, данное разложение имеет несколько следствий:

- Поскольку $\boldsymbol{S}$ невырождена, по предположению 3.5, из данного разложения следует, что $\sigma^2 > 0$ и $\boldsymbol{\Sigma}_{\boldsymbol{xx}}$ невырождена.

- Оценка, использующая данное разложение $\boldsymbol{S}$, имеет вид:

$$\widehat{\boldsymbol{S}} = \widehat{\sigma}^2\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i' = \widehat{\sigma}^2\,\boldsymbol{S}_{\boldsymbol{xx}}, \tag{3.8.2}$$

где $\widehat{\sigma}^2$ — некоторая состоятельная оценка, определяемая ниже. Из эргодической стационарности, $\boldsymbol{S_{xx}} \to_{\text{a.s.}} \boldsymbol{\Sigma_{xx}}$. Таким образом, при состоятельной $\widehat{\sigma}^2$ нам не нужно требование состоятельности четвертых моментов (предположение 3.6), чтобы обеспечить состоятельность $\widehat{\boldsymbol{S}}$.

Излишне говорить, что все представленные до сих пор результаты верны при этом дополнительном предположении условной гомоскедастичности. Но многие результаты и формулы могут быть упрощены заменой $\boldsymbol{S}$ на $\sigma^2\boldsymbol{\Sigma_{xx}}$, а выражения (3.5.10) для $\widehat{\boldsymbol{S}}$ на (3.8.2). Эти упрощения собраны в данном параграфе.

Эффективный GMM становится 2SLS

В эффективном GMM-оценивании взвешивающей матрицей является $\widehat{\boldsymbol{S}}^{-1}$. Если взять $\widehat{\boldsymbol{S}}$ из (3.8.2), то GMM-оценка станет равной:

$$\begin{aligned}\widehat{\delta}(\widehat{\boldsymbol{S}}^{-1}) &= [\boldsymbol{S}'_{\boldsymbol{xz}}(\widehat{\sigma}^2\,\boldsymbol{S_{xx}})^{-1}\,\boldsymbol{S_{xz}}]^{-1}\,\boldsymbol{S}'_{\boldsymbol{xz}}(\widehat{\sigma}^2\,\boldsymbol{S_{xx}})^{-1}\,\boldsymbol{s_{xy}} = \\ &= (\boldsymbol{S}'_{\boldsymbol{xz}}\,\boldsymbol{S}^{-1}_{\boldsymbol{xx}}\,\boldsymbol{S_{xz}})^{-1}\,\boldsymbol{S}'_{\boldsymbol{xz}}\,\boldsymbol{S}^{-1}_{\boldsymbol{xx}}\,\boldsymbol{s_{xy}} = \\ &= \widehat{\delta}(\boldsymbol{S}^{-1}_{\boldsymbol{xx}}) \equiv \widehat{\boldsymbol{\delta}}_{\text{2SLS}}\end{aligned} \tag{3.8.3}$$

и не зависящей от $\widehat{\sigma}^2$. В общем случае основной задачей первого шага эффективной двухшаговой процедуры GMM является получение состоятельной оценки $\boldsymbol{S}$. При условной гомоскедастичности в первом шаге необходимости нет, так как второй шаг сводится к $\widehat{\delta}(\boldsymbol{S}^{-1}_{\boldsymbol{xx}})$ — оценке GMM с $\boldsymbol{S_{xx}}$ вместо $\widehat{\boldsymbol{S}}$. Эта оценка, $\widehat{\delta}_{2SLS}$, называется **двухшаговой МНК** (*Two-Stage Least Squares*) (2SLS или TSLS)[1]. Это же уравнение можно оценить методом максимального правдоподобия. Параграф 8.6 будет включать в себя ML аналог 2SLS, называемый оценкой максимального правдоподобия с ограниченной информацией.

Выражение для $\text{Avar}(\widehat{\delta}_{2SLS})$ может быть получено подстановкой (3.8.1) в (3.5.13):

$$\text{Avar}(\widehat{\boldsymbol{\delta}}_{\text{2SLS}}) = \sigma^2\cdot(\boldsymbol{\Sigma}'_{\boldsymbol{xz}}\,\boldsymbol{\Sigma}^{-1}_{\boldsymbol{xx}}\,\boldsymbol{\Sigma_{xz}})^{-1}. \tag{3.8.4}$$

Естественной оценкой этого является

$$\widehat{\text{Avar}(\widehat{\boldsymbol{\delta}}_{\text{2SLS}})} = \widehat{\sigma}^2\cdot(\boldsymbol{S}'_{\boldsymbol{xz}}\,\boldsymbol{S}^{-1}_{\boldsymbol{xx}}\,\boldsymbol{S_{xz}})^{-1}. \tag{3.8.5}$$

Что касается $\widehat{\sigma}^2$, рассмотрим выборочную дисперсию ошибок 2SLS:

$$\widehat{\sigma}^2 \equiv \frac{1}{n}\sum_{i=1}^{n}(y_i - \boldsymbol{z}'_i\widehat{\boldsymbol{\delta}}_{\text{2SLS}})^2 \tag{3.8.6}$$

[1] Оценка впервые предложена в: [Theil, 1953].

(некоторые авторы при вычислении $\widehat{\sigma}^2$ делят сумму квадратов на $n - L$, а не на n). В силу утверждения 3.2, (3.8.6) $\to_{\mathrm{p}} \widehat{\sigma}^2$, если $\mathrm{E}(\boldsymbol{z}_i \boldsymbol{z}_i')$ существует и конечна. Тем самым $\widehat{\boldsymbol{S}}$, определенная в (3.8.2), с данной $\widehat{\sigma}^2$ является состоятельной для $\boldsymbol{S}$.

Подставляя (3.8.2) в (3.8.15) и (3.5.16), получим следующие выражения для t-статистики и статистики Вальда:

$$t_\ell = \frac{\widehat{\delta}_{2\mathrm{SLS},\ell} - \overline{\delta}_\ell}{SE_\ell} \quad \text{с} \quad SE_\ell = \sqrt{\frac{\widehat{\sigma}^2}{n} \cdot \left((\boldsymbol{S}_{\boldsymbol{xz}}' \boldsymbol{S}_{\boldsymbol{xx}}^{-1} \boldsymbol{S}_{\boldsymbol{xz}})^{-1}\right)_{\ell\ell}} \tag{3.8.7}$$

$$W = n \cdot \frac{\boldsymbol{a}(\widehat{\boldsymbol{\delta}}_{2\mathrm{SLS}})' \left[\boldsymbol{A}(\widehat{\boldsymbol{\delta}}_{2\mathrm{SLS}}) (\boldsymbol{S}_{\boldsymbol{xz}}' \boldsymbol{S}_{\boldsymbol{xx}}^{-1} \boldsymbol{S}_{\boldsymbol{xz}})^{-1} \boldsymbol{A}(\widehat{\boldsymbol{\delta}}_{2\mathrm{SLS}})'\right]^{-1} \boldsymbol{a}(\widehat{\boldsymbol{\delta}}_{2\mathrm{SLS}})}{\widehat{\sigma}^2}. \tag{3.8.8}$$

J становится статистикой Саргана

Задав $\widehat{\boldsymbol{W}} = (\widehat{\sigma}^2 \boldsymbol{S}_{\boldsymbol{xx}})^{-1}$, получим следующее выражение для расстояния (3.4.6):

$$J\left(\widetilde{\boldsymbol{\delta}}, (\widehat{\sigma}^2 \cdot \boldsymbol{S}_{\boldsymbol{xx}})^{-1}\right) = n \cdot \frac{(\boldsymbol{s}_{\boldsymbol{xy}} - \boldsymbol{S}_{\boldsymbol{xz}} \widetilde{\boldsymbol{\delta}})' \boldsymbol{S}_{\boldsymbol{xx}}^{-1} (\boldsymbol{s}_{\boldsymbol{xy}} - \boldsymbol{S}_{\boldsymbol{xz}} \widetilde{\boldsymbol{\delta}})}{\widehat{\sigma}^2}. \tag{3.8.9}$$

Из утверждения 3.6 тогда следует, что это расстояние, вычисленное для эффективной GMM-оценки при условной гомоскедастичности, $\widehat{\boldsymbol{\delta}}_{2\mathrm{SLS}}$, имеет асимптотическое распределение χ^2. Это расстояние называется **статистикой Саргана** (*Sargan's statistic*) [Sargan, 1958]:

$$\text{статистика Саргана} = n \cdot \frac{(\boldsymbol{s}_{\boldsymbol{xy}} - \boldsymbol{S}_{\boldsymbol{xz}} \widehat{\boldsymbol{\delta}}_{2\mathrm{SLS}})' \boldsymbol{S}_{\boldsymbol{xx}}^{-1} (\boldsymbol{s}_{\boldsymbol{xy}} - \boldsymbol{S}_{\boldsymbol{xz}} \widehat{\boldsymbol{\delta}}_{2\mathrm{SLS}})}{\widehat{\sigma}^2}. \tag{3.8.10}$$

Резюмируем полученные нами до этого результаты.

Утверждение 3.9 (асимптотические свойства 2SLS):

(a) *При предположениях 3.1–3.4, оценка 2SLS (3.8.3) состоятельна. При добавлении предположения 3.5, эта оценка асимптотически нормальна с асимптотической дисперсией* (3.5.1), *где* $\boldsymbol{W} = (\sigma^2 \boldsymbol{\Sigma}_{\boldsymbol{xx}})^{-1}$. *При добавлении к предположениям 3.1–3.5 предположения 3.7 (условная гомоскедастичность) эта оценка будет эффективной GMM-оценкой.*

Более того, если $\mathrm{E}(\boldsymbol{z}_i \boldsymbol{z}_i')$ *существует и конечна*[1], *то*

(b) *(3.8.5) — состоятельная оценка асимптотической дисперсии,*

(c) t_ℓ *в* (3.8.7) $\to_d N(0,1)$, W *в* (3.8.8) $\to_d \chi^2(\#\boldsymbol{r})$ *и*

[1] Это дополнительное предположение требуется для состоятельности $\widehat{\sigma}^2$, см. утверждение 3.2.

(d) статистика Саргана (3.8.10) $\to_d \chi^2(K - L)$.

Утверждение 3.8 показывает, что LR-статистика, являющаяся разностью значений J, полученных с учетом нулевой гипотезы и без, имеет асимптотическое распределение χ^2. Поскольку J можно записать как (3.8.9), мы получаем:

$$LR = n \times \\ \times \frac{(\boldsymbol{s_{xy}} - \boldsymbol{S_{xz}}\overline{\boldsymbol{\delta}})'\boldsymbol{S_{xx}^{-1}}(\boldsymbol{s_{xy}} - \boldsymbol{S_{xz}}\overline{\boldsymbol{\delta}}) - (\boldsymbol{s_{xy}} - \boldsymbol{S_{xz}}\widehat{\boldsymbol{\delta}}_{2\mathrm{SLS}})'\boldsymbol{S_{xx}^{-1}}(\boldsymbol{s_{xy}} - \boldsymbol{S_{xz}}\widehat{\boldsymbol{\delta}}_{2\mathrm{SLS}})}{\widehat{\sigma}^2}, \tag{3.8.11}$$

где $\overline{\boldsymbol{\delta}}$ — ограниченная 2SLS-оценка, минимизирующая (3.8.9) при нулевой гипотезе[1]. В утверждении 3.8 использование одинаковых $\widehat{\boldsymbol{S}}$ гарантирует неотрицательность статистики в конечных выборках. Здесь деление на одну и ту же $\widehat{\sigma}^2$ гарантирует неотрицательность статистики. Если гипотеза линейна, то тогда эта LR-статистика численно равна статистике Вальда W.

Свойства 2SLS на конечных выборках

Имеется довольно большое количество литературы, посвященной распределению оценки 2SLS в конечных выборках (см., например: [Judge et al., 1985, Section 15.4] и [Staiger and Stock, 1997, Section 1]). Некоторые исследования выводят точное распределение этой оценки в конечных выборках, в то время как другие используют метод Монте-Карло для различных DGP. Аналитические результаты, однако, неприменимы в эмпирических исследованиях, так как они выводятся при жестких ограничениях фиксированности инструментов и нормальности ошибок, а аналитические выражения не поддаются вычислениям.

Для случая единственного регрессора и единственного (стохастического) инструмента с нормальными ошибками [Nelson and Starz, 1990] выводят точное распределение оценки 2SLS в конечных выборках, относительно простое и легко вычисляемое. Они также показывают, что, если инструмент «слабый», в смысле низкой объяснительной силы на первом шаге регрессии регрессора на инструмент, основная часть распределения ошибки оценки $\widehat{\delta}_{2SLS} - \delta$ располагается далеко от нуля при недостаточно большом объеме выборки.

Их работа показывает необходимость сообщения значения R^2 для первого шага оценивания; если R^2 мал, то следует ожидать плохой аппроксимации распределения оценки 2SLS (пример приведен в части (g)

[1]Эта статистика впервые представлена в: [Gallant and Jorgenson, 1979].

эмпирического упражнения). Недавно [Staiger and Stock, 1997] предложили альтернативное асимптотическое приближение распределения оценок 2SLS в конечных выборках, других оценок и статистик для случая «слабых» инструментов. Их метод заключается в рассмотрении последовательности моделей, в которых коэффициенты при инструментах сходятся к нулю (аналитическое упражнение 10 выводит данный тип асимптотики для простого случая одного регрессора и одного инструмента).

Альтернативные выводы 2SLS

Если определить матрицы данных как

$$\underset{(n\times K)}{\mathbf{X}} = \begin{bmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_n' \end{bmatrix}, \quad \underset{(n\times L)}{\mathbf{Z}} = \begin{bmatrix} \mathbf{z}_1' \\ \mathbf{z}_2' \\ \vdots \\ \mathbf{z}_n' \end{bmatrix}, \quad \underset{(n\times 1)}{\mathbf{y}} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},$$

то легко увидеть, что оценки 2SLS и соответствующие статистики могут быть переписаны как

$$\begin{aligned}\widehat{\boldsymbol{\delta}}_{2\mathrm{SLS}} &= [\mathbf{Z}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}]^{-1}\mathbf{Z}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \\ &= (\mathbf{Z}'\mathbf{P}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{P}\mathbf{y},\end{aligned} \tag{3.8.3'}$$

где $\mathbf{P} \equiv \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ — проекционная матрица.

$$\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\delta}}_{2\mathrm{SLS}})} = n\cdot\widehat{\sigma}^2\cdot[\mathbf{Z}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}]^{-1} = n\cdot\widehat{\sigma}^2\cdot(\mathbf{Z}'\mathbf{P}\mathbf{Z})^{-1}, \tag{3.8.5'}$$

$$\widehat{\sigma}^2 \equiv \frac{\widehat{\boldsymbol{\varepsilon}}'\widehat{\boldsymbol{\varepsilon}}}{n}, \quad \text{где } \widehat{\boldsymbol{\varepsilon}} \equiv \mathbf{y} - \mathbf{Z}\widehat{\boldsymbol{\delta}}_{2\mathrm{SLS}}. \tag{3.8.6'}$$

$$t_\ell = \frac{\widehat{\delta}_{2\mathrm{SLS},\ell} - \overline{\delta}_\ell}{\sqrt{\widehat{\sigma}^2\cdot\left([\mathbf{Z}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}]^{-1}\right)_{\ell\ell}}}, \tag{3.8.7'}$$

$$W = \frac{\mathbf{a}(\widehat{\boldsymbol{\delta}}_{2\mathrm{SLS}})'\left[\mathbf{A}(\widehat{\boldsymbol{\delta}}_{2\mathrm{SLS}})(\mathbf{Z}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z})^{-1}\mathbf{A}(\widehat{\boldsymbol{\delta}}_{2\mathrm{SLS}})'\right]^{-1}\mathbf{a}(\widehat{\boldsymbol{\delta}}_{2\mathrm{SLS}})}{\widehat{\sigma}^2}. \tag{3.8.8'}$$

$$J(\widetilde{\boldsymbol{\delta}}, (\widehat{\sigma}^2\cdot\mathbf{S}_{\mathbf{xx}})^{-1}) = \frac{(\mathbf{y} - \mathbf{Z}\widetilde{\boldsymbol{\delta}})'\mathbf{P}(\mathbf{y} - \mathbf{Z}\widetilde{\boldsymbol{\delta}})}{\widehat{\sigma}^2}, \tag{3.8.9'}$$

$$\text{Статистика Саргана} = \frac{\widehat{\boldsymbol{\varepsilon}}'\mathbf{P}\widehat{\boldsymbol{\varepsilon}}}{\widehat{\sigma}^2}. \tag{3.8.10'}$$

Используя эти формулы, можно получить два других вывода оценки 2SLS.

2SLS как IV-оценка

Пусть $\widehat{z}_i$ $(L \times 1)$ — вектор L инструментов (которые будут сгенерированы из $\boldsymbol{x}_i$ по описанному ниже способу) для L регрессоров и пусть $\widehat{\boldsymbol{Z}}$ — матрица этих инструментов размера $n \times L$. То есть i-я строка $\widehat{\boldsymbol{Z}}$ — это $\widehat{z}_i'$. IV-оценка $\boldsymbol{\delta}$ с $\widehat{z}_i$ в качестве инструментов есть, по (3.4.4),

$$\widehat{\delta}_{\mathrm{IV}} = \left(\frac{1}{n}\sum_{i=1}^{n}\widehat{z}_i z_i'\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\widehat{z}_i \cdot y_i = (\widehat{\boldsymbol{Z}}'\boldsymbol{Z})^{-1}\,\widehat{\boldsymbol{Z}}'\boldsymbol{y}. \tag{3.8.12}$$

Теперь мы сгенерируем эти инструменты следующим образом: ℓ-й инструмент является подобранным значением регрессии $z_{i\ell}$ (ℓ-го регрессора) на $\boldsymbol{x}_i$. Вектор длины n из подобранных значений равен $\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})\boldsymbol{X}'\boldsymbol{z}_\ell$, где $\boldsymbol{z}_\ell$ — вектор длины n значений ℓ-го регрессора (ℓ-й столбец матрицы $\boldsymbol{Z}$). Таким образом, $n \times L$ матрица инструментов имеет вид:

$$\widehat{\boldsymbol{Z}} = (\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'z_1, \ldots, \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'z_L) = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z} = \boldsymbol{P}\boldsymbol{Z}, \tag{3.8.13}$$

где $\boldsymbol{P}$ — проекционная матрица. Подстановка этого выражения в (3.8.12) дает 2SLS-оценку (см. (3.8.3′)).

2SLS как две регрессии

Вместо подстановки сгенерированных инструментов $\widehat{z}_i$ в выражение для IV-оценивания уравнения $y_i = z_i'\delta + \varepsilon_i$ рассмотрим регрессию y_i на $\widehat{z}_i$. Оценка коэффициентов имеет вид:

$$(\widehat{\boldsymbol{Z}}'\widehat{\boldsymbol{Z}})^{-1}\widehat{\boldsymbol{Z}}'\boldsymbol{y} = (\boldsymbol{Z}'\boldsymbol{P}'\boldsymbol{P}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{P}\boldsymbol{y} = (\boldsymbol{Z}'\boldsymbol{P}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{P}\boldsymbol{y},$$

поскольку $\boldsymbol{P}$ симметрична и идемпотентна.

Это 2SLS-оценка. Поэтому оценки 2SLS могут быть получены за два шага: оценкой регрессии L регрессоров на $\boldsymbol{x}_i$ и получением подобранных значений $\widehat{z}_i$; и последующей оценкой регрессии y_i на эти подобранные значения.

Для предопределенных регрессоров, входящих в $\boldsymbol{x}_i$, нет необходимости в первом шаге, так как подобранное значение будет самим этим регрессором. Действительно, если $z_{i\ell}$ предопределен и включен в $\boldsymbol{x}_i$ как k-й инструмент, вектор длины n подобранных значений для регрессора ℓ равен $\boldsymbol{P}\boldsymbol{z}_\ell$, где $\boldsymbol{z}_\ell$ — вектор длины n, чьим i-м элементом является $z_{i\ell}$. Но так как $\boldsymbol{z}_\ell$ также является k-м столбцом $\boldsymbol{X}$, $\boldsymbol{P}\boldsymbol{z}_\ell = \boldsymbol{P}\boldsymbol{x}_k$. Поскольку $\boldsymbol{P}$ — проекционная матрица, то $\boldsymbol{P}\boldsymbol{x}_k = \boldsymbol{x}_k$.

Этот вывод 2SLS полезен, так как он оправдывает название оценки. Но есть один подводный камень. При выполнении регрессии y_i на $\widehat{z}_i$ на втором шаге стандартные ошибки автоматически рассчитываются МНК-программами на основе ошибок $\boldsymbol{y} - \widehat{\boldsymbol{Z}}\widehat{\delta}_{2SLS}$, которые не равны остаткам 2SLS $\boldsymbol{y}-\boldsymbol{Z}\widehat{\boldsymbol{\delta}}_{2SLS}$. Поэтому соответствующие стандартные ошибки и асимптотические дисперсии не могут использоваться для статистических выводов.

Когда регрессоры предопределены

Когда все регрессоры являются предопределенными, а ошибки условно гомоскедастичны, существует тесная связь между функцией расстояния J эффективной GMM-оценки и суммой квадратов остатков (SSR). Из (3.8.9′):

$$
\begin{aligned}
J(\widetilde{\boldsymbol{\delta}}, (\widehat{\sigma}^2 \cdot \boldsymbol{S}_{\boldsymbol{xx}})^{-1}) &= \frac{(\boldsymbol{y}-\boldsymbol{Z}\widetilde{\boldsymbol{\delta}})'\boldsymbol{P}(\boldsymbol{y}-\boldsymbol{Z}\widetilde{\boldsymbol{\delta}})}{\widehat{\sigma}^2} = \\
&= \frac{\boldsymbol{y}'\boldsymbol{P}\boldsymbol{y} - 2\boldsymbol{y}'\boldsymbol{P}\boldsymbol{Z}\widetilde{\boldsymbol{\delta}} + \widetilde{\boldsymbol{\delta}}'\boldsymbol{Z}'\boldsymbol{P}\boldsymbol{Z}\widetilde{\boldsymbol{\delta}}}{\widehat{\sigma}^2} = \\
&= \frac{\boldsymbol{y}'\boldsymbol{P}\boldsymbol{y} - 2\boldsymbol{y}'\boldsymbol{Z}\widetilde{\boldsymbol{\delta}} + \widetilde{\boldsymbol{\delta}}'\boldsymbol{Z}'\boldsymbol{Z}\widetilde{\boldsymbol{\delta}}}{\widehat{\sigma}^2} \\
&\quad (\text{так как } \boldsymbol{P}\boldsymbol{Z} = \boldsymbol{Z}, \text{ когда } \boldsymbol{z}_i \subset \boldsymbol{x}_i) \\
&= \frac{(\boldsymbol{y}-\boldsymbol{Z}\widetilde{\boldsymbol{\delta}})'(\boldsymbol{y}-\boldsymbol{Z}\widetilde{\boldsymbol{\delta}})}{\widehat{\sigma}^2} - \frac{\boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{P}\boldsymbol{y}}{\widehat{\sigma}^2} = \\
&= \frac{(\boldsymbol{y}-\boldsymbol{Z}\widetilde{\boldsymbol{\delta}})'(\boldsymbol{y}-\boldsymbol{Z}\widetilde{\boldsymbol{\delta}})}{\widehat{\sigma}^2} - \frac{(\boldsymbol{y}-\widehat{\boldsymbol{y}})'(\boldsymbol{y}-\widehat{\boldsymbol{y}})}{\widehat{\sigma}^2},
\end{aligned} \tag{3.8.14}
$$

где $\widehat{\boldsymbol{y}} = \boldsymbol{P}\boldsymbol{y}$ — вектор подобранных значений МНК-оценки без ограничений. Поскольку вычитаемое справа слагаемое не зависит от $\widetilde{\boldsymbol{\delta}}$, минимизация J сводится к минимизации суммы квадратов остатков (SSR) $(\boldsymbol{y}-\boldsymbol{Z}\widetilde{\boldsymbol{\delta}})'(\boldsymbol{y}-\boldsymbol{Z}\widetilde{\boldsymbol{\delta}})$. Из этого следует, что

- Эффективной GMM-оценкой является МНК-оценка (что справедливо и без условной гомоскедастичности так как $\boldsymbol{z}_i = \boldsymbol{x}_i$).

- Ограниченной GMM-оценкой с учетом ограничений нулевой гипотезы является МНК-оценка с ограничениями (с SSR, а не J в качестве целевой функции).

- Статистика Вальда, численно равная LR-статистике, может быть рассчитана как нормированная на $\widehat{\sigma}^2$ разность SSR с учетом ограничений нулевой гипотезы и без. Данный результат подтверждает получение статистики Вальда в соответствии с принципом отношения правдоподобий в параграфе 2.6.

Тестирование подмножества условий ортогональности

В параграфе 3.6 нами была представлена C-статистика для тестирования подмножества условий ортогональности. Она использовала две разные эффективные GMM-оценки одного и того же уравнения: одну с использованием полного набора инструментов $\boldsymbol{x}_i$, другую — с использованием подмножества инструментов $\boldsymbol{x}_{i1}$. Для получения выражения для этой статистики при условной гомоскедастичности предположим, что $\boldsymbol{X}_1$ — матрица размера $(n \times K_1)$, i-я строка которой равна $\boldsymbol{x}'_{i1}$. Поскольку при условной гомоскедастичности эффективной GMM-оценкой является 2SLS-оценка, эти две GMM-оценки задаются выражениями

$$\widehat{\boldsymbol{\delta}} = (\boldsymbol{Z}'\boldsymbol{P}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{P}\boldsymbol{y} \quad \text{с} \quad \boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}', \tag{3.8.15}$$

$$\overline{\boldsymbol{\delta}} = (\boldsymbol{Z}'\boldsymbol{P}_1\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{P}_1\boldsymbol{y} \quad \text{с} \quad \boldsymbol{P}_1 = \boldsymbol{X}_1(\boldsymbol{X}'_1\boldsymbol{X}_1)^{-1}\boldsymbol{X}'_1. \tag{3.8.16}$$

А C-статистика становится разностью двух статистик Саргана:

$$C = \frac{\widehat{\boldsymbol{\varepsilon}}'\boldsymbol{P}\widehat{\boldsymbol{\varepsilon}} - \overline{\boldsymbol{\varepsilon}}'\boldsymbol{P}_1\overline{\boldsymbol{\varepsilon}}}{\widehat{\sigma}^2}, \tag{3.8.17}$$

где

$$\widehat{\boldsymbol{\varepsilon}} \equiv \boldsymbol{y} - \boldsymbol{Z}\widehat{\boldsymbol{\delta}}, \quad \overline{\boldsymbol{\varepsilon}} \equiv \boldsymbol{y} - \boldsymbol{Z}\overline{\boldsymbol{\delta}}, \quad \widehat{\sigma}^2 \equiv \frac{\widehat{\boldsymbol{\varepsilon}}'\widehat{\boldsymbol{\varepsilon}}}{n}.$$

Как показано в случае без условной гомоскедастичности, C гарантированно будет неотрицательной при использовании одной и той же матрицы $\widehat{S}$, что при условной гомоскедастичности эквивалентно использованию одной и той же оценки $\widehat{\sigma}^2$, для нормировки как $\widehat{\boldsymbol{\varepsilon}}'\boldsymbol{P}\widehat{\boldsymbol{\varepsilon}}$, так и $\overline{\boldsymbol{\varepsilon}}'\boldsymbol{P}_1\overline{\boldsymbol{\varepsilon}}$ (см. (3.8.17)). По утверждению 3.7, C имеет асимптотическое распределение $\chi^2(K - K_1)$.

Вероятно, более популярной является статистика, предложенная [Hausman and Taylor, 1980] и детально изученная [Newey, 1985]. $\widehat{\boldsymbol{\delta}}$ асимптотически более эффективна, чем $\overline{\boldsymbol{\delta}}$, так как использует больше условий ортогональности. Тем самым $\mathrm{Avar}(\overline{\boldsymbol{\delta}}) \geq \mathrm{Avar}(\widehat{\boldsymbol{\delta}})$. Более того, как вам нужно будет показать (аналитическое упражнение 9), при той же гипотезе, гарантирующей асимптотическое распределение χ^2 у C-статистики,

$$\mathrm{Avar}(\overline{\boldsymbol{\delta}} - \widehat{\boldsymbol{\delta}}) = \mathrm{Avar}(\overline{\boldsymbol{\delta}}) - \mathrm{Avar}(\widehat{\boldsymbol{\delta}}). \tag{3.8.18}$$

(Если вы уже сталкивались с тестом Хаусмана в контексте ML, вы можете узнать в нем GMM версию принципа Хаусмана). По утверждению 3.9, $\mathrm{Avar}(\widehat{\boldsymbol{\delta}})$ состоятельно оценивается как

$$\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\delta}})} \equiv n \cdot \widehat{\sigma}^2 \cdot (\boldsymbol{Z}'\boldsymbol{P}\boldsymbol{Z})^{-1}. \tag{3.8.19}$$

Похожим образом, $\mathrm{Avar}(\overline{\boldsymbol{\delta}})$ состоятельно оценивается как

$$\widehat{\mathrm{Avar}(\overline{\boldsymbol{\delta}})} \equiv n \cdot \widehat{\sigma}^2 \cdot (\boldsymbol{Z}'\boldsymbol{P}_1\boldsymbol{Z})^{-1}. \tag{3.8.20}$$

Здесь, как при оценивании C-статистики, используется одна и та же оценка $\widehat{\sigma}^2$ для гарантирования неотрицательности статистики. Итоговая оценка $\mathrm{Avar}(\overline{\boldsymbol{\delta}} - \widehat{\boldsymbol{\delta}})$ равна:

$$\widehat{\mathrm{Avar}(\overline{\boldsymbol{\delta}} - \widehat{\boldsymbol{\delta}})} = \widehat{\sigma}^2 \cdot n \cdot \left[(\boldsymbol{Z}'\boldsymbol{P}_1\boldsymbol{Z})^{-1} - (\boldsymbol{Z}'\boldsymbol{P}\boldsymbol{Z})^{-1}\right]. \tag{3.8.21}$$

Хаусман и Тэйлор [Hausman and Taylor, 1980] показали, что

- в конечных выборках эта матрица положительно полуопределена (неотрицательно определена), но необязательно невырождена;
- для любой обобщенной обратной[1] к этой матрице **статистика Хаусмана** (*Hausman statistic*)

$$\begin{aligned} H &\equiv \sqrt{n}(\overline{\boldsymbol{\delta}} - \widehat{\boldsymbol{\delta}})'\left\{\widehat{\sigma}^2 \cdot n \cdot \left[(\boldsymbol{Z}'\boldsymbol{P}_1\boldsymbol{Z})^{-1} - (\boldsymbol{Z}'\boldsymbol{P}\boldsymbol{Z})^{-1}\right]\right\}^{-}\sqrt{n}(\overline{\boldsymbol{\delta}} - \widehat{\boldsymbol{\delta}}) = \\ &= \frac{(\overline{\boldsymbol{\delta}} - \widehat{\boldsymbol{\delta}})'\left[(\boldsymbol{Z}'\boldsymbol{P}_1\boldsymbol{Z})^{-1} - (\boldsymbol{Z}'\boldsymbol{P}\boldsymbol{Z})^{-1}\right]^{-}(\overline{\boldsymbol{\delta}} - \widehat{\boldsymbol{\delta}})}{\widehat{\sigma}^2} \end{aligned} \tag{3.8.22}$$

  инвариантна относительно выбора обобщенной обратной и имеет асимптотическое распределение χ^2 с числом степеней свободы, равным $\min(K - K_1, L - s)$, где

  $s = \#\boldsymbol{z}_i \cap \boldsymbol{x}_{i1} =$
  = количество регрессоров, используемых в качестве инструментов.

Какова связь между C и H при условной гомоскедастичности? Можно показать (см.: [Newey, 1985]), что: если $K - K_1 \leq L - s$, так что H и C имеют одинаковое число степеней свободы, то тогда H и C численно равны. В противном случае эти статистики не равны и имеют разное число степеней свободы.

Частый случай, в котором $K - K_1 \leq L - s$, — это когда $\boldsymbol{x}_{i2}$ является подмножеством $\boldsymbol{z}_i$, то есть подозрительные инструменты являются регрессорами. В этом случае (3.8.17) и (3.8.22) — два способа рассчитать одну и ту же статистику. В случае $K - K_1 > L - s$ число степеней свободы H меньше, чем у C. По этой причине тест Хаусмана, в отличие от C-теста, не состоятелен против некоторых локальных альтернатив[2].

---

[1] **Обобщенной обратной,** $\boldsymbol{A}^-$, для матрицы $\boldsymbol{A}$ называется любая матрица, такая, что $\boldsymbol{A}\boldsymbol{A}^-\boldsymbol{A} = \boldsymbol{A}$. Если матрица $\boldsymbol{A}$ квадратная и невырожденная, то $\boldsymbol{A}^-$ единственная и равна $\boldsymbol{A}^{-1}$.

[2] Статистику Хаусмана можно обобщить на случай условной гетероскедастичности, но это не применяется на практике, так как степени свободы зависят от неизвестных значений матриц $\boldsymbol{\Sigma}_{\boldsymbol{xz}}$ и $\boldsymbol{S}$. (См.: [Newey, 1985].)

Тестирование условной гомоскедастичности

Как было показано в параграфе 2.7, для случая МНК с предопределенными регрессорами существует удобный тест условной гомоскедастичности nR^2. В текущем случае статистика nR^2, полученная регрессированием квадрата 2SLS остатков на константу и моменты второго порядка инструментальных переменных, *не имеет* ожидаемого асимптотического распределения. Тестовая статистика, имеющая асимптотическое распределение χ^2, существует, но чрезвычайно сложна. (См.: [White, 1982].)

Тестирование на наличие автокорреляции

Для случая МНК в параграфе 2.10 были разработаны тесты на наличие автокорреляции в ошибке. Точнее, при выполнении предположения (2.10.15) (которое сильнее, чем предположения 2.3 и 3.3) и (2.10.16) (которое сильнее предположений 2.7 и 3.7 условной гомоскедастичности), модифицированная статистика Бокса — Пирса Q (2.10.20) может быть использована для тестирования нулевой гипотезы об отсутствии автокорреляции в ошибке, и эта статистика асимптотически эквивалентна (в более строгом смысле сходимости разности к нулю по вероятности) статистике nR^2 из регрессии (2.10.21). Может ли этот тест быть распространен на случай эндогенных регрессоров z_i? Если инструменты $\boldsymbol{x}_i$ удовлетворяют (2.10.15) и (2.10.16), то тогда доказательство в приложении 2.B к главе 2 может быть обобщено для получения модифицированной статистики Q, которая будет иметь асимптотическое распределение χ^2 при нулевой гипотезе об отсутствии автокорреляции. Однако выражение для такой статистики более сложное, чем (2.10.20), и поэтому не представлено здесь. При этом непохоже, чтобы эта модифицированная Q-статистика была асимптотически эквивалентной статистике nR^2 из регрессии (2.10.21).

Контрольные вопросы

1. (GMM с условной гомоскедастичностью.) В эффективном двухшаговом GMM-оценивании $\widehat{S}$ рассчитывается на втором шаге по (3.5.10) с использованием остатков первого шага. Выведите асимптотическую дисперсию двухшаговой оценки для случая условной гомоскедастичности. Равна ли она (3.8.4)? **Указание:** *При условной гомоскедастичности* (3.5.10) $\to_p$ $\to_p \sigma^2 \boldsymbol{\Sigma}_{\boldsymbol{xx}}$.

2. (2SLS без условной гомоскедастичности.) Состоятельна ли оценка 2SLS при отсутствии условной гомоскедастичности? Выведите plim (3.8.2) и $\mathrm{Avar}(\widehat{\boldsymbol{\delta}})$ без предположения об условной гомоскедастичности. Является ли 2SLS столь же эффективным, как и двухшаговый GMM (то есть асимптотическая дисперсия у него так же мала, как и (3.5.13))?

Указание: *2SLS-оценка является GMM-оценкой с выбором $\widehat{W}$, и она не обязана быть эффективной при отсутствии условной гомоскедастичности.*

3. (GLS-интерпретация 2SLS.) Проверьте, что 2SLS-оценку можно записать как GLS-оценку, если S_{xz} и s_{xy} интерпретировать как матрицу значений регрессоров и вектор значений зависимой переменной, а S_{xx} — как ковариационную матрицу вектора ошибок.

4. Приведите два примера, в которых $\widehat{\delta}_{2SLS}$ и $\widehat{\delta}(\widehat{S}^{-1})$ асимптотически эквивалентны, то есть
$$\sqrt{n}[\widehat{\delta}_{2SLS} - \widehat{\delta}(\widehat{S}^{-1})] \underset{p}{\to} \mathbf{0}$$
Указание: *Ключевые слова: «условная гомоскедастичность» и «точно идентифицируема».*

5. Предположим, что уравнение идентифицируемо точно. Покажите, что 2SLS (3.8.3) сводится к IV (3.4.4).

6. (Статистика Саргана как nR^2.) Докажите, что статистика Саргана (3.8.10′) равна nR^2_{uc}, где R^2_{uc} — нецентрированный коэффициент детерминации из регрессии $\widehat{\varepsilon}$ на X. **Указание:** *Контрольный вопрос 8 из параграфа 1.2.*

7. (z_i как строгое подмножество x_i.) Предположим, что z_i является строгим подмножеством x_i. То есть x_i содержит в дополнение к предопределенным регрессорам некоторые другие предопределенные переменные. Мы показали, что эффективная GMM-оценка является МНК при условной гомоскедастичности. Сохранится ли этот результат без условной гомоскедастичности? **Указание:** *Контрольный вопрос 5 из параграфа 3.5.*

3.9. Применение: отдача от образования

Начиная с работы [Mincer, 1958] связь между уровнем заработной платы и образованием была объектом большого количества теоретических и эмпирических исследований. Такое внимание может показаться странным, так как объяснение положительного влияния выглядит очевидным: образование увеличивает производительность труда индивида. Однако есть и другие объяснения. Например, модель сигналов на рынке труда Спенса утверждает, что более образованные индивиды получают большую заработную плату исключительно по причине того, что образование используется в качестве сигнала о более высоких способностях; несмотря на то что образование не увеличивает вероятность более высокого заработка, между этими показателями есть корреляция по причине влияния на них третьей переменной — способностей. В данном параграфе показано, как отделить влияние образования от влияния способностей с помощью методик из данной главы. Одной из ранних, посвященных этой проблеме, является работа Грилихеса [Griliches, 1976].

Данные NLS-Y

Данные, использованные Грилихесом, — Young Men's Cohort из National Longitudinal Survey (NLS-Y). Опрос в данной когорте проводился в 1966 году среди 5225 респондентов в возрасте 14–24 лет и был повторен через один и через два года. К 1969 году около четверти исходной выборки было потеряно, но 2026 респондентов сообщили уровень своей зарплаты в 1969 году и имели достаточно полные данные для вывода необходимых для исследования переменных. Привлекательной особенностью этих данных было включение в них двух показателей способностей. Первый — оценка в тесте Knowledge of the World of Work, полученная респондентами в 1966 году. Второй — уровень IQ. Все участники опроса, закончившие 9-й класс к 1966 году, подписывали документы, разрешавшие администрации школ передавать организаторам опроса данные по успеваемости. Итоговый опрос, проведенный в 1968 году, принес данные по умственным способностям респондентов, сгруппированные в аналоги уровня IQ. Из указанных 2026 респондентов уровень IQ имеется только для 1362, что говорит о том, что итоговый опрос покрывает только две трети исходной выборки. Наше обсуждение касается результатов Грилихеса, основанных на этой меньшей выборке. В табл. 3.1 приведены средние и стандартные отклонения ключевых переменных:

Таблица 3.1

Характеристики респондентов из NLS-Y

Средние и стандартные отклонения (в скобках)

Переменная	
Размер выборки	1,362
Возраст в 1969 году	22,3
	(3,2)
Образование в годах в 1969 году (S)	12,5
	(1.9)
Логарифм почасовой оплаты (в центах) в 1969 году (LW)	5,68
	(0.40)
Оценка в тесте Knowledge of the World of Work (KWW)	35.1
	(7,9)
Уровень IQ (IQ)	97.7
	(15.3)
Опыт работы в 1969 году	3.7
	(2.8)

Источник: [Griliches, 1976, Table 1]

Полулогарифмическое уравнение заработной платы

В литературе обычно оценивается следующее уравнение для заработной платы:

$$LW = \alpha + \beta S + \gamma A + \boldsymbol{\delta}'\boldsymbol{h} + \varepsilon, \tag{3.9.1}$$

где LW — логарифм заработной платы индивида, S — образование в годах, A — мера способностей, $\boldsymbol{h}$ — вектор наблюдаемых характеристик индивида (например, опыт работы и место жительства), $\boldsymbol{\delta}$ — соответствующий вектор коэффициентов, ε — ненаблюдаемая ошибка с нулевым математическим ожиданием (в данном параграфе индекс индивида i будет опускаться для упрощения записи). Полулогарифмическая спецификация для S часто оправдывается линейной связью между логарифмом заработной платы и образованием, установленной в качестве стилизованного факта на больших панелях данных (например, Current Population Survey)[1].

Коэффициент β показывает процентное изменение заработной платы при увеличении срока образования индивида на один год. Тем самым он показывает предельную отдачу от инвестирования в человеческий капитал, которая должна быть величиной того же порядка, что и доходность от инвестирования в финансовые активы. Мы предполагаем, что переменные S, A, $\boldsymbol{h}$ не коррелируют с ошибкой регрессии ε; поэтому МНК является подходящим методом оценивания при включении способностей A в модель наряду с образованием S и $\boldsymbol{h}$. В оставшейся части параграфа мы рассмотрим то, что Грилихес назвал «смещением из-за способностей», — смещение МНК-оценки коэффициента β при невключении A в уравнение или включении вместо нее какой-то неполноценной меры способностей.

Смещение из-за невключения переменной

Иногда наборы данных не включают в себя меры способностей (например, Current Population Survey). Каковы последствия игнорирования и невключения способностей в уравнение заработной платы? Из параграфа 2.9 нам известно, что регрессия LW на константу, S и $\boldsymbol{h}$ дает состоятельную оценку соответствующей МНК-проекции:

$$\begin{aligned}\widehat{\mathrm{E}}^*(LW|1, S, \boldsymbol{h}) &= \widehat{\mathrm{E}}^*(\alpha + \beta S + \gamma A + \boldsymbol{\delta}'\boldsymbol{h} + \varepsilon|1, S, \boldsymbol{h}) \quad (\text{из } (3.9.1)) = \\ &= \alpha + \beta S + \boldsymbol{\delta}'\boldsymbol{h} + \gamma\,\widehat{\mathrm{E}}^*(A|1, S, \boldsymbol{h}) + \widehat{\mathrm{E}}^*(\varepsilon|1, S, \boldsymbol{h}).\end{aligned} \tag{3.9.2}$$

Поскольку все регрессоры в (3.9.1) предопределены, имеем $\mathrm{E}(\varepsilon) = 0$, $\mathrm{E}(S\varepsilon) = 0$ и $\mathrm{E}(\boldsymbol{h}\varepsilon) = \boldsymbol{0}$. Поэтому $\widehat{\mathrm{E}}^*(\varepsilon|1, S, \boldsymbol{h}) = 0$. Если обозначить

[1] Вопросы функциональной формы рассматриваются в: [Card, 1995].

$\widehat{\mathrm{E}}^*(A|1, S, \boldsymbol{h}) = \theta_1 + \theta_S S + \boldsymbol{\theta}_{\boldsymbol{h}}' \boldsymbol{h}$, то (3.9.2) преобразуется в

$$\widehat{\mathrm{E}}^*(LW|1, S, \boldsymbol{h}) = (\alpha + \gamma\theta_1) + (\beta + \gamma\theta_S)S + (\boldsymbol{\delta} + \gamma\cdot\boldsymbol{\theta}_{\boldsymbol{h}})'\boldsymbol{h}.$$

Тем самым МНК-оценка коэффициента может оказаться асимптотически смещенной для всех включенных регрессоров. Это явление известно как **смещение из-за невключения переменной** (*omitted variable bias*). В частности,

$$\operatorname{plim} \widehat{\beta}_{\mathrm{OLS}} = \beta + \gamma\theta_S.$$

То есть МНК-оценка коэффициента при образовании S включает непрямой эффект способностей через образование ($\gamma\theta_S$) наряду с прямым эффектом β. Если θ_S положительна, то оценка коэффициента $\widehat{\beta}_{OLS}$ смещена вверх.

Используя описанную выше выборку из NLS-Y, состоящую из 1362 индивидов, Грилихес оценил уравнение заработной платы для 1969 года. Список переменных, включенных в $\boldsymbol{h}$, здесь представлен не будет; достаточно сказать, что он включает в себя опыт работы в годах и некоторые дамми-переменные для региона и размера города. Оценка Грилихеса для коэффициента при уровне образования S в случае игнорирования способностей представлена в строке 1 табл. 3.2. Оценка смотрится хорошо: эта точечная оценка похожа на доходность, которую можно получить от вложения в финансовые активы; причем эта оценка достаточно точна, о чем говорит высокое значение t-статистики.

IQ как мера способностей

Как уже было сказано, в NLS-Y представлены две меры способностей: IQ и KWW (оценка в тесте *Knowledge of the World of Work* в 1966 году). Поскольку респонденты NLS-Y имели в 1966 году возраст не менее 14 лет, результаты данного теста отражают влияние уже полученного на тот момент образования и не могут служить мерой чистых способностей. Уровень IQ не имеет этого недостатка. Если бы оценка IQ была идеальной мерой способностей индивида, то можно было бы получить состоятельную оценку уравнения (3.9.1) простой подстановкой IQ вместо A. МНК-оценки Грилихеса при включении IQ в модель представлены в строке 2 табл. 3.2. Теперь коэффициент при образовании ниже, что подтверждает наше предположение о том, что коэффициент при образовании при невключении в модель меры способностей включает в себя также влияние способностей на размер заработной платы.

Таблица 3.2

Оценки коэффициентов

Уравнение: $LW = \alpha + \beta S + \gamma IQ +$ прочие переменные

Строка	**Метод оценивания**	**Коэффициент**		**SER**	R^2	**Эндогенные регрессоры**	**Исключенные предопределенные регрессоры**
		S	***IQ***				
1	МНК	0,065 (13,2)		0,332	0,309	нет	
2	МНК	0,059 (10,7)	0,0019 (2,8)	0,331	0,313	нет	
3	2SLS	0,052 (7,0)	0,0038 (2,4)	0,332		IQ	MED, KWW, возраст, возраст в квадрате, общие характеристики

Источник: строка 1 — уравнение (B1) в: [Griliches, 1976, Table 2];
строка 2 — уравнение (B3) в табл. 2 там же;

Ошибки в переменных

Разумеется, уровень IQ не может быть безошибочной мерой способностей. Если η — ошибка измерения, то IQ связан с A как

$$IQ = \phi + A + \eta, \tag{3.9.3}$$

с $\mathrm{E}(\eta) = 0$. Интерпретация η как ошибки измерения позволяет предположить, что она не коррелирует с A, S, $\boldsymbol{h}$ и ошибкой регрессии ε. Подставляя (3.9.3) в (3.9.1), получаем:

$$LW = (\alpha - \gamma\phi) + \beta S + \gamma IQ + \boldsymbol{\delta}'\boldsymbol{h} + (\varepsilon - \gamma\eta). \tag{3.9.4}$$

Проиллюстрируем теперь на этом примере, что если хотя бы одна из включенных в модель переменных измерена с ошибкой, то оценки коэффициентов у всех переменных могут быть асимптотически смещенными.

Для изучения последствий включения в модель измеренной с ошибкой меры IQ рассмотрим следующую МНК-проекцию:

$$\begin{aligned}\widehat{\mathrm{E}}^*(LW|1, S, IQ, \boldsymbol{h}) = (\alpha - \gamma\phi) + \beta S + \gamma IQ + \boldsymbol{\delta}'\boldsymbol{h} + \\ + \widehat{\mathrm{E}}^*(\varepsilon|1, S, IQ, \boldsymbol{h}) - \gamma\,\widehat{\mathrm{E}}^*(\eta|1, S, IQ, \boldsymbol{h}).\end{aligned} \tag{3.9.5}$$

Рассмотрим $\widehat{\mathrm{E}}^*(\varepsilon|1, S, IQ, \boldsymbol{h})$ в (3.9.5). Поскольку $\mathrm{E}(\varepsilon) = 0$, и S и $\boldsymbol{h}$ не коррелированы с ε, $(1, S, \boldsymbol{h})$ ортогональны ε. IQ также ортогонален ε, так как

$$\begin{aligned}\mathrm{E}(IQ\,\varepsilon) &= \mathrm{E}[(\phi + A + \eta)\,\varepsilon] \quad (\text{из } (3.9.3) = \\ &= \mathrm{E}(\eta\,\varepsilon) \quad (\text{поскольку } \mathrm{E}(\varepsilon) = 0,\ \mathrm{Cov}(A, \varepsilon) = 0) = \\ &= 0.\end{aligned}$$

Тем самым $\widehat{\mathrm{E}}^*(\varepsilon|1, S, IQ, \boldsymbol{h}) = 0$ в (3.9.5).

Таким образом, смещения, если таковые есть, равны $-\gamma\,\widehat{\mathrm{E}}^*(\eta|1, S, IQ, \boldsymbol{h})$ в (3.9.5). Пусть $(\theta_S, \theta_{IQ}, \boldsymbol{\theta_h})$ — коэффициенты проекции $(S, IQ, \boldsymbol{h})$ в $\widehat{\mathrm{E}}^*(\eta|1, S, IQ, \boldsymbol{h})$. По формуле (2.9.7) из главы 2

$$\begin{bmatrix}\theta_S \\ \theta_{IQ} \\ \boldsymbol{\theta_h}\end{bmatrix} = \begin{bmatrix}\mathrm{Var}(S) & \mathrm{Cov}(S, IQ) & \mathrm{Cov}(S, \boldsymbol{h}') \\ \mathrm{Cov}(IQ, S) & \mathrm{Var}(IQ) & \mathrm{Cov}(IQ, \boldsymbol{h}') \\ \mathrm{Cov}(\boldsymbol{h}, S) & \mathrm{Cov}(\boldsymbol{h}, IQ) & \mathrm{Var}(\boldsymbol{h})\end{bmatrix}^{-1} \begin{bmatrix}\mathrm{Cov}(S, \eta) \\ \mathrm{Cov}(IQ, \eta) \\ \mathrm{Cov}(\boldsymbol{h}, \eta)\end{bmatrix}. \tag{3.9.6}$$

В этом выражении $\mathrm{Cov}(S, \eta)$ и $\mathrm{Cov}(\boldsymbol{h}, \eta)$ равны нулю по предположению. Однако $\mathrm{Cov}(IQ, \eta)$ не равна нулю, так как

$$\begin{aligned}\mathrm{Cov}(IQ, \eta) &= \mathrm{E}(IQ\,\eta) \quad (\text{поскольку } \mathrm{E}(\eta) = 0) = \\ &= \mathrm{E}[(\phi + A + \eta)\,\eta] \quad \text{из } (3.9.3) = \\ &= \phi\,\mathrm{E}(\eta) + \mathrm{E}(A\,\eta) + \mathrm{E}(\eta^2) \\ &= \mathrm{Var}(\eta) \quad (\text{так как } \mathrm{E}(\eta) = 0 \text{ и } \mathrm{Cov}(A, \eta) = 0).\end{aligned}$$

То есть ошибка измерения, не коррелируя с истинным значением, с необходимостью коррелирует с измеренным значением. С учетом того, что

$$\mathrm{Cov}(S, \eta) = 0, \quad \mathrm{Cov}(IQ, \eta) = \mathrm{Var}(\eta), \quad \text{и} \quad \mathrm{Cov}(\boldsymbol{h}, \eta) = \boldsymbol{0},$$

коэффициенты проекции можно записать как

$$\begin{bmatrix} \theta_S \\ \theta_{IQ} \\ \boldsymbol{\theta_h} \end{bmatrix} = \mathrm{Var}(\eta) \cdot \boldsymbol{a},$$

где

$$\boldsymbol{a} \equiv \text{второй столбец} \begin{bmatrix} \mathrm{Var}(S) & \mathrm{Cov}(S, IQ) & \mathrm{Cov}(S, \boldsymbol{h}') \\ \mathrm{Cov}(IQ, S) & \mathrm{Var}(IQ) & \mathrm{Cov}(IQ, \boldsymbol{h}') \\ \mathrm{Cov}(\boldsymbol{h}, S) & \mathrm{Cov}(\boldsymbol{h}, IQ) & \mathrm{Var}(\boldsymbol{h}) \end{bmatrix}^{-1}.$$

Тем самым в регрессии LW на константу, S, IQ и $\boldsymbol{h}$

$$\operatorname{plim} \widehat{\beta}_{\mathrm{OLS}} = \beta - \gamma \cdot \mathrm{Var}(\eta) \cdot (\text{1-й элемент } \boldsymbol{a}), \tag{3.9.7a}$$

$$\operatorname{plim} \widehat{\gamma}_{\mathrm{OLS}} = \gamma - \gamma \cdot \mathrm{Var}(\eta) \cdot (\text{2-й элемент } \boldsymbol{a}). \tag{3.9.7b}$$

Поскольку второй элемент $\boldsymbol{a}$ положителен (это диагональный элемент матрицы, обратной к ковариационной матрице), МНК-оценка коэффициента при способности будет смещена вниз. Однако направление смещения коэффициента при образовании зависит от знака первого элемента $\boldsymbol{a}$. Обычно $\mathrm{Cov}(S, IQ)$ положительна, так что, за исключением нетипично сильной корреляции (S, IQ) с $\boldsymbol{h}$, первый элемент $\boldsymbol{a}$ будет отрицательным. Тем самым можно предположить, что $\widehat{\beta}_{OLS}$ для образования будет смещена вверх.

Коррекция смещения с помощью 2SLS

Для учета смещения Грилихес применяет к уравнению заработной платы 2SLS. Чтобы сделать это, следует указать список инструментов. В набор инструментов можно включить предопределенные регрессоры $(1, S, \boldsymbol{h})$. Дополнительные инструменты для IQ, включенные в набор, — возраст, возраст в квадрате, KWW, образование матери и некоторые другие характеристики индивида (например, занятость отца). Эти переменные также рассматриваются как предопределенные. 2SLS-оценка Грилихеса для этой спецификации приведена в строке 3 табл. 3.2[1]. В соответствии с нашим предсказанием того, что МНК-оценка для коэффициента при мере способностей смещена вниз, МНК-оценка из строки 2, равная 0,0019, меньше, чем приведенная в строке 3 2SLS-оценка, равная 0,0038. Наше

[1] R^2 не приведено, так как сумму квадратов ошибок в случае 2SLS, в отличие от случая МНК, нельзя разделить между «объясненной» и «необъясненной» изменчивостью.

предположение о том, что оценка $\widehat{\beta}_{OLS}$ коэффициента при образовании при включении IQ смещена вверх, также подтверждается данными, так как оценка β, равная 0,059, в строке 2 выше, чем оценка 0,052 в строке 3.

В итоге «смещение из-за способностей», исследованное Грилихесом, заключается в смещении оценки коэффициента при образовании вверх при невключении IQ и смещении ее в неизвестном направлении (вероятно, вверх) при неточном измерении включенной переменной (в данном случае IQ). 2SLS дает решение, но этот метод зависит от предположения, что использованные инструменты (такие как MED и KWW) не коррелируют с ненаблюдаемыми детерминантами заработной платы ε и η. Эти детерминанты могут включать в себя такие персональные характеристики, как усердие. Представляется небезосновательным предположить, что KWW зависит от таких характеристик. Если это так, то KWW не может быть инструментом. Если пойти дальше, то можно утверждать, что образование матери также непредопределенная переменная, так как некоторые характеристики матери, влиявшие на уровень ее образования и ценящиеся на рынке труда, могут быть унаследованы детьми. Являются ли данные переменные предопределенными, можно проверить с использованием статистики Саргана, но в статье Грилихеса эти статистики не приведены (и сама работа написана задолго до того, как это стало общей практикой).

Дальнейшее развитие

Возможно, из-за того, что выбор инструментов, коррелированных с IQ, но не коррелированных в ненаблюдаемыми характеристиками, является сложной задачей, учет «смещения из-за способностей» продолжает привлекать большое внимание в литературе. Существует довольно много литературы, начиная с [Behrman et al., 1980], в которой производится сравнение однояйцевых близнецов с разным уровнем образования для учета генетических характеристик и влияния семьи. Можно также рассматривать одного и того же человека в разные моменты времени. Необходимая для обоих этих случаев методика, оценка с фиксированными эффектами, будет рассмотрена в главе 5.

Другой важной проблемой является эндогенность уровня образования S. Если предположить, что ошибка включает в себя набор ненаблюдаемых характеристик индивида, которые могут повлиять на выбор уровня образования, то S уже следует рассматривать как эндогенную переменную. Но, опять же, выбор подходящих инструментов для S, не коррелированных с ненаблюдаемыми характеристиками, но коррелированных с уровнем образования, очень сложен. Последние работы в этой области можно рассматривать как поиск детерминантов уровня образования, имеющих мало общего с подобными ненаблюдаемыми характеристиками. Например, [Angrist and Krueger, 1991] используют переменную, показывающую, воспользовался ли индивид законом об обязательном школьном образовании.

Наконец, снова привлекают внимание работы, посвященные связи между качеством школы и заработком (см., например: [Card and Krueger, 1996]).

Контрольные вопросы

1. Перечислите все предположения о средних и ковариациях $(S, A, IQ, KWW, \boldsymbol{h}, \varepsilon, \eta)$, сделанные в данном параграфе. Изучите, как Грилихес формулирует эти предположения. Какие из них являются ключевыми для состоятельности 2SLS-оценки?

2. Предположим, что зависимость между IQ и A задается не (3.9.3), а выражением
$$IQ = \phi + \pi A + \eta.$$
Остается ли состоятельной 2SLS-оценка коэффициента β при уровне образования? [Ответ: Да.]

3. В модели сигналов на рынке труда Спенса индивиды разделены на две группы: малоспособных, выбирающих низкий уровень образования, и очень способных, выбирающих высокий уровень образования. Образование не увеличивает трудоспособность индивидов, так что размер заработной платы определяется только способностями индивида. Если бы мы имели данные по заработной плате, способностям и образованию для выборки, включающей индивидов обоих типов, могли бы мы тогда проверить гипотезу о том, что уровень образования сам по себе не приводит к более высокой заработной плате? **Указание:** *Возможно наличие проблемы мультиколлинеарности.*

Набор задач для главы 3

Аналитические упражнения

1. Докажите, что симметричная и идемпотентная матрица положительно полуопределена.

2. Мы хотим доказать утверждение 3.4. Чтобы избежать ненужных осложнений, рассмотрим случай $K = 1$ и $L = 1$ (так что $\boldsymbol{x}_i$ и $\boldsymbol{z}_i$ скаляры). То есть (3.5.5) принимает вид:
$$\widehat{\varepsilon}_i^2 = \varepsilon_i^2 - 2(\widehat{\delta} - \delta)z_i\varepsilon_i + (\widehat{\delta} - \delta)^2 z_i^2,$$
а $\widehat{\boldsymbol{S}}$ в (3.5.10) принимает вид:
$$\frac{1}{n}\sum_{i=1}^{n} x_i^2\varepsilon_i^2 - 2(\widehat{\delta} - \delta)\frac{1}{n}\sum_{i=1}^{n} z_i x_i^2 \varepsilon_i + (\widehat{\delta} - \delta)^2 \frac{1}{n}\sum_{i=1}^{n} z_i^2 x_i^2. \qquad (*)$$

(a) Покажите, что первая составляющая в (*) сходится по вероятности к S.

(b) Используя неравенство Коши — Шварца

$$\mathrm{E}(|X \cdot Y|) \le \sqrt{\mathrm{E}(X^2) \cdot \mathrm{E}(Y^2)},$$

покажите, что $\mathrm{E}(z_i x_i^2 \varepsilon_i)$ существует и конечно. **Указание:** $z_i x_i^2 \varepsilon_i$ *является произведением* $x_i \varepsilon_i$ *и* $z_i x_i$.

Факт: $\mathrm{E}(x)$ существует и конечно тогда и только тогда, когда $\mathrm{E}(|x|) < \infty$. Где используется предположение 3.6?

(c) Покажите, что вторая составляющая в (*) сходится к нулю по вероятности.

(d) Покажите, что третья составляющая в (*) сходится к нулю по вероятности.

3. Мы хотим доказать алгебраический результат (3.5.11). Для этого требуется доказать, что $\boldsymbol{A} - \boldsymbol{B}$ является положительно полуопределенной матрицей, где

$$\boldsymbol{A} \equiv (\boldsymbol{\Sigma}'_{\boldsymbol{xz}} \boldsymbol{W} \boldsymbol{\Sigma}_{\boldsymbol{xz}})^{-1} \boldsymbol{\Sigma}'_{\boldsymbol{xz}} \boldsymbol{W} \boldsymbol{S} \boldsymbol{W} \boldsymbol{\Sigma}_{\boldsymbol{xz}} (\boldsymbol{\Sigma}'_{\boldsymbol{xz}} \boldsymbol{W} \boldsymbol{\Sigma}_{\boldsymbol{xz}})^{-1}$$

и

$$\boldsymbol{B} \equiv (\boldsymbol{\Sigma}'_{\boldsymbol{xz}} \boldsymbol{S}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{xz}})^{-1}.$$

Из матричной алгебры известно, что если $\boldsymbol{A}$ и $\boldsymbol{B}$ положительно определенные и симметричные матрицы, то $\boldsymbol{A} - \boldsymbol{B}$ будет положительно полуопределенной тогда и только тогда, когда $\boldsymbol{B}^{-1} - \boldsymbol{A}^{-1}$ положительно полуопределена. Тем самым нужно доказать, что

$$\boldsymbol{Q} \equiv \boldsymbol{\Sigma}'_{\boldsymbol{xz}} \boldsymbol{S}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{xz}} - \boldsymbol{\Sigma}'_{\boldsymbol{xz}} \boldsymbol{W} \boldsymbol{\Sigma}_{\boldsymbol{xz}} (\boldsymbol{\Sigma}'_{\boldsymbol{xz}} \boldsymbol{W} \boldsymbol{S} \boldsymbol{W} \boldsymbol{\Sigma}_{\boldsymbol{xz}})^{-1} \boldsymbol{\Sigma}'_{\boldsymbol{xz}} \boldsymbol{W} \boldsymbol{\Sigma}_{\boldsymbol{xz}}$$

является положительно полуопределенной.

(a) Поскольку $\boldsymbol{S}$ $(K \times K)$ положительно определена, то существует невырожденная матрица $\boldsymbol{C}$ $(K \times K)$ такая, что $\boldsymbol{C}'\boldsymbol{C} = \boldsymbol{S}^{-1}$. Проверьте, что $\boldsymbol{Q}$ можно записать в виде:

$$\boldsymbol{Q} = \boldsymbol{H}' \boldsymbol{M}_{\boldsymbol{G}} \boldsymbol{H},$$

где

$$\boldsymbol{H} = \boldsymbol{C} \boldsymbol{\Sigma}_{\boldsymbol{xz}}, \quad \boldsymbol{M}_{\boldsymbol{G}} = \boldsymbol{I}_K - \boldsymbol{G}(\boldsymbol{G}'\boldsymbol{G})^{-1}\boldsymbol{G}', \quad \boldsymbol{G} = \boldsymbol{C}'^{-1} \boldsymbol{W} \boldsymbol{\Sigma}_{\boldsymbol{xz}}.$$

Указание: $\boldsymbol{C}^{-1}\boldsymbol{C}'^{-1} = \boldsymbol{S}$.

(b) Покажите, что Q положительно полуопределена. **Указание:** *Сначала покажите, что M_G симметрична и идемпотентна. Как было показано в упражнении 1, симметричная идемпотентная матрица положительно полуопределена.*

4. Предположим, что z_i является строгим подмножеством x_i. Что меньше в матричном смысле: асимптотическая дисперсия МНК-оценки или асимптотическая дисперсия эффективной двухшаговой GMM-оценки? **Указание:** *Без потери общности можно предположить, что z_i — первые L элементов K-мерного вектора x_i. Если матрица W $(K \times K)$ дана как*

$$W = \begin{bmatrix} A^{-1} & 0 \\ 0 & 0 \end{bmatrix},$$

где

$$A \equiv \mathrm{E}(\varepsilon_i^2 z_i z_i') = \text{главная } (L \times L)\text{-подматрица матрицы } S,$$

то асимптотическая дисперсия МНК-оценки может быть записана как в формуле (3.5.1). *Если вы сделали упражнение 3, то можете проверить, что* (3.5.11) *не требует положительной определенности W если $\Sigma_{xz}' W S W \Sigma_{xz}$ невырождена.*

5. (Необязательное упражнение.) Докажите утверждение 3.6, предпринимая следующие шаги:

(a) Покажите, что $g_n(\widehat{(}\delta(\widehat{S}^{-1})) = \widehat{B}\overline{g}$, где

$$\widehat{B} \equiv I_K - S_{xz}(S_{xz}'\widehat{S}S_{xz})^{-1}S_{xz}'\widehat{S}^{-1}.$$

(b) Поскольку $\widehat{S}$ положительно определена, существует невырожденная матрица C $(K \times K)$ такая, что $C'C = \widehat{S}^{-1}$. Пусть $A \equiv CS_{xz}$. Покажите, что $\widehat{B}'\widehat{S}^{-1}\widehat{B} = C'MC$ где $M \equiv I_K - A(A'A)^{-1}A'$. Какой ранг у M? **Указание:** *Ранг идемпотентной матрицы равен ее следу.*

(c) Пусть $v \equiv \sqrt{n}C\overline{g}$. Покажите, что $v \to_{\mathrm{d}} N(0, I_K)$.

(d) Покажите, что $J(\widehat{\delta}(\widehat{S}^{-1}), \widehat{S}^{-1}) = v'Mv$. Каково его асимптотическое распределение?

6. Покажите, что J-статистика из утверждения 3.6 численно равна:

$$n \cdot s_{xy}'\widehat{B}'\widehat{S}^{-1}\widehat{B}s_{xy},$$

где $\widehat{B}$ определена в упражнении 5a). **Указание:** *Из упражнения 5 следует $J = n\overline{g}'\widehat{B}'\widehat{S}^{-1}\widehat{B}\overline{g}$. Покажите, что $\widehat{B}\overline{g} = \widehat{B}s_{xy}$.*

7. (Необязательное упражнение.) Докажите утверждение 3.7, предпринимая следующие шаги. Обозначения, не вводимые в этом упражнении, те же самые, что и в упражнении 5.

   (a) Докажите, что $\boldsymbol{g}_{1n}(\overline{\boldsymbol{\delta}}) = \widehat{\boldsymbol{B}}_1\overline{\boldsymbol{g}}_1$ и $\widehat{\boldsymbol{B}}_1'(\widehat{\boldsymbol{S}}_{11})^{-1}\widehat{\boldsymbol{B}}_1 = \boldsymbol{C}_1'\boldsymbol{M}_1\boldsymbol{C}_1$,

$$\text{где}\quad \overline{\boldsymbol{g}}_1 \equiv \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{i1}\cdot\varepsilon_i,$$

$$\widehat{\boldsymbol{B}}_1 \equiv \boldsymbol{I}_{K_1} - \boldsymbol{S}_{\boldsymbol{x}_1\boldsymbol{z}}[\boldsymbol{S}_{\boldsymbol{x}_1\boldsymbol{z}}'(\widehat{\boldsymbol{S}}_{11})^{-1}\boldsymbol{S}_{\boldsymbol{x}_1\boldsymbol{z}}]^{-1}\boldsymbol{S}_{\boldsymbol{x}_1\boldsymbol{z}}'(\widehat{\boldsymbol{S}}_{11})^{-1},$$

$$\boldsymbol{C}_1'\boldsymbol{C}_1 \equiv (\widehat{\boldsymbol{S}}_{11})^{-1},$$

$$\boldsymbol{M}_1 \equiv \boldsymbol{I}_{K_1} - \boldsymbol{A}_1(\boldsymbol{A}_1'\boldsymbol{A}_1)\boldsymbol{A}_1',$$

$$\boldsymbol{A}_1 \equiv \boldsymbol{C}_1\boldsymbol{S}_{\boldsymbol{x}_1\boldsymbol{z}}.$$

   (b) Докажите, что $\operatorname{rank}(\boldsymbol{M}) = K - L$, $\operatorname{rank}(\boldsymbol{M}_1) = K_1 - L$.

   (c) В предыдущем упражнении было доказано, что $J = \boldsymbol{v}'\boldsymbol{M}\boldsymbol{v}$, где $\boldsymbol{v} \equiv \sqrt{n}\boldsymbol{C}\overline{\boldsymbol{g}}$. Покажите, что $J_1 = \boldsymbol{v}_1'\boldsymbol{M}_1\boldsymbol{v}_1$, где $\boldsymbol{v}_1 \equiv \sqrt{n}\boldsymbol{C}_1\overline{\boldsymbol{g}}_1$.

   Для продолжения определим $\boldsymbol{F}$ ($K \times K_1$) как

$$\boldsymbol{F} = \begin{bmatrix}\boldsymbol{I}_{K_1}\\ \boldsymbol{0}\end{bmatrix}\begin{matrix}\}\quad K_1 \text{ строк},\\ \}\quad K - K_1 \text{ строк}.\end{matrix}$$

   Тогда $\boldsymbol{x}_{i1} = \boldsymbol{F}'\boldsymbol{x}_i$, $\boldsymbol{S}_{\boldsymbol{x}_1\boldsymbol{z}} = \boldsymbol{F}'\boldsymbol{S}_{\boldsymbol{x}\boldsymbol{z}}$, $\overline{\boldsymbol{g}}_i = \boldsymbol{F}'\overline{\boldsymbol{g}}_i$.

   (d) Докажите, что $J - J_1 = \boldsymbol{v}'(\boldsymbol{M} - \boldsymbol{D})\boldsymbol{v}$, где

$$\boldsymbol{D} = \boldsymbol{C}'^{-1}\boldsymbol{F}\boldsymbol{C}_1'\boldsymbol{M}_1\boldsymbol{C}_1\boldsymbol{F}'\boldsymbol{C}^{-1}.$$

   (e) Докажите, что $\boldsymbol{D}$ симметрична и идемпотентна с рангом $K_1 - L$. **Указание:**

$$\boldsymbol{F}'\boldsymbol{C}^{-1}\boldsymbol{C}'^{-1}\boldsymbol{F} = \boldsymbol{F}'(\boldsymbol{C}\boldsymbol{C}')^{-1}\boldsymbol{F} = \boldsymbol{F}'\widehat{\boldsymbol{S}}\boldsymbol{F} = \widehat{\boldsymbol{S}}_{11},$$

$$\boldsymbol{C}_1\widehat{\boldsymbol{S}}_{11}\boldsymbol{C}_1' = \boldsymbol{C}_1(\boldsymbol{C}_1'\boldsymbol{C}_1)^{-1}\boldsymbol{C}_1' = \boldsymbol{I}_{K_1}.$$

   (f) Докажите, что $\boldsymbol{A}'\boldsymbol{D} = \boldsymbol{0}$. **Указание:**

$$\boldsymbol{A}'\boldsymbol{D} = (\boldsymbol{S}_{\boldsymbol{x}\boldsymbol{z}}'\boldsymbol{C}')(\boldsymbol{C}'^{-1}\boldsymbol{F}\boldsymbol{C}_1'\boldsymbol{M}_1\boldsymbol{C}_1\boldsymbol{F}'\boldsymbol{C}^{-1}) = (\boldsymbol{S}_{\boldsymbol{x}\boldsymbol{z}}'\boldsymbol{F}\boldsymbol{C}_1'\boldsymbol{M}_1)(\boldsymbol{C}_1\boldsymbol{F}'\boldsymbol{C}^{-1}),$$

$$\boldsymbol{S}_{\boldsymbol{x}\boldsymbol{z}}'\boldsymbol{F}\boldsymbol{C}_1' = \boldsymbol{A}_1',$$

   так что

$$\boldsymbol{S}_{\boldsymbol{x}\boldsymbol{z}}'\boldsymbol{F}\boldsymbol{C}_1'\boldsymbol{M}_1 = \boldsymbol{A}_1'\boldsymbol{M}_1.$$

   (g) Докажите, что $\boldsymbol{M} - \boldsymbol{D}$ симметрична и идемпотентна с рангом $K - K_1$.

(h) Докажите желаемый результат: $J - J_1 \to_d \chi^2(K - K_1)$.

(i) На шаге d) установлено, что $C = n\overline{g}'C'(M - D)C\overline{g}$. Покажите, что C можно переписать в виде:

$$n \cdot s'_{xy} C'(M - D) C s_{xy}.$$

Указание: $g_n(\widehat{\delta}) = \widehat{B}s_{xy}$. *Таким образом, численное значение C было бы таким же, если бы мы изначально заменили $\overline{g}$ на s_{xy} в a).*

(j) Покажите, что $M - D$ положительно полуопределена. Поэтому квадратичная форма C будет неотрицательной на любой выборке.

8. (Необязательное упражнение, доказательство численной эквивалентности в утверждении 3.8.) В эффективной GMM с ограничениями функция J минимизируется при ограничении $R\widetilde{\delta} = r$. Сейчас не будем требовать, чтобы $\widehat{W}$ была равной $\widehat{S}^{-1}$, и запишем Лагранжиан в виде:

$$\mathcal{L} = n \cdot (s_{xy} - S_{xz}\widetilde{\delta})'\widehat{W}(s_{xy} - S_{xz}\widetilde{\delta}) + \lambda'(R\widetilde{\delta} - r),$$

где λ — $\#r$-мерный вектор множителей Лагранжа (напоминаем: r имеет размер $\#r \times 1$, R — $\#r \times L$, $\widetilde{\delta}$ — $L \times 1$). Пусть $\overline{\delta}$ — решение задачи минимизации с ограничениями (если $\widehat{W} = \widehat{S}^{-1}$, то $\overline{\delta}$ — эффективная GMM-оценка).

(a) Пусть $\widehat{\delta}(\widehat{W})$ — GMM-оценка без ограничений, минимизирующая $J(\widetilde{\delta}, \widehat{W})$. Покажите, что

$$\overline{\delta} = \widehat{\delta}(\widehat{W}) - (S'_{xz}\widehat{W}S_{xz})^{-1}R'[R(S'_{xz}\widehat{W}S_{xz})^{-1}R']^{-1}[R\widehat{\delta}(\widehat{W}) - r].$$
$$\lambda = 2n \cdot [R(S'_{xz}\widehat{W}S_{xz})^{-1}R']^{-1}[R\widehat{\delta}(\widehat{W}) - r].$$

Указание: *Условия первого порядка имеют вид:*

$$-2nS'_{xz}\widehat{W}s_{xz} + 2n(S'_{xz}\widehat{W}S_{xz})\overline{\delta} + R'\lambda = 0.$$

Совмещая их с ограничением $R\widetilde{\delta} = r$, решим полученную систему относительно λ и $\overline{\delta}$.

(b) Покажите, что

$$J(\widetilde{\delta}, \widehat{W}) = J(\widehat{\delta}(\widehat{W}), \widehat{W}) + n \cdot (\widehat{\delta}(\widehat{W}) - \overline{\delta})'(S'_{xz}\widehat{W}S_{xz})(\widehat{\delta}(\widehat{W}) - \overline{\delta}).$$

Указание:

$$\begin{aligned} J(\widetilde{\delta}, \widehat{W}) &= n \cdot (s_{xy} - S_{xz}\overline{\delta})'\widehat{W}(s_{xy} - S_{xz}\overline{\delta}) = \\ &= n \cdot \left[(s_{xy} - S_{xz}\widehat{\delta}(\widehat{W})) + S_{xz}(\widehat{\delta}(\widehat{W}) - \widetilde{\delta})\right]'\widehat{W} \times \\ &\times \left[(s_{xy} - S_{xz}\widehat{\delta}(\widehat{W})) + S_{xz}(\widehat{\delta}(\widehat{W}) - \overline{\delta})\right]. \end{aligned}$$

Используйте условия первого порядка для GMM без ограничений:

$$\mathbf{S}'_{xz}\widehat{\mathbf{W}}(\mathbf{s}_{xy} - \mathbf{S}_{xz}\widehat{\boldsymbol{\delta}}(\widehat{\mathbf{W}})) = \mathbf{0}.$$

(c) Пусть $\widehat{\mathbf{W}} = \widehat{\mathbf{S}}^{-1}$. Покажите, что статистика Вальда W (3.5.8) численно равна LR (3.7.2).

9. (Принцип Хаусмана в GMM.) Пусть $\widehat{\boldsymbol{\delta}}_1 \equiv \widehat{\boldsymbol{\delta}}(\widehat{\mathbf{W}}_1)$ и $\widehat{\boldsymbol{\delta}}_2 \equiv \widehat{\boldsymbol{\delta}}(\widehat{\mathbf{W}}_2)$ — две GMM-оценки с разными взвешивающими матрицами $\widehat{\mathbf{W}}_1$ и $\widehat{\mathbf{W}}_2$.

   (a) Покажите, что

$$\begin{bmatrix} \sqrt{n}(\widehat{\boldsymbol{\delta}}_1 - \boldsymbol{\delta}) \\ \sqrt{n}(\widehat{\boldsymbol{\delta}}_2 - \boldsymbol{\delta}) \end{bmatrix} \xrightarrow[\mathrm{d}]{} N\left(\mathbf{0}, \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}\right),$$

   где

$$\begin{aligned}
\mathbf{A}_{11} &= \mathbf{Q}_1^{-1}\boldsymbol{\Sigma}'_{xz}\mathbf{W}_1\mathbf{S}\mathbf{W}_1\boldsymbol{\Sigma}_{xz}\mathbf{Q}_1^{-1}, \\
\mathbf{A}_{12} &= \mathbf{Q}_1^{-1}\boldsymbol{\Sigma}'_{xz}\mathbf{W}_1\mathbf{S}\mathbf{W}_2\boldsymbol{\Sigma}_{xz}\mathbf{Q}_2^{-1}, \\
\mathbf{A}_{21} &= \mathbf{Q}_2^{-1}\boldsymbol{\Sigma}'_{xz}\mathbf{W}_2\mathbf{S}\mathbf{W}_1\boldsymbol{\Sigma}_{xz}\mathbf{Q}_1^{-1}, \\
\mathbf{A}_{22} &= \mathbf{Q}_2^{-1}\boldsymbol{\Sigma}'_{xz}\mathbf{W}_2\mathbf{S}\mathbf{W}_2\boldsymbol{\Sigma}_{xz}\mathbf{Q}_2^{-1}, \\
\mathbf{Q}_1 &= \boldsymbol{\Sigma}'_{xz}\mathbf{W}_1\boldsymbol{\Sigma}_{xz}, \\
\mathbf{Q}_2 &= \boldsymbol{\Sigma}'_{xz}\mathbf{W}_2\boldsymbol{\Sigma}_{xz}, \\
\operatorname*{plim}_{n\to\infty}\widehat{\mathbf{W}}_j &= \mathbf{W}_j \ (j = 1, 2).
\end{aligned}$$

   (b) Пусть $\mathbf{q} \equiv \widehat{\boldsymbol{\delta}}_1 - \widehat{\boldsymbol{\delta}}_2$. Покажите, что $\sqrt{n}\mathbf{q} \to_{\mathrm{d}} N(\mathbf{0}, \mathrm{Avar}(\mathbf{q}))$, где

$$\mathrm{Avar}(\mathbf{q}) = \mathbf{A}_{11} + \mathbf{A}_{22} - \mathbf{A}_{12} - \mathbf{A}_{21}.$$

   (c) Пусть $\widehat{\mathbf{W}}_2 = \widehat{\mathbf{S}}^{-1}$, так что $\widehat{\boldsymbol{\delta}}_2$ — эффективная GMM-оценка. Покажите, что

$$\mathrm{Avar}(\mathbf{q}) = \mathbf{A}_{11} - (\boldsymbol{\Sigma}'_{xz}\mathbf{S}^{-1}\boldsymbol{\Sigma}_{xz})^{-1} = \mathrm{Avar}(\widehat{\boldsymbol{\delta}}(\widehat{\mathbf{W}}_1)) - \mathrm{Avar}(\widehat{\boldsymbol{\delta}}(\widehat{\mathbf{S}}^{-1})).$$

10. (Слабые инструменты, вопрос, восходящий к М. Уотсону.) Рассмотрим модель

$$y_i = z_i\delta + \varepsilon_i \ (i = 1, 2, \ldots, n),$$

где y_i, z_i и ε_i — скалярные случайные величины. Пусть x_i — скалярная инструментальная переменная, связанная с z_i следующим уравнением:

$$z_i = x_i\beta + v_i.$$

Пусть

$$\boldsymbol{\eta}_i \equiv \begin{bmatrix} \varepsilon_i \\ v_i \end{bmatrix}, \ \mathbf{g}_i \equiv \boldsymbol{\eta}_i \cdot x_i \equiv \begin{bmatrix} g_{1i} \\ g_{2i} \end{bmatrix} = \begin{bmatrix} x_i\varepsilon_i \\ x_i v_i \end{bmatrix}.$$

Предположим, что

(1) $\{x_i, \eta_i\}$ — эргодический стационарный процесс.

(2) $\boldsymbol{g}_i$ — последовательность мартингал-разностей с положительно определенной матрицей $\mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_i') = \boldsymbol{S}$.

(3) $\mathrm{E}(x_i^2) = \sigma_x^2 > 0$.

(4) $\beta \neq 0$.

Пусть $\widehat{\delta} \equiv s_{xz}^{-1} s_{xy}$, где

$$s_{xz} \equiv \frac{1}{n}\sum_{i=1}^{n} x_i z_i, \quad s_{xy} \equiv \frac{1}{n}\sum_{i=1}^{n} x_i y_i.$$

(a) Докажите, что $\sigma_{xz} \equiv \mathrm{E}(x_i z_i) \neq 0$ (то есть ранговое условие идентификации выполнено).

(b) Докажите, что $\widehat{\delta} \to_{\mathrm{p}} \delta$.

Мы хотим сейчас рассмотреть случай, когда ранговое условие удовлетворяется едва-едва, то есть σ_{xz} очень близко к нулю. Одним из способов добиться этого является замена предположения 4) на 4′) $\beta = 1/\sqrt{n}$.

В оставшейся части этой задачи будем считать, что выполнены предположения (1)–(3) и 4′) (это предположение и последующие вопросы основаны на [Staiger and Stock, 1997]). Заметим, что $\widehat{\delta} - \delta = s_{xz}^{-1}\overline{g}_1$, где $\overline{g}_1$ — выборочное среднее $g_{1i} \equiv x_i\varepsilon_i$.

(c) Покажите, что $s_{xz} \to_{\mathrm{p}} 0$.

(d) Покажите, что $\sqrt{n}s_{xz} \to_{\mathrm{d}} \sigma_x^2 + a$, где a имеет распределение $N(0, s_{22})$ и s_{22} — (2,2) элемент $\boldsymbol{S}$.

(e) Покажите, что $\widehat{\delta} - \delta \to_{\mathrm{d}} (\sigma_x^2 + a)^{-1} b$, где (a, b) имеют совместное нормальное распределение с нулевым средним и ковариационной матрицей $\boldsymbol{S}$.

(f) Состоятельна ли $\widehat{\delta}$?

Эмпирические упражнения

Перед выполнением прочтите [Griliches, 1976]. Мы будем оценивать уравнение заработной платы, оцененное Грилихесом, используя выборку из NLS-Y [Blackburn and Neumark, 1992][1]. NLS-Y представляет **панельные данные** (*panel data*), в которых в различные моменты времени были опрошены одни и те же молодые люди (мы не будем, однако, использовать в этом упражнении панельную структуру этого множества данных).

[1] См.: [Blackburn and Neumark, 1992, Section III] для более детального описания данных.

Выборка содержит информацию об этих индивидах в два момента времени: в самом раннем году с доступной информацией о заработной плате и в 1980 году. В файле GRILIC.ASC представлены данные *RNS, RNS80, MRT, MRT80, SMSA, SMSA80, MED, IQ, KWW, YEAR, AGE, AGE80, S, S80, EXPR, EXPR80, TENURE, TENURE80, LW* и *LW80* (в таком порядке, столбцы представляют переменные). Переменная *YEAR* — год, соответствующий первому моменту времени. Переменные без «80» относятся к первому моменту времени, с «80» — к 1980 году. Определения переменных для первого момента:

RNS — дамми для проживания в южных штатах;

MRT — дамми для семейного статуса (1 для женатых);

SMSA — дамми для проживания в крупных городах;

MED — образование матери в годах;

KWW — оценка в тесте *Knowledge of the World of Work*;

IQ — уровень IQ;

AGE — возраст индивида;

S — образование в годах;

EXPR — опыт работы в годах;

TENURE — срок пребывания в должности в годах:

LW — логарифм заработной платы.

Выборка Блэкмана — Ньюмарка содержит 815 наблюдений после исключения темнокожих индивидов. Также удалены случаи с отсутствующей информацией об образовании матери, что сократило выборку до 758 наблюдений.

(a) Рассчитайте средние и стандартные отклонения всех представленных переменных (в том числе для 1980 года) и подготовьте таблицу наподобие табл. 1 Грилихеса. Также рассчитайте корреляцию между *IQ* и *S* (используйте команду `MSD(CORR)` для TSP или `CMOM (PRINT,CORR)` для RATS).

Поскольку год первого наблюдения заработной платы различен для всех индивидов, величина заработной платы будет включать эффект года. Создайте фиктивные переменные для каждого года с 1966-го по 1973-й. (Замечание: в 1972 году наблюдений нет.) Они будут включены в уравнение заработной платы для учета эффекта года.

(b) Рассмотрим уравнение заработной платы (опуская индекс индивида i)

$$LW = \beta S + \gamma IQ + \boldsymbol{\delta}'\boldsymbol{h} + \varepsilon,$$

где LW — логарифм заработной платы, S — образование в годах, и

$$\boldsymbol{h} \equiv (EXPR, TENURE, RNS, SMSA, \text{указанные дамми})'.$$

(Включать константу нет необходимости, так как она является линейной комбинацией дамми-переменных.) Подготовьте таблицу, подобную табл. 3.2. При 2SLS-оценивании уравнения заработной платы список инструментов должен состоять из предопределенных регрессоров (S и $\boldsymbol{h}$) и исключенных предопределенных переменных (*MED, KWW, MRT, AGE*). Согласуются ли относительные величины трех оценок коэффициента для образования с предсказаниями, основанными на смещениях, вызываемых пропуском переменной и ошибками в переменных?

***Совет при работе с пакетом* TSP:** Для выполнения 2SLS используйте команду `INST` или `2SLS`.

***Совет при работе с пакетом* RATS:** Используйте команду `LINREG` с опцией `INST`. Для определения списка инструментов используйте команду `INSTRUMENTS`.

(c) Для вашего 2SLS-оценивания уравнения заработной платы из пункта b) рассчитайте статистику Саргана (она должна быть равна 87,655). Каким должно быть число степеней свободы? Вычислите p-значение.

***Совет при работе с пакетом* TSP**: Статистика Саргана может быть рассчитана как `@phi/(@ssr/@nob)`, где переменные с «@» являются результатами выполнения команд TSP. `@nob` — число наблюдений, `@phi` — $\widehat{\boldsymbol{\varepsilon}}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\varepsilon}}$, где $\widehat{\boldsymbol{\varepsilon}}$ — остатки 2SLS-оценивания, `@ssr` — $\widehat{\boldsymbol{\varepsilon}}'\widehat{\boldsymbol{\varepsilon}}$.

***Совет при работе с пакетом* RATS:** Статистика Саргана может быть рассчитана как `%uzwzu/(%rss/%nobs)`, где переменные с «%» являются результатами команд RATS. `%nobs` — число наблюдений, `%uzwzu` — $\widehat{\boldsymbol{\varepsilon}}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\varepsilon}}$, `%rss` — $\widehat{\boldsymbol{\varepsilon}}'\widehat{\boldsymbol{\varepsilon}}$.

(d) Получите 2SLS-оценку, реально оценивая две регрессии. Проверьте, что полученные на втором шаге стандартные ошибки *отличаются* от полученных в b).

(e) Грилихес упоминает, что образование также может быть эндогенной переменной. Как он аргументирует это? Оцените уравнение заработной платы методом 2SLS, рассматривая S и IQ как эндогенные переменные. Что происходит с коэффициентом при образовании? Как можно объяснить различие оценок коэффициента при образовании здесь и в пункте b)? Вычислите статистику Саргана (она должна быть равна 13,268) и ее p-значение.

(f) Оцените уравнение заработной платы с помощью GMM, рассматривая образование как предопределенную переменную аналогично пункту b). Меньше ли стандартные ошибки 2SLS, чем стандартные ошибки GMM? Проверьте предопределенность образования с помощью C-статистики. Для оценки C нужно получить две GMM-оценки с и без образования в качестве инструмента. Используйте одно и то же $\widehat{S}$. Для расчета $\widehat{S}$ используйте остатки 2SLS из пункта b). C-статистика должна быть равной 58,168.

***Совет при работе с пакетом* TSP:** Вычислить $\widehat{S}$ довольно затруднительно. Следующий код, хотя и не слишком элегантный, выполнит необходимый расчет. Здесь `res0` — остатки 2SLS из пункта b).

```
smpl 1 758;
x0=res0*s;
x1=res0*expr;
x2=res0*tenure;
x3=res0*rns;
x4=res0*smsa;
x5=res0*y66;
...
x11=res0*y73;
x12=res0*med;
x13=res0*kww;
x14=res0*mrt;
x15=res0*age;
msd(noprint,moment) x0-x15;
```

`@mom` — результат выполнения команды `MSD` — и есть $\widehat{S}$. Для реализации GMM используйте команду `GMM`. Опция `COVOC=matrix name` заставляет эту команду использовать заданную матрицу как $\widehat{S}$. J-статистика рассчитывается как `@nob*@phi`.

***Совет при работе с пакетом* RATS:** `LINREG` с опциями `INST` и `ROBUSTERRORS` реализует GMM для линейных уравнений. Опция `WMATRIX=matrixname` требует $\widehat{S}^{-1}$, а не $\widehat{S}$. Для вычисления $\widehat{S}^{-1}$ сделайте следующее. Здесь `res0` — остатки 2SLS из пункта b).

```
mcov / res0
# s expr tenure rns smsa y66 ...y73
```

```
  med kww mrt age
compute sinv=inv(%cmom)
```

`sinv` является $\widehat{S}^{-1}$. J-статистика задается как `%uzwzu`.

(g) Вернитесь к уравнению заработной платы, которое вы оценили в пункте e) методом 2SLS, считая уровень обучения эндогенным. Предметом беспокойства является большое значение статистики Саргана (и J-статистики из пункта f)). Исключим MED и KWW из списка инструментов. Удовлетворяется ли еще порядковое условие идентификации? Что происходит с 2SLS-оценками? (Коэффициент при образовании будет равен -529.3%!) В чем, по-вашему, проблема? **Указание:** *Имеется два эндогенных регрессора S и IQ. MRT и AGE, единственные исключенные предопределенные переменные. Если MRT и AGE не имеют предсказательной силы в регрессии первого шага S и IQ на инструменты, то тогда подобранные значения S и IQ будут весьма близки к линейным комбинациям включенных предопределенных переменных. Это приведет к проблеме мультиколлинеарности в регрессии второго шага. Убедитесь в предсказательной силе MRT и AGE в регрессии первого шага.*

Ответы на избранные вопросы

Аналитические упражнения

4. Асимптотическая дисперсия МНК-оценки равна (после подстановки $\boldsymbol{x} = \boldsymbol{z}$ в (3.5.1))

$$\Sigma_{zz}^{-1} A \Sigma_{zz}^{-1} = (\Sigma_{zz} A^{-1} \Sigma_{zz})^{-1},$$

где $\boldsymbol{A} \equiv \mathrm{E}(\varepsilon_i^2 \boldsymbol{z}_i \boldsymbol{z}_i')$. Асимптотическая дисперсия двухшагового GMM (согласно (3.5.13)) равна:

$$(\Sigma_{xz}' S^{-1} \Sigma_{xz})^{-1},$$

где $\boldsymbol{S} \equiv \mathrm{E}(\varepsilon_i^2 \boldsymbol{x}_i \boldsymbol{x}_i')$. Если определить $\boldsymbol{W}$, как в указании, то

$$\boldsymbol{WSW} = \boldsymbol{W} \quad \text{и} \quad \Sigma_{xz}' W \Sigma_{xz} = \Sigma_{zz} A \Sigma_{zz}.$$

Таким образом, (3.5.1) сводится к асимптотической дисперсии МНК-оценки. По (3.5.11) она не меньше, чем $(\Sigma_{xz}' S^{-1} \Sigma_{xz})^{-1}$, — асимптотическая дисперсия двухшаговой GMM-оценки.

Таблица 3.3

Оценки параметров, выборка Блэкмана — Ньюмарка

Уравнение: зависимая переменная = LW, регрессоры: S, IQ, $EXPR$, $TENURE$ и другие контрольные переменные.

№	Метод оценки	Коэффициент при				SER	R^2	Тест на сверхидентифицируемость	Эндогенные переменные	Исключенные предопределенные переменные, кроме региональных дамми
		S	IQ	$EXPR$	$TENURE$					
1	МНК	0,070	—	0,030	0,043	0,328	0,425	—	нет	—
		(0,0067)	—	(0,0065)	(0,0075)					
2	МНК	0,062	0,0027	0,031	0,042	0,326	0,430	—	нет	—
		(0,0073)	(0,0010)	(0,0065)	(0,0075)					
3	2SLS	0,069	0,0002	0,030	0,043	0,328	—	87,6	IQ	MED, KWW,
		(0,013)	(0,0039)	(0,0066)	(0,0076)			($p = 0{,}0000$)		AGE, MRT
4	2SLS	0,172	−0,009	0,049	0,042	0,380	—	13,3	S, IQ	MED, KWW,
		(0,021)	(0,0047)	(0,0082)	(0,0088)			($p = 0{,}00131$)		AGE, MRT
5	GMM	0,176	−0,009	0,050	0,043	0,379	—	11,6	S, IQ	MED, KWW,
		(0,021)	(0,0049)	(0,0080)	(0,0095)			($p = 0{,}00303$)		AGE, MRT
6	2SLS	0,117	0,002	0,033	0,0051	0,380	—	14,9	$S80$, IQ	MED, KWW,
		(0,027)	(0,0050)	(0,0052)	(0,0029)			($p = 0{,}00058$)		$AGE80$, $MRT80$

Указание: *стандартные ошибки в скобках. Строка 6 для 1980 года.*

Эмпирические упражнения

(b) См. табл. 3.3, строки 1–3.

(e) Он указывает три причины эндогенности образования. Во-первых, ошибка измерения η может быть связана с успехами в обучении. Во-вторых, как мы аргументировали ранее, выбор S зависит от ненаблюдаемых характеристик. В-третьих, образование может быть измерено с ошибкой. Общая формула для асимптотического смещения 2SLS-оценки такова:

$$\operatorname{plim} \widehat{\delta}_{2SLS} - \boldsymbol{\delta} = (\boldsymbol{\Sigma}'_{\boldsymbol{xz}} \boldsymbol{\Sigma}^{-1}_{\boldsymbol{xx}} \boldsymbol{\Sigma}_{\boldsymbol{xz}})^{-1} \boldsymbol{\Sigma}'_{\boldsymbol{xz}} \boldsymbol{\Sigma}^{-1}_{\boldsymbol{xx}} \operatorname{E}(\boldsymbol{x}_i \cdot \varepsilon_i).$$

Если S эндогенно, то набор инструментов $\boldsymbol{x}_i$ для 2SLS-оценивания ошибочно включает в себя не предопределенную переменную. Поэтому один из элементов $\operatorname{E}(\boldsymbol{x}_i \cdot \varepsilon_i)$ не равен нулю. Тогда из формулы следует, что асимптотическое смещение может быть по крайней мере у некоторых коэффициентов.

(g) Без MED и KWW в качестве инструментов 2SLS-оценка становится неточной. Порядковое условие удовлетворяется как равенство. Вина лежит на низкой предсказательной силе MRT (посмотрите на ее t-значение в регрессиях первого шага для S и IQ).

Литература

Angrist, J., and A. Krueger, 1991, "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics*, 106, 979–1014.

Behrman, J., Z. Hrubec, P. Taubman, and T. Wales, 1980, *Socioeconomic Success: A Study of the Effects of Genetic Endowments, Family Environment, and Schooling*, Amsterdam: North-Holland.

Blackburn, M., and D. Neumark, 1992, "Unobserved Ability, Efficiency Wages, and Interindustry Wage Differentials," *Quarterly Journal of Economics*, 107, 1421–1436.

Card, D., 1995, "Earnings, Schooling, and Ability Revisited," *Research in Labor Economics*, 14, 23–48.

Card, D., and A. Krueger, 1996, "Labor Market Effects of School Quality: Theory and Evidence," Working Paper 5450, NBER.

Eichenbaum, M., L. Hansen, and K. Singleton, 1988, "A Time Series Analysis of Representative Agent Models of Consumption and Leisure Choice under Uncertainty," *Quarterly Journal of Economics*, 103, 51–78.

Friedman, M., 1957, *A Theory of the Consumption Function*, Princeton: Princeton University Press.

Gallant, R., and D. Jorgenson, 1979, "Statistical Inference for a System of Simultaneous, Non-linear, Implicit Equations in the Context of Instrumental Variable Estimation," *Journal of Economics*, 11, 275–302.

Griliches, Z., 1976, "Wages of Very Young Men," *Journal of Political Economy*, 84, S69–S85.

Haavelmo, T., 1943, "The Statistical Implications of a System of Simultaneous Equations," *Econometrica*, 11, 1–12.

Hansen, L., 1982, "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

Hausman, J., and W. Taylor, 1980, "Comparing Specification Tests and Classical Tests," MIT Working Paper No. 266.

Judge, G., W. Griffiths, R. Hill, H. Lütkepohl, and T. Lee, 1985, *The Theory and Practice of Econometrics* (2d ed.), New York: Wiley.

Mincer, J., 1958, "Investment in Human Capital and Personal Income Distribution," *Journal of Political Economy*, 66, 281–302.

Nelson, C., and R. Startz, 1990, "Some Further Results on the Exact Small Sample Properties of the Instrumental Variable Estimator," *Econometrica*, 58, 967–976.

Newey, W., 1985, "Generalized Method of Moments Specification Testing," *Journal of Economics*, 29, 229–256.

Newey, W., and D. McFadden, 1994, "Large Sample Estimation and Hypothesis Testing," Chapter 36 in R. Engle and D. McFadden (eds.), *Handbook of Econometrics*, Volume IV, New York: North-Holland.

Sargan, D., 1958, "The Estimation of Economic Relationships Using Instrumental Variables," *Econometrica*, 26, 393–415.

Searle, S., 1982, *Matrix Algebra Useful for Statistics*, New York: Wiley.

Staiger, D., and J. Stock, 1997, "Instrumental Variables Regression with Weak Instruments," *Econometrica*, 65, 557–586.

Theil, H., 1953, "Repeated Least-Squares Applied to Complete Equation Systems," MS., The Hague: Central Planning Bureau.

White, H., 1982, "Instrumental Variables Regression with Independent Observations," *Econometrica*, 50, 483–499.

———, 1984, *Asymptotic Theory for Econometricians*, Academic.

Working, E. J., 1927, "What Do 'Statistical Demand' Curves Show?" *Quarterly Journal of Economics*, 41, 212–235.

Глава 4. Обобщенный метод моментов для систем уравнений

Эта глава посвящена оцениванию обобщенным методом моментов нескольких уравнений одновременно. Зная, как оценивать обобщенным методом моментов (GMM) одно уравнение, остается сделать лишь небольшой шаг для перехода к системе уравнений, несмотря на кажущуюся сложность системы обозначений, предназначенной для контроля за каждым уравнением. Это определяется тем, что GMM-оценка нескольких уравнений может быть представлена в виде оценки обобщенным методом моментов одного уравнения, если надлежащим образом специфицировать матрицы и векторы, входящие в формулу для GMM-оценки единственного уравнения. В таком случае мы можем разрабатывать асимптотическую теорию оценивания систем уравнений обобщенным методом моментов в почти готовом виде.

Использование GMM-оценки для нескольких уравнений дает значительные выгоды. В случае условной гомоскедастичности она редуцируется к инструментальной оценке с полной информацией (FIVE), которая, в свою очередь, редуцируется к оценке трехшаговым методом наименьших квадратов (3SLS), если набор инструментов является общим для всех уравнений. Если также предположить, что все регрессоры являются предопределенными, то тогда 3SLS редуцируется к оценке кажущихся несвязанными регрессий (SUR), которая, в свою очередь, редуцируется к оценке многомерной регрессии, если все уравнения имеют один и тот же набор регрессоров.

Мы также покажем, что систему нескольких уравнений можно записать в виде системы с перекрестными ограничениями равенства коэффициентов во всех уравнениях. GMM-оценка для такой системы опять является частным случаем обычной GMM-оценки для одного уравнения. GMM-оценка, в случае когда все регрессоры являются предопределенными, а ошибки — условно гомоскедастичными, носит название оценки со случайными эффектами (RE-оценки). Таким образом, SUR и RE являются эквивалентными оценками.

Прикладной раздел данной главы посвящен оцениванию модели спроса на взаимосвязанные факторы посредством SUR/RE. Система функций спроса получается из транслогарифмической функции издержек, которая является обобщением линейной логарифмической функций издержек, рассмотренной в главе 1.

Здесь мы не рассматриваем оценивание систем уравнений методом максимального правдоподобия (ML). Подробно изложение данного метода представлено в параграфе 8.5.

Если вам мешает запутанность обозначений в этой главе, вы можете просто положить M (число уравнений) равным 2. Это никак не препятствует общности любого из результатов данной главы.

4.1. Модель с несколькими уравнениями

Чтобы иметь возможность работать с более чем одним уравнением, нам необходимо усложнить обозначения, введя дополнительный подстрочный индекс, обозначающий конкретное уравнение. Например, i наблюдение зависимой переменной m-го уравнения будет обозначаться y_{im}.

Линейность

Рассмотрим систему из M линейных уравнений, каждое из которых аналогично линейному уравнению в предположении 3.1:

Предположение 4.1 (линейность): *Имеется M линейных уравнений*

$$y_{im} = \mathbf{z}'_{im}\boldsymbol{\delta}_m + \varepsilon_{im} \quad (m = 1, 2, \ldots M; i = 1, 2, \ldots, n), \qquad (4.1.1)$$

где n — размер выборки, $\mathbf{z}_{im}$ — L_m-мерный вектор регрессоров, $\boldsymbol{\delta}_m$ — соответствующий вектор коэффициентов, ε_{im} — ненаблюдаемая ошибка в уравнении m.

Модель не включает в себя никаких предположений о корреляции ошибок **между уравнениями** (*interequation*) (или **одновременной корреляции** (*contemporaneous correlation*)). Также не постулируется никаких *априорных* ограничений на коэффициенты из разных уравнений. То есть модель не предполагает **перекрестных** ограничений на коэффициенты. Чтобы проиллюстрировать это, рассмотрим

Пример 4.1 (уравнение заработной платы): В примере Грилихеса из главы 3 мы оценивали уравнение заработной платы. В одной из спецификаций Грилихес добавляет к ней уравнение для KWW (результат теста на «знание рынка труда»):

$$LW_i = \phi_1 + \beta_1 S_i + \gamma_1 IQ_i + \pi EXPR_i + \varepsilon_{i1},$$
$$KWW_i = \phi_2 + \beta_2 S_i + \gamma_2 IQ_i + \varepsilon_{i2},$$

где (для освежения вашей памяти) LW_i — логарифм заработной платы i-го индивида в первый год опроса, S_i — количество лет обучения в школе, IQ_i — IQ, а $EXPR_i$ — стаж, измеренный в годах. В этом примере $M = 2$, $L_1 = 4$, $L_2 = 3$, $y_{i1} = LW_i$, $y_{i2} = KWW_i$ и

$$\mathbf{z}_{i1} = \begin{bmatrix} 1 \\ S_i \\ IQ_i \\ EXPR_i \end{bmatrix}, \quad \mathbf{z}_{i2} = \begin{bmatrix} 1 \\ S_i \\ IQ_i \end{bmatrix}, \quad \boldsymbol{\delta}_1 = \begin{bmatrix} \phi_1 \\ \beta_1 \\ \gamma_1 \\ \pi \end{bmatrix}, \quad \boldsymbol{\delta}_2 = \begin{bmatrix} \phi_2 \\ \beta_2 \\ \gamma_2 \end{bmatrix}. \qquad (4.1.2)$$

Ошибки ε_{i1} и ε_{i2} могут быть коррелированными. Корреляция может возникнуть, к примеру, если какая-то ненаблюдаемая переменная влияет на заработную плату и на результат теста. Также в модели нет никаких перекрестных ограничений типа $\beta_1 = \beta_2$.

Перекрестные ограничения естественно возникают в случае **панельных данных** (*panel data*), когда одно и то же соотношение может оцениваться для различных моментов времени.

Пример 4.2 (уравнение заработной платы для двух лет): Данные NLS-Y, которые были использованы в примере Грилихеса, в действительности содержат информацию и для 1980 года, так что мы можем оценить два уравнения: одно — для первого момента времени (скажем, для 1969 года) и второе — для 1980 года:

$$LW69_i = \phi_1 + \beta_1 S69_i + \gamma_1 IQ_i + \pi EXPR69_i + \varepsilon_{i1},$$
$$LW80_i = \phi_2 + \beta_2 S80_i + \gamma_2 IQ_i + \pi EXPR80_i + \varepsilon_{i2}.$$

На языке нескольких уравнений $S69_i$ (образование в 1969 году) и $S80_i$ (образование в 1980 году) суть две разные переменные. Было бы естественным допустить случай, когда коэффициенты остаются неизменными во времени: $\phi_1 = \phi_2$, $\beta_1 = \beta_2$, $\gamma_1 = \gamma_2$. Это есть набор линейных перекрестных ограничений.

Как будет показано ниже, в параграфе 4.3, тестирование (линейных и нелинейных) перекрестных ограничений является достаточно несложным делом. Оценивание при наличии ограничения, состоящего в равенстве коэффициентов в разных уравнениях, будет рассмотрено в параграфе 4.6.

Стационарность и эргодичность

Пусть x_{im} есть вектор из K_m инструментов для уравнения m. Хотя это и нечасто встречается на практике, количество инструментов K_m может меняться от уравнения к уравнению.

Пример 4.3 (продолжение примера 4.1): Предположим, что переменная IQ_i является эндогенной в обоих уравнениях, а также что имеется переменная MED_i (образование матери), являющаяся предопределенной в первом уравнении (так что $\mathrm{E}(MED_i\varepsilon_{i1}) = 0$). Тогда инструментами для уравнения для LW являются $(1, S_i, EXPR_i, MED_i)$. Если эти инструменты ортогональны также и к ε_{i2}, то уравнение для KWW может быть оценено с использованием того же набора инструментов, что и уравнение для LW. В данном примере $K_1 = K_2 = 4$ и

$$\boldsymbol{x}_{i1} = \boldsymbol{x}_{i2} = \begin{bmatrix} 1 \\ S_i \\ EXPR_i \\ MED_i \end{bmatrix}. \tag{4.1.3}$$

Модификация предположения 3.2 для системы из нескольких уравнений выглядит следующим образом.

Предположение 4.2 (эргодическая стационарность):
Пусть $\boldsymbol{w}_i$ образован различными и не являющимися константами элементами $(y_{i1}, \ldots, y_{iM}, \boldsymbol{z}_{i1}, \ldots, \boldsymbol{z}_{iM}, \boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{iM})$. Тогда $\{\boldsymbol{w}_i\}$ является совместно стационарным и эргодическим.

Данное предположение является более сильным, чем предположение о стационарности и эргодичности для каждого уравнения системы в отдельности. Даже если процесс $\{y_{im}, \boldsymbol{x}_{im}, \boldsymbol{z}_{im}\}$ является стационарным и эргодическим для каждого уравнения в отдельности, то, строго говоря, из этого необязательно следует, что более крупный процесс $\boldsymbol{w}_i$, являющийся объединением отдельных процессов, является (совместно) стационарным и эргодическим (см. пример 3.2 из главы 3, в котором рассмотрен нестационарный векторный процесс, компоненты которого являются одномерными стационарными процессами). На практике эту разницу сложно заметить, так как обычно уравнения состоят из одних и тех же переменных.

Пример 4.4 (продолжение примера 4.1): Поскольку $\boldsymbol{z}_{i1}$, $\boldsymbol{z}_{i2}$, $\boldsymbol{x}_{i1}$ и $\boldsymbol{x}_{i2}$, например, имеют несколько общих элементов, $\boldsymbol{w}_i$ будет состоять всего из шести членов: $(LW_i, KWW_i, S_i, IQ_i, EXPR_i, MED_i)$. Предположение об эргодической стационарности для первого уравнения состоит в том, что $\{LW_i, S_i, IQ_i, EXPR_i, MED_i\}$ является совместно стационарным и эргодическим. Предположение 4.2 требует, чтобы совместный процесс включал в себя KWW_i.

Условия ортогональности

Условия ортогональности для системы из M уравнений есть просто совокупность условий ортогональности для отдельных уравнений.

Предположение 4.3 (условия ортогональности): *Для каждого уравнения m K_m переменных в $\boldsymbol{x}_{im}$ являются предопределенными. То есть выполнены следующие условия ортогональности:*

$$\mathrm{E}(\boldsymbol{x}_{im} \cdot \varepsilon_{im}) = \mathbf{0} \quad (m = 1, 2, \ldots, M).$$

Таким образом, в итоге мы имеем $\sum_m K_m$ условий ортогональности. Если мы положим

$$\underset{(\sum_{m=1}^{M} K_m \times 1)}{\boldsymbol{g}_i} \equiv \begin{bmatrix} \boldsymbol{x}_{i1} \cdot \varepsilon_{i1} \\ \vdots \\ \boldsymbol{x}_{iM} \cdot \varepsilon_{iM} \end{bmatrix}, \tag{4.1.4}$$

то все эти условия ортогональности можно записать компактно как $\mathrm{E}(\boldsymbol{g}_i) = 0$. Заметим, что мы не предполагаем никаких «перекрестных» ортогональностей: например, $\boldsymbol{x}_{i1}$ не обязан быть ортогональным ε_{i2}, хотя требуется, чтобы он был ортогональным ε_{i1}. Однако если переменная включена и в $\boldsymbol{x}_{i1}$ и в $\boldsymbol{x}_{i2}$, то она будет ортогональна и ε_{i1} и ε_{i2}.

Пример 4.5 (продолжение примера 4.1): В данном примере $\sum_m K_m = 8$ $(= 4 + 4)$ и $\boldsymbol{g}_i$ есть

$$\boldsymbol{g}_i = \begin{bmatrix} \varepsilon_{i1} \\ S_i\varepsilon_{i1} \\ EXPR_i\varepsilon_{i1} \\ MED_i\varepsilon_{i1} \\ \varepsilon_{i2} \\ S_i\varepsilon_{i2} \\ EXPR_i\varepsilon_{i2} \\ MED_i\varepsilon_{i2} \end{bmatrix}.$$

Поскольку векторы $\boldsymbol{x}_{i1}$ и $\boldsymbol{x}_{i2}$ имеют один и тот же набор инструментов, каждый инструмент ортогонален и ε_{i1} и ε_{i2}.

Идентифицируемость

Установив условия ортогональности для системы из нескольких уравнений, мы можем получить условие идентифицируемости аналогично случаю одного уравнения. Модификация равенства (3.3.1) для случая нескольких уравнений выглядит следующим образом:

$$\boldsymbol{g}(\boldsymbol{w}_i; \boldsymbol{\delta}) \equiv \begin{bmatrix} \boldsymbol{x}_{i1} \cdot (y_{i1} - \boldsymbol{z}'_{i1}\boldsymbol{\delta}_1) \\ \vdots \\ \boldsymbol{x}_{iM} \cdot (y_{iM} - \boldsymbol{z}'_{iM}\boldsymbol{\delta}_M) \end{bmatrix}, \tag{4.1.5}$$

где $\boldsymbol{w}_i$ соответствует предположению 4.2, а вектор $\boldsymbol{\delta}$ без индекса есть штабелированный вектор коэффициентов

$$\underset{(\sum_{m=1}^{M} L_m \times 1)}{\boldsymbol{\delta}} = \begin{bmatrix} \boldsymbol{\delta}_1 \\ \vdots \\ \boldsymbol{\delta}_M \end{bmatrix}. \tag{4.1.6}$$

Таким образом, условия ортогональности могут быть записаны как $\mathrm{E}[\boldsymbol{g}(\boldsymbol{w}_i; \boldsymbol{\delta})] = \mathbf{0}$. Вектор коэффициентов является идентифицируемым, если $\widetilde{\boldsymbol{\delta}} = \boldsymbol{\delta}$ есть единственное решение системы уравнений

$$\mathrm{E}[\boldsymbol{g}(\boldsymbol{w}_i; \widetilde{\boldsymbol{\delta}})] = \mathbf{0}. \tag{4.1.7}$$

Используя определение (4.1.5), мы можем переписать левую часть этого уравнения, $\mathrm{E}[\boldsymbol{g}(\boldsymbol{w}_i; \widetilde{\boldsymbol{\delta}})]$, как

$$\begin{aligned}
\mathrm{E}[\boldsymbol{g}(\boldsymbol{w}_i; \widetilde{\boldsymbol{\delta}})] &\equiv \begin{bmatrix} \mathrm{E}[\boldsymbol{x}_{i1} \cdot (y_{i1} - \boldsymbol{z}'_{i1}\widetilde{\boldsymbol{\delta}}_1)] \\ \vdots \\ \mathrm{E}[\boldsymbol{x}_{iM} \cdot (y_{iM} - \boldsymbol{z}'_{iM}\widetilde{\boldsymbol{\delta}}_M)] \end{bmatrix} = \\
&= \begin{bmatrix} \mathrm{E}(\boldsymbol{x}_{i1} \cdot y_{i1}) \\ \vdots \\ \mathrm{E}(\boldsymbol{x}_{iM} \cdot y_{iM}) \end{bmatrix} - \begin{bmatrix} \mathrm{E}(\boldsymbol{x}_{i1}\boldsymbol{z}'_{i1})\widetilde{\boldsymbol{\delta}}_1 \\ \vdots \\ \mathrm{E}(\boldsymbol{x}_{iM}\boldsymbol{z}'_{iM})\widetilde{\boldsymbol{\delta}}_M \end{bmatrix} = \\
&= \begin{bmatrix} \mathrm{E}(\boldsymbol{x}_{i1} \cdot y_{i1}) \\ \vdots \\ \mathrm{E}(\boldsymbol{x}_{iM} \cdot y_{iM}) \end{bmatrix} - \begin{bmatrix} \mathrm{E}(\boldsymbol{x}_{i1}\boldsymbol{z}'_{i1}) & & \\ & \ddots & \\ & & \mathrm{E}(\boldsymbol{x}_{iM}\boldsymbol{z}'_{iM}) \end{bmatrix} \begin{bmatrix} \widetilde{\boldsymbol{\delta}}_1 \\ \vdots \\ \widetilde{\boldsymbol{\delta}}_M \end{bmatrix} \equiv \\
&\equiv \boldsymbol{\sigma}_{xy} - \boldsymbol{\Sigma}_{xz}\widetilde{\boldsymbol{\delta}},
\end{aligned} \tag{4.1.8}$$

где

$$\boldsymbol{\sigma}_{xy} \equiv \begin{bmatrix} \mathrm{E}(\boldsymbol{x}_{i1} \cdot y_{i1}) \\ \vdots \\ \mathrm{E}(\boldsymbol{x}_{iM} \cdot y_{iM}) \end{bmatrix}, \quad \boldsymbol{\Sigma}_{xz} \equiv \begin{bmatrix} \mathrm{E}(\boldsymbol{x}_{i1}\boldsymbol{z}'_{i1}) & & \\ & \ddots & \\ & & \mathrm{E}(\boldsymbol{x}_{iM}\boldsymbol{z}'_{iM}) \end{bmatrix}. \tag{4.1.9}$$

(Третье равенство в (4.1.8) получено применением формулы (А.4) из приложения А для умножения блочных матриц.) Следовательно, система уравнений, определяющая $\widetilde{\boldsymbol{\delta}}$, может быть записана как

$$\boldsymbol{\Sigma}_{xz}\widetilde{\boldsymbol{\delta}} = \boldsymbol{\sigma}_{xy}, \tag{4.1.10}$$

что *абсолютно аналогично* по форме равенству (3.3.4), определяющему $\widetilde{\boldsymbol{\delta}}$ в случае единственного уравнения!

Из обсуждения идентифицируемости в случае одного уравнения следует, что необходимым и достаточным условием является полный ранг матрицы $\boldsymbol{\Sigma}_{xz}$. Но так как матрица $\boldsymbol{\Sigma}_{xz}$ является блочно-диагональной, то это условие эквивалентно следующему.

Предположение 4.4 (ранговое условие идентифицируемости): *Для каждого* m $(= 1, 2, \ldots, M)$ *матрица* $\mathrm{E}(\boldsymbol{x}_{im}\boldsymbol{z}'_{im})$ $(K_m \times L_m)$ *имеет полный столбцовый ранг.*

Несложно понять, почему условие идентифицируемости выглядит именно так. Мы можем единственным образом определить все векторы коэффициентов $(\boldsymbol{\delta}_1 \ldots \boldsymbol{\delta}_M)$ тогда и только тогда, когда вектор $\boldsymbol{\delta}_m$ определен единственным образом, что может случиться, если и только если предположение 3.4 выполнено для всех уравнений. Ранговое условие в данном случае выглядит просто, так как нет никаких *априорных* перекрестных ограничений. Идентифицируемость для случая, когда коэффициенты одинаковы во всех уравнениях, будет рассмотрена в параграфе 4.6.

Предположение для асимптотической нормальности

Как и для случая одного уравнения, для обеспечения асимптотической нормальности условия ортогональности должны быть усилены:

Предположение 4.5 ($\boldsymbol{g}_i$ является мартингал-разностью с конечными вторыми моментами): $\{\boldsymbol{g}_i\}$ *является совместной мартингал-разностью. Матрица* $\mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}'_i)$ *невырождена.*

Те же самые замечания, которые мы делали относительно совместной эргодической стационарности, применимы и здесь: приведенное предположение является более сильным, чем просто требование выполнения аналогичного предположения для всех уравнений по отдельности. Как и в предыдущих главах, мы используем символ $\boldsymbol{S}$ для $\mathrm{Avar}(\overline{\boldsymbol{g}})$, асимптотической дисперсии $\overline{\boldsymbol{g}}$ (то есть дисперсии предельного распределения $\sqrt{n}\overline{\boldsymbol{g}}$). Согласно ЦПТ для стационарных и эргодических мартингал-разностей, эта дисперсия равна $\mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}'_i)$. Эта матрица имеет блочную структуру:

$$\boldsymbol{S} = \underset{(\sum_{m=1}^{M} K_m \times \sum_{m=1}^{M} K_m)}{\mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}'_i)} = \begin{bmatrix} \mathrm{E}(\varepsilon_{i1}\varepsilon_{i1}\boldsymbol{x}_{i1}\boldsymbol{x}'_{i1}) & \cdots & \mathrm{E}(\varepsilon_{i1}\varepsilon_{iM}\boldsymbol{x}_{i1}\boldsymbol{x}'_{iM}) \\ \vdots & & \vdots \\ \mathrm{E}(\varepsilon_{iM}\varepsilon_{i1}\boldsymbol{x}_{iM}\boldsymbol{x}'_{i1}) & \cdots & \mathrm{E}(\varepsilon_{iM}\varepsilon_{iM}\boldsymbol{x}_{iM}\boldsymbol{x}'_{iM}) \end{bmatrix}. \tag{4.1.11}$$

То есть блок (m, h) матрицы $\boldsymbol{S}$ есть $\mathrm{E}(\varepsilon_{im}\varepsilon_{ih}\boldsymbol{x}_{im}\boldsymbol{x}'_{ih})$ $(m, h = 1, 2, \ldots, M)$.

Таким образом, модель из нескольких уравнений представляет собой систему уравнений, для каждого из которых мы делаем те же предположения, которые мы делали для модели единственного уравнения, добавляя в надлежащих случаях требование совместности (такое, как совместная стационарность).

Связь с «полной» системой одновременных уравнений

Модель из нескольких уравнений, рассматриваемая во многих учебниках и называемая «полной» системой одновременных уравнений[1], включает в себя некоторые дополнительные предположения. Эти предположения *не* требуются для оценки методом GMM, однако если это представляет интерес, удовлетворить его можно в первых двух подпараграфах параграфа 8.5.

4.2. GMM для системы уравнений

Получение выражения для GMM-оценки для нескольких уравнений может быть проведено аналогично случаю одного уравнения. Пусть $\widetilde{\boldsymbol{\delta}}$ — гипотетическое значение параметра $\boldsymbol{\delta}$, и определим $\boldsymbol{g}_n(\widetilde{\boldsymbol{\delta}})$ посредством (3.4.1). Определение GMM-оценки в данном случае такое же, как и в (3.4.5), однако взвешивающая матрица $\widehat{\boldsymbol{W}}$ имеет теперь размер $\sum_m K_m \times$ $\times \sum_m K_m$. В предыдущем параграфе мы могли переписать $\mathrm{E}[\boldsymbol{g}(\boldsymbol{w}_i; \widetilde{\boldsymbol{\delta}})]$ для случая нескольких уравнений как линейную функцию от $\widetilde{\boldsymbol{\delta}}$ (см. (4.1.8)). Мы можем сделать то же самое и сейчас для выборочного аналога $\boldsymbol{g}_n(\widetilde{\boldsymbol{\delta}})$:

$$\underset{(\sum_{m=1}^{M} K_m \times 1)}{\boldsymbol{g}_n(\widetilde{\boldsymbol{\delta}})} = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{i1}\cdot(y_{i1} - \boldsymbol{z}_{i1}'\widetilde{\boldsymbol{\delta}}_1) \\ \vdots \\ \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{iM}\cdot(y_{iM} - \boldsymbol{z}_{iM}'\widetilde{\boldsymbol{\delta}}_M) \end{bmatrix} =$$

$$= \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{i1}\cdot y_{i1} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{iM}\cdot y_{iM} \end{bmatrix} - \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{i1}\boldsymbol{z}_{i1}'\widetilde{\boldsymbol{\delta}}_1 \\ \vdots \\ \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{iM}\boldsymbol{z}_{iM}'\widetilde{\boldsymbol{\delta}}_M \end{bmatrix} =$$

$$= \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{i1}\cdot y_{i1} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{iM}\cdot y_{iM} \end{bmatrix} - \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{i1}\boldsymbol{z}_{i1}' & & \\ & \ddots & \\ & & \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_{iM}\boldsymbol{z}_{iM}' \end{bmatrix} \times$$

$$\times \begin{bmatrix} \widetilde{\boldsymbol{\delta}}_1 \\ \vdots \\ \widetilde{\boldsymbol{\delta}}_M \end{bmatrix} \equiv \boldsymbol{s}_{\boldsymbol{xy}} - \boldsymbol{S}_{\boldsymbol{xz}}\widetilde{\boldsymbol{\delta}}, \qquad (4.2.1)$$

[1] Полная система одновременных уравнений содержит столько уравнений, сколько имеется эндогенных переменных. — *Прим. науч. ред. перевода.*

где

$$s_{xy} \equiv \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n x_{i1}\cdot y_{i1} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^n x_{iM}\cdot y_{iM} \end{bmatrix}, \quad S_{xz} \equiv \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n x_{i1}z'_{i1} & & \\ & \ddots & \\ & & \frac{1}{n}\sum_{i=1}^n x_{iM}z'_{iM} \end{bmatrix}. \tag{4.2.2}$$

И опять, по форме это *то же самое*, что и выражение для $g_n(\widetilde{\delta})$ для случая одного уравнения.

Отсюда следует, что результаты параграфа 3.4, касающиеся вывода GMM-оценки, применимы и в случае нескольких уравнений. В частности, простое воспроизведение (3.4.8) и (3.4.11) дает:

GMM-оценка для системы уравнений:

$$\widehat{\delta}(\widehat{W}) = (S'_{xz}\widehat{W}S_{xz})^{-1}S'_{xz}\widehat{W}s_{xy}, \tag{4.2.3}$$

ее ошибка оценивания:

$$\widehat{\delta}(\widehat{W}) - \delta = (S'_{xz}\widehat{W}S_{xz})^{-1}S'_{xz}\,\widehat{W}\overline{g}. \tag{4.2.4}$$

Особенности формулы GMM-оценки системы уравнений таковы: (i) s_{xz} является штабелированным вектором, как в (4.2.2); (ii) S_{xz} является блочно-диагональной матрицей, как в (4.2.2); (iii) соответственно, взвешивающая матрица $\widehat{W}$ имеет размер $\sum_m K_m \times \sum_m K_m$ и (iv) $\overline{g}$ в (4.2.4), выборочное среднее для g_i, является штабелированным вектором:

$$\overline{g} \equiv \frac{1}{n}\sum_{i=1}^n g_i = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n x_{i1}\cdot\varepsilon_{i1} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^n x_{iM}\cdot\varepsilon_{iM} \end{bmatrix} = g_n(\delta). \tag{4.2.5}$$

Используя все эти особенности, необходимо будет переписать формул для GMM-оценки для системы уравнений (4.2.3). Если $\widehat{W}_{mh}(K_m \times K$ есть (m,h) блок матрицы $\widehat{W}$ $(m,h = 1,2,\dots,M)$, то тогда (4.2.3) мож записать подробно как

$$\widehat{\delta}(\widehat{W}) = \begin{bmatrix} \widehat{\delta}_1(\widehat{W}) \\ \vdots \\ \widehat{\delta}_M(\widehat{W}) \end{bmatrix} = \left(\begin{bmatrix} \frac{1}{n}\sum_{i=1}^n z_{i1}x'_{i1} & & \\ & \ddots & \\ & & \frac{1}{n}\sum_{i=1}^n z_{iM}x'_{iM} \end{bmatrix} \times \right.$$

$$\times \begin{bmatrix} \widehat{W}_{11} & \cdots & \widehat{W}_{1M} \\ \vdots & & \vdots \\ \widehat{W}_{M1} & \cdots & \widehat{W}_{MM} \end{bmatrix} \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n x_{i1}z'_{i1} & & \\ & \ddots & \\ & & \frac{1}{n}\sum_{i=1}^n x_{iM}z'_{i} \end{bmatrix}$$

$$\times \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} z_{i1}x'_{i1} & & \\ & \ddots & \\ & & \frac{1}{n}\sum_{i=1}^{n} z_{iM}x'_{iM} \end{bmatrix} \begin{bmatrix} \widehat{W}_{11} & \cdots & \widehat{W}_{1M} \\ \vdots & & \vdots \\ \widehat{W}_{M1} & \cdots & \widehat{W}_{MM} \end{bmatrix} \times$$

$$\times \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} x_{i1}\cdot y_{i1} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} x_{iM}\cdot y_{iM} \end{bmatrix} = \begin{bmatrix} A_{11} & \cdots & A_{1M} \\ \vdots & & \vdots \\ A_{M1} & \cdots & A_{MM} \end{bmatrix}^{-1} \begin{bmatrix} B_1 \\ \vdots \\ B_M \end{bmatrix}, \qquad (4.2.6)$$

где

$$A_{11} = \left(\frac{1}{n}\sum_{i=1}^{n} z_{i1}x'_{i1}\right)\widehat{W}_{11}\left(\frac{1}{n}\sum_{i=1}^{n} x_{i1}z'_{i1}\right),$$

$$A_{1M} = \left(\frac{1}{n}\sum_{i=1}^{n} z_{i1}x'_{i1}\right)\widehat{W}_{1M}\left(\frac{1}{n}\sum_{i=1}^{n} x_{iM}z'_{iM}\right),$$

$$A_{M1} = \left(\frac{1}{n}\sum_{i=1}^{n} z_{iM}x'_{iM}\right)\widehat{W}_{M1}\left(\frac{1}{n}\sum_{i=1}^{n} x_{i1}z'_{i1}\right),$$

$$A_{MM} = \left(\frac{1}{n}\sum_{i=1}^{n} z_{iM}x'_{iM}\right)\widehat{W}_{MM}\left(\frac{1}{n}\sum_{i=1}^{n} x_{iM}z'_{iM}\right),$$

$$B_1 = \left(\frac{1}{n}\sum_{i=1}^{n} z_{i1}x'_{i1}\right)\widehat{W}_{11}\left(\frac{1}{n}\sum_{i=1}^{n} x_{i1}y_{i1}\right) +$$

$$+ \cdots + \left(\frac{1}{n}\sum_{i=1}^{n} z_{i1}x'_{i1}\right)\widehat{W}_{1M}\left(\frac{1}{n}\sum_{i=1}^{n} x_{iM}y_{iM}\right).$$

$$B_M = \left(\frac{1}{n}\sum_{i=1}^{n} z_{iM}x'_{iM}\right)\widehat{W}_{M1}\left(\frac{1}{n}\sum_{i=1}^{n} x_{i1}\, y_{i1}\right) +$$

$$+ \cdots + \left(\frac{1}{n}\sum_{i=1}^{n} z_{iM}x'_{iM}\right)\widehat{W}_{MM}\left(\frac{1}{n}\sum_{i=1}^{n} x_{iM}\, y_{iM}\right).$$

Для перехода к последней строке используйте формулы (A.6) и (A.7) для блочных матриц из приложения A. Если вы находите эту работу слишком сложной, просто положите $M = 2$.

4.3. Теория больших выборок

Теория больших выборок для обобщенного метода моментов (GMM) для одного уравнения была сформулирована в утверждениях 3.1–3.8. Перенос ее на модель с несколькими уравнениями — это просто вопрос механического замещения. При условии что δ, $\boldsymbol{\Sigma}_{xz}$, $\boldsymbol{S}$, g_i, s_{xz} и $\boldsymbol{S}_{xz}$ определены, как в предыдущих параграфах, и что «предположение 3.x» заменяется на «предположение 4.х», утверждения 3.1, 3.3 и 3.5–3.8 справедливы и для случая системы нескольких уравнений. Нужно лишь сделать несколько комментариев относительно их интерпретации.

(Тестирование гипотез.) В рассматриваемом сейчас случае нескольких уравнений $\boldsymbol{\delta}$ представляет собой штабелированный вектор, составленный из коэффициентов разных уравнений. Суть утверждений 3.3 и 3.8 о тестировании гипотез состоит в том, что мы можем тестировать перекрестные ограничения.

Пример 4.6 (продолжение примера 4.2): В случае состоящей из двух уравнений системы уравнений заработной платы для двух лет штабелированный вектор коэффициентов $\boldsymbol{\delta}$ есть:

$$\boldsymbol{\delta} = (\phi_1, \beta_1, \gamma_1, \pi_1, \phi_2, \beta_2, \gamma_2, \pi_2)'.$$

Было бы интересно проверить гипотезу $H_0 : \beta_1 = \beta_2$ и $\pi_1 = \pi_2$ о том, что премии за образование и стаж неизменны во времени. Эту гипотезу можно записать в виде $\boldsymbol{R\delta} = \boldsymbol{r}$, где

$$\boldsymbol{R} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{bmatrix}, \quad \boldsymbol{r} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Можно также проверять и нелинейные перекрестные ограничения; для этого следует просто использовать пункт (с) утверждения 3.3 или утверждение 3.8.

(Тест на сверхидентифицирующие ограничения.) Количество условий ортогональности равно $\sum_m K_m$, а количество коэффициентов равно $\sum_m L_m$. Соответственно, число степеней свободы J-статистики из утверждения 3.6 равно $\sum_m K_m - \sum_m L_m$, а для C-статистики из утверждения 3.7 оно же есть общее количество вызывающих сомнение инструментов из разных уравнений.

Утверждение 3.2 не может быть напрямую использовано для случая системы из нескольких уравнений, однако очевидно, как его можно адаптировать.

Утверждение 4.1 (состоятельное оценивание перекрестных моментов одновременных ошибок): *Пусть $\widehat{\boldsymbol{\delta}}_m$ — состоятельная оценка $\boldsymbol{\delta}_m$ и пусть $\widehat{\varepsilon}_{im} \equiv y_{im} - \boldsymbol{z}'_{im}\widehat{\boldsymbol{\delta}}_m$ — подразумеваемый остаток для $m = 1, 2, \ldots, M$. В предположениях 4.1 и 4.2, а также в предположении, что $\mathrm{E}(\boldsymbol{z}_{im}\boldsymbol{z}'_{ih})$ существует и конечно для всех m, h $(= 1, 2, \ldots, M)$,*

$$\widehat{\sigma}_{mh} \xrightarrow[p]{} \sigma_{mh}, \qquad \text{где} \tag{4.3.1}$$

$$\widehat{\sigma}_{mh} \equiv \frac{1}{n}\sum_{i=1}^{n} \widehat{\varepsilon}_{im}\widehat{\varepsilon}_{ih} \quad \text{и} \quad \sigma_{mh} \equiv \mathrm{E}(\varepsilon_{im}\varepsilon_{ih})$$

при условии, что $\mathrm{E}(\varepsilon_{im}\varepsilon_{ih})$ существует и конечно.

Доказательство весьма похоже на доказательство утверждения 3.2 и весьма утомительно, поэтому мы оставляем его в качестве самостоятельного упражнения.

Остается утверждение 4.2, касающееся состоятельного оценивания $\boldsymbol{S}$. Предположение, соответствующее предположению 3.6 (о конечности четвертых моментов), таково:

Предположение 4.6 (конечные четвертые моменты): $\mathrm{E}[(x_{imk} \cdot z_{ihj})^2]$ *существует и конечно для всех* k $(= 1, 2, \ldots, K_m)$, j $(= 1, 2, \ldots, L_h)$, m, h $(= 1, 2, \ldots, M)$, *где* x_{imk} *есть* k*-й элемент* $\boldsymbol{x}_{im}$ *и* z_{ihj} *есть* j*-й элемент* $\boldsymbol{z}_{ih}$.

Формула (3.5.10) для состоятельной оценки матрицы $\boldsymbol{S}$ в случае системы нескольких уравнений имеет вид:

$$\widehat{\boldsymbol{S}} = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} \widehat{\varepsilon}_{i1}\widehat{\varepsilon}_{i1}\boldsymbol{x}_{i1}\boldsymbol{x}'_{i1} & \cdots & \frac{1}{n}\sum_{i=1}^{n} \widehat{\varepsilon}_{i1}\widehat{\varepsilon}_{iM}\boldsymbol{x}_{i1}\boldsymbol{x}'_{iM} \\ \vdots & & \vdots \\ \frac{1}{n}\sum_{i=1}^{n} \widehat{\varepsilon}_{iM}\widehat{\varepsilon}_{i1}\boldsymbol{x}_{iM}\boldsymbol{x}'_{i1} & \cdots & \frac{1}{n}\sum_{i=1}^{n} \widehat{\varepsilon}_{iM}\widehat{\varepsilon}_{iM}\boldsymbol{x}_{iM}\boldsymbol{x}'_{iM} \end{bmatrix} \tag{4.3.2}$$

для некоторой состоятельной оценки $\widehat{\varepsilon}_{im}$ ошибки ε_{im}.

Утверждение 4.2 (состоятельная оценка матрицы $\boldsymbol{S}$, асимптотическая дисперсия вектора $\overline{\boldsymbol{g}}$): *Пусть* $\widehat{\boldsymbol{\delta}}_m$ *— состоятельная оценка вектора* $\boldsymbol{\delta}_m$ *и пусть* $\widehat{\varepsilon}_{im} \equiv y_{im} - \boldsymbol{z}'_{im}\widehat{\boldsymbol{\delta}}_m$ *представляет собой подразумеваемый остаток для* $m = 1, 2, \ldots, M$. *В предположениях 4.1, 4.2 и 4.6* $\widehat{\boldsymbol{S}}$ *из* (4.3.2) *является состоятельной оценкой для матрицы* $\boldsymbol{S}$.

Мы не будем приводить доказательство, поскольку метод доказательства очень схож с использованным при доказательстве утверждения 2.4.

Поскольку $\widehat{\boldsymbol{S}}$ — состоятельная оценка $\boldsymbol{S}$, GMM-оценка с использованием $\widehat{\boldsymbol{S}}^{-1}$ в качестве взвешивающей матрицы, $\widehat{\delta}(\widehat{\boldsymbol{S}}^{-1})$, является эффективной GMM-оценкой с минимальной асимптотической дисперсией. Для удобства здесь воспроизводятся формулы (3.5.13) и (3.5.14) для асимптотической дисперсии и ее оценки:

$$\mathrm{Avar}(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1})) = (\boldsymbol{\Sigma}'_{\boldsymbol{xz}}\boldsymbol{S}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{xz}})^{-1}. \tag{4.3.3}$$

$$\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}))} = (\boldsymbol{S}'_{\boldsymbol{xz}}\widehat{\boldsymbol{S}}^{-1}\boldsymbol{S}_{\boldsymbol{xz}})^{-1}. \tag{4.3.4}$$

Резюмируя, GMM-оценка системы уравнений имеет следующие специфические свойства: $\Sigma_{\boldsymbol{xz}}$ — блочно-диагональная матрица из (4.1.9), $\boldsymbol{S}$ задается уравнением (4.1.11), $\boldsymbol{s}_{\boldsymbol{xz}}$ и $\boldsymbol{S}_{\boldsymbol{xz}}$ берутся из уравнения (4.2.2), а $\widehat{\boldsymbol{S}}$ — из уравнения (4.3.2). Для получения начальной оценки $\widehat{\boldsymbol{\delta}}_m$, необходимой для вычисления $\widehat{\varepsilon}_{im}$ и $\widehat{\boldsymbol{S}}$, мы можем использовать FIVE-оценку (будет рассмотрена позже) или эффективную GMM-оценку для одного уравнения, примененную к каждому уравнению в отдельности. Разумеется, что при условии состоятельности выбор начальной оценки никак не влияет на асимптотическое распределение эффективной GMM-оценки.

В табл. 4.1 сведены все основные результаты данного параграфа.

Таблица 4.1

GMM для системы уравнений в формате случая единственного уравнения

Выборочный аналог
условий ортогональности: $\boldsymbol{g}_n(\widetilde{\boldsymbol{\delta}}) = \boldsymbol{s}_{xy} - \boldsymbol{S}_{xz}\widetilde{\boldsymbol{\delta}} = \mathbf{0}$.
GMM-оценка: $\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}) = (\boldsymbol{S}'_{xz}\widehat{\boldsymbol{W}}\boldsymbol{S}_{xz})^{-1}\boldsymbol{S}'_{xz}\widehat{\boldsymbol{W}}\boldsymbol{s}_{xy}$.
Ее выборочная ошибка: $\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}) = (\boldsymbol{S}'_{xz}\widehat{\boldsymbol{W}}\boldsymbol{S}_{xz})^{-1}\boldsymbol{S}'_{xz}\widehat{\boldsymbol{W}}\boldsymbol{s}_{xy}$.
Асимптотическая дисперсия
оптимальной GMM-оценки: $\mathrm{Avar}(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1})) = (\boldsymbol{\Sigma}'_{xz}\boldsymbol{S}^{-1}\boldsymbol{\Sigma}_{xz})^{-1}$.
Ее оценка: $\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}))} = (\boldsymbol{S}'_{xz}\widehat{\boldsymbol{S}}^{-1}\boldsymbol{S}_{xz})^{-1}$.
J-статистика: $J(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}), \widehat{\boldsymbol{S}}^{-1}) \equiv n \cdot \boldsymbol{g}_n(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}))'\widehat{\boldsymbol{S}}^{-1}\boldsymbol{g}_n(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}))$.

	GMM для одного уравнения, применительно к интересующему нас уравнению	**GMM для системы уравнений**
$\boldsymbol{g}_i$	$\boldsymbol{x}_i \cdot \varepsilon_i$	(4.1.4)
$\boldsymbol{\delta}$	$\boldsymbol{\delta}$	(4.1.6)
$\boldsymbol{s}_{xy}$	$\frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_i \cdot y_i$	(4.2.2)
$\boldsymbol{S}_{xz}$	$\frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{z}'_i$	(4.2.2)
Размер $\boldsymbol{W}$	$K \times K$	$\sum_m K_m \times \sum_m K_m$
$\boldsymbol{\Sigma}_{xz}$	$\mathrm{E}(\boldsymbol{x}_i\boldsymbol{z}'_i)$	(4.1.9)
$\boldsymbol{S}(\equiv Avar(\overline{\boldsymbol{g}}))$	$\mathrm{E}(\varepsilon_i^2\boldsymbol{x}_i\boldsymbol{x}'_i)$	(4.1.11)
$\widehat{\boldsymbol{S}}$	$\frac{1}{n}\sum_{i=1}^n \hat{\varepsilon}_i^2\boldsymbol{x}_i\boldsymbol{x}'_i$	(4.3.2)
При каких условиях оценка состоятельна?	3.1–3.4	4.1–4.4
При каких условиях оценка асимптотически нормальна?	3.1–3.5	4.1–4.5
При каких условиях $\widehat{\boldsymbol{S}} \to_p \boldsymbol{S}$?	3.1, 3.2, 3.6, $\mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}'_i)$ конечно	4.1, 4.2, 4.6, $\mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}'_i)$ конечно
Степени свободы J	$K - L$	$\sum_m(K_m - L_m)$

4.4. Сравнение оценивания одного уравнения и системы уравнений

Очевидной альтернативой GMM для системы уравнений, оценивающему штабелированный вектор коэффициентов $\boldsymbol{\delta}(\Sigma_m L_m \times 1)$ совместно, является применение GMM для одного уравнения к каждому уравнению в отдельности и последующее штабелирование оценок коэффициентов. Если взвешивающая матрица в случае обычного GMM для m-го уравнения есть $\widehat{\boldsymbol{W}}_{mm}$ $(K_m \times K_m)$, то получающийся в результате (оценивания каждого уравнения в отдельности) штабелированный вектор коэффициентов $\boldsymbol{\delta}$ может быть записан как GMM-оценка для системы уравнений $\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}})$ из (4.2.3), где взвешивающая матрица $\widehat{\boldsymbol{W}}$ размерности $\sum_m K_m \times \sum_m K_m$ есть блочно-диагональная матрица, m-й блок которой есть $\widehat{\boldsymbol{W}}_{mm}$:

$$\widehat{\boldsymbol{W}} = \begin{bmatrix} \widehat{\boldsymbol{W}}_{11} & & \\ & \ddots & \\ & & \widehat{\boldsymbol{W}}_{MM} \end{bmatrix}. \tag{4.4.1}$$

Это равенство следует из того, что обе матрицы, $\boldsymbol{S_{xz}}(\Sigma_m K_m \times \Sigma_m L_m)$ и $\widehat{\boldsymbol{W}}$, являются блочно-диагональными. (Читателю следует подставить выражение (4.2.2) для $\boldsymbol{S}'_{xz}$ и (4.4.1) для $\widehat{\boldsymbol{W}}$ в (4.2.3), чтобы проверить, что $\boldsymbol{S}'_{\boldsymbol{xz}}\widehat{\boldsymbol{W}}\boldsymbol{S_{xz}}$ и, следовательно, $(\boldsymbol{S}'_{\boldsymbol{xz}}\widehat{\boldsymbol{W}}\boldsymbol{S_{xz}})^{-1}\boldsymbol{S}'_{\boldsymbol{xz}}\widehat{\boldsymbol{W}}$ являются блочно-диагональными, используя формулы (A.5) и (A.9) из приложения A. Для простоты можно положить $M = 2$.)

Таким образом, разница между использованием GMM для системы уравнений и GMM для одного уравнения при оценивании штабелированного вектора δ состоит в выборе взвешивающей матрицы $\widehat{\boldsymbol{W}}$ размера $\Sigma_m K_m \times \Sigma_m K_m$. Иначе говоря, GMM, применяемый последовательно к каждому уравнению, является частным случаем GMM для системы уравнений.

Когда они «эквивалентны»?

Как видно из главы 3, если уравнение точно идентифицируемо, то выбор взвешивающей матрицы $\widehat{\boldsymbol{W}}$ не имеет значения, поскольку GMM-оценка, независимо от выбора $\widehat{\boldsymbol{W}}$, численно равна IV-оценке. То же самое верно и для GMM для системы уравнений: если каждое уравнение системы точно идентифицируемо, то есть $L_m = K_m$ для всех m, то тогда $\boldsymbol{S_{xz}}$ из (4.2.3) — квадратная матрица и GMM-оценка при любом выборе $\widehat{\boldsymbol{W}}$ становится IV-оценкой системы уравнений $\boldsymbol{S}_{\boldsymbol{xz}}^{-1}\boldsymbol{s_{xy}}$ (и, так как матрица $\boldsymbol{S}_{xz}$ является блочно-диагональной, полученная оценка представляет собой совокупность IV-оценок для каждого уравнения).

С другой стороны, если хотя бы одно из M уравнений сверхидентифицируемо, выбор матрицы $\widehat{\boldsymbol{W}}(\sum_m K_m \times \sum_m K_m)$ влияет на численное значение GMM-оценки. Рассмотрим эффективную GMM-оценку для $\boldsymbol{\delta}\,(\sum_m L_m \times 1)$, полученную оцениванием каждого уравнения в отдельности, причем каждое уравнение оценено при оптимальном выборе $\widehat{\boldsymbol{W}}_{mm}$ $(m = 1, 2, \ldots, M)$. Эффективна ли эта оценка так же, как и оценка GMM всей системы уравнений одновременно? Оценку, полученную в результате GMM-оценивания каждого уравнения в отдельности, можно записать как $\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}})$, где взвешивающая матрица $\widehat{\boldsymbol{W}}$ размера $\sum_m K_m \times \times \sum_m L_m$ удовлетворяет соотношению:

$$\operatorname*{plim}_{n\to\infty} \widehat{\boldsymbol{W}} = \begin{bmatrix} \left(\mathrm{E}(\varepsilon_{i1}^2 \boldsymbol{x}_{i1}\boldsymbol{x}'_{i1})\right)^{-1} & & \\ & \ddots & \\ & & \left(\mathrm{E}(\varepsilon_{iM}^2 \boldsymbol{x}_{iM}\boldsymbol{x}'_{iM})\right)^{-1} \end{bmatrix}. \tag{4.4.2}$$

Поскольку этот предел по вероятности не обязательно равен $\boldsymbol{S}^{-1}$ (если только не выполняется условие (4.4.3) ниже), данная оценка, как правило, менее эффективна, чем одновременная GMM-оценка системы уравнений $\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1})$, где $\widehat{\boldsymbol{S}}$ есть состоятельная оценка $\boldsymbol{S}$. Но если уравнения являются «несвязанными»[1] в том смысле, что

$$\mathrm{E}(\varepsilon_{im}\varepsilon_{ih}\boldsymbol{x}_{im}\boldsymbol{x}'_{ih}) = \mathbf{0} \quad \text{для всех } m \neq h \ (= 1, 2, \ldots, M), \tag{4.4.3}$$

то $\boldsymbol{S}$ становится блочно-диагональной и для матрицы $\widehat{\boldsymbol{W}}$, определенной выше, $\operatorname{plim} \widehat{\boldsymbol{W}} = \boldsymbol{S}^{-1}$. Так как в этом случае $\widehat{\boldsymbol{W}} - \boldsymbol{S}^{-1} \to_{\mathrm{p}} 0$, мы имеем (как было доказано в контрольном вопросе 3 из параграфа 3.5)

$$\sqrt{n}\,\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}) - \sqrt{n}\,\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}) \underset{\mathrm{p}}{\to} \mathbf{0}.$$

Из леммы 2.4(а) следует, что асимптотическое распределение $\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}})$ такое же, как и у $\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1})$, поэтому эффективная GMM-оценка каждого уравнения поочередно есть не что иное, как эффективная GMM-оценка всей системы одновременно.

Подытожим сказанное выше:

Утверждение 4.3 (эквивалентность GMM-оценивания поочередно каждого уравнения и всей системы одновременно):

(a) *Если все уравнения точно идентифицируемы, то GMM-оценивание поочередно каждого уравнения и GMM-оценивание всей системы одновременно дают численно одинаковые оценки, эквивалентные IV-оценке.*

(b) *Если хотя бы одно из уравнений сверхидентифицируемо, но уравнения «не связаны» в смысле (4.4.3), то GMM-оценивание поочередно каждого уравнения и GMM-оценивание всей системы одновременно асимптотически эквивалентны в том смысле, что умноженная на $\sqrt{n}$ разность этих оценок сходится к нулю по вероятности.*

[1] При условной гомоскедастичности это условие превращается в более знакомое условие $\mathrm{E}(\varepsilon_{im}\varepsilon_{ih}) = 0$. См. следующий параграф.

В двух описанных случаях совместное оценивание не дает никакого выигрыша в эффективности. Более того, если *априорно* известно, что уравнения «не связаны», то тогда предпочтительнее оценивать каждое уравнение в отдельности, так как использование априорного знания, выражающегося в занулении всех элементов матрицы $\widehat{\boldsymbol{W}}$ вне диагональных блоков, весьма вероятно, приведет к улучшению свойств оценки на конечных выборках.

Совместное оценивание может быть рискованным

За исключением случаев (a) и (b), совместное оценивание асимптотически более эффективно. Даже если интерес представляет только оценивание одного конкретного уравнения (как, например, уравнения для LW из примера 4.1), вы можете обычно выиграть в асимптотической эффективности, комбинируя его с другими уравнениями (см., однако, аналитическое упражнение 10 в качестве примера, когда совместное оценивание не влечет за собой увеличение асимптотической эффективности, даже если добавляемые уравнения не являются «несвязанными»).

Следует, однако, сделать некоторые оговорки. Во-первых, оценки коэффициентов интересующего нас уравнения могут иметь лучшие свойства на конечных выборках без совместного оценивания с другими уравнениями. Во-вторых, асимптотические результаты предполагают, что модель **правильно специфицирована** (*correctly specified*), то есть все предположения модели выполнены. Если модель специфицирована неправильно, то ни GMM-оценка каждого уравнения поочередно, ни GMM-оценка всей системы одновременно не гарантируют даже состоятельности. А шансы неправильности спецификации возрастают при добавлении уравнений к системе.

Чтобы проиллюстрировать второе замечание, допустим, что предположение 4.3 не выполняется из-за невыполнения условий ортогональности для M-го уравнения и только для него: $\mathrm{E}(\boldsymbol{x}_{iM}\varepsilon_{iM}) \neq \mathbf{0}$. Это может произойти, например, в случае пропуска переменной, которая должна быть включена в качестве регрессора. Чтобы увидеть, что это ведет к возможной потере состоятельности во всех уравнениях, посмотрим на формулу (4.2.4) для ошибки оценивания, где $\overline{\boldsymbol{g}}$ задается выражением (4.2.5). M-й блок матрицы $\mathrm{plim}_{n\to\infty}\overline{\boldsymbol{g}}$ не равен нулю. Поскольку $\mathrm{plim}\,\boldsymbol{S}_{\boldsymbol{xz}} = \boldsymbol{\Sigma}_{\boldsymbol{xz}}$ при условии эргодической стационарности и $\widehat{\boldsymbol{W}} = \boldsymbol{W}$, по предположению, асимптотическое смещение равно:

$$\operatorname*{plim}_{n\to\infty} \widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}) - \boldsymbol{\delta} = (\boldsymbol{\Sigma}'_{\boldsymbol{xz}}\boldsymbol{W}\boldsymbol{\Sigma}_{\boldsymbol{xz}})^{-1}\boldsymbol{\Sigma}'_{\boldsymbol{xz}}\boldsymbol{W}\begin{bmatrix} 0 \\ \vdots \\ 0 \\ \mathrm{E}(\boldsymbol{x}_{iM}\cdot\varepsilon_{iM}) \end{bmatrix}. \qquad (4.4.4)$$

У эффективной GMM-оценки системы уравнений W, и, следовательно, $(\Sigma'_{xz} W \Sigma_{xz})^{-1} \Sigma'_{xz} W$, в общем случае не являются блочно-диагональными, так что любой элемент из $\mathrm{plim}_{n\to\infty} \widehat{\delta}(\widehat{W}) - \delta$ может быть отличным от нуля. Даже для коэффициентов интересующих нас правильно специфицированных уравнений асимптотическое смещение может быть ненулевым. То есть в случае совместного оценивания смещение, вызванное неправильной спецификацией одного уравнения, распространяется на всю систему. Эта проблема не возникает в случае GMM-оценивания каждого уравнения поочередно, при котором матрица $\widehat{W}$ имеет блочно-диагональную структуру.

Контрольные вопросы

1. Выразите 2SLS-оценку каждого уравнения поочередно в форме (4.2.3), специфицируя подходящим образом $\widehat{W}_{mm}$ в (4.4.1).
   [Ответ: $\widehat{W}_{mm} = (\frac{1}{n}\sum_{i=1}^{n} x_{im} x'_{im})^{-1}$.]

4.5. Частные случаи GMM-оценивания системы уравнений: FIVE, 3SLS и SUR

Условная гомоскедастичность

В случае условной гомоскедастичности можно легко получить несколько важных оценок в качестве частных случаев GMM-оценки системы уравнений. Модификация условной гомоскедастичности для систем уравнений такова:

Предположение 4.7 (условная гомоскедастичность):

$$\mathrm{E}(\varepsilon_{im}\,\varepsilon_{ih} | x_{im}, x_{ih}) = \sigma_{mh}$$

для всех $m, h = 1, 2, \ldots, M$.

В этом случае безусловный перекрестный момент $\mathrm{E}(\varepsilon_{im}\varepsilon_{ih})$ равен σ_{mh} в соответствии с формулой полного математического ожидания. Вам должно быть несложно показать, что

блок (m, h) из $\mathrm{E}(g_i g_i')$, представленный в (4.1.11), $=$

$$= \mathrm{E}(\varepsilon_{im}\varepsilon_{ih} x_{im} x'_{ih}) =$$

$$= \mathrm{E}[\mathrm{E}(\varepsilon_{im}\varepsilon_{ih} x_{im} x'_{ih} | x_{im}, x_{ih})]$$

(по закону полного математического ожидания) $=$

$$= \mathrm{E}[\mathrm{E}(\varepsilon_{im}\varepsilon_{ih}|\boldsymbol{x}_{im}, \boldsymbol{x}_{ih})\boldsymbol{x}_{im}\boldsymbol{x}'_{ih}]$$

(ввиду линейности условного математического ожидания) =

$$= \mathrm{E}[\sigma_{mh}\boldsymbol{x}_{im}\boldsymbol{x}'_{ih}]$$

(ввиду условной гомоскедастичности) =

$$= \sigma_{mh}\,\mathrm{E}(\boldsymbol{x}_{im}\boldsymbol{x}'_{ih}) \quad \text{(ввиду линейности математического ожидания).} \tag{4.5.1}$$

Тогда $\boldsymbol{S}$ из (4.1.11) можно записать как

$$\boldsymbol{S} = \begin{bmatrix} \sigma_{11}\,\mathrm{E}(\boldsymbol{x}_{i1}\boldsymbol{x}'_{i1}) & \cdots & \sigma_{1M}\,\mathrm{E}(\boldsymbol{x}_{i1}\boldsymbol{x}'_{iM}) \\ \vdots & & \vdots \\ \sigma_{M1}\,\mathrm{E}(\boldsymbol{x}_{iM}\boldsymbol{x}'_{i1}) & \cdots & \sigma_{MM}\,\mathrm{E}(\boldsymbol{x}_{iM}\boldsymbol{x}'_{iM}) \end{bmatrix}. \tag{4.5.2}$$

Поскольку, по предположению 4.5, $\boldsymbol{S}$ конечна, из данной декомпозиции вытекает, что $\mathrm{E}(\boldsymbol{x}_{im}\boldsymbol{x}'_{ih})$ существует и конечно для всех m, h $(= 1, 2, \ldots, M)$.

Инструментальная оценка с полной информацией (FIVE)

Оценка для $\boldsymbol{S}$, использующая структуру четвертых моментов, показанную в (4.5.2), такова:

$$\widehat{\boldsymbol{S}} = \begin{bmatrix} \widehat{\sigma}_{11}\cdot\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{i1}\boldsymbol{x}'_{i1}\right) & \cdots & \widehat{\sigma}_{1M}\cdot\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{i1}\boldsymbol{x}'_{iM}\right) \\ \vdots & & \vdots \\ \widehat{\sigma}_{M1}\cdot\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{iM}\boldsymbol{x}'_{i1}\right) & \cdots & \widehat{\sigma}_{MM}\cdot\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{iM}\boldsymbol{x}'_{iM}\right) \end{bmatrix}, \tag{4.5.3}$$

где для некоторой состоятельной оценки $\widehat{\boldsymbol{\delta}}_m$ вектора $\boldsymbol{\delta}_m$

$$\widehat{\sigma}_{mh} \equiv \frac{1}{n}\sum_{i=1}^{n}\widehat{\varepsilon}_{im}\,\widehat{\varepsilon}_{ih},\ \widehat{\varepsilon}_{im} \equiv y_{im} - \boldsymbol{z}'_{im}\widehat{\boldsymbol{\delta}}_m \quad (m, h = 1, 2, \ldots, M). \tag{4.5.4}$$

Согласно утверждению 4.1, $\widehat{\sigma}_{mh} \to_{\mathrm{p}} \sigma_{mh}$ при условии (в дополнение к предположениям 4.1 и 4.2), что $\mathrm{E}(\boldsymbol{z}_{im}\boldsymbol{z}'_{ih})$ конечно. Из эргодической стационарности следует, что $\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{im}\boldsymbol{x}'_{ih}$ сходится по вероятности к $\mathrm{E}(\boldsymbol{x}_{im}\boldsymbol{x}'_{ih})$, которое, как было только что отмечено, существует и конечно. Следовательно, (4.5.3) состоятельна для $\boldsymbol{S}$ без условия на четвертые моменты (предположение 4.6).

FIVE-оценка (*FIVE estimator*) вектора $\boldsymbol{\delta}$, обозначаемая $\widehat{\boldsymbol{\delta}}_{FIVE}$, есть

$$\widehat{\boldsymbol{\delta}}_{FIVE} \equiv \widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}).$$

где $\widehat{\boldsymbol{S}}$ задается посредством (4.5.3), чтобы использовать условную гомоскедастичность[1]. Будучи GMM-оценкой для системы уравнений при некоторой выбранной $\widehat{\boldsymbol{W}}$, она является состоятельной и асимптотически нормальной согласно утверждению 3.1, адаптированному к случаю системы уравнений. Поскольку $\widehat{\boldsymbol{S}}$ — состоятельная оценка для $\boldsymbol{S}$, она является эффективной согласно утверждению 3.5. Таким образом, мы доказали

Утверждение 4.4 (свойства FIVE-оценок на больших выборках): *Предположим, что выполняются предположения 4.1–4.5 и 4.7. Кроме того, предположим, что* $\mathrm{E}(z_{im}z'_{ih})$ *существует и конечно для всех* m, h $(= 1, 2, \ldots, M)$.[2] *Пусть* $\boldsymbol{S}$ *и* $\widehat{\boldsymbol{S}}$ *соответствуют* (4.5.2) *и* (4.5.3). *Тогда*

(a) $\widehat{\boldsymbol{S}} \to_p \boldsymbol{S}$.

(b) $\widehat{\boldsymbol{\delta}}_{\text{FIVE}}$, *определяемая как* $\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1})$, *является состоятельной, асимптотически нормальной и эффективной с* $\mathrm{Avar}(\widehat{\boldsymbol{\delta}}_{\text{FIVE}})$ *из* (4.3.3).

(c) Оценка асимптотической дисперсии, приведенная в (4.3.4), *является состоятельной для* $\mathrm{Avar}(\widehat{\boldsymbol{\delta}}_{\text{FIVE}})$.

(d) (Статистика Саргана.)

$$J(\widehat{\boldsymbol{\delta}}_{\text{FIVE}}, \widehat{\boldsymbol{S}}^{-1}) \equiv n \cdot \boldsymbol{g}_n(\widehat{\boldsymbol{\delta}}_{\text{FIVE}})' \, \widehat{\boldsymbol{S}}^{-1} \, \boldsymbol{g}_n(\widehat{\boldsymbol{\delta}}_{\text{FIVE}}) \underset{d}{\to} \chi^2\Big(\sum_m (K_m - L_m)\Big),$$

где $\boldsymbol{g}_n(\cdot)$ *задается в* (4.2.1).

Обычно начальная оценка $\widehat{\boldsymbol{\delta}}_m$, необходимая для получения $\widehat{\boldsymbol{S}}$, является 2SLS-оценкой.

Трехшаговый метод наименьших квадратов (3SLS)

Когда набор инструментов один и тот же для всех уравнений, формула для FIVE-оценки может быть упрощена. Упрощенная формула определяет **3SLS-оценку** (*3SLS-estimator*), обозначаемую как $\widehat{\boldsymbol{\delta}}_{3SLS}$[3]. Для этого положим

$$\underset{(M\times 1)}{\boldsymbol{\varepsilon}_i} \equiv \begin{bmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iM} \end{bmatrix}. \tag{4.5.5}$$

[1]Это оценка из [Brundy and Jorgenson, 1971].

[2]Это предположение необходимо для состоятельности $\widehat{\sigma_{mh}}$. См. утверждение 4.1.

[3]Это оценка из [Zellner and Theil, 1962].

Тогда матрица перекрестных моментов для ε_i размера $M \times M$, обозначаемая как $\mathbf{\Sigma}$, может быть записана как

$$\underset{(M\times M)}{\mathbf{\Sigma}} \equiv \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1M} \\ \vdots & & \vdots \\ \sigma_{M1} & \cdots & \sigma_{MM} \end{bmatrix} = \mathrm{E}(\varepsilon_i \varepsilon_i'). \tag{4.5.6}$$

Чтобы состоятельно оценить $\mathbf{\Sigma}$, нам требуется начальная состоятельная оценка δ_m для получения остатка $\widehat{\varepsilon}_{im}$. Термин 3SLS связан с тем, что в качестве начальной оценки для $\boldsymbol{\delta}_m$ используется 2SLS-оценка. При наличии остатков, полученных таким путем, естественной оценкой для $\mathbf{\Sigma}$ является

$$\widehat{\mathbf{\Sigma}} \equiv \begin{bmatrix} \widehat{\sigma}_{11} & \cdots & \widehat{\sigma}_{1M} \\ \vdots & & \vdots \\ \widehat{\sigma}_{M1} & \cdots & \widehat{\sigma}_{MM} \end{bmatrix} = \frac{1}{n}\sum_{i=1}^{n} \widehat{\varepsilon}_i \widehat{\varepsilon}_i'. \tag{4.5.7}$$

Это матрица, в которой элемент (m, h) является оцененным перекрестным моментом, заданным в (4.5.4).

Если $\boldsymbol{x}_i(= \boldsymbol{x}_{i1} = \boldsymbol{x}_{i2} = \cdots = \boldsymbol{x}_{iM})$ — общий набор инструментов размерности K, то $\boldsymbol{g}_i$ из (4.1.6), $\boldsymbol{S}$ из (4.5.2) и $\widehat{\boldsymbol{S}}$ из (4.5.3) могут быть компактно записаны с использованием произведения Кронекера[1] как

$$\underset{(MK\times 1)}{\boldsymbol{g}_i} = \boldsymbol{\varepsilon}_i \otimes \boldsymbol{x}_i. \tag{4.5.8}$$

$$\underset{(MK\times MK)}{\boldsymbol{S}} = \underset{(M\times M)}{\mathbf{\Sigma}} \otimes \underset{(K\times K)}{\mathrm{E}(\boldsymbol{x}_i \boldsymbol{x}_i')}. \tag{4.5.9}$$

так что $\boldsymbol{S}^{-1} = \mathbf{\Sigma}^{-1} \otimes [\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')]^{-1}$,

$$\widehat{\boldsymbol{S}} = \widehat{\mathbf{\Sigma}} \otimes \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'\right), \tag{4.5.10}$$

так что $\widehat{\boldsymbol{S}}^{-1} = \widehat{\mathbf{\Sigma}}^{-1} \otimes (\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i')^{-1}$. Из представления в виде произведения Кронекера (4.5.9) для $\boldsymbol{S}$ и невырожденности $\boldsymbol{S}$ (согласно предположению 4.5) вытекает, что обе матрицы $\mathbf{\Sigma}$ и $\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')$ невырождены.

Для дальнейшего преобразования формулы для 3SLS-оценки мы должны вернуться к соотношению (4.2.6), которое, при $\widehat{\boldsymbol{W}} = \widehat{\boldsymbol{S}}^{-1}$, описывает эффективную GMM-оценку для системы уравнений. Формула (4.5.10) предполагает, что $\widehat{\boldsymbol{W}}$ в (4.2.6) такова, что

$$\widehat{\boldsymbol{W}}_{mh} \ (\equiv (m, h) \text{ блок } \widehat{\boldsymbol{W}}) = \widehat{\sigma}^{mh} \cdot \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'\right)^{-1}. \tag{4.5.11}$$

[1]Если вы не знакомы с записью произведения Кронекера, посмотрите формулы (А.11) и (А.15) приложения А.

где $\widehat{\sigma}^{mh}$ есть элемент (m, h) матрицы $\widehat{\mathbf{\Sigma}}^{-1}$. Подставляя это в (4.2.6), получаем:

$$\widehat{\boldsymbol{\delta}}_{3SLS} = \begin{bmatrix} \widehat{\sigma}^{11}\widehat{\boldsymbol{A}}_{11} & \cdots & \widehat{\sigma}^{1M}\widehat{\boldsymbol{A}}_{1M} \\ \vdots & & \vdots \\ \widehat{\sigma}^{M1}\widehat{\boldsymbol{A}}_{M1} & \cdots & \widehat{\sigma}^{MM}\widehat{\boldsymbol{A}}_{MM} \end{bmatrix}^{-1} \begin{bmatrix} \widehat{\sigma}^{11}\widehat{\boldsymbol{c}}_{11} + \cdots + \widehat{\sigma}^{1M}\widehat{\boldsymbol{c}}_{1M} \\ \vdots \\ \widehat{\sigma}^{M1}\widehat{\boldsymbol{c}}_{M1} + \cdots + \widehat{\sigma}^{MM}\widehat{\boldsymbol{c}}_{MM} \end{bmatrix}, \tag{4.5.12}$$

где

$$\widehat{\boldsymbol{A}}_{mh} \equiv \underset{(L_m \times K)}{\Big(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_{im}\boldsymbol{x}_i'\Big)} \underset{(K \times K)}{\Big(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i'\Big)^{-1}} \underset{(K \times L_h)}{\Big(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{z}_{ih}'\Big)}, \tag{4.5.13}$$

$$\widehat{\boldsymbol{c}}_{mh} \equiv \underset{(L_m \times K)}{\Big(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_{im}\boldsymbol{x}_i'\Big)} \underset{(K \times K)}{\Big(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i'\Big)^{-1}} \underset{(K \times 1)}{\Big(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \cdot y_{ih}\Big)}. \tag{4.5.14}$$

Аналогично, подставляя в (4.3.3) выражение для $\mathbf{\Sigma}_{\boldsymbol{xz}}$ из (4.1.9) и выражение для $\boldsymbol{S}$ из (4.5.9) и используя формулу (A.6) из приложения A, получим выражение для $\mathrm{Avar}(\widehat{\boldsymbol{\delta}}_{3SLS})$:

$$\mathrm{Avar}(\widehat{\boldsymbol{\delta}}_{3SLS}) = \begin{bmatrix} \sigma^{11}\boldsymbol{A}_{11} & \cdots & \sigma^{1M}\boldsymbol{A}_{1M} \\ \vdots & & \vdots \\ \sigma^{M1}\boldsymbol{A}_{M1} & \cdots & \sigma^{MM}\boldsymbol{A}_{MM} \end{bmatrix}^{-1}, \tag{4.5.15}$$

где

$$\boldsymbol{A}_{mh} \equiv \mathrm{E}(\boldsymbol{z}_{im}\boldsymbol{x}_i')\left[\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')\right]^{-1}\mathrm{E}(\boldsymbol{x}_i\boldsymbol{z}_{ih}') \tag{4.5.16}$$

и σ^{mh} есть элемент (m, h) матрицы $\mathbf{\Sigma}^{-1}$. Эта асимптотическая дисперсия состоятельно оценивается посредством

$$\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\delta}}_{3SLS})} = \begin{bmatrix} \widehat{\sigma}^{11}\widehat{\boldsymbol{A}}_{11} & \cdots & \widehat{\sigma}^{1M}\widehat{\boldsymbol{A}}_{1M} \\ \vdots & & \vdots \\ \widehat{\sigma}^{M1}\widehat{\boldsymbol{A}}_{M1} & \cdots & \widehat{\sigma}^{MM}\widehat{\boldsymbol{A}}_{MM} \end{bmatrix}^{-1}. \tag{4.5.17}$$

(Этот выборочный аналог можно также получить напрямую подстановкой (4.2.2) и (4.5.10) в (4.3.4).)

В большинстве учебников приводятся даже еще более привлекательные выражения с использованием матриц данных (таких как матрица $\boldsymbol{X}$ размера $n \times K$). Вывод таких выражений мы оставляем в качестве аналитического упражнения к этой главе. Подытожим полученные результаты:

Утверждение 4.5 (Свойства 3SLS-оценки на больших выборках): *Допустим, что предположения 4.1–4.5 и 4.7 выполняются, и предположим, что* $\boldsymbol{x}_{im} = \boldsymbol{x}_i$ *(общий набор инструментов). Предположим, далее,* *что* $\mathrm{E}(\boldsymbol{z}_{im}\boldsymbol{z}'_{ih})$ *существует и конечно для всех* m, h $(= 1, 2, \ldots, M)$. *Пусть* $\widehat{\boldsymbol{\Sigma}}$ *есть матрица размера* $M \times M$ *оцененных перекрестных моментов ошибок, полученная в* (4.5.7) *с использованием 2SLS-остатков. Тогда*

(a) $\widehat{\boldsymbol{\delta}}_{\mathrm{3SLS}}$ *из* (4.5.12) *является состоятельной, асимптотически нормальной и эффективной оценкой с асимптотической дисперсией* $\mathrm{Avar}(\widehat{\boldsymbol{\delta}}_{\mathrm{3SLS}})$, *заданной в* (4.5.15).

(b) Оценка асимптотической дисперсии (4.5.17) *является состоятельной оценкой для* $\mathrm{Avar}(\widehat{\boldsymbol{\delta}}_{\mathrm{3SLS}})$.

(c) (Статистика Саргана.)

$$J(\widehat{\boldsymbol{\delta}}_{\mathrm{3SLS}}, \widehat{\boldsymbol{S}}^{-1}) \equiv n \cdot \boldsymbol{g}_n(\widehat{\boldsymbol{\delta}}_{\mathrm{3SLS}})' \, \widehat{\boldsymbol{S}}^{-1} \, \boldsymbol{g}_n(\widehat{\boldsymbol{\delta}}_{\mathrm{3SLS}}) \underset{d}{\to} \chi^2\Big(MK - \sum_m L_m\Big),$$

где $\widehat{\boldsymbol{S}} = \widehat{\boldsymbol{\Sigma}} \otimes \left(\frac{1}{n}\sum_i \boldsymbol{x}_i\boldsymbol{x}'_i\right)$, K — *число общих инструментов и* $\boldsymbol{g}_n(\cdot)$ *задается в* (4.2.1).

Кажущиеся несвязанными регрессии (SUR)

3SLS-формула может быть упрощена, если

$$\boldsymbol{x}_i = \text{объединение } (\boldsymbol{z}_{i1}, \ldots, \boldsymbol{z}_{iM}). \tag{4.5.18}$$

Это равносильно условию:

$$\mathrm{E}(\boldsymbol{z}_{im} \cdot \varepsilon_{ih}) = \mathbf{0} \quad (m, h = 1, 2, \ldots, M) \tag{4.5.18$'$}$$

То есть предопределенные регрессоры должны удовлетворять условию перекрестной ортогональности: они должны быть предопределенными не только в своем уравнении ($\mathrm{E}(\boldsymbol{z}_{im} \cdot \varepsilon_{im}) = 0$), но также и для других уравнений ($\mathrm{E}(\boldsymbol{z}_{im} \cdot \varepsilon_{ih}) = 0$ для $m \neq h$). Эта упрощенная формула называется **SUR-оценкой** (*SUR estimator*) и обозначается как $\widehat{\boldsymbol{\delta}}_{\mathrm{SUR}}$[1].

Пример 4.7 (пример 4.1 с измененной спецификацией инструментов): Если (4.5.18) выполняется для системы из двух уравнений из примера 4.1, $\boldsymbol{x}_i$ есть объединение $\boldsymbol{z}_{i1}$ и $\boldsymbol{z}_{i2}$, то тогда

$$\boldsymbol{x}_i = \begin{bmatrix} 1 \\ S_i \\ IQ_i \\ EXPR_i \end{bmatrix}. \tag{4.5.19}$$

[1]Эта оценка была утверждена в [Zellner, 1962].

Условия ортогональности $\mathrm{E}(\boldsymbol{x}_i \cdot \varepsilon_{i1}) = 0$ и $\mathrm{E}(\boldsymbol{x}_i \cdot \varepsilon_{i2}) = 0$ превращаются в

$$\mathrm{E}\begin{bmatrix} \varepsilon_{i1} \\ S_i\,\varepsilon_{i1} \\ IQ_i\,\varepsilon_{i1} \\ EXPR_i\,\varepsilon_{i1} \end{bmatrix} = \mathbf{0} \quad \text{и} \quad \mathrm{E}\begin{bmatrix} \varepsilon_{i2} \\ S_i\,\varepsilon_{i2} \\ IQ_i\,\varepsilon_{i2} \\ EXPR_i\,\varepsilon_{i2} \end{bmatrix} = \mathbf{0}, \tag{4.5.20}$$

что отличается от декларирования того, что регрессоры являются предопределенными в обоих уравнениях:

$$\mathrm{E}\begin{bmatrix} \varepsilon_{i1} \\ S_i\,\varepsilon_{i1} \\ IQ_i\,\varepsilon_{i1} \\ EXPR_i\,\varepsilon_{i1} \end{bmatrix} = \mathbf{0} \quad \text{и} \quad \mathrm{E}\begin{bmatrix} \varepsilon_{i2} \\ S_i\,\varepsilon_{i2} \\ IQ_i\,\varepsilon_{i2} \end{bmatrix} = \mathbf{0}. \tag{4.5.21}$$

Разница между (4.5.20) и (4.5.21) состоит в том, что уравнение KWW сверхидентифицируемо с $EXPR$ как дополнительным инструментом.

Так как SUR есть частный случай 3SLS, формулы (4.5.12), (4.5.15) и (4.5.17) для $\widehat{\delta}_{3\mathrm{SLS}}$ применимы к $\widehat{\delta}_{\mathrm{SUR}}$. Из (4.5.18) следует, что в полученных ранее выражениях для $\widehat{\boldsymbol{A}}_{mh}$, $\widehat{\boldsymbol{c}}_{mh}$ и $\boldsymbol{A}_{mh}$, $\boldsymbol{x}_i$ «исчезают»:

$$\widehat{\boldsymbol{A}}_{mh} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_{im}\boldsymbol{z}'_{ih}, \tag{4.5.13$'$}$$

$$\widehat{\boldsymbol{c}}_{mh} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_{im} \cdot y_{ih}, \tag{4.5.14$'$}$$

$$\boldsymbol{A}_{mh} = \mathrm{E}(\boldsymbol{z}_{im}\boldsymbol{z}'_{ih}). \tag{4.5.16$'$}$$

Докажем (4.5.16′). Без ограничения общности предположим, что $\boldsymbol{z}_{im}$ есть первые L_m элементов из K элементов вектора $\boldsymbol{x}_i$. Пусть $\boldsymbol{D}(K \times L_m)$ — первые L_m столбцов $\boldsymbol{I}_K$. Тогда

$$\boldsymbol{z}_{im} = \boldsymbol{D}'\boldsymbol{x}_i, \ \mathrm{E}(\boldsymbol{z}_{im}\boldsymbol{x}'_i) = \boldsymbol{D}'\,\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}'_i), \ \boldsymbol{D}'\,\mathrm{E}(\boldsymbol{x}_i\boldsymbol{z}'_{ih}) = \mathrm{E}(\boldsymbol{z}_{im}\boldsymbol{z}'_{ih}).$$

Подставляя это в (4.5.16), получим (4.5.16′). Доказательство равенств (4.5.13′) и (4.5.14′), касающихся выборочных моментов, проводится аналогично случаю с $\boldsymbol{z}_{im} = \boldsymbol{D}'\boldsymbol{x}_i$. (В одной из задач предлагается провести альтернативное доказательство с использованием оператора проектирования.)

В методе 3SLS в качестве начальной состоятельной оценки $\widehat{\boldsymbol{\Sigma}}$ использовалась 2SLS-оценка. Но 2SLS в случае вложенности регрессоров во множество инструментов превращается в OLS (как показано в контрольном вопросе 7 параграфа 3.8). Таким образом, для SUR начальной оценкой служит OLS-оценка. Из вышесказанного следует

Утверждение 4.6 (свойства SUR на больших выборках): *Пусть предположения 4.1–4.5 и 4.7 выполняются при* $\boldsymbol{x}_i$ = *объединение* $(\boldsymbol{z}_{i1}, \ldots, \boldsymbol{z}_{iM})$. *Пусть* $\widehat{\boldsymbol{\Sigma}}$ *— матрица оцененных кросс-моментов ошибок размерности* $M \times M$, *полученная в (4.5.7) с использованием OLS-остатков. Тогда*

(a) $\widehat{\boldsymbol{\delta}}_{\text{SUR}}$ *из* (4.5.12) *вместе с* $\widehat{\boldsymbol{A}}_{mh}$ *и* $\widehat{\boldsymbol{c}}_{mh}$ *из (4.5.13′) и (4.5.14′) для* $m, h = 1, \ldots, M$ *есть состоятельная, асимптотически нормальная и эффективная оценка с* $\text{Avar}(\widehat{\boldsymbol{\delta}}_{\text{SUR}})$ *из* (4.5.15), *где* $\boldsymbol{A}_{mh}$ *задается в (4.5.16′).*

(b) *Оценка асимптотической дисперсии* (4.5.17), *где* $\widehat{\boldsymbol{A}}_{mh}$ *задается в (4.5.13′), есть состоятельная оценка для* $\text{Avar}(\widehat{\boldsymbol{\delta}}_{\text{SUR}})$.

(c) *(Статистика Саргана.)*

$$J(\widehat{\boldsymbol{\delta}}_{\text{SUR}}, \widehat{\boldsymbol{S}}^{-1}) \equiv n \cdot \boldsymbol{g}_n(\widehat{\boldsymbol{\delta}}_{\text{SUR}})' \, \widehat{\boldsymbol{S}}^{-1} \, \boldsymbol{g}_n(\widehat{\boldsymbol{\delta}}_{\text{SUR}}) \underset{d}{\to} \chi^2\Big(MK - \sum_m L_m\Big),$$

где $\widehat{\boldsymbol{S}} = \widehat{\boldsymbol{\Sigma}} \otimes (\frac{1}{n}\sum_i \boldsymbol{x}_i \boldsymbol{x}_i')$, K *— количество общих инструментов и* $\boldsymbol{g}_n(\cdot)$ *задается в* (4.2.1).

(В противоположность утверждению 4.5 здесь, благодаря (4.5.18), не требуется условия конечности $\mathrm{E}(\boldsymbol{z}_{im}\boldsymbol{z}_{ih}')$, поскольку это следует из условия невырожденности $\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')$, что, в свою очередь, следует из невырожденности $\boldsymbol{S}$ [из предположения 4.5] и того факта, что $\boldsymbol{S}$ может быть записана как $\boldsymbol{\Sigma} \otimes (\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i')$.) Статистика Саргана для SUR редко рассматривается на практике.

SUR и OLS

Поскольку регрессоры предопределены, систему можно оценить с помощью OLS для каждого уравнения в отдельности. Тогда почему SUR лучше OLS? Как мы только что видели, при условной гомоскедастичности эффективная GMM-оценка системы уравнений есть FIVE-оценка, которая численно равна SUR-оценке при условии (4.5.18). Как показано в главе 3, при условной гомоскедастичности эффективная GMM-оценка одного уравнения есть 2SLS-оценка, которая, в свою очередь, численно эквивалентна OLS-оценке, так как регрессоры являются предопределенными при (4.5.18). Это соотношение показано на рис. 4.1. Следовательно, соотношение между SUR и OLS для каждого уравнения в отдельности точно такое же, как и соотношение между GMM для системы уравнений и GMM для каждого уравнения в отдельности. Как обсуждалось в параграфе 4.4, следует рассмотреть два случая.

(a) Каждое уравнение идентифицируемо точно. Поскольку общий набор инструментов является объединением всех регрессоров, то это может быть возможно, только если во всех уравнениях одни и те же регрессоры, то есть $z_{im} = x_i$ для всех m. SUR в таком случае называется **многомерной регрессией** (*multivariate regressions*). Мы показали в параграфе 4.4, что оба эффективных GMM, для системы уравнений и для каждого уравнения в отдельности, численно эквивалентны оценке методом IV для точно идентифицируемой системы. Так как регрессоры предопределены, мы можем заключить, что GMM-оценка модели многомерной регрессии есть OLS-оценка каждого уравнения в отдельности[1]. Это можно проверить, подставив (4.5.13′) и (4.5.14′) (с $z_{im} = x_i$) в выражение для точечной оценки (4.5.12). А подставив (4.5.16′) и (4.5.13′) (опять же с $z_{im} = x_i$) в (4.5.15) и (4.5.17), мы получим, что для многомерной регрессии выполнено:

$$\operatorname{Avar}(\widehat{\boldsymbol{\delta}}) = \boldsymbol{\Sigma} \otimes \left[\operatorname{E}(\boldsymbol{x}_i \boldsymbol{x}_i')\right]^{-1}. \tag{4.5.22}$$

$$\widehat{\operatorname{Avar}(\widehat{\boldsymbol{\delta}})} = \widehat{\boldsymbol{\Sigma}} \otimes \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'\right)^{-1}. \tag{4.5.23}$$

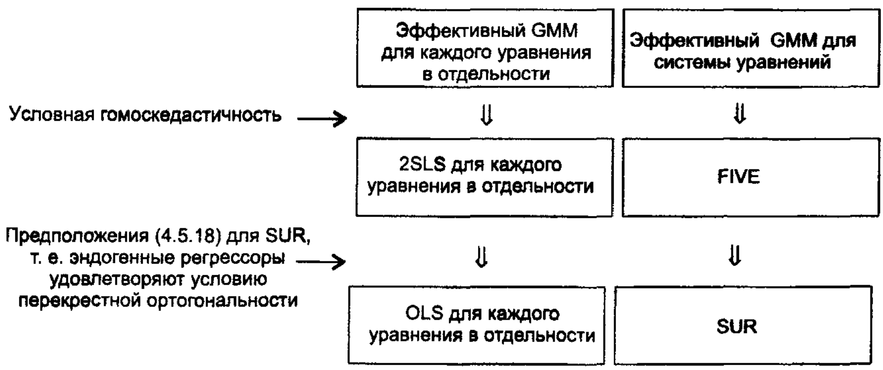

Рис. 4.1. OLS и GMM

(b) По крайней мере одно уравнение сверхидентифицируемо. Тогда SUR-оценивание более эффективно, чем оценивание методом OLS каждого уравнения в отдельности, если только уравнения не являются «не связанными» друг с другом в смысле (4.4.3). В случае условной гомоскедастичности и общего набора инструментов (4.4.3) превращается в

$$\sigma_{mh}\operatorname{E}(\boldsymbol{x}_i \boldsymbol{x}_i') = \mathbf{0} \text{ для всех } m \neq h.$$

Поскольку $\operatorname{E}(\boldsymbol{x}_i \boldsymbol{x}_i')$ не может быть нулевой матрицей (она предполагается невырожденной), уравнения являются «не связанными» друг

[1] В главе 8 будет показано, что эта оценка является также оценкой максимального правдоподобия.

с другом, только если $\sigma_{mh} = 0$. Следовательно, SUR более эффективна, чем OLS, если $\sigma_{mh} \neq 0$ для некоторой пары (m, h). Если $\sigma_{mh} = 0$ для всех (m, h), то эти две оценки асимптотически эквивалентны (разность оценок, умноженная на $\sqrt{n}$, стремится к нулю).

Эффективность SUR можно также увидеть, если посмотреть на эту модель как на многомерную регрессию с **априорными** исключающими ограничениями. В качестве примера расширим систему из двух уравнений из примера 4.3 следующим образом:

$$LW_i = \phi_1 + \beta_1 S_i + \gamma_1 IQ_i + \pi EXPR_i + \varepsilon_{i1},$$
$$KWW_i = \phi_2 + \beta_2 S_i + \gamma_2 IQ_i + \pi_2 EXPR_i + \varepsilon_{i2}.$$

Здесь общий набор инструментов $(1, S_i, IQ_i, EXPR_i)$. Поэтому эта система из двух уравнений есть многомерная регрессия с одинаковыми регрессорами в обоих уравнениях. Но если накладывается априорное ограничение в виде равенства π_2 нулю и, следовательно, $EXPR_i$ исключается из второго уравнения, то указанная модель превращается в SUR-модель. SUR-оценка более эффективна по сравнению с оценкой многомерной регрессии, поскольку она использует исключающие ограничения.

Соотношение между различными методами оценивания, рассмотренными в данном параграфе, показано в табл. 4.2.

Контрольные вопросы

1. (FIVE без условной гомоскедастичности.) Состоятельна ли оценка $\widehat{S}$ в (4.5.3) для S без требования условной гомоскедастичности? [Ответ: Нет. Почему?] Состоятельна ли оценка (4.3.2) для (4.5.2) при условной гомоскедастичности? [Ответ: Да. Почему?]

2. (FIVE без условной гомоскедастичности.) Является ли оценка FIVE состоятельной и асимптотически нормальной без требования условной гомоскедастичности? А эффективной? **Указание:** *FIVE-оценка есть эффективная GMM-оценка системы уравнений* (4.2.3) *для некоторой* $\widehat{\boldsymbol{W}}$. Удовлетворяет ли $\widehat{\boldsymbol{W}}$ условию эффективности $\widehat{\boldsymbol{W}} = \boldsymbol{S}^{-1}$?

3. (Когда FIVE-оценка и 2SLS-оценка каждого уравнения в отдельности численно одинаковы.) Предположим, что все уравнения идентифицируемы точно. Являются ли в таком случае FIVE-оценка и 2SLS-оценка каждого уравнения в отдельности численно эквивалентными (то есть умноженная на $\sqrt{n}$ оценок разность этих оценок стремится к нулю)? А численно равными? **Указание:** *Обе оценки FIVE и 2SLS сводятся к IV.*

4. (Когда FIVE-оценка и 2SLS-оценка каждого уравнения в отдельности асимптотически эквивалентны.) Если ошибки являются ортогональными ($\sigma_{mn} = 0$ для $m \neq h$), то FIVE-оценка и 2SLS-оценка каждого уравнения в отдельности

Таблица 4.2

Соотношения между оценками системы уравнений

	GMM для нескольких уравнений	**FIVE**	**3SLS**	**SUR**	**Многомерная регрессия**
Модель	Предположения 4.1–4.6	Предположения 4.1–4.5, 4.7 $\mathrm{E}(\boldsymbol{z}_{im}\boldsymbol{z}_{ih}')$ конечно	Предположения 4.1–4.5, 4.7 $\mathrm{E}(\boldsymbol{z}_{im}\boldsymbol{z}_{ih}')$ конечно $\boldsymbol{x}_{im} = \boldsymbol{x}_i$ для всех m	Предположения 4.1–4.5, 4.7 $\boldsymbol{x}_{im} = \boldsymbol{x}_i$ для всех m, $\boldsymbol{x}_i$ = объединение $\boldsymbol{z}_{i1}, \ldots, \boldsymbol{z}_{iM}$	Предположения 4.1–4.5, 4.7, $\boldsymbol{x}_{im} = \boldsymbol{x}_i$ для всех m, $\boldsymbol{z}_{im} = \boldsymbol{x}_i$ для всех m
$\boldsymbol{S}(\equiv \mathrm{Avar}(\overline{\boldsymbol{g}}))$	(4.1.11)	(4.5.2)	$\Sigma \otimes \mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')$	$\Sigma \otimes \mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')$	неактуально
$\widehat{\boldsymbol{S}}$	(4.3.2)	(4.5.3)	$\widehat{\Sigma} \otimes (\boldsymbol{n}^{-1}\Sigma_i\boldsymbol{x}_i\boldsymbol{x}_i')$ $\widehat{\Sigma}$ из 2SLS-остатков	$\widehat{\Sigma} \otimes (\boldsymbol{n}^{-1}\Sigma_i\boldsymbol{x}_i\boldsymbol{x}_i')$ $\widehat{\Sigma}$ из OLS-остатков	неактуально
$\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1})$	упрощений нет	упрощений нет	(4.5.12) с (4.5.13), (4.5.13)	(4.5.12) с (4.5.13′), (4.5.14′)	поочередное OLS-оценивание каждого уравнения
$\mathrm{Avar}\,\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1})$	(4.3.3)	(4.3.3)	(4.5.15) с (4.5.16)	(4.5.15) с (4.5.16′)	формула для OLS
$\widehat{\mathrm{Avar}}\,\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1})$	(4.3.4)	(4.3.4)	(4.5.17) с (4.5.13)	(4.5.17) с (4.5.13′)	формула для OLS

асимптотически эквивалентны (то есть разность этих оценок, умноженная на $\sqrt{n}$, стремится к нулю). **Указание:** *Наложим требование условной гомоскедастичности на утверждение 4.3(b). При условной гомоскедастичности эффективный GMM для системы уравнений сводится к FIVE и эффективный GMM для каждого уравнения в отдельности сводится к 2SLS для каждого уравнения в отдельности. Также при условной гомоскедастичности уравнения являются «несвязанными» в смысле (4.4.3), если $\sigma_{mn} = 0$.*

5. (Условная гомоскедастичность при предположениях SUR.) Проверьте, что при предположении SUR (4.5.18) предположение 4.7 принимает вид:

$$\mathrm{E}(\varepsilon_{im}\varepsilon_{ih}|\boldsymbol{z}_{i1},\ldots,\boldsymbol{z}_{iM}) = \sigma_{mh} \text{ для всех } m,h = 1,2,\ldots,M.$$

6. Что произойдет с SUR-оценкой из примера 4.7, если $\boldsymbol{x}_i$ будет включать в себя также MED_i? **Указание:** *$\boldsymbol{x}_i$ по-прежнему «исчезает» из* (4.5.13), (4.5.14) *и* (4.5.16).

   А что произойдет со статистикой Саргана?

7. (Идентифицируемость в SUR.) В утверждении 4.6 условие идентифицируемости (предположение 4.4) не является необходимым, так как оно вытекает из других предположений. Проверьте, что предположение 4.4 вытекает из предположения 4.5 и 4.7. **Указание:** *Из предположения 4.5 и разложения Кронекера для $\boldsymbol{S}$ следует, что $\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')$ невырождена. $\boldsymbol{z}_{im}$ есть подмножество $\boldsymbol{x}_i$.*

8. (SUR без условной гомоскедастичности.)

   (a) Являются ли SUR-оценки состоятельными и асимптотически нормальными без требования условной гомоскедастичности? **Указание:** *SUR-оценка по-прежнему представляет собой GMM-оценку системы уравнений, так что можно использовать утверждение 3.1.*

   (b) Можно ли свести SUR к многомерной регрессии и в случае, когда система идентифицирована точно, то есть точно идентифицировано каждое уравнение (так что $\boldsymbol{z}_{im} = \boldsymbol{x}_i$), без требования условной гомоскедастичности? [Ответ: Да.]

9. (Роль условий «перекрестной» ортогональности.) Предположим, что вместо (4.5.18′) условия ортогональности выглядят как $\mathrm{E}(\boldsymbol{z}_{im}\cdot\varepsilon_{im}) = \boldsymbol{0}$ для $m = 1,2,\ldots,M$. Является ли SUR-оценка состоятельной в этом случае? [Ответ: Необязательно.] Найдите эффективную GMM-оценку системы уравнений без требования условной гомоскедастичности. [Ответ: Это OLS.]

10. Проверьте, что статистика Саргана для многомерной регрессии равняется нулю.

4.6. Общие коэффициенты

На практике, особенно в работе с панельными данными, часто приходится иметь дело с частным случаем системы уравнений, в котором количество регрессоров одинаково во всех уравнениях (соответственно, с одинаковым набором коэффициентов). Такая модель представляет собой частный случай модели из нескольких уравнений из параграфа 4.1. Здесь показано, как применять GMM к нескольким уравнениям, налагая общие ограничения на коэффициенты. В итоге будет показано, что модель с кажущимся наличием ограничений (*seemingly restrictive model*) в действительности включает в себя как частный случай модель из нескольких уравнений *без* общих ограничений на коэффициенты.

Модель с общими коэффициентами

Система уравнений с общими коэффициентами превращается в

Предположение 4.1′ (линейность): *Оцениванию подлежат M линейных уравнений:*

$$y_{im} = z'_{im}\delta + \varepsilon_{im}(m = 1, 2, \ldots, M; i = 1, 2, \ldots, n), \tag{4.6.1}$$

где n — размер выборки, z_{im} — L-мерный вектор регрессоров, $\boldsymbol{\delta}$ — L-мерный вектор коэффициентов, общий для всех уравнений, а ε_{im} — ненаблюдаемая ошибка в m-м уравнении.

Среди других предположений модели нескольких уравнений, предположений 4.2–4.6, только предположение 4.4 (идентифицируемость) требует изменений, отражающих наличие ограничения в виде общих коэффициентов. Вариантом $\boldsymbol{g}(\boldsymbol{w}_i; \widetilde{\boldsymbol{\delta}})$ для нескольких уравнений является:

$$\boldsymbol{g}(\boldsymbol{w}_i; \widetilde{\boldsymbol{\delta}}) \equiv \begin{bmatrix} \boldsymbol{x}_{i1} \cdot (y_{i1} - \boldsymbol{z}'_{i1}\widetilde{\boldsymbol{\delta}}) \\ \vdots \\ \boldsymbol{x}_{iM} \cdot (y_{iM} - \boldsymbol{z}'_{iM}\widetilde{\boldsymbol{\delta}}) \end{bmatrix}, \tag{4.6.2}$$

так что $\mathrm{E}\,[\boldsymbol{g}(\boldsymbol{w}_i; \widetilde{\boldsymbol{\delta}})]$ превращается в

$$\mathrm{E}[\boldsymbol{g}(\boldsymbol{w}_i; \widetilde{\boldsymbol{\delta}})] = \begin{bmatrix} \mathrm{E}(\boldsymbol{x}_{i1} \cdot y_{i1}) \\ \vdots \\ \mathrm{E}(\boldsymbol{x}_{iM} \cdot y_{iM}) \end{bmatrix} - \begin{bmatrix} \mathrm{E}(\boldsymbol{x}_{i1} \cdot \boldsymbol{z}'_{i1})\widetilde{\boldsymbol{\delta}} \\ \vdots \\ \mathrm{E}(\boldsymbol{x}_{iM} \cdot \boldsymbol{z}'_{iM})\widetilde{\boldsymbol{\delta}} \end{bmatrix} =$$

$$= \underset{(\sum_{m=1}^{M} K_m \times 1)}{\boldsymbol{\sigma}_{\boldsymbol{xy}}} - \underset{(\sum_{m=1}^{M} K_m \times L)}{\boldsymbol{\Sigma}_{\boldsymbol{xz}}} \underset{(L \times 1)}{\widetilde{\boldsymbol{\delta}}}. \tag{4.6.3}$$

где

$$\boldsymbol{\sigma}_{\boldsymbol{xy}} \equiv \begin{bmatrix} \mathrm{E}(\boldsymbol{x}_{i1} \cdot y_{i1}) \\ \vdots \\ \mathrm{E}(\boldsymbol{x}_{iM} \cdot y_{iM}) \end{bmatrix}, \quad \boldsymbol{\Sigma}_{\boldsymbol{xz}} \equiv \begin{bmatrix} \mathrm{E}(\boldsymbol{x}_{i1}\boldsymbol{z}'_{i1}) \\ \vdots \\ \mathrm{E}(\boldsymbol{x}_{iM}\boldsymbol{z}'_{iM}) \end{bmatrix}. \tag{4.6.4}$$

То есть $\boldsymbol{\Sigma}_{\boldsymbol{xy}}$ теперь является штабелированной, а не блочно-диагональной матрицей. С этим изменением система уравнений, определяющая $\widetilde{\boldsymbol{\delta}}$, снова есть (4.1.10), так что условием идентифицируемости является

Предположение 4.4′ (идентифицируемость при общих коэффициентах): *Матрица $\boldsymbol{\Sigma_{xz}}$ размерности $\sum_{m=1}^{M} K_m \times L$, определенная в (4.6.4), имеет полный ранг.*

Это условие слабее, чем предположение 4.4, которое требует, чтобы каждое уравнение системы было идентифицируемо. На самом деле достаточным условием идентифицируемости является полный столбцовый ранг матрицы $\mathrm{E}(\boldsymbol{x}_{im}\boldsymbol{z}'_{im})$ *для некоторого*[1] m. Коэффициенты во всех уравнениях одни и те же благодаря именно *априорным* ограничениям. Система может быть идентифицируемой, даже если ни одно уравнение в отдельности не является идентифицируемым.

GMM-оценка

Оставляем читателю проверить, что $\boldsymbol{g}_n(\widetilde{\boldsymbol{\delta}})$ можно записать как $\boldsymbol{s_{zx}}-\boldsymbol{S_{zx}}\widetilde{\boldsymbol{\delta}}$ с

$$\boldsymbol{s_{xy}} \equiv \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_{i1}\cdot y_{i1} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_{iM}\cdot y_{iM} \end{bmatrix}, \quad \boldsymbol{S_{xz}} \equiv \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_{i1}\boldsymbol{z}'_{i1} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_{iM}\boldsymbol{z}'_{iM} \end{bmatrix}, \tag{4.6.5}$$

так что GMM-оценка со взвешивающей матрицей $\widehat{\boldsymbol{W}}$ размера $\sum_m K_m \times \sum_m K_m$ есть (4.2.3) с ошибкой оценки из (4.2.4). При условии, что $\boldsymbol{\Sigma_{xz}}$ и $\boldsymbol{S_{xz}}$ переопределены соответствующим образом и $\boldsymbol{\delta}$ интерпретируется как набор общих коэффициентов, теория больших выборок для модели с несколькими уравнениями та же самая, что и для случая самостоятельных коэффициентов из параграфа 4.3 с заменой предположения 4.1 на предположение 4.1′, предположения 4.4 на предположение 4.4′ и остатками $\widehat{\varepsilon}_i$, переопределенными как $y_{im} - \boldsymbol{z}'_{im}\widehat{\boldsymbol{\delta}}$ (а не $y_{im} - \boldsymbol{z}'_{im}\widehat{\boldsymbol{\delta}}_m$) для некоторой состоятельной оценки L-мерного вектора общих коэффициентов $\boldsymbol{\delta}$.

Чтобы связать данную GMM-оценку с другими популярными оценками, имеющимися в литературе, нужно подробно расписать формулу для данной GMM-оценки. Подставив (4.6.5) в (4.2.3), получим:

$$\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{W}}) = \left(\begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_{i1}\boldsymbol{x}'_{i1} & \cdots & \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_{iM}\boldsymbol{x}'_{iM} \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{W}}_{11} & \cdots & \widehat{\boldsymbol{W}}_{1M} \\ \vdots & & \vdots \\ \widehat{\boldsymbol{W}}_{M1} & \cdots & \widehat{\boldsymbol{W}}_{MM} \end{bmatrix} \times \right.$$

[1]**Доказательство:** Если $\mathrm{E}(\boldsymbol{x}_{im}\boldsymbol{z}'_{im})$ имеет полный столбцовый ранг, то ее ранг равен L. Поскольку ранг матрицы равен количеству линейно независимых строк, $\mathrm{E}(\boldsymbol{x}_{im}\boldsymbol{z}'_{im})$ имеет L линейно независимых строк. Поэтому $\boldsymbol{\Sigma_{xz}}$ имеет не менее L линейно независимых строк.

$$\times \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_{i1}\boldsymbol{z}_{i1}' \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_{iM}\boldsymbol{z}_{iM}' \end{bmatrix} \Bigg)^{-1} \left[\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_{i1}\boldsymbol{x}_{i1}' \;\cdots\; \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_{iM}\boldsymbol{x}_{iM}'\right]$$

$$\times \begin{bmatrix} \widehat{\boldsymbol{W}}_{11} & \cdots & \widehat{\boldsymbol{W}}_{1M} \\ \vdots & & \vdots \\ \widehat{\boldsymbol{W}}_{M1} & \cdots & \widehat{\boldsymbol{W}}_{MM} \end{bmatrix} \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_{i1}\cdot y_{i1} \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_{iM}\cdot y_{iM} \end{bmatrix} =$$

$$= \left[\sum_{m=1}^{M}\sum_{h=1}^{M}\left\{\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_{im}\boldsymbol{x}_{im}'\right)\widehat{\boldsymbol{W}}_{mh}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{ih}\boldsymbol{z}_{ih}'\right)\right\}\right]^{-1} \times$$

$$\times \left[\sum_{m=1}^{M}\sum_{h=1}^{M}\left\{\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_{im}\boldsymbol{x}_{im}'\right)\widehat{\boldsymbol{W}}_{mh}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{ih}\cdot y_{ih}\right)\right\}\right], \qquad (4.6.6)$$

где $\widehat{\boldsymbol{W}}_{mh}$ есть блок (m, h) матрицы $\widehat{\boldsymbol{W}}$. (Для получения второго равенства используется формула (A.8) из приложения A.) Эффективную GMM-оценку можно получить, если заменить в данном выражении $\widehat{\boldsymbol{W}}$ на матрицу, обратную матрице $\widehat{\boldsymbol{S}}$, определенной в (4.3.2).

Введение требования условной гомоскедастичности

Эффективную GMM-оценку можно получить, положив в (4.6.6) $\widehat{\boldsymbol{W}} = \widehat{\boldsymbol{S}}^{-1}$, где $\widehat{\boldsymbol{S}}$ есть состоятельная оценка матрицы $\boldsymbol{S}$. Как и ранее, мы можем наложить требование условной гомоскедастичности, чтобы свести GMM-оценку к более известным оценкам. Если задать $\widehat{\boldsymbol{S}}$ соотношением (4.5.3), то получится FIVE-оценка.

Если мы предположим еще, что набор инструментов одинаков для всех уравнений, то тогда, как и ранее, эту $\widehat{\boldsymbol{S}}$ можно представить как произведение Кронекера:

$$\widehat{\boldsymbol{S}} = \widehat{\boldsymbol{\Sigma}} \otimes \left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\right),$$

так что $\widehat{\boldsymbol{W}}_{mh}$ в (4.6.6) задается соотношением (4.5.11), что дает нам так называемую 3SLS-оценку с общими коэффициентами:

$$\left[\sum_{m=1}^{M}\sum_{h=1}^{M}\left\{\widehat{\sigma}^{mh}\cdot\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_{im}\boldsymbol{x}_i'\right)\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{z}_{ih}'\right)\right\}\right]^{-1} \times$$

$$\times \left[\sum_{m=1}^{M}\sum_{h=1}^{M}\left\{\widehat{\sigma}^{mh}\cdot\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_{im}\boldsymbol{x}_i'\right)\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\cdot y_{ih}\right)\right\}\right]. \qquad (4.6.7)$$

где $\widehat{\sigma}^{mh}$ — блок (m, h) матрицы $\widehat{\Sigma}^{-1}$. Если, в дополнение, выполняется условие SUR (4.5.18), то тогда происходит «исчезновение $\boldsymbol{x}$» в (4.6.7) и эффективная GMM-оценка превращается в так называемую (по историческим причинам) **оценку со случайными эффектами** (*random-effects estimator*):

$$\widehat{\boldsymbol{\delta}}_{\mathrm{RE}} = \left[\sum_{m=1}^{M}\sum_{h=1}^{M} \widehat{\sigma}^{mh}\Big(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_{im}\boldsymbol{z}'_{ih}\Big)\right]^{-1} \sum_{m=1}^{M}\sum_{h=1}^{M} \widehat{\sigma}^{mh}\Big(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_{im}\cdot y_{ih}\Big). \tag{4.6.8}$$

Ее асимптотическая дисперсия

$$\mathrm{Avar}(\widehat{\boldsymbol{\delta}}_{\mathrm{RE}}) = \left[\sum_{m=1}^{M}\sum_{h=1}^{M} \sigma^{mh}\, \mathrm{E}(\boldsymbol{z}_{im}\boldsymbol{z}'_{ih})\right]^{-1}. \tag{4.6.9}$$

(Чтобы получить это выражение, вернемся к общей формуле (4.3.3) для асимптотической дисперсии эффективной GMM-оценки. Положим $\boldsymbol{\Sigma}_{\boldsymbol{xz}}$ как в (4.6.4), $\boldsymbol{S} = \boldsymbol{\Sigma} \otimes \mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}'_i)$, воспользуемся формулой (A.8) из приложения A для вычисления произведения матриц и обнаружим «исчезновение $\boldsymbol{x}$», что превращает (4.5.16) в (4.5.16′) на с. 309 при условии (4.5.18) для SUR.) Для этой асимптотической дисперсии можно получить состоятельную оценку:

$$\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\delta}}_{\mathrm{RE}})} = \left[\sum_{m=1}^{M}\sum_{h=1}^{M} \widehat{\sigma}^{mh}\Big(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_{im}\boldsymbol{z}'_{ih}\Big)\right]^{-1}. \tag{4.6.10}$$

Подытожим результат об оценке случайных эффектов.

Утверждение 4.7 (свойства на больших выборках оценки случайных эффектов): *Предположим, что предположения 4.1′, 4.2, 4.3, 4.4′, 4.5 и 4.7 выполняются с* $\boldsymbol{x}_i$ = *объединение* $(\boldsymbol{z}_{i1}, \ldots, \boldsymbol{z}_{iM})$. *Пусть* $\boldsymbol{\Sigma}$ — *матрица размера* $M\times M$, *элемент* (m, h) *которой равен* $\mathrm{E}(\varepsilon_{im}\,\varepsilon_{ih})$, *и пусть* $\widehat{\boldsymbol{\Sigma}}$ *является состоятельной оценкой* $\boldsymbol{\Sigma}$. *Тогда*

(a) *Оценка* $\widehat{\boldsymbol{\delta}}_{\mathrm{RE}}$, *заданная в* (4.6.8), *является состоятельной, асимптотически нормальной и эффективной с* $\mathrm{Avar}(\widehat{\boldsymbol{\delta}}_{\mathrm{RE}})$, *заданной в* (4.6.9).

(b) *Оцененная асимптотическая дисперсия* (4.6.10) *является состоятельной для* $\mathrm{Avar}(\widehat{\boldsymbol{\delta}}_{\mathrm{RE}})$.

(c) *(Статистика Саргана.)*

$$J(\widehat{\boldsymbol{\delta}}_{\mathrm{RE}}, \widehat{\boldsymbol{S}}^{-1}) \equiv n\cdot \boldsymbol{g}_n(\widehat{\boldsymbol{\delta}}_{\mathrm{RE}})'\,\widehat{\boldsymbol{S}}^{-1}\,\boldsymbol{g}_n(\widehat{\boldsymbol{\delta}}_{\mathrm{RE}}) \underset{d}{\to} \chi^2(MK-L).$$

где $\widehat{\boldsymbol{S}} = \widehat{\boldsymbol{\Sigma}} \otimes (\frac{1}{n}\sum \boldsymbol{x}_i\boldsymbol{x}'_i)$, K — *количество общих инструментов и* $\boldsymbol{g}_n(\widetilde{\boldsymbol{\delta}}) = \boldsymbol{s}_{\boldsymbol{xy}} - \boldsymbol{S}_{\boldsymbol{xz}}\widetilde{\boldsymbol{\delta}}$ *с* $\boldsymbol{s}_{\boldsymbol{xy}}$, *а* $\boldsymbol{S}_{\boldsymbol{xz}}$ *задано в* (4.6.5).

Ниже мы перепишем формулы (4.6.8)–(4.6.10) в более элегантном виде.

OLS в модели пула

Для SUR из предыдущего параграфа мы построили $\widehat{\mathbf{\Sigma}}$ из остатков, полученных OLS для каждого уравнения. Состоятельность — это все, что требуется для $\widehat{\mathbf{\Sigma}}$, поэтому аналогичная процедура для $\widehat{\mathbf{\Sigma}}$ работает здесь точно так же, но распределение $\widehat{\delta}_{\mathrm{RE}}$ на конечных выборках можно улучшить, если мы используем априорное ограничение о том, что коэффициенты одинаковы во всех уравнениях в оценивании $\widehat{\mathbf{\Sigma}}$. Приравняем поэтому $\widehat{\boldsymbol{W}}$ в GMM-оценивании на первом шаге для получения начальной состоятельной оценки $\boldsymbol{\delta}$

$$\boldsymbol{I}_M \otimes \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'\right)^{-1},$$

а не

$$\widehat{\mathbf{\Sigma}}^{-1} \otimes \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'\right)^{-1}.$$

Эта оценка дается формулой (4.6.8) с $\widehat{\sigma}^{mh} = 1$ для $m = h$ и 0 для $m \neq h$, что можно записать как

$$\begin{aligned}\widehat{\boldsymbol{\delta}}_{\text{pooled OLS}} &= \left[\sum_{m=1}^{M}\left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_{im}\boldsymbol{z}_{im}'\right)\right]^{-1} \sum_{m=1}^{M}\left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_{im}\cdot y_{im}\right) = \\ &= \left(\sum_{i=1}^{n}\sum_{m=1}^{M} \boldsymbol{z}_{im}\boldsymbol{z}_{im}'\right)^{-1} \sum_{i=1}^{n}\sum_{m=1}^{M} \boldsymbol{z}_{im}\cdot y_{im}, \end{aligned} \tag{4.6.11}$$

что является просто OLS-оценкой по выборке объема Mn, где наблюдения из разных уравнений объединяются (*are pooled across equations*). По этой причине оценка называется **объединенной OLS** (*pooled OLS*) оценкой. Из этого выражения следует, что используемыми условиями ортогональности являются условия

$$\mathrm{E}(\boldsymbol{z}_{i1}\cdot \varepsilon_{i1} + \cdots + \boldsymbol{z}_{iM}\cdot \varepsilon_{iM}) = \mathbf{0}, \tag{4.6.12}$$

которые не включают «кросс»-ортогональности $\mathrm{E}(\boldsymbol{z}_{im}\cdot \varepsilon_{ih}) = 0$ $(m \neq h)$. В качестве аналитического упражнения остается показать, что объединенная OLS-оценка является GMM-оценкой, использующей (4.6.12).

Поскольку объединенную OLS-оценку легко вычислить, она является популярной опцией для тех исследователей, которые не хотят программировать получение оценок. Эта оценка также робастна к нарушению условий «кросс»-ортогональности. Однако важно иметь в виду то, что стандартные ошибки, которые приводятся в пакетах программ и которые не принимают во внимание перекрестные моменты внутри уравнений (σ_{mh}), являются смещенными. Для объединенной OLS-оценки, которая является GMM-оценкой с неоптимальным выбором $\widehat{\boldsymbol{W}}$, корректной формулой для асимптотической дисперсии является (3.5.1) в утверждении 3.1. Полагая $\boldsymbol{W} = \boldsymbol{I}_M \otimes [\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')]^{-1}$, $\boldsymbol{S} = \boldsymbol{\Sigma} \otimes \mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')$, $\boldsymbol{\Sigma}_{\boldsymbol{xz}}$ как в (4.6.4) и снова наблюдая исчезновение $\boldsymbol{x}$, мы получаем:

$$\begin{aligned}\mathrm{Avar}(\widehat{\boldsymbol{\delta}}_{\text{pooled OLS}}) = &\left[\sum_{m=1}^{M} \mathrm{E}(\boldsymbol{z}_{im}\boldsymbol{z}_{im}')\right]^{-1} \times \\ &\times \sum_{m=1}^{M}\sum_{h=1}^{M} \sigma_{mh}\,\mathrm{E}(\boldsymbol{z}_{im}\boldsymbol{z}_{ih}') \left[\sum_{m=1}^{M} \mathrm{E}(\boldsymbol{z}_{im}\boldsymbol{z}_{im}')\right]^{-1}, \end{aligned} \tag{4.6.13}$$

которая состоятельно оценивается посредством

$$\left(\sum_{m=1}^{M}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_{im}\boldsymbol{z}_{im}'\right)^{-1} \sum_{m=1}^{M}\sum_{h=1}^{M}\widehat{\sigma}_{mh}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_{im}\boldsymbol{z}_{ih}'\left(\sum_{m=1}^{M}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_{im}\boldsymbol{z}_{im}'\right)^{-1}. \tag{4.6.14}$$

Поскольку объединенная OLS является состоятельной оценкой, соответствующие остатки можно использовать для вычисления $\widehat{\sigma}_{mh}$ $(m, h = 1, \ldots, M)$ в этом выражении. Корректными стандартными ошибками для объединенной OLS будут квадратные корни из (умноженных на $1/n$) диагональных элементов этой матрицы.

Приведение формул к более красивому виду

Приведенные формулы для оценки случайных эффектов и объединенной OLS-оценки вывести достаточно просто, но выглядят они с двойным и тройным суммированием довольно сложными. Их можно, однако, привести к более красивому виду — и в то же время к более полезному для программирования, — если ввести новые матричные обозначения. Пусть

$$\underset{(M\times 1)}{\boldsymbol{y}_i} = \begin{bmatrix} y_{i1} \\ \vdots \\ y_{iM} \end{bmatrix}, \quad \underset{(M\times L)}{\boldsymbol{Z}_i} = \begin{bmatrix} \boldsymbol{z}_{i1}' \\ \vdots \\ \boldsymbol{z}_{iM}' \end{bmatrix}, \quad \underset{(M\times 1)}{\boldsymbol{\varepsilon}_i} = \begin{bmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iM} \end{bmatrix}, \tag{4.6.15}$$

так что систему из M уравнений с общими коэффициентами в предположении 4.1$'$ можно записать как

$$\boldsymbol{y}_i = \boldsymbol{Z}_i\boldsymbol{\delta} + \boldsymbol{\varepsilon}_i \quad (i = 1, 2, \ldots, n). \tag{4.6.1$'$}$$

Приведение к более красивому виду сложных формул использует следующие алгебраические результаты:

$$\sum_{m=1}^{M} \boldsymbol{z}_{im}\boldsymbol{z}'_{im} = \boldsymbol{Z}'_i\boldsymbol{Z}_i, \quad \sum_{m=1}^{M} \boldsymbol{z}_{im} \cdot y_{im} = \boldsymbol{Z}'_i\boldsymbol{y}_i. \tag{4.6.16a}$$

$$\sum_{m=1}^{M}\sum_{h=1}^{M} c_{mh} \cdot \boldsymbol{z}_{im} \cdot y_{ih} = \boldsymbol{Z}'_i\boldsymbol{C}\boldsymbol{y}_i, \quad \sum_{m=1}^{M}\sum_{h=1}^{M} c_{mh} \cdot \boldsymbol{z}_{im}\boldsymbol{z}'_{ih} = \boldsymbol{Z}'_i\boldsymbol{C}\boldsymbol{Z}_i \tag{4.6.16b}$$

для любой $M \times M$ матрицы $\boldsymbol{C} = (c_{mh})$. Возьмем теперь тройное суммирование в (4.6.8):

$$\sum_{m=1}^{M}\sum_{h=1}^{M} \widehat{\sigma}^{mh} \cdot \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_{im} \cdot y_{ih}\right) =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{m=1}^{M}\sum_{h=1}^{M} \widehat{\sigma}^{mh} \cdot \boldsymbol{z}_{im} \cdot y_{ih}\right)$$

(изменяя порядок суммирования) =

$$= \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{Z}'_i\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{y}_i \quad (\text{по (4.6.16b) с } \boldsymbol{C} = \widehat{\boldsymbol{\Sigma}}^{-1}). \tag{4.6.17}$$

Аналогично по (4.6.16b) с $\boldsymbol{C} = \widehat{\boldsymbol{\Sigma}}^{-1}$, другие тройные суммирования в (4.6.8) принимают вид:

$$\sum_{m=1}^{M}\sum_{h=1}^{M} \widehat{\sigma}^{mh}\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_{im}\boldsymbol{z}'_{ih} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{Z}'_i\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{Z}_i. \tag{4.6.18}$$

Таким же образом можно записать двойные и тройные суммирования в (4.6.9) и (4.6.10). Мы имеем тогда переписанные формулы для случайных эффектов в виде:

$$\widehat{\boldsymbol{\delta}}_{\mathrm{RE}} = \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{Z}'_i\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{Z}_i\right)^{-1}\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{Z}'_i\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{y}_i, \tag{4.6.8'}$$

$$\mathrm{Avar}(\widehat{\boldsymbol{\delta}}_{\mathrm{RE}}) = \left(\mathrm{E}(\boldsymbol{Z}'_i\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}_i)\right)^{-1}, \tag{4.6.9'}$$

$$\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\delta}}_{\mathrm{RE}})} = \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{Z}'_i\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{Z}_i\right)^{-1}. \tag{4.6.10'}$$

Используя (4.6.16a), формулы для объединенной OLS-оценки можно записать как

$$\widehat{\boldsymbol{\delta}}_{\text{pooled OLS}} = \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{Z}_i'\boldsymbol{Z}_i\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{Z}_i'\boldsymbol{y}_i, \tag{4.6.11'}$$

$$\operatorname{Avar}(\widehat{\boldsymbol{\delta}}_{\text{pooled OLS}}) = [\mathrm{E}(\boldsymbol{Z}_i'\boldsymbol{Z}_i)]^{-1}\,\mathrm{E}(\boldsymbol{Z}_i'\boldsymbol{\Sigma}\boldsymbol{Z}_i)[\mathrm{E}(\boldsymbol{Z}_i'\boldsymbol{Z}_i)]^{-1}, \tag{4.6.13'}$$

$$\widehat{\operatorname{Avar}(\widehat{\boldsymbol{\delta}}_{\text{pooled OLS}})} = \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{Z}_i'\boldsymbol{Z}_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{Z}_i'\widehat{\boldsymbol{\Sigma}}\boldsymbol{Z}_i\right)\left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{Z}_i'\boldsymbol{Z}_i\right)^{-1}. \tag{4.6.14'}$$

Эти формулы не просто симпатичны; как будет видно в следующем параграфе, они будут полезны для обработки определенных классов панельных данных (так называемых **несбалансированных панельных** (*unbalanced panel*)) данных).

Ограничение, которое не является ограничивающим

Хотя кажется, что модель в этом параграфе с предположением об общих коэффициентах является частным случаем модели из параграфа 4.1, последнюю можно привести к формату модели данного параграфа при надлежащем переопределении регрессоров. Рассмотрим, например, пример 4.1. Вместо определения $\boldsymbol{z}_{i1}$ и $\boldsymbol{z}_{i2}$ как в (4.1.2), определим их как

$$\boldsymbol{z}_{i1} = \begin{bmatrix} 1 \\ S_i \\ IQ_i \\ EXPR_i \\ 0 \\ 0 \\ 0 \end{bmatrix}, \qquad \boldsymbol{z}_{i2} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ S_i \\ IQ_i \end{bmatrix} \qquad \text{с} \quad \boldsymbol{\delta} = \begin{bmatrix} \phi_1 \\ \beta_1 \\ \gamma_1 \\ \pi \\ \phi_2 \\ \beta_2 \\ \gamma_2 \end{bmatrix}.$$

Тогда система из двух уравнений соответствует формату предположения 4.1′. (См. контрольный вопрос 5 ниже для демонстрации того же самого в более общем случае.)

Более того, систему, в которой общие коэффициенты имеет только некоторое подмножество переменных, можно записать в виде предположения 4.1′. Рассмотрим пример 4.2. Допустим, что мы хотим *априорно* предположить, что коэффициенты при образовании, IQ и стаже остаются постоянными во времени (так что $\beta_1 = \beta_2 = \beta$, $\gamma_1 = \gamma_2 = \gamma$, $\pi_1 = \pi_2 = \pi$), но константы различны. Преобразование осуществляется следующим образом:

$$z_{i1} = \begin{bmatrix} 1 \\ 0 \\ S69_i \\ IQ_i \\ EXPR69_i \end{bmatrix}, \quad z_{i2} = \begin{bmatrix} 0 \\ 1 \\ S80_i \\ IQ_i \\ EXPR80_i \end{bmatrix}, \quad \delta = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \beta \\ \gamma \\ \pi \end{bmatrix}. \tag{4.6.19}$$

Следовательно, ограничение, требующее общих коэффициентов, не является ограничивающим. Мы могли бы получить GMM-оценку для модели из параграфа 4.1 как частный случай GMM-оценки с ограничением, требующим общих коэффициентов. Но модель без такого ограничения, возможно, является более простым введением в многомерные уравнения.

Контрольные вопросы

1. (Идентификация с общими коэффициентами.) Пусть z_{imh} является j-м элементом $\boldsymbol{z}_{im}$. То есть z_{imj} является j-м регрессором в m-м уравнении. Предположим, что $z_{im1} = 1$ для всех i, m. Какое из предположений, сделанных для этой модели, нарушается, если $z_{im2} = 1$ для всех i, m? **Указание:** *Посмотрите на ранговое условие.* Изменится ли ваш ответ, если $z_{im2} = m$ для всех i, m? **Указание:** *Должен измениться.*

2. (Важность «кросс»-ортогональностей.) Предположим, что $\mathrm{E}(\boldsymbol{Z}_{im} \cdot \varepsilon_{ih}) = 0$ для $m = h$, но не обязательно для $m \neq h$. Будет ли оценка случайных эффектов состоятельной? **Указание:** *Ошибка оценки случайных эффектов равна:*

$$\widehat{\delta}_{\mathrm{RE}} - \boldsymbol{\delta} = \left[\sum_{m=1}^{M}\sum_{h=1}^{M} \widehat{\sigma}^{mh}\left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_{im}\boldsymbol{z}'_{ih}\right)\right]^{-1} \sum_{m=1}^{M}\sum_{h=1}^{M} \widehat{\sigma}^{mh}\left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_{im} \cdot \varepsilon_{ih}\right).$$

3. (Неэффективность объединенной OLS-оценки.) Выведите эффективную GMM-оценку для $\boldsymbol{\delta}$, которая использует тот факт, что $\mathrm{E}(\boldsymbol{z}_{im} \cdot \varepsilon_{im}) = 0$ для всех m. Является ли она той же самой, что и объединенная OLS-оценка? [Ответ: Нет.] Является ли она такой же, что и оценка случайных эффектов? [Ответ: Нет.]

4. ($\boldsymbol{\Sigma}_{xz}$, $\boldsymbol{s}_{xz}$, $\boldsymbol{S}_{xz}$ с общими инструментами в кронекеровских произведениях.) Предположим, что инструменты во всех уравнениях одинаковы: $\boldsymbol{x}_{im} = \boldsymbol{x}_i$. Покажите, что $\boldsymbol{\Sigma}_{xz}$ в (4.6.4) можно записать как $\mathrm{E}(\boldsymbol{Z}_i \otimes \boldsymbol{x}_i)$, где $\boldsymbol{Z}_i$ является $M \times L$ матрицей, определенной в (4.6.15). Также $\boldsymbol{s}_{xz} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{y}_i \otimes \boldsymbol{x}_i$, $\boldsymbol{S}_{xz} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{Z}_i \otimes \boldsymbol{x}_i$. **Указание:** $\boldsymbol{x}_i\boldsymbol{z}'_{im} = \boldsymbol{z}'_{im} \otimes \boldsymbol{x}_i$.

5. (Ограничение, не являющееся ограничивающим.) Чтобы исключить возможную путаницу, возникающую из того факта, что $\boldsymbol{z}_{im}$ в параграфе 4.1 и $\boldsymbol{z}_{im}$ в этом параграфе различны, запишем систему из M уравнений (4.1.11) как

$$y_{im} = \boldsymbol{z}^{*\prime}_{im}\boldsymbol{\delta}^{*}_{m} + \varepsilon_{im} \quad (m = 1, 2, \ldots, M;\ i = 1, 2, \ldots, n), \tag{1}$$

где $\boldsymbol{z}^{*}_{im}$ и $\boldsymbol{\delta}^{*}_{m}$ являются L_m-мерными.

(a) (Предположение 4.1 как частный случай предположения 4.1′.) Проверьте, что (1) можно записать как (4.6.1), если определить

$$
\underset{(\sum_{m=1}^{M} L_m \times 1)}{z_{im}} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ z^*_{im} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \begin{array}{l} \} \ L_1 \text{ строк} \\ \qquad \vdots \\ \} \ L_{m-1} \text{ строк} \\ \} \ L_m \text{ строк} \\ \} \ L_{m+1} \text{ строк} \\ \qquad \vdots \\ \} \ L_M \text{ строк} \end{array}
$$

$$
\text{и} \quad \underset{(\sum_{m=1}^{M} L_m \times 1)}{\delta} = \begin{bmatrix} \delta^*_1 \\ \vdots \\ \delta^*_M \end{bmatrix}. \tag{2}
$$

(b) (Предположение 4.4 как частный случай предположения 4.4′.) Проверьте, что, когда z_{im} определяется как в (2) с z^*_{im}, интерпретируемым как z_{im} в параграфе 4.1, предположение 4.4′ становится предположением 4.4.

(c) (Формула GMM.) При такой же интерпретации проверьте, что матрица S_{xz} в (4.6.5) становится блочно-диагональной, m-й блок которой равен $\frac{1}{n}\Sigma_i x_{im} z^{*\prime}_{im}$, так что (4.6.6) становится (4.2.6), если z_{im} в (4.2.6) понимается как z^*_{im}.

6. (Идентификация в RE-модели.) Предположение 4.4′ в утверждении 4.7 фактически не является необходимым. Проверьте, что из предположений 4.5 и 4.7 следует предположение 4.4′. **Указание:** *Если* $\mathrm{E}(x_i x_i')$ *невырождена, то предположение 4.4′ выполняется, как показано в контрольном вопросе 7 параграфа 4.5. Предположение 4.4′ более слабое, чем предположение 4.4.*

7. (Только для тех, кто знаком с кронекеровским произведением.)

(a) Покажите, что $\mathrm{E}(Z_i'\Sigma^{-1}Z_i)$ в (4.6.9′) на с. 321 для оценки случайных эффектов является невырожденной, если выполняются предположения утверждения 4.7. **Указание:**

$$
\mathrm{E}(Z_i'\Sigma^{-1}Z_i) = \sum_{m=1}^{M}\sum_{h=1}^{M} \sigma^{mh}\,\mathrm{E}(z_{im}z_{ih}') =
$$

$$
= \sum_{m=1}^{M}\sum_{h=1}^{M} \sigma^{mh}\,\mathrm{E}(z_{im}x_i')[\mathrm{E}(x_i x_i')]^{-1}\,\mathrm{E}(x_i z_{ih}') =
$$

(x_i может снова появиться, поскольку x_i включает z_{im}, z_{ih})

$$
= \Sigma_{xz}'(\Sigma^{-1} \otimes [\mathrm{E}(x_i x_i')]^{-1})\Sigma_{xz}.
$$

Σ_{xz} имеет полный полный столбцовый ранг по предположению 4.4′.

(b) Покажите, что $\mathrm{E}(Z_i'Z_i)$ в (4.6.13′) на с. 322 для объединенного OLS в модели пула является обратимой. **Указание:** *Замените* Σ^{-1} *на* I.

4.7. Приложение. Спрос на взаимосвязанные факторы

Транслогарифмическая функция издержек (*translog cost function*) представляет собой обобщение функции издержек вида Кобба — Дугласа (логлинейной), рассмотренной в параграфе 1.7. Привлекательной чертой транслогарифмической спецификации является то, что уравнения спроса на факторы, минимизирующие издержки (преобразованные в уравнения для долей затрат на факторы в издержках), являются линейными по логарифмам выпуска и ценам факторов. При этом коэффициенты представляют собой (часть) параметров функции издержек, определяющих технологию. Эти параметры технологии могут быть оценены из системы уравнений для долей затрат. Одними из первых эту идею воплотили [Berndt and Wood, 1975] и [Christensen and Greene, 1976]. Этот параграф посвящен изложению основ их подхода. Данный подход также может быть применен для анализа потребительского спроса.

Чтобы лучше сконцентрироваться на интересующем нас вопросе и следовать логике этих основополагающих работ, дальнейшее изложение выстроено при предположении об условной гомоскедастичности. Кроме того, мы будем предполагать, что регрессоры (логарифмы цен факторов и логарифм выпуска) в системе уравнений для долей являются предопределенными. Как мы уже обсуждали в параграфе 1.7, это предположение является вполне обоснованным для отрасли производства и распределения электроэнергии США до недавнего дерегулирования отрасли. Следовательно, подходящим методом оценивания нескольких уравнений является модель многомерной регрессии.

Транслогарифмическая функция издержек

Мы уже частично совершили переход от логлинейной к транслогарифмической функции издержек: в эмпирическом упражнении главы 1 мы реализовали идею добавить квадрат логарифма выпуска в логлинейную функцию издержек (см. модель 4). Если этот подход расширить и добавить квадраты и перекрестные произведения логарифмов всех аргументов, то логлинейная функция издержек с тремя аргументами (1.7.4) становится транслогарифмической функцией издержек:

$$\log(C) = \alpha_0 + \sum_{j=1}^{3} \alpha_j \log(p_j) + \frac{1}{2}\sum_{j=1}^{3}\sum_{k=1}^{3} \gamma_{jk} \log(p_j)\log(p_k) +$$
$$+ \alpha_Q \log(Q) + \frac{1}{2}\gamma_{QQ}(\log(Q))^2 + \sum_{j=1}^{3} \gamma_{jQ} \log(p_j)\log(Q) + \varepsilon. \quad (4.7.1)$$

Здесь, как и в оставшейся части параграфа, мы опускаем индекс наблюдений «i». В этом выражении слагаемое

$$\frac{1}{2}\sum_{j=1}^{3}\sum_{k=1}^{3}\gamma_{jk}\log(p_j)\log(p_k)$$

является квадратичной формой, представляющей вклад факторов второго порядка. Без потери общности можно предположить, что матрица коэффициентов квадратичной формы размера 3×3, (γ_{jk}) симметрична[1]:

$$\gamma_{jk} = \gamma_{kj} \quad (j, k = 1, 2, 3). \tag{4.7.2}$$

В случае логлинейной функции издержек, рассмотренной в параграфе 1.7, отдача от масштаба может быть вычислена как единица, деленная на эластичность издержек по выпуску. Если это определение применить к транслогарифмической функции издержек, получим, что

$$\begin{aligned}\text{отдача от масштаба} &= \frac{1}{\partial\log(C)/\partial log(Q)} = \\ &= \frac{1}{\alpha_Q + \gamma_{QQ}\log(Q) + \sum_{j=1}^{3}\gamma_{jQ}\log(p_j)}.\end{aligned} \tag{4.7.3}$$

Доли факторов

Связь между параметрами функции издержек и спросом на факторы определяется **леммой Шепарда** (*Shephard's Lemma*) из микроэкономики. Пусть x_j — спрос на фактор j, минимизирующий функцию издержек при фиксированных ценах факторов (p_1, p_2, p_3) и выпуске Q. То есть $\sum_{j=1}^{3} p_j x_j = C$. Лемма гласит, что

$$\frac{\partial C}{\partial p_j} = x_j. \tag{4.7.4}$$

Используя выражение

$$\frac{\partial\log(C)}{\partial\log(p_j)} = \frac{p_j}{C}\cdot\frac{\partial C}{\partial p_j},$$

лемму можно переформулировать как равенство частной производной логарифма функции издержек доле фактора, а именно

$$\frac{\partial\log(C)}{\partial\log(p_j)} = \frac{p_j x_j}{C}. \tag{4.7.5}$$

[1]Пусть $\boldsymbol{x}$ — n-мерный вектор. Тогда $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}$ — соответствующая квадратичная форма. Поскольку $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}'\boldsymbol{A}'\boldsymbol{x}$, то $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}'\left[(\boldsymbol{A}+\boldsymbol{A}')/2\right]\boldsymbol{x}$. Таким образом, если матрица $\boldsymbol{A}$ несимметрична, ее можно заменить симметричной матрицей $(\boldsymbol{A}+\boldsymbol{A}')/2$, не изменив при этом значение квадратичной формы.

В случае транслогарифмической функции (4.7.1) частная производная логарифма функции затрат вычисляется легко:

$$\frac{\partial \log(C)}{\partial \log(p_j)} = \alpha_j + \sum_{k=1}^{3} \gamma_{jk} \log(p_k) + \gamma_{jQ} \log(Q) \quad (j = 1, 2, 3). \tag{4.7.6}$$

Совмещая это с (4.7.5) и определяя доли затрат на факторы как $s_j \equiv \equiv p_j x_j / C$, получим следующую систему для долей затрат на факторы для транслогарифмической функции:

$$s_j = \alpha_j + \sum_{k=1}^{3} \gamma_{jk} \log(p_k) + \gamma_{jQ} \log(Q) \quad (j = 1, 2, 3). \tag{4.7.7}$$

На эту систему уравнений наложены **перекрестные ограничения** (*cross-equation restrictions*), состоящие в том, что коэффициент перед $\log(p_k)$ в уравнении для s_j равен коэффициенту перед $\log(p_j)$ в уравнении для s_k для $j \neq k$. До конца параграфа мы будем называть эти ограничения ограничениями симметрии, хотя они и не являются следствием предположения о симметрии (4.7.2). Эти перекрестные ограничения являются следствием вычислений: проверьте, что без предположения о симметрии коэффициент при $\log(p_2)$ в уравнении для s_1, например, будет равен $(\gamma_{12} + \gamma_{21})/2$ и будет совпадать с коэффициентом при $\log(p_1)$ в уравнении для s_2.

Эластичности замещения

В литературе, посвященной производственной функции, большое внимание уделяется оцениванию эластичностей замещения (точнее, частных эластичностей замещения Хикса — Аллена) между различными факторами производства[1]. Как показано в [Uzawa, 1962], эластичность замещения между факторами j и k, обозначаемая как η_{jk}, связана с функцией издержек C посредством следующего соотношения:

$$\eta_{jk} = \frac{C \cdot \dfrac{\partial^2 C}{\partial p_j \partial p_k}}{\dfrac{\partial C}{\partial p_j} \cdot \dfrac{\partial C}{\partial p_k}}. \tag{4.7.8}$$

Для транслогарифмической функции издержек применение леммы Шепарда дает следующий результат:

$$\eta_{jk} = \begin{cases} \dfrac{\gamma_{jk} + s_j s_k}{s_j s_k} & \text{для } j \neq k \\[2ex] \dfrac{\gamma_{jj} + {s_j}^2 - s_j}{{s_j}^2} & \text{для } j = k. \end{cases} \tag{4.7.9}$$

[1]История развития данного направления изложена в [Berndt, 1991, Section 9.2].

где s_j — доля затрат на фактор j в общих издержках. В частном случае логлинейной функции издержек, где $\gamma_{jk} = 0$ для всех j, k, получаем $\eta_{jk} = 1$ для $j \neq k$, что подтверждает то, что для технологии Кобба — Дугласа эластичности замещения между любыми двумя факторами равны единице.

Свойства функции издержек

Как хорошо известно из микроэкономики (см., например: [Varian, 1992, Chapter 5]), функция издержек обладает следующими свойствами.

1. (Однородность.) Однородность первой степени по ценам факторов.

2. (Монотонность.) Неубывание по ценам факторов, для чего необходимо, чтобы $\partial C/\partial p_j \geq 0$ для всех j.

3. (Вогнутость.) Вогнутость по ценам факторов, для чего необходимо, чтобы гессиан (матрица вторых производных) функции издержек C,

$$\boldsymbol{H} \equiv (\partial^2 C/\partial p_j \partial p_k),$$

был отрицательно полуопределен. Из (4.7.8) следует, что требование вогнутости выполнено если и только если матрица эластичностей замещения (η_{jk}) отрицательно полуопределена[1].

В случае транслогарифмической функции издержек эти свойства можно переформулировать следующим образом.

1. *Однородность:* Несложно проверить (см. контрольные вопросы), что при введении ограничений симметрии на (γ_{jk}) условие однородности может быть переписано как:

$$\begin{cases} \alpha_1 + \alpha_2 + \alpha_3 = 1, \\ \gamma_{11} + \gamma_{12} + \gamma_{13} = 0, \\ \gamma_{12} + \gamma_{22} + \gamma_{23} = 0, \\ \gamma_{13} + \gamma_{23} + \gamma_{33} = 0, \\ \gamma_{1Q} + \gamma_{2Q} + \gamma_{3Q} = 0. \end{cases} \tag{4.7.10}$$

Мы будем пользоваться предположением об однородности при оценке уравнений для долей.

[1]Пусть $\boldsymbol{H}$ — гессиан функции издержек C и $\boldsymbol{V}$ — диагональная матрица, j-й элемент которой равен $\partial C/\partial p_j$. Тогда матрица эластичностей замещения (η_{jk}), определенная в (4.7.8), может быть переписана в более компактном виде: $(\eta_{jk}) = C \cdot \boldsymbol{V}^{-1}\boldsymbol{H}\boldsymbol{V}^{-1}$. Поскольку $\boldsymbol{V}$ — диагональная матрица, то $\boldsymbol{V}^{-1}\boldsymbol{H}\boldsymbol{V}^{-1}$ отрицательно полуопределена, если и только если отрицательно полуопределена матрица $\boldsymbol{H}$.

2. *Монотонность:* Принимая во внимание лемму Шепарда, свойство монотонности эквивалентно требованию, что правые части уравнений для долей (4.7.7) неотрицательны для любых комбинаций цен факторов и выпуска. Для любых значений коэффициентов всегда возможно выбрать цены факторов и объем выпуска таким образом, чтобы это требование нарушалось. Следовательно, мы не можем ввести требование монотонности для уравнений для долей затрат. Однако возможно проверить, выполнено ли требование монотонности для заданных значений (*relevant range*). Иными словами, выполнено ли это требование для всех комбинаций цен факторов и объемов выпуска, присутствующих в конкретном наборе данных. Мы сделаем это для оцененных параметров ниже.

3. *Вогнутость:* Вогнутость требует, чтобы матрица эластичностей замещения (4.7.9) была отрицательно полуопределена для любой комбинации долей затрат на факторы. Можно показать, что необходимым и достаточным условием вогнутости является отрицательная полуопределенность матрицы (γ_{jk}) размера[1] 3×3.

   В частности, диагональные элементы $(\gamma_{11}, \gamma_{22}, \gamma_{33})$ должны быть неположительными. Условие отрицательной полуопределенности матрицы (γ_{jk}) представляет собой набор неравенств для ее элементов. Оценивание при введении ограничений в виде неравенств возможно (см.: [Jorgenson, 1986, Section 2.4.5] для описания методики), но мы не будем рассматривать его здесь.

Стохастические спецификации

Транслогарифмическая функция издержек (4.7.1) включает в себя (аддитивную) ошибку, ε, в то время как уравнения долей затрат, выведенные из нее, не включают ошибку. При использовании нашей базы данных уравнения для долей не выполняются в точности для каждой фирмы, следовательно, мы должны включить ошибки. Обычным является включение случайного возмущения в каждое из уравнений для долей. Существуют два способа обоснования подобной стохастической спецификации. Один из них — это ошибки оптимизации: фирмы совершают

[1] Доказательство: Обозначим через $\boldsymbol{s}$ вектор размера 3×1, j-й элемент которого равен s_j, а $\boldsymbol{S}$ — диагональная матрица, j-й элемент которой равен s_j. Тогда матрица эластичностей замещения, определенная в (4.7.9), может быть записана как $(\eta_{jk}) = \boldsymbol{S}^{-1}[(\gamma_{jk}) + \boldsymbol{s}\boldsymbol{s}' - \boldsymbol{S}]\boldsymbol{S}^{-1}$. Для некоторой комбинации долей затрат выражение $\boldsymbol{s}\boldsymbol{s}' - \boldsymbol{S}$ равно нулю (например, для $\boldsymbol{s} = (1, 0, 0)'$). То есть необходимым условием отрицательной полуопределенности матрицы (η_{jk}) является отрицательная полуопределенность матрицы (γ_{jk}). Это условие также является и достаточным, так как матрица $\boldsymbol{s}\boldsymbol{s}' - \boldsymbol{S}$ отрицательно полуопределена, и сумма двух отрицательно определенных матриц также является отрицательно полуопределенной.

случайные ошибки при выборе комбинаций факторов производства, минимизирующих издержки. Но тогда фактические общие издержки будут выше затрат, соответствующих функции издержек, поскольку при выводе функции издержек не предусматриваются ошибки оптимизации. Другим обоснованием наличия ошибок является допущение о том, что α_j в уравнениях долей является стохастическим и различным для различных фирм. В таком случае константой в уравнении доли фактора j в общих издержках будет математическое ожидание α_j, а ошибкой — отклонение α_j от своего математического ожидания. Но это также означает, что α_j в функции издержек следует интерпретировать как случайную величину.

Поэтому в обоих случаях внутренняя согласованность не позволила бы нам включить функцию издержек в систему уравнений для долей с аддитивными ошибками, чтобы оценивать все уравнения совместно. До конца параграфа мы будем рассматривать систему уравнений, состоящую только из уравнений для долей затрат факторов. Однако это значит, что мы можем оценить лишь часть параметров функции издержек. Два параметра (α_Q и γ_{QQ}), которые отсутствуют в функциях спроса на факторы, не могут быть оценены. Так как отдача от масштаба зависит от них (см. (4.7.3)), она также не может быть оценена.

Природа ограничений

После добавления аддитивных ошибок в систему уравнений для долей издержек на факторы производства система уравнений (4.7.7) записывается следующим образом:

$$\begin{cases} s_1 = \alpha_1 + \gamma_{11}\log(p_1) + \gamma_{12}\log(p_2) + \gamma_{13}\log(p_3) + \gamma_{1Q}\log(Q) + \varepsilon_1, \\ s_2 = \alpha_2 + \gamma_{21}\log(p_1) + \gamma_{22}\log(p_2) + \gamma_{23}\log(p_3) + \gamma_{2Q}\log(Q) + \varepsilon_2, \\ s_3 = \alpha_3 + \gamma_{31}\log(p_1) + \gamma_{32}\log(p_2) + \gamma_{33}\log(p_3) + \gamma_{3Q}\log(Q) + \varepsilon_3. \end{cases} \tag{4.7.11}$$

В этой системе пятнадцать коэффициентов, или параметров. Перекрестные ограничения (выраженные в том, что матрица логарифмов коэффициентов при ценах размера 3×3, γ_{jk}, симметрична), и ограничения однородности (4.7.10) вместе дают восемь ограничений. Таким образом, пятнадцать параметров могут быть описаны семью неизвестными параметрами. Чтобы понять это, полезно разделить ограничения на следующие три группы.

$$\text{ограничения в виде суммы:} \quad \begin{cases} \alpha_1 + \alpha_2 + \alpha_3 = 1, \\ \gamma_{11} + \gamma_{21} + \gamma_{31} = 0, \\ \gamma_{12} + \gamma_{22} + \gamma_{32} = 0, \\ \gamma_{13} + \gamma_{23} + \gamma_{33} = 0, \\ \gamma_{1Q} + \gamma_{2Q} + \gamma_{3Q} = 0. \end{cases} \tag{4.7.12}$$

$$\text{однородность:} \begin{cases} \gamma_{11} + \gamma_{12} + \gamma_{13} = 0, \\ \gamma_{21} + \gamma_{22} + \gamma_{23} = 0, \\ \gamma_{31} + \gamma_{32} + \gamma_{33} = 0, \end{cases} \tag{4.7.13}$$

$$\text{симметрия:} \begin{cases} \gamma_{12} = \gamma_{21}, \\ \gamma_{13} = \gamma_{31}, \\ \gamma_{23} = \gamma_{32}. \end{cases} \tag{4.7.14}$$

Здесь одиннадцать уравнений, но только восемь из них являются ограничительными: учитывая ограничения в виде сумм, одно из трех ограничений однородности следует из двух других; при условии ограничений в виде сумм и однородности два из трех ограничений симметрии следуют из третьего (проверьте это).

Модель многомерной регрессии с перекрестными ограничениями

Введение ограничений однородности (4.7.13) осуществляется непосредственно, поскольку они не являются перекрестными. Мы можем устранить три параметра, скажем $(\gamma_{13}, \gamma_{23}, \gamma_{33})$, и получить следующую систему:

$$\begin{cases} s_1 = \alpha_1 + \gamma_{11}\log(p_1/p_3) + \gamma_{12}\log(p_2/p_3) + \gamma_{1Q}\log(Q) + \varepsilon_1, \\ s_2 = \alpha_2 + \gamma_{21}\log(p_1/p_3) + \gamma_{22}\log(p_2/p_3) + \gamma_{2Q}\log(Q) + \varepsilon_2, \\ s_3 = \alpha_3 + \gamma_{31}\log(p_1/p_3) + \gamma_{32}\log(p_2/p_3) + \gamma_{3Q}\log(Q) + \varepsilon_3. \end{cases} \tag{4.7.15}$$

Особенность этой системы (вне зависимости от введения требования однородности) в том, что сумма зависимых переменных (s_1, s_2, s_3) дает единицу для всех наблюдений. Учитывая это, а также ограничения в виде сумм, получаем, что сумма ошибок $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ равна нулю для любого наблюдения. То есть ковариационная матрица ошибок $\boldsymbol{\Sigma} \equiv \mathrm{Var}(\varepsilon_1, \varepsilon_2, \varepsilon_3)$ размера 3×3 вырождена. Вспомним, что в модели многомерной регрессии (которая является частным случаем SUR-модели, описанной в утверждении 4.6) матрица $\boldsymbol{S}$, определенная в (4.1.11), может быть записана как Кронекерово произведение: $\boldsymbol{S} = \boldsymbol{\Sigma} \otimes \mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')$. Таким образом, $\boldsymbol{S}$ становится вырожденной, что нарушает предположение 4.5. С этой проблемой вырожденности обычно справляются путем исключения одного уравнения из системы и оценивания оставшихся уравнений (ниже этот вопрос рассмотрен более подробно). Коэффициенты опущенного уравнения могут быть вычислены с использованием оцененных коэффициентов и ограничений в виде сумм. Таким образом, учитываются как ограничения в виде сумм, так и ограничения однородности.

Как уже упоминалось выше, при учете ограничений в виде сумм и однородности остается лишь одно ограничение симметрии. Например, если мы исключили из системы коэффициенты $(\gamma_{13}, \gamma_{23}, \gamma_{33})$ при введении ограничений однородности (как это было сделано в (4.7.15)) и если мы опустили третье уравнение при введении ограничений в виде сумм, то тогда единственное остающееся ограничение симметрии — это $\gamma_{12} = \gamma_{21}$.

При введении этого перекрестного ограничения система принимает вид:

$$\begin{cases} s_1 = \alpha_1 + \gamma_{11}\log(p_1/p_3) + \gamma_{12}\log(p_2/p_3) + \gamma_{1Q}\log(Q) + \varepsilon_1, \\ s_2 = \alpha_2 + \gamma_{12}\log(p_1/p_3) + \gamma_{22}\log(p_2/p_3) + \gamma_{2Q}\log(Q) + \varepsilon_2. \end{cases} \tag{4.7.16}$$

Поскольку регрессоры являются предопределенными и поскольку уравнения содержат общий набор регрессоров, система может быть оценена с помощью модели многомерной регрессии с перекрестным ограничением симметрии. Это оценивание с ограничениями может быть преобразовано в безусловное оценивание после приведения к виду «с общими коэффициентами», изложенному в параграфе 4.6. То есть система (4.7.16) может быть переписана в форме «с общими коэффициентами» $\boldsymbol{y}_i = \boldsymbol{Z}_i\boldsymbol{\delta} + \boldsymbol{\varepsilon}_i$ (см. (4.6.1′) на с. 320), если мы определим матрицы следующим образом (мы все еще опускаем индекс «i» для наблюдений в $s_1, s_2, \log(p1/p3), \log(p2/p3), \log(Q)$):

$$\boldsymbol{y}_i = \begin{bmatrix} s_1 \\ s_2 \end{bmatrix},$$

$$\boldsymbol{Z}_i = \begin{bmatrix} 1 & 0 & \log(p_1/p_3) & \log(p_2/p_3) & 0 & \log(Q) & 0 \\ 0 & 1 & 0 & \log(p_1/p_3) & \log(p_2/p_3) & 0 & \log(Q) \end{bmatrix},$$

$$\boldsymbol{\delta}' = \begin{bmatrix} \alpha_1 & \alpha_2 & \gamma_{11} & \gamma_{12} & \gamma_{22} & \gamma_{1Q} & \gamma_{2Q} \end{bmatrix}. \tag{4.7.17}$$

Таким образом, оценивание многомерной регрессии при условии симметрии равносильно оцениванию методом случайных эффектов без ограничений.

Это оценивание методом случайных эффектов указанной системы из двух уравнений дает состоятельные оценки следующих семи параметров:

$$\alpha_1; \alpha_2, \gamma_{11}, \gamma_{12}; \gamma_{22}; \gamma_{1Q}; \gamma_{2Q}.$$

Остальные из пятнадцати параметров уравнений для долей затрат,

$$\alpha_3, \gamma_{13}; \gamma_{21}, \gamma_{23}, \gamma_{31}; \gamma_{32}; \gamma_{32}; \gamma_{33}, \gamma_{3Q},$$

могут быть вычислены с использованием ограничений в виде сумм (4.7.12), однородности (4.7.13) и симметрии (4.7.14). Например, γ_{33} можно вычислить так:

$$\gamma_{33} = -\gamma_{31} - \gamma_{32} = -\gamma_{13} - \gamma_{23} = (\gamma_{11} + \gamma_{12}) + (\gamma_{12} + \gamma_{22}) = \gamma_{11} + 2\gamma_{12} + \gamma_{22}.$$

То есть точечная оценка γ_{33} равна:

$$\widehat{\gamma}_{33} = \widehat{\gamma}_{11} + 2\widehat{\gamma}_{12} + \widehat{\gamma}_{22},$$

где $(\widehat{\gamma}_{11}, \widehat{\gamma}_{12}, \widehat{\gamma}_{22})$ — оценки параметров, полученные при оценивании методом случайных эффектов. Стандартная ошибка $\widehat{\gamma}_{33}$ может быть вычислена применением «дельта-метода» (см. лемму 2.5) к состоятельной оценке $\text{Avar}(\widehat{\gamma}_{11}, \widehat{\gamma}_{12}, \widehat{\gamma}_{22})$, полученной при оценивании методом случайных эффектов.

Какое уравнение исключать

Имеет ли значение, какое из уравнений системы исключать? Очевидно, что ответ был бы отрицательным, если бы не было перекрестных ограничений, так как многомерная регрессия численно эквивалентна OLS-оцениванию каждого уравнения в отдельности. Посмотрим, выполняется ли численная эквивалентность при наличии перекрестного ограничения. Пусть $\mathbf{\Sigma}^*$ — ковариационная матрица ошибок размера 2×2 для системы из двух уравнений, полученной при исключении одного уравнения из системы (4.7.15). Это соответствующая подматрица ковариационной матрицы ошибок $\mathbf{\Sigma}$ размера 3×3. (Например, если исключить третье уравнение, то тогда $\mathbf{\Sigma}^*$ — ведущая подматрица размера 2×2 матрицы $\mathbf{\Sigma}$ размера 3×3.) Чтобы реализовать оценивание методом случайных эффектов, нам необходимо найти состоятельную оценку $(\widehat{\mathbf{\Sigma}}^*)$ матрицы $\widehat{\mathbf{\Sigma}}$. Существуют два способа вычисления $\widehat{\mathbf{\Sigma}}^*$:

- (OLS-оценивание каждого уравнения.) Оцените два уравнения по отдельности, таким образом игнорируя перекрестное ограничение, а затем используйте полученные остатки для вычисления $\widehat{\mathbf{\Sigma}}^*$. Равносильно можно произвести раздельное OLS-оценивание *трех* уравнений и использовать полученные при этом остатки для вычисления $\widehat{\mathbf{\Sigma}}$ (оценки ковариационной матрицы $\mathbf{\Sigma}$ размера 3×3), а затем извлечь необходимую подматрицу матрицы $\widehat{\mathbf{\Sigma}}$. Если матрица $\widehat{\mathbf{\Sigma}}^*$ вычисляется подобным образом, то численная эквивалентность гарантирована (вы проверите это в эмпирическом упражнении), и не имеет значения, какое из уравнений системы исключать.

- Получите $\widehat{\mathbf{\Sigma}}^*$ с помощью некоторого метода, учитывающего перекрестные ограничения, связывающего разные уравнения (например, объединенного OLS, примененного к системе, приведенной к форме с общими коэффициентами). Численная эквивалентность в этом случае не гарантируется.

Численная эквивалентность выполнена в случае применения OLS для каждого уравнения по отдельности, поскольку матрица $\widehat{\mathbf{\Sigma}}$, из которой извлекается $\widehat{\mathbf{\Sigma}}^*$, является вырожденной в конечных выборках. В больших выборках оценка $\widehat{\mathbf{\Sigma}}^*$ становится состоятельной, и, даже если $\widehat{\mathbf{\Sigma}}^*$ получается не из OLS-оценивания уравнений по отдельности, это взаимоотношение между $\widehat{\mathbf{\Sigma}}^*$ и $\widehat{\mathbf{\Sigma}}$ выполняется асимптотически, так как матрица $\mathbf{\Sigma}$ вырожденная. То есть инвариантность выполняется асимптотически, если $\widehat{\mathbf{\Sigma}}^*$ состоятельна. В частности, асимптотическая дисперсия оценок параметров не зависит от выбора исключаемого уравнения.

Результаты

Рассмотренная в главе 1 работа Нерлова использовала кросс-секционные данные за 1995 год для компаний США, занимающихся производством и распределением электроэнергии. [Christensen and Greene, 1976] анализировали похожие данные за 1970 год для 99 фирм, используя при этом источники данных, которые позволили им получить более точные меры для уровня заработной платы и цен на топливо (см. параграф V их работы для более подробного ознакомления). Их набор данных включает информацию о долях затрат на факторы, так что уравнения для долей затрат можно оценить. Средние значения и стандартные отклонения для некоторых переменных из их набора данных представлены в табл. 4.3, из которой видно, что в общих затратах на производство электроэнергии значительную долю составляют затраты на топливо.

Таблица 4.3

Простые статистики (объем выборки = 99)

	Производство электроэнергии, млрд кВт/ч	**Доля затрат труда**	**Доля затрат капитала**	**Доля затрат на топливо**
Среднее	9,0	0,141	0,227	0,633
Стандартное отклонение	10,3	0,059	0,062	0,092

Для расчета $\widehat{\Sigma}^*$ мы используем остатки, полученные после OLS-оценки каждого уравнения по отдельности, так что оценки, получаемые методом случайных эффектов, численно инвариантны к выбору удаляемого уравнения. Эти оценки приведены в табл. 4.4 наряду с $\widehat{\Sigma}$, полученной в результате применения OLS ко всем уравнениям по отдельности. Индекс фактора производства j равен 1 для труда, 2 — для капитала и 3 — для топлива. Некоторые (на самом деле все) диагональные элементы матрицы оценок коэффициентов при ценах $(\widehat{\gamma}_{jk})$ положительны. То есть матрица не является отрицательно полуопределенной, что противоречит вогнутости[1].

Даже если $(\widehat{\gamma}_{jk})$ не является отрицательно полуопределенной, соответствующие эластичности замещения, рассчитываемые по формуле (4.7.9), могут быть отрицательно полуопределенными. Поскольку эта формула выводилась в предположении отсутствия ошибок оптимизации, для s_j и s_k в этой формуле должны использоваться оценки долей затрат, а не

[1]Такое «тестирование» вогнутости может быть формализовано. Используя дельта-метод, мы можем найти асимптотическое распределение характеристических корней $(\widehat{\gamma}_{jk})$. Нулевая гипотеза вогнутости состоит в том, что все эти характеристические корни неположительны.

фактические доли затрат. Эластичности, усредненные по фирмам, приведены в табл. 4.5. (Они являются внедиагональными элементами симметричной матрицы эластичностей замещения размера 3×3, которая была найдена таким образом.) Указанные три фактора производства являются субститутами, поскольку эластичности замещения положительны. Однако степень их взаимозаменимости намного меньше по сравнению с логлинейной технологией. Требование вогнутости нарушается даже для имеющихся данных: для 20 фирм из 99, представленных в выборке, матрица эластичностей замещения не является отрицательно полуопределенной. Монотонность, напротив, выполняется для рассматриваемого набора данных: оцененные доли затрат неотрицательны для всех факторов производства и для всех фирм в выборке.

Таблица 4.4

Оценки методом случайных эффектов

Параметр	Точечная оценка	Стандартная ошибка	t-значение
α_1	−0,132	0,106	−1,25
α_2	0,318	0,085	3,75
α_3	0,813	0,094	8,69
γ_{11}	0,084	0,020	4,19
γ_{12}	−0,023	0,016	−1,46
γ_{13}	−0,060	0,015	−3,92
γ_{22}	0,122	0,020	6,19
γ_{23}	−0,099	0,017	−5,75
γ_{33}	0,159	0,023	6,90
γ_{1Q}	−0,0211	0,0025	−8,55
γ_{2Q}	−0,0086	0,0030	−2,87
γ_{3Q}	0,0297	0,0037	7,98

$$\widehat{\boldsymbol{\Sigma}} \text{ полученная посредством поочередного}$$
$$\text{OLS-оценивания каждого уравнения} =$$
$$= \begin{bmatrix} 0{,}00173 & -0{,}000171 & -0{,}00156 \\ -0{,}000171 & 0{,}00253 & -0{,}00236 \\ -0{,}00156 & -0{,}00236 & 0{,}00391 \end{bmatrix}.$$

Таблица 4.5

Эластичности замещения

Труд—капитал	Капитал—топливо	Труд—топливо
0,17	0,29	0,27

Контрольные вопросы

1. Выведите (4.7.6). Проверьте, что симметрия (γ_{jk}) используется при выводе.

2. Выведите ограничения однородности (4.7.10). **Указание:** *Для того чтобы транслогарифмическая функция издержек (4.7.1) была однородной первой степени по ценам факторов, необходимо и достаточно, чтобы для любого скаляра* $\lambda \neq 0$

$$\sum_{j=1}^{3} \alpha_j \log(\lambda \cdot p_j) + \frac{1}{2} \sum_{j=1}^{3} \sum_{k=1}^{3} \gamma_{jk} \log(\lambda \cdot p_j) \log(\lambda \cdot p_k) +$$
$$+ \alpha_Q \log(Q) + \frac{1}{2} \gamma_{QQ} (\log(Q))^2 + \sum_{j=1}^{3} \gamma_{jQ} \log(\lambda \cdot p_j) \log(Q) =$$
$$= \log(\lambda) + \sum_{j=1}^{3} \alpha_j \log(p_j) + \frac{1}{2} \sum_{j=1}^{3} \sum_{k=1}^{3} \gamma_{jk} \log(p_j) \log(p_k) +$$
$$+ \alpha_Q \log(Q) + \frac{1}{2} \gamma_{QQ} (\log(Q))^2 + \sum_{j=1}^{3} \gamma_{jQ} \log(p_j) \log(Q).$$

3. (Вывод ограничений.) Используя ограничения в виде сумм (4.7.12), два из трех уравнений (4.7.13) и одно из трех уравнений (4.7.14), выведите два других уравнения в (4.7.14).

4. (Для OLS вырожденность соблюдается.) Рассмотрим OLS-оценивание каждого из трех уравнений (4.7.15) по отдельности. Покажите, что полученные оценки коэффициентов удовлетворяют ограничениям в виде сумм. Покажите, что ковариационная матрица ошибок размера 3×3, вычисленная с помощью OLS-оценивания уравнений по отдельности, является вырожденной. **Указание:** *Пусть* $\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3$ *— n-мерные векторы трех зависимых переменных (где n — размер выборки), пусть $\boldsymbol{X}$ — матрица данных размера $n \times K$, содержащая K общих регрессоров ($K = 4$ в (4.7.15)). Заметьте, что* $\boldsymbol{y}_1 + \boldsymbol{y}_2 + \boldsymbol{y}_3 = \mathbf{1}$. *Оценки коэффициентов, полученные при OLS-оценивании каждого уравнения по отдельности, равны* $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}_j$, $j = 1, 2, 3$. *Три вектора остатков размера $n \times 1$ равны* $\boldsymbol{M}\boldsymbol{y}_j$, $j = 1, 2, 3$, *где* $\boldsymbol{M} \equiv \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$.

Набор задач для главы 4

Аналитические упражнения

1. (Представление 3SLS в матричном виде.) В модели 3SLS набор инструментальных переменных одинаков для всех уравнений. Пусть K — количество общих инструментов. Введем следующие опреде-

ления:

$$\underset{n\times K}{\boldsymbol{X}} = \begin{bmatrix} \boldsymbol{x}_1' \\ \vdots \\ \boldsymbol{x}_n' \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{1K} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nK} \end{bmatrix},$$

$$\underset{(n\times L_m)}{\boldsymbol{Z}_m} = \begin{bmatrix} \boldsymbol{z}_{1m}' \\ \vdots \\ \boldsymbol{z}_{nm}' \end{bmatrix}, \quad \underset{(Mn\times\sum_{m=1}^M L_m)}{\boldsymbol{Z}} = \begin{bmatrix} \boldsymbol{Z}_1 & & \\ & \ddots & \\ & & \boldsymbol{Z}_M \end{bmatrix},$$

$$\underset{(\sum_{m=1}^M L_m\times 1)}{\boldsymbol{\delta}} = \begin{bmatrix} \boldsymbol{\delta}_1 \\ \vdots \\ \boldsymbol{\delta}_M \end{bmatrix},$$

$$\underset{(Mn\times 1)}{\boldsymbol{y}} = \begin{bmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_M \end{bmatrix}, \quad \underset{(n\times 1)}{\boldsymbol{y}_m} = \begin{bmatrix} y_{1m} \\ \vdots \\ y_{nm} \end{bmatrix}.$$

$$\underset{(Mn\times 1)}{\boldsymbol{\varepsilon}} = \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_M \end{bmatrix}, \quad \underset{(n\times 1)}{\boldsymbol{\varepsilon}_m} = \begin{bmatrix} \varepsilon_{1m} \\ \vdots \\ \varepsilon_{nm} \end{bmatrix}.$$

Покажите следующее:

- Предположение 4.1 можно записать как $\boldsymbol{y} = \boldsymbol{Z\delta} + \boldsymbol{\varepsilon}$.
- $\widehat{\boldsymbol{S}}$ в (4.5.10) принимает вид $\widehat{\boldsymbol{S}} = \widehat{\boldsymbol{\Sigma}} \otimes \frac{1}{n}\boldsymbol{X}'\boldsymbol{X}$.
- $\widehat{\boldsymbol{\delta}}_{3SLS} = [\boldsymbol{Z}'(\widehat{\boldsymbol{\Sigma}}^{-1} \otimes \boldsymbol{P})\boldsymbol{Z}]^{-1}\boldsymbol{Z}'(\widehat{\boldsymbol{\Sigma}}^{-1} \otimes \boldsymbol{P})\boldsymbol{y}$, где $\boldsymbol{P} \equiv \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$.
- $\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\delta}}_{3SLS})} = n \cdot [\boldsymbol{Z}'(\widehat{\Sigma}^{-1} \otimes \boldsymbol{P})\boldsymbol{Z}]^{-1}$.
- $\widehat{\boldsymbol{A}}_{mh}$ в (4.5.13) принимает вид $\widehat{\boldsymbol{A}}_{mh} = \frac{1}{n}\boldsymbol{Z}_m'\boldsymbol{P}\boldsymbol{Z}_h$.
- $\widehat{\boldsymbol{c}}_{mh}$ в (4.5.14) принимает вид $\widehat{\boldsymbol{c}}_{mh} = \frac{1}{n}\boldsymbol{Z}_m'\boldsymbol{P}\boldsymbol{y}_h$.
- $J(\widehat{\boldsymbol{\delta}}_{3SLS}, \widehat{\boldsymbol{S}}^{-1}) = (\boldsymbol{y} - \boldsymbol{Z}\widehat{\boldsymbol{\delta}}_{3SLS})'(\widehat{\boldsymbol{\Sigma}}^{-1} \otimes \boldsymbol{P})(\boldsymbol{y} - \boldsymbol{Z}\widehat{\boldsymbol{\delta}}_{3SLS})$.

  Указание:

  $$\frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_i\boldsymbol{x}_i' = \frac{1}{n}\boldsymbol{X}'\boldsymbol{X}, \; \boldsymbol{s_{xy}} = (\boldsymbol{I}_M \otimes \boldsymbol{X})'\boldsymbol{y}, \; \boldsymbol{S_{xz}} = \frac{1}{n}(\boldsymbol{I}_M \otimes \boldsymbol{X})'\boldsymbol{Z}.$$

  Используйте формулы (A.11) и (A.13)–(A.15) приложения A.

2. (Представление в матричном виде SUR, RE и объединенного OLS.)

   (a) Докажите, что (4.5.13) сводится к (4.5.13′) (на с. 309), а (4.5.14) сводится к (4.5.14′) при предположении SUR (4.5.18). **Указание:** *Если* $\boldsymbol{P} \equiv \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$, *то* $\widehat{\boldsymbol{A}}_{mh} = \frac{1}{n}\boldsymbol{Z}_m'\boldsymbol{P}\boldsymbol{Z}_h, \widehat{\boldsymbol{c}}_{mh} = \frac{1}{n}\boldsymbol{Z}_m'\boldsymbol{P}\boldsymbol{y}_h$, $\boldsymbol{P}\boldsymbol{Z}_m = \boldsymbol{Z}_m, \boldsymbol{Z}_m$, *если столбцы* $\boldsymbol{Z}_m$ *сформированы из столбцов* $\boldsymbol{X}$.

(b) Покажите, что для SUR

$$\text{оценка} = [Z'(\widehat{\Sigma}^{-1} \otimes I_n)Z]^{-1}Z'(\widehat{\Sigma}^{-1} \otimes I_n)y,$$
$$\text{ее } \widehat{\text{Avar}} = n \cdot [Z'(\widehat{\Sigma}^{-1} \otimes I_n)Z]^{-1}.$$

(c) Пусть (только для этой части вопроса) $z_{im} = x_i$ для всех m. Тогда $Z = I_M \otimes X$. Покажите, что SUR сводится к оценке многомерной регрессии

$$[I_M \otimes (X'X)^{-1}X']y.$$

Проверьте, что m-й K-мерный блок этого MK-мерного вектора — это $(X'X)^{-1}X'y_m$.

Теперь определим Z так:

$$\underset{(Mn\times L)}{Z} \equiv \begin{bmatrix} Z_1 \\ \vdots \\ Z_M \end{bmatrix}. \qquad (*)$$

Тогда предположение 4.1′ (линейные уравнения с общими коэффициентами) может быть переписано как $y = Z\delta + \varepsilon$, где y и ε — те же, что и в аналитическом упражнении 1, и δ — L-мерный вектор общих коэффициентов.

(d) Покажите, что для оценки методом случайных эффектов верны формулы, выведенные вами в пункте (b) для SUR, и Z определяется как в (*). **Указание:** *Эта оценка приведена в* (4.6.8), *а ее* $\widehat{\text{Avar}}$ — *в* (4.6.10). *Используйте формулу (A.8) приложения A.*

(e) Покажите, что

$$\widehat{\delta}_{pooled\ OLS} \text{ в (4.6.11) } = (Z'Z)^{-1}Z'y.$$

$$\text{Avar}(\widehat{\widehat{\delta}_{pooled\ OLS}}) \text{ в (4.6.14) } = n \cdot (Z'Z)^{-1}[Z'(\widehat{\Sigma} \otimes I_n)Z](Z'Z)^{-}$$

(f) В матрице Z размера $Mn\times L$, определенной в (*), строки упорядочены сначала по номерам уравнений (m), а затем по наблюдениям (i). Теперь упорядочим строки сначала по наблюдениям (i), что приведет к переопределению Z:

$$\underset{(Mn\times L)}{Z} \equiv \begin{bmatrix} Z_1 \\ \vdots \\ Z_M \end{bmatrix}.$$

где $\boldsymbol{Z}_i(i = 1, 2, \ldots, n)$ — матрица размера $M \times L$, определенная в (4.6.15). Перегруппируйте ряды $\boldsymbol{y}$ и ε аналогичным образом:

$$\underset{(Mn\times 1)}{\boldsymbol{y}} = \begin{bmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_n \end{bmatrix}, \quad \underset{(Mn\times 1)}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

Как изменятся формулы для RE и объединенного OLS, которые вы вывели в пунктах (d) и (e)?

3. (GLS-интерпретация SUR и RE.) SUR-модель состоит из предположений 4.1–4.5, 4.7 и (4.5.18) (что $\boldsymbol{x}_i$ = объединение $(\boldsymbol{z}_{i1}, \ldots, \boldsymbol{z}_{iM})$). Уточним модель усилением предположения 4.3 (условия ортогональности) следующим образом:

$$\mathrm{E}(\varepsilon_{im}|\boldsymbol{z}_{i1}, \ldots, \boldsymbol{z}_{iM}) = 0 \quad (m = 1, \ldots, M),$$

а предположение 4.2 (эргодическая стационарность) как i.i.d.

$$\{y_{i1}, \ldots, y_{iM}, \boldsymbol{z}_{i1}, \ldots, \boldsymbol{z}_{iM}\}.$$

(a) (Просто.) Объясните, почему эти предположения являются более строгими.

(b) Пусть $\boldsymbol{\varepsilon}(Mn \times 1)$ и $\boldsymbol{Z}(Mn \times \sum_m L_m)$ определяются так же, как и в аналитическом упражнении 1 выше. Покажите, что в уточненной модели выполнено

$$\mathrm{E}(\boldsymbol{\varepsilon}|\boldsymbol{Z}) = \boldsymbol{0} \text{ и } \mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\boldsymbol{Z}) = \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n.$$

Указание: *Сначала необходимо показать, что* $\mathrm{E}(\varepsilon_{im}|\boldsymbol{Z}) = 0$. *Блок* (m, n) *матрицы* $\mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\boldsymbol{Z})$ *— это* $\mathrm{E}(\boldsymbol{\varepsilon}_m\boldsymbol{\varepsilon}_h'|\boldsymbol{Z})$. *Необходимо показать, что эта матрица равна* $\sigma_{mh}\boldsymbol{I}_n$.

(c) (Свойства SUR в конечных выборках.) Пусть матрица $\boldsymbol{\Sigma}$ известна, так что состоятельные оценки $\widehat{\sigma}_{mh}$ в формулах SUR в утверждении 4.6 можно отождествить с истинными значениями σ_{mh}.

(i) Проверьте, что эта модель удовлетворяет предположениям GLS параграфа 1.6 (то есть предположениям 1.1–1.3 и (1.6.1)) с объемом выборки Mn.

(ii) Проверьте, что SUR-оценка является GLS-оценкой. **Указание:** *В качестве* $\boldsymbol{X}$ *в параграфе 1.6 здесь выступает* $\boldsymbol{Z}$. *Покажите, что* $\sigma^2\boldsymbol{V}$ *здесь принимает вид* $\boldsymbol{\Sigma} \otimes \boldsymbol{I}_n$. *Используйте ваш ответ в пункте 2(b).*

(iii) Проверьте, что SUR-оценка является несмещенной в конечных выборках, то есть $\mathrm{E}(\widehat{\boldsymbol{\delta}}_{\mathrm{SUR}} - \boldsymbol{\delta}|\boldsymbol{Z}) = \mathbf{0}$.

(iv) Выражение для $\mathrm{Var}(\widehat{\boldsymbol{\delta}}_{\mathrm{SUR}}|\boldsymbol{Z})$ приведено в утверждении 1.7 параграфа 1.6. Покажите, что

$$\mathrm{Avar}(\widehat{\boldsymbol{\delta}}_{\mathrm{SUR}}) = \operatorname*{plim}_{n\to\infty} n \cdot \mathrm{Var}(\widehat{\boldsymbol{\delta}}_{\mathrm{SUR}}|\boldsymbol{Z}).$$

Указание: *С учетом того что* $\mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|\boldsymbol{Z}) = \boldsymbol{\Sigma} \otimes \boldsymbol{I}_n$, *матрица* $\mathrm{Var}(\widehat{\boldsymbol{\delta}}_{\mathrm{SUR}}|\boldsymbol{Z})$ *является обратной к матрице,* (m, h) *блок которой равен* $\sigma^{mh}\boldsymbol{Z}_m'\boldsymbol{Z}_h$. $\boldsymbol{Z}_m'\boldsymbol{Z}_h = \sum_i \boldsymbol{z}_{im}\boldsymbol{z}_{ih}'$.

(d) (Свойства RE в конечных выборках.) Наложим теперь на модель условие, состоящее в том, что вектор коэффициентов одинаков для всех уравнений, и $\boldsymbol{Z}$ определяется как в (*) аналитического упражнения 2. Как и в (c), предположим, что $\boldsymbol{\Sigma}$ известна, так что состоятельная оценка $\widehat{\sigma}_{mh}$ в формулах RE утверждения 4.7 может быть заменена истинным значением σ_{mh}.

(i) Проверьте, что такая модель с общими коэффициентами удовлетворяет предположениям GLS параграфа 1.6. В частности, предположение 1.1 может быть переписано как $\boldsymbol{y} = \boldsymbol{Z}\boldsymbol{\delta} + \boldsymbol{\varepsilon}$, где $\boldsymbol{y}$ и $\boldsymbol{\varepsilon}$ — те же, что и в аналитическом упражнении 2.

(ii) Проверьте, что оценка методом случайных эффектов является GLS-оценкой.

(iii) Проверьте, что оценка методом случайных эффектов является несмещенной в конечных выборках.

(iv) Покажите, что

$$\mathrm{Avar}(\widehat{\boldsymbol{\delta}}_{\mathrm{RE}}) = \operatorname*{plim}_{n\to\infty} n \cdot \mathrm{Var}(\widehat{\boldsymbol{\delta}}_{\mathrm{RE}}|\boldsymbol{Z}).$$

Указание: $\mathrm{Var}(\widehat{\boldsymbol{\delta}}_{RE}|\boldsymbol{Z})$ *— обратная к матрице* $\sum_m \sum_h \sigma^{mh}\boldsymbol{Z}_m'\boldsymbol{Z}_h$.

4. (Стандартные ошибки, полученные при использовании GMM отдельно для каждого уравнения и при использовании GMM для нескольких уравнений.) В параграфе 4.4 мы показали, что Avar эффективной оценки, полученной применением GMM к каждому уравнению по отдельности, не меньше (в матричном смысле), чем Avar эффективной оценки, полученной применением GMM для нескольких уравнений. В этом упражнении мы покажем, что то же самое верно и для $\widehat{\mathrm{Avar}}$. Пусть $\widehat{\boldsymbol{S}}$ и $\boldsymbol{S}_{\boldsymbol{xz}}$ те же, что и в (4.3.2) и (4.2.2) соответственно.

(a) (Тривиально.) Проверьте, что при соответствующих предположениях $(\boldsymbol{S}'_{\boldsymbol{xz}}\widehat{\boldsymbol{S}}^{-1}\boldsymbol{S}_{\boldsymbol{xz}})^{-1}$ является состоятельной оценкой асимптотической дисперсии эффективной GMM-оценки для нескольких уравнений. Что это за предположения?

(b) Пусть $\widetilde{\boldsymbol{\delta}}_m$ — эффективная GMM-оценка $\boldsymbol{\delta}_m$ для одного уравнения, а $\widetilde{\boldsymbol{\delta}}$ — вектор, сформированный как $(\widetilde{\boldsymbol{\delta}}_1, \ldots, \widetilde{\boldsymbol{\delta}}_M)$. $\widetilde{\boldsymbol{\delta}}$ — эффективная оценка для $\boldsymbol{\delta}$, получаемая GMM для каждого уравнения в отдельности. Покажите, что $\text{Avar}(\widetilde{\boldsymbol{\delta}}_m)$ состоятельно оценивается посредством

$$\left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z}_{im}\boldsymbol{x}'_{im}\widehat{\boldsymbol{W}}_{mm}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{im}\boldsymbol{z}'_{im}\right)^{-1}, \qquad (*)$$

где

$$\widehat{\boldsymbol{W}}_{mm} = \left(\frac{1}{n}\sum_{i=1}^{n}\widehat{\varepsilon}_{im}^{2}\boldsymbol{x}_{im}\boldsymbol{x}'_{im}\right)^{-1}.$$

(c) Проверьте, что $(*)$ — это блок (m, m) матрицы

$$(\boldsymbol{S}'_{\boldsymbol{xz}}\widehat{\boldsymbol{W}}\boldsymbol{S}_{\boldsymbol{xz}})^{-1}\boldsymbol{S}'_{\boldsymbol{xz}}\widehat{\boldsymbol{W}}\widehat{\boldsymbol{S}}\widehat{\boldsymbol{W}}\boldsymbol{S}_{\boldsymbol{xz}}(\boldsymbol{S}'_{\boldsymbol{xz}}\widehat{\boldsymbol{W}}\boldsymbol{S}_{\boldsymbol{xz}})^{-1}, \qquad (**)$$

где $\widehat{\boldsymbol{W}}$ — блочно-диагональная матрица, m-й диагональный блок которой — это $\widehat{\boldsymbol{W}}_{mm}$.

(d) Утверждение 3.5 — это алгебраический результат, который применим не только к теоретическим моментам, так что из него следует, что

$$(**) \geq (\boldsymbol{S}'_{\boldsymbol{xz}}\widehat{\boldsymbol{S}}^{-1}\boldsymbol{S}_{\boldsymbol{xz}})^{-1}.$$

Считая это известным, покажите следующее: если одни и те же остатки используются при вычислении $\widehat{\boldsymbol{W}}$ для GMM, применяемого к уравнениям по отдельности, и при вычислении $\widehat{\boldsymbol{S}}$ для GMM для нескольких уравнений, то тогда для каждого коэффициента стандартная ошибка, полученная при применении эффективного GMM для нескольких уравнений, не превышает стандартной ошибки, полученной при применении GMM к уравнениям по отдельности. **Указание:** *Стандартные ошибки — это квадратные корни из диагональных элементов оцененной асимптотической ковариационной матрицы, деленные на объем выборки.*

(e) Сохраняются ли результаты, полученные в (d), если модель неправильно специфицирована в том или ином отношении? **Указание:** *То, что вы доказали, является алгебраическим результатом.*

(f) Выполните шаги (a)–(d) для FIVE и для 2SLS, применяемого для каждого уравнения отдельно. **Указание:** *Для этого необходимо только переопределить* $\widehat{\boldsymbol{S}}$ *и* m*-й диагональный блок* $\widehat{\boldsymbol{W}}_{mm}$ *матрицы* $\widehat{\boldsymbol{W}}$. $\widehat{\boldsymbol{W}}_{mm}$ *для 2SLS — это матрица, обратная к*

$$\frac{1}{n}\sum_{i=1}^{n}\widehat{\varepsilon}_{im}^{2}\cdot\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{im}\boldsymbol{x}_{im}',$$

а $\widehat{\boldsymbol{S}}$ *для FIVE — это* (4.5.3).

5. (Идентификация и ограничения, связывающие разные уравнения.) Рассмотрим уравнение заработной платы для 1969 и 1980 годов:

$$LW69_i = \beta_0 + \beta_1 \cdot S69_i + \beta_2 \cdot IQ_i + \varepsilon_{i1},$$
$$LW80_i = \beta_0 + \beta_1 \cdot S80_i + \beta_2 \cdot IQ_i + \varepsilon_{i2}.$$

Условия ортогональности записываются следующим образом:

$$\mathrm{E}(\varepsilon_{i1}) = 0, \quad \mathrm{E}(\varepsilon_{i2}) = 0, \quad \mathrm{E}(MED_i\varepsilon_{i1}) = 0, \quad \mathrm{E}(MED_i\varepsilon_{i2}) = 0.$$

Пусть в нашем распоряжении имеется случайная выборка для переменных $(LW69,\ LW80, S69, S80, IQ, MED)$.

(a) Является ли уравнение для $LW69$ идентифицируемым, если рассматривать его в изоляции от других? **Указание:** *Проверьте для уравнения* $LW69$ *порядковое условие идентифицируемости.*

(b) Запишите указанные четыре условия ортогональности в терминах $(\beta_0, \beta_1, \beta_2)$ как систему уравнений, линейных по β. Сколько уравнений получилось? Запишите ранговое условие идентифицируемости в терминах соответствующих перекрестных моментов. (Этот пример показывает, что даже если каждое уравнение системы неидентифицируемо, если рассматривать его в изоляции, система может быть идентифицируемой благодаря наличию перекрестных ограничений, требующих, чтобы β были одинаковыми для всех уравнений.)

(c) Покажите, что система не может быть идентифицируемой, если IQ и MED не коррелированы между собой.

6. (Дополнительное.) Докажите утверждение 4.1. **Указание:** *Обобщите доказательство утверждения 3.2.*

7. (Численная эквивалентность 2SLS и 3SLS.) Пусть набор инструментов в системе уравнений $y_{im} = \boldsymbol{Z}_{im}'\boldsymbol{\delta}_m + \varepsilon_{im}$ $(i = 1, \ldots, n;$

$m = 1, \dots, M$) одинаков для всех уравнений (то есть $\boldsymbol{x}_{im} = \boldsymbol{x}_i$). (Здесь и далее предполагается, что все условия применимости 3SLS выполнены.)

(a) Если в правых частях всех уравнений стоят одни и те же переменные (то есть $\boldsymbol{z}_{im} = \boldsymbol{z}_i$), то оценки 2SLS и 3SLS численно совпадают. Докажите это. **Указание:** *Пусть* $\boldsymbol{B} \equiv \frac{1}{n}\sum_i \boldsymbol{x}_i \boldsymbol{z}_i'$. *3SLS-оценка — это* (4.2.3), *где* $\boldsymbol{S}_{xz} = \boldsymbol{I}_M \otimes \boldsymbol{B}$, $\widehat{\boldsymbol{W}} = \widehat{\boldsymbol{\Sigma}}^{-1} \otimes \boldsymbol{S}_{xx}^{-1}$, $\boldsymbol{S}_{xx} \equiv \frac{1}{n}\sum_i \boldsymbol{x}_i \boldsymbol{x}_i'$.

(b) Если ошибки не являются условно гомоскедастичными, то остается ли верным результат пункта (a) для GMM-оценки для каждого уравнения по отдельности и для GMM-оценки для всех уравнений вместе? **Указание:** *Нет. Почему? Можете ли вы записать* $\widehat{\boldsymbol{S}}^{-1}$ *в виде Кронекерова произведения без предположения об условной гомоскедастичности?*

8. (SUR — это больше, чем предположение о предопределенности регрессоров в каждом из уравнений.) Рассмотрим оценивание системы из M уравнений: $y_{im} = \boldsymbol{z}_{im}'\boldsymbol{\delta}_m + \varepsilon_{im}$ ($i = 1, \dots, n$; $m = 1, \dots, M$). Условия ортогональности: $\mathrm{E}(\boldsymbol{z}_{im} \cdot \varepsilon_{im}) = \boldsymbol{0}$ ($m = 1, 2, \dots, M$).

(a) Сводится ли эффективная GMM-оценка $\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1})$ для нескольких уравнений к OLS-оцениванию каждого уравнения отдельно? Изменится ли ваш вывод, если ошибки условно гомоскедастичны? **Указание:** *Каков размер* $\boldsymbol{S}_{xz}$? $\widehat{\boldsymbol{S}}$?

(b) Объясните, почему эффективная GMM-оценка, которую вы вывели в пункте (a) при предположении об условной гомоскедастичности отличается от SUR? **Указание:** *Разница в спецификации условий ортогональности.*

9. (Дополнительное, опасность совместного оценивания.) Взято из [Greene, 1997, p. 706]. Рассмотрим SUR-модель из двух уравнений:

$$
\begin{aligned}
y_{i1} &= \beta_1 x_{i1} + \varepsilon_{i1}, \\
y_{i2} &= \beta_2 x_{i2} + \beta_3 x_{i3} + \varepsilon_{i2}.
\end{aligned}
$$

Для простоты предположим, что $\mathrm{E}(\varepsilon_{i1}^2)$, $\mathrm{E}(\varepsilon_{i1}\varepsilon_{i2})$ и $\mathrm{E}(\varepsilon_{i2}^2)$ известны. Предположим теперь, что вы применяете SUR, но ошибочно не включаете x_{i3} во второе уравнение. Какое влияние это окажет на состоятельность оценки β_1? **Указание:** *Пусть* $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{\varepsilon}_1$

и ε_2 — n-мерные векторы соответствующих переменных и

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{bmatrix}, \quad \overline{\boldsymbol{X}} = \begin{bmatrix} \boldsymbol{x}_1 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{x}_2 \end{bmatrix},$$

$$\boldsymbol{d} = \begin{bmatrix} \mathbf{0} \\ \boldsymbol{x}_3 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, \quad \boldsymbol{\Sigma} = \begin{bmatrix} \mathrm{E}(\varepsilon_{i1}^2) & \mathrm{E}(\varepsilon_{i1}\varepsilon_{i2}) \\ \mathrm{E}(\varepsilon_{i2}\varepsilon_{i1}) & \mathrm{E}(\varepsilon_{i2}^2) \end{bmatrix}.$$

(Так что, например, $\overline{\boldsymbol{X}}$ — вектор размера $2n \times 2$, и β — вектор размера 2×1.) Покажите, что

(a) *Указанная система двух уравнений может быть записана как $\boldsymbol{y} = \overline{\boldsymbol{X}}\boldsymbol{\beta} + \beta_3 \cdot \boldsymbol{d} + \boldsymbol{\varepsilon}$.*

(b) *SUR-оценка $\boldsymbol{\beta}$ в случае, когда переменная x_{i3} ошибочно пропущена, равна:*

$$\widehat{\boldsymbol{\beta}}_{\mathrm{SUR}} = \left[\overline{\boldsymbol{X}}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\overline{\boldsymbol{X}}\right]^{-1} \overline{\boldsymbol{X}}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{y}.$$

(c) *Используя формулы (A.6) и (A.7) приложения A, покажите, что*

$$\widehat{\boldsymbol{\beta}}_{\mathrm{SUR}} = \boldsymbol{\beta} + \beta_3 \cdot \boldsymbol{a} + \boldsymbol{b},$$

где

$$\boldsymbol{a} = \begin{bmatrix} \sigma^{11}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{i1}^2 & \sigma^{12}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{i1}x_{i2} \\ \sigma^{12}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{i2}x_{i1} & \sigma^{22}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{i2}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sigma^{12}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{i1}x_{i3} \\ \sigma^{22}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{i2}x_{i3} \end{bmatrix},$$

$$\boldsymbol{b} = \begin{bmatrix} \sigma^{11}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{i1}^2 & \sigma^{12}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{i1}x_{i2} \\ \sigma^{12}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{i2}x_{i1} & \sigma^{22}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{i2}^2 \end{bmatrix}^{-1} \times$$

$$\times \begin{bmatrix} \sigma^{11}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{i1}\varepsilon_{i1} + \sigma^{12}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{i1}\varepsilon_{i2} \\ \sigma^{21}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{i2}\varepsilon_{i1} + \sigma^{22}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_{i2}\varepsilon_{i2} \end{bmatrix}.$$

(d) *Докажите, что $\operatorname{plim} \boldsymbol{b} = \mathbf{0}$. Будет ли $\operatorname{plim} \boldsymbol{a} = \mathbf{0}$?*

10. (По желанию, добавление точно идентифицируемых уравнений не увеличивает эффективность.) Рассмотрите модель 3SLS с двумя уравнениями, в которой второе уравнение идентифицируемо точно (то есть $L_2 = K$). Пусть $\widehat{\boldsymbol{\delta}}_{1,\mathrm{3SLS}}$ — это 3SLS-оценка коэффициентов первого уравнения, а $\widehat{\boldsymbol{\delta}}_{1,\mathrm{2SLS}}$ — это 2SLS-оценка этих же коэффициентов.

(a) Запишите $\mathrm{Avar}(\widehat{\boldsymbol{\delta}}_{1,\mathrm{2SLS}})$ в терминах σ_{11} и $\boldsymbol{A}_{11}$, определенной в (4.5.16).

(b) Покажите, что $\mathrm{Avar}(\widehat{\boldsymbol{\delta}}_{1,\mathrm{3SLS}}) = \mathrm{Avar}(\widehat{\boldsymbol{\delta}}_{1,\mathrm{2SLS}})$. **Указание:** *Используя формулу Фробениуса для нахождения обратной матрицы для блочных матриц (см. формулу (A.10) приложения A), запишите левый верхний (размера $L_1 \times L_1$) блок $\mathrm{Avar}(\widehat{\boldsymbol{\delta}}_{1,\mathrm{3SLS}})$ как $\boldsymbol{G}^{-1}$, где*

$$\boldsymbol{G} = \sigma^{11}\boldsymbol{A}_{11} - \left(\frac{\sigma^{12}\cdot\sigma^{21}}{\sigma^{22}}\right)\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}.$$

Но в соответствии с (4.5.16)

$$\begin{aligned}
&\boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21} = \\
&= \mathrm{E}(\boldsymbol{z}_{i1}\boldsymbol{x}_i')[\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')]^{-1}\,\mathrm{E}(\boldsymbol{x}_i\boldsymbol{z}_{i2}')[\mathrm{E}(\boldsymbol{z}_{i2}\boldsymbol{x}_i')[\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')]^{-1}\,\mathrm{E}(\boldsymbol{x}_i\boldsymbol{z}_{i2}')]^{-1}\times \\
&\qquad \times \mathrm{E}(\boldsymbol{z}_{i2}\boldsymbol{x}_i')[\mathrm{E}(\boldsymbol{z}_{i2}\boldsymbol{x}_i')]^{-1}\,\mathrm{E}(\boldsymbol{x}_i\boldsymbol{z}_{i1}') = \\
&= \mathrm{E}(\boldsymbol{z}_{i1}\boldsymbol{x}_i')[\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')]^{-1}\,\mathrm{E}(\boldsymbol{x}_i\boldsymbol{z}_{i1}') \\
&(\textit{так как } \#\boldsymbol{z}_{i2}(=L_2) = \#\boldsymbol{x}_i(=K)) \\
&= \boldsymbol{A}_{11}.
\end{aligned}$$

Таким образом,

$$\boldsymbol{G} = \left[\sigma^{11} - \left(\frac{\sigma^{12}\cdot\sigma^{21}}{\sigma^{22}}\right)\right]\boldsymbol{A}_{11} = \frac{\boldsymbol{A}_{11}}{\sigma_{11}},$$

$$(\textit{потому что } \sigma^{11} - \left(\frac{\sigma^{12}\cdot\sigma^{21}}{\sigma^{22}}\right) = 1/\sigma_{11}).$$

. (Линейные комбинации условий ортогональности.) Для модели с несколькими уравнениями из утверждения 4.7 выведите эффективную GMM-оценку, которая использует следующие условия ортогональности:

$$\mathrm{E}(\boldsymbol{z}_{i1}\cdot\varepsilon_{i1} + \cdots + \boldsymbol{z}_{iM}\cdot\varepsilon_{iM}) = 0. \tag{1}$$

Указание: *Выборочный аналог левой части этих условий ортогональности выглядит следующим образом:*

$$\begin{aligned}
\boldsymbol{g}_n(\widetilde{\boldsymbol{\delta}}) \equiv \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_{i1}\cdot y_{i1} + \cdots + \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_{iM}\cdot y_{iM} - \\
- \left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_{i1}\cdot\boldsymbol{z}_{i1}' + \cdots + \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_{iM}\cdot\boldsymbol{z}_{iM}'\right)\widetilde{\boldsymbol{\delta}}.
\end{aligned}$$

Количество условий ортогональности равно количеству коэффициентов. Эффективная GMM-оценка является IV-оценкой, которая является решением $\boldsymbol{g}_n(\widetilde{\boldsymbol{\delta}}) = \mathbf{0}$.

Эмпирические упражнения

В файле GREENE.ASC содержатся данные по следующим переменным:

Столбец 1: идентификатор фирмы в выборке Нерлова 1955.

Столбец 2: издержки в миллионах долларов в ценах 1970 года.

Столбец 3: выпуск в миллионах киловатт-часов.

Столбец 4: цена трудовых ресурсов.

Столбец 5: цена капитала.

Столбец 6: цена топлива.

Столбец 7: доля издержек на труд.

Столбец 8: доля издержек на капитал.

В выборке 99 наблюдений. Доля издержек на топливо может быть вычислена как единица минус сумма долей издержек на труд и на капитал. Эти данные были использованы в работе [Christensen and Greene, 1976].

В тексте третье уравнение системы (для доли затрат на топливо) не использовалось при оценивании модели со случайными эффектами. Чтобы проверить, что результаты будут численно одинаковыми, не используйте второе уравнение (для доли затрат на капитал) и устраните $(\gamma_{12}, \gamma_{22}, \gamma_{32})$ при введении ограничения однородности. Результатом этого будет система из двух уравнений:

$$\begin{cases} s_1 = \alpha_1 + \gamma_{11}\log(p_1/p_2) + \gamma_{13}\log(p_3/p_2) + \gamma_{1Q}\log(Q) + \varepsilon_1 \\ \text{(доля издержек на труд).} \\ s_3 = \alpha_3 + \gamma_{31}\log(p_1/p_2) + \gamma_{33}\log(p_3/p_2) + \gamma_{3Q}\log(Q) + \varepsilon_3 \\ \text{(доля издержек на топливо).} \end{cases}$$

Назовем это *системой без ограничений*. *Система с ограничениями* накладывает условие симметричности $\gamma_{13} = \gamma_{31}$ на систему уравнений без ограничений.

(a) Проверьте, что вы можете воспроизвести значения простых выборочных статистик, приведенные в тексте.

(b) (Совпадение численных результатов.) Оцените модель со случайными эффектами (то есть многомерную регрессию при ограничении, связывающем разные уравнения системы) для системы с ограничениями. Сможете ли вы воспроизвести оценки коэффициентов, приведенные в тексте? Вы должны сначала применить OLS к каждому уравнению системы без ограничений по отдельности, чтобы вычислить $\widehat{\Sigma}^*$. (Замечание относительно вычислений: запрограммировать совместное распределение гораздо проще, используя программные модули, такие как TSP и RATS. Однако эти программные модули не выдают статистику Саргана для SUR.)

Совет для пакета GAUSS: Оценка модели со случайными эффектами может быть получена с помощью операций над матрицами данных, но мы советуем использовать циклы «do». Используйте (4.6.8′) на с. 321.

Совет для пакета RATS: Используйте процедуру NLSYS, которая разработана для нелинейных уравнений, но которая может быть также использована для линейных уравнений. Она может также быть использована для вычисления $\widehat{\Sigma}^*$, как это сделано в следующих кодах:

```
Оценивание модели с исключением уравнения для доли
затрат на капитал
* Определите параметры
nonlin al af gll glf gfl gff gly gfy
* Определите систему из 2 уравнений
frml labor sl = al+gll*log(pl/pk)+glf*log(pf/pk)
+gly*log(kwh) frml fuel sf =
af+gfl*log(pl/pk)+gff*log(pf/pk) +gfy*log(kwh)
* Задайте начальные значения для параметров
compute al=af=gll=glf=gfl=gff=gly=gfy=0
* Теперь оцените многомерную регрессию
* без ограничений, связывающих разные уравнения
системы.
nlsystem(outsigma=sigmahat) / labor fuel
```

Здесь `sigmahat` — матрица $\widehat{\Sigma}^*$ размера 2×2. Оценивание этой матрицы производится с помощью следующих кодов:

```
* Теперь введите ограничения, связывающие разные
уравнения системы,
* на уравнение для доли затрат на топливо (set gfl=glf)
* Определите параметры.
nonlin al af gll glf gff gly gfy
* Введите условие симметрии.
frml fuelb sf = af+glf*log(pl/pk)+gff*log(pf/pk)
+gfy*log(kwh)
* Многомерная регрессия с ковариационной матрицей,
* вычисляемой по остаткам, полученным
* при оценивании модели OLS для каждого уравнения
в отдельности.
nlsystem(isigma=sigmahat) / labor fuelb
```

Совет для пакета TSP. Пример кодов TSP, которые выполняют то же самое, что и коды RATS выше:

```
? Оцените систему без уравнения
? для доли издержек на капитал.
frml labor sl=al+gll*log(pl/pk)+glf*log(pf/pk)
+gly*log(kwh);
frml fuel sf=af+gfl*log(pl/pk)+gff*log(pf/pk)
+gfy*log(kwh);
? Положите значения параметров равными нулю.
param al af gll glf gfl gff gly gfy;
? Многомерная регрессия без симметрии.
sur labor fuel;
? Оцените ковариационную матрицу ошибок по остаткам.
mform sigmahat=@covu;
?
? Оцените многомерную регрессию
? с ограничением, связывающим разные уравнения систе-
мы.
? Введите условие симметрии.
frml fuel sf=af+glf*log(pl/pk)+gff*log(pf/pk)
+gfy*log(kwh);
? Многомерная регрессия при условии симметрии.
lsq(maxitw=0,wname=sigmahat) labor fuel;
```

Это дает RE-оценку семи свободных параметров

$$\alpha_1, \alpha_3, \gamma_{11}, \gamma_{13}, \gamma_{33}, \gamma_{1Q}, \gamma_{3Q}.$$

Используйте, как объяснено в тексте, ограничения в виде сумм, однородности и симметрии, чтобы вычислить точечные оценки и их стандартные ошибки для оставшихся восьми параметров. Этому должна помочь программа ANALYZ в TSP.

(c) (По желанию для пользователей программы GAUSS.) Вычислите статистику Саргана. **Указание:** *Используйте формулу из утверждения 4.7.* Вектор $\boldsymbol{x}_i$ должен быть таким: константа, $\ln(p_1/p_2)$, $\ln(p_3/p_2)$, $\ln(Q)$. Для $\widehat{\boldsymbol{\Sigma}}^*$ используйте оценку, полученную из системы без ограничений. $\boldsymbol{s_{xy}} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{y}_i \otimes \boldsymbol{x}_i$ и $\boldsymbol{S_{xz}} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{Z}_i \otimes \boldsymbol{x}_i$. Значение статистики должно быть равным 0,63313 с p-значением 0,42621.

(d) (Проверка симметрии.) Учитывая, что статистика Саргана равна 0,63313, произведите тестирование следующей гипотезы. Пусть поддерживаемая гипотеза — система без ограничений. Нулевая гипотеза состоит в выполнении ограничения симметрии $\gamma_{13} = \gamma_{31}$. **Указание:**

Согласно утверждению 3.8 разность между статистиками Саргана при наличии и при отсутствии симметрии имеет асимптотическое распределение хи-квадрат. Чему равна статистика Саргана для системы без ограничения?

(e) (Тест Вальда на симметрию.) Проверьте ту же нулевую гипотезу с помощью статистики Вальда. Таким образом вы проверяете ограничение симметрии в системе без ограничений. Проверьте, что эта тестовая статистика численно равняется статистике Саргана для системы с ограничениями, которую вы оценивали в пункте (b).

Совет для пакета TSP: Просто вставьте строчку

```
analyz wald wald=glf-gfl;
```

сразу после кодов для оценивания без ограничений.

(f) Используя оценки параметров, полученные в пункте (b), рассчитайте среднюю по 99 фирмам эластичность замещения труда капиталом. У вас получился тот же результат, что в тексте?

Ответы на избранные вопросы

Аналитические упражнения

3b.

$$\begin{aligned}
&\mathrm{E}(\varepsilon_{im}|\boldsymbol{Z}) = \\
&= \mathrm{E}(\varepsilon_{im}|\boldsymbol{z}_{11},\dots,\boldsymbol{z}_{1M},\boldsymbol{z}_{21},\dots,\boldsymbol{z}_{2M},\dots,\boldsymbol{z}_{n1},\dots,\boldsymbol{z}_{nM}) = \\
&\text{(по определению } \boldsymbol{Z}) \\
&= \mathrm{E}(\varepsilon_{im}|\boldsymbol{z}_{11},\dots,\boldsymbol{z}_{1M}) \text{ (по предположению i.i.d.)} = \\
&= 0 \text{ (по усиленным условиям ортогональности)}.
\end{aligned}$$

Элемент (i,j) квадратной матрицы $\mathrm{E}(\boldsymbol{\varepsilon}_m\boldsymbol{\varepsilon}_h'|\boldsymbol{Z})$ размера $n \times n$ — это $\mathrm{E}(\varepsilon_{im}\varepsilon_{jh}|\boldsymbol{Z})$. По предположению i.i.d., он равен нулю для $j \neq i$. Для $j = i$

$$\begin{aligned}
&\mathrm{E}(\varepsilon_{im}\varepsilon_{ih}|\boldsymbol{Z}) = \\
&= \mathrm{E}(\varepsilon_{im}\varepsilon_{ih}|\boldsymbol{z}_{11},\dots,\boldsymbol{z}_{1M},\boldsymbol{z}_{21},\dots,\boldsymbol{z}_{2M},\dots,\boldsymbol{z}_{n1},\dots,\boldsymbol{z}_{nM}) = \\
&= \mathrm{E}(\varepsilon_{im}\varepsilon_{ih}|\boldsymbol{z}_{i1},\dots,\boldsymbol{z}_{iM}) \text{ (по предположению i.i.d.)}.
\end{aligned}$$

Поскольку $\boldsymbol{x}_{im} = \boldsymbol{x}_i$, а $\boldsymbol{x}_i$ — это набор $(\boldsymbol{z}_{i1},\dots,\boldsymbol{z}_{iM})$ в модели SUR, предположение условной гомоскедастичности (предположение 4.7) гласит, что $\mathrm{E}(\varepsilon_{im}\varepsilon_{ih}|\boldsymbol{z}_{i1},\dots,\boldsymbol{z}_{iM}) = \sigma_{mh}$.

5b.

$$\begin{bmatrix} 1 & \mathrm{E}(S69) & \mathrm{E}(IQ) \\ 1 & \mathrm{E}(S80) & \mathrm{E}(IQ) \\ \mathrm{E}(MED) & \mathrm{E}(S69 \cdot MED) & \mathrm{E}(IQ \cdot MED) \\ \mathrm{E}(MED) & \mathrm{E}(S80 \cdot MED) & \mathrm{E}(IQ \cdot MED) \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} =$$

$$= \begin{bmatrix} \mathrm{E}(LW69) \\ \mathrm{E}(LW80) \\ \mathrm{E}(LW69 \cdot MED) \\ \mathrm{E}(LW80 \cdot MED) \end{bmatrix}.$$

Условие идентифицируемости системы состоит в том, что матрица коэффициентов размера 4×3 имеет полный столбцовый ранг.

5c. Если IQ и MED не коррелированы между собой, то тогда $\mathrm{E}(IQ \cdot MED) = \mathrm{E}(IQ) \cdot \mathrm{E}(MED)$ и третий столбец матрицы коэффициентов равен $\mathrm{E}(IQ)$, умноженному на первый столбец. В этом случае матрица не может иметь полный столбцовый ранг.

6. $\widehat{\varepsilon}_{im} = y_{im} - \boldsymbol{z}'_{im}\widehat{\boldsymbol{\delta}}_m = \varepsilon_{im} - \boldsymbol{z}'_{im}(\widehat{\boldsymbol{\delta}}_m - \boldsymbol{\delta}_m)$. Поэтому

$$\frac{1}{n}\sum_{i=1}^{n}[\varepsilon_{im} - \boldsymbol{z}'_{im}(\widehat{\boldsymbol{\delta}}_m - \boldsymbol{\delta}_m)][\varepsilon_{ih} - \boldsymbol{z}'_{ih}(\widehat{\boldsymbol{\delta}}_h - \boldsymbol{\delta}_h)] = (1) + (2) + (3) + (4),$$

где

$$(1) = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_{im}\varepsilon_{ih},$$

$$(2) = -(\widehat{\boldsymbol{\delta}}_m - \boldsymbol{\delta}_m)'\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_{im} \cdot \varepsilon_{ih},$$

$$(3) = -(\widehat{\boldsymbol{\delta}}_h - \boldsymbol{\delta}_h)'\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_{ih} \cdot \varepsilon_{im},$$

$$(4) = (\widehat{\boldsymbol{\delta}}_m - \boldsymbol{\delta}_m)'\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_{im}\boldsymbol{z}'_{ih}\right)(\widehat{\boldsymbol{\delta}}_h - \boldsymbol{\delta}_h).$$

Как обычно, по предположению 4.1 и 4.2, $(1) \to_{\mathrm{p}} \sigma_{mh} (\equiv \mathrm{E}(\varepsilon_{im}\varepsilon_{ih}))$. Для (4), согласно предположению 4.2 и предположению о том, что $\mathrm{E}(\boldsymbol{z}_{im}\boldsymbol{z}'_{ih})$ конечно, $\frac{1}{n}\sum_i \boldsymbol{z}_{im}\boldsymbol{z}'_{ih}$ сходится по вероятности к (конечной) матрице. Таким образом, $(4) \to_{\mathrm{p}} 0$.

Что касается (2), то по неравенству Коши — Шварца

$$\mathrm{E}(|z_{imj} \cdot \varepsilon_{ih}|) \le \sqrt{\mathrm{E}(z^2_{imj}) \cdot \mathrm{E}(\varepsilon^2_{ih})},$$

где z_{imj} — j-й элемент $\mathbf{z}_{im}$. Тогда $\mathrm{E}(\mathbf{z}_{im} \cdot \varepsilon_{ih})$ конечна и (2) $\to_{\mathrm{p}} 0$, потому что $\widehat{\boldsymbol{\delta}}_m - \boldsymbol{\delta}_m \to_{\mathrm{p}} \mathbf{0}$. Аналогичным образом (3) $\to_{\mathrm{p}} 0$.

Литература

Berndt, E., 1991, *The Practice of Econometrics*, New York: Addison-Wesley.

Berndt, E., and D. Wood, 1975, "Technology, Prices, and the Derived Demand for Energy," *Review of Economics and Statistics*, 57, 259–268.

Brundy, J., and D. Jorgenson, 1971, "Efficient Estimation of Simultaneous Systems by Instrumental Variables," *Review of Economics and Statistics*, 53, 207–224.

Christensen, L., and W. Greene, 1976, "Economies of Scale in U.S. Electric Power Generation," *Journal of Political Economy*, 84, 655–676.

Greene, W., 1997, *Econometric Analysis* (3d ed.), New Jersey: Prentice Hall.

Jorgenson, D., 1986, "Econometric Methods for Modeling Producer Behavior," in Z. Griliches and M. Intriligator (eds.), *Handbook of Econometrics*, Volume III, Amsterdam: Elsevier Science Publishers, 1841–1915.

Uzawa, H., 1962, "Production Function with Constant Elasticities of Substitution," *Review of Economic Studies*, 29, 291–299.

Varian, H., 1992, *Microeconomic Analysis* (3d ed.), New York: Norton.

Zellner, A., 1962, "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American Statistical Association*, 63, 502–511.

Zellner, A., and H. Theil, 1962, "Three-Stage Least Squares: Simultaneous Estimation of Simultaneous Equations," *Econometrica*, 30, 54–78.

Глава 5. Панельные данные

Лонгитюдные (*longitudinal*), или **панельные** (*panel*), данные представляют повторные наблюдения для некоторого набора кросс-секционных единиц. Как предполагает название, панель имеет два измерения, одно — для кросс-секционных единиц, второе — для наблюдений. Второе измерение обычно является временем, но бывают и исключения. Например, кросс-секционная выборка близнецов является панелью, в которой вторую размерность образуют братья и сестры. В этой главе должно приниматься во внимание различие между кросс-секционными единицами и наблюдениями — терминами, которые до сих пор использовались как синонимы. Когда это будет уместно, мы будем использовать для кросс-секционной единицы термин «группа».

Во многих приложениях мы сталкиваемся с **моделью компонент ошибки** (*error-components model*). В таких моделях ошибка в уравнении имеет компоненту, которая является общей для всех наблюдений (инвариантной к времени, если вторая размерность в панели является временем), но которая может не быть ортогональной регрессорам. Существует доступный метод, называемый оценкой фиксированных эффектов, который позволяет нам оценивать такую модель без помощи инструментальных переменных. Эта оценка и оценка случайных эффектов из предыдущей главы являются двумя основными методами в эконометрике панельных данных.

Необязательный параграф этой главы, параграф 5.3, показывает, как модифицировать оценку фиксированных эффектов, когда число наблюдений изменяется от группы к группе. В параграфе, посвященном приложениям, рассматривается эмпирика роста, тема, которая в последнее время привлекает большое внимание.

Как выясняется, оценка фиксированных эффектов и другие оценки, рассматриваемые в этой главе, являются частными случаями GMM-оценок при соответствующем преобразовании модели компонент ошибки. GMM-интерпретация оценки фиксированных эффектов продолжается в аналитическом упражнении к этой главе, где вам будет предложено найти оценку, которая будет асимптотически более эффективной, чем оценка фиксированных эффектов.

5.1. Модель компонент ошибки

Нашей отправной точкой является модель с несколькими уравнениями с одинаковыми коэффициентами из утверждения 4.7. Чтобы упростить обсуждение, мы предположим, что выборка является случайной. Это предположение выполняется для многих популярных панелей, таких как Panel Study of Income Dynamics (PSID) и National Longitudinal Survey (NLS). В этих панелях большое количество — сотни или даже тысячи — кросс-секционных единиц извлечены из генеральной совокупности случайным образом.

Для этой ситуации случайных выборок предположения утверждения 4.7 можно переформулировать (см. (4.6.15) в отношении определения $\boldsymbol{y}_i$ $(M \times 1)$, $\boldsymbol{Z}_i$ $(M \times L)$, ε_i $(M \times 1)$):

$$\text{(система из } M \text{ линейных уравнений)} \quad \boldsymbol{y}_i = \boldsymbol{Z}_i\boldsymbol{\delta} + \boldsymbol{\varepsilon}_i, \tag{5.1.1}$$

$$\text{(случайная выборка)} \quad \{\boldsymbol{y}_i, \boldsymbol{Z}_i\} \text{ являются i.i.d.}, \tag{5.1.2}$$

$$\text{(«предположение SUR»)} \quad \mathrm{E}(\boldsymbol{z}_{im} \cdot \varepsilon_{ih}) = \boldsymbol{0} \quad \text{для } m, h = 1, 2, \ldots, M. \tag{5.1.3}$$

то есть $\mathrm{E}(\boldsymbol{\varepsilon}_i \otimes \boldsymbol{x}_i) = \boldsymbol{0}$, где $\boldsymbol{x}_i = (\boldsymbol{z}_{i1}, \ldots, \boldsymbol{z}_{iM})$,

$$\text{(идентификация)} \quad \mathrm{E}(\boldsymbol{Z}_i \otimes \boldsymbol{x}_i) \text{ имеет полный столбцовый ранг}^1, \tag{5.1.4}$$

$$\text{(условная гомоскедастичность)} \quad \mathrm{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'|\boldsymbol{x}_i) = \mathrm{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i') \equiv \boldsymbol{\Sigma}, \tag{5.1.5}$$

$$\text{(невырожденность } \mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_i')) \quad \mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_i') \text{ невырождена, где } \boldsymbol{g}_i \equiv \boldsymbol{\varepsilon}_i \otimes \boldsymbol{x}_i. \tag{5.1.6}$$

Как отмечалось в параграфе 4.5, поскольку при условной гомоскедастичности $\mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_i') = \boldsymbol{\Sigma} \otimes \mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')$, (5.1.6) эквивалентно условию:

$$\boldsymbol{\Sigma} \text{ и } \mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i') \text{ невырождены.} \tag{5.1.6$'$}$$

При этих предположениях оценка случайных эффектов $\hat{\boldsymbol{\delta}}_{RE}$, определенная в (4.6.8′) на с. 321, является эффективной GMM-оценкой.

[1] Матрицу $\boldsymbol{\Sigma}_{xz}$ в предположении 4.4′ можно записать как $\mathrm{E}(\boldsymbol{Z}_i \otimes \boldsymbol{x}_i)$; см. контрольный вопрос 4 в параграфе 4.6.

Компоненты ошибки

Данная модель, очень часто использующаяся для панельных данных, одна из тех, в которых предполагается, что ненаблюдаемые ошибки ε_{im} состоят из двух компонент:

$$\varepsilon_{im} = \alpha_i + \eta_{im}. \tag{5.1.7}$$

Первая ненаблюдаемая компонента α_i, без индекса m и, следовательно, являющаяся общей для всех M уравнений в i-м наблюдении, называется **индивидуальным эффектом** (*individual effect*), **индивидуальной неоднородностью** (*individual heterogeneity*) или **фиксированным эффектом** (*fixed effect*). Если мы определим

$$\underset{(M\times 1)}{\boldsymbol{\eta}_i} \equiv \begin{bmatrix} \eta_{i1} \\ \vdots \\ \eta_{iM} \end{bmatrix},$$

матричное представление системы из M уравнений будет иметь вид:

$$\begin{bmatrix} y_{i1} \\ \vdots \\ y_{iM} \end{bmatrix} = \begin{bmatrix} \mathbf{z}'_{i1}\boldsymbol{\delta} \\ \vdots \\ \mathbf{z}'_{iM}\boldsymbol{\delta} \end{bmatrix} + \begin{bmatrix} \alpha_i \\ \vdots \\ \alpha_i \end{bmatrix} + \begin{bmatrix} \eta_{i1} \\ \vdots \\ \eta_{iM} \end{bmatrix}$$

или

$$\mathbf{y}_i = \mathbf{Z}_i\boldsymbol{\delta} + \mathbf{1}_M \cdot \alpha_i + \boldsymbol{\eta}_i \quad (i = 1, 2, \ldots, n), \tag{5.1.1'}$$

где $\mathbf{1}_M$ является M-мерным вектором из единиц.

Условие ортогональности (5.1.3) выполняется, если регрессоры системы ортогональны обеим компонентам ошибки, то есть если

$$\mathrm{E}(\mathbf{z}_{im} \cdot \alpha_i) = \mathbf{0} \quad \text{для } m = 1, 2, \ldots, M \tag{5.1.8a}$$

и

$$\mathrm{E}(\mathbf{z}_{im} \cdot \eta_{ih}) = \mathbf{0} \quad \text{для } m, h = 1, 2, \ldots, M \tag{5.1.8b}$$

Однако во многих приложениях, которые используют панельные данн предположение (5.1.8a) не является разумным. Это потому, что фи рованный эффект представляет некоторую постоянную характерис индивидуальной экономической единицы, что иллюстрируется сл щими примерами:

Пример 5.1 (производственная функция с неоднородностью

В логлинейной производственной функции (3.2.13) из параграфа 3 положим, что эффективность фирм u_i остается постоянной во Тогда уравнение для года m имеет вид:

$$\log(Q_{im}) = \phi_0 + \phi_1 \log(L_{im}) + u_i + v_{im}.$$

где Q_{im} является выпуском фирмы i в год m, L_{im} является количеством затраченного труда в год m и v_{im} является технологическим шоком в год m. Это уравнение можно записать как (5.1.1′), полагая

$$y_{im} = \log(Q_{im}),\ \ z_{im} = (1, \log(L_{im}))',\ \ \boldsymbol{\delta} = (\phi_0, \phi_1)',\ \ \alpha_i = u_i,\ \ \eta_{im} = v_{im}.$$

Также

$$\boldsymbol{x}_i = (1, \log(L_{i1}), \ldots, \log(L_{iM}))'.$$

При совершенной конкуренции индивидуальный эффект u_i будет положительно коррелирован с количеством затраченного труда, поскольку эффективные фирмы, u_i которых выше, чем их среднее значение, будут нанимать большее количество рабочей силы для расширения (см. формулу (3.2.14)). Если v_{im} представляет неожиданные шоки, которые являются непредвиденными для фирмы в момент выбора входных показателей, разумно предположить, что v_{im} не коррелирована с регрессорами.

Пример 5.2 (уравнение заработной платы): Для простоты удалим стаж из уравнения заработной платы в примере 4.2, но предположим, что, в дополнение к данным за 1969 и 1980 годы, доступны также данные за 1982 год. Предположим также, что коэффициенты при образовании и при IQ остаются постоянными во времени, но константа от времени зависит (например, вследствие влияния на заработную плату делового цикла). С учетом декомпозиции ошибки (5.1.7), система из трех уравнений принимает вид:

$$\begin{cases} LW69_i = \phi_1 + \beta S69_i + \gamma IQ_i + \alpha_i + \eta_{i1}, \\ LW80_i = \phi_2 + \beta S80_i + \gamma IQ_i + \alpha_i + \eta_{i2}, \\ LW82_i = \phi_3 + \beta S82_i + \gamma IQ_i + \alpha_i + \eta_{i3}. \end{cases}$$

Как было отмечено в конце параграфа 4.6, даже если некоторое подмножество коэффициентов изменяется от уравнения к уравнению, эту систему можно записать как несколько уравнений с общими коэффициентами. В данном примере это достигается с помощью

$$\boldsymbol{Z}_i = \begin{bmatrix} \boldsymbol{z}'_{i1} \\ \boldsymbol{z}'_{i2} \\ \boldsymbol{z}'_{i3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & S69_i & IQ_i \\ 0 & 1 & 0 & S80_i & IQ_i \\ 0 & 0 & 1 & S82_i & IQ_i \end{bmatrix},\quad \boldsymbol{\delta}' = (\phi_1, \phi_2, \phi_3, \beta, \gamma). \qquad (5.1.9)$$

Также

$$\boldsymbol{x}_i = (1, S69_i, S80_i, S82_i, IQ_i)'.$$

Ошибка включает в себя детерминанты заработной платы, не учтенные переменными образования и IQ. Разумно предположить, что они разделяются между перманентными чертами индивида (эта часть обозначается как α_i), которые влияли бы на выбор образования индивидом, и другими факторами (эта часть обозначается как η_{im}), не коррелированными с регрессорами, такими как ошибки измерения в логарифме уровня заработной платы.

Групповые средние

К счастью, существует популярная оценка, называемая **оценкой фиксированных эффектов** (*fixed-effects estimator*), которая робастна к нарушению предположения (5.1.8a), то есть которая является состоятельной, даже когда регрессоры не являются ортогональными фиксированному эффекту α_i. Предваряя обсуждение в следующем параграфе, заметим, что эта оценка применяется к системе из M уравнений, преобразованной из первоначальной системы (5.1.1′). Матрица, используемая для преобразования, является аннулятором, ассоциированным с $\mathbf{1}_M$:

$$\underset{(M\times M)}{\boldsymbol{Q}} \equiv \boldsymbol{I}_M - \mathbf{1}_M(\mathbf{1}'_M\mathbf{1}_M)^{-1}\mathbf{1}'_M = = \boldsymbol{I}_M - \frac{1}{M}\mathbf{1}_M\mathbf{1}'_M = = \boldsymbol{I}_M - \begin{bmatrix} 1/M & \cdots & 1/M \\ \vdots & & \vdots \\ 1/M & \cdots & 1/M \end{bmatrix}. \tag{5.1.10}$$

Что эта матрица делает — извлекает **отклонения от групповых средних** (*deviations from group means*). Например, умножение M-мерного вектора $\boldsymbol{y}_i$ слева на $\boldsymbol{Q}$ приводит к

$$\tilde{\boldsymbol{y}}_i \equiv \boldsymbol{Q}\boldsymbol{y}_i = \begin{bmatrix} y_{i1} - \overline{y}_i \\ \vdots \\ y_{iM} - \overline{y}_i \end{bmatrix} = \boldsymbol{y}_i - \mathbf{1}_M \cdot \overline{y}_i, \tag{5.1.11}$$

где

$$\overline{y}_i = \frac{1}{M}\mathbf{1}'_M\boldsymbol{y}_i = \frac{1}{M}\sum_{m=1}^{M} y_{im}$$

является **групповым средним** (*group mean*) для зависимой переменной. Заметим, что это среднее по m, *не* по i; оно специфично для каждой группы (i). Аналогично и с регрессорами: каждый столбец $\boldsymbol{Q}\boldsymbol{Z}_i$ является вектором отклонений для соответствующего столбца в $\boldsymbol{Z}_i$.

Репараметризация

Неудивительно, однако, что робастность к нарушению (5.1.8a) имеет свою цену: некоторые из параметров модели могут стать неидентифицируемыми после такого преобразования посредством $\boldsymbol{Q}$. Приведем два примера.

- Очевидный случай — коэффициент при переменной одинаков для всех уравнений. Например, IQ_i в примере 5.2 является общим регрессором (см. (5.1.9)), так что столбец в $\boldsymbol{Z}_i$, соответствующий IQ_i (пятый столбец), равен $\mathbf{1}_M \cdot IQ_i$, и пятый столбец в $\boldsymbol{QZ}_i$ является вектором из нулей. Следовательно, коэффициент при IQ, γ, не может быть идентифицирован после такого преобразования. Чтобы отделить коэффициент при общих регрессорах от остальных, пусть $\boldsymbol{b}_i$ является вектором общих регрессоров, и запишем тогда $M \times L$ матрицу регрессоров $\boldsymbol{Z}_i$ как

$$\boldsymbol{Z}_i = (\boldsymbol{F}_i \;\vdots\; \mathbf{1}_M \boldsymbol{b}_i'); \tag{5.1.12}$$

вектор коэффициентов $\boldsymbol{\delta}$ разобьем соответственно:

$$\boldsymbol{\delta} = \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\gamma} \end{bmatrix}.$$

Вектор коэффициентов $\boldsymbol{\gamma}$ не может быть идентифицирован после преобразования.

- Также может случиться, что, в дополнение к $\boldsymbol{\gamma}$, некоторые коэффициенты в $\boldsymbol{\beta}$ неидентифицируемы, и необходима репараметризация. Рассмотрим пример 5.2, где каждое уравнение имеет различные константы. Одну из констант нужно удалить из $\boldsymbol{F}_i$ и включить в $\boldsymbol{b}_i$, поскольку можно создать общую константу, беря линейную комбинацию констант, специфичных для каждого уравнения. Одна из репараметризаций определяет

$$\boldsymbol{F}_i = \begin{bmatrix} 1 & 0 & S69_i \\ 0 & 1 & S80_i \\ 0 & 0 & S82_i \end{bmatrix}, \quad \boldsymbol{b}_i = (1, IQ_i)', \tag{5.1.13}$$

так что

$$\boldsymbol{Z}_i = \begin{bmatrix} 1 & 0 & S69_i & 1 & IQ_i \\ 0 & 1 & S80_i & 1 & IQ_i \\ 0 & 0 & S82_i & 1 & IQ_i \end{bmatrix}, \quad \boldsymbol{\beta} = (\phi_1 - \phi_3, \phi_2 - \phi_3, \beta)', \quad \boldsymbol{\gamma} = (\phi_3, \gamma)'. \tag{5.1.14}$$

В более общем случае условие идентификации для оценивания фиксированных эффектов состоит в том, что $\boldsymbol{F}_i$ и $\boldsymbol{b}_i$ определяются так, чтобы

$$\mathrm{E}(\boldsymbol{Q}\boldsymbol{F}_i \otimes \boldsymbol{x}_i) \text{ имела полный столбцовый ранг,} \tag{5.1.15}$$

где $\boldsymbol{x}_i =$ объединение $(\boldsymbol{z}_{i1}, \ldots, \boldsymbol{z}_{iM})$. Почему это является условием идентификации? Потому что (как вы увидите более ясно в аналитическом упражнении 2) оценка фиксированных эффектов является частным случаем оценки случайных эффектов, примененной к преобразованной регрессии $\boldsymbol{Qy}$ на $\boldsymbol{QF}$. (5.1.15) является просто адаптацией условия идентификации (5.1.4) к преобразованной регрессии.

С $\boldsymbol{Z}_i$, разделенной таким образом между $\boldsymbol{F}_i$ и $\boldsymbol{b}_i$, систему из M уравнений (5.1.1′) на с. 354 можно записать как

$$\begin{aligned} \boldsymbol{y}_i &= \boldsymbol{F}_i\boldsymbol{\beta} + \mathbf{1}_M \cdot \boldsymbol{b}_i'\boldsymbol{\gamma} + \mathbf{1}_M \cdot \alpha_i + \boldsymbol{\eta}_i \quad (i = 1, 2, \ldots, n), \text{ или} \\ y_{im} &= \boldsymbol{f}_{im}'\boldsymbol{\beta} + \boldsymbol{b}_i'\boldsymbol{\gamma} + \alpha_i + \eta_{im} \ (i = 1, 2, \ldots, n;\ m = 1, \ldots, M), \end{aligned} \tag{5.1.1''}$$

где $\boldsymbol{f}_{im}'$ является m-й строкой в $\boldsymbol{F}_i$. **Модель компонент ошибки** (*error-components model*) состоит из предположения (5.1.1″) (M линейных уравнений), (5.1.2) (случайная выборка), (5.1.8) (предположение SUR с двумя компонентами ошибки), (5.1.4) (идентификация), (5.1.5) (условная гомоскедастичность), (5.1.6) (невырожденность $\mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_i')$) и (5.1.15) (идентификация фиксированных эффектов). Мы включаем дополнительное условие идентификации (5.1.15) для оценивания с фиксированными эффектами, так что обе оценки, случайных и фиксированных эффектов определены для той же самой модели.

Контрольные вопросы

1. (Проверка предположений утверждения 4.7.) Проверьте, что условия утверждения 4.7 удовлетворяются, когда выполнены (5.1.1)–(5.1.6). Покажите, что условие идентификации (5.1.4) выполняется, когда $\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')$ невырождена. **Указание:** *$\boldsymbol{z}_{im}$ является подмножеством $\boldsymbol{x}_i$. Таким образом,* (5.1.4) *является излишним.*

2. (Неединственность репараметризации.) (5.1.13) не является единственной репараметризацией для примера 5.2. Если $\boldsymbol{F}_i$ определяется как

$$\boldsymbol{F}_i = \begin{bmatrix} 0 & 0 & S69_i \\ 1 & 0 & S80_i \\ 0 & 1 & S82_i \end{bmatrix},$$

то как следует определить $\boldsymbol{b}_i$ и $\boldsymbol{\delta}$?

3. Без репараметризации $\boldsymbol{F}_i$ в примере 5.2 равна:

$$\boldsymbol{F}_i = \begin{bmatrix} 1 & 0 & 0 & S69_i \\ 0 & 1 & 0 & S80_i \\ 0 & 0 & 1 & S82_i \end{bmatrix}$$

с $\boldsymbol{b}_i = IQ_i$. Проверьте, что (5.1.15) не удовлетворяется.

5.2. Оценка фиксированных эффектов

Формула

Как уже было отмечено, оценка фиксированных эффектов определяется для преобразованной системы. Умножая обе части (5.1.1″) на с. 358 слева на $\boldsymbol{Q}$ и используя тот факт, что $\boldsymbol{Q}\mathbf{1}_M = \mathbf{0}$, мы получаем

$$\boldsymbol{Q}\boldsymbol{y}_i = \boldsymbol{Q}\boldsymbol{F}_i\boldsymbol{\beta} + \boldsymbol{Q}\boldsymbol{\eta}_i$$

или

$$\widetilde{\boldsymbol{y}}_i = \widetilde{\boldsymbol{F}}_i\boldsymbol{\beta} + \widetilde{\boldsymbol{\eta}}_i \quad (i = 1, 2, \ldots, n), \tag{5.2.1}$$

где

$$\underset{(M\times 1)}{\widetilde{\boldsymbol{y}}_i} \equiv \boldsymbol{Q}\boldsymbol{y}_i, \quad \underset{(M\times \#\boldsymbol{\beta})}{\widetilde{\boldsymbol{F}}_i} \equiv \boldsymbol{Q}\boldsymbol{F}_i, \quad \underset{(M\times 1)}{\widetilde{\boldsymbol{\eta}}_i} \equiv \boldsymbol{Q}.\boldsymbol{\eta}_i. \tag{5.2.2}$$

Таким образом, фиксированный эффект α_i и общие регрессоры выпадают из преобразованных уравнений. Сформируем теперь объединенную выборку преобразованных $\boldsymbol{y}_i$ и $\boldsymbol{F}_i$ как

$$\underset{(Mn\times 1)}{\widetilde{\boldsymbol{y}}} \equiv \begin{bmatrix} \widetilde{\boldsymbol{y}}_1 \\ \vdots \\ \widetilde{\boldsymbol{y}}_n \end{bmatrix}, \quad \underset{(Mn\times \#\boldsymbol{\beta})}{\widetilde{\boldsymbol{F}}} \equiv \begin{bmatrix} \widetilde{\boldsymbol{F}}_1 \\ \vdots \\ \widetilde{\boldsymbol{F}}_n \end{bmatrix}. \tag{5.2.3}$$

Оценка фиксированных эффектов (*fixed-effects estimator*) для $\boldsymbol{\beta}$, обозначаемая как $\widehat{\boldsymbol{\beta}}_{\text{FE}}$, является объединенной OLS-оценкой, то есть OLS-оценкой, примененной к объединенной выборке $(\widetilde{\boldsymbol{y}}, \widetilde{\boldsymbol{F}})$ объема MN. Следовательно,

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}_{\text{FE}} &\equiv (\widetilde{\boldsymbol{F}}'\widetilde{\boldsymbol{F}})^{-1}\widetilde{\boldsymbol{F}}'\widetilde{\boldsymbol{y}} = \\
&= \Big(\frac{1}{n}\sum_{i=1}^{n}\widetilde{\boldsymbol{F}}_i'\widetilde{\boldsymbol{F}}_i\Big)^{-1}\frac{1}{n}\sum_{i=1}^{n}\widetilde{\boldsymbol{F}}_i'\widetilde{\boldsymbol{y}}_i = \\
&= \Big[\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{Q}\boldsymbol{F}_i)'(\boldsymbol{Q}\boldsymbol{F}_i)\Big]^{-1}\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{Q}\boldsymbol{F}_i)'(\boldsymbol{Q}\boldsymbol{y}_i) = \\
&= \Big(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{F}_i'\boldsymbol{Q}\boldsymbol{Q}\boldsymbol{F}_i\Big)^{-1}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{F}_i'\boldsymbol{Q}\boldsymbol{Q}\boldsymbol{y}_i = \\
&= \Big(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{F}_i'\boldsymbol{Q}\boldsymbol{F}_i\Big)^{-1}\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{F}_i'\boldsymbol{Q}\boldsymbol{y}_i
\end{aligned}$$

$$\text{(поскольку } \boldsymbol{Q} \text{ симметрична и идемпотентна).} \tag{5.2.4}$$

Подставляя верхнее уравнение (5.2.1) в (5.2.4), можно получить ошибку оценки как

$$\widehat{\beta}_{\mathrm{FE}} - \beta = \Big(\frac{1}{n}\sum_{i=1}^{n}\widetilde{F}_i'\widetilde{F}_i\Big)^{-1}\frac{1}{n}\sum_{i=1}^{n}\widetilde{F}_i'\widetilde{\eta}_i = $$
$$= \Big(\frac{1}{n}\sum_{i=1}^{n}F_i'QF_i\Big)^{-1}\frac{1}{n}\sum_{i=1}^{n}F_i'Q\eta_i. \tag{5.2.5}$$

Будучи основанной на отклонениях от групповых средних, оценка фиксированных эффектов известна также как **внутри-оценка** (*within estimator*), или **ковариационная оценка** (*covariance estimator*). Другое, и, возможно, более популярное, ее получение заключается в применении OLS к уровням, то есть не в отклонениях, но со специфическими дамми для групп (i), добавленными к списку регрессоров. По этой причине эта оценка называется также **OLS-дамми** (*least squares dummy variables —* **LSDV**) оценкой. LSDV интерпретация позволяет нам увидеть, какие нужны дополнительные предположения для модели компонент ошибки при разработке теории конечных выборок для этой оценки. Это остается в качестве аналитического упражнения 1.

Свойства на больших выборках

Как вы докажете в аналитическом упражнении 2, оценку фиксированных эффектов можно также записать как GMM-оценку, что дает возможность непосредственной разработки теории для больших выборок для оценки фиксированных эффектов, что подытоживается в утверждении 5.1 ниже. Доказательство этого утверждения, однако, можно выполнить более просто, изучая выражение (5.2.5) для ошибки оценки.

Утверждение 5.1 (свойства оценки фиксированных эффектов на больших выборках): *Рассмотрим модель компонент ошибки (состоящую из* (5.1.1″) *на с. 358,* (5.1.2), (5.1.8), (5.1.4)–(5.1.6) *и* (5.1.15)), *но ослабим предположение SUR* (5.1.8), *требуя только, чтобы*

$$\mathrm{E}(\boldsymbol{f}_{im}\cdot\eta_{ih}) = \mathbf{0} \quad \text{для всех } m, h \ (= 1, 2, \ldots, M). \tag{5.2.6}$$

где $\boldsymbol{f}_{im}'$ является m-й строкой $\boldsymbol{F}_i$. Определим $\widetilde{\boldsymbol{y}}_i$, $\widetilde{\boldsymbol{F}}_i$, и $\widetilde{\boldsymbol{\eta}}_i$ посредством (5.2.2). *Тогда:*

(a) оценка фиксированных эффектов (5.2.4) *состоятельна и асимптотически нормальна с*

$$\mathrm{Avar}(\widehat{\boldsymbol{\beta}}_{\mathrm{FE}}) = \big[\mathrm{E}(\widetilde{\boldsymbol{F}}_i'\widetilde{\boldsymbol{F}}_i)\big]^{-1}\mathrm{E}\big[\widetilde{\boldsymbol{F}}_i'\,\mathrm{E}(\widetilde{\boldsymbol{\eta}}_i\widetilde{\boldsymbol{\eta}}_i')\widetilde{\boldsymbol{F}}_i\big]\big[\mathrm{E}(\widetilde{\boldsymbol{F}}_i'.\widetilde{\boldsymbol{F}}_i)\big]^{-1}. \tag{5.2.7}$$

(b) Эта асимптотическая дисперсия состоятельно оценивается как

$$\widehat{\mathrm{Avar}(\widehat{\beta}_{\mathrm{FE}})} = \Big(\frac{1}{n}\sum_{i=1}^{n}\widetilde{F}_i'\widetilde{F}_i\Big)^{-1}\Big(\frac{1}{n}\sum_{i=1}^{n}\widetilde{F}_i'\widehat{V}\widetilde{F}_i\Big)\Big(\frac{1}{n}\sum_{i=1}^{n}\widetilde{F}_i'\widetilde{F}_i\Big)^{-1}, \quad (5.2.8)$$

где $\widehat{V}$ является выборочной матрицей из перекрестных моментов преобразованных остатков, связанных с оценкой фиксированных эффектов:

$$\underset{(M\times M)}{\widehat{V}} \equiv \frac{1}{n}\sum_{i=1}^{n}(\widetilde{y}_i - \widetilde{F}_i\widehat{\beta}_{\mathrm{FE}})(\widetilde{y}_i - \widetilde{F}_i\widehat{\beta}_{\mathrm{FE}})'. \quad (5.2.9)$$

Единственная нетривиальная часть доказательства состоит в том, чтобы показать:

(1) $\mathrm{E}(\widetilde{F}_i'\widetilde{F}_i)$ является невырожденной (и, следовательно, обратимой).

(2) $\mathrm{E}(F_i'Q\eta_i) = \mathbf{0}$ (что нужно для состоятельности) и

(3) $\mathrm{E}(\widetilde{F}_i'\widetilde{\eta}_i\widetilde{\eta}_i'\widetilde{F}_i)$ (что является Avar для $\frac{1}{n}\sum_{i=1}^{n}\widetilde{F}_i'\widetilde{\eta}_i) = \mathrm{E}\big[\widetilde{F}_i'\,\mathrm{E}(\widetilde{\eta}_i\widetilde{\eta}_i')\widetilde{F}_i\big]$.

Оставшаяся часть доказательства является простым применением LLN Колмогорова и CLT Линдеберга — Леви.

Доказательство того, что из (5.1.15) следует (1), остается в качестве дополнительного аналитического упражнения.

Для доказательства (2) используйте формулу (4.6.16b) и перепишите математическое ожидание как

$$\begin{aligned}\mathrm{E}(F_i'Q\eta_i) &= \mathrm{E}\Big(\sum_{m=1}^{M}\sum_{h=1}^{M} q_{mh} f_{im}\cdot\eta_{ih}\Big) = \qquad (q_{mh} \equiv (m,h) \text{ элемент } Q)\\ &= \sum_{m=1}^{M}\sum_{h=1}^{M} q_{mh}\,\mathrm{E}(f_{im}\cdot\eta_{ih}).\end{aligned} \quad (5.2.10)$$

Оно равно нулю вследствие условий ортогональности (5.2.6).

Доказательство (3) заключается в следующем. Сначала заметим, что

$$\mathrm{E}(\widetilde{F}_i'\widetilde{\eta}_i\widetilde{\eta}_i'\widetilde{F}_i) = \mathrm{E}\big[\,\widetilde{F}_i'\,\mathrm{E}(\widetilde{\eta}_i\widetilde{\eta}_i'|x_i)\widetilde{F}_i\big].$$

(Это следует из закона полных математических ожиданий и того факта, что x_i [= объединение $(z_{i1},\dots,z_{iM})$] включает в себя элементы из F_i и $\widetilde{F}_i$ является функцией от F_i.) Но $\mathrm{E}(\widetilde{\eta}_i\widetilde{\eta}_i'|x_i)$ равно безусловному среднему $\mathrm{E}(\widetilde{\boldsymbol{\eta}}_i\widetilde{\boldsymbol{\eta}}_i')$, поскольку

$$\begin{aligned}
&\mathrm{E}(\widetilde{\boldsymbol{\eta}}_i\widetilde{\boldsymbol{\eta}}_i'|\boldsymbol{x}_i) = \mathrm{E}(\widetilde{\boldsymbol{\varepsilon}}_i\widetilde{\boldsymbol{\varepsilon}}_i'|\boldsymbol{x}_i)\\
&(\text{так как } \widetilde{\boldsymbol{\varepsilon}}_i \equiv \boldsymbol{Q}\boldsymbol{\varepsilon}_i = \boldsymbol{Q}(\mathbf{1}_M \cdot \alpha_i + \boldsymbol{\eta}_i) = \boldsymbol{Q}\boldsymbol{\eta}_i = \widetilde{\boldsymbol{\eta}}_i)\\
&= \boldsymbol{Q}\,\mathrm{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'|\boldsymbol{x}_i)\boldsymbol{Q} =\\
&= \boldsymbol{Q}\,\mathrm{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i')\boldsymbol{Q} \quad (\text{согласно (5.1.5)}) =\\
&= \mathrm{E}(\widetilde{\boldsymbol{\varepsilon}}_i\widetilde{\boldsymbol{\varepsilon}}_i') =\\
&= \mathrm{E}(\widetilde{\boldsymbol{\eta}}_i\widetilde{\boldsymbol{\eta}}_i') \quad (\text{так как } \widetilde{\boldsymbol{\varepsilon}}_i = \widetilde{\boldsymbol{\eta}}_i).
\end{aligned} \tag{5.2.11}$$

Беглый взгляд на выражение для асимптотической дисперсии должен побудить вас предположить, что существует какая-то другая оценка, которая является асимптотически более эффективной, потому что асимптотическую дисперсию эффективной оценки можно записать в виде обратной матрицы. В самом деле, как вы докажете в аналитическом упражнении 2, существует оценка, которая асимптотически более эффективна, чем оценка фиксированных эффектов.

Отступление: когда η_i является сферической

Обычно в модели компонент ошибки дополнительно предполагается, что вторая компонента ошибки $\boldsymbol{\eta}_i$ является сферической:

$$\mathrm{E}(\boldsymbol{\eta}_i\boldsymbol{\eta}_i') = \sigma_\eta^2 \boldsymbol{I}_M, \text{ так что } \mathrm{E}(\widetilde{\boldsymbol{\eta}}_i\widetilde{\boldsymbol{\eta}}_i') = \sigma_\eta^2\boldsymbol{Q}. \tag{5.2.12}$$

Если, как во многих панельных данных, m является временем, то это предположение часто называют предположением об «отсутствии сериальной корреляции». Отсутствие сериальной корреляции в этом случае не следует путать с условием, что $\boldsymbol{\eta}_i$ не коррелирована с $\boldsymbol{\eta}_j$ для $i \neq j$. Последнее является следствием нашего поддерживаемого предположения о том, что выборка является i.i.d.

Подставляя это в (5.2.7) и замечая, что $\boldsymbol{Q}\widetilde{\boldsymbol{F}}_i = \widetilde{\boldsymbol{F}}_i$, мы обнаруживаем, что выражение для асимптотической дисперсии упрощается до

$$\mathrm{Avar}(\widehat{\boldsymbol{\beta}}_{\mathrm{FE}}) = \sigma_\eta^2 \cdot \left[\mathrm{E}(\widetilde{\boldsymbol{F}}_i'\widetilde{\boldsymbol{F}}_i)\right]^{-1}. \tag{5.2.13}$$

Если существует состоятельная оценка $\widehat{\sigma}_\eta^2$ для σ_η^2, то эта асимптотическая дисперсия состоятельно оценивается как

$$\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\beta}}_{\mathrm{FE}})} = \widehat{\sigma}_\eta^2 \cdot \left(\frac{1}{n}\sum_{i=1}^{n} \widetilde{\boldsymbol{F}}_i'\widetilde{\boldsymbol{F}}_i\right)^{-1}, \tag{5.2.14}$$

что равно n, умноженному на $\widehat{\sigma}_\eta^2 \cdot (\widetilde{\boldsymbol{F}}'\widetilde{\boldsymbol{F}})^{-1}$, то есть n, умноженному на обычное выражение для оцененной матрицы дисперсии, когда OLS применяется к объединенной выборке $(\widetilde{\boldsymbol{y}}, \widetilde{\boldsymbol{F}})$ объема Mn.

Чтобы расширить аналогию с OLS, определим SSR как

$$SSR = (\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{F}}\widehat{\boldsymbol{\beta}}_{\mathrm{FE}})'(\widetilde{\boldsymbol{y}} - \widetilde{\boldsymbol{F}}\widehat{\boldsymbol{\beta}}_{\mathrm{FE}}) = \sum_{i=1}^{n}(\widetilde{\boldsymbol{y}}_i - \widetilde{\boldsymbol{F}}_i\widehat{\boldsymbol{\beta}}_{\mathrm{FE}})'(\widetilde{\boldsymbol{y}}_i - \widetilde{\boldsymbol{F}}_i\widehat{\boldsymbol{\beta}}_{\mathrm{FE}}). \tag{5.2.15}$$

В объединенной OLS на отклонения от групповых средних есть $\#\boldsymbol{\beta}$ коэффициентов, которые нужно оценить. Обычная OLS-оценка дисперсии ошибок равна:

$$\frac{SSR}{Mn - \#\boldsymbol{\beta}}. \tag{5.2.16}$$

Она, однако, *не* является состоятельной для σ_η^2. Вас попросят показать в аналитическом упражнении 3, что состоятельной оценкой является

$$\frac{SSR}{Mn - n - \#\boldsymbol{\beta}}. \tag{5.2.17}$$

(Вообще-то, и $SSR/(Mn - n)$ также является состоятельной.) Причина того, что из знаменателя (5.2.16) нужно вычесть n, заключается в том, что M преобразованных уравнений не являются линейно независимыми; для обеих частей M преобразованных уравнений сумма по i всегда равна нулю, как вы сможете увидеть, умножая обе части (5.2.1) слева на $\mathbf{1}'_M$ и используя то, что $\mathbf{1}'_M\boldsymbol{Q} = \mathbf{0}'$. Эффективный объем объединенной выборки равен $Mn - n$, а не Mn. Использование (5.2.16) вместо (5.2.17) является очень распространенной ошибкой, которая приводит к недооценке стандартных ошибок и переоценке t-значений. Например, если $M = 3$, $n = 1000$ и $\#\boldsymbol{\beta} = 5$, то t-значения будут переоценены примерно на 23% (= квадратному корню из 2995/1995 минус единица).

Сравнение случайных эффектов с фиксированными

Вернемся теперь к общему случаю с отсутствием ограничений на $\mathrm{E}(\boldsymbol{\eta}_i\boldsymbol{\eta}_i')$, так что ошибки могут иметь произвольную структуру «сериальной корреляции» вдоль m. Сравнивая (5.2.6) с (5.1.8), мы видим, что условия ортогональности, *не* используемые в оценке фиксированных эффектов, имеют вид:

$$\mathrm{E}(\boldsymbol{f}_{im}\cdot\alpha_i) = \mathbf{0} \text{ для всех } m, \quad \mathrm{E}(\boldsymbol{b}_i\cdot\alpha_i) = \mathbf{0}, \quad \mathrm{E}(\boldsymbol{b}_i\cdot\eta_{im}) = \mathbf{0} \text{ для всех } m. \tag{5.2.18}$$

То есть оценка фиксированных эффектов является робастной к нарушению (5.2.18). Пусть $\widehat{\boldsymbol{\beta}}_{\mathrm{RE}}$ является элементом в $\widehat{\boldsymbol{\delta}}_{\mathrm{RE}}$ (оценка случайных эффектов, примененная к оригинальной системе (5.1.1″) из M уравнений на с. 358, который соответствует $\boldsymbol{\beta}$:

$$\widehat{\boldsymbol{\delta}}_{\mathrm{RE}} \equiv \begin{bmatrix}\widehat{\boldsymbol{\beta}}_{\mathrm{RE}}\\ \widehat{\boldsymbol{\gamma}}_{\mathrm{RE}}\end{bmatrix}. \tag{5.2.19}$$

Также пусть $\mathrm{Avar}(\widehat{\beta}_{\mathrm{RE}})$ является асимптотической дисперсией для $\widehat{\beta}_{\mathrm{RE}}$. Это ведущая подматрица в $\mathrm{Avar}(\widehat{\delta}_{\mathrm{RE}})$ (заданная в (4.6.9′) на с. 321), соответствующая β.

Оценка случайных эффектов $\widehat{\beta}_{\mathrm{RE}}$ является эффективной, в то время как $\widehat{\beta}_{\mathrm{FE}}$ является состоятельной, но неэффективной. Однако, если (5.2.18) нарушается, больше нет гарантии того, что $\widehat{\beta}_{\mathrm{RE}}$ будет состоятельной, в то время как $\widehat{\beta}_{\mathrm{FE}}$ остается состоятельной. Таким образом, естественным тестом для (5.2.18) будет рассмотрение разности между этими двумя оценками:

$$\widehat{q} \equiv \widehat{\beta}_{\mathrm{FE}} - \widehat{\beta}_{\mathrm{RE}}. \tag{5.2.20}$$

Легко доказать, что вектор $\sqrt{n}\widehat{q}$ асимптотически нормален. Принцип Хаусмана (первоначально разработанный для ML-оценок, но, как доказано в аналитическом упражнении 9 к главе 3, применимый и к GMM-оценкам) предполагает, что

$$\mathrm{Avar}(\widehat{q}) = \mathrm{Avar}(\widehat{\beta}_{\mathrm{FE}}) - \mathrm{Avar}(\widehat{\beta}_{\mathrm{RE}}). \tag{5.2.21}$$

(Поэтому нет необходимости включать асимптотическую ковариацию между $\widehat{\beta}_{\mathrm{FE}}$ и $\widehat{\beta}_{\mathrm{RE}}$.) По утверждению 4.7, ведущая подматрица (4.6.9′) на с. 321 дает состоятельную оценку для $\mathrm{Avar}(\widehat{\beta}_{\mathrm{RE}})$. Из утверждения 5.1 следует, что $\mathrm{Avar}(\widehat{\beta}_{\mathrm{FE}})$ состоятельно оценивается посредством (5.2.8). Следовательно, состоятельная оценка для $\mathrm{Avar}(\widehat{q})$ является их разностью. Запишем ее как $\widehat{\mathrm{Avar}(\widehat{q})}$. В приложении доказывается, что (i) $\mathrm{Avar}(\widehat{q})$ является невырожденной (и, следовательно, положительно определенной) и (ii) статистика Хаусмана, определенная ниже, гарантированно является неотрицательной для любой выборки $\{y_i, Z_i\}$. Подведем итог.

Утверждение 5.2 (тест Хаусмана на спецификацию): *Допустим, что выполняются предположения модели компонент ошибки* ((5.1.1″) *(на с. 358)*, (5.1.2), (5.1.8), (5.1.4)–(5.1.6) *и* (5.1.15)). *Определим* $\widehat{q}$ *и* $\widehat{\mathrm{Avar}(\widehat{q})}$ *как только что было описано. Тогда статистика Хаусмана*

$$H \equiv n \cdot \widehat{q}'\left(\widehat{\mathrm{Avar}(\widehat{q})}\right)^{-1}\widehat{q} \tag{5.2.22}$$

асимптотически распределена как хи-квадрат с $\#\beta$ *степенями свободы. Она неотрицательна в конечных выборках.*

Этот тест является тестом на спецификацию, поскольку он может обнаружить нарушение (5.2.18), которое является частью поддерживаемых предположений модели компоненты ошибки. Более точно рассмотрим последовательность локальных альтернатив, которые удовлетворяют всем предположениям модели компонент ошибки, за исключением (5.2.18)[1].

[1]Понятие локальных альтернатив вводится в параграфе 2.4.

При некоторых дополнительных технических предположениях можно показать, что статистика Хаусмана сходится вдоль этих последовательностей локальных альтернатив к нецентральному χ^2 распределению с некоторым параметром нецентральности[1]. То есть статистика Хаусмана имеет мощность против локальных альтернатив, при которых оценка случайных эффектов несостоятельна.

Ослабление условной гомоскедастичности

Получить распределение оценки фиксированных эффектов на больших выборках нетрудно и без условной гомоскедастичности. Как будет видно в следующем параграфе, обобщить результаты для больших выборок с целью охвата несбалансированных панелей на самом деле легче без условной гомоскедастичности.

Утверждение 5.3 (оценка фиксированных эффектов без условной гомоскедастичности): *Уберем условную гомоскедастичность (5.1.5) из предположений утверждения 5.1. Определим $\widetilde{\boldsymbol{y}}_i$, $\widetilde{\boldsymbol{F}}_i$ и $\widetilde{\boldsymbol{\eta}}_i$ посредством* (5.2.2). *Тогда:*

(a) оценка фиксированных эффектов (5.2.4) *состоятельна и асимптотически нормальна с*

$$\operatorname{Avar}(\widehat{\boldsymbol{\beta}}_{\mathrm{FE}}) = \left[\mathrm{E}(\widetilde{\boldsymbol{F}}_i'\widetilde{\boldsymbol{F}}_i)\right]^{-1} \mathrm{E}(\widetilde{\boldsymbol{F}}_i'\widetilde{\boldsymbol{\eta}}_i\widetilde{\boldsymbol{\eta}}_i'\widetilde{\boldsymbol{F}}_i) \left[\mathrm{E}(\widetilde{\boldsymbol{F}}_i'\widetilde{\boldsymbol{F}}_i)\right]^{-1}. \tag{5.2.23}$$

(b) Если кроме этого выполняется соответствующее предположение о конечности четвертого момента (которое является адаптацией предположения 4.6), то асимптотическая дисперсия состоятельно оценивается как

$$\widehat{\operatorname{Avar}(\widehat{\boldsymbol{\beta}}_{\mathrm{FE}})} = \left(\frac{1}{n}\sum_{i=1}^{n}\widetilde{\boldsymbol{F}}_i'\widetilde{\boldsymbol{F}}_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}(\widetilde{\boldsymbol{F}}_i'\check{\boldsymbol{\eta}}_i\check{\boldsymbol{\eta}}_i'\widetilde{\boldsymbol{F}}_i)\right)\left(\frac{1}{n}\sum_{i=1}^{n}\widetilde{\boldsymbol{F}}_i'\widetilde{\boldsymbol{F}}_i\right)^{-1}. \tag{5.2.24}$$

где $\check{\boldsymbol{\eta}}_i \equiv \widetilde{\boldsymbol{y}}_i - \widetilde{\boldsymbol{F}}_i\widehat{\boldsymbol{\beta}}_{\mathrm{FE}}$ (это $\check{\boldsymbol{\eta}}_i$ не нужно путать с $\widetilde{\boldsymbol{\eta}}_i$ в (5.2.1)*).*

И опять, есть два способа доказать это. Запишите оценку фиксированных эффектов как GMM-оценку и затем примените теорию больших выборок для GMM-оценки. Или вы можете доказать это непосредственно, рассматривая выражение (5.2.4) для ошибки оценки. Единственная нетривиальная часть доказательства — показать, что средняя сумма в выражении для $\widehat{\operatorname{Avar}(\widehat{\boldsymbol{\beta}}_{\mathrm{FE}})}$ сходится по вероятности к ее популяционному аналогу, $\mathrm{E}(\widetilde{\boldsymbol{F}}_i'\widetilde{\boldsymbol{\eta}}_i\widetilde{\boldsymbol{\eta}}_i'\widetilde{\boldsymbol{F}}_i)$. Требующийся метод доказательства

---

[1] См.: [Newey, 1985]. Требуемое техническое предположение является предположением 5 в: [Newey, 1985].

очень похож на тот, который используется для доказательства утверждения 4.2, и поэтому не повторяется здесь.

Контрольные вопросы

1. Должны ли мы, для того чтобы выполнялись результаты для больших выборок в этом параграфе, предполагать, что α_i и η_i ортогональны друг другу?

2. (Удаление дублирующих уравнений.) Как упоминалось в основном тексте, M преобразованных уравнений не являются линейно независимыми, в том смысле, что любое из M преобразованных уравнений является линейной комбинацией остальных $M-1$ уравнений. Поэтому рассмотрим формирование объединенной выборки объема $Mn-n$, удаляя одно из преобразованных уравнений (скажем, последнее уравнение) для каждой группы (i). Является ли OLS-оценка в этой меньшей объединенной выборке оценкой фиксированных эффектов? [Ответ: Нет.]

3. (Важность перекрестной ортогональности.) Предположим, что (5.2.6) выполняется для $m = h$, но не обязательно для $m \neq h$. Будет ли оценка фиксированных эффектов обязательно состоятельной? [Ответ: Нет.]

4. Проверьте, что $\mathrm{E}(\widetilde{\boldsymbol{\eta}}_i\widetilde{\boldsymbol{\eta}}_i')$ является вырожденной матрицей. (Однако матрицы $\mathrm{E}[\widetilde{\boldsymbol{F}}_i'\,\mathrm{E}(\widetilde{\boldsymbol{\eta}}_i\widetilde{\boldsymbol{\eta}}_i')\widetilde{\boldsymbol{F}}_i]$ в (5.2.7) и $\mathrm{E}(\widetilde{\boldsymbol{F}}_i'\widetilde{\boldsymbol{\eta}}_i\widetilde{\boldsymbol{\eta}}_i'\widetilde{\boldsymbol{F}}_i)$ в (5.2.23) являются невырожденными, как вас попросят показать в дополнительном аналитическом упражнении.)

5. (Состоятельное оценивание ковариаций ошибок.) Чтобы доказать, что (5.2.9) является состоятельной оценкой для $\mathrm{E}(\widetilde{\boldsymbol{\eta}}_i\widetilde{\boldsymbol{\eta}}_i')$, вы должны применить утверждение 4.1 к преобразованной системе из M уравнений. Проверьте, что условия утверждения здесь выполняются. В частности, выполняется ли требование того, чтобы $\mathrm{E}(\widetilde{\boldsymbol{f}}_{im}\widetilde{\boldsymbol{f}}_{ih}')$ (где $\widetilde{\boldsymbol{f}}_{im}$ является m-й строкой $\widetilde{\boldsymbol{F}}_i$) было конечным, в модели компонент ошибки? **Указание:** *$\widetilde{\boldsymbol{f}}_{im}$ является линейным преобразованием $\boldsymbol{x}_i$. Из* (5.1.5) *и* (5.1.6) *следует, что* $\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')$ *невырождена.*

6. (Что тестирует статистика Хаусмана?) Для простоты предположим, что матрица $\boldsymbol{\Sigma}(\equiv \mathrm{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'))$ известна. Также для простоты предположим, что общих регрессоров нет, так что $\boldsymbol{z}_{im} = \boldsymbol{f}_{im}$, и (5.2.18) редуцируется к (5.1.8a): $\mathrm{E}(\boldsymbol{z}_{im}\cdot\alpha_i) = \mathbf{0}$ для всех m. Это является ограничением, *не* используемым в оценке фиксированных эффектов. Является ли оценка случайных эффектов обязательно несостоятельной, когда (5.1.8a) не выполняется, но выполняются все другие предположения модели компонент ошибки (такие как (5.1.8b))? **Указание:** *Покажите, что без* (5.1.8a)

$$\operatorname{plim}\widehat{\boldsymbol{\delta}}_{\mathrm{RE}} - \boldsymbol{\delta} = \mathrm{E}(\boldsymbol{Z}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}_i)^{-1}\,\mathrm{E}(\boldsymbol{Z}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_i),$$

где $\boldsymbol{\varepsilon}_i = \mathbf{1}\cdot\alpha_i + \boldsymbol{\eta}_i$. При условии, что (5.1.8b) *выполняется,* (5.1.8a) *является достаточным для того, чтобы* $\mathrm{E}(\boldsymbol{Z}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_i)$ *было равно нулю. Является ли оно необходимым?* Таким образом, строго говоря, тест Хаусмана не является тестом для (5.1.8a); он является тестом для $\mathrm{E}(\boldsymbol{Z}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}_i) = \mathbf{0}$.

5.3. Несбалансированные панели (необязательный параграф)

Применяя модель нескольких уравнений к панельным данным, мы сделали неявное, но важное предположение о том, что переменные наблюдаемы для всех m, так что число наблюдений для каждой рассматриваемой единицы одинаковое (M). Такая панель называется **сбалансированной панелью** (*balanced panel*). Однако доступных сбалансированных множеств панельных данных не существует из-за того, что одни обследуемые единицы выходят из обследования, а другие входят в него. Неизбежно некоторые фирмы исчезают из обследуемой выборки из-за банкротств и слияний до окончания опросов, в то время как те фирмы, которых не было в начале исследования, могут быть включены в него позже. Некоторые из домохозяйств, первоначально участвовавших в опросе, не доводят его до конца или из-за исчезновения домохозяйства, или из-за того, что респондентам надоедает отвечать на одни и те же вопросы снова и снова. Панель идентичных братьев и сестер является несбалансированной, если она включает в себя, наряду с парами близнецов, также идентичные тройни (и даже семерни).

В этом параграфе показано, что оценка фиксированных эффектов легко обобщается на несбалансированные панели. Этот оптимистичный вывод, однако, зависит от предположения о том, что нахождение обследуемой единицы в выборке не зависит от ошибки. Без этого предположения оценки уже не являются состоятельными. Этот феномен называется **смещением из-за отбора** (*selectivity bias*). Всестороннее обсуждение того, почему возникает смещение из-за отбора и как это можно исправить, переносится в параграф 8.2, поскольку этот вопрос не является специфичным для панельных данных.

Обнуление пропущенных наблюдений

Чтобы справиться с пропущенными наблюдениями, удобно определить дамми-переменную, d_{im},

$$d_{im} = \begin{cases} 1 & \text{если наблюдение } m \text{ находится в выборке} \\ 0 & \text{в противном случае} \end{cases} \tag{5.3.1}$$

и

$$\underset{(M\times 1)}{\boldsymbol{d}_i} = \begin{bmatrix} d_{i1} \\ \vdots \\ d_{iM} \end{bmatrix}, \quad M_i = \sum_{m=1}^{M} d_{im} = \#\ \text{наблюдений для группы } i. \tag{5.3.2}$$

Если наблюдение m пропущено для группы i, заполним m-е строки $\boldsymbol{y}_i$, $\boldsymbol{F}_i$ и $\boldsymbol{\eta}_i$ нулями так, чтобы они продолжали иметь фиксированный размер.

То есть

$$\underset{(M\times 1)}{\boldsymbol{y}_i} = \begin{bmatrix} d_{i1}\cdot y_{i1} \\ \vdots \\ d_{iM}\cdot y_{iM} \end{bmatrix}, \quad \underset{(M\times \#\boldsymbol{\beta})}{\boldsymbol{F}_i} = \begin{bmatrix} d_{i1}\cdot \boldsymbol{f}'_{i1} \\ \vdots \\ d_{iM}\cdot \boldsymbol{f}'_{iM} \end{bmatrix}, \quad \underset{(M\times 1)}{\boldsymbol{\eta}_i} = \begin{bmatrix} d_{i1}\cdot \eta_{i1} \\ \vdots \\ d_{iM}\cdot \eta_{iM} \end{bmatrix}. \tag{5.3.3}$$

Таким образом, для каждого i и m либо все элементы вектора $(y_{im}, \boldsymbol{f}_{im})$ наблюдаемы, либо не наблюдаем ни один. Мы не рассматриваем случай, в котором некоторые, но не все, элементы $(y_{im}, \boldsymbol{f}_{im})$ наблюдаемы.

В этих новых обозначениях, (5.1.1″) на с. 358 принимает вид:

$$\boldsymbol{y}_i = \boldsymbol{F}_i\boldsymbol{\beta} + \boldsymbol{d}_i\cdot \boldsymbol{b}'_i\boldsymbol{\gamma} + \boldsymbol{d}_i\cdot\alpha_i + \boldsymbol{\eta}_i \quad (i = 1, 2, \ldots, n). \tag{5.3.4}$$

В случае сбалансированных панелей, где $\boldsymbol{d}_i = \mathbf{1}_M$, преобразующая матрица $\boldsymbol{Q}$ в формуле фиксированных эффектов (5.2.4) является аннулятором, соответствующим $\mathbf{1}_M$ (см. (5.1.10)). При наличии пропущенных наблюдений мы используем в качестве преобразующей матрицы аннулятор, соответствующий

$$\begin{aligned} \underset{(M\times M)}{\boldsymbol{Q}} &= \boldsymbol{I}_M - \boldsymbol{d}_i(\boldsymbol{d}'_i\boldsymbol{d}_i)^{-1}\boldsymbol{d}'_i = \\ &= \boldsymbol{I}_M - \frac{1}{M_i}\boldsymbol{d}_i\boldsymbol{d}'_i \quad (\text{поскольку } \boldsymbol{d}'_i\boldsymbol{d}_i = M_i). \end{aligned} \tag{5.3.5}$$

Эта матрица зависит от i, поскольку $\boldsymbol{d}_i$ для разных i различны. Однако мы оставляем эту зависимость $\boldsymbol{Q}$ от i неявной для экономности обозначений. С таким образом переопределенными $\boldsymbol{y}_i$, $\boldsymbol{F}_i$, $\boldsymbol{\eta}_i$ и $\boldsymbol{Q}$ формула для оценки фиксированных эффектов остается той же самой, что и (5.2.4). Подставляя (5.3.4) в эту формулу и замечая, что $\boldsymbol{Q}\boldsymbol{d}_i = \mathbf{0}$, мы видим, что выражение для ошибки оценки точно такое же, как и (5.2.5).

Обнуление против сжатия

Как было видно в параграфе 5.2 для случая сбалансированных панелей, оценку фиксированных эффектов можно вычислить как OLS-оценку по объединенной выборке (размера Mn) отклонений от групповых средних. Как и выше, пусть $\widetilde{\boldsymbol{y}}_i$, $\widetilde{\boldsymbol{F}}_i$ и $\widetilde{\boldsymbol{\eta}}_i$ являются преобразованиями $\boldsymbol{y}_i$, $\boldsymbol{F}_i$, $\boldsymbol{\eta}_i$ через $\boldsymbol{Q}$ соответственно. Например, $\widetilde{\boldsymbol{y}}_i$ выглядит как

$$\underset{(M\times 1)}{\widetilde{\boldsymbol{y}}_i} \equiv \boldsymbol{Q}\boldsymbol{y}_i = \begin{bmatrix} d_{i1}\cdot y_{i1} - d_{i1}\cdot\overline{y}_i \\ \vdots \\ d_{iM}\cdot y_{iM} - d_{iM}\cdot\overline{y}_i \end{bmatrix},$$

где

$$\bar{y}_i \equiv \frac{1}{M_i} \boldsymbol{d}_i' \boldsymbol{y}_i =$$
$$= \frac{1}{M_i} \times (\text{сумма } y_{im} \text{ по тем } m, \text{ для которых данные доступны}). \quad (5.3.6)$$

Таким образом, если наблюдение m для группы i доступно, m-й элемент $\widetilde{\boldsymbol{y}}_i$ все еще является отклонением y_{im} от группового среднего, но групповое среднее вычисляется по доступным наблюдениям. Поскольку $d_{im} = 0$, если наблюдение m недоступно, строки в $\widetilde{\boldsymbol{y}}_i$, соответствующие пропущенным наблюдениям, равны нулю. Такую же структуру имеют и столбцы $\widetilde{\boldsymbol{F}}_i$. Таким образом, эти строки можно удалить из объединенной выборки без влияния на численное значение оценки фиксированных эффектов. То есть мы можем до формирования объединенной выборки «сжать» $(\widetilde{\boldsymbol{y}}_i, \widetilde{\boldsymbol{F}}_i)$, удаляя строки, заполненные нулями.

Отсутствие смещения вследствие отбора

Оценка фиксированных эффектов, адаптированная таким образом к несбалансированным панелям, остается состоятельной и асимптотически нормальной, если условия ортогональности (5.2.6) в утверждении 5.1 модифицируются как

$$\text{(отсутствие смещения вследствие отбора)}$$
$$\mathrm{E}(\boldsymbol{f}_{im} \cdot \eta_{ih} | \boldsymbol{d}_i) = \boldsymbol{0} \quad (m, h = 1, 2, \ldots, M). \quad (5.3.7)$$

Как показано в следующем утверждении, это условие является ключевым для гарантирования состоятельности оценки фиксированных эффектов. По закону полных математических ожиданий, это условие является более сильным, чем условие (5.2.6), требующее, чтобы было равно нулю *без*условное математическое ожидание. Легко построить пример, где (5.3.7) не выполняется, поскольку характер выбора $\boldsymbol{d}_i$ зависит от ошибки $\boldsymbol{\eta}_i$. Заметим, что (5.3.7) не включает фиксированный эффект α_i. Таким образом, если зависимость отбора в выборку от ошибок реализуется только через фиксированный эффект, проблема смещения из-за такого отбора не возникает.

Если условия ортогональности (5.2.6) усиливаются до (5.3.7), то тогда выводы утверждения 5.3 переносятся на несбалансированные панели:

Утверждение 5.4 (свойства на больших выборках оценки фиксированных эффектов для несбалансированной панели): *Рассмотрим модель компонент ошибки, состоящей из* (5.3.4), (5.1.2), (5.1.4)–(5.1.6), (5.1.15) *и* (5.3.7) *с* $(\boldsymbol{y}_i, \boldsymbol{F}_i, \boldsymbol{\eta}_i)$, *как в* (5.3.3). *Как и выше, определим*

$(\widetilde{\boldsymbol{y}}_i, \widetilde{\boldsymbol{F}}_i, \widetilde{\boldsymbol{\eta}}_i)$ посредством (5.2.2), но с $\boldsymbol{Q}$, заданной в (5.3.5). Тогда выводы утверждения 5.3 остаются в силе.

Здесь мы показываем, что $\mathrm{E}(\boldsymbol{F}_i'\boldsymbol{Q}\boldsymbol{\eta}_i) = \boldsymbol{0}$. (Другие части доказательства, являющиеся применением LLN Колмогорова и CLT Линдеберга — Леви, те же самые, что и в доказательстве утверждения 5.3.) Поскольку

$$\boldsymbol{F}_i'\boldsymbol{Q}\boldsymbol{\eta}_i = \sum_{m=1}^{M}\sum_{h=1}^{M} q_{mh}^{(i)} \cdot d_{im} \cdot d_{ih} \cdot \boldsymbol{f}_{im} \cdot \eta_{ih}, \tag{5.3.8}$$

где $q_{mh}^{(i)}$ — (m, h)-й элемент $\boldsymbol{Q}$, примененной к i-му уравнению, мы имеем:

$$\mathrm{E}(\boldsymbol{F}_i'\boldsymbol{Q}\boldsymbol{\eta}_i) = \sum_{m=1}^{M}\sum_{h=1}^{M} \mathrm{E}(q_{mh}^{(i)} \cdot d_{im} \cdot d_{ih} \cdot \boldsymbol{f}_{im} \cdot \eta_{ih}). \tag{5.3.9}$$

Если нет смещения из-за отбора, так что $\mathrm{E}(\boldsymbol{f}_{im} \cdot \eta_{ih} | \boldsymbol{d}_i) = \boldsymbol{0}$, то (5.3.9) равно нулю, так как

$$\begin{aligned} &\mathrm{E}(q_{mh}^{(i)} \cdot d_{im} \cdot d_{ih} \cdot \boldsymbol{f}_{im} \cdot \eta_{ih}) = \\ &= \mathrm{E}\left[q_{mh}^{(i)} \cdot d_{im} \cdot d_{ih} \cdot \mathrm{E}(\boldsymbol{f}_{im} \cdot \eta_{ih} | \boldsymbol{d}_i)\right] \\ &\text{(поскольку } q_{mh}^{(i)} \text{ является функцией от } \boldsymbol{d}_i\text{)}. \end{aligned} \tag{5.3.10}$$

Интересно, что обобщение на несбалансированную панель легче без условной гомоскедастичности. Чтобы (5.2.7) из утверждения 5.1 было асимптотической дисперсией оценки при условной гомоскедастичности, мы должны были бы предположить, что второй момент $\widetilde{\boldsymbol{\eta}}_i$, условный относительно $\boldsymbol{x}_i$ *и* $\boldsymbol{d}_i$, равен безусловному второму моменту, а состоятельная оценка этого второго момента была бы немного сложнее.

Контрольные вопросы

1. Для $\boldsymbol{d}_i = (1, 0, 1)'$ запишите $\boldsymbol{Q}$.
2. (Альтернативное получение оценки фиксированных эффектов для несбалансированных панелей.) Пусть $\boldsymbol{y}_i^*$ является M_i-мерным вектором, полученным из M-мерного вектора $\boldsymbol{y}_i$ путем удаления строк, для которых нет наблюдений. Точно так же создадим $\boldsymbol{F}_i^*$ $(M_i \times \#\beta)$ из $\boldsymbol{F}_i$ и $\boldsymbol{d}_i^*$ из $\boldsymbol{d}_i$ (так что $\boldsymbol{d}_i^*$ является M_i-мерным вектором из единиц). Пусть $\boldsymbol{Q}^*$ $(M_i \times M_i)$ является аннулятором для $\boldsymbol{d}_i^*$. Проверьте, что оценку фиксированных эффектов можно также записать как (5.2.4) с помощью этих матриц и векторов со звездочками.
3. В параграфе 4.6 описывается объединенная OLS-оценка в уровнях для системы (4.6.1′) на с. 320.

(a) Как бы вы модифицировали оценку (4.6.11′) на с. 322, чтобы охватить несбалансированные панели? **Указание:** *Определите* $\boldsymbol{y}_i$ *и* $\boldsymbol{Z}_i$ *(*$M \times L$*) через «обнуление».*

(b) Чему равна асимптотическая дисперсия при отсутствии условной гомоскедастичности?
[Ответ:$[\mathrm{E}(\boldsymbol{Z}_i'\boldsymbol{Z}_i)]^{-1}\,[\mathrm{E}(\boldsymbol{Z}_i'\varepsilon_i\varepsilon_i'\boldsymbol{Z}_i)]^{-1}\,[\mathrm{E}(\boldsymbol{Z}_i'\boldsymbol{Z}_i)]^{-1}$.]

(c) Как вы бы оценивали асимптотическую дисперсию? **Указание:** *Замените в утверждении 5.3* $\widetilde{\boldsymbol{F}}_i$ *на* $\boldsymbol{Z}_i$, *и* $\check{\boldsymbol{\eta}}_i$ *на* ε_i.

5.4. Приложение: Международные различия в темпах роста

Растут ли бедные экономики быстрее, чем богатые? Если это так, то насколько быстрее? Эти вопросы рассмотрены в быстро увеличивающейся в последнее время литературе по росту[1]. В этом параграфе выводится ключевое уравнение, которое было использовано для объяснения экономического роста в различных странах, и обсуждаются способы его оценивания. Фактическое оценивание нескольких спецификаций этого уравнения вынесено в эмпирическое упражнение.

Вывод оцениваемого уравнения

В неоклассической теории роста выпуск на эффективную единицу труда в период t, обозначаемый как $q(t)$ и определенный более точно в (5.4.4) ниже, сходится к стационарному уровню q^*. Логлинейная аппроксимация в окрестности стационарного состояния дает

$$\frac{\mathrm{d}\log(q(t))}{\mathrm{d}t} = \lambda \cdot [\log(q^*) - \log(q(t))]. \tag{5.4.1}$$

Оценивание λ, **скорости конвергенции** (*speed of convergence*), было целью в недавней литературе по росту. Из (5.4.1) следует, что для любых двух точек t_{m-1} и t_m на оси времени

$$\log(q(t_m)) = (1-\rho)\cdot\log(q^*) + \rho\cdot\log(q(t_{m-1})). \tag{5.4.2}$$

где

$$\rho \equiv \exp[-\lambda\cdot(t_m - t_{m-1})]. \tag{5.4.3}$$

(Поскольку в наших данных t_m равномерно распределены по времени, ρ не зависит от m.) Выпуск на единицу эффективного труда определяется как

$$q(t) \equiv \frac{Y(t)}{A(t)\,L(t)}. \tag{5.4.4}$$

[1] Весьма обширным обзором литературы является [Barro and Sala-i-Martin, 1995].

где $Y(t)$ равно агрегированному выпуску, $L(t)$ равно агрегированным отработанным часам и $A(t)$ равно уровню трудоинтенсивного технологического прогресса. Предполагая, что $A(t)$ растет с постоянным темпом g (так что $A(t) = A(0)\exp(gt)$), из (5.4.4) следует:

$$\log(q(t)) = \log\left(\frac{Y(t)}{L(t)}\right) - \log(A(0)) - g \cdot t. \tag{5.4.5}$$

Подставляя это в (5.4.2), мы получаем:

$$\log\left(\frac{Y(t_m)}{L(t_m)}\right) = \rho \cdot \log\left(\frac{Y(t_{m-1})}{L(t_{m-1})}\right) + (1-\rho) \cdot [\log(q^*) + \log(A(0))] + \phi_m, \tag{5.4.6}$$

где

$$\phi_m \equiv g \cdot (t_m - \rho \cdot t_{m-1}). \tag{5.4.7}$$

Эквивалентную форму уравнения (5.4.6) можно получить, вычитая из обеих его частей $\log \frac{Y(t_{m-1})}{L(t_{m-1})}$:

$$\log\left(\frac{Y(t_m)}{L(t_m)}\right) - \log\left(\frac{Y(t_{m-1})}{L(t_{m-1})}\right) =$$
$$= (\rho - 1)\log\left(\frac{Y(t_{m-1})}{L(t_{m-1})}\right) + (1-\rho) \cdot [\log(q^*) + \log(A(0))] + \phi_m. \tag{5.4.6$'$}$$

Поскольку $\rho < 1$ при $\lambda > 0$, это уравнение говорит, что уровень выпуска на душу населения имеет отрицательное влияние на последующий рост, бедные страны должны расти быстрее. Эта форма уравнения интуитивно более понятна, но для целей выбора корректного метода оценивания более поучительным является (5.4.6), а не (5.4.6′).

Уравнение (5.4.6) должно выполняться для каждой страны i[1]. В оставшейся части этого параграфа мы предполагаем, что скорость конвергенции (λ) и темп роста трудоинтенсивного технологического прогресса (g) одинаковы для разных стран. Тогда (5.4.6) для страны i можно записать как

$$y_{im} = \phi_m + \rho y_{i,m-1} + \alpha_i. \tag{5.4.8}$$

где

$$\begin{aligned} y_{im} &= \log\left(\frac{Y(t_m)}{L(t_m)}\right) \quad \text{для страны } i, \\ \alpha_i &= (1-\rho) \times \{\log(q^*) + \log(A(0))\} \quad \text{для страны } i. \end{aligned} \tag{5.4.9}$$

[1] Мы получили уравнение, описывающее конвергенцию, предполагая закрытую экономику. Это предположение сложно оправдать, но можно получить аналогичное уравнение при свободной мобильности физического капитала, если человеческий капитал является другим фактором производства и не является свободно мобильным на международном уровне. См., например: [Barro and Sala-i-Martin, 1995].

Добавление ошибок

В прикладной эконометрике довольно стандартной является практика, когда на основе теории выводится уравнение без включения ошибки, а ошибка добавляется затем для оценивания. Эмпирика по росту не является исключением; ошибка η_{im} добавляется в (5.4.8), чтобы получить

$$y_{im} = \phi_m + \rho y_{i,m-1} + \alpha_i + \eta_{im}. \qquad (5.4.10)$$

Одной из естественных интерпретаций η_{im} являются бизнес-циклы, которые в макроэкономике определяются как разность между фактическим выпуском и потенциальным выпуском. Теория роста говорит о потенциальном выпуске. Несоответствие между потенциальным и фактическим выпусками становится ошибкой в оцениваемом уравнении, выведенном из теории роста. Но тогда существовала бы проблема ошибок в переменных, которая делает регрессоры эндогенными[1]. В оставшейся части параграфа мы следуем мейнстриму в литературе по росту и игнорируем проблему эндогенности.

Другой проблемой является возможная корреляция в η_{im} между странами (i). Эта пространственная (или географическая) корреляция не возникала бы, если бы выборка была случайной. Но выборка стран не является случайной; она представляет генеральную совокупность. Феномен, известный как международные бизнес-циклы, предполагает, что пространственная корреляция должна быть положительной. Статистические выводы, рассматривающие страны как если бы они были независимыми точками данных, завышают статистическую значимость. Эта проблема также по большей части игнорируется в литературе и будет игнорироваться и в нашем обсуждении.

Трактовка α_i

Как ясно из (5.4.9), компонента α_i в уравнении (5.4.10) зависит от стационарного уровня q^* и исходного уровня технологии $A(0)$. Как трактовать α_i в оценивании, зависит от того, какой версии неоклассической теории роста вы придерживаетесь. Согласно модели оптимального роста Касса — Купманса, единственным фактором, определяющим стационарный уровень q^*, является темп роста труда. Таким образом, если технологическое знание свободно перетекает на международном уровне, так что $A(0)$ является одинаковым для разных стран, то международные отличия в α_i

[1]Пусть теория роста пытается объяснить x_{im} — логарифм потенциального выпуска в год t_m для страны i, так что (5.4.8) выполняется для x_{im}, а не для y_{im}. Пусть логарифм фактического выпуска y_{im} соотносится с x_{im} как $y_{im} = x_{im} + \eta_{im}$. Тогда получается (5.4.10), но с ошибками $\eta_{im} - \rho\eta_{i,m-1}$. Так что проблему могут создавать не только ошибки в переменных, но также и то, что ошибки имеют структуру скользящего среднего.

полностью объясняются темпом роста труда. Следовательно, оставляя в стороне проблему эндогенности, упомянутую только что, OLS-оценивание (5.4.10) обеспечивает состоятельность, если темпы роста труда входят в качестве дополнительного регрессора для контроля α_i. Если вы придерживаетесь теории роста Солоу — Свана, норма сбережения является другим фактором, определяющим q^*, поэтому регрессия должна включать норму сбережения наряду с темпом роста труда.

Большая доля современной литературы по росту под рубрикой «условная конвергенция» включает эти и другие переменные, которые могут влиять на q^* или $A(0)$ — такие как мера политической стабильности и степень финансового посредничества — чтобы контролировать α_i. В эмпирическом упражнении вас попросят оценить ρ и, следовательно, скорость конвергенции λ, используя подход условной конвергенции.

Однако, независимо от того, насколько изобретательно выбраны переменные для контроля α_i, некоторая компонента в α_i, имеющая дело с особенностями, уникальными для страны, останется необъясненной и попадет в ошибку. Будучи частью α_i, она коррелирована с регрессором $y_{i,m-1}$ в (5.4.10). OLS-оценивание, даже если уравнение включает целый ряд переменных для контроля α_i, не обеспечит состоятельность. Альтернативный подход состоит в том, чтобы трактовать α_i как ненаблюдаемый фиксированный эффект.

Состоятельное оценивание скорости конвергенции

Если выпуск и другие переменные наблюдаются в $M+1$ моментах времени, $t_0, t_1, t_2, \ldots, t_M$, то уравнение (5.4.10) доступно для $m = 1, \ldots, M$. Метод фиксированных эффектов можно применить к этой системе из M уравнений, но он не дает состоятельности. Причина состоит в том, что эта система является «динамической», в том смысле, что некоторые из регрессоров являются зависимыми переменными для других уравнений системы. Чтобы увидеть это, посмотрим на первые два уравнения системы:

$$\begin{aligned} y_{i1} &= \phi_1 + \rho y_{i0} + \alpha_i + \eta_{i1}, \\ y_{i2} &= \phi_2 + \rho y_{i1} + \alpha_i + \eta_{i2}. \end{aligned}$$

Предположим, что ошибка η_{i1} ортогональна регрессорам (которые суть константа и y_{i0}) в первом уравнении, так что $\mathrm{E}(\eta_{i1}) = 0$ и $\mathrm{E}(y_{i0}\eta_{i1}) = 0$. Если мы сделаем еще предположение, что $\mathrm{E}(\alpha_i\eta_{i1}) = 0$, то из первого уравнения

$$\mathrm{E}(y_{i1}\,\eta_{i1}) = \mathrm{Var}(\eta_{i1}) \neq 0.$$

Но y_{i1} является одним из регрессоров во втором уравнении. Следовательно, если ошибка η_{im} ортогональна регрессорам для каждого уравнения, «кросс»-ортогональности обязательно нарушатся[1]. Как подчеркнуто в параграфе 5.2 (см. контрольный вопрос 3), если «кросс»-ортогональности не выполняются, оценка фиксированных эффектов не обязательно будет состоятельной. Относительно точного выражения для смещения см. аналитическое упражнение 4 в этой главе.

Один из способов достижения состоятельности таков. Возьмем первые разности M уравнений, получая $M-1$ уравнений:

$$\begin{aligned} y_{i2} - y_{i1} &= \mu_1 + (y_{i1} - y_{i0})\rho + (\eta_{i2} - \eta_{i1}), \\ y_{i3} - y_{i2} &= \mu_2 + (y_{i2} - y_{i1})\rho + (\eta_{i3} - \eta_{i2}), \\ &\dots \\ y_{iM} - y_{i,M-1} &= \mu_{M-1} + (y_{i,M-1} - y_{i,M-2})\rho + (\eta_{iM} - \eta_{i,M-1}), \end{aligned} \tag{5.4.11}$$

где

$$\mu_m \equiv \phi_{m+1} - \phi_m.$$

Оценка случайных эффектов и OLS для каждого уравнения не дают состоятельности, поскольку регрессоры не являются ортогональными ошибкам. Например, в первом уравнении разность $y_{i1} - y_{i0}$ неортогональна $\eta_{i2} - \eta_{i1}$, поскольку $\mathrm{E}(y_{i1} \cdot \eta_{i1}) \neq 0$, как только что было видно. Состоятельное оценивание можно осуществить посредством многомерного GMM, если есть релевантные инструменты. В дополнительном эмпирическом упражнении к этой главе вас попросят использовать норму сбережения в качестве инструмента для эндогенных регрессоров. При условии, что норма сбережений не коррелирует с η_{im}, эта оценка является робастной к проблеме эндогенности, упомянутой выше.

Контрольные вопросы

1. Предположение, что g (темп роста технологического прогресса) является общим для всех стран, необходимо для получения (5.4.8). Должны ли мы предполагать, что g является постоянным во времени? [Ответ: Нет.]

2. (Идентификация.) Для простоты пусть $M = 3$, так что (5.4.11) является системой из двух уравнений.

   (a) Запишите (5.4.11) как систему из нескольких уравнений с общими коэффициентами, надлежащим образом определяя y_{i1}, y_{i2}, $\boldsymbol{z}_{i1}$, $\boldsymbol{z}_{i2}$, $\boldsymbol{\delta}$ в (4.6.1). **Указание:** *$\boldsymbol{z}_{i2}$, например, равно* $(0, 1, y_{i2} - y_{i1})'$.

   (b) Если s_{im} является инструментом для $y_{im} - y_{i,m-1}$, вектором инструментов для m-го уравнения является $\boldsymbol{x}_{im} = (1, s_{im})'$ для $m = 1, 2$. Покажите, что система неидентифицируема, если $\mathrm{Cov}(s_{im}, y_{im} - y_{i,m-1}) = 0$

[1] Это было впервые отмечено в: [Nickell, 1981].

для всех m. **Указание:**

$$\Sigma = \begin{bmatrix} 1 & 0 & \mathrm{E}(y_{i1} - y_{i0}) \\ \mathrm{E}(s_{i1}) & 0 & \mathrm{E}[s_{i1}(y_{i1} - y_{i0})] \\ 0 & 1 & \mathrm{E}(y_{i2} - y_{i1}) \\ 0 & \mathrm{E}(s_{i2}) & \mathrm{E}[s_{i2}(y_{i2} - y_{i1})] \end{bmatrix}.$$

Приложение 5.А. Распределение статистики Хаусмана

В этом приложении доказывается, что $\mathrm{Avar}(\widehat{q})$ в (5.2.21) является положительно определенной матрицей и статистика Хаусмана (5.2.22) гарантированно неотрицательна в любых конечных выборках.

Как показано в параграфе 4.6,

$$\widehat{\delta}_{\mathrm{RE}} = (S'_{xz}\widehat{S}^{-1}S_{xz})^{-1}S'_{xz}\widehat{S}^{-1}s_{xy}, \tag{5.A.1}$$

где

$$\underset{(MK\times L)}{S_{xz}} \equiv \frac{1}{n}\sum_{i=1}^{n} Z_i \otimes x_i, \qquad \underset{(MK\times MK)}{\widehat{S}} \equiv \widehat{\Sigma}\otimes S_{xx},$$

$$\underset{(K\times K)}{S_{xx}} \equiv \frac{1}{n}\sum_{i=1}^{n} x_i x_i', \qquad s_{xy} \equiv \frac{1}{n}\sum_{i=1}^{n} y_i \otimes x_i,$$

$K \equiv \#x_i$, $L \equiv \#z_{im}$ (так что Z_i имеет размер $M \times L$).

Пусть

$$Z_i = (F_i \,\vdots\, 1b_i'),\ p \equiv \#\beta,\ \underset{(L\times p)}{D} = \begin{bmatrix} I_p \\ 0 \end{bmatrix},$$

так что $F_i = Z_i D$. Тогда можно довольно легко показать, что

$$\widehat{\beta}_{\mathrm{FE}} = (D'S'_{xz}\widehat{W}S_{xz}D)^{-1}D'S'_{xz}\widehat{W}s_{xy}. \tag{5.A.2}$$

где

$$\widehat{W} = Q \otimes S_{xx}^{-1},\ Q \equiv \text{аннулятор для } 1_M.$$

Из (5.А.1) и (5.А.2) следует, что

$$\widehat{q} = (D'S'_{xz}\widehat{W}S_{xz}D)^{-1}D'S'_{xz}\widehat{W}g_n(\widehat{\delta}_{\mathrm{RE}}),\ g_n(\widehat{\delta}_{\mathrm{RE}}) = s_{xy} - S_{xz}\widehat{\delta}_{\mathrm{RE}}. \tag{5.A.3}$$

(Чтобы вывести это, вспомним из параграфа 4.6, что

$$S'_{xz}\widehat{W}^{-1}S_{xz} = \frac{1}{n}\sum_{i=1}^{n} Z_i' Q Z_i.$$

Но поскольку $QZ_i = (QF_i \,\vdots\, 0)$, это равно

$$\begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{F}_i'\boldsymbol{Q}\boldsymbol{F}_i & 0 \\ \mathbf{0} & 0 \end{bmatrix}, \tag{5.A.4}$$

так что $\boldsymbol{S}_{\boldsymbol{xz}}'\widehat{\boldsymbol{W}}\boldsymbol{S}_{\boldsymbol{xz}} = \boldsymbol{D}\boldsymbol{D}'\boldsymbol{S}_{\boldsymbol{xz}}'\widehat{\boldsymbol{W}}\boldsymbol{S}_{\boldsymbol{xz}} = \boldsymbol{S}_{\boldsymbol{xz}}'\widehat{\boldsymbol{W}}\boldsymbol{S}_{\boldsymbol{xz}}\boldsymbol{D}\boldsymbol{D}'$.)

Как показано в аналитическом упражнении 5 в главе 3,

$$g_n(\widehat{\delta}_{\mathrm{RE}}) = \widehat{\boldsymbol{B}}\overline{\boldsymbol{g}},\ \widehat{\boldsymbol{B}} = \boldsymbol{I}_{MK} - \boldsymbol{S}_{\boldsymbol{xz}}(\boldsymbol{S}_{\boldsymbol{xz}}'\widehat{\boldsymbol{S}}^{-1}\boldsymbol{S}_{\boldsymbol{xz}})^{-1}\boldsymbol{S}_{\boldsymbol{xz}}'\widehat{\boldsymbol{S}}^{-1}. \tag{5.A.5}$$

Подставляя это в (5.A.3), мы получаем

$$\sqrt{n}\widehat{\boldsymbol{q}} = (\boldsymbol{D}'\boldsymbol{S}_{\boldsymbol{xz}}'\widehat{\boldsymbol{W}}\boldsymbol{S}_{\boldsymbol{xz}}\boldsymbol{D})^{-1}\boldsymbol{D}'\boldsymbol{S}_{\boldsymbol{xz}}'\widehat{\boldsymbol{W}}\widehat{\boldsymbol{B}}\sqrt{n}\overline{\boldsymbol{g}}. \tag{5.A.6}$$

Асимптотическая дисперсия этого выражения равна:

$$\begin{aligned}\mathrm{Avar}(\widehat{\boldsymbol{q}}) = (\boldsymbol{D}'\boldsymbol{\Sigma}_{\boldsymbol{xz}}'\boldsymbol{W}\boldsymbol{\Sigma}_{\boldsymbol{xz}}\boldsymbol{D})^{-1}\times \\ \times \boldsymbol{D}'\boldsymbol{\Sigma}_{\boldsymbol{xz}}'\boldsymbol{W}\boldsymbol{B}\boldsymbol{S}\boldsymbol{B}'\boldsymbol{W}\boldsymbol{\Sigma}_{\boldsymbol{xz}}\boldsymbol{D}(\boldsymbol{D}'\boldsymbol{\Sigma}_{\boldsymbol{xz}}'\boldsymbol{W}\boldsymbol{\Sigma}_{\boldsymbol{xz}}\boldsymbol{D})^{-1},\end{aligned} \tag{5.A.7}$$

где $\boldsymbol{B} = \mathrm{plim}\,\widehat{\boldsymbol{B}}$, $\boldsymbol{W} = \mathrm{plim}\,\widehat{\boldsymbol{W}} = \boldsymbol{Q}\otimes\boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1}$ и $\boldsymbol{S} \equiv \mathrm{Avar}(\widehat{\boldsymbol{g}})$. Здесь $\boldsymbol{D}'\boldsymbol{\Sigma}_{\boldsymbol{xz}}'\boldsymbol{W}\boldsymbol{\Sigma}_{\boldsymbol{xz}}\boldsymbol{D}$ обратима, поскольку легко показать, что

$$\boldsymbol{D}'\boldsymbol{\Sigma}_{\boldsymbol{xz}}'\boldsymbol{W}\boldsymbol{\Sigma}_{\boldsymbol{xz}}\boldsymbol{D} = \mathrm{E}(\boldsymbol{F}_i'\boldsymbol{Q}\boldsymbol{F}_i), \tag{5.A.8}$$

которая, как доказано в аналитическом упражнении 5, невырождена. (5.A.7) состоятельно оценивается посредством

$$\begin{aligned}\widehat{\mathrm{Avar}(\widehat{\boldsymbol{q}})} = (\boldsymbol{D}'\boldsymbol{\Sigma}_{\boldsymbol{xz}}'\widehat{\boldsymbol{W}}\boldsymbol{\Sigma}_{\boldsymbol{xz}}\boldsymbol{D})^{-1}\times \\ \times \boldsymbol{D}'\boldsymbol{\Sigma}_{\boldsymbol{xz}}'\widehat{\boldsymbol{W}}\widehat{\boldsymbol{B}}\widehat{\boldsymbol{S}}\widehat{\boldsymbol{B}}'\widehat{\boldsymbol{W}}\boldsymbol{\Sigma}_{\boldsymbol{xz}}\boldsymbol{D}(\boldsymbol{D}'\boldsymbol{\Sigma}_{\boldsymbol{xz}}'\widehat{\boldsymbol{W}}\boldsymbol{\Sigma}_{\boldsymbol{xz}}\boldsymbol{D})^{-1}.\end{aligned} \tag{5.A.9}$$

Простые, но длительные матричные вычисления показывают, что (5.A.7) равно $\mathrm{Avar}(\widehat{\beta}_{\mathrm{FE}}) - \boldsymbol{D}'\,\mathrm{Avar}(\widehat{\delta}_{\mathrm{RE}})\boldsymbol{D}$, а (5.A.9) — $\widehat{\mathrm{Avar}(\widehat{\beta}_{\mathrm{FE}})} - \boldsymbol{D}'\widehat{\mathrm{Avar}(\widehat{\delta}_{\mathrm{RE}})}\boldsymbol{D}$. В любой конечной выборке (5.A.9) положительно полуопределена. Поэтому статистика Хаусмана гарантированно неотрицательна.

Теперь мы покажем, что матрица (5.A.7) невырождена. В соответствии с (5.1.6), $\boldsymbol{S}$ положительно определена. Следовательно, ранг $\mathrm{Avar}(\widehat{\boldsymbol{q}})$ равен рангу

$$\boldsymbol{B}'(\boldsymbol{Q}\otimes\boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1})\boldsymbol{\Sigma}_{\boldsymbol{xz}}\boldsymbol{D}. \tag{5.A.10}$$

Поскольку столбцы $\boldsymbol{S}^{-1}\boldsymbol{\Sigma}_{\boldsymbol{xz}}$ формируют базис для столбцового нуль-пространства $\boldsymbol{B}$ и поскольку $(\boldsymbol{Q}\otimes\boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1})\boldsymbol{\Sigma}_{\boldsymbol{xz}}\boldsymbol{D}$ имеет полный столбцовый ранг по (5.1.15), по лемме А.5 в [Newey, 1985][1], ранг (5.A.10) равен:

$$\mathrm{rank}[(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1})\boldsymbol{\Sigma}_{\boldsymbol{xz}} \,\vdots\, (\boldsymbol{Q}\otimes\boldsymbol{\Sigma}_{\boldsymbol{xx}}^{-1})\boldsymbol{\Sigma}_{\boldsymbol{xz}}\boldsymbol{D}] - L, \tag{5.A.11}$$

где был использован тот факт, что $\boldsymbol{S} = \boldsymbol{\Sigma}\otimes\boldsymbol{\Sigma}_{\boldsymbol{xx}}$.

[1] Эта лемма утверждает следующее: Пусть $\boldsymbol{A}$ — $k\times l$ матрица, $\boldsymbol{B}$ — $l\times m$ матрица и $\boldsymbol{C}$ — $l\times n$ матрица. Если столбцы в $\boldsymbol{C}$ формируют базис столбцового нуль-пространства матрицы $\boldsymbol{A}$ и $\mathrm{rank}(\boldsymbol{B}) = m$, то $\mathrm{rank}(\boldsymbol{AB}) = \mathrm{rank}(\boldsymbol{C}\vdots\boldsymbol{B}) - n$. Здесь $k = l = MK$, $m = p$ и $n = L$.

Далее, $\boldsymbol{x}_i$ содержит в себе все уникальные переменные в $\boldsymbol{Z}_i$. Сделаем перестановку в $\boldsymbol{x}_i$:

$$\boldsymbol{x}_i = (\boldsymbol{f}'_{i1}, \ldots, \boldsymbol{f}'_{iM}, \boldsymbol{b}'_i)'. \tag{5.A.12}$$

Тогда

$$\boldsymbol{z}'_{im} = \boldsymbol{x}'_i \boldsymbol{H}_m, \ \boldsymbol{H}_m \equiv \begin{bmatrix} \boldsymbol{e}_m \otimes \boldsymbol{I}_p & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{L-p} \end{bmatrix}, \ \boldsymbol{e}_m = m\text{-й столбец в } \boldsymbol{I}_M, \tag{5.A.13}$$

так что $\boldsymbol{\Sigma}_{\boldsymbol{xz}}$ можно записать как

$$\boldsymbol{\Sigma}_{\boldsymbol{xz}} = (\boldsymbol{I}_m \otimes \boldsymbol{\Sigma}_{\boldsymbol{xx}})\boldsymbol{H}, \tag{5.A.14}$$

где $\boldsymbol{H}' = (\boldsymbol{H}'_1, \ldots, \boldsymbol{H}'_M)$. Подставляя (5.A.14) в (5.A.11), находим, что ранг $\mathrm{Avar}(\widehat{\boldsymbol{q}})$ равен

$$\mathrm{rank}[(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_K)\boldsymbol{H} \ \vdots \ (\boldsymbol{Q} \otimes \boldsymbol{I}_K)\boldsymbol{H}\boldsymbol{D}] - L.$$

m-е K строк $(MK \times (L+p))$ матрицы $[(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_K)\boldsymbol{H} \ \vdots \ (\boldsymbol{Q} \otimes \boldsymbol{I}_K)\boldsymbol{H}\boldsymbol{D}]$ имеют вид:

$$\begin{bmatrix} \begin{pmatrix} \sigma^{m1} \\ \vdots \\ \sigma^{mM} \end{pmatrix} \otimes \boldsymbol{I}_p & \mathbf{0} & \begin{pmatrix} q_{m1} \\ \vdots \\ q_{mM} \end{pmatrix} \otimes \boldsymbol{I}_p \\ \mathbf{0} & \left(\sum_{h=1}^{M} \sigma^{mh}\right) \boldsymbol{I}_{L-p} & \mathbf{0} \end{bmatrix}. \tag{5.A.15}$$

Следовательно, $\mathrm{rank}[(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_K)\boldsymbol{H} \ \vdots \ (\boldsymbol{Q} \otimes \boldsymbol{I}_K)\boldsymbol{H}\boldsymbol{D}] = L+p$, если $\mathrm{vec}(\boldsymbol{\Sigma}^{-1})$ и $\mathrm{vec}(\boldsymbol{Q})$ не являются линейно зависимыми; то есть если $\boldsymbol{\Sigma}^{-1}$ не пропорциональна $\boldsymbol{Q}$. Но поскольку $\mathrm{rank}(\boldsymbol{\Sigma}^{-1}) = M \neq \mathrm{rank}(\boldsymbol{Q}) = M - 1$, $\boldsymbol{\Sigma}^{-1}$ не может быть пропорциональна $\boldsymbol{Q}$.

Набор задач для главы 5

Аналитические упражнения

1. (Оценка фиксированных эффектов как LSDV-оценка.) В (5.1.1″) на с. 358 переопределим α_i как сумму старого α_i и $\boldsymbol{b}'_i\boldsymbol{\gamma}$, так что (5.1.1″) принимает вид:

$$\boldsymbol{y}_i = \boldsymbol{F}_i\boldsymbol{\beta} + \mathbf{1}_M \cdot \alpha_i + \boldsymbol{\eta}_i \quad (i = 1, 2, \ldots, n). \tag{1}$$

Определим следующие штабелированные векторы и матрицы:

$$\underset{(Mn\times 1)}{\boldsymbol{y}} = \begin{bmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_n \end{bmatrix}, \quad \underset{(Mn\times \#\boldsymbol{\beta})}{\boldsymbol{F}} = \begin{bmatrix} \boldsymbol{F}_1 \\ \vdots \\ \boldsymbol{F}_n \end{bmatrix}, \quad \underset{(Mn\times 1)}{\boldsymbol{\eta}} = \begin{bmatrix} \boldsymbol{\eta}_1 \\ \vdots \\ \boldsymbol{\eta}_n \end{bmatrix}, \quad \underset{(n\times 1)}{\boldsymbol{\alpha}} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix},$$

$$\underset{(Mn\times(n+\#\beta))}{W} = (D \vdots F), \quad \underset{((n+\#\beta)\times 1)}{\theta} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}.$$

$$\underset{(Mn\times n)}{D} = I_n \otimes 1_M = \begin{bmatrix} 1_M & & & \\ & 1_M & & \\ & & \ddots & \\ & & & 1_M \end{bmatrix}, \qquad (2$$

так что систему (1) из M уравнений для $i = 1, 2, \ldots, n$ можно записать как

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1_M \cdot \alpha_1 \\ 1_M \cdot \alpha_2 \\ \vdots \\ 1_M \cdot \alpha_n \end{bmatrix} + \begin{bmatrix} F_1\beta \\ F_2\beta \\ \vdots \\ F_n\beta \end{bmatrix} + \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{bmatrix},$$

или

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1_M & & & \\ & 1_M & & \\ & & \ddots & \\ & & & 1_M \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} + \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_n \end{bmatrix} \beta + \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_n \end{bmatrix},$$

или

$$y = D\alpha + F\beta + \eta = W\theta + \eta. \qquad (3)$$

OLS-оценка для θ в объединенной выборке объема Mn, (y, W), равна $(W'W)^{-1}W'y$.

(a) Покажите, что оценка фиксированных эффектов, заданная в (5.2.4), представляет последние $\#\beta$ элементов этого вектора. **Указание:** *Используйте теорему Фриша — Во о регрессии с блочной структурой данных (см. аналитическое упражнение 4(b) в главе 1). На первом шаге проводится регрессия y на D и проводится регрессия каждого столбца F на D. Если мы определим аннулятор для D как*

$$\underset{(Mn\times Mn)}{M_D} \equiv I_{Mn} - D(D'D)^{-1}D', \qquad (4)$$

то векторы остатков можно записать как $M_D y$ и $M_D F$. На втором шаге проводится регрессия $M_D y$ на $M_D F$. Покажите, что оцененный регрессионный коэффициент можно записать как

$$(F'M_D F)^{-1}F'M_D y. \qquad (5)$$

Покажите, что $M_D = I_n \otimes Q$ (где Q — аннулятор для 1_M, как в тексте), и используйте этот факт, чтобы показать, что (5) действительно является оценкой фиксированных эффектов.

Альтернативно вы можете пройти весь путь обратно к нормальным уравнениям. Если $\boldsymbol{a}$ и $\boldsymbol{b}$ являются OLS-оценками $\boldsymbol{\alpha}$ и $\boldsymbol{\beta}$, то $(\boldsymbol{a}, \boldsymbol{b})$ удовлетворяет системе из $n + \#\boldsymbol{\beta}$ линейных одновременных уравнений:

$$\begin{bmatrix} \boldsymbol{D}'\boldsymbol{D} & \boldsymbol{D}'\boldsymbol{F} \\ \boldsymbol{F}'\boldsymbol{D} & \boldsymbol{F}'\boldsymbol{F} \end{bmatrix} \begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{bmatrix} = \begin{bmatrix} \boldsymbol{D}'\boldsymbol{y} \\ \boldsymbol{F}'\boldsymbol{y} \end{bmatrix}. \tag{6}$$

Возьмите первые n уравнений и решите их относительно $\boldsymbol{a}$, считая $\boldsymbol{b}$ заданным. Это должно привести к $\boldsymbol{a} = (\boldsymbol{D}'\boldsymbol{D})^{-1}(\boldsymbol{D}'\boldsymbol{y} - \boldsymbol{D}'\boldsymbol{F}\boldsymbol{b})$. Затем подставьте это выражение в оставшиеся уравнения и решите их относительно $\boldsymbol{b}$. (Вы можете выполнить то же самое, используя формулу для обращения блочных матриц.)

(b) Покажите:

$$a_i = \overline{y}_i - \left(\frac{1}{M}\sum_{m=1}^{M} \boldsymbol{f}'_{im}\right)\widehat{\boldsymbol{\beta}}_{\text{FE}},$$

где a_i является OLS-оценкой для α_i, $\boldsymbol{f}'_{im}$ является m-й строкой в $\boldsymbol{F}_i$ и $\overline{y}_i = \frac{1}{M}(y_{i1} + \cdots + y_{iM})$. **Указание:** *Используйте нормальные уравнения для $\boldsymbol{\alpha}$.*

(c) Предположим:

(i) $\{\boldsymbol{y}_i, \boldsymbol{F}_i\}$ является i.i.d.,

(ii) $\mathrm{E}(\boldsymbol{\eta}_i|\boldsymbol{F}_i) = \boldsymbol{0}$ (что является более сильным, чем условия ортогональности (5.1.8b)),

(iii) $\mathrm{E}(\boldsymbol{\eta}_i\boldsymbol{\eta}'_i|\boldsymbol{F}_i) = \sigma_\eta^2 \boldsymbol{I}_M$ (предположение о сферичности ошибок) и

(iv) $\boldsymbol{W}$ имеет полный столбцовый ранг.

Покажите, что предположения классической модели регрессии выполняются для объединенной регрессии (3). **Указание:** *Вы должны показать, что $\boldsymbol{W}$ является строго экзогенной в (3), то есть*

$$\mathrm{E}(\boldsymbol{\eta}|\boldsymbol{W}) = \underset{Mn\times 1}{\boldsymbol{0}}.$$

Также покажите, что $\mathrm{E}(\boldsymbol{\eta}\boldsymbol{\eta}'|\boldsymbol{W}) = \sigma_\eta^2 \boldsymbol{I}_{Mn}$.

Разработайте для $\boldsymbol{a}$ и $\boldsymbol{b}$ теорию *конечных выборок.* Например, являются ли они несмещенными? Пусть SSR является суммой квадратов остатков в объединенной регрессии (3). Является ли эта SSR той же самой, что и SSR в регрессии «внутри» (5.2.15)?

2. (GMM-интерпретация FE.) Хорошо известно, что для случая двух уравнений ($M = 2$) оценка фиксированных эффектов является OLS-оценкой в регрессии первых разностей $y_{i2} - y_{i1}$ на $\boldsymbol{z}_{i2} - \boldsymbol{z}_{i1}$. Этот вопрос обобщает доказательство на случай произвольного M.

Мы хотим записать оценку фиксированных эффектов (5.2.4) в GMM-формате:

$$(\boldsymbol{S}'_{\boldsymbol{xz}}\widehat{\boldsymbol{W}}\boldsymbol{S}_{\boldsymbol{xz}})^{-1}\boldsymbol{S}'_{\boldsymbol{xz}}\widehat{\boldsymbol{W}}\boldsymbol{s}_{\boldsymbol{xy}} \tag{7}$$

для надлежащим образом определенных $\boldsymbol{S}_{\boldsymbol{xz}}$, $\widehat{\boldsymbol{W}}$ и $\boldsymbol{s}_{\boldsymbol{xy}}$. Пусть $\boldsymbol{C}$ — такая матрица, что

$\boldsymbol{C}$ является $M \times (M-1)$ матрицей полного столбцового ранга, удовлетворяющей $\boldsymbol{C}'\mathbf{1}_M = \mathbf{0}$. (8)

Известным примером такой матрицы $\boldsymbol{C}$ является матрица первых разностей

$$\underset{M\times(M-1)}{\boldsymbol{C}} = \begin{bmatrix} -1 & 0 & 0 & 0 & \dots & 0 \\ 1 & -1 & 0 & 0 & \dots & 0 \\ 0 & 1 & -1 & 0 & \dots & 0 \\ & & \ddots & \ddots & & \\ 0 & \dots & 0 & 1 & -1 & 0 \\ 0 & \dots & 0 & 0 & 1 & -1 \\ 0 & \dots & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Другим примером является $M \times (M-1)$ матрица, построенная удалением одного столбца, скажем, последнего, из матрицы $\boldsymbol{Q}$, аннулятора для $\mathbf{1}_M$.

(a) Проверьте, что эти две матрицы удовлетворяют (8).

(b) Для любого заданного выбора $\boldsymbol{C}$ проверьте, что вы можете получить следующую систему $M-1$ преобразованных уравнений, умножая обе части системы M уравнений (5.1.1″) (на с. 358) на $\boldsymbol{C}'$ слева:

$$\widehat{\boldsymbol{y}}_i = \widehat{\boldsymbol{F}}_i\boldsymbol{\beta} + \widehat{\boldsymbol{\eta}}_i, \tag{9}$$

где

$$\underset{((M-1)\times 1)}{\widehat{\boldsymbol{y}}_i} = \boldsymbol{C}'\boldsymbol{y}_i, \quad \underset{((M-1)\times\#\boldsymbol{\beta})}{\widehat{\boldsymbol{F}}_i} = \boldsymbol{C}'\boldsymbol{F}_i, \quad \underset{((M-1)\times 1)}{\widehat{\boldsymbol{\eta}}_i} = \boldsymbol{C}'\boldsymbol{\eta}_i.$$

(c) (Дополнительно.) Проверьте следующее:

Если $\{\boldsymbol{y}_i, \boldsymbol{Z}_i\}$ удовлетворяет предположениям модели компонент ошибки (которыми являются (5.1.1″), (5.1.2), (5.1.8), (5.1.4), (5.1.5), (5.1.15) и (5.1.6) или (5.1.6′) на с. 353), то $\{\widehat{\boldsymbol{y}}_i, \widehat{\boldsymbol{F}}_i\}$ удовлетворяет предположениям модели случайных эффектов, то есть для системы (9) из $M-1$ уравнений,

(случайная выборка) $\{\widehat{\boldsymbol{y}}_i, \widehat{\boldsymbol{F}}_i\}$ является i.i.d.,

(условия ортогональности) $\mathrm{E}(\widehat{\boldsymbol{\eta}}_i \otimes \boldsymbol{x}_i) = \mathbf{0}$ с $\boldsymbol{x}_i$ = объединение $(\boldsymbol{z}_{i1}, \dots, \boldsymbol{z}_{iM})$ = уникальные (различные) элементы $(\boldsymbol{F}_i, \boldsymbol{b}_i)$, (идентификация) $\mathrm{E}(\widehat{\boldsymbol{F}}_i \otimes \boldsymbol{x}_i)$ имеет полный столбцовый ранг,

(условная гомоскедастичность) $\mathrm{E}(\widehat{\boldsymbol{\eta}}_i\widehat{\boldsymbol{\eta}}_i'|\boldsymbol{x}_i) = \mathrm{E}(\widehat{\boldsymbol{\eta}}_i\widehat{\boldsymbol{\eta}}_i')$, последняя матрица невырождена,

(невырожденность $\mathrm{E}(\widehat{\boldsymbol{g}}_i\widehat{\boldsymbol{g}}_i')$) $\mathrm{E}(\widehat{\boldsymbol{g}}_i\widehat{\boldsymbol{g}}_i')$ невырождена, $\widehat{\boldsymbol{g}}_i \equiv \widehat{\boldsymbol{\eta}}_i \otimes \boldsymbol{x}_i$.

Указание: *Для условий ортогональности заметим, что* (5.1.8b) *можно записать как* $\mathrm{E}(\boldsymbol{\eta}_i \otimes \boldsymbol{x}_i) = \mathbf{0}$ *и что* $\widehat{\boldsymbol{\eta}}_i \otimes \boldsymbol{x}_i = (\boldsymbol{C}' \otimes \boldsymbol{I}_K)(\boldsymbol{\eta}_i \otimes \otimes \boldsymbol{x}_i)$. *Условие идентификации эквивалентно* (5.1.15) *(относительно доказательства см. ответы на вопросы). Чтобы показать, что* $\mathrm{E}(\widehat{\boldsymbol{\eta}}_i\widehat{\boldsymbol{\eta}}_i'|\boldsymbol{x}_i)$ *не зависит от* $\boldsymbol{x}_i$, *используйте тот факт, что* $\widehat{\boldsymbol{\eta}}_i = \boldsymbol{C}'\boldsymbol{\eta}_i = \boldsymbol{C}'\boldsymbol{\varepsilon}_i$, *поскольку* $\boldsymbol{\varepsilon}_i = \mathbf{1}_M \cdot \alpha_i + \boldsymbol{\eta}_i$. *При заданной условной гомоскедастичности мы имеем*

$$\mathrm{E}(\widehat{\boldsymbol{g}}_i\widehat{\boldsymbol{g}}_i') = \mathrm{E}(\widehat{\boldsymbol{\eta}}_i\widehat{\boldsymbol{\eta}}_i') \otimes \mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i').$$

Поэтому последнее условие (невырожденность $\mathrm{E}(\widehat{\boldsymbol{g}}_i\widehat{\boldsymbol{g}}_i')$*) является просто условием того, что* $\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')$ *невырождена.*

(d) Покажите, что оценку фиксированных эффектов (5.2.4) можно записать в виде GMM-оценки (7) с

$$\boldsymbol{S}_{\boldsymbol{xz}} = \frac{1}{n}\sum_{i=1}^{n} \widehat{\boldsymbol{F}}_i \otimes \boldsymbol{x}_i, \quad \boldsymbol{s}_{\boldsymbol{xy}} = \frac{1}{n}\sum_{i=1}^{n} \widehat{\boldsymbol{y}}_i \otimes \boldsymbol{x}_i, \tag{10}$$

$$\widehat{\boldsymbol{W}} = (\boldsymbol{C}'\boldsymbol{C})^{-1} \otimes \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i'\right)^{-1}.$$

Указание:

Используйте следующий результат из матричной алгебры: Если $\boldsymbol{C}$ *удовлетворяет условию (8), то*

$$\boldsymbol{C}(\boldsymbol{C}'\boldsymbol{C})^{-1}\boldsymbol{C}' = \boldsymbol{Q} \tag{11}$$

(аннулятор для $\mathbf{1}_M$*). Элементы* $\boldsymbol{f}_{im}$ *являются подмножеством элементов* $\boldsymbol{x}_i$. *Пронаблюдайте «исчезновение* $\boldsymbol{x}$*». Поэтому, например,*

$$\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{f}_{im}\boldsymbol{x}_i' \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i'\right)^{-1} \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{f}_{ih}' = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{f}_{im}\boldsymbol{f}_{ih}'.$$

(e) Пусть $\widehat{\boldsymbol{\Psi}}$ $((M-1) \times (M-1))$ является состоятельной оценкой $\boldsymbol{\Psi} \equiv \mathrm{E}(\widehat{\boldsymbol{\eta}}_i\widehat{\boldsymbol{\eta}}_i')$. Покажите, что эффективная GMM-оценка задается как

$$\widehat{\boldsymbol{\beta}} = \left(\frac{1}{n}\sum_{i=1}^{n} \widehat{\boldsymbol{F}}_i'\widehat{\boldsymbol{\Psi}}^{-1}\widehat{\boldsymbol{F}}_i\right)^{-1} \widehat{\boldsymbol{F}}_i'\widehat{\boldsymbol{\Psi}}^{-1}\widehat{\boldsymbol{y}}_i. \tag{12}$$

Покажите, что

$$\mathrm{Avar}(\widehat{\boldsymbol{\beta}}) = \left[\mathrm{E}(\widehat{\boldsymbol{F}}_i'\widehat{\boldsymbol{\Psi}}^{-1}\widehat{\boldsymbol{F}}_i)\right]^{-1}. \tag{13}$$

Указание: *Примените утверждение 4.7, чтобы представить систему из $M-1$ уравнений.*

(f) Покажите, что следующая матрица является состоятельной оценкой для $\boldsymbol{\Psi}$:

$$\underset{((M-1)\times(M-1))}{\widehat{\boldsymbol{\Psi}}} = \frac{1}{n}\sum_{i=1}^{n}(\widehat{\boldsymbol{y}}_i - \widehat{\boldsymbol{F}}_i\widehat{\boldsymbol{\beta}}_{\mathrm{FE}})(\widehat{\boldsymbol{y}}_i - \widehat{\boldsymbol{F}}_i\widehat{\boldsymbol{\beta}}_{\mathrm{FE}})'. \tag{14}$$

(g) (Статистика Саргана.) Выведите статистику Саргана, связанную с эффективной GMM-оценкой. **Указание:** *В утверждении 4.7(c) положите*

$$\widehat{\boldsymbol{S}} = \widehat{\boldsymbol{\Psi}} \otimes \left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x}_i\boldsymbol{x}_i'\right),\quad \boldsymbol{g}_n(\widehat{\boldsymbol{\beta}}) = \boldsymbol{s}_{\boldsymbol{xy}} - \boldsymbol{S}_{\boldsymbol{xz}}\widehat{\boldsymbol{\beta}}.$$

(h) Предположим, что $\boldsymbol{\eta}_i$ является сферической и что

$$\mathrm{E}(\boldsymbol{\eta}_i\boldsymbol{\eta}_i') = \sigma_\eta^2\boldsymbol{I}_M. \tag{15}$$

Проверьте, что если $\widehat{\sigma}_\eta^2$ является состоятельной оценкой σ_η^2, то

$$\widehat{\boldsymbol{\Psi}} = \widehat{\sigma}_\eta^2\boldsymbol{C}'\boldsymbol{C} \tag{16}$$

является состоятельной для $\boldsymbol{\Psi}$. Проверьте, что эффективная GMM-оценка (12) с таким выбором $\widehat{\boldsymbol{\Psi}}$ численно равна оценке фиксированных эффектов.

(i) (Инвариантность выбора $\boldsymbol{C}$.) Проверьте: если $M\times(M-1)$ матрица $\boldsymbol{C}$ удовлетворяет (8), то выполняется $\boldsymbol{B} = \boldsymbol{C}\boldsymbol{A}$, где $\boldsymbol{A}$ является невырожденной матрицей $(M-1)\times(M-1)$. Замените всюду в этом вопросе $\boldsymbol{C}$ на $\boldsymbol{B}$ и проверьте, что выбор $\boldsymbol{C}$ не изменяет никакие результаты этого вопроса.

3. Пусть SSR — сумма квадратов остатков, определенная в (5.2.15). Для модели компонент ошибки с предположением сферичности $\mathrm{E}(\boldsymbol{\eta}_i\boldsymbol{\eta}_i') = \sigma_\eta^2\boldsymbol{I}_M$ докажите, что

$$\operatorname*{plim}_{n\to\infty}\frac{SSR}{n} = (M-1)\cdot\sigma_\eta^2.$$

Указание:

$$
\begin{aligned}
SSR &= \sum_{i=1}^{n}[\boldsymbol{Q}(\boldsymbol{y}_i - \boldsymbol{F}_i\widehat{\boldsymbol{\beta}}_{FE})]'[\boldsymbol{Q}(\boldsymbol{y}_i - \boldsymbol{F}_i\widehat{\boldsymbol{\beta}}_{FE})] = \\
&= \sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{F}_i\widehat{\boldsymbol{\beta}}_{FE})'\boldsymbol{Q}(\boldsymbol{y}_i - \boldsymbol{F}_i\widehat{\boldsymbol{\beta}}_{FE}) \\
&\qquad \textit{(поскольку } \boldsymbol{Q} \textit{ симметрична и идемпотентна)} \\
&= \sum_{i=1}^{n}(\boldsymbol{y}_i - \boldsymbol{F}_i\widehat{\boldsymbol{\beta}}_{FE})'\boldsymbol{C}(\boldsymbol{C}'\boldsymbol{C})^{-1}\boldsymbol{C}'(\boldsymbol{y}_i - \boldsymbol{F}_i\widehat{\boldsymbol{\beta}}_{FE}) = \qquad \textit{(из (11))} \\
&= \sum_{i=1}^{n}\widehat{\boldsymbol{v}}_i'(\boldsymbol{C}'\boldsymbol{C})^{-1}\widehat{\boldsymbol{v}}_i \quad (\widehat{\boldsymbol{v}}_i \equiv \boldsymbol{C}'\boldsymbol{y}_i - \boldsymbol{C}'\boldsymbol{F}_i\widehat{\boldsymbol{\beta}}_{\mathrm{FE}}) = \\
&= \sum_{i=1}^{n}\mathrm{trace}[\widehat{\boldsymbol{v}}_i'(\boldsymbol{C}'\boldsymbol{C})^{-1}\widehat{\boldsymbol{v}}_i] = \\
&= \sum_{i=1}^{n}\mathrm{trace}[(\boldsymbol{C}'\boldsymbol{C})^{-1}\widehat{\boldsymbol{v}}_i\widehat{\boldsymbol{v}}_i'] = \\
&= n\cdot\mathrm{trace}[(\boldsymbol{C}'\boldsymbol{C})^{-1}\frac{1}{n}\sum_{i=1}^{n}\widehat{\boldsymbol{v}}_i\widehat{\boldsymbol{v}}_i']
\end{aligned}
$$

(поскольку оператор следа линейный).

С обычным техническим условием ($\mathrm{E}(\boldsymbol{f}_{im}\boldsymbol{f}_{ih}')$ существует и конечно для всех m, h, что получается, если $\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')$ существует и конечно) теперь можно стандартным образом показать, что

$$\frac{1}{n}\sum_{i=1}^{n}\widehat{\boldsymbol{v}}_i\widehat{\boldsymbol{v}}_i' \underset{p}{\to} \mathrm{E}(\boldsymbol{v}_i\boldsymbol{v}_i') \quad \textit{при } n \to \infty,$$

где $\boldsymbol{v}_i \equiv \boldsymbol{C}'\boldsymbol{y}_i - \boldsymbol{C}'\boldsymbol{F}_i\boldsymbol{\beta}$. Покажите, что $\mathrm{E}(\boldsymbol{v}_i\boldsymbol{v}_i') = \sigma_\eta^2 \cdot \boldsymbol{C}'\boldsymbol{C}$.

4. (Несостоятельность оценки фиксированных эффектов для динамических моделей.) Рассмотрим «динамическую» модель:

$$y_{im} = \alpha_i + \rho \cdot y_{i,m-1} + \eta_{im} \quad (m = 1, 2, \ldots, M; i = 1, 2, \ldots, n).$$

Предположим, что у нас есть случайная выборка по $(y_{i0}, y_{i1}, \ldots, y_{iM})$. Мы предполагаем, что $\mathrm{E}(\alpha_i\eta_{im}) = 0$ и $\mathrm{E}(y_{i0}\eta_{im}) = 0$ для всех m. Также предположим, что не существует кросс-секционной корреляции: $\mathrm{E}(\eta_{im}\eta_{ih}) = 0$ для $m \neq h$ и что второй момент одинаков для всех уравнений: $\mathrm{E}(\eta_{im}^2) = \sigma_\eta^2$.

(a) Запишите систему M уравнений как (5.1.1″) на с. 358, надлежащим образом определяя $\boldsymbol{y}_i$, $\boldsymbol{F}_i$ и $\boldsymbol{b}_i$.

(b) Покажите: $\mathrm{E}(y_{im} \cdot \eta_{ih}) = 0$, если $m < h$. **Указание:**

$$y_{im} = \eta_{im} + \rho\eta_{i,m-1} + \cdots + \rho^{m-1}\eta_{i1} + \frac{1-\rho^m}{1-\rho}\alpha_i + \rho^m y_{i0}.$$

(c) Проверьте, что $\mathrm{E}(y_{im} \cdot \eta_{ih}) = \rho^{m-h}\sigma_\eta^2$ для $m \geq h$. (Дополнительно.) Покажите, что

$$\mathrm{E}(\boldsymbol{F}_i'\boldsymbol{Q}\boldsymbol{\eta}_i) = -\frac{\sigma_\eta^2}{M}\frac{(M-1) - M\rho + \rho^M}{(1-\rho)^2}.$$

(d) Следовательно, если $\mathrm{E}(\boldsymbol{F}_i'\boldsymbol{Q}\boldsymbol{F}_i)$ невырождена и если $(M-1)-M\rho+\rho^M \neq 0$, то оценка фиксированных эффектов для этой динамической модели *несостоятельна*. Какое предположение в утверждении 5.1, гарантирующее состоятельность оценки фиксированных эффектов, нарушено?

5. (Дополнительно, идентификация гарантирует невырожденность.)

(a) Покажите, что матрица $\mathrm{E}(\widetilde{\boldsymbol{F}}_i'\widetilde{\boldsymbol{F}}_i)$ в (5.2.7) обратима при предположении (5.1.15) о том, что $\mathrm{E}(\boldsymbol{Q}\boldsymbol{F}_i \otimes \boldsymbol{x}_i)$ имеет полный столбцовый ранг. **Указание:**

$$\mathrm{E}(\widetilde{\boldsymbol{F}}_i'\widetilde{\boldsymbol{F}}_i) = \mathrm{E}(\boldsymbol{F}_i'\boldsymbol{Q}\boldsymbol{F}_i) = \sum_{m=1}^{M}\sum_{h=1}^{M} q_{mh}\,\mathrm{E}(\boldsymbol{f}_{im}\boldsymbol{f}_{ih}').$$

где $q_{mh} = (m,h)$-й элемент $\boldsymbol{Q}$. Покажите, что

$$\mathrm{E}(\boldsymbol{f}_{im}\boldsymbol{f}_{ih}') = \mathrm{E}(\boldsymbol{f}_{im}\boldsymbol{x}_i')[\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')]^{-1}\,\mathrm{E}(\boldsymbol{x}_i\boldsymbol{f}_{ih}').$$

(Напомним, что элементы $\boldsymbol{f}_{im}$ включены в $\boldsymbol{x}_i$.) Наконец, покажите, что $\boldsymbol{x}_i\boldsymbol{f}_{im}' = \boldsymbol{f}_{im}' \otimes \boldsymbol{x}_i$, $(\boldsymbol{F}_i \otimes \boldsymbol{x}_i)' = (\boldsymbol{f}_{i1} \otimes \boldsymbol{x}_i', \ldots, \boldsymbol{f}_{iM} \otimes \boldsymbol{x}_i')$ и

$$\begin{aligned}\mathrm{E}(\widetilde{\boldsymbol{F}}_i'\widetilde{\boldsymbol{F}}_i) &= \mathrm{E}(\boldsymbol{F}_i \otimes \boldsymbol{x}_i)'\left[\boldsymbol{Q} \otimes [\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')]^{-1}\right]\mathrm{E}(\boldsymbol{F}_i \otimes \boldsymbol{x}_i) = \\ &= \mathrm{E}(\boldsymbol{Q}\boldsymbol{F}_i \otimes \boldsymbol{x}_i)'(\boldsymbol{I}_M \otimes [\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')]^{-1})\,\mathrm{E}(\boldsymbol{Q}\boldsymbol{F}_i \otimes \boldsymbol{x}_i).\end{aligned}$$

(b) Используйте аналогичные аргументы, чтобы показать, что матрица $\mathrm{E}[\widetilde{\boldsymbol{F}}_i'\,\mathrm{E}(\widetilde{\boldsymbol{\eta}}_i\widetilde{\boldsymbol{\eta}}_i')\widetilde{\boldsymbol{F}}_i]$ в (5.2.7) является невырожденной. **Указание:** *Поскольку*

$$\widetilde{\boldsymbol{\eta}}_i = \boldsymbol{Q}\boldsymbol{\eta}_i = \boldsymbol{Q}\boldsymbol{\varepsilon}_i,$$

мы имеем

$$\widetilde{\boldsymbol{F}}_i'\,\mathrm{E}(\widetilde{\boldsymbol{\eta}}_i\widetilde{\boldsymbol{\eta}}_i')\widetilde{\boldsymbol{F}}_i = \widetilde{\boldsymbol{F}}_i'\boldsymbol{Q}\,\mathrm{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i')\boldsymbol{Q}\widetilde{\boldsymbol{F}}_i = \widetilde{\boldsymbol{F}}_i'\,\mathrm{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i')\widetilde{\boldsymbol{F}}_i.$$

Также

$$\mathrm{E}[\widetilde{\boldsymbol{F}}_i'\,\mathrm{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i')\widetilde{\boldsymbol{F}}_i] = \mathrm{E}(\boldsymbol{Q}\boldsymbol{F}_i \otimes \boldsymbol{x}_i)'(\mathrm{E}(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i') \otimes [\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')]^{-1})\,\mathrm{E}(\boldsymbol{Q}\boldsymbol{F}_i \otimes \boldsymbol{x}_i).$$

(c) Покажите, что $\mathrm{E}(\widehat{\boldsymbol{F}}_i'\boldsymbol{\Psi}^{-1}\widehat{\boldsymbol{F}}_i)$ в (13) в аналитическом упражнении 2 является невырожденной. **Указание:** *Как показано в аналитическом упражнении 2(c),* $\mathrm{E}(\boldsymbol{Q}\boldsymbol{F}_i \otimes \boldsymbol{x}_i)$ *имеет полный столбцовый ранг тогда и только тогда, когда* $\mathrm{E}(\boldsymbol{C}'\boldsymbol{F}_i \otimes \boldsymbol{x}_i)$ *имеет полный столбцовый ранг.*

6. (Дополнительно, альтернативное условие идентификации.) Рассмотрим модель компонент ошибки (5.1.1″) на с. 358. Пусть $p \equiv \#\boldsymbol{\beta}$. Если $\boldsymbol{A}_m$ является матрицей, которая выхватывает $\boldsymbol{f}_{im}$ из $\boldsymbol{x}_i$, ее можно записать как

$$\underset{(K\times p)}{\boldsymbol{A}_m} = \begin{bmatrix} \boldsymbol{e}_m \otimes \boldsymbol{I}_p \\ \boldsymbol{0} \end{bmatrix},$$

где $\boldsymbol{e}_m$ является m-м столбцом $\boldsymbol{I}_M$.

(a) Пусть $\boldsymbol{A}$ $(KM \times p)$ является матрицей, полученной штабелированием $(\boldsymbol{A}_1, \dots, \boldsymbol{A}_M)$. Покажите, что $\boldsymbol{F}_i = (\boldsymbol{I}_M \otimes \boldsymbol{x}_i')\boldsymbol{A}$.

(b) Покажите, что ранговое условие (5.1.15), объединенное с условием, что $\mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')$ является невырожденной (что вытекает из (5.1.5) и (5.1.6)), можно записать как

$$\operatorname{rank}[\boldsymbol{Q} \otimes \boldsymbol{I}_K)\boldsymbol{A}] = p \quad (\equiv \#\boldsymbol{\beta}).$$

(Этот результат будет использован в следующем упражнении.) **Указание:**

$$\begin{aligned} \boldsymbol{F}_i \otimes \boldsymbol{x}_i &= \left([(\boldsymbol{I}_M \otimes \boldsymbol{x}_i')\boldsymbol{A}]\right) \otimes \boldsymbol{x}_i = \\ &= [(\boldsymbol{I}_M \otimes \boldsymbol{x}_i') \otimes \boldsymbol{x}_i](\boldsymbol{A} \otimes 1) = \\ &= (\boldsymbol{I}_M \times \boldsymbol{x}_i' \otimes \boldsymbol{x}_i)\boldsymbol{A} = \\ &= [\boldsymbol{I}_M \otimes (\boldsymbol{x}_i\boldsymbol{x}_i')]\boldsymbol{A}. \end{aligned}$$

Поэтому

$$\begin{aligned} \widetilde{\boldsymbol{F}}_i \otimes \boldsymbol{x}_i &= (\boldsymbol{Q} \otimes \boldsymbol{I}_K)(\boldsymbol{F}_i \otimes \boldsymbol{x}_i) = \\ &= [\boldsymbol{Q} \otimes (\boldsymbol{x}_i\boldsymbol{x}_i')]\boldsymbol{A} = \\ &= (\boldsymbol{I}_M \otimes \boldsymbol{x}_i\boldsymbol{x}_i')(\boldsymbol{Q} \otimes \boldsymbol{I}_K)\boldsymbol{A}. \end{aligned}$$

Поэтому $\mathrm{E}(\widetilde{\boldsymbol{F}}_i \otimes \boldsymbol{x}_i) = [\boldsymbol{I}_M \otimes \mathrm{E}(\boldsymbol{x}_i\boldsymbol{x}_i')](\boldsymbol{Q} \otimes \boldsymbol{I}_K)\boldsymbol{A}$.

7. (Дополнительно, проверка невырожденности.) Докажите, что матрица $\mathrm{E}(\widetilde{\boldsymbol{F}}_i'\widetilde{\boldsymbol{\eta}}_i\widetilde{\boldsymbol{\eta}}_i'\widetilde{\boldsymbol{F}}_i)$ в (5.2.23) является невырожденной, выполнив следующие шаги:

(a) Запишите $\widetilde{F}_i'\widetilde{\eta}_i$ как линейную функцию от $g_i \equiv \varepsilon \otimes x_i$. **Указание:** *Пусть A определяется, как в предыдущем упражнении.*

$$\begin{aligned}\widetilde{F}_i'\widetilde{\eta}_i &= A'(I_M \otimes x_i)Q\eta_i = \\ &= A'(I_M \otimes x_i)Q\varepsilon_i = \quad (\text{поскольку } \varepsilon_i = \mathbf{1}_M \cdot \alpha_i + \eta_i) = \\ &= A'(I_M \otimes x_i)(Q\varepsilon_i \otimes 1) = \\ &= A'(Q\varepsilon_i \otimes x_i) = \\ &= A'(Q \otimes I_K)(\varepsilon_i \otimes x_i) = \\ &= A'(Q \otimes I_K)g_i.\end{aligned}$$

(b) Докажите желаемый результат. **Указание:** $\mathrm{E}(g_i g_i')$ *положительно определена по* (5.1.6).

Эмпирические упражнения

В этом упражнении вы оцените скорость конвергенции, используя несколько различных методов оценивания. Мы используем международную панель, построенную в [Summers and Heston, 1991], которая стала стандартным набором данных для изучения развития стран. Панель Саммерса — Хестона (также называемая Penn World Table) включает основные макроэкономические переменные, измеряемые на постоянной основе практически для всех стран мира.

Файл SUM_HES.ASC извлечен из Penn World Table версии 5.6 за 1960–1985 с домашней страницы NBER (www.nber.com). Файл организован следующим образом. Для каждой страны информация по стране содержится в нескольких записях (строках). Первая запись содержит идентификационный код страны и две дамми, одна — для коммунистического режима (называемая как COM ниже) и другая — для организации стран — экспортеров нефти (называемая как $OPEC$ ниже). Вторая запись содержит год, население (в тысячах), реальный ВВП на душу населения (в долларах США за 1985 год) и норму сбережения (в процентах) для 1960 года. Третья запись содержит то же самое за 1961 год и так далее. Двадцать седьмая запись для страны относится к 1985 году. Первые несколько записей файла выглядят как:

```
  1 0 0
1960     10800     1723     19.9
1961     11016     1599     21.1
                  ⋮
1984     21173     2962     26.4
1985     21848     2988     26.0
  2 0 0
1960      4816      931      3.3
1961      4918      976      2.9
```

Файл охватывает все 125 стран, для которых доступны данные. (Соответствие между ID страны и названием страны содержится в `country.asc`. Страна 1 — это Алжир, страна 2 — Ангола и т. д. Заметим, что соответствие в версии 5.6 Penn World Table отличается от [Summers and Heston, 1991].) GDP на душу населения — это реальный ВВП страны, пересчитанный в доллары США на основе паритета покупательной способности валюты страны против доллара за 1985 год (переменная «RGDPCH» в Penn World Table). Норма сбережения измеряется как отношение реальных инвестиций к реальному ВВП («I» в Penn World Table). Для понимания того, как построены переменные, см. параграфы II и III (в частности, III.D) в [Summers and Heston, 1991].

Первым вопросом, который мы изучаем эмпирически, является условная конвергенция. (На этом этапе было бы полезно бегло просмотреть работу [Mankiw, Romer, and Weil, 1992].) Пусть s_i является нормой сбережения страны i, а n_i является темпом роста населения страны (который мы будем считать темпом роста трудовых ресурсов). В модели роста Солоу — Свана с производственной функцией Кобба — Дугласа равновесный уровень выпуска на эффективную единицу труда, обозначенный в основном тексте как q^*, является линейной функцией $\log(s_i) - \log(n_i + g + \delta)$, где g — темп трудоинтенсивного технологического прогресса, а δ — норма амортизации основного капитала. Тогда, предполагая что $A(0)$, начальный уровень технологии, одинаков для всех стран, уравнение (5.4.10) можно записать как

$$y_{im} = \phi_m + \rho y_{i,m-1} + \gamma_1 \log(s_i) + \gamma_2 \log(n_i + g + \delta) + \eta_{im}. \qquad (1)$$

Из модели роста следует, что $\gamma_1 = -\gamma_2$, но мы не будем накладывать это условие. Это является спецификацией, оцененной в табл. IV в [Mankiw et al., 1992]. Как и в [Mankiw et al.], предположим, что $g + \delta$ равно 5% для всех стран, и возьмем s_i как норму сбережения, усредненную за 1960–1985 годы. Мы измеряем n_i как среднегодовой темп роста населения за тот же период ([Mankiw et al.] используют средний темп роста населения в трудоспособном возрасте). Поскольку наша выборка из 125 стран включает (бывшие) коммунистические страны и OPEC, мы добавляем COM (= 1, если коммунистические, 0 — в противном случае) и $OPEC$ (= 1, если страна в OPEC, 0 — в противном случае) в уравнение:

$$\begin{aligned} y_{im} = \phi_m + \rho y_{i,m-1} + \gamma_1 \log(s_i) + \gamma_2 \log(n_i + g + \delta) + \\ + \gamma_3 COM_i + \gamma_4 OPEC_i + \eta_{im}. \end{aligned} \qquad (2)$$

(a) Вычитая $y_{i,m-1}$ из обеих частей, мы можем переписать (2) как

$$\begin{aligned} y_{im} - y_{i,m-1} = \phi_m + (\rho - 1) y_{i,m-1} + \gamma_1 \log(s_i) + \gamma_2 \log(n_i + g + \delta) + \\ + \gamma_3 COM_i + \gamma_4 OPEC_i + \eta_{im}. \end{aligned} \qquad (2')$$

Для этой спецификации положим $M = 1$ (только одно уравнение для оценивания) и возьмем $t_0 = 1960$ и $t_1 = 1985$, так что зависимая переменная $y_{i1} - y_{i0}$ является накопленным ростом между 1960 и 1985 годами. Постройте график $y_{i1} - y_{i0}$ против y_{i0} для 125 стран, включенных в выборку. Есть ли какая-либо связь между начальным ВВП (y_{i0}) и последующим ростом ($y_{i1} - y_{i0}$)? Предполагая напрямую, что ошибки ортогональны регрессорам, оцените уравнение (2′) посредством OLS. Вычислите значение λ (скорость конвергенции), вытекающее из OLS-оценки для ρ. (Она должна быть менее 1% в год.) Вычислите асимптотическую стандартную ошибку (то есть умноженный на $1/\sqrt{n}$ квадратный корень из асимптотической дисперсии) вашей оценки λ. **Указание:** *Из* (5.4.3) *следует, что* $\widehat{\lambda} = -\log(\widehat{\rho})/25$. *Из дельта-метода (лемма 2.5 в параграфе 2.1)*

$$\mathrm{Avar}(\widehat{\lambda}) = \frac{\mathrm{Avar}(\widehat{\rho})}{(25\rho)^2}$$

(поскольку производная $\log(\rho)$ *равна* $1/\rho$*). Поэтому состоятельная оценка* $\mathrm{Avar}(\widehat{\lambda})$ *равна* $\widehat{\mathrm{Avar}(\widehat{\rho})}/(25\widehat{\rho})^2$. *Стандартная ошибка* $\widehat{\lambda}$ *равна умноженному на* $1/\sqrt{n}$ *квадратному корню из этой величины. Поэтому: стандартная ошибка* $\widehat{\lambda} =$ *(стандартная ошибка* $\widehat{\rho}$*)/(25*$\widehat{\rho}$*).*

Можете ли вы подтвердить вывод о том, что «если страна не изменяет своих инвестиций и темпов роста населения, то должна быть устойчивой тенденция более быстрого роста бедных стран по сравнению с богатыми» [Mankiw et al., 1992, p. 428].

(b) (Оценивание методом фиксированных эффектов.) Полагая $M = 25$, $t_0 = 1960, t_1 = 1961, \ldots, t_{25} = 1985$, мы можем сформировать систему из 25 уравнений с (5.4.10) в качестве m-го уравнения. Оцените эту систему методом фиксированных эффектов, предполагая, что ошибка сферическая в смысле (5.2.12). Чему равно выведенное значение λ? (Оно должно быть около 6,4% в год.) (Дополнительно: Вычислите стандартные ошибки без предположения о сферичности ошибок. Как подчеркивается в тексте, оценка фиксированных эффектов *не является* состоятельной. Мы тем не менее слепо применяем метод фиксированных эффектов только для получения опыта программирования.)

Совет для пакета TSP: Командой TSP для оценки фиксированных эффектов (с предположением о сферичности ошибок) является PANEL. Она, однако, не принимает организацию данных в SUM_HES.ASC. Она требует, чтобы ID группы (страны) и год были включены в строки, содержащие информацию по группе—году. Файл данных, который PANEL принимает, называется `FIXED.ASC`. Его первые несколько записей имеют вид:

Ответы на избранные вопросы

Аналитические упражнения

1c. Решение имеет вид:

$$\begin{aligned}
\mathrm{E}(\boldsymbol{\eta}_i|\boldsymbol{W}) &= \mathrm{E}(\boldsymbol{\eta}_i|\boldsymbol{F}) = \quad \text{(поскольку } \boldsymbol{D} \text{ является матрицей констант)}\\
&= \mathrm{E}(\boldsymbol{\eta}_i|\boldsymbol{F}_1,\dots,\boldsymbol{F}_n) =\\
&= \mathrm{E}(\boldsymbol{\eta}_i|\boldsymbol{F}_i)\\
&\text{(поскольку } \boldsymbol{\eta}_i \text{ не зависит от } (\boldsymbol{F}_1,\dots,\boldsymbol{F}_{i-1},\boldsymbol{F}_{i+1},\dots,\boldsymbol{F}_n)\\
&\text{по предположению i.i.d.)}\\
&= \mathbf{0}.
\end{aligned}$$

Следовательно, регрессоры являются экзогенными. Также

$$\begin{aligned}
\mathrm{E}(\boldsymbol{\eta}_i\boldsymbol{\eta}_i'|\boldsymbol{W}) &= \mathrm{E}(\boldsymbol{\eta}_i\boldsymbol{\eta}_i'|\boldsymbol{F}) =\\
&= \mathrm{E}(\boldsymbol{\eta}_i\boldsymbol{\eta}_i'|\boldsymbol{F}_i) =\\
&= \sigma_\eta^2\boldsymbol{I}_M \quad \text{(по предположению о сферичности ошибок).}
\end{aligned}$$

В силу предположения i.i.d., $\mathrm{E}(\boldsymbol{\eta}_i\boldsymbol{\eta}_i'|\boldsymbol{W}) = \mathbf{0}$ для $i \neq j$. Поэтому $\mathrm{E}(\boldsymbol{\eta}\boldsymbol{\eta}'|\boldsymbol{W}) = \sigma_\eta^2\boldsymbol{I}_{Mn}$.

2c. Докажите, что «$\mathrm{E}(\boldsymbol{Q}\boldsymbol{F}_i \otimes \boldsymbol{x}_i)$ имеет полный столбцовый ранг» $\Leftrightarrow$ $\Leftrightarrow$ «$\mathrm{E}(\boldsymbol{C}'\boldsymbol{F}_i \otimes \boldsymbol{x}_i)$ имеет полный столбцовый ранг»:

Пусть $\boldsymbol{A} \equiv \mathrm{E}(\boldsymbol{Q}\boldsymbol{F}_i \otimes \boldsymbol{x}_i)$ и $\boldsymbol{B} \equiv \mathrm{E}(\boldsymbol{C}'\boldsymbol{F}_i \otimes \boldsymbol{x}_i)$. Существует матрица $\boldsymbol{L}$ размера $M \times (M-1)$ полного столбцового ранга, такая что $\boldsymbol{Q} = \boldsymbol{L}\boldsymbol{C}'$ (это фактически $\boldsymbol{C}(\boldsymbol{C}'\boldsymbol{C})^{-1}$, как следует из (11)). Пусть $\boldsymbol{D} \equiv \boldsymbol{L} \otimes \boldsymbol{I}_K$. Тогда $\boldsymbol{A} = \boldsymbol{D}\boldsymbol{B}$. Поскольку ранг произведения двух матриц не превосходит ранга любой из этих двух матриц,

$$\operatorname{rank}(\boldsymbol{A}) \leq \operatorname{rank}(\boldsymbol{B}). \tag{1}$$

Далее, рассмотрим произведение $\boldsymbol{D}'\boldsymbol{A} = \boldsymbol{D}'\boldsymbol{D}\boldsymbol{B}$.

$$\operatorname{rank}(\boldsymbol{D}'\boldsymbol{D}\boldsymbol{B}) \leq \operatorname{rank}(\boldsymbol{D}'\boldsymbol{A}) \leq \operatorname{rank}(\boldsymbol{A}). \tag{2}$$

Но поскольку $\boldsymbol{L}$ имеет полный столбцовый ранг, $\boldsymbol{D}'\boldsymbol{D}$ является невырожденной. Так как умножение на невырожденную матрицу не изменяет ранг,

$$\operatorname{rank}(\boldsymbol{D}'\boldsymbol{D}\boldsymbol{B}) = \operatorname{rank}(\boldsymbol{B}). \tag{3}$$

Из (1)–(3) следует, что $\operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{B})$. Поскольку $\boldsymbol{A}$ и $\boldsymbol{B}$ имеют одинаковое число столбцов, из этого следует нужный вывод.

4a.

$$\boldsymbol{y}_i = \begin{bmatrix} y_{i1} \\ \vdots \\ y_{iM} \end{bmatrix}, \quad \boldsymbol{F}_i = \begin{bmatrix} y_{i0} \\ \vdots \\ y_{i,M-1} \end{bmatrix}.$$

Вектора $\boldsymbol{b}_i$ нет.

4d. Поскольку $\mathrm{E}(\eta_{ih} \cdot y_{im}) \neq 0$ для $m \geq h$, некоторые из «кросс»-ортогональностей не выполняются.

Литература

Barro, R., and X. Sala-i-Martin, 1995, *Economic Growth*, New York: McGraw-Hill.

Mankiw, G., D. Romer, and D. Weil, 1992, "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics*, 107, 407–437.

Newey, W., 1985, "Generalized Method of Moments Specification Testing," *Journal of Economics*, 29, 229–256.

Nickell, S., 1981, "Biases in Dynamic Models with Fixed Effects," *Econometrica*, 49, 1417–1426.

Summers, R., and A. Heston, 1991, "The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950–1988," *Quarterly Journal of Economics*, 106, 327–368.

Глава 6. Сериальная корреляция

Эта глава обобщает GMM-модель из главы 3, чтобы включить сериальную корреляцию в произведение вектора инструментов и ошибки. Для этого нам нужно обобщить центральную предельную теорему из главы 2 на сериально коррелированные процессы. Такое обобщение возможно при определенных условиях, ограничивающих степень сериальной корреляции. Соответствующее условие является прозрачным для случайных процессов, называемых линейными. Линейные процессы важны и сами по себе, и в этой главе их рассмотрению посвящены четыре параграфа. В прикладном параграфе обобщенная модель GMM используется для тестирования гипотезы эффективности рынка, на этот раз валютного.

Замечание по обозначениям: Поскольку проблемы, рассматриваемые здесь, являются специфическими для временных рядов, в этой главе мы используем индекс «t» вместо «i».

6.1. Моделирование сериальной корреляции: линейные процессы

В параграфе 2.2 были введены ковариационно стационарные процессы и их автоковариации. В частности, (скалярный или одномерный) **процесс белого шума** (*white noise process*) $\{\varepsilon_t\}$ является ковариационно стационарным процессом с нулевым средним и с отсутствием сериальной корреляции:

$$\mathrm{E}(\varepsilon_t) = 0, \quad \mathrm{E}(\varepsilon_t^2) = \sigma^2 > 0, \quad \mathrm{E}(\varepsilon_t \varepsilon_{t-j}) = 0 \quad \text{для } j \neq 0.$$

Очень важный класс ковариационно стационарных процессов, называемых **линейными процессами** (*linear processes*), можно получить, беря скользящее среднее процесса белого шума. Этот параграф изучает общие свойства линейных процессов, используя инструмент, называемый **фильтром** (*filter*).

Поскольку текущее значение линейного процесса может зависеть от возможно бесконечного количества прошлых значений процесса белого шума, удобно, чтобы процесс белого шума и линейный процесс, который он порождает, были определены для $t = 0, -1, -2, \ldots$, так же как и для $t = 1, 2, \ldots$. Интуитивно ковариационно стационарный процесс, определенный для всех целых ($t = 0, \pm 1, \pm 2, \ldots$), является процессом, начинающимся настолько давно, что его среднее и автоковариации стабилизировались к инвариантным ко времени константам.

MA(q)

Самым простым примером линейных процессов, которые проявляют сериальную корреляцию, является процесс скользящего среднего конечного порядка. Процесс $\{y_t\}$ называется **процессом скользящего среднего q-го порядка** (*q-th order moving average process*) (MA(q)), если он может быть записан как взвешенное среднее текущего и q самых недавних значений процесса белого шума:

$$y_t = \mu + \theta_0 \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \quad \text{с } \theta_0 = 1. \tag{6.1.1}$$

Очевидно, процесс скользящего среднего является ковариационно стационарным со средним μ. Легко проверить, что автоковариация j-го порядка, $\gamma_j (\equiv \mathrm{E}[(y_t - \mu)(y_{t-j} - \mu)])$, имеет вид:

$$\gamma_j = (\theta_j \theta_0 + \theta_{j+1}\theta_1 + \cdots + \theta_q \theta_{q-j})\sigma^2 = \sigma^2 \sum_{k=0}^{q-j} \theta_{j+k}\theta_k$$

$$\text{для } j = 0, 1, \ldots, q, \tag{6.1.2a}$$

$$\gamma_j = 0 \quad \text{для } j > q, \tag{6.1.2b}$$

где $\sigma^2 \equiv \mathrm{E}(\varepsilon_t^2)$. (Эта формула также охватывает γ_{-j}, поскольку $\gamma_j = \gamma_{-j}$.) Например, для $q = 1$ мы имеем:

$$\gamma_0 = (\theta_0^2 + \theta_1^2)\sigma^2 = (1 + \theta_1^2)\sigma^2$$

(полагая $j = 0$ и $q = 1$ в (6.1.2a) и замечая, что $\theta_0 = 1$).

$$\gamma_1 = \gamma_{-1} = (\theta_1 \theta_0)\sigma^2 = \theta_1 \sigma^2$$

(полагая $j = 1$ и $q = 1$ в (6.1.2a) и замечая, что $\theta_0 = 1$).

$$\gamma_j = 0 \quad \text{для } j = \pm 2, \pm 3, \ldots.$$

Полный профиль автоковариаций, $\{\gamma_j\}$, описывается $q + 1$ параметрами ($\theta_1, \theta_2, \ldots, \theta_q$ и σ^2), а коррелограмма $\{\rho_j\}$ (где $\rho_j \equiv \gamma_j / \gamma_0$) — q параметрами ($\theta_1, \theta_2, \ldots, \theta_q$).

MA(∞) *как предел в среднем квадратическом*

Как иллюстрирует приведенный выше пример, сериальная корреляция в процессах MA(q) полностью пропадает после q лагов. Хотя некоторые временные ряды обладают этим свойством (мы увидим такой пример в приложении ниже), конечно, желательно, чтобы была возможность моделировать сериальную корреляцию, которая этим свойством не обладает. Очевидная идея — сделать y_t зависящим от бесконечного прошлого, заменяя сумму конечного количества слагаемых в (6.1.1),

$$\theta_0\varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q},$$

на бесконечную сумму

$$\psi_0\varepsilon_t + \psi_1\varepsilon_{t-1} + \cdots = \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j}, \tag{6.1.3}$$

где $\{\psi_j\}$ — последовательность действительных чисел.

Для того чтобы эта идея работала, мы должны быть уверены в том, что эта сумма бесконечно большого количества случайных величин определена, а именно что частичная сумма

$$\sum_{j=0}^{n}\psi_j\varepsilon_{t-j} \tag{6.1.4}$$

при $n \to \infty$ сходится (в определенном смысле) к некоторой случайной величине. Одно из популярных условий, при которых это реализуется, состоит в том, что последовательность действительных чисел $\{\psi_j\}$ является **абсолютно суммируемой** (*absolutely summable*):

$$\sum_{j=0}^{\infty}|\psi_j| < \infty. \tag{6.1.5}$$

(Это стандартное сокращенное выражение в математическом анализе. Уравнение (6.1.5) следует читать так: частичная сумма $\sum_{j=0}^{n}|\psi_j|$ сходится к некоторому действительному числу [то есть к конечному пределу] при $n \to \infty$, а бесконечная сумма в (6.1.5) *по определению* обозначает это действительное число.) Чтобы последовательность $\{\psi_j\}$ была абсолютно суммируемой, необходимо (но, конечно, не достаточно), чтобы $\psi_j \to 0$ при $j \to \infty$. Таким образом, абсолютная суммируемость требует, чтобы влияние прошлых шоков, представленное посредством ψ_j, в конце концов, затухало.

Как заявляется в следующем утверждении, для любого заданного t частичная сумма (6.1.4) сходится в среднем квадратическом к некоторой случайной величине. Этот предел в среднем квадратическом является единственным в следующем смысле. Если существует другая случайная величина, к которой сходится в среднем квадратическом указанная частичная сумма, то тогда (как вы покажете в аналитическом упражнении 1) эти два предела в среднем квадратическом равны с вероятностью единица. Мы *определяем* бесконечную сумму (6.1.3) как единственный предел в среднем квадратическом частичной суммы (6.1.4) и говорим, что «эта бесконечная сумма сходится в среднем квадратическом».

Утверждение 6.1 (MA(∞) с абсолютно суммируемыми коэффициентами): *Пусть $\{\varepsilon_t\}$ является белым шумом и $\{\psi_j\}$ является последовательностью действительных чисел, которая абсолютно суммируема. Тогда*

(a) *Для каждого t*

$$y_t = \mu + \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \tag{6.1.6}$$

сходится в среднем квадратическом. $\{y_t\}$ является ковариационно стационарным. (Процесс $\{y_t\}$ называется ***процессом скользящего среднего бесконечного порядка*** *(MA(∞)) (infinite-order moving-average process).*

(b) *Математическое ожидание y_t равно μ. Автоковариации $\{\gamma_j\}$ задаются как*

$$\begin{aligned}\gamma_j &= (\psi_j\psi_0 + \psi_{j+1}\psi_1 + \psi_{j+2}\psi_2 + \cdots)\sigma^2 = \\ &= \sigma^2 \sum_{k=0}^{\infty} \psi_{j+k}\psi_k \quad (j = 0, 1, 2, \dots).\end{aligned} \tag{6.1.7}$$

(c) *Автоковариации абсолютно суммируемы:*

$$\sum_{j=0}^{\infty} |\gamma_j| < \infty. \tag{6.1.8}$$

(d) *Если, в дополнение, $\{\varepsilon_t\}$ являются i.i.d, то тогда процесс $\{y_t\}$ является (строго) стационарным и эргодическим.*

Этот результат включает процессы MA(q) в качестве частного случая с $\psi_j = \theta_j$ для $j = 0, 1, \dots, q$ и $\psi_j = 0$ для $j > q$. Очевидно, что такая последовательность абсолютно суммируема. Доказательство частей (a)–(c) является аналитическим упражнением 2. Доказательство (a) будет использовать известный из математического анализа факт о конвергации последовательности Коши (см. контрольный вопрос 1 о последовательности Коши). Часть (b) является очевидной экстраполяцией на бесконечность соответствующих результатов о среднем и автоковариациях процессов MA конечного порядка, но эта экстраполяция предполагает перестановку математических ожиданий и суммирований. Например, чтобы доказать, что средним MA(∞)-процесса является μ, вам придется показать, что математическое ожидание предела в среднем квадратическом (6.1.3) равно пределу математического ожидания (6.1.4). Тот факт, что эта операция легитимна при сходимости в среднем квадратическом, будет использован для доказательства части (b). Интуитивно часть (c) говорит о том, что если влияние прошлых шоков, представленных ψ_j, затухает достаточно быстро, чтобы $\{\psi_j\}$ была абсолютно суммируемой, то тогда и сериальная корреляция затухает быстро. Абсолютная суммируемость автоковариаций будет играть ключевую роль в теории ковариационно стационарных процессов, разрабатываемой в этой главе. Для доказательства (d) см., например: [Hannan, 1970, p. 204, Theorem 3][1].

Процесс MA(∞) в (6.1.6) называется **односторонним** (*one-sided*), поскольку скользящее среднее не включает будущее значение ε_t. Это частный случай **линейного процесса** (*linear process*), который определяется как двустороннее скользящее среднее вида:

$$y_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j} \quad \text{с} \quad \sum_{j=-\infty}^{\infty} |\psi_j| < \infty.$$

Указанное утверждение (и все результаты в этом параграфе) можно легко распространить на общие линейные процессы. Мы, однако, фокусируемся на односторонних скользящих средних только потому, что двусторонние скользящие средние редко встречаются в экономике.

Утверждение 6.1 может быть обобщено на случай, когда процесс $\{y_t\}$ является скользящим средним общего ковариационно стационарного процесса, а не процесса белого шума. В частности, абсолютная суммируемость автоковариаций переживает операцию взятия абсолютно суммируемого взвешенного среднего.

Утверждение 6.2 (Фильтрация ковариационно стационарного процесса): *Пусть $\{x_t\}$ является ковариационно стационарным процессом, а $\{h_j\}$ — абсолютно суммируемой последовательностью действительных чисел. Тогда*

(a) Для каждого t бесконечная сумма

$$y_t = \sum_{j=0}^{\infty} h_j x_{t-j} \tag{6.1.9}$$

[1] Абсолютной суммируемости $\{\psi_j\}$ более чем достаточно для выполнения части (a). Как показано в аналитическом упражнении 2, суммируемости с квадратом, которая является более слабой, чем абсолютная суммируемость, достаточно, чтобы гарантировать, что бесконечная сумма также определена как предел в среднем квадратическом. Однако для выполнения других частей этого утверждения необходима абсолютная суммируемость.

сходится в среднем квадратическом. Процесс $\{y_t\}$ является ковариационно стационарным.

(b) *Если, кроме того, автоковариации процесса $\{x_t\}$ абсолютно суммируемы, то тогда то же свойство выполняется и для автоковариаций процесса $\{y_t\}$.*

За доказательством мы отсылаем, например, к Фуллеру: [Fuller, 1996, Theorem 4.3.1], но основная техника здесь та же самая, что и в доказательстве утверждения 6.1. Получение автоковариаций $\{y_t\}$, которые обобщают (6.1.7), является аналитическим упражнением 3.

Фильтры

Операция взятия взвешенного среднего (возможно бесконечного количества) последовательных значений процесса, как в (6.1.6) и (6.1.9), называется **фильтрацией** (*filtering*). Ее можно записать в компактном виде, если использовать так называемый **оператор запаздывания** (*lag operator*) L, который определяется соотношением $L^j x_t = x_{t-j}$. Мы сейчас введем некоторые понятия, связанные с оператором запаздывания, которые будут полезны в анализе временных рядов[1].

Определение фильтров

Для произвольно заданной последовательности действительных чисел, $(\alpha_0, \alpha_1, \alpha_2, \dots)$, определим **фильтр** (*filter*) как

$$\alpha(L) \equiv \alpha_0 + \alpha_1 L + \alpha_2 L^2 + \cdots . \tag{6.1.10}$$

Если мы просто механически, не задумываясь, применим определение оператора запаздывания к процессу $\{x_t\}$, то получим:

$$\begin{aligned} \alpha(L)x_t &= \alpha_0 x_t + \alpha_1 L x_t + \alpha_2 L^2 x_t + \cdots = \\ &= \alpha_0 x_t + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots = \\ &= \sum_{j=0}^{\infty} \alpha_j x_{t-j}. \end{aligned} \tag{6.1.11}$$

Это и есть определение $\alpha(L)x_t$. Если $x_t = c$ (константа), то $\alpha(L)c = \alpha(1)c = c\sum_{j=0}^{\infty}\alpha_j$. Если $\alpha_j \neq 0$ для $j = p$ и $\alpha_j = 0$ для $j > p$, фильтр сводится к **лаговому полиному p-й степени** (*p-th degree lag polynomial*):

$$\alpha(L) = \alpha_0 + \alpha_1 L + \cdots + \alpha_p L^p.$$

[1] Для более полного обсуждения фильтров см., например: [Gourieroux and Monfort, 1997, Sections 5.2 and 7.4].

Когда он применяется к входному процессу, то создает взвешенное среднее текущего и p самых недавних значений процесса. Утверждение 6.2 гарантирует нам, что объект $\alpha(L)x_t$, определенный в (6.1.11), является случайной величиной, формирующей ковариационно стационарный процесс, если последовательность $\{\alpha_j\}$ является абсолютно суммируемой и если входной процесс $\{x_t\}$ является ковариационно стационарным. Фильтр с абсолютно суммируемыми коэффициентами будет называться (в этой книге) **абсолютно суммируемым фильтром** (*absolutely summable filter*). Если использовать модный язык теории множеств, то абсолютно суммируемый фильтр является отображением множества ковариационно стационарных процессов в себя.

Произведение фильтров

Пусть $\{\alpha_j\}$ и $\{\beta_j\}$ являются двумя произвольными последовательностями действительных чисел. Определим последовательность $\{\delta_j\}$ соотношением:

$$\begin{gathered}\alpha_0\beta_0 = \delta_0,\\ \alpha_0\beta_1 + \alpha_1\beta_0 = \delta_1,\\ \alpha_0\beta_2 + \alpha_1\beta_1 + \alpha_2\beta_0 = \delta_2,\\ \cdots,\\ \alpha_0\beta_j + \alpha_1\beta_{j-1} + \alpha_2\beta_{j-2} + \cdots + \alpha_{j-1}\beta_1 + \alpha_j\beta_0 = \delta_j,\\ \cdots.\end{gathered} \tag{6.1.12}$$

Последовательность $\{\delta_j\}$, построенная по этой запутанной формуле, называется **сверткой** (*convolution*) последовательностей $\{\alpha_j\}$ и $\{\beta_j\}$. Пусть $\alpha(L) \equiv \alpha_0 + \alpha_1 L + \alpha_2 L^2 + \cdots$, $\beta(L) \equiv \beta_0 + \beta_1 L + \beta_2 L^2 + \cdots$ и $\delta(L) \equiv \equiv \delta_0 + \delta_1 L + \delta_2 L^2 + \cdots$ являются соответствующими фильтрами. Мы определяем **произведение** (*product*) двух фильтров $\alpha(L)$ и $\beta(L)$ как этот самый фильтр $\delta(L)$ и записываем его как

$$\delta(L) = \alpha(L)\,\beta(L).$$

Например, для $\alpha(L) = 1 + \alpha_1 L$ и $\beta(L) = 1 + \beta_1 L$ формула свертки приводит к

$$(1 + \alpha_1 L)(1 + \beta_1 L) = 1 + (\alpha_1 + \beta_1)L + \alpha_1\beta_1 L^2.$$

(Если вы знакомы с произведением двух полиномов обычного вида из математического анализа, вы можете сразу видеть, что это определение для фильтров полностью аналогично.) Как ясно из определения, фильтры **коммутативны** (*commutative*):

$$\alpha(L)\beta(L) = \beta(L)\alpha(L).$$

Коммутативность, однако, не будет переноситься на матричные фильтры (см. параграф 6.3 ниже).

Если две последовательности $\{\alpha_j\}$ и $\{\beta_j\}$ абсолютно суммируемы и если процесс $\{x_t\}$ ковариационно стационарный, то утверждение 6.2 гарантирует нам, что процесс $\{\beta(L)x_t\}$ ковариационно стационарный, и поэтому $\alpha(L)\beta(L)x_t$ является случайной величиной, равной $\delta(L)x_t$. Таким образом, если $\alpha(L)$ и $\beta(L)$ абсолютно суммируемы и процесс $\{x_t\}$ ковариационно стационарный, то

$$«\alpha(L)\,\beta(L) = \delta(L)» \Rightarrow «\alpha(L)\,\beta(L)x_t = \delta(L)x_t». \qquad (6.1.13)$$

(Кстати, последовательность $\{\delta_j\}$ абсолютно суммируема (см., например: [Fuller, 1996, p. 30]).)

Обращения

Особый интерес представляет случай, когда $\delta(L) = 1$ (весьма частный случай фильтра), так что два фильтра $\alpha(L)$ и $\beta(L)$ удовлетворяют соотношению $\alpha(L)\beta(L) = 1$. Мы говорим, что $\beta(L)$ является **обращением** (*inverse*) фильтра $\alpha(L)$ и обозначаем его как $\alpha(L)^{-1}$ или $1/\alpha(L)$. То есть

$\alpha(L)^{-1}$ является фильтром, удовлетворяющим соотношению

$$\alpha(L)\,\alpha(L)^{-1} = 1. \qquad (6.1.14)$$

При условии, что $\alpha_0 \neq 0$ в $\alpha(L) = \alpha_0 + \alpha_1 L + \cdots$, обращение для $\alpha(L)$ может быть определено для произвольной последовательности $\{\alpha_j\}$, поскольку множество уравнений (6.1.12) с $\delta_0 = 1$, $\delta_j = 0$ $(j = 1, 2, \dots)$ может быть последовательно разрешено относительно $\{\beta_j\}$:

$$\beta_0 = \frac{1}{\alpha_0}, \quad \beta_1 = -\frac{\alpha_1\beta_0}{\alpha_0} = -\frac{\alpha_1}{\alpha_0^2} \quad \text{и т. д.}$$

В качестве примера рассмотрим фильтр $1 - L$. Обратный ему фильтр задается как

$$(1 - L)^{-1} = 1 + L + L^2 + L^3 + \cdots.$$

Последовательность коэффициентов здесь, очевидно, не является абсолютно суммируемой. Этот пример иллюстрирует тот факт, что обратный фильтр может не быть абсолютно суммируемым.

В качестве контрольного вопроса 3 оставляем показать, что

$$\alpha(L)\,\alpha(L)^{-1} = \alpha(L)^{-1}\alpha(L) \quad \text{(поэтому обращение коммутативно)}, \qquad (6.1.15a)$$

$$«\alpha(L)\,\beta(L) = \delta(L)» \Leftrightarrow «\beta(L) = \alpha(L)^{-1}\delta(L)» \Leftrightarrow «\alpha(L) = \delta(L)\,\beta(L)^{-1}», \qquad (6.1.15b)$$

при условии, что $\alpha_0 \neq 0$ и $\beta_0 \neq 0$. Следовательно, например, чтобы решить $\alpha(L)\beta(L) = \delta(L)$ относительно $\beta(L)$, вы «умножаете слева» на обращение $\alpha(L)$. Мы заметили выше, что коммутативность необязательно выполняется для матричных фильтров. Как будет показано в параграфе 6.3, для обращений коммутативность обобщается на матричные фильтры.

Обращение лаговых полиномов

В следующем параграфе о процессах ARMA будет необходимо вычислить обращение лагового полинома p-й степени $\phi(L)$

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p.$$

Такое обращение может быть определено, поскольку $\phi_0 = 1 \neq 0$. Пусть $\psi(L) \equiv \phi(L)^{-1}$. Тогда, по определению,

$$\phi(L)\,\psi(L) = 1.$$

Формула свертки (6.1.12) предлагает алгоритм решения этого уравнения относительно $\psi(L)$. Полагая в этой формуле $\alpha(L) = \phi(L)$, $\beta(L) = \psi(L)$ и $\delta(L) = 1$, получаем:

$$\begin{array}{rlllllll}
\text{константа:} & \psi_0 & & & & & & = 1, \\
L: & \psi_1 & - \phi_1\psi_0 & & & & & = 0, \\
L^2: & \psi_2 & - \phi_1\psi_1 & - \phi_2\psi_0 & & & & = 0, \\
\cdots & & & \cdots & & & & \\
L^{p-1}: & \psi_{p-1} & - \phi_1\psi_{p-2} & - \phi_2\psi_{p-3} & - \cdots & - \phi_{p-1}\psi_0 & & = 0, \\
L^p: & \psi_p & - \phi_1\psi_{p-1} & - \phi_2\psi_{p-2} & - \cdots & - \phi_{p-1}\psi_1 & - \phi_p\psi_0 & = 0, \\
L^{p+1}: & \psi_{p+1} & - \phi_1\psi_p & - \phi_2\psi_{p-1} & - \cdots & - \phi_{p-1}\psi_2 & - \phi_p\psi_1 & = 0, \\
L^{p+2}: & \psi_{p+2} & - \phi_1\psi_{p+1} & - \phi_2\psi_p & - \cdots & - \phi_{p-1}\psi_3 & - \phi_p\psi_2 & = 0, \\
& & & & \cdots & & &
\end{array} \tag{6.1.16}$$

Эти уравнения могут быть легко решены последовательно относительно $(\psi_0, \psi_1, \psi_2, \ldots)$ как

$$\psi_0 = 1, \quad \psi_1 = \phi_1, \quad \psi_2 = \phi_2 + \phi_1^2, \quad \text{и т. д.}$$

Также заметим, что для достаточно большого j (фактически для $j \geq p$) $\{\psi_j\}$ следует однородному разностному уравнению p-го порядка

$$\psi_j - \phi_1\psi_{j-1} - \phi_2\psi_{j-2} - \cdots - \phi_{p-1}\psi_{j-p+1} - \phi_p\psi_{j-p} = 0. \tag{6.1.17}$$

Поэтому, вычислив первые p коэффициентов $(\psi_0, \psi_1, \ldots, \psi_{p-1})$, мы можем использовать это однородное разностное уравнение p-го порядка для порождения оставшихся коэффициентов (ψ_j, $j \geq p$) с этими p коэффициентами в качестве начальных значений.

Как мы знаем из теории однородных разностных уравнений (кратко описанной в аналитическом упражнении 4), последовательность решений $\{\psi_j\}$ уравнения (6.1.17), в конце концов, начинает убывать в геометрической прогрессии, если выполняется то, что известно как **условие стабильности** (*stability condition*). Это условие формулируется следующим образом:

Все корни полиномиального уравнения p-го порядка от z

$$\phi(z) = 0, \quad \text{где} \quad \phi(z) \equiv 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p \tag{6.1.18}$$

больше 1 по абсолютной величине.

Корни полиномиального уравнения $\phi(z) = 0$ могут быть комплексными. Если ваше владение комплексными числами ослабло, напомним, что комплексное число z может быть записано как

$$z = a + bi,$$

где a и b — два действительных числа, а $i = \sqrt{-1}$. Поэтому комплексное число может быть представлено как точка на двумерной диаграмме с горизонтальной осью, измеряющей a (действительная часть), и вертикальной осью, измеряющей b (мнимая часть). Его абсолютное значение равно $|z| = \sqrt{a^2 + b^2}$, то есть расстоянию между началом координат и точкой, представляющей z. Мы говорим, что z лежит вне единичного круга, если $|z| > 1$. Поэтому условие стабильности, сформулированное выше, может быть иначе сформулировано как требование того, что корни указанного уравнения лежат вне единичного круга. Поскольку $z = \lambda$ является решением уравнения $\phi(z) = 0$ всякий раз, когда $z = 1/\lambda$ является решением уравнения $z^p - \phi_1 z^{p-1} - \phi_2 z^{p-2} - \cdots - \phi_{p-1} z - \phi_p = 0$, условие стабильности может быть эквивалентно сформулировано следующим образом:

Все корни полиномиального уравнения p-го порядка

$$z^p - \phi_1 z^{p-1} - \phi_2 z^{p-2} - \cdots - \phi_{p-1} z - \phi_p = 0 \tag{6.1.18$'$}$$

меньше 1 по абсолютному значению (то есть лежат *внутри* единичного круга).

При этом условии, как вам будет предложено доказать (аналитическое упражнение 4), существуют два действительных числа, $A > 0$ и $0 < b < 1$, таких что

$$|\psi_j| < Ab^j \quad \text{для всех } j. \tag{6.1.19}$$

При выполнении (6.1.19) абсолютная суммируемость $\{\psi_j\}$ следует из

$$\sum_{j=0}^{\infty}|\psi_j| < \sum_{j=0}^{\infty} Ab^j = \frac{A}{1-b} < \infty.$$

Таким образом, мы доказали

Утверждение 6.3 (абсолютно суммируемые обращения лаговых полиномов): *Рассмотрим лаговый полином p-й степени $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p$, и положим $\psi(L) \equiv \phi(L)^{-1}$. Если соответствующее полиномиальное уравнение p-й степени $\phi(z) = 0$ удовлетворяет условию стабильности (6.1.18) или (6.1.18′), то последовательность коэффициентов $\{\psi_j\}$ для $\psi(L)$ ограничена по абсолютному значению геометрически убывающей прогрессией (как в* (6.1.19)*) и, следовательно, абсолютно суммируема.*

Для иллюстрации рассмотрим лаговый полином степени единица, $1 - \phi L$. Корень соответствующего полиномиального уравнения $1 - \phi z = 0$ равен $1/\phi$. Условие стабильности заключается в том, что $|1/\phi| > 1$ или $|\phi| < 1$. Фильтр

$$1 + \phi L + \phi^2 L^2 + \cdots \tag{6.1.20}$$

очевидно является обращением рассматриваемого полинома. Последовательность коэффициентов $\{\phi^j\}$ действительно ограничена по абсолютному значению геометрически убывающей последовательностью (например, положим $A = 1.1$ и $b = \phi$).

Контрольные вопросы

1. (Последовательность Коши.) Последовательность действительных чисел, $\{\alpha_n\}$, называют **последовательностью Коши** (*Cauchy sequence*), если $|\alpha_m - \alpha_n| \to 0$ при $m, n \to \infty$. Важный результат математического анализа заключается в том, что α_n сходится к действительному числу тогда и только тогда, когда эта последовательность является последовательностью Коши. Принимая это как данность, покажите, что $\{\gamma_j\}$ абсолютно суммируема тогда и только тогда, когда $\sum_{j=n+1}^{m}|\gamma_j| \to 0$ при $m, n \to \infty$. (Вы можете предположить без потери общности, что $m > n$.)

2. (Как фильтрация изменяет автоковариации.) Пусть $h(L) = h_0 + h_1 L$ и $y_t = h(L)x_t$, где процесс $\{x_t\}$ ковариационно стационарный. Проверьте, что (с автоковариациями процесса $\{x_t\}$, равными $\{\gamma_j^x\}$) автоковариации процесса $\{y_t\}$ задаются соотношением:

$$\gamma_j^y = \sum_{k=0}^{1}\sum_{l=0}^{1} h_k h_l \gamma_{j-k+l}^x.$$

3. Докажите (6.1.15b). **Указание:** *Вот доказательство того, что*

$$«\alpha(L)\,\beta(L) = \delta(L)» \Rightarrow «\beta(L) = \alpha(L)^{-1}\delta(L)»:$$

$$\alpha(L)^{-1}\delta(L) = \alpha(L)^{-1}\alpha(L)\beta(L) = \alpha(L)\alpha(L)^{-1}\beta(L) = \beta(L).$$

4. Удовлетворяет ли $\phi(z) = 1 - \frac{3}{4}z + \frac{9}{16}z^2$ условию стабильности? **Указание:** *Корни* $\phi(z) = 0$ *равны* $2(1 \pm \sqrt{3}i)/3$.

5. (Коэффициенты полиномиальных уравнений.) Рассмотрим полиномиальное уравнение $\phi(z) = 0$, где $\phi(z) = 1 - \phi_1 z - \phi_2 z^2$, и пусть (λ_1, λ_2) являются двумя его корнями. Проверьте, что $\phi_1 = 1/\lambda_1 + 1/\lambda_2$, $\phi_2 = -1/(\lambda_1 \cdot \lambda_2)$. **Указание:** *Если* λ_1 *и* λ_2 *являются двумя корнями для* $\phi(z) = 0$, $\phi(z)$ *может быть записан как* $\phi(z) = (1 - \frac{1}{\lambda_1}z)(1 - \frac{1}{\lambda_2}z)$.

6. (Обращение $1 - \phi L$.) Убедитесь, что (6.1.20) является обращением для $1 - \phi L$, проверяя формулу свертки (6.1.12) с $\delta_0 = 1$, $\delta_j = 0$ для $j \geq 1$.

6.2. Процессы ARMA

После введения понятия фильтров мы можем с легкостью и изяществом изучать класс линейных процессов, называемых процессами ARMA, которые являются параметризацией коэффициентов процесса MA(∞).

AR(1) *и его* MA(∞)*-представление*

Процесс авторегрессии первого порядка (AR(1)) (*first-order autoregressive process*) удовлетворяет следующему стохастическому разностному уравнению:

$$y_t = c + \phi y_{t-1} + \varepsilon_t \quad \text{или} \quad y_t - \phi y_{t-1} = c + \varepsilon_t \quad \text{или} \quad (1 - \phi L)y_t = c + \varepsilon_t, \tag{6.2.1}$$

где $\{\varepsilon_t\}$ является белым шумом. Если $\phi \neq 1$, положим $\mu \equiv c/(1 - \phi)$ и перепишем это уравнение как

$$(y_t - \mu) - \phi \cdot (y_{t-1} - \mu) = \varepsilon_t \quad \text{или} \quad (1 - \phi L)(y_t - \mu) = \varepsilon_t. \tag{6.2.1$'$}$$

Как будет видно через мгновение, именно μ, а не c является математическим ожиданием процесса y_t, если y_t ковариационно стационарный. По этой причине мы будем называть (6.2.1$'$) формой в отклонениях от среднего. Скользящее среднее является последовательными значениями $\{y_t\}$, а не $\{\varepsilon_t\}$. Указанное разностное уравнение называется стохастическим из-за наличия случайной величины ε_t.

Мы ищем ковариационно стационарное решение $\{y_t\}$ этого стохастического разностного уравнения. Это решение зависит от того, является ли $|\phi|$ меньшим, равным или большим, чем 1.

Случай 1: $|\phi| < 1$

Решение можно легко получить, используя обращение $(1 - \phi L)^{-1}$, заданное выражением (6.1.20). Так как этот фильтр абсолютно суммируем, когда $|\phi| < 1$, мы можем применить его к обеим сторонам AR(1)-уравнения (6.2.1′), получая в результате:

$$(1 - \phi L)^{-1}(1 - \phi L)(y_t - \mu) = (1 - \phi L)^{-1}\varepsilon_t.$$

Эта операция легитимна, поскольку благодаря утверждению 6.2 обе стороны этого уравнения определены как пределы в среднем квадратическом. По определению обращений (6.1.14), $(1 - \phi L)(1 - \phi L)^{-1} = 1$, а по коммутативности обращения, $(1 - \phi L)^{-1}(1 - \phi L) = 1$. Следовательно, если $\{y_t\}$ ковариационно стационарный, левая сторона равна $y_t - \mu$ по (6.1.13). Поэтому

$$y_t - \mu = (1 - \phi L)^{-1}\varepsilon_t = (1 + \phi L + \phi^2 L^2 + \cdots)\varepsilon_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$$

$$\text{или} \quad y_t = \mu + \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}. \tag{6.2.2}$$

То, что мы показали, — это то, что если $\{y_t\}$ является ковариационно стационарным решением стохастического разностного уравнения (6.2.1) или (6.2.1′), то y_t имеет **представление в виде скользящего среднего** (*moving-average representation*) как в (6.2.2). Обратно, если процесс y_t имеет представление (6.2.2), то он удовлетворяет указанному разностному уравнению. Это можно легко получить, подставляя (6.2.2) в (6.2.1′):

$$(1 - \phi L)(y_t - \mu) = (1 - \phi L)(1 - \phi L)^{-1}\varepsilon_t =$$
$$= \varepsilon_t \quad (\text{поскольку } (1 - \phi L)(1 - \phi L)^{-1} = 1 \text{ по определению обращения}).$$

Таким образом, процесс $\{y_t\}$, заданный в (6.2.2), является *единственным* ковариационно стационарным решением стохастического дифференциального уравнения, если $|\phi| < 1$. Математическим ожиданием этого процесса, согласно утверждению 6.1(b), является μ.

Условие $|\phi| < 1$, являющееся условием стабильности (6.1.18), связанным с полиномиальным уравнением первой степени $1 - \phi z = 0$, называется **условием стационарности** (*stationarity condition*) в текущем контексте процессов авторегрессии. Интуитивно этот результат говорит о том, что соответствующий процесс, если он начинается очень давно, стабилизируется, при условии, что $|\phi| < 1$, так что влияние прошлого геометрически затухает с течением времени.

Случай 2: $|\phi| > 1$

Сдвигая время вперед на один период (то есть заменяя «t» на «$t+1$»), умножая обе стороны на ϕ^{-1} и производя перестановку, стохастическое разностное уравнение (6.2.1′) можно записать в виде:

$$y_t - \mu = \phi^{-1}(y_{t+1} - \mu) - \phi^{-1}\varepsilon_{t+1}. \tag{6.2.3}$$

Применение того же рассуждения, что и выше, но на этот раз с перемещением в обратном направлении во времени, показывает, что единственным ковариационно стационарным решением является здесь

$$y_t = \mu - \sum_{j=1}^{\infty} \phi^{-j}\varepsilon_{t+j}. \tag{6.2.4}$$

То есть текущее значение y является скользящим средним *будущих* значений ε. Эта бесконечная сумма определена, поскольку последовательность $\{\phi^{-j}\}$ абсолютно суммируема, если $|\phi| > 1$.

Случай 3: $|\phi| = 1$

Рассматриваемое стохастическое разностное уравнение не имеет ковариационно стационарного решения. Например, если $\phi = 1$, это стохастическое разностное уравнение принимает вид:

$$\begin{aligned} y_t &= c + y_{t-1} + \varepsilon_t = \\ &= c + (c + y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t, \quad \text{поскольку } y_{t-1} = c + y_{t-2} + \varepsilon_{t-1} = \\ &= c + (c + (c + y_{t-3} + \varepsilon_{t-2}) + \varepsilon_{t-1}) + \varepsilon_t, \quad \text{и т. д.} \end{aligned}$$

Повторяя этот тип последовательных подстановок j раз, мы получаем:

$$y_t - y_{t-j} = c \cdot j + (\varepsilon_t + \varepsilon_{t-1} + \cdots + \varepsilon_{t-j+1}).$$

(Если $\{\varepsilon_t\}$ является независимым белым шумом, этот процесс является **случайным блужданием со сносом c** (*random walk with drift* c).) Предположим, вопреки нашему заявлению, что этот процесс ковариационно стационарный. Тогда дисперсия $y_t - y_{t-j}$ равна $2(\gamma_0 - \gamma_j)$. Дисперсия правой части равна $\sigma^2 \cdot j$. Поэтому

$$2 \cdot (\gamma_0 - \gamma_j) = \sigma^2 \cdot j \quad \text{или}$$

$$\rho_j = 1 - \frac{\sigma^2}{2\gamma_0} \cdot j < -1 \quad \text{для достаточно больших } j$$

$$(\text{вспомните: } \rho_j = \gamma_j/\gamma_0).$$

Это является противоречием, поскольку коэффициенты автокорреляции не могут быть больше единицы по абсолютному значению. Поэтому процесс $\{y_t\}$ не может быть ковариационно стационарным. Процессы с $\phi = 1$ являются примерами «процессов с единичным корнем» и будут изучаться в главе 9.

Решение (6.1.4) для случая $|\phi| > 1$, в котором текущее значение y связано с будущими значениями порождающего $\{\varepsilon_t\}$, непригодно для экономики, которая не наделяет экономических агентов совершенным предвидением. В дальнейшем, если не указано противное, мы будем использовать термин «AR(1)-процесс» для единственного ковариационно стационарного решения AR(1)-уравнений (6.2.1) или (6.2.1′) при выполнении условия стационарности.

Автоковариации AR(1)

Автоковариации $\{\gamma_j\}$ для AR(1) (для случая $|\phi| < 1$) могут быть вычислены двумя способами. Первый способ заключается в представлении AR(1) в виде MA(∞), как в (6.2.2), и использовании (6.1.7). Поскольку для AR(1) $\psi_j = \phi^j$, мы имеем

$$\begin{aligned} \gamma_j &= (\phi^j + \phi^{j+2} + \phi^{j+4} + \cdots)\sigma^2 = \frac{\phi^j \sigma^2}{1-\phi^2} \\ \text{и} \quad \rho_j &\ (\equiv \gamma_j/\gamma_0) = \phi^j \qquad (j = 0, 1, \ldots). \end{aligned} \tag{6.2.5}$$

Поэтому коррелограмма процесса AR(1) очень простая: она убывает по j с постоянной геометрической скоростью.

Второй способ основан на использовании так называемых **уравнений Юла — Уокера** (*Yule — Walker equations*); см. аналитическое упражнение 5 об использовании уравнений Юла — Уокера при вычислении автоковариаций.

AR(p) *и его* MA(∞)*-представление*

Все результаты, только что полученные для AR(1), могут быть обобщены на AR(p), **процесс авторегрессии p-го порядка** (*p-th order autoregressive process*), который удовлетворяет стохастическому разностному уравнению p-го порядка:

$$\begin{aligned} y_t &= c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t, \quad \text{или} \\ y_t - \phi_1 y_{t-1} - \cdots - \phi_p y_{t-p} &= c + \varepsilon_t, \quad \text{или} \\ \phi(L) y_t = c + \varepsilon_t \quad &\text{с } \phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p, \end{aligned} \tag{6.2.6}$$

где $\{\varepsilon_t\}$ — белый шум и $\phi_p \neq 0$. Очевидно, $\phi(1) = 1 - \phi_1 - \cdots - \phi_p$. Если $\phi(1) \neq 0$, положим $\mu \equiv c/(1-\phi_1 - \cdots - \phi_p) = c/\phi(1)$. Как показано ниже, μ является математическим ожиданием процесса y_t, если y_t является ковариационно стационарным. Выражая отсюда c, AR(p)-уравнение (6.2.6) можно эквивалентно записать в форме отклонений от среднего:

$$(y_t - \mu) - \phi_1 \cdot (y_{t-1} - \mu) - \cdots - \phi_p \cdot (y_{t-p} - \mu) = \varepsilon_t$$
$$\text{или} \quad \phi(L)(y_t - \mu) = \varepsilon_t. \tag{6.2.6$'$}$$

Следующее утверждение является обобщением на процесс AR(p) результата, полученного нами для процесса AR(1).

Утверждение 6.4 (AR(p) как MA(∞) с абсолютно суммируемыми коэффициентами): *Предположим, что полином p-й степени $\phi(z)$ удовлетворяет условию стационарности (стабильности)* (6.1.18) *или* (6.1.18$'$). Тогда

(a) Единственное ковариационно стационарное решение стохастического разностного уравнения p-го порядка (6.2.6) или (6.2.6$'$) имеет MA(∞)-представление

$$y_t = \mu + \psi(L)\varepsilon_t, \quad \psi(L) = \psi_0 + \psi_1 L + \psi_2 L^2 + \psi_3 L^3 + \cdots, \tag{6.2.7}$$

где $\psi(L) = \phi(L)^{-1}$. Последовательность коэффициентов $\{\psi_j\}$ ограничена геометрически убывающей последовательностью и, следовательно, абсолютно суммируема.

(b) Математическое ожидание μ указанного процесса задается как

$$\mu = \phi(1)^{-1}c, \quad \text{где } c \text{ является константой в (6.2.6).} \tag{6.2.8}$$

(c) $\{\gamma_j\}$ ограничены по абсолютному значению последовательностью, которая геометрически убывает по j. Следовательно, автоковариации абсолютно суммируемы.

Если не указано противное, мы будем использовать термин «процесс AR(p)» для единственного ковариационно стационарного решения (6.2.7) AR(p)-уравнения (6.2.6), удовлетворяющего условию стационарности. Абсолютная суммируемость обращения $\phi(L)^{-1}$ в части (a) этого утверждения немедленно следует из утверждения 6.3. При наличии абсолютной суммируемости первая часть (a) может быть доказана с использованием той же аргументации, которую мы использовали для AR(1); просто умножьте обе стороны (6.2.6$'$) на $\phi(L)^{-1}$ и заметьте, что

$$\phi(L)^{-1}\phi(L) = \phi(L)\,\phi(L)^{-1} = 1.$$

Часть (b) немедленно следует из утверждения 6.1(b). Часть (c) можно легко доказать, объединяя (a) и утверждение 6.1(b). В аналитическом упражнении 7 вам будет предложено дать альтернативное доказательство этого утверждения без использования фильтров, но с использованием инструмента под названием **сопровождающая форма** (*companion form*).

Как и для AR(1), автоковариации могут быть получены двумя способами. Первый способ использует представление MA(∞). Коэффициенты $\{\phi_j\}$ в представлении MA являются коэффициентами в обращении лагового полинома $\phi(L)$, а алгоритм для вычисления таких коэффициентов уже был представлен ранее (см. (6.1.16)). При заданных $\{\psi_j\}$ используйте (6.1.7) для вычисления автоковариаций $\{\gamma_j\}$. Второй метод использует уравнения Юла — Уокера.

ARMA(p, q)

Процесс ARMA(p,q) объединяет AR(p) и MA(q):

$$\begin{aligned} y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \\ + \theta_0 \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}, \end{aligned}$$

или

$$\begin{aligned} y_t - \phi_1 y_{t-1} - \phi_2 y_{t-2} - \cdots - \phi_p y_{t-p} = \\ = c + \theta_0 \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}. \end{aligned}$$

или

$$\begin{aligned} \phi(L) y_t = c + \theta(L)\varepsilon_t, \quad \phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p, \\ \theta(L) = \theta_0 + \theta_1 L + \cdots + \theta_q L^q, \end{aligned} \tag{6.2.9}$$

где $\{\varepsilon_t\}$ — белый шум. Если $\phi(1) \neq 0$, положим $\mu = c/\phi(1)$. Форма в отклонениях от среднего имеет вид:

$$\begin{aligned} (y_t - \mu) - \phi_1 \cdot (y_{t-1} - \mu) - \cdots - \phi_p \cdot (y_{t-p} - \mu) = \\ = \theta_0 \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \end{aligned}$$

или

$$\phi(L)(y_t - \mu) = \theta(L)\varepsilon_t. \tag{6.2.9'}$$

Это все еще стохастическое разностное уравнение p-го порядка, но с сериально коррелированным порождающим процессом $\{\theta(L)\varepsilon_t\}$ вместо процесса белого шума $\{\varepsilon_t\}$. Утверждение 6.4 легко обобщается до следующего:

Утверждение 6.5 (ARMA(p,q) как MA(∞) с абсолютно суммируемыми коэффициентами): *Предположим, что полином p-й степени $\phi(z)$ удовлетворяет условию стационарности (стабильности)* (6.1.18) *или* (6.1.18$'$). *Тогда*

(a) Единственное ковариационно стационарное решение стохастического разностного уравнения p-го порядка (6.2.9) *или* (6.2.9$'$) *имеет*

MA(∞)-*представление*

$$y_t = \mu + \psi(L)\varepsilon_t, \quad \psi(L) = \psi_0 + \psi_1 L + \psi_2 L^2 + \psi_3 L^3 + \cdots, \tag{6.2.10}$$

где $\psi(L) \equiv \phi(L)^{-1}\theta(L)$. *Последовательность коэффициентов* $\{\psi_j\}$ *ограничена геометрически убывающей последовательностью и, следовательно, абсолютно суммируема.*

(b) Математическое ожидание μ *указанного процесса задается как*

$$\mu = \phi(1)^{-1}c, \quad \text{где } c \text{ является константой в (6.2.9).} \tag{6.2.11}$$

(c) $\{\gamma_j\}$ *ограничены по абсолютному значению последовательностью, которая геометрически убывает по* j. *Следовательно, автоковариации абсолютно суммируемы.*

Если не указано противное, мы будем использовать термин «процесс ARMA» для единственного решения (6.2.10) ARMA-уравнения (6.2.9) или (6.2.9′), которое удовлетворяет условию стационарности. Представление MA(∞) можно легко получить, умножая обе стороны (6.2.9′) на обращение $\phi(L)^{-1}$ и замечая, что $\phi(L)^{-1}\phi(L) = \phi(L)\phi(L)^{-1} = 1$. Для оставшейся части (a) и части (b) единственная нетривиальная часть доказательства заключается в том, чтобы показать, что $\{\psi_j\}$ ограничена по абсолютному значению геометрически убывающей последовательностью. В случае AR(p) мы использовали алгоритм, основанный на множестве уравнений (6.1.16), чтобы решить $\phi(L)\psi(L) = 1$ относительно $\psi(L)$. Этот алгоритм может быть легко обобщен. На этот раз $\psi(L) \equiv \equiv \phi(L)^{-1}\theta(L)$ или

$$\phi(L)\,\psi(L) = \theta(L).$$

Чтобы решить это относительно $\psi(L)$, применим формулу свертки (6.1.12) с $\alpha(L) = \phi(L)$, $\beta(L) = \psi(L)$ и $\delta(L) = \theta(L)$ (вместо 1). Результирующее множество уравнений то же самое, что и (6.1.16), за исключением того, что правая часть уравнения для константы равна θ_0, а не 1, правая часть уравнения для L равна θ_1, а не 0, правая часть уравнения для L^2 равна θ_2, и так далее, вплоть до уравнения для L^q, правая часть которого равна θ_q. Правые части оставшихся уравнений равны 0. Как и в случае процесса AR(p), эти уравнения могут быть легко разрешены последовательно относительно $(\psi_0, \psi_1, \psi_2, \ldots)$; при $\theta_0 = 1$ имеем:

$$\psi_0 = \theta_0 = 1, \quad \psi_1 = \theta_1 + \phi_1, \quad \psi_2 = \theta_2 + \phi_2 + \theta_1\phi_1 + \phi_1^2 \quad \text{и т. д.}$$

И здесь, как в случае AR(p), для достаточно большого j (фактически для $j \geq \max\{p, q+1\}$, как вы проверите в контрольном вопросе 5), $\{\psi_j\}$ следует тому же самому однородному разностному уравнению p-го порядка (6.1.17). Поэтому $\{\psi_j\}$ ограничена по абсолютной величине геометрически убывающей последовательностью и, следовательно, абсолютно суммируема.

Снова существуют два способа получить выражение для автоковариаций в терминах $(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma^2)$. Вы можете вычислить коэффициенты $\{\psi_j\}$ в МА-представлении как только что описано, а затем использовать (6.1.7). Другой метод основан на уравнениях Юла — Уокера.

ARMA(p, q) *с общими корнями*

Предположим, что в уравнении ARMA(p,q) (6.2.9′) $\phi(L)$ удовлетворяет условию стационарности, как в утверждении 6.5, но предположим, что один из корней $\phi(z) = 0$ является также и корнем $\theta(z) = 0$. Если этот корень равен λ, полиномы $\phi(z)$ и $\theta(z)$ имеют факторизации:

$$\phi(z) = \alpha(z)\,\phi^*(z), \quad \theta(z) = \alpha(z)\,\theta^*(z) \quad \text{с } \alpha(z) = 1 - \lambda^{-1}z.$$

Обращение для $\phi(L)$ равно $\phi^*(L)^{-1}\alpha(L)^{-1}$. Поэтому $\psi(L)$ в утверждении 6.5 можно записать как

$$\psi(L) = \phi(L)^{-1}\theta(L) = \phi^*(L)^{-1}\alpha(L)^{-1}\alpha(L)\,\theta^*(L) = \phi^*(L)^{-1}\theta^*(L), \tag{6.2.12}$$

и это показывает, что уравнение ARMA(p,q) (6.2.9′) и более простое уравнение ARMA($p-1, q-1$)

$$\phi^*(L)(y_t - \mu) = \theta^*(L)\varepsilon_t$$

имеют в качестве единственного ковариационно стационарного решения один и тот же процесс[1]. Из соображений экономности, уравнения ARMA с общими корнями редко используются для параметризации ковариационно стационарных процессов.

Обратимость

Для уравнения ARMA(p,q) (6.2.9) если $\phi(z) = 0$ и $\theta(z) = 0$ не имеют общих корней и если $\theta(z)$ удовлетворяет условию стабильности (6.1.18) или (6.1.18′), то ARMA-процесс называется **обратимым** (*invertible*), а условие стабильности для $\theta(z)$ называется **условием обратимости** (*invertibility condition*)[2]. Поскольку $\theta(L)^{-1}$ абсолютно суммируема, если $\theta(z)$

[1]Поскольку мы определили фильтры для последовательностей действительных чисел, обсуждение в тексте предполагает, что корень λ вещественный. Если λ комплексный, то должен существовать другой общий корень, который является комплексно сопряженным для λ. Пусть $\lambda_1 = a + bi$ и $\lambda_2 = a - bi$ являются такой парой комплексных корней. Тогда, если мы определим

$$\alpha(z) = 1 - (\lambda_1^{-1} + \lambda_2^{-1})z + \frac{1}{(\lambda_1 \cdot \lambda_2)}z^2 = 1 - \frac{2a}{(a^2 + b^2)}z + \frac{1}{(a^2 + b^2)}z^2.$$

приведенное в тексте обсуждение сохраняется.

[2]Легко спутать обратимость и существование обращения. Поскольку $\theta_0 = 1 \neq 0$, $\theta(L)$ имеет обращение без условия обратимости. Обращение, однако, может не быть абсолютно суммируемым, если не выполнено условие обратимости.

удовлетворяет условию обратимости (стабильности), мы можем умножить обе стороны ARMA уравнения (6.2.9) на $\theta(L)^{-1}$ и использовать соотношение $\theta(L)^{-1}c = c\theta(L)^{-1} = c/\theta(1)$, чтобы получить

$$\theta(L)^{-1}\phi(L)y_t = \frac{c}{\theta(1)} + \varepsilon_t. \qquad (6.2.13)$$

Таким образом, рассматриваемый процесс имеет представление в виде процесса авторегрессии бесконечного порядка (AR(∞)). Его получение не требует выполнения условия стационарности (того, что $\phi(z)$ удовлетворяет условию стабильности). Если оба $\phi(L)$ и $\theta(L)$ удовлетворяют условию стабильности, то процесс ARMA(p, q) имеет и MA(∞), и AR(∞)-представления.

Производящая функция автоковариаций и спектр

Весьма полезным способом компактной записи информации о последовательности автоковариаций $\{\gamma_j\}$ ковариационно стационарного процесса $\{y_t\}$ является **производящая функция автоковариаций** (*autocovariance-generating function*):

$$\begin{aligned} g_Y(z) &\equiv \sum_{j=-\infty}^{\infty} \gamma_j z^j = \\ &= \gamma_0 + \sum_{j=1}^{\infty} \gamma_j \cdot (z^j + z^{-j}) \quad (\text{так как } \gamma_{-j} = \gamma_j). \end{aligned} \qquad (6.2.14)$$

Аргументом этой функции (z) является комплексный скаляр. Поскольку суммирование включает бесконечно много слагаемых, возникает вопрос: определена ли такая функция. Известный результат из математического анализа говорит о том, что указанная бесконечная сумма определена по крайней мере для $|z| = 1$ (единичная окружность), если $\{\gamma_j\}$ абсолютно суммируема. Например, для $z = 1$

$$g_Y(1) = \sum_{j=-\infty}^{\infty} \gamma_j = \gamma_0 + 2\sum_{j=1}^{\infty} \gamma_j. \qquad (6.2.15)$$

Это выражение очевидно определено вследствие сходимости бесконечной суммы[1].

Любое комплексное число на единичной окружности может быть представлено как

$$z = \cos(\omega) - i\sin(\omega) = e^{-i\omega}, \qquad (6.2.16)$$

[1]Факт: Если $\{\gamma_j\}$ абсолютно суммируема, то она суммируема (то есть частичная сумма $\{\gamma_j\}$ сходится к конечному пределу).

где $i = \sqrt{-1}$ и ω является взятым со знаком «минус» углом в радианах, который z образует с действительной осью. Если производящая функция автоковариаций вычислена для этого z и разделена на 2π, результирующая функция от ω,

$$s_Y(\omega) \equiv \frac{1}{2\pi} g_Y(\cos(\omega) - i\sin(\omega)) = \frac{1}{2\pi} g_Y(e^{-i\omega}) \tag{6.2.17}$$

называется **(энергетическим) спектром** (*(power) spectrum*), или **спектральной плотностью, функцией спектральной плотности** (*spectral density (function)*)-процесса $\{y_t\}$, а ω называют **частотой** (*frequency*).

Можно показать, что в случае абсолютно суммируемых автоковариаций все эти автоковариации могут быть вычислены на основе спектра (отсюда и термин «производящая функция автоковариаций»). Поэтому существует взаимооднозначное отображение между последовательностью $\{\gamma_j\}$ и $g_Y(z)$ или $s_Y(\omega)$. Таким образом, **подход во временной области** (*time domain approach*), который концентрируется на последовательности $\{\gamma_j\}$ и представляет именно то, что мы делали до сих пор, и **подход в частотной области** (frequency domain approach), который основан на интерпретации спектра, *эквивалентны*, хотя некоторые результаты могут быть более легко установлены при одном подходе, нежели при другом. В этой книге подход в частотной области не изучается далее того, что уже сказано.

Если $\{y_t\}$ — белый шум, то очевидно, что $g_Y(z)$ постоянна и равна дисперсии. Для MA(1) $g_Y(z)$ получается простой подстановкой (6.1.2) для $q = 1$ в определение (6.2.14):

$$\begin{aligned} g_Y(z) &= \gamma_{-1}z^{-1} + \gamma_0 + \gamma_1 z = \\ &= (\theta_1\sigma^2)z^{-1} + (1+\theta_1^2)\sigma^2 + (\theta_1\sigma^2)z = \\ &= \sigma^2 \cdot [\theta_1 z^{-1} + (1+\theta_1^2) + \theta_1 z] = \\ &= \sigma^2 \cdot (1+\theta_1 z)(1+\theta_1 z^{-1}). \end{aligned}$$

Поэтому

$$g_Y(z) = \sigma^2\theta(z)\,\theta(z^{-1}). \tag{6.2.18}$$

где $\theta(z) = 1 + \theta_1 z$. Это последнее выражение обобщается на случай MA(q), где $\theta(z) = 1 + \theta_1 z + \theta_2 z^2 + \dots + \theta_q z^q$. (Вы можете проверить это путем проведения умножения в этом выражении для $g_Y(z)$ в случае MA(q), приводя подобные по степеням z и наблюдая коэффициенты при z^j и z^{-j}.) Это выражение также годится и для процесса MA(∞), если его коэффициенты абсолютно суммируемы, как нам гарантирует следующий результат.

Утверждение 6.6 (производящая функция автоковариаций для профильтрованного процесса): *Пусть $\{\varepsilon_t\}$ является белым шумом и пусть $\psi(L) = \psi_0 + \psi_1 L + \psi_2 L^2 + \cdots$ является абсолютно суммируемым фильтром (то есть последовательность $\{\psi_j\}$ абсолютно суммируема). Тогда* *производящая функция автоковариаций* $\mathrm{MA}(\infty)$*-процесса* $\{y_t\}$, *где* $y_t = \mu + \psi(L)\varepsilon_t$, *равна:*

$$g_Y(z) = \sigma^2\, \psi(z)\, \psi(z^{-1}). \tag{6.2.19}$$

В более общем случае, пусть $\{x_t\}$ *является ковариационно стационарным процессом с абсолютно суммируемыми автоковариациями и* $g_X(z)$ *является производящей функцией автоковариаций для* $\{x_t\}$. *Производящая функция автоковариаций профильтрованного ряда* $\{y_t\}$, *где* $y_t = h(L)x_t$ *(с абсолютно суммируемой* $\{h_j\}$*) равна:*

$$g_Y(z) = h(z)\, g_X(z)\, h(z^{-1}). \tag{6.2.20}$$

Мы не будем доказывать этот результат; см., например, теорему 4.3.1 в [Fuller, 1996].

Поскольку процессы $\mathrm{AR}(p)$ и $\mathrm{ARMA}(p,q)$ имеют представление $\mathrm{MA}(\infty)$, их производящие функции автоковариаций могут быть легко получены из (6.2.19). Так как $\psi(L) = 1/\phi(L)$ для $\mathrm{AR}(p)$ и $\psi(L) = \theta(L)/\phi(L)$ для $\mathrm{ARMA}(p,q)$, мы имеем

$$\mathrm{AR}(p)\colon\ g_Y(z) = \sigma^2 \frac{1}{\phi(z)\,\phi(z^{-1})}. \tag{6.2.21}$$

$$\mathrm{ARMA}(p,q)\colon\ g_Y(z) = \sigma^2 \frac{\theta(z)\,\theta(z^{-1})}{\phi(z)\,\phi(z^{-1})}. \tag{6.2.22}$$

Контрольные вопросы

1. Если $\{y_t\}$ является ковариационно стационарным решением $\mathrm{AR}(1)$-уравнения (6.2.1) с $|\phi| < 1$, то что представляет собой

$$\widehat{\mathrm{E}}^*(y_t\,|\,1, y_{t-1})$$

(проекция методом наименьших квадратов)? **Указание:** $\mathrm{E}(y_{t-1}\varepsilon_t) = 0$.

Зависят ли коэффициенты этой проекции от t? Верно ли, что $\widehat{\mathrm{E}}^*(y_t|1, y_{t-1}) = \mathrm{E}(y_t|y_{t-1})$? **Указание:** $\mathrm{E}(y_{t-1}\varepsilon_t) = 0$, *но означает ли это, что* $\mathrm{E}(\varepsilon_t|y_{t-1}) = 0$?

Что представляет собой $\widehat{\mathrm{E}}^*(y_t|1, y_{t-1}, y_{t-2})$? Предположим теперь, что $\{y_t\}$ — ковариационно стационарное решение уравнения $\mathrm{AR}(1)$, когда $|\phi| > 1$. Верно ли, что $\mathrm{E}(y_{t-1}\varepsilon_t) = 0$? [Ответ: Нет.] Верно ли, что $\widehat{\mathrm{E}}^*(y_t|1, y_{t-1}) = c + \phi y_{t-1}$, как было в случае $|\phi| < 1$?

2. ($\mathrm{AR}(1)$ с $\phi = -1$.) Для простоты предположим, что $c = 0$ в $\mathrm{AR}(1)$-уравнении (6.2.1). Покажите, что $\mathrm{AR}(1)$-уравнение с $\phi = -1$ не имеет ковариационно стационарного решения. **Указание:** *Тот же самый вид последовательной подстановки, который мы использовали для случая* $\phi = 1$, *приводит к соотношению:*

$$y_t - (-1)^j y_{t-j} = \varepsilon_t - \varepsilon_{t-1} + \cdots + (-1)^{j-1}\varepsilon_{t-j+1}.$$

Из этого соотношения получаем:

$$(-1)^j \rho_j = 1 - \frac{\sigma^2}{2\gamma_0} \cdot j.$$

3. (Об утверждении 6.4.) Откуда мы знаем, что $\phi(1) \neq 0$ в (6.2.8)? **Указание:** *Используйте условие стационарности.*

   Докажите часть (b) утверждения 6.4. **Указание:** *Возьмите математическое ожидание от обеих сторон* (6.2.6) *и используйте ковариационную стационарность процесса* $\{y_t\}$.

4. (Об утверждении 6.5.) Проверьте, что $y_t - \mu = \phi(L)^{-1}\theta(L)\varepsilon_t$ является решением уравнения ARMA(p, q).

5. ($\{\psi_j\}$ для ARMA(p, q).) В случае AR(p) мы использовали уравнение (6.1.12) для вычисления MA-коэффициентов $\{\psi_j\}$. Запишите соответствующие уравнения для ARMA$(3, 1)$. С какого лага j стартует последовательность $\{\psi_j\}$, следующая соотношению:

$$\psi_j - \phi_1\psi_{j-1} - \phi_2\psi_{j-2} - \phi_3\psi_{j-3} = 0,$$

   что является (6.1.17) для $p = 3$? Сделайте то же самое для ARMA$(1, 2)$. С какого лага j стартует последовательность $\{\psi_j\}$, следующая соотношению (6.1.17) для $p = 1$? [Ответ: $j = 3$.]

6. (Как фильтрация изменяет спектр.) Пусть фильтр $h(L)$ абсолютно суммируем. По утверждению 6.2, если процесс $\{x_t\}$ является ковариационно стационарным с абсолютно суммируемыми автоковариациями, то это выполняется и для процесса $\{y_t\}$, где $y_t = h(L)x_t$. Используя (6.2.20), проверьте, что

$$s_Y(\omega) = h(e^{-i\omega})s_X(\omega)h(e^{i\omega}).$$

7. (Спектр является вещественнозначным.) Покажите, что спектр является вещественнозначным. **Указание:** *Подставьте* (6.2.16) *в* (6.2.14). *Используйте следующие факты из комплексного анализа:* $[\cos(\omega) - i\sin(\omega)]^j = \cos(j\omega) - i\sin(j\omega)$; $\sin(-\omega) = -\sin(\omega)$.

6.3. Векторные процессы

Все понятия, введенные до сих пор, легко обобщить на векторные процессы.

Векторный процесс белого шума (*vector white noise process*) $\{\boldsymbol{\varepsilon}_t\}$ является совместно ковариационно стационарным процессом, удовлетворяющим соотношениям

$$\mathrm{E}(\boldsymbol{\varepsilon}_t) = \mathbf{0}, \quad \mathrm{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \boldsymbol{\Omega} \text{ (положительно определена)}$$
$$\text{и} \quad \mathrm{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_{t-j}') = \mathbf{0} \quad \text{для } j \neq 0. \tag{6.3.1}$$

Поскольку $\boldsymbol{\Omega}$ не ограничена диагональным видом, между элементами $\boldsymbol{\varepsilon}_t$ могут быть одновременные корреляции. Полная корреляция между элементами $\boldsymbol{\varepsilon}_t$ исключается, поскольку от матрицы $\boldsymbol{\Omega}$ требуется положительная определенность.

Векторный MA(∞)**-процесс** (*vector* MA(∞) *process*) является очевидной векторной версией (6.1.6):

$$\boldsymbol{y}_t = \boldsymbol{\mu} + \sum_{j=0}^{\infty} \boldsymbol{\Psi}_j \boldsymbol{\varepsilon}_{t-j} \quad \text{с } \boldsymbol{\Psi}_0 = \boldsymbol{I}, \tag{6.3.2}$$

где $\{\boldsymbol{\Psi}_j\}$ — квадратные матрицы. Последовательность матриц коэффициентов называется абсолютно суммируемой, если каждый элемент абсолютно суммируем. То есть, если ψ_{klj} является (k,l)-м элементом $\boldsymbol{\Psi}_j$,

$$«\{\boldsymbol{\Psi}_j\} \text{ — абсолютно суммируема»} \Leftrightarrow «\sum_{j=0}^{\infty} |\psi_{k\ell j}| < \infty \text{ для всех } (k,\ell)». \tag{6.3.3}$$

При таком определении утверждение 6.1 очевидным образом обобщается на многомерный случай. В частности, если $\boldsymbol{\Gamma}_j (\equiv \mathrm{E}[(\boldsymbol{y}_t - \boldsymbol{\mu})(\boldsymbol{y}_{t-j} - \boldsymbol{\mu})'])$ — автоковариационная матрица j-го порядка, то выражение для автоковариаций в части (b) утверждения 6.1 принимает вид:

$$\boldsymbol{\Gamma}_j = \sum_{k=0}^{\infty} \boldsymbol{\Psi}_{j+k} \boldsymbol{\Omega} \boldsymbol{\Psi}_k' \quad (j = 0, 1, 2, \ldots). \tag{6.3.4}$$

(Эта формула распространяется также и на $j = -1, -2, \ldots$, поскольку $\boldsymbol{\Gamma}_{-j} = \boldsymbol{\Gamma}_j'$.)

Многомерный фильтр (*multivariate filter*) может быть записан как

$$\boldsymbol{H}(L) = \boldsymbol{H}_0 + \boldsymbol{H}_1 L + \boldsymbol{H}_2 L^2 + \cdots,$$

где $\{\boldsymbol{H}_j\}$ — последовательность (необязательно квадратных) матриц. Поэтому, если H_{klj} является (k,l)-м элементом $\boldsymbol{H}_j$, (k,l)-й элемент $\boldsymbol{H}(L)$ равен:

$$H_{kl0} + H_{kl1} L + H_{kl2} L^2 + \cdots.$$

Многомерная версия утверждения 6.2 очевидна с $\boldsymbol{y}_t = \boldsymbol{H}(L)\boldsymbol{x}_t$.

Произведение фильтров

Пусть $\boldsymbol{A}(L)$ и $\boldsymbol{B}(L)$ — два фильтра, где $\boldsymbol{A}_j$ имеет размер $m \times r$, а $\boldsymbol{B}_j$ имеет размер $r \times s$, так что можно определить произведение матриц $\boldsymbol{A}_j \boldsymbol{B}_k$. Произведение двух фильтров, $\boldsymbol{D}(L) = \boldsymbol{A}(L)\,\boldsymbol{B}(L)$, является фильтром размера $m \times n$, последовательность матриц коэффициентов которого $\{\boldsymbol{D}_j\}$ задается многомерной версией формулы свертки (6.1.12):

$$\begin{gathered}
\boldsymbol{A}_0 \boldsymbol{B}_0 = \boldsymbol{D}_0, \\
\boldsymbol{A}_0 \boldsymbol{B}_1 + \boldsymbol{A}_1 \boldsymbol{B}_0 = \boldsymbol{D}_1, \\
\boldsymbol{A}_0 \boldsymbol{B}_2 + \boldsymbol{A}_1 \boldsymbol{B}_1 + \boldsymbol{A}_2 \boldsymbol{B}_0, = \boldsymbol{D}_2 \\
\cdots \\
\boldsymbol{A}_0 \boldsymbol{B}_j + \boldsymbol{A}_1 \boldsymbol{B}_{j-1} + \boldsymbol{A}_2 \boldsymbol{B}_{j-2} + \cdots + \boldsymbol{A}_{j-1} \boldsymbol{B}_1 + \boldsymbol{A}_j \boldsymbol{B}_0 = \boldsymbol{D}_j, \\
\cdots .
\end{gathered} \tag{6.3.5}$$

Обращение

Пусть $\boldsymbol{A}(L)$ и $\boldsymbol{B}(L)$ — два фильтра, матрицы коэффициентов которых квадратные. $\boldsymbol{B}(L)$ называется **обратным** (*inverse*) для $\boldsymbol{A}(L)$ и обозначается $\boldsymbol{A}(L)^{-1}$, если

$$\boldsymbol{A}(L)\,\boldsymbol{B}(L) = \boldsymbol{I}. \tag{6.3.6}$$

Для любой произвольной последовательности квадратных матриц $\{\boldsymbol{A}_j\}$ обращение для $\boldsymbol{A}(L)$ существует, если $\boldsymbol{A}_0$ невырождена. Легко показать, что обращения коммутативны (см. контрольный вопрос 2), поэтому

$$\boldsymbol{A}(L)\,\boldsymbol{A}(L)^{-1} = \boldsymbol{A}(L)^{-1}\boldsymbol{A}(L).$$

Абсолютно суммируемые обращения лаговых полиномов

Лаговый полином p-й степени (*p-th degree lag matrix polynomial*) имеет вид:

$$\boldsymbol{\Phi}(L) = \boldsymbol{I}_r - \boldsymbol{\Phi}_1 L - \boldsymbol{\Phi}_2 L^2 - \cdots - \boldsymbol{\Phi}_p L^p, \tag{6.3.7}$$

где $\{\boldsymbol{\Phi}_j\}$ является последовательностью $r \times r$ матриц с $\boldsymbol{\Phi}_p \neq \mathbf{0}$. Как следует из теории однородных разностных уравнений, условие стабильности для многомерного случая заключается в следующем:

Все корни детерминантного уравнения

$$|\boldsymbol{I} - \boldsymbol{\Phi}_1 z - \cdots - \boldsymbol{\Phi}_p z^p| = 0 \tag{6.3.8}$$

больше единицы по абсолютной величине (то есть лежат вне единичного круга).

Или эквивалентно

Все корни детерминантного уравнения

$$|\boldsymbol{I} z^p - \boldsymbol{\Phi}_1 z^{p-1} - \cdots - \boldsymbol{\Phi}_p| = 0 \tag{6.3.8'}$$

меньше единицы по абсолютному значению (то есть лежат внутри единичного круга).

В качестве примера рассмотрим лаговый полином первого порядка $\mathbf{\Phi}(L) = \boldsymbol{I} - \mathbf{\Phi}_1 L$, где $\mathbf{\Phi}_1 = (\phi_{kl})$ является 2×2-матрицей. Уравнение (6.3.8) может быть записано как

$$\begin{vmatrix} 1 - \phi_{11}z & -\phi_{12}z \\ -\phi_{21}z & 1 - \phi_{22}z \end{vmatrix} = 1 - (\phi_{11} + \phi_{22})z + (\phi_{11}\phi_{22} - \phi_{21}\phi_{12})z^2 = 0.$$

С обобщенным таким образом условием стабильности утверждение 6.3 обобщается очевидным образом. Пусть $\mathbf{\Psi}(L) = \mathbf{\Phi}(L)^{-1}$. Каждая компонента последовательности матриц коэффициентов $\{\mathbf{\Psi}_j\}$ будет ограничена по абсолютному значению геометрически убывающей последовательностью.

Многомерным аналогом процесса $\mathrm{AR}(p)$ является **процесс векторной авторегрессии p-го порядка ($\mathrm{VAR}(p)$)** (*vector autoregressive process of p-th order*). Он является единственным ковариационно стационарным решением при условии стационарности (6.3.8) для следующего векторного стохастического разностного уравнения:

$$\begin{gathered} \boldsymbol{y}_t - \mathbf{\Phi}_1 \boldsymbol{y}_{t-1} - \dots - \mathbf{\Phi}_p \boldsymbol{y}_{t-p} = \boldsymbol{c} + \boldsymbol{\varepsilon}_t \quad \text{или} \quad \mathbf{\Phi}(L)(\boldsymbol{y}_t - \boldsymbol{\mu}) = \boldsymbol{\varepsilon}_t, \\ \text{где} \quad \mathbf{\Phi}(L) = \boldsymbol{I}_r - \mathbf{\Phi}_1 L - \mathbf{\Phi}_2 L^2 - \dots - \mathbf{\Phi}_p L^p \quad \text{и} \quad \boldsymbol{\mu} = \mathbf{\Phi}(1)^{-1}\boldsymbol{c}. \end{gathered} \tag{6.3.9}$$

причем $\mathbf{\Phi}_p \neq \mathbf{0}$. Например, двумерный процесс $\mathrm{VAR}(1)$ может быть записан в полном объеме как система двух уравнений с общими регрессорами:

$$\begin{aligned} y_{1t} &= c_1 + \phi_{11} y_{1,t-1} + \phi_{12} y_{2,t-1} + \varepsilon_{1t}, \\ y_{2t} &= c_2 + \phi_{21} y_{1,t-1} + \phi_{22} y_{2,t-1} + \varepsilon_{1t}, \quad \text{где } \mathbf{\Phi}_1 = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}. \end{aligned}$$

В более общем случае m-мерный $\mathrm{VAR}(p)$ описывается множеством из m уравнений с $Mp + 1$ общими регрессорами. Утверждение 6.4 непосредственно обобщается на многомерный случай. В частности, каждый элемент $\mathbf{\Gamma}_j$ ограничен по абсолютному значению геометрически убывающей последовательностью.

Векторные авторегрессии являются очень популярным инструментом для анализа динамических взаимосвязей между ключевыми макропеременными. Глубокий анализ можно найти в [Hamilton, 1994, Chapter 11].

Многомерным аналогом $\mathrm{ARMA}(p, q)$ служит **векторный $\mathrm{ARMA}(p, q)$ ($\mathrm{VARMA}(p, q)$)** (*vector ARMA(p,q)*). Он является единственным ковариационно стационарным решением при условии стационарности для следующего векторного стохастического разностного уравнения:

$$\begin{gathered}\boldsymbol{y}_t = \boldsymbol{c} + \boldsymbol{\Phi}_1\boldsymbol{y}_{t-1} + \boldsymbol{\Phi}_2\boldsymbol{y}_{t-2} + \cdots + \boldsymbol{\Phi}_p\boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t + \\ + \boldsymbol{\Theta}_1\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\Theta}_2\boldsymbol{\varepsilon}_{t-2} + \cdots + \boldsymbol{\Theta}_q\boldsymbol{\varepsilon}_{t-q} \quad \text{или} \\ \boldsymbol{\Phi}(L)(\boldsymbol{y}_t - \boldsymbol{\mu}) = \boldsymbol{\Theta}(L)\boldsymbol{\varepsilon}_t,\end{gathered}$$

где

$$\boldsymbol{\Phi}(L) = \boldsymbol{I} - \boldsymbol{\Phi}_1 L - \cdots - \boldsymbol{\Phi}_p L^p, \quad \boldsymbol{\Theta}(L) = \boldsymbol{I} + \boldsymbol{\Theta}_1 L + \cdots + \boldsymbol{\Theta}_q L^q,$$
$$\boldsymbol{\mu} = \boldsymbol{\Phi}(1)^{-1}\boldsymbol{c}, \tag{6.3.10}$$

и при этом $\boldsymbol{\Phi}(L)_p \neq \boldsymbol{0}$ и $\boldsymbol{\Theta}_q \neq \boldsymbol{0}$. Утверждение 6.5 обобщается очевидным способом.

Производящая функция автоковариаций

Многомерная версия производящей функции автоковариаций для векторного ковариационно стационарного процесса $\{\boldsymbol{y}_t\}$ имеет вид:

$$\boldsymbol{G_Y}(z) = \sum_{j=-\infty}^{\infty} \boldsymbol{\Gamma}_j z^j = \boldsymbol{\Gamma}_0 + \sum_{j=1}^{\infty}(\boldsymbol{\Gamma}_j z^j + \boldsymbol{\Gamma}'_j z^{-j}) \quad (\text{поскольку } \boldsymbol{\Gamma}_{-j} = \boldsymbol{\Gamma}'_j). \tag{6.3.11}$$

Спектр векторного процесса $\{\boldsymbol{y}_t\}$ определяется как

$$\boldsymbol{s_Y}(\omega) = \frac{1}{2\pi}\boldsymbol{G_Y}(e^{-i\omega}). \tag{6.3.12}$$

Утверждение 6.6 легко обобщается. В частности, если $\boldsymbol{H}(L)$ является абсолютно суммируемым фильтром $r \times s$, а $\boldsymbol{G_X}(z)$ является производящей функцией автоковариаций s-мерного ковариационно стационарного процесса $\{\boldsymbol{x}_t\}$ с абсолютно суммируемыми автоковариациями, то производящая функция автоковариаций для $\boldsymbol{y}_t = \boldsymbol{H}(L)\boldsymbol{x}_t$ задается как

$$\underset{(r\times r)}{\boldsymbol{G_Y}(z)} = \underset{(r\times s)}{\boldsymbol{H}(z)}\,\underset{(s\times s)}{\boldsymbol{G_x}(z)}\,\underset{(s\times r)}{\boldsymbol{H}(z^{-1})'}. \tag{6.3.13}$$

В качестве частных случаев этого соотношения мы имеем:

$$\text{векторный белый шум}: \boldsymbol{G_Y}(z) = \boldsymbol{\Omega}. \tag{6.3.14}$$

$$\text{VMA}(\infty): \boldsymbol{G_Y}(z) = \boldsymbol{\Psi}(z)\,\boldsymbol{\Omega}\,\boldsymbol{\Psi}(z^{-1})'. \tag{6.3.15}$$

$$\text{VAR}(p): \boldsymbol{G_Y}(z) = [\boldsymbol{\Phi}(z)^{-1}]\,\boldsymbol{\Omega}\,[\boldsymbol{\Phi}(z^{-1})^{-1}]'. \tag{6.3.16}$$

$$\text{VARMA}(p,q): \boldsymbol{G_Y}(z) = [\boldsymbol{\Phi}(z)^{-1}]\,[\boldsymbol{\Theta}(z)]\,\boldsymbol{\Omega}\,[\boldsymbol{\Theta}(z^{-1})]'\,[\boldsymbol{\Phi}(z^{-1})^{-1}]'. \tag{6.3.17}$$

Например, если $\boldsymbol{y}_y = \boldsymbol{\varepsilon}_t + \boldsymbol{\Theta}_1\boldsymbol{\varepsilon}_{t-1}$,

$$\boldsymbol{G_Y}(z) = (\boldsymbol{I} + \boldsymbol{\Theta}_1 z)\boldsymbol{\Omega}(\boldsymbol{I} + \boldsymbol{\Theta}'_1 z^{-1}).$$

Контрольные вопросы

1. Проверьте (6.3.4) для векторного MA(1).

2. (Коммутативность обращений.) Проверьте коммутативность обращений, то есть «$\boldsymbol{A}(L)\,\boldsymbol{B}(L) = \boldsymbol{I}$» $\Leftrightarrow$ «$\boldsymbol{B}(L)\,\boldsymbol{A}(L) = \boldsymbol{I}$» для двух квадратных фильтров размерности r с $|\boldsymbol{A}_0| \neq 0$ и $|\boldsymbol{B}_0| \neq 0$. **Указание:** *Положите $\boldsymbol{D}_0 = \boldsymbol{I}$, $\boldsymbol{D}_j = \boldsymbol{0}$ ($j \geq 1$) в системе (6.3.5) и решите ее относительно $\boldsymbol{B}$. Покажите, что этот $\boldsymbol{B}(L)$ удовлетворяет соотношению $\boldsymbol{B}(L)\,\boldsymbol{A}(L) = \boldsymbol{I}$.*

3. (Отсутствие коммутативности для многомерных фильтров.) Предложите пример, где $\boldsymbol{A}(L)$ и $\boldsymbol{B}(L)$ — 2×2, но $\boldsymbol{A}(L)\,\boldsymbol{B}(L) \neq \boldsymbol{B}(L)\,\boldsymbol{A}(L)$. **Указание:** *Как насчет $\boldsymbol{A}(L) = \boldsymbol{I} + \boldsymbol{A}_1 L$ и $\boldsymbol{B}(L) = \boldsymbol{I} + \boldsymbol{B}_1 L$? Найдите две квадратные матрицы $\boldsymbol{A}_1$ и $\boldsymbol{B}_1$, такие что $\boldsymbol{A}_1\,\boldsymbol{B}_1 \neq \boldsymbol{B}_1\,\boldsymbol{A}_1$.*

4. Проверьте многомерную версию (6.1.15b), а именно:
$$«\boldsymbol{A}(L)\,\boldsymbol{B}(L) = \boldsymbol{D}(L)» \Leftrightarrow «\boldsymbol{B}(L) = \boldsymbol{A}(L)^{-1}\boldsymbol{D}(L)» \Leftrightarrow$$
$$\Leftrightarrow «\boldsymbol{A}(L) = \boldsymbol{D}(L)\,\boldsymbol{B}(L)^{-1}»,$$
при условии невырожденности $\boldsymbol{A}_0$ и $\boldsymbol{B}_0$. Поэтому, чтобы решить $\boldsymbol{A}(L)\,\boldsymbol{B}(L) = \boldsymbol{D}(L)$, «умножьте обе стороны слева» на $\boldsymbol{A}(L)^{-1}$.

5. Покажите, что $[\boldsymbol{A}(L)\,\boldsymbol{B}(L)]^{-1} = \boldsymbol{B}(L)^{-1}\,\boldsymbol{A}(L)^{-1}$ при условии невырожденности $\boldsymbol{A}_0$ и $\boldsymbol{B}_0$. **Указание:** *По определению,*
$$\boldsymbol{A}(L)\,\boldsymbol{B}(L)[\boldsymbol{A}(L)\,\boldsymbol{B}(L)]^{-1} = \boldsymbol{I}.$$

6.4. Оценивание авторегрессий

Авторегрессионные процессы (авторегрессии) популярны в эконометрике, не только потому что они имеют естественную интерпретацию, но также потому, что они легко оцениваются. Этот параграф рассматривает оценивание авторегрессий. Оценивание ARMA-процессов будет затронуто в конце параграфа.

Оценивание AR(1)

Вспомним, что AR(1) является MA(∞) с абсолютно суммируемыми коэффициентами. Поэтому, если $\{\varepsilon_t\}$ является независимым белым шумом (последовательностью i.i.d. случайных величин с нулевым средним и конечной дисперсией), он является строго стационарным и эргодическим (утверждение 6.1(d)). Полагая $\boldsymbol{x}_t = (1, y_{t-1})'$ и $\boldsymbol{\beta} = (c, \phi)'$, AR(1)-уравнение (6.2.1) можно записать как уравнение регрессии
$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + \varepsilon_t. \tag{6.4.1}$$

Мы предполагаем, что выборка состоит из $(y_0, y_1, y_2, \ldots, y_n)$, включая y_0, так что (6.4.1) для $t = 1$ (которое имеет y_0 в правой части) может быть включено в оценивание. Поэтому период выборки с $t = 1$ по n.

Мы теперь покажем, что все условия утверждения 2.5 об асимптотических свойствах OLS-оценки для β при наличии условной гомоскедастичности здесь удовлетворяются. Во-первых, очевидно, что линейность (предположение 2.1) удовлетворяется. Во-вторых, как только что было замечено, процесс $\{y_t, \boldsymbol{x}_t\}$ совместно стационарный и эргодический (предположение 2.2). В-третьих, поскольку y_{t-1} (непостоянный регрессор в $\boldsymbol{x}_t$) является функцией от $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}$, он не зависит от ошибки ε_t по предположению i.i.d. для $\{\varepsilon_t\}$. Таким образом,

$$\mathrm{E}(\varepsilon_t^2 | \boldsymbol{x}_t) = \sigma^2, \tag{6.4.2}$$

то есть ошибка условно гомоскедастична (предположение 2.7). В-четвертых, если $\boldsymbol{g}_t \equiv \boldsymbol{x}_t \cdot \varepsilon_t = (\varepsilon_t, y_{t-1}\varepsilon_t)'$, то $\{\boldsymbol{g}_t\}$ является мартингал-разностью, как требуется в предположении 2.5, поскольку

$$\begin{aligned}
&\text{первый элемент } \mathrm{E}(\boldsymbol{g}_t | \boldsymbol{g}_{t-1}, \boldsymbol{g}_{t-2}, \dots) = \\
&= \mathrm{E}(\varepsilon_t | \boldsymbol{g}_{t-1}, \boldsymbol{g}_{t-2}, \dots) = \\
&= \mathrm{E}(\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, y_{t-2}\varepsilon_{t-1}, y_{t-3}\varepsilon_{t-2}, \dots) = \\
&= 0 \quad (\text{поскольку } \{\varepsilon_t\} \text{ является i.i.d. и } y_{t-j} \text{ является функцией от} \\
&\qquad (\varepsilon_{t-j}, \varepsilon_{t-j-1}, \dots))
\end{aligned}$$

и

$$\begin{aligned}
&\text{второй элемент } \mathrm{E}(\boldsymbol{g}_t | \boldsymbol{g}_{t-1}, \boldsymbol{g}_{t-2}, \dots) = \\
&= \mathrm{E}(\varepsilon_t y_{t-1} | \boldsymbol{g}_{t-1}, \boldsymbol{g}_{t-2}, \dots) = \\
&= \mathrm{E}[\mathrm{E}(\varepsilon_t y_{t-1} | y_{t-1}, \boldsymbol{g}_{t-1}, \boldsymbol{g}_{t-2}, \dots) | \boldsymbol{g}_{t-1}, \boldsymbol{g}_{t-2}, \dots] \\
&\qquad (\text{по закону повторных математических ожиданий}) \\
&= \mathrm{E}[y_{t-1}\, \mathrm{E}(\varepsilon_t | y_{t-1}, \boldsymbol{g}_{t-1}, \boldsymbol{g}_{t-2}, \dots) | \boldsymbol{g}_{t-1}, \boldsymbol{g}_{t-2}, \dots] = \\
&= 0 \quad (\text{так как } \mathrm{E}(\varepsilon_t | y_{t-1}, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots, y_{t-2}\varepsilon_{t-1}, y_{t-3}\varepsilon_{t-2}, \dots) = 0).
\end{aligned} \tag{6.4.3}$$

В частности, $\mathrm{E}(\varepsilon_t y_{t-1}) = 0$, поэтому $\boldsymbol{x}_t$ ортогонален ошибке (предположение 2.3). Наконец, ранговое условие о матрице моментов $\mathrm{E}(\boldsymbol{x}_t \boldsymbol{x}_t')$ (предположение 2.4) удовлетворяется, поскольку определитель для

$$\mathrm{E}(\boldsymbol{x}_t \boldsymbol{x}_t') = \begin{bmatrix} 1 & \mu \\ \mu & \gamma_0 + \mu^2 \end{bmatrix} \tag{6.4.4}$$

равен $\gamma_0 > 0$. Это также означает, что матрица $\mathrm{E}(\boldsymbol{g}_t \boldsymbol{g}_t')$ невырожденная, поскольку при условной гомоскедастичности $\mathrm{E}(\boldsymbol{g}_t \boldsymbol{g}_t') = \sigma^2\, \mathrm{E}(\boldsymbol{x}_t \boldsymbol{x}_t')$. Поэтому удовлетворяется условие невырожденности в предположении 2.5.

Таким образом, AR(1) с процессом независимого белого шума удовлетворяет всем предположениям утверждения 2.5, так что части (а)–(с) этого утверждения остаются здесь верными с $(c, \phi)'$ в качестве $\boldsymbol{\beta}$, $(1, y_{t-1})'$ в качестве $\boldsymbol{x}_t$ и σ^2 в качестве дисперсии ошибок. В частности, OLS-оценка дисперсии ошибок,

$$s^2 = \frac{1}{n-2}\sum_{t=1}^{n} e_t^2,\ e_t \equiv y_t - \widehat{c} - \widehat{\phi} y_{t-1},\ \text{где } (\widehat{c}, \widehat{\phi}) \text{ — OLS-оценки},$$

является состоятельной для σ^2.

Оценивание AR(p)

Аналогично AR(1), авторегрессионные коэффициенты $(\phi_1, \ldots, \phi_p)$ можно оценить состоятельно, оценивая регрессию y_t на константу и лагированные y, $(y_{t-1}, \ldots, y_{t-p})$. Мы предполагаем, что выборка состоит из $(y_{-p+1}, y_{-p+2}, \ldots, y_0, y_1, \ldots, y_n)$, включая p лагированных значений, предшествующих y_1, так что уравнение AR(p) может быть записано начиная с $t = 1$ и оценивание производится по периоду с $t = 1$ до n.

Утверждение 6.7 (оценивание AR-коэффициентов): *Пусть $\{y_t\}$ является AR(p)-процессом, следующим (6.2.6) с условием стационарности (6.1.18). Предположим далее, что $\{\varepsilon_t\}$ является независимым белым шумом. Тогда OLS-оценка для $(c, \phi_1, \phi_2, \ldots, \phi_p)$, полученная путем регрессии y_t на константу и p лагированных значений y для периода выборки $t = 1, 2, \ldots, n$, является состоятельной и асимптотически нормальной. Если $\boldsymbol{\beta} = (c, \phi_1, \phi_2, \ldots, \phi_p)'$, $\boldsymbol{x}_t = (1, y_{t-1}, \ldots, y_{t-p})'$, и $\widehat{\boldsymbol{\beta}}$ — OLS-оценка для $\boldsymbol{\beta}$, то мы имеем:*

$$\mathrm{Avar}(\widehat{\boldsymbol{\beta}}) = \sigma^2\, \mathrm{E}(\boldsymbol{x}_t \boldsymbol{x}_t')^{-1},$$

что состоятельно оценивается как

$$\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\beta}})} = s^2 \cdot \left(\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{x}_t'\right)^{-1}.$$

где s^2 — OLS-оценка дисперсии ошибок, заданная как

$$s^2 = \frac{1}{n-p-1}\sum_{t=1}^{n}(y_t - \widehat{c} - \widehat{\phi}_1 y_{t-1} - \cdots - \widehat{\phi}_p y_{t-p})^2. \tag{6.4.5}$$

Доказательство заключается в проверке утверждения 2.5, что может быть сделано аналогично тому, как мы это делали для AR(1). Трудность состоит в том, чтобы показать, что матрица $\mathrm{E}(\boldsymbol{x}_t \boldsymbol{x}_t')$ невырождена (предположение 2.4), что эквивалентно требованию того, чтобы автоковариационная матрица $\mathrm{Var}(y_t, \ldots, y_{t-p+1})$ размерности $p \times p$ являлась невырожденной (см. контрольный вопрос 1(b) ниже). Можно показать, что автоковариационная матрица ковариационно стационарного процесса является невырожденной для любого p, если $\gamma_0 > 0$ и $\gamma_j \to 0$ при[1] $j \to \infty$. Поэтому предположение 2.4 также выполняется.

Мы не рассматривали оценивание AR(p) методом максимального правдоподобия, но для тех из вас, кто интересуется связью с ML, OLS-оценка численно совпадает с оценкой условного гауссовского ML и асимптотически эквивалентна оценке безусловного гауссовского ML. См. параграф 8.7.

Выбор глубины запаздываний

Все эти приятные результаты об оценивании авторегрессий предполагают, что порядок (глубина запаздываний) авторегрессии, p, известен. Как мы должны действовать, если порядок p неизвестен? Прежде всего ясно, что утверждение 6.7, которое основывается на авторегрессиях p-го порядка, применимо к авторегрессиям более низкого порядка, скажем, $r < p$. То есть, даже если $\phi_r \neq 0$, но $\phi_{r+1} = \phi_{r+2} = \cdots = \phi_p = 0$, это утверждение по-прежнему применимо, до тех пор пока $(\phi_1, \phi_2, \ldots, \phi_r)$ удовлетворяет условию стационарности. В частности, OLS-оценки для $(\phi_{r+1}, \phi_{r+2}, \ldots, \phi_p)$ будут сходиться к истинным нулевым значениям. Чтобы представить то же доказательство иначе, предположим, что истинный порядок авторегрессии равен p (так что $\phi_p \neq 0$), и предположим, что мы знаем о p то, что он меньше или равен некоторому известному целому числу $p_{\max}$. Полагая r в только что приведенном доказательстве равным p, а p равным $p_{\max}$, мы видим, что утверждение 6.7 применимо к авторегрессиям с $p_{\max}$ лагами: если $\{y_t\}$ является стационарным AR(p)-процессом и если мы оцениваем регрессию y_t на константу и $p_{\max}$ лагов от y, то OLS-оценки коэффициентов будут состоятельными и асимптотически нормальными.

Естественно думать, что истинная глубина запаздываний может быть оценена по указанной OLS-оценке для $(\phi_1, \ldots, \phi_{p_{\max}})$. Мы рассмотрим два класса правил для определения глубины запаздываний. Первое — это

Последовательное t-правило от общего к частному: *Начните с авторегрессии с $p_{\max}$ лагами. Если последний лаг значим на некотором предварительно специфицированном уровне значимости (скажем, 10%), то закончите процедуру и положите глубину запаздываний равной $p_{\max}$. В противном случае удалите последний лаг, переоцените авторегрессию с меньшим на единицу количеством лагов и повторите тот же самый тест. Если такой процесс продолжился, до тех пор пока в авторегрессии остался*

[1]См., например, Proposition 5.1.1 в [Brockwell and Davis, 1991].

только один лаг, и если этот лаг является незначимым, то тогда положите глубину запаздываний равной 0 (отсутствие лагов).

Это правило обладает следующими свойствами, которые будут показаны через мгновение. Поскольку t-тест является состоятельным, глубина запаздываний, выбранная таким образом, никогда не будет меньше, чем p (истинная глубина запаздываний) на больших выборках. Однако вероятность избыточности (того, что глубина запаздываний выбрана большей, чем истинная глубина запаздываний p) отлична от нуля даже на больших выборках. То есть если $\widehat{p}$ — глубина запаздываний, выбранная указанным последовательным t-правилом, то

$$\lim_{n\to\infty} \operatorname{Prob}(\widehat{p} < p) = 0 \quad \text{и} \quad \lim_{n\to\infty} \operatorname{Prob}(\widehat{p} > p) > 0. \tag{6.4.6}$$

Чтобы пояснить это, предположим, что $p = 2$ и $p_{\max} = 3$. Последовательное t-правило начинается с трех лагов в авторегрессии:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \varepsilon_t.$$

Мы проверяем нулевую гипотезу о том, что $\phi_3 = 0$, на 10%-ном уровне значимости. С вероятностью 10% мы отвергаем (верную) нулевую гипотезу и полагаем глубину запаздываний равной 3 или принимаем нулевую гипотезу с вероятностью 90% на больших выборках. В случае принятия нулевой гипотезы мы полагаем $\phi_3 = 0$, оцениваем авторегрессию с двумя лагами и тестируем гипотезу о том, что $\phi_2 = 0$. Поскольку в действительности $\phi_2 \neq 0$ (так как $p = 2$), абсолютное значение t-статистики, основанной на OLS-оценке для ϕ_2, может быть очень велико на больших выборках, так что мы никогда не примем (неверную) нулевую гипотезу о том, что второй лаг нулевой. Следовательно, в этом примере $\operatorname{Prob}(\widehat{p} = 2) = 90\%$ и $\operatorname{Prob}(\widehat{p} = 3) = 10\%$ на больших выборках.

Есть две незначительные вариации в выборе периода выборки с данными $(y_1, \ldots, y_n)$. Одна всюду использует одинаковый период $t = p_{\max} + 1, p_{\max} + 2, \ldots, n$. Другая полагает период выборки «эластичным», расширяя его при все меньшем и меньшем количестве включенных лагов. То есть, когда оцениваемая авторегрессия имеет j лагов, периодом выборки является $t = j + 1, j + 2, \ldots, n$.

Второе правило для выбора глубины запаздываний использует или **информационный критерий Акаике (AIC)** (*Akaike information criterion*) или **информационный критерий Шварца (SIC)** (*Schwartz information criterion*), также известный как **байесовский информационный критерий (BIC)** (*Bayesian information criterion*). Эта процедура полагает глубину запаздываний равной значению j, которое минимизирует

$$\log(SSR_j/n) + (j+1)C(n)/n \tag{6.4.7}$$

по $j = 0, 1, 2, \ldots, p_{\max}$, где SSR_j является суммой квадратов остатков для авторегрессии с j лагами. В этой формуле «$j + 1$» входит в качестве числа параметров (коэффициентов при константе и j лагов). Для AIC $C(n) = 2$, в то время как для BIC $C(n) = \log(n)$. Несколько комментариев об этом правиле, основанном на информационных критериях:

- Поскольку количество параметров возрастает вместе с j, первое слагаемое целевой функции (6.4.7) уменьшается, так как подгонка уравнения становится лучше, но второе слагаемое увеличивается. Информационный критерий обеспечивает баланс между лучшей подгонкой и экономностью модели.

- Без верхней границы $p_{\max}$ может быть выбрано смехотворно большое значение j (очевидно, информационный критерий достигает минимума $-\infty$ при $j + 1 = n$, поскольку $SSR_j = 0$ для $j = n - 1$). В текущем контексте, где истинный DGP является авторегрессией конечного порядка, разумной верхней границей является максимально возможная глубина запаздываний.

- И здесь существуют два способа устанавливать период выборки. Вы можете использовать всюду фиксированный период выборки $t = p_{\max}+1, p_{\max}+2, \ldots, n$, в этом случае SSR_j равна сумме $n-p_{\max}$ квадратов остатков. Или вы можете использовать «эластичный» период выборки, в этом случае SSR_j равна сумме $n - j$ квадратов остатков. В любом случае есть еще один вариант, который заключается в замене n в информационном критерии (6.4.7) на фактический объем выборки. Поэтому вы заменяете n на $n - p_{\max}$, если используется фиксированная выборка. На основе симуляций и некоторых теоретических соображений [Ng and Perron, 2000] рекомендуют использовать фиксированный, а не «эластичный» период выборки, с заменой n на $n - p_{\max}$ в информационном критерии.

Пусть $\widehat{p}_{\text{AIC}}$ и $\widehat{p}_{\text{BIC}}$ являются количествами запаздываний, выбранными посредством AIC и BIC соответственно. Как и глубина запаздываний, выбранная посредством t-правила от общего к частному, они есть функции от данных и, следовательно, являются случайными величинами. Для них можно показать следующее (см., например: [Lütkepohl, 1993, Section 4.3]:

(1) $\widehat{p}_{\text{BIC}} \le \widehat{p}_{\text{AIC}}$, если объем выборки не очень мал. (Если используется фиксированная выборка $t = p_{\max} + 1, \ldots, n$ и «$n - p_{\max}$» заменяет n в (6.4.7), это верно для любой конечной выборки с $n \ge p_{\max} + 8$.) Это является алгебраическим результатом и выполняется для любой выборки y.

(2) Предположим, что $\{y_t\}$ — стационарный AR(p) и ε_t — независимый белый шум с конечным четвертым моментом. Тогда

$$\operatorname*{plim}_{n\to\infty} \widehat{p}_{\mathrm{BIC}} = p, \quad \lim_{n\to\infty} \mathrm{Prob}(\widehat{p}_{\mathrm{AIC}} < p) = 0, \quad \lim_{n\to\infty} \mathrm{Prob}(\widehat{p}_{\mathrm{AIC}} > p) > 0. \tag{6.4.8}$$

То есть, точно так же, как и последовательное t-правило от общего к частному, правило AIC имеет положительную вероятность завышения значения p. В противоположность им правило BIC является состоятельным. Кроме того, [Hannan and Deistler 1988, Theorem 5.4.1(c) and Remark 1] показали, что состоятельность $\widehat{p}_{\mathrm{BIC}}$ выполняется, когда верхняя граница $p_{\max}$ сделана возрастающей со скоростью $\log(n)$ (то есть когда она полагается равной $[c\log(n)]$ [целой частью $c\log(n)$] для любого заданного $c > 0$). Это означает, что вам не нужно знать верхнюю границу для состоятельного оценивания порядка авторегрессии посредством BIC.

Оценивание VAR

Оценивание коэффициентов VAR столь же просто. Если $\boldsymbol{y}_t$ является M-мерным вектором, VAR(p) (6.3.9) является системой из M уравнений и может быть записана в виде:

$$y_{tm} = \boldsymbol{x}_t'\boldsymbol{\delta}_m + \varepsilon_{tm} \quad (m = 1, 2, \dots, M), \tag{6.4.9}$$

где y_{tm} является m-м элементом в $\boldsymbol{y}_t$, ε_{tm} — m-м элементом в $\boldsymbol{\varepsilon}_t$, и

$$\underset{(Mp+1)\times 1}{\boldsymbol{x}_t} = \begin{bmatrix} 1 \\ \boldsymbol{y}_{t-1} \\ \boldsymbol{y}_{t-2} \\ \vdots \\ \boldsymbol{y}_{t-p} \end{bmatrix}, \quad \boldsymbol{\delta}_m = \begin{bmatrix} c_m \\ \boldsymbol{\phi}_{1m} \\ \boldsymbol{\phi}_{2m} \\ \vdots \\ \boldsymbol{\phi}_{pm} \end{bmatrix},$$

$$c_m = m\text{-й элемент вектора } \boldsymbol{c}, \quad \boldsymbol{\phi}_{jm}' = m\text{-я строка матрицы } \boldsymbol{\Phi}_j.$$

Эти M уравнений имеют одно и то же множество регрессоров. Легко проверить, что $\boldsymbol{x}_t$ ортогонален ошибке в каждом уравнении и что ошибки условно гомоскедастичны (доказательство очень похоже на доказательство для AR(p)). Таким образом, указанная система из M уравнений с i.i.d. $\{\boldsymbol{\varepsilon}_t\}$ удовлетворяет условиям многомерной регрессионной модели из параграфа 4.5. Поэтому эффективное GMM-оценивание осуществляется очень просто: достаточно применить OLS поочередно к каждому уравнению системы. Выражения для асимптотической дисперсии и ее состоятельной оценки могут быть получены из утверждения 4.6 с $\boldsymbol{z}_{im} = \boldsymbol{x}_t$ (так что (4.5.13′) на с. 309 становится равным $\frac{1}{n}\sum_t \boldsymbol{x}_t\boldsymbol{x}_t'$). Если $\boldsymbol{\delta}$ является $(Mp+1)M$-мерным штабелированным вектором, созданным из $(\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_M)$, и если $\widehat{\boldsymbol{\delta}}$ является оценкой для $\boldsymbol{\delta}$, получаемой последовательным применением OLS к отдельным уравнениям, то мы имеем из (4.5.17):

$$\widehat{\operatorname{Avar}(\widehat{\boldsymbol{\delta}})} = \widehat{\boldsymbol{\Omega}} \otimes \left(\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{x}_t'\right)^{-1}, \tag{6.4.10}$$

где

$$\widehat{\boldsymbol{\Omega}} \equiv \frac{1}{n - Mp - 1}\sum_{t=1}^{n} \widehat{\boldsymbol{\varepsilon}}_t \widehat{\boldsymbol{\varepsilon}}_t', \quad \widehat{\boldsymbol{\varepsilon}}_t = \boldsymbol{y}_t - \widehat{\boldsymbol{c}} - \widehat{\boldsymbol{\Phi}}_1 \boldsymbol{y}_{t-1} - \cdots - \widehat{\boldsymbol{\Phi}}_p \boldsymbol{y}_{t-p}. \tag{6.4.11}$$

Если мы не знаем глубину запаздываний p, оцениваем ее по данным. В правиле, основанном на информационном критерии, глубина запаздываний, которую мы выбираем, равна значению k, которое минимизирует

$$\log\left(\left|\frac{1}{n}\sum_{t=1}^{n} \widehat{\boldsymbol{\varepsilon}}_t \widehat{\boldsymbol{\varepsilon}}_t'\right|\right) + (kM^2 + M) \cdot \frac{C(n)}{n} \tag{6.4.12}$$

по $k = 0, 1, \ldots, p_{\max}$. Здесь, как и в одномерном случае, $C(n) = 2$ для AIC и $\log(n)$ для BIC. Точно такие же результаты, перечисленные ранее как (1) и (2) для одномерного случая, выполняются и для многомерного случая (см., например: [Lütkepohl, 1993, Section 4.3].

Оценивание ARMA(p, q)

Полагая

$$u_t \equiv \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q},$$
$$\boldsymbol{z}_t = (1, y_{t-1}, y_{t-2}, \ldots, y_{t-p})', \quad \boldsymbol{\delta} = (c, \phi_1, \phi_2, \ldots, \phi_p)',$$

одномерное уравнение ARMA(p, q) (6.2.9) можно записать в виде:

$$y_t = \boldsymbol{z}_t' \boldsymbol{\delta} + u_t. \tag{6.4.13}$$

Компонента скользящего среднего в ARMA создает две проблемы. Во-первых, очевидно, что ошибки u_t сериально коррелированы (фактически они образуют процесс MA(q)). Во-вторых, поскольку лагированные зависимые переменные, включенные в $\boldsymbol{z}_t$, коррелированы с лагом ε, включенным в ошибки, регрессоры $\boldsymbol{z}_t$ не являются ортогональными ошибке u_t. То есть сериальная корреляция при наличии лагированных зависимых переменных приводит к смещенному оцениванию. Со второй проблемой можно было бы справиться с помощью подходящих лагированных зависимых переменных $(y_{t-q-1}, y_{t-q-2}, \ldots)$ в качестве инструментов. Что касается первой проблемы, метод, который будет разработан позднее в этой главе, предоставляет состоятельные оценки вектора коэффициентов при наличии сериальной корреляции. При заданной состоятельной оценке δ, MA-параметры (θ-ы) могут быть оценены по остаткам. Эта процедура, однако, не является эффективной[1]. Мы не будем рассматривать тему эффективного оценивания процессов ARMA(p, q) с гауссовскими ошибками. Интересующиеся читатели могут обратиться, например, к [Fuller, 1996, Chapter 8] или [Hamilton, 1994, Chapter 5].

Контрольные вопросы

1. (Avar для AR-коэффициентов.)

   (a) Для AR(1) покажите, что

$$\text{Avar}(\widehat{\phi}) = \frac{\sigma^2}{\gamma_0} = 1 - \phi^2,$$

   где $\widehat{\phi}$ — OLS-оценка для ϕ. **Указание:** *Это равно умноженному на* σ^2 $(2, 2)$*-му элементу обращения* (6.4.4).

   (b) Для AR(p) пусть $\widehat{\boldsymbol{\phi}}$ является OLS-оценкой для $(\phi_1, \phi_2, \dots, \phi_p)'$. Покажите, что

$$\text{Avar}(\widehat{\boldsymbol{\phi}}) = \sigma^2 \boldsymbol{V}^{-1},$$

   где $\boldsymbol{V}$ является автоковариационной матрицей размера $p \times p$:

$$\boldsymbol{V} = \text{Var}(y_t, \dots, y_{t-p+1}) = \begin{bmatrix} \gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \gamma_1 & \cdots & \gamma_{p-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \gamma_{p-2} & \cdots & \gamma_1 & \gamma_0 & \gamma_1 \\ \gamma_{p-1} & \cdots & \gamma_2 & \gamma_1 & \gamma_0 \end{bmatrix}.$$

   Указание: *Если* $\boldsymbol{x}_t$ *является тем же, что и в утверждении 6.7, то*

$$\text{E}(\boldsymbol{x}_t \boldsymbol{x}_t') = \begin{bmatrix} 1 & \mu \mathbf{1}' \\ \mu \mathbf{1} & \boldsymbol{V} + \mu^2 \mathbf{1}\mathbf{1}' \end{bmatrix},$$

   где $\mathbf{1}$ — $p \times 1$ *вектор из единиц. Покажите, что обращение этой матрицы задается как*

$$\boldsymbol{A} \equiv \begin{bmatrix} a & \boldsymbol{c}' \\ \boldsymbol{c} & \boldsymbol{V}^{-1} \end{bmatrix}, \text{ где } a \equiv 1 + \mu^2 \mathbf{1}' \boldsymbol{V}^{-1} \mathbf{1},\ \boldsymbol{c} \equiv -\mu \boldsymbol{V}^{-1} \mathbf{1}.$$

   Это также показывает, что матрица $\text{E}(\boldsymbol{x}_t \boldsymbol{x}_t')$ *является невырожденной тогда и только тогда, когда это выполняется для* $\boldsymbol{V}$. *(Если* $\boldsymbol{V}$ *невырожденная, то* $\text{E}(\boldsymbol{x}_t \boldsymbol{x}_t')$ *невырожденная из-за ее обратимости. Если* $\boldsymbol{V}$ *вырожденная, то существует* p*-мерный вектор* $\boldsymbol{d} \neq \mathbf{0}$, *такой что* $\boldsymbol{V}\boldsymbol{d} = \mathbf{0}$ *и* $\text{E}(\boldsymbol{x}_t \boldsymbol{x}_t')\boldsymbol{f} = \mathbf{0}$, *где* $\boldsymbol{f}' = (-\mu \mathbf{1}'\boldsymbol{d}, \boldsymbol{d}')$.*)*

[1]Эта модель приводит к бесконечно большому количеству условий ортогональности, $\text{E}(u_t \cdot y_s) = 0$ для всех $s \leq t - q - 1$. Процедура GMM может использовать только конечное количество условий ортогональности.

2. (Оцененные коэффициенты VAR(1).) Рассмотрим VAR(1) без константы: $\mathbf{y}_t = \mathbf{A}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t$. Проверьте, что оценка для $\mathbf{A}$ в многомерной регрессии при использовании выборки $(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_n)$ (так что период выборки — $t = 2, 3, \dots, n$) равна:

$$\widehat{\mathbf{A}} = \left(\sum_{t=2}^{n} \mathbf{y}_t \mathbf{y}'_{t-1}\right) \left(\sum_{t=2}^{n} \mathbf{y}_{t-1} \mathbf{y}'_{t-1}\right)^{-1}.$$

Указание: *Вектор коэффициентов в m-м уравнении равен m-й строке* $\mathbf{A}$.

6.5. Асимптотика для выборочных средних сериально коррелированных процессов

Этот параграф изучает асимптотические свойства выборочного среднего

$$\overline{y} \equiv \frac{1}{n} \sum_{t=1}^{n} y_t$$

для сериально коррелированных процессов. (Следует иметь в виду, что $\overline{y}$ зависит от n, хотя обозначение не указывает на это явно.) Для состоятельности выборочного среднего у нас уже есть достаточное условие в эргодической теореме главы 2. Первое утверждение в этом параграфе предлагает другое достаточное условие в форме ограничений на ковариационно стационарные процессы. Мы также предлагаем две CLT для сериально коррелированных процессов: одну — для линейных процессов, а другую — для эргодических стационарных процессов. Глава 2 включает CLT для эргодических стационарных процессов, но она не учитывает сериальную корреляцию, поскольку предполагается, что процесс является мартингал-разностью. CLT для эргодических стационарных процессов этого параграфа обобщает ее, допуская сериальную корреляцию.

LLN для ковариационно стационарных процессов

Утверждение 6.8 (LLN для ковариационно стационарных процессов с затухающими автоковариациями): *Пусть $\{y_t\}$ является ковариационно стационарным процессом с математическим ожиданием μ и $\{\gamma_j\}$ являются автоковариациями процесса $\{y_t\}$. Тогда*

(a) $\overline{y} \to_{m.s.} \mu$ *при* $n \to \infty$, *если* $\lim_{j\to\infty} \gamma_j = 0$.

(b)

$$\lim_{n\to\infty} \mathrm{Var}(\sqrt{n}\,\overline{y}) = \sum_{j=-\infty}^{\infty} \gamma_j < \infty, \quad \text{если } \{\gamma_j\} \text{ суммируема}^1. \qquad (6.5.1)$$

[1] [Anderson, 1971, Lemma 8.3.1, p. 460]. Абсолютная суммируемость $\{\gamma_j\}$ не является необходимой.

Поскольку из сходимости в среднем квадратическом вытекает сходимость по вероятности, часть (a) показывает, что достаточно очень слабых условий на ковариационно стационарный процесс, чтобы выборочное среднее было состоятельным для μ. Для доказательства части (a), по слабому закону больших чисел Чебышева (см. параграф 2.1), достаточно показать, что $\lim_{n\to\infty} \mathrm{Var}(\overline{y}) = 0$, что довольно легко (аналитическое упражнение 9), после того как получено выражение для $\mathrm{Var}(\overline{y})$. В отличие от случая с отсутствием сериальной корреляции в $\{y_t\}$, вычисление дисперсии для $\overline{y}$ является более громоздким, поскольку должны быть приняты во внимание ковариации между двумя различными слагаемыми:

$$
\begin{aligned}
n\cdot \mathrm{Var}(\overline{y}) &= \mathrm{Var}(\sqrt{n}\,\overline{y}) = \\
&= \frac{1}{n}[\mathrm{Cov}(y_1, y_1 + \cdots + y_n) + \cdots + \mathrm{Cov}(y_n, y_1 + \cdots + y_n)] = \\
&= \frac{1}{n}\,[(\gamma_0 + \gamma_1 + \cdots + \gamma_{n-2} + \gamma_{n-1}) + (\gamma_1 + \gamma_0 + \gamma_1 + \cdots + \gamma_{n-2}) + \\
&\qquad + \cdots + (\gamma_{n-1} + \gamma_{n-2} + \cdots + \gamma_1 + \gamma_0)] \\
&= \frac{1}{n}\,[n\gamma_0 + 2(n-1)\gamma_1 + \cdots + 2(n-j)\gamma_j + \cdots + 2\gamma_{n-1}] = \\
&= \gamma_0 + 2\sum_{j=1}^{n-1}\Big(1 - \frac{j}{n}\Big)\gamma_j. \qquad (6.5.2)
\end{aligned}
$$

(Если вам сложно получить это, положите n, скажем, 3 для проверки.) Для каждого фиксированного $j \geq 1$ коэффициент при γ_j в (6.5.2) сходится к 1 при $n \to \infty$, так что все автоковариации в конечном счете входят в сумму с коэффициентом, равным единице. Это не является достаточным для того, чтобы показать желаемый результат (6.5.1). Доказательство части (b) является аналитическим упражнением 10.

Для ковариационно стационарного процесса $\{y_t\}$ мы определяем **долговременную дисперсию** (*long-run variance*) как предел $\mathrm{Var}(\sqrt{n}\,\overline{y})$ при $n \to \infty$ (если он существует). Поэтому часть (b) утверждения говорит о том, что долговременная дисперсия равна $\sum_{j=-\infty}^{\infty}\gamma_j$, что, в свою очередь, равно значению производящей функции автоковариаций $g_Y(z)$ при $z = 1$ (см. (6.2.15)) или умноженному на 2π значению спектра на нулевой частоте (см. (6.2.17)).

Две центральные предельные теоремы

Мы представляем две CLT, которые охватывают сериальную корреляцию. Первая является обобщением CLT Линдеберга — Леви.

Утверждение 6.9 (CLT для MA(∞)): *Пусть $y_t = \mu + \sum_{j=0}^{\infty}\psi_j\varepsilon_{t-j}$, где $\{\varepsilon_t\}$ — независимый белый шум, и $\sum_{j=0}^{\infty}|\psi_j| < \infty$. Тогда*

$$\sqrt{n}\,(\bar{y}-\mu) \underset{d}{\to} N\Big(0, \sum_{j=-\infty}^{\infty} \gamma_j\Big). \tag{6.5.3}$$

Мы не будем доказывать этот результат (для доказательства см.: [Anderson, 1971, Theorem 7.7.8] или [Brockwell and Davis, 1991, Theorem 7.1.2]), но мы не должны удивляться тому, что $\text{Avar}(\bar{y})$ (дисперсия предельного распределения $\sqrt{n}\,(\bar{y}-\mu)$) равна долговременной дисперсии $\sum_{j=-\infty}^{\infty}\gamma_j$. Мы знаем из леммы 2.1, что «$x_n \to_{\text{d}} x$», «$\text{Var}(x_n) \to c < \infty$», «$\text{E}(|x_n|^{2+\delta}) < M$» $\Rightarrow$ «$\text{Var}\,x = c$». Положим здесь $x_n = \sqrt{n}\,(\bar{y}-\mu)$. По (6.5.1), $\text{Var}(x_n)$ сходится к конечному пределу $\sum\gamma_j$. Поэтому дисперсия предельного распределения $\sqrt{n}\,(\bar{y}-\mu)$ равна $\sum\gamma_j$.

По утверждению 6.1(d), процесс в утверждении 6.9 является (стационарным и) эргодическим. Вспомните, что эргодичность ограничивает сериальную корреляцию требованием того, что две случайные величины, отстоящие достаточно далеко друг от друга в последовательности, являются почти независимыми. Но эргодическая стационарность сама по себе не является достаточной для гарантирования асимптотической нормальности $\bar{y}$. Возникает естественный вопрос: существуют ли дополнительные ограничения на эргодичность, если не считать линейные процессы, как в утверждении 6.9, которые обеспечивают асимптотическую нормальность? Ограничение, которое становится все более популярным — это то, что в этой книге называется **условием Гордина** (*Gordin's condition*) для эргодических стационарных процессов. У него есть три части. Первые две части следующие:

(a) $\text{E}(y_t^2) < \infty$. (Это является ограничением на [строго] стационарные процессы, поскольку, строго говоря, стационарные процессы могут не иметь конечных вторых моментов.)

(b) $\text{E}(y_t|y_{t-j}, y_{t-j-1}, \ldots) \underset{\text{m.s.}}{\to} 0$ при $j \to \infty$. (Так как процесс $\{y_t\}$ стационарный, это условие эквивалентно тому, что $\text{E}(y_0|y_{-j}, y_{-j-1}, \ldots) \underset{\text{m.s.}}{\to} 0$ при $j \to \infty$. Поскольку $\text{E}(y_t|y_{t-j}, y_{t-j-1}, \ldots)$ является случайной величиной, эта сходимость не может быть обычной сходимостью для действительных чисел.)

Из этого условия вытекает, что безусловное среднее равно нулю[1], $\text{E}(y_t) = 0$. Поскольку прогноз (условное математическое ожидание) основывается на все меньшем и меньшем количестве информации, он должен приближаться к прогнозу, основанному на отсутствии информации, то есть к безусловному математическому ожиданию. Для подготовки к третьей

[1]Относительно доказательства см.: Lemma 5.14 в [White, 1984].

части положим $I_t \equiv (y_t, y_{t-1}, y_{t-2}, \dots)$ и запишем y_t как

$$\begin{aligned} y_t &= y_t - [\mathrm{E}(y_t|I_{t-1}) - \mathrm{E}(y_t|I_{t-1})] - [\mathrm{E}(y_t|I_{t-2}) - \mathrm{E}(y_t|I_{t-2})] - \dots - \\ &\quad -[\mathrm{E}(y_t|I_{t-j}) - \mathrm{E}(y_t|I_{t-j})] = \\ &= [y_t - \mathrm{E}(y_t|I_{t-1})] + [\mathrm{E}(y_t|I_{t-1}) - \mathrm{E}(y_t|I_{t-2})] + \dots + \\ &\quad +[\mathrm{E}(y_t|I_{t-j+1}) - \mathrm{E}(y_t|I_{t-j})] + \mathrm{E}(y_t|I_{t-j}) = \\ &= (r_{t0} + r_{t1} + \dots + r_{t,j-1}) + \mathrm{E}(y_t|y_{t-j}, y_{t-j-1}, \dots), \end{aligned}$$

где

$$r_{tk} \equiv \mathrm{E}(y_t|I_{t-k}) - \mathrm{E}(y_t|I_{t-k-1}).$$

Это r_{tk} является изменением (пересмотром) ожиданий относительно y_t при увеличении информационного множества с I_{t-k-1} до I_{t-k}. По (b), $y_t - (r_{t0} + r_{t1} + \dots + r_{t,j-1})$ сходится в среднем квадратическом к нулю при $j \to \infty$ для каждого t. Поэтому y_t можно записать как

$$y_t = \sum_{j=0}^{\infty} r_{tj}.$$

Эта сумма называется **телескопической суммой** (*telescoping sum*). Третьей частью условия Гордина является

c) $\sum_{j=0}^{\infty} [\mathrm{E}(r_{tj}^2)]^{1/2} < \infty$.

Телескопическая сумма указывает, как «шоки», представленные $(r_{t0}, \dots)$, влияют на текущее значение y. Условие (c), грубо говоря, показывает, что шоки, которые произошли давно, не оказывают непропорционально большого влияния. В таком виде это условие ограничивает степень сериальной автокорреляции в $\{y_t\}$.

Для понимания указанного условия возьмем пример AR(1) с нулевым средним и с независимыми ошибками:

$$y_t = \phi y_{t-1} + \varepsilon_t, \quad |\phi| < 1.$$

$\{\varepsilon_t\}$ — независимый белый шум с $\sigma^2 = \mathrm{Var}(\varepsilon)$.

Очевидно, что условие Гордина (a) удовлетворяется. Условие (b) также удовлетворяется, поскольку

$$\mathrm{E}(y_t|y_{t-j}, y_{t-j-1}, \dots) = \phi^j y_{t-j} \tag{6.5.4}$$

и $\mathrm{E}[(\phi^j y_{t-j})^2] \to 0$ при $j \to \infty$. Пересмотр ожиданий r_{tj} может быть записан как

$$r_{tj} = \phi^j y_{t-j} - \phi^{j+1} y_{t-j-1} = \phi^j (y_{t-j} - \phi y_{t-j-1}) = \phi^j \varepsilon_{t-j}. \tag{6.5.5}$$

Поэтому телескопическая сумма для AR(1) является MA(∞)-представлением. Условие (c) удовлетворяется, поскольку $[\mathrm{E}(r_{tj}^2)]^{1/2} = |\phi|^j\sigma$ и

$$\sum_{j=0}^{\infty} |\phi|^j \sigma = \frac{\sigma}{1-|\phi|} < \infty.$$

Искомая CLT такова:

Утверждение 6.10 (CLT Гордина для эргодических стационарных процессов с нулевым средним[1]): *Предположим, что процесс* $\{y_t\}$ *стационарный и эргодический, и предположим, что выполняется условие Гордина. Тогда* $\mathrm{E}(y_t) = 0$, *автоковариации* $\{\gamma_j\}$ *абсолютно суммируемы и*

$$\sqrt{n}\,\overline{y} \underset{d}{\to} N\Big(0, \sum_{j=-\infty}^{\infty} \gamma_j\Big).$$

Поскольку условие Гордина удовлетворяется, если $\{y_t\}$ является мартингал-разностью, которая не проявляет сериальной корреляции, утверждение 6.10 является обобщением CLT для эргодической стационарной мартингал-разности из главы 2.

Многомерное обобщение

Обобщение приведенных выше результатов на многомерный случай производится непосредственно. Пусть

$$\overline{\boldsymbol{y}} \equiv \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{y}_t$$

является выборочным средним векторного процесса $\{\boldsymbol{y}_t\}$.

- Многомерная версия утверждения 6.8 заключается в том, что
  (a) $\overline{\boldsymbol{y}} \to_{\mathrm{m.s.}} \boldsymbol{\mu}$, если каждый диагональный элемент в $\boldsymbol{\Gamma}_j$ стремится к нулю при $j \to \infty$, и
  (b) предел $\lim_{n\to\infty} \mathrm{Var}(\sqrt{n}\,\overline{\boldsymbol{y}})$ (который является **долговременной ковариационной матрицей** (*long-run covariance matrix*) для $\{\boldsymbol{y}_t\}$) равен $\sum_{j=-\infty}^{\infty} \boldsymbol{\Gamma}_j$, если $\{\boldsymbol{\Gamma}_j\}$ является суммируемой (то есть если каждая компонента $\boldsymbol{\Gamma}_j$ является суммируемой).

[1] Этот результат, восходящий к [Gordin, 1969], был переформулирован как Theorem 5.15 в: [White, 1984]. Утверждение об абсолютной суммируемости автоковариаций отмечается в сноске 19 в: [Hansen, 1982].

- Таким образом, долговременная ковариационная матрица $\{\boldsymbol{y}_t\}$ может быть записана как

$$\boldsymbol{G_Y}(1) = 2\pi \boldsymbol{s_Y}(0) = \sum_{j=-\infty}^{\infty} \boldsymbol{\Gamma}_j = \boldsymbol{\Gamma}_0 + \sum_{j=1}^{\infty}(\boldsymbol{\Gamma}_j + \boldsymbol{\Gamma}'_j), \tag{6.5.6}$$

где $\boldsymbol{G_Y}(z)$ (производящая функция автоковариаций) и $\boldsymbol{s_Y}(\omega)$ (спектр) для векторного процесса $\{\boldsymbol{y}_t\}$ определяются в параграфе 6.3.

- Многомерная версия утверждения 6.9 заключается в том, что

$$\sqrt{n}\,(\overline{\boldsymbol{y}} - \boldsymbol{\mu}) \underset{\mathrm{d}}{\to} N\Big(\mathbf{0}, \sum_{j=-\infty}^{\infty} \boldsymbol{\Gamma}_j\Big), \tag{6.5.7}$$

если $\{\boldsymbol{y}_t\}$ является векторным MA(∞) с абсолютно суммируемыми коэффициентами, а $\{\boldsymbol{\varepsilon}_t\}$ является векторным независимым белым шумом.

- Многомерное обобщение условия Гордина на эргодические стационарные процессы очевидно[1]:

Условие Гордина на эргодические стационарные процессы

(a) $\mathrm{E}(\boldsymbol{y}_t\boldsymbol{y}'_t)$ *существует и конечно.*

(b) $\mathrm{E}(\boldsymbol{y}_t|\boldsymbol{y}_{t-j}, \boldsymbol{y}_{t-j-1}, \ldots) \to_{m.s.} \mathbf{0}$ *при* $j \to \infty$.

(c) $\sum_{j=0}^{\infty}[\mathrm{E}(\boldsymbol{r}'_{tj}\boldsymbol{r}_{tj})]^{1/2}$ *конечна, где*

$$\boldsymbol{r}_{tj} \equiv \mathrm{E}(\boldsymbol{y}_t|\boldsymbol{y}_{t-j}, \boldsymbol{y}_{t-j-1}, \ldots) - \mathrm{E}(\boldsymbol{y}_t|\boldsymbol{y}_{t-j-1}, \boldsymbol{y}_{t-j-2}, \ldots).$$

- Многомерная версия утверждения 6.10 очевидна: Предположим, что условие Гордина выполняется для векторного эргодического стационарного процесса $\{\boldsymbol{y}_t\}$. Тогда $\mathrm{E}(\boldsymbol{y}_t) = \mathbf{0}$, $\{\boldsymbol{\Gamma}_j\}$ является абсолютно суммируемой и

$$\sqrt{n}\,\overline{\boldsymbol{y}} \underset{\mathrm{d}}{\to} N\Big(\mathbf{0}, \sum_{j=-\infty}^{\infty} \boldsymbol{\Gamma}_j\Big). \tag{6.5.8}$$

Контрольные вопросы

1. Покажите, что $\mathrm{Var}(\sqrt{n}\,\overline{y}) = \mathbf{1}'\,\mathrm{Var}(y_t, y_{t-1}, \ldots, y_{t-n+1})\mathbf{1}/n$, где $\mathrm{Var}(y_t, y_{t-1}, y_{t-2}, \ldots, y_{t-n+1})$ — автоковариационная матрица размера $n \times n$ ковариационно стационарного процесса $\{y_t\}$.

[1] Сформулировано как Assumption 3.5 в [Hansen, 1982].

2. ($\mathrm{Avar}(\overline{y})$ для $\mathrm{MA}(\infty)$.) Проверьте, что $\mathrm{Avar}(\overline{y})$ для процесса $\mathrm{MA}(\infty)$ в утверждении 6.9 равна
$$\sigma^2 \cdot \Big(\sum_{j=0}^{\infty} \psi_j\Big)^2.$$
Указание: *Она должна быть равна* $g_Y(1)$ *для* $MA(\infty)$.

3. (Асимптотика $\overline{y}$ для AR(1).) Рассмотрим AR(1)-процесс $y_t = c + \phi y_{t-1} + \varepsilon_t$ с $|\phi| < 1$. Верно ли, что $\overline{y} \to_{\mathrm{p}} \mu$? Добавьте предположение, что $\{\varepsilon_t\}$ является процессом независимого белого шума. Вычислите $\mathrm{Avar}(\overline{y})$. [Ответ: $[(1+\phi)/(1-\phi)]\gamma_0$.]

4. (Долгосрочная ковариационная матрица профильтрованного ряда.)
Пусть $\boldsymbol{H}(L) = \boldsymbol{H}_0 + \boldsymbol{H}_1 L$, и $\{\boldsymbol{x}_t\}$ — ковариационно стационарный процесс с абсолютно суммируемыми автоковариациями. Проверьте, что долговременная ковариационная матрица для $\boldsymbol{y}_t = \boldsymbol{H}(L)\boldsymbol{x}_t$ задается как
$$(\boldsymbol{H}_0 + \boldsymbol{H}_1)\boldsymbol{G}_{\boldsymbol{x}}(1)(\boldsymbol{H}_0' + \boldsymbol{H}_1').$$
Указание: *Используйте* (6.3.13) *и* (6.5.6).

5. (Долгосрочная дисперсия процесса «MA с единичным корнем».) Какова долговременная дисперсия процесса $y_t = \varepsilon_t - \varepsilon_{t-1}$, где ε_t является белым шумом? [Ответ: равна нулю.]

6.6. Включение сериальной корреляции в GMM

Модель и асимптотические результаты

Сейчас у нас есть все инструменты, необходимые для включения сериальной корреляции в модель GMM с одним уравнением главы 3. Вспомним, что $\boldsymbol{g}_t$ из главы 3 (с индексом «i», замененным на «t») является K-мерным вектором, определенным как $\boldsymbol{x}_t \cdot \varepsilon_t$, произведение K-мерного вектора инструментов $\boldsymbol{x}_t$ и скалярной ошибки ε_t. По предположению 3.3 (условия ортогональности), среднее $\boldsymbol{g}_t$ равно нулю. Мы предполагали в предположении 3.5, что $\{\boldsymbol{g}_t\}$ является мартингал-разностью. Поскольку при этом предположении не допускается сериальная корреляция, матрица $\boldsymbol{S}$, определенная как асимптотическая дисперсия $\overline{\boldsymbol{g}}(\equiv \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{g}_t)$, была просто дисперсией $\boldsymbol{g}_t$. Теперь мы допускаем сериальную корреляцию в $\{\boldsymbol{g}_t\}$, ослабляя предположение 3.5 следующим образом:

Предположение 3.5′ (условие Гордина, ограничивающее степень сериальной корреляции):$\{\boldsymbol{g}_t\}$ *удовлетворяет условию Гордина. Его долговременная ковариационная матрица невырождена.*

Тогда по многомерной версии утверждения 6.10, $\sqrt{n}\,\overline{\boldsymbol{g}}$ сходится к нормальному распределению. Дисперсия этого предельного распределения (а именно $\mathrm{Avar}(\overline{g})$), обозначенная как $\boldsymbol{S}$, равна долговременной ковариационной матрице,

$$\boldsymbol{S} = \sum_{j=-\infty}^{\infty} \boldsymbol{\Gamma}_j = \boldsymbol{\Gamma}_0 + \sum_{j=1}^{\infty}(\boldsymbol{\Gamma}_j + \boldsymbol{\Gamma}'_j). \tag{6.6.1}$$

где $\boldsymbol{\Gamma}_j$ — автоковариационная матрица j-го порядка

$$\boldsymbol{\Gamma}_j = \mathrm{E}(g_t g'_{t-j}) \quad (j = 0, \pm 1, \pm 2, \ldots). \tag{6.6.2}$$

(Так как $\mathrm{E}(g_t) = \mathbf{0}$ по условиям ортогональности, здесь нет необходимости вычитать среднее из g_t.)

То есть $\boldsymbol{S}(\equiv \mathrm{Avar}(\overline{g}))$ равна $\boldsymbol{\Gamma}_0$ при предположении 3.5 и равна $\sum_{j=-\infty}^{\infty} \boldsymbol{\Gamma}_j$ при предположении 3.5′. Это *единственное* различие между моделями GMM с наличием и отсутствием сериальной корреляции. Соответственно переносятся все результаты, которые мы разработали в параграфах 3.5–3.7 при условии, что $\boldsymbol{S}$ задается сейчас соотношением (6.6.1) и $\widehat{\boldsymbol{S}}$, некоторая состоятельная оценка для $\boldsymbol{S}$, переопределяется, чтобы принять во внимание сериальную корреляцию. Более конкретно:

- GMM-оценка $\widehat{\delta}(\widehat{\boldsymbol{W}})$ остается состоятельной и асимптотически нормальной. Ее асимптотическая дисперсия состоятельно оценивается посредством (3.5.2) при условии, что $\widehat{\boldsymbol{S}}$ является состоятельной оценкой для $\boldsymbol{S}$, определенной в (6.6.1). Об этой оцененной асимптотической дисперсии иногда говорят как о состоятельной к гетероскедастичности и автокорреляции **(HAC)**.

- GMM-оценка достигает минимальной дисперсии, при $\operatorname*{plim}_{n\to\infty} \widehat{\boldsymbol{W}} = \boldsymbol{S}^{-1}$, где $\boldsymbol{S}$ задается как в (6.6.1).

- Двухшаговая процедура, описанная в параграфе 3.5, все еще дает эффективную GMM-оценку при условии что $\widehat{\boldsymbol{S}}$, вычисленная на *шаге 1*, является состоятельной для $\boldsymbol{S}$.

- С такой, надлежащим образом определенной $\widehat{\boldsymbol{S}}$, выражения для статистик t, W, J, C и LR в параграфах 3.5–3.7 остаются в силе; эти статистики сохраняют те же асимптотические распределения и при наличии сериальной корреляции.

Оценивание S, когда автоковариации обращаются в нуль после конечного количества лагов

Все эти результаты подразумевают, что у вас есть состоятельная оценка, $\widehat{\boldsymbol{S}}$, долговременной ковариационной матрицы $\boldsymbol{S}$. Получение такой оценки требует некоторого интеллектуального усилия, особенно когда автоковариации не обращаются в нуль после конечного количества лагов.

Мы начинаем наш поиск с состоятельного оценивания отдельных автоковариаций. Естественной оценкой является

$$\widehat{\mathbf{\Gamma}}_j = \frac{1}{n}\sum_{t=j+1}^{n} \widehat{\boldsymbol{g}}_t \widehat{\boldsymbol{g}}'_{t-j} \quad (j = 0, 1, \ldots, n-1), \tag{6.6.3}$$

где

$$\widehat{\boldsymbol{g}}_t \equiv \boldsymbol{x}_t \cdot \widehat{\varepsilon}_t, \quad \widehat{\varepsilon}_t = y_t - \boldsymbol{z}'_t\widehat{\boldsymbol{\delta}}, \quad \widehat{\boldsymbol{\delta}} \text{ состоятельна для } \boldsymbol{\delta}.$$

Если бы у нас были действительные значения $\boldsymbol{g}_t$ вместо оцененных значений $\widehat{\boldsymbol{g}}_t$ в этой формуле, то $\widehat{\mathbf{\Gamma}_j}$ была бы состоятельной для $\mathbf{\Gamma}_j$ при предположениях 3.1 и 3.2, обеспечивающих эргодическую стационарность для $\{\boldsymbol{g}_t\}$. Но мы должны использовать для вычисления автоковариаций не истинный, а оцененный ряд $\{\widehat{\boldsymbol{g}}\}$. По этой причине нам нужно подходящее условие на четвертый момент (о формулировании которого мы не будем беспокоиться) для оцененных автоковариаций, чтобы они были состоятельными. Доказательство состоятельности здесь не дается, поскольку оно очень близко доказательству утверждения 3.4.

При заданных оцененных автоковариациях следует рассмотреть две возможности вычисления $\widehat{\boldsymbol{S}}$. Если мы *априори* знаем, что $\mathbf{\Gamma}_j = \mathbf{0}$ для $j > q$, где q известно и конечно, то ясно, что $\boldsymbol{S}$ может быть состоятельно оценена как

$$\widehat{\boldsymbol{S}} = \widehat{\mathbf{\Gamma}}_0 + \sum_{j=1}^{q}(\widehat{\mathbf{\Gamma}}_j + \widehat{\mathbf{\Gamma}}'_j) = \sum_{j=-q}^{q}\widehat{\mathbf{\Gamma}}_j \quad (\text{напомним: } \widehat{\mathbf{\Gamma}}_{-j} = \widehat{\mathbf{\Gamma}}'_j). \tag{6.6.4}$$

Более сложным является другой случай, когда мы не знаем значения q (которое может быть, а может и не быть конечным). Как мы можем оценить долговременную дисперсионную матрицу, которая включает возможно бесконечно много параметров? Есть два подхода.

Использование ядер для оценивания S

Недавно были предложены несколько процедур для оценивания долговременной дисперсионной матрицы[1] $\boldsymbol{S}$. Так называемые **ядерные оценки** (*kernel-based estimators*) (или «непараметрические») могут быть представлены как взвешенное среднее оцененных автоковариаций:

$$\widehat{\boldsymbol{S}} = \sum_{j=-n+1}^{n-1} k\left(\frac{j}{q(n)}\right) \cdot \widehat{\mathbf{\Gamma}}_j. \tag{6.6.5}$$

[1] Для того чтобы результаты, сформулированные в этом подпараграфе и в следующем, выполнялись, нам необходимо наложить ряд технических условий (в дополнение к условию Гордина выше), которые еще более ограничивают характер сериальной корреляции. Относительно формулирования этих условий и более детального обсуждения темы, рассматриваемой в этом подпараграфе и в следующем, см.: [Den Haan and Levin, 1996b].

Здесь функция $k(\cdot)$, которая задает веса автоковариациям, называется **ядром** (*kernel*), а $q(n)$ называется **шириной окна** (*bandwidth*). Ширина окна может зависеть от объема выборки n. Оценка (6.6.4) является частным случаем ядерной оценки с $q(n) = q$ и

$$k(x) = \begin{cases} 1 & \text{для } |x| \leq 1, \\ 0 & \text{для } |x| > 1. \end{cases} \tag{6.6.6}$$

Это ядро называется **усеченным ядром** (*truncated kernel*). В данном контексте неизвестного q мы можем использовать усеченное ядро с шириной окна $q(n)$, которая возрастает вместе с объемом выборки. Если $q(n)$ увеличивается до бесконечности, в вычисление $\widehat{\boldsymbol{S}}$ может быть включено все больше и больше (и в конечном итоге все) автоковариаций. Однако на конечных выборках не гарантируется положительная полуопределенность этой усеченной ядерной оценки. Это является проблемой, поскольку тогда оцененная асимптотическая дисперсия GMM-оценки может не быть положительно полуопределенной.

[Newey and West, 1987] заметили, что ядерная оценка может быть сделана неотрицательно определенной на конечных выборках, если ее ядро является **ядром Бартлетта** (*Bartlett kernel*):

$$k(x) = \begin{cases} 1 - |x| & \text{для } |x| \leq 1, \\ 0 & \text{для } |x| > 1. \end{cases} \tag{6.6.7}$$

Ядерная оценка для $\boldsymbol{S}$, основанная на ядре Бартлетта, называется (в эконометрике) **оценкой Ньюи — Веста** (*Newey — West estimator*). Например, для $q(n) = 3$ ядерная оценка включает автоковариации до двух (*не* трех) лагов:

$$\widehat{\boldsymbol{S}} = \widehat{\boldsymbol{\Gamma}}_0 + \left(\frac{2}{3}\right)(\widehat{\boldsymbol{\Gamma}}_1 + \widehat{\boldsymbol{\Gamma}}_1') + \left(\frac{1}{3}\right)(\widehat{\boldsymbol{\Gamma}}_2 + \widehat{\boldsymbol{\Gamma}}_2'). \tag{6.6.8}$$

В своем наиболее полном исследовании ядерного оценивания долговременной дисперсии [Andrews, 1991] проанализировал положительно полуопределенные ядра, а именно класс ядер (включая ядро Бартлетта), которые приводят к положительной полуопределенности $\widehat{\boldsymbol{S}}$. Он показал, что при некоторых соответствующих условиях регулярности, таких как скорость, с которой $\widehat{\boldsymbol{S}}$ сходится (в среднем квадратическом) к $\boldsymbol{S}$, зависит от ядра и $q(n)$. Самая высокая возможная скорость равна $n^{2/5}$ (то есть последовательность $n^{2/5} \cdot (\widehat{\boldsymbol{S}} - \boldsymbol{S})$ остается стохастически ограниченной)[1],

[1]Последовательность случайных величин $\{x_n\}$ называется **стохастически ограниченной** (*stochastically bounded*), что записывается как $x_n = O_p$, если для любого ε существует $M > 0$, такое что $\text{Prob}(|x_n| > M) < \varepsilon$ для всех n. Если последовательность сходится по распределению к некоторой случайной величине, то эта последовательность стохастически ограниченна.

что происходит, когда $q(n)$ растет со скоростью $n^{1/5}$ (то есть $q(n)/n^{1/5}$ остается ограниченной), а ядро принадлежит определенному подмножеству положительно полуопределенных ядер. Ядро Бартлетта не входит в это подмножество. Для ядра Бартлетта самая высокая возможная скорость равна $n^{1/3}$, что достигается при росте $q(n)$ со скоростью $n^{1/3}$. Следовательно, по крайней мере на асимптотических основаниях, положительно полуопределенные ядра из этого подмножества должны быть предпочтительнее ядра Бартлетта. Пример таких ядер — **квадратичное спектральное (QS)** (*quadratic spectral*) ядро:

$$k(x) = \frac{25}{12\pi^2 x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right). \tag{6.6.9}$$

Поскольку в QS-ядре $k(x) \neq 0$ для $|x| > 1$, все оцененные автоковариации $\widehat{\mathbf{\Gamma}}_j$ $(j = 0, 1, \ldots, n-1)$ входят в вычисление $\widehat{\boldsymbol{S}}$, даже если $q(n) < n-1$.

Надо отметить, что знания скорости роста недостаточно для определения значения ширины окна для конкретных данных конечной длины. [Andrews, 1991] рассмотрел также формулу для определения ширины окна на основе данных, которая не только растет с оптимальной скоростью, но и минимизирует некоторый соответствующий критерий (асимптотическую среднеквадратическую ошибку). Но даже в этой формуле некоторые значения параметров должны быть предложены пользователем, так что эта формула на самом деле не предоставляет автоматический выбор ширины окна на основе данных.

VARHAC

Другая процедура для оценивания $\boldsymbol{S}$, называемая состоятельной к гетероскедастичности и автокорреляции VAR (**VARHAC**) оценкой, была предложена в [Den Haan and Levin, 1996a]. Идея заключается в том, чтобы подогнать VAR конечного порядка к K-мерному ряду $\{\widehat{\boldsymbol{g}}\}$, а затем построить долговременную ковариационную матрицу, подразумеваемую оцененной VAR. Это вовсе не означает, что $\boldsymbol{g}_t$ предполагается VAR конечного порядка. Фактически в этой процедуре (а также и в процедуре, использующей ядро) $\boldsymbol{g}_t$ даже не обязательно должен быть линейным процессом, не говоря уже об авторегрессии конечного порядка. Процедура VARHAC состоит из двух шагов.

Шаг 1: **Выбор глубины запаздываний для каждого уравнения VAR.** В этом шаге для каждого уравнения VAR вы определяете глубину запаздываний на основе Байесовского информационного критерия (BIC). Пусть $\widehat{g}_{kt}$ — k-й элемент K-мерного вектора $\widehat{\boldsymbol{g}}_t$. Тогда k-е уравнение VAR может быть записано в виде:

$$\widehat{g}_{kt} = \phi_{11}^{(k)}\widehat{g}_{1,t-1} + \phi_{12}^{(k)}\widehat{g}_{1,t-2} + \cdots + \phi_{1p}^{(k)}\widehat{g}_{1,t-p} +$$

$$+ \phi_{21}^{(k)}\widehat{g}_{2,t-1} + \phi_{22}^{(k)}\widehat{g}_{2,t-2} + \cdots + \phi_{2p}^{(k)}\widehat{g}_{2,t-p}$$

$$\ldots$$

$$\phi_{K1}^{(k)}\widehat{g}_{K,t-1} + \phi_{K2}^{(k)}\widehat{g}_{K,t-2} + \cdots + \phi_{Kp}^{(k)}\widehat{g}_{K,t-p} + e_{kt} =$$

$$= \boldsymbol{\phi}_1^{(k)}\widehat{\boldsymbol{g}}_{t-1} + \cdots + \boldsymbol{\phi}_p^{(k)}\widehat{\boldsymbol{g}}_{t-p} + e_{kt}, \qquad (6.6.10)$$

где

$$\underset{(1\times K)}{\boldsymbol{\phi}_j^{(k)}} = (\phi_{1j}^{(k)}, \ldots, \phi_{Kj}^{(k)}), \quad (j = 1, 2, \ldots, p).$$

При выборе глубины запаздываний по BIC должна быть специфицирована максимальная глубина запаздываний $p_{\max}$. В параграфе 6.4, когда BIC вводился в контексте оценивания истинного порядка (глубины запаздываний) авторегрессии конечного порядка, $p_{\max}$ полагался некоторому целому числу, о котором было известно, что оно больше или равно истинной глубине запаздываний. В текущем контексте, где $\boldsymbol{g}_t$ не обязательно является VAR конечного порядка, [Den Haan and Levin, 1996a, Theorem 2(c)] показывают, что оценка VARHAC является состоятельной для $\boldsymbol{S}$, когда $p_{\max}$ растет со скоростью $n^{1/3}$, и рекомендуют положить $p_{\max}$ равным $[n^{1/3}]$ (целая часть $n^{1/3}$). Чтобы подчеркнуть ее зависимость от объема выборки, мы обозначаем максимальную глубину запаздываний как $p_{\max}(n)$. Пусть $SSR_p^{(k)}$ является суммой квадратов остатков (SSR) при OLS-оценивании (6.6.10) (k-е уравнение в VAR) для $t = p_{\max}(n) + 1, p_{\max}(n) + 2, \ldots, n$[1]. Для $p = 0$ $SSR_p^{(k)} = \sum_{t=p_{\max}+1}^{n} (\widehat{g}_{kt})^2$, SSR в регрессии $\widehat{g}_{kt}$ при отсутствии регрессоров. Поскольку в k-м уравнении имеется pK коэффициентов, критерий BIC равен:

$$\log(SSR_p^{(k)}/n) + pK \cdot \log(n)/n. \qquad (6.6.11)$$

(в этом выражении n можно было бы заменить на фактический объем выборки $n - p_{\max}$ с отсутствием асимптотических последствий.) Пусть $p(k)$ равно значению p, которое минимизирует информационный критерий по $p = 0, 1, \ldots, p_{\max}(n)$ для k-го уравнения VAR. Также пусть $\widehat{\boldsymbol{\phi}}_j^{(k)}$ $(j = 1, 2, \ldots, p(k))$ является OLS-оценкой VAR-коэффициентов $\boldsymbol{\phi}_j^{(k)}$ $(j = 1, 2, \ldots, p(k))$, когда (6.6.10) оценивается посредством OLS для $p = p(k)$.

Шаг 2: **Вычисление подразумеваемой долговременной дисперсии.** Пусть

$$P \equiv \max\{p(1), p(2), \ldots, p(K)\},$$

[1] Так что период выборки фиксируется на протяжении всего процесса выбора глубины запаздываний.

наибольшая глубина запаздываний в K уравнениях. По построению, $P \leq p_{\max}(n)$. *Шаг 1* дает оцененную VAR, которая может быть записана как

$$\underset{(K\times 1)}{\widehat{\boldsymbol{g}}_t} = \underset{(K\times K)}{\widehat{\boldsymbol{\Phi}}_1}\underset{(K\times 1)}{\widehat{\boldsymbol{g}}_{t-1}} + \cdots + \underset{(K\times K)}{\widehat{\boldsymbol{\Phi}}_P}\underset{(K\times 1)}{\widehat{\boldsymbol{g}}_{t-P}} + \underset{(K\times 1)}{\widehat{\boldsymbol{e}}_t}$$
$$(t = p_{max}(n) + 1, p_{max}(n) + 2, \ldots, n). \quad (6.6.12)$$

Поскольку глубина запаздываний отличается для разных уравнений, некоторые строки в $\widehat{\boldsymbol{\Phi}}_p$ $(p = 1, 2, \ldots, P)$ могут быть нулевыми. Строки $\widehat{\boldsymbol{\Phi}}_p$ связаны с $\widehat{\boldsymbol{\phi}}_j^{(k)}$ $(j = 1, 2, \ldots, p(k))$, полученными в *шаге 1*, следующим образом:

$$k\text{-я строка в } \underset{(K\times K)}{\widehat{\boldsymbol{\Phi}}_p} = \begin{cases} \widehat{\boldsymbol{\phi}}_p^{(k)} & \text{for } 1 \leq p \leq p(k), \\ \mathbf{0} & \text{for } p(k) < p \leq P. \end{cases} \quad (6.6.13)$$

Пусть $\widehat{\boldsymbol{\Omega}}$ является ковариационной матрицей остатков VAR:

$$\underset{(K\times K)}{\widehat{\boldsymbol{\Omega}}} \equiv \frac{1}{n} \sum_{t=p_{max}(n)+1}^{n} \widehat{\boldsymbol{e}}_t \widehat{\boldsymbol{e}}_t'. \quad (6.6.14)$$

Вспомним, что долговременная ковариационная матрица VAR задается соотношением (6.3.16) для $z = 1$. Оценка для $\boldsymbol{S}$, подразумеваемая оцененной VAR (6.6.12), равна тогда:

$$\widehat{\boldsymbol{S}} = \left[\boldsymbol{I}_K - \sum_{p=1}^{P} \widehat{\boldsymbol{\Phi}}_p\right]^{-1} \widehat{\boldsymbol{\Omega}} \left[\boldsymbol{I}_K - \sum_{p=1}^{P} \widehat{\boldsymbol{\Phi}}_p'\right]^{-1}. \quad (6.6.15)$$

Эта VARHAC-оценка положительно полуопределена по построению. [Den Haan and Levin, 1996a] показали, что (при подходящих условиях регулярности) эта оценка сходится к $\boldsymbol{S}$ с более высокой скоростью, чем любая положительно определенная ядерная оценка, для почти любой автоковариационной структуры процесса $\boldsymbol{g}_t$ (который необязательно является VAR конечного порядка). В частности, если $\boldsymbol{g}_t$ является VARMA конечного порядка, то порядок запаздываний, выбранной по BIC, растет с логарифмической скоростью, и VARHAC-оценка сходится со скоростью, сколь угодно близкой к $n^{1/2}$, что быстрее, чем максимальная скорость $n^{2/3}$, достигаемая положительно полуопределенной ядерной оценкой.

Контрольные вопросы

1. Что является эффективной GMM-оценкой,когда $x_t = z_t$?

2. (Последствие игнорирования сериальной корреляции.) Рассмотрим эффективную GMM-оценку, которая предполагает, что $\{g_t\}$ сериально *не*коррелированна. Будет ли эта оценка состоятельной, если $\{g_t\}$ на самом деле сериально коррелирована? Будет ли она эффективной?

6.7. Оценивание при условной гомоскедастичности (дополнительно)

Мы видели в параграфе 3.8, что при условной гомоскедастичности GMM-оценка редуцируется к 2SLS-оценке. Этот параграф показывает, как 2SLS можно обобщить для включения сериальной корреляции. При этом станет ясно, что GLS-оценка не является GMM-оценкой с сериальной корреляцией. Последняя половина этого параграфа изучает соотношение между GLS и GMM.

Ядерное оценивание S при условной гомоскедастичности

Соотношение между сериальной корреляцией в $g_t \equiv x_t \cdot \varepsilon_t$ и сериальной корреляцией в ошибках ε_t становится более ясным при наличии условной гомоскедастичности. Пусть

$$\omega_j \equiv \mathrm{E}(\varepsilon_t \varepsilon_{t-j}) \quad \text{для всех } j. \tag{6.7.1}$$

Поскольку $\{\varepsilon_t\}$ является стационарным по предположениям 3.1 и 3.2, ω_j не зависит от t. Если $\mathrm{E}(\varepsilon_t) = 0$ (что имеет место, когда x_t включает константу), то ω_j равна автоковариации $\{\varepsilon_t\}$ j-го порядка. Предположим, что ошибки условно гомоскедастичны:

условная гомоскедастичность:

$$\mathrm{E}(\varepsilon_t \varepsilon_{t-j} | x_t, x_{t-j}) = \omega_j \quad (j = 0, \pm 1, \pm 2, \dots). \tag{6.7.2}$$

При этом дополнительном условии к Γ_j можно применить обычные аргументы, использующие закон полных математических ожиданий:

$$\begin{aligned}
\Gamma_j &= \mathrm{E}(g_t g'_{t-j}) \quad (\text{поскольку } \mathrm{E}(g_t) = 0) = \\
&= \mathrm{E}(\varepsilon_t \varepsilon_{t-j} x_t x'_{t-j}) = \\
&= \mathrm{E}[\mathrm{E}(\varepsilon_t \varepsilon_{t-j} | x_t, x_{t-j}) x_t x'_{t-j}] \\
&\quad (\text{по закону полных математических ожиданий}) \\
&= \omega_j \, \mathrm{E}(x_t x'_{t-j}).
\end{aligned} \tag{6.7.3}$$

Таким образом, если $\mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}'_{t-j})$ не равно нулю, $\{\boldsymbol{g}_t\}$ сериально коррелированы тогда и только тогда, когда $\{\varepsilon_t\}$ сериально коррелированы.

Для оценивания $\boldsymbol{\Gamma}_j$ мы используем тот факт, что $\boldsymbol{\Gamma}_j$ является произведением двух вторых моментов, $\mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}'_{t-j})$ и $\mathrm{E}(\varepsilon_t\varepsilon_{t-j})$. Естественной оценкой для $\mathrm{E}(\varepsilon_t\varepsilon_{t-j})$ является $\frac{1}{n}\sum_{t=j+1}^{n}\boldsymbol{x}_t\boldsymbol{x}'_{t-j}$. Легко показать, что ω_j состоятельно оценивается посредством $\frac{1}{n}\sum_{t=j+1}^{n}\widehat{\varepsilon}_t\widehat{\varepsilon}_{t-j}$, где $\widehat{\varepsilon}_t$ является остатком для некоторой состоятельной оценки (доказательство почти то же самое, как в случае с отсутствием сериальной корреляции, см. утверждение 3.2). Таким образом, естественной оценкой для $\boldsymbol{\Gamma}_j$ при условной гомоскедастичности является

$$\widehat{\boldsymbol{\Gamma}}_j = \Big(\frac{1}{n}\sum_{t=j+1}^{n}\widehat{\varepsilon}_t\widehat{\varepsilon}_{t-j}\Big)\Big(\frac{1}{n}\sum_{t=j+1}^{n}\boldsymbol{x}_t\boldsymbol{x}'_{t-j}\Big). \tag{6.7.4}$$

При наличии этих оцененных автоковариаций ядерное оценивание $\widehat{\boldsymbol{S}}$ производится точно так же, как и в случае отсутствия условной гомоскедастичности, описанном в предыдущем параграфе.

Представление оцененной долговременной дисперсии через матрицу данных

С целью соотнесения GMM-оценки с сериальной корреляцией, но с условной гомоскедастичностью и обычных оценок полезно представление $\widehat{\boldsymbol{S}}$ через матрицу данных. Определяя $n \times n$ матрицу $\widehat{\boldsymbol{\Omega}}$ соответствующим образом, мы можем записать $\widehat{\boldsymbol{S}}$ как

$$\widehat{\boldsymbol{S}} = \frac{1}{n}\boldsymbol{X}'\widehat{\boldsymbol{\Omega}}\boldsymbol{X}, \tag{6.7.5}$$

где $\boldsymbol{X}$ является матрицей данных размера $n \times K$, t-я строка которой равна $\boldsymbol{x}'_t$. Матрица $\widehat{\boldsymbol{\Omega}}$ выглядит так же, как автоковариационная матрица $\{\varepsilon_t\}$:

$$\underset{(n\times n)}{\widehat{\boldsymbol{\Omega}}} = \begin{bmatrix} \widehat{\omega}_0 & \widehat{\omega}_1 & \widehat{\omega}_2 & \cdots & \widehat{\omega}_{n-1} \\ \widehat{\omega}_1 & \widehat{\omega}_0 & \widehat{\omega}_1 & \cdots & \widehat{\omega}_{n-2} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \widehat{\omega}_{n-2} & \cdots & \widehat{\omega}_1 & \widehat{\omega}_0 & \widehat{\omega}_1 \\ \widehat{\omega}_{n-1} & \cdots & \widehat{\omega}_2 & \widehat{\omega}_1 & \widehat{\omega}_0 \end{bmatrix}. \tag{6.7.6}$$

Если мы *априорно* знаем, что $\mathrm{Cov}(\varepsilon_t, \varepsilon_{t-j}) = 0$ (так что $\boldsymbol{\Gamma}_j = \boldsymbol{0}$) для $j > q$, то специфицируем элементы этой $\widehat{\boldsymbol{\Omega}}$ как

$$\widehat{\omega}_j = \begin{cases} \frac{1}{n}\sum_{t=j+1}^{n}\widehat{\varepsilon}_t\widehat{\varepsilon}_{t-j}, & \text{если } 0 \le j \le q, \\ 0, & \text{если } j > q. \end{cases} \tag{6.7.7}$$

Тогда $\widehat{\boldsymbol{S}}$ в (6.6.4) может быть записана как (6.7.5). Если мы не знаем, что существует такое q, то используем ядерную оценку для $\boldsymbol{S}$. Для ядра Бартлетта положим

$$\widehat{\omega}_j = \begin{cases} \left[1 - \dfrac{j}{q(n)}\right] \dfrac{1}{n} \displaystyle\sum_{t=j+1}^{n} \widehat{\varepsilon}_t \widehat{\varepsilon}_{t-j}, & \text{если } 0 \le j \le q(n) - 1, \\ 0, & \text{если } j > q(n) - 1. \end{cases} \tag{6.7.8}$$

Тогда состоятельная оценка (6.6.3) с (6.6.4) равна (6.7.5).

Теперь легко показать следующее обобщение 2SLS-оценки с целью охвата сериальной корреляции.

- Эффективная GMM-оценка при условной гомоскедастичности может быть записана как

$$\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}) = [\boldsymbol{Z}'\boldsymbol{X}(\boldsymbol{X}'\widehat{\boldsymbol{\Omega}}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}]^{-1}\boldsymbol{Z}'\boldsymbol{X}(\boldsymbol{X}'\widehat{\boldsymbol{\Omega}}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}, \tag{6.7.9}$$

  где $\boldsymbol{Z}$ и $\boldsymbol{y}$ — матрицы данных для регрессоров и зависимой переменной. Это обобщает выражение (3.8.3′) на с. 259.

- Состоятельная оценка $\mathrm{Avar}(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}))$ редуцируется к

$$\begin{aligned} \widehat{\mathrm{Avar}(\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{S}}^{-1}))} &= \left[\frac{1}{n}\boldsymbol{Z}'\boldsymbol{X}\left(\frac{1}{n}\boldsymbol{X}'\widehat{\boldsymbol{\Omega}}\boldsymbol{X}\right)^{-1}\frac{1}{n}\boldsymbol{X}'\boldsymbol{Z}\right]^{-1} = \\ &= n \cdot \left[\boldsymbol{Z}'\boldsymbol{X}(\boldsymbol{X}'\widehat{\boldsymbol{\Omega}}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}\right]^{-1}. \end{aligned} \tag{6.7.10}$$

  Поэтому квадратные корни из диагональных элементов

$$\left[\boldsymbol{Z}'\boldsymbol{X}(\boldsymbol{X}'\widehat{\boldsymbol{\Omega}}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Z}\right]$$

  являются стандартными ошибками. Это обобщает выражение (3.8.5′) на с. 259.

Связь с GLS

Процедура для включения сериальной корреляции в GMM-оценивание, которую мы описали, отличается от GLS-процедуры. Чтобы увидеть это отличие более ясно, рассмотрим случай, когда $\boldsymbol{x}_t = \boldsymbol{z}_t$. Если имеется L регрессоров, то $\boldsymbol{X}'\boldsymbol{Z}$ и $\boldsymbol{X}'\widehat{\boldsymbol{\Omega}}\boldsymbol{X}$ в (6.7.9) являются матрицами размера $L \times L$, поэтому эффективной GMM-оценкой, которая использует условия ортогональности $\mathrm{E}(\boldsymbol{z}_t \cdot \varepsilon_t) = \boldsymbol{0}$, является OLS-оценка $\widehat{\boldsymbol{\delta}}_{OLS} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{y}$. Состоятельная оценка ее асимптотической дисперсии получается, если положить $\boldsymbol{X} = \boldsymbol{Z}$ в (6.7.10):

$$\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\delta}}_{\mathrm{OLS}})} = n \cdot (\boldsymbol{Z}'\boldsymbol{Z})^{-1}(\boldsymbol{Z}'\widehat{\boldsymbol{\Omega}}\boldsymbol{Z})(\boldsymbol{Z}'\boldsymbol{Z})^{-1}. \tag{6.7.11}$$

Это может показаться странным, поскольку мы знаем из параграфа 1.6, что GLS-оценка является более эффективной, чем OLS, когда ошибки сериально коррелированы. Этот результат относится к конечным выборкам, но при условии, что у нас есть состоятельная оценка $\widehat{\boldsymbol{\Omega}}$ для $\boldsymbol{\Omega}$, мы можем взять формулу GLS и оценить $\boldsymbol{\delta}$ как

$$\widehat{\boldsymbol{\delta}}_{\mathrm{GLS}} = (\boldsymbol{Z}'\widehat{\boldsymbol{\Omega}}^{-1}\boldsymbol{Z})^{-1}\boldsymbol{Z}'\widehat{\boldsymbol{\Omega}}^{-1}\boldsymbol{y}.$$

Не является ли эта оценка асимптотически более эффективной, чем OLS, в том смысле, что ее асимптотическая дисперсия меньше? Проблема с GLS заключается в том, что состоятельность *не* гарантируется, если регрессоры просто предопределенные, как показано ниже. Как мы подчеркивали в главе 2, во многих моделях временных рядов регрессоры не являются строго экзогенными. Отсюда следует, что ***GLS не должен использоваться для коррекции сериальной корреляции в ошибках для моделей, не имеющих строгой экзогенности.*** В частности, для случая $\boldsymbol{x}_t = \boldsymbol{z}_t$ корректная процедура для коррекции сериальной корреляции заключается в том, чтобы оставить точечную оценку неизменной и в то же время включить сериальную корреляцию в оценку асимптотической дисперсии.

То, что GLS-оценка в общем случае является несостоятельной, можно увидеть следующим образом. Чтобы сосредоточиться на решаемой проблеме, предположим, что истинная автоковариационная матрица пропорциональна некоторой известной $n \times n$ матрице $\boldsymbol{V}$:

$$\mathrm{Var}(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n) = \sigma^2 \cdot \boldsymbol{V}.$$

Как показано в параграфе 1.6, GLS-оценка коэффициента $\boldsymbol{\delta}$ в регрессионной модели $\boldsymbol{y} = \boldsymbol{Z}\boldsymbol{\delta} + \boldsymbol{\varepsilon}$ может быть вычислена как OLS-оценка в преобразованной модели $\boldsymbol{C}\boldsymbol{y} = \boldsymbol{C}\boldsymbol{Z}\boldsymbol{\delta} + \boldsymbol{C}\boldsymbol{\varepsilon}$, где $\boldsymbol{C}$ является «квадратным корнем» из $\boldsymbol{V}^{-1}$, так что $\boldsymbol{C}'\boldsymbol{C} = \boldsymbol{V}^{-1}$. Но посмотрим на t-е наблюдение преобразованной модели, которое может быть записано как

$$\widetilde{y}_t = \widetilde{\boldsymbol{z}}_t'\boldsymbol{\delta} + \widetilde{\varepsilon}_t, \tag{6.7.12}$$

где

$$\widetilde{y}_t = c_{t1}y_1 + \cdots + c_{tn}y_n,$$
$$\widetilde{\boldsymbol{z}}_t = c_{t1}\boldsymbol{z}_1 + \cdots + c_{tn}\boldsymbol{z}_n,$$
$$\widetilde{\varepsilon}_t = c_{t1}\varepsilon_1 + \cdots + c_{tn}\varepsilon_n,$$
$$(c_{t1}, \ldots, c_{tn}) \text{ — } t\text{-я строка в } \boldsymbol{C}.$$

Для состоятельности GLS-оценки необходимо, чтобы $\mathrm{E}(\widetilde{\boldsymbol{z}}_t \cdot \widetilde{\varepsilon}_t) = \boldsymbol{0}$. Левая часть этого условия может быть записана как

$$\mathrm{E}(\widetilde{\boldsymbol{z}}_t \cdot \widetilde{\varepsilon}_t) = (\mathrm{a}) + (\mathrm{b}),$$

$$(\mathrm{a}) = \sum_{s=1}^{n} c_{ts}^2 \,\mathrm{E}(\boldsymbol{z}_s \cdot \varepsilon_s), \quad (\mathrm{b}) = \sum_{s \neq v} c_{ts} \cdot c_{tv} \,\mathrm{E}(\boldsymbol{z}_s \cdot \varepsilon_v).$$

По условию ортогональности $\mathrm{E}(\boldsymbol{z}_t \cdot \varepsilon_t) = 0$, (a) равно нулю. Что касается (b), то поскольку регрессоры просто предопределены и не предполагаются строго экзогенными, то нет гарантии того, что $\mathrm{E}(\boldsymbol{z}_s \cdot \varepsilon_v)$ равно нулю для $s \neq v$. Поэтому (b) не равно нулю, если c не принимают определенные значения, для того чтобы компенсировать ненулевые слагаемые.

Существует один важный частный случай, в котором GLS-оценка состоятельна, — это случай, когда ошибки образуют процесс авторегрессии конечного порядка. Чтобы взять самый простой случай, предположим, что $\{\varepsilon_t\}$ является $\mathrm{AR}(1)$: $\varepsilon_t = \phi\varepsilon_{t-1} + \eta_t$. Его автоковариационная матрица равна:

$$\mathrm{Var}(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n) = \sigma^2 \cdot \boldsymbol{V}, \quad \boldsymbol{V} = \frac{1}{1-\phi^2} \begin{bmatrix} 1 & \phi & \phi^2 & \cdots & \phi^{n-1} \\ \phi & 1 & \phi & \cdots & \phi^{n-2} \\ & & \ddots & & \\ \phi^{n-2} & \cdots & \phi & 1 & \phi \\ \phi^{n-1} & \cdots & \phi^2 & \phi & 1 \end{bmatrix}.$$

Легко проверить, что квадратный корень из $\boldsymbol{V}^{-1}$, $\boldsymbol{C}$ (удовлетворяющий $\boldsymbol{C}'\boldsymbol{C} = \boldsymbol{V}^{-1}$), имеет вид:

$$\boldsymbol{C} = \begin{bmatrix} \sqrt{1-\phi^2} & 0 & 0 & \cdots & 0 & 0 \\ -\phi & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\phi & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\phi & 1 \end{bmatrix}. \tag{6.7.13}$$

Поэтому t-е преобразованное уравнение имеет вид:

$$y_t - \phi y_{t-1} = (\boldsymbol{z}_t - \phi \boldsymbol{z}_{t-1})'\boldsymbol{\delta} + \eta_t. \tag{6.7.14}$$

GLS-оценка для $\boldsymbol{\delta}$, которая является OLS-оценкой в регрессии $y_t - \phi y_{t-1}$ на $\boldsymbol{z}_t - \phi \boldsymbol{z}_{t-1}$, является состоятельной, если η_t ортогональна *и* $\boldsymbol{z}_t$, *и* $\boldsymbol{z}_{t-1}$. Однако эти условия ортогональности отличаются от условий ортогональности для непреобразованной модели, которые заключаются в том, что $\mathrm{E}(\boldsymbol{z}_t \cdot \varepsilon_t) = \boldsymbol{0}$.

Контрольные вопросы

1. Проверьте (6.7.5) для $n = 3$, $q = 1$.

2. (Статистика Саргана.) Получите обобщение статистики Саргана (формула (3.8.10′) на с. 259) на случай сериально коррелированных ошибок.

3. Предположим, что $\boldsymbol{z}_t$ является собственным подмножеством $\boldsymbol{x}_t$. Мы видели в параграфе 3.8, что оценка 2SLS редуцируется к OLS-оценке. Верно ли это также и здесь? То есть редуцируется ли (6.7.9) к $(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{y}$? [Ответ: Нет.]

4. (Представление $\widehat{\boldsymbol{S}}$ через матрицу данных без условной гомоскедастичности.) Определите $\widehat{\boldsymbol{\Omega}}$ подходящим образом так, чтобы (6.7.5) было равно ядерной оценке Бартлетта для $\widehat{\boldsymbol{S}}$ без условной гомоскедастичности, а именно (6.6.5) с (6.6.7). **Указание:** *Запишите* $\widehat{\boldsymbol{\Omega}}$ *как (6.7.6). Если глубина запаздываний* q *известна, то* $\widehat{\omega}_{ij} = \widehat{\varepsilon}_i\widehat{\varepsilon}_j$ *для* $|i-j| \le q$ *и равна 0 для* $|i-j| > q$, *где* $\widehat{\omega}_{ij}$ *является* (i,j)*-м элементом в* $\widehat{\boldsymbol{\Omega}}$. *Как вы определите* $\widehat{\omega}_{ij}$ *для ядерной оценки Бартлетта?*

5. Рассмотрим простую регрессионную модель: $y_t = \beta z_t + \varepsilon_t$, где $\{\varepsilon_t\}$ является AR(1), $\varepsilon_t = \phi\varepsilon_{t-1} + \eta_t$ с известным ϕ. Предположим, что $\mathrm{E}(z_t\varepsilon_t) = 0$, $\mathrm{E}(z_t\eta_t) = 0$, $\mathrm{E}(z_{t-1}\eta_t) = 0$, и условную гомоскедастичность. Также, для упрощения, пусть $\mathrm{E}(z_t) = 0$. Покажите, что

$$\mathrm{Avar}(\widehat{\beta}_{OLS}) = \frac{\sigma_\eta^2}{\sigma_z^2}\frac{1+2\sum_{j=1}^{\infty}\phi^j\rho_j}{1-\phi^2}, \quad \mathrm{Avar}(\widehat{\beta}_{GLS}) = \frac{\sigma_\eta^2}{\sigma_z^2}\frac{1}{(1+\phi^2)-2\phi\rho_1},$$

где $\sigma_z^2 \equiv \mathrm{Var}(z_t)$, $\sigma_\eta^2 \equiv \mathrm{Var}(\eta_t)$, $\rho_j \equiv \mathrm{Corr}(z_t, z_{t-j})$. Найдите такую конфигурацию для $(\phi, \{\gamma_j\})$, чтобы $\mathrm{Avar}(\widehat{\beta}_{OLS}) > \mathrm{Avar}(\widehat{\beta}_{GLS})$. Объясните, почему это соответствует тому факту, что $\widehat{\beta}_{OLS}$ является эффективной GMM-оценкой.

6.8. Приложение: Форвардные обменные курсы как оптимальные предикторы

На валютных рынках процентная разница между форвардным и спотовым обменными курсами называется **форвардной премией** (*forward premium*). Соотношение между форвардной премией и ожидаемым темпом укрепления/ослабления валюты в течение срока действия форвардного контракта является предметом большого количества эмпирических исследований. Самая простая гипотеза об этом соотношении заключается в том, что эти два значения одинаковы. В данном параграфе мы проверяем эту гипотезу, привлекая те же самые недельные данные, которые были использованы в [Bekaert and Hodrick, 1993].

Данные охватывают период с 1975 по 1989 год и включают следующие переменные:

S_t = спотовый обменный курс в единицах иностранной валюты за доллар (например, 125 иен за доллар) в пятницу t-й недели,

R_t = 30-дневный форвардный обменный курс в пятницу t-й недели, а именно цена доллара в единицах иностранной валюты, поставляемой через 30 дней после пятницы t-й недели,

$S30_t$ = спотовая ставка на дату поставки 30-дневного форвардного контракта, созданного в пятницу t-й недели.

Важной особенностью этих данных является то, что срок погашения контракта охватывает несколько интервалов выборки; датой поставки является дата после пятницы $t+4$-й недели, но до пятницы $t+5$-й недели.

Гипотеза эффективности рынка

Чтобы наметить предположения, лежащие в основе этой гипотезы, рассмотрим три стратегии инвестирования долларовых средств на 30 дней в момент t. Первая заключается в инвестировании в инструмент внутреннего денежного рынка (например, казначейские векселя США), срок погашения которого составляет 30 дней. Пусть i_t является нормой доходности за 30-дневный период. Вторая инвестиционная стратегия заключается в конвертации долларов в единицы иностранной валюты, покупкой 30-дневных инструментов денежного рынка, номинированных в иностранной валюте, процентная ставка которой обозначается как i_t^*, и конвертацией обратно в доллары в момент погашения. Норма доходности этой инвестиционной стратегии равна:

$$S_t \cdot (1 + i_t^*)/S30_t.$$

Эта стратегия включает риск изменения курса валюты, поскольку в момент инвестирования, t, вы не знаете значения $S30_t$, которое будет являться спотовой ставкой через 30 дней. Третья инвестиционная стратегия отличается от второй тем, что она хеджирует против риска изменения курса путем покупки долларового форварда по цене F_t в тот же самый момент, когда долларовые средства конвертируются в единицы иностранной валюты. Эта хеджированная иностранная инвестиция обеспечивает гарантированную доходность, равную:

$$S_t \cdot (1 + i_t^*)/F_t.$$

Поскольку первая и третья стратегии не включают риска, арбитраж требует, чтобы нормы доходности были одинаковыми:

$$1 + i_t = S_t \cdot (1 + i_t^*)/F_t.$$

Беря логарифмы от обеих частей и используя аппроксимацию $\log(1+x) \approx \approx x$, мы получаем **уравнение покрытого процентного паритета** (*covered interest parity*):

$$i_t = s_t - f_t + i_t^*, \quad \text{или} \quad f_t - s_t = i_t^* - i_t. \tag{6.8.1}$$

где строчные буквы для S и F обозначают натуральные логарифмы. То есть форвардная премия, $f_t - s_t$, равна дифференциалу процентных ставок. Это соотношение выполняется почти точно в реальных данных. Фактически люди обычно вычисляют форвардную премию как дифференциал процентных ставок.

Норма доходности во второй стратегии не определена, но, если инвесторы нейтральны к риску, ее ожидаемая доходность равна гарантированной норме доходности. Кроме того, если их ожидания рациональны, ожидаемая доходность является условным математическим ожиданием, основанным на доступной информации о валюте, так что

$$\mathrm{E}[S_t \cdot (1 + i_t^*)/S30_t | I_t] = S_t \cdot (1 + i_t^*)/F_t, \tag{6.8.2}$$

где $\mathrm{E}(\cdot|I_t)$ является оператором математического ожидания, а I_t является информационным множеством по состоянию на дату t. Поскольку S_t, F_t и i_t^* известны в дату t, это равенство редуцируется к

$$\mathrm{E}(F_t/S30_t | I_t) = 1.$$

И опять, по логарифмической аппроксимации, $F_t/S30_t \approx 1 + f_t - s30_t$. Следовательно,

$$\mathrm{E}(s30_t | I_t) = f_t, \quad \text{или} \quad \mathrm{E}(\varepsilon_t | I_t) = 0, \quad \text{где } \varepsilon_t \equiv s30_t - f_t. \tag{6.8.3}$$

Это гипотеза эффективности рынка. Как ясно из вывода, она предполагает нейтральность к риску и рациональные ожидания. Эту гипотезу иногда описывают как то, что форвардная ставка является оптимальным прогнозом будущей спотовой ставки, поскольку условное математическое ожидание является оптимальным прогнозом в том смысле, что оно минимизирует среднеквадратичную ошибку (см. утверждение 2.7).

Тестирование равенства нулю безусловного среднего

Беря безусловное математическое ожидание от обеих сторон (6.8.3), мы получаем безусловное соотношение:

$$\mathrm{E}(s30_t) = \mathrm{E}(f_t), \quad \text{или} \quad \mathrm{E}(\varepsilon_t) = 0. \tag{6.8.4}$$

То есть прогноз должен быть в среднем правильным. Рис. 6.1 изображает ошибку прогноза ε_t для обменного курса иена/доллар. Она похожа на стационарный ряд, и мы исходим из предположения, что это так[1].

[1]Используя покрытый процентный паритет, получаем $\varepsilon_t = (s30_t - s_t) - (i_t^* - i_t)$. Если вы верите, что и 30-дневная скорость изменения $s30_t - s_t$ и дифференциал процентных ставок $i_t^* - i_t$ являются стационарными, то вы должны верить и в то, что ошибка прогноза является стационарной.

Поскольку ошибка прогноза наблюдаема, мы можем проверить гипотезу о том, что безусловное среднее равно нулю. В этом заключается отличие от упражнения, посвященного работе Фамы из главы 2, где ошибка прогноза (относительно будущей инфляции) наблюдаема только с точностью до константы.

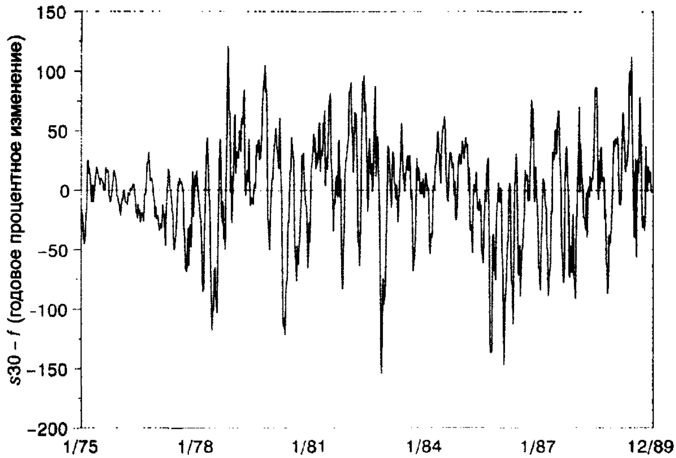

Рис. 6.1. Ошибка прогноза, иена/доллар

Для проверки гипотезы эффективности рынка нам нужно получить асимптотическое распределение выборочного среднего $\bar{\varepsilon}$ при этой гипотезе. Эта задача в определенной степени сложна, поскольку последовательность $\{\varepsilon_t\}$ сериально коррелирована. Чтобы увидеть, почему ошибка прогноза имеет сериальную корреляцию, заметим, что I_t, включая $\{\varepsilon_{t-5}, \varepsilon_{t-6}, \dots\}$, не включает $\{\varepsilon_t, \dots, \varepsilon_{t-4}\}$, поскольку ε_t, которая зависит от $s30_t$, не может наблюдаться до момента $t+4$. Поэтому

$$\mathrm{E}(\varepsilon_t | \varepsilon_{t-5}, \varepsilon_{t-6}, \dots) = 0. \tag{6.8.5}$$

Из этого следует, что $\mathrm{Cov}(\varepsilon_t, \varepsilon_{t-j}) = 0$ для $j \geq 5$, но необязательно для $j \leq 4$. В упражнении, посвященном работе Фамы, ex post реальная процентная ставка (которая равна прогнозу инфляции плюс константа) была сериально некоррелирована, поскольку срок погашения процентной ставки и интервал выборки были одинаковыми. Здесь же ошибка прогноза имеет сериальную корреляцию, поскольку срок погашения форвардного контракта (30 дней) длиннее, чем интервал выборки (неделя). Оцененная коррелограмма изображена вместе с полосой в две стандартные ошибки на рис. 6.2[1]. Стандартная ошибка вычисляется в предположении, что ошибка прогноза является i.i.d. Поэтому она равна $1/\sqrt{n}$, как

[1]Хотя теория подразумевает, что популяционное среднее равно нулю, я вычел выборочное среднее при вычислении выборочных автокорреляций.

в табл. 2.1. Поведение автокорреляции в целом соответствует гипотезе: она остается значительно выше двух стандартных ошибок для первых четырех лагов, а впоследствии, в основном, лежит внутри полосы.

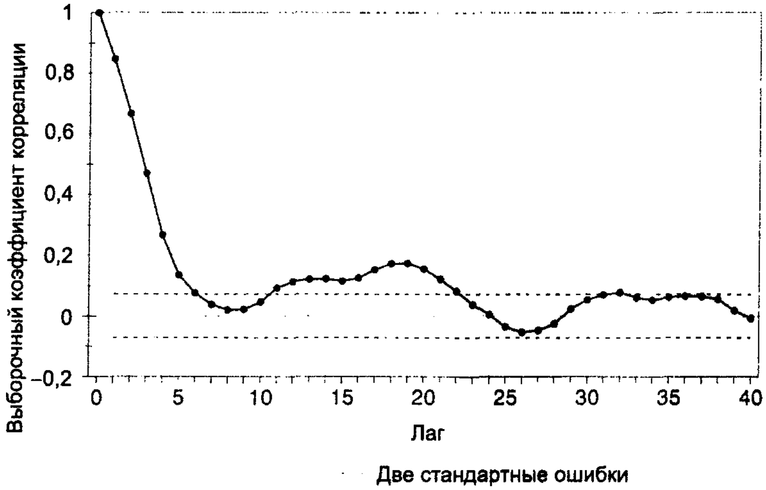

Рис. 6.2. Коррелограмма $s30 - f$, иена/доллар

При нулевой гипотезе об эффективности рынка, поскольку сериальная корреляция исчезает после конечного лага, легко видеть, что $\{\varepsilon_t\}$ удовлетворяет существенной части условия Гордина, ограничивающей сериальную корреляцию. По закону повторных математических ожиданий, из (6.8.5) следует $\mathrm{E}(\varepsilon_t|\varepsilon_{t-j}, \varepsilon_{t-j-1}, \dots) = 0$ для $j \geq 5$, из чего немедленно вытекает часть (b) условия Гордина. Поскольку пересмотр ожиданий, r_{tj}, равен нулю для $j \geq 5$, часть (c) выполняется, при условии, что r_{tj} $(0 \leq j \leq 4)$ имеет конечные четвертые моменты. Тогда из утверждения 6.10 следует, что

$$\frac{\sqrt{n}\,\bar{\varepsilon}}{\sqrt{\gamma_0 + 2\sum_{j=1}^{4}\gamma_j}} = \frac{\bar{\varepsilon}}{\sqrt{\left(\gamma_0 + 2\sum_{j=1}^{4}\gamma_j\right)/n}} \underset{d}{\to} N(0,1). \tag{6.8.6}$$

Кроме того, знаменатель состоятельно оценивается путем замены γ_j на ее выборочный аналог $\widehat{\gamma}_j$. Поэтому отношение

$$\frac{\bar{\varepsilon}}{\sqrt{\left(\widehat{\gamma}_0 + 2\sum_{j=1}^{4}\widehat{\gamma}_j\right)/n}} \tag{6.8.7}$$

также является асимптотически нормальным. Знаменатель этого выражения будет называться стандартной ошибкой выборочного среднего.

В табл. 6.1 приведены результаты тестирования следствия из гипотезы, состоящего в том, что безусловное среднее ошибки прогноза равно нулю, для трех курсов иностранных валют по отношению к доллару. Эти три валюты — немецкая марка, британский фунт и японская иена. В первых двух столбцах таблицы приведены среднее и стандартное отклонение фактической скорости изменения спотовой ставки, $s30_t - s_t$, и форвардной премии, $f_t - s_t$. (Поскольку берется спотовая ставка по отношению к доллару, положительная скорость изменения означает, что доллар укрепляется.) Месячные скорость изменения и форвардная премия были умножены на 1200, чтобы выразить их в годовых процентах. Гипотеза эффективности рынка заключается в том, что форвардная премия является ожидаемой скоростью изменения спотовой ставки. Очевидно, что фактическая скорость изменения намного более волатильна, чем форвардная премия. Третий столбец сообщает выборочное среднее для ε_t, которое равно разности между первыми двумя столбцами для выборочного среднего. Числа в скобках под выборочными средними представляют стандартные ошибки, вычисленные, как описывалось выше. Ни в одном случае нет значимого свидетельства против нулевой гипотезы о нулевом среднем.

Таблица 6.1

Средние скоростей изменения, форвардных премий и разности между ними: 1975–1989 годы

Обменный курс	Средние и стандартные отклонения		
	$s30 - s$ (фактическая скорость изменения)	**$f - s$ (ожидаемая скорость изменения)**	**Разность (неожидаемая скорость изменения)**
иена/доллар	−4,98	−3,74	−1,25
	(41,6)	(3,65)	(42,4)
			стандартная ошибка = 3,56
немецкая марка/доллар	−1,78	−3,91	2,13
	(40,6)	(2,17)	(41,1)
			стандартная ошибка = 3,26
фунт стерлингов/доллар	3,59	2,16	1,43
	(39,2)	(3,49)	(39,9)
			стандартная ошибка = 3,29

Примечание: $s30 - s$ — скорость изменения в течение 30 дней и $f - s$ — форвардная премия, выраженные в годовых процентных изменениях. Стандартные отклонения показаны в скобках. Стандартные ошибки — значения знаменателя в (6.8.7). Данные недельные с 1975 по 1989 г. Объем выборки — 778.

Регрессионные тесты

Безусловный тест, однако, не использует то следствие эффективности рынка, которое заключается в том, что ошибка прогноза ортогональна обусловливающей информации I_t. Вероятно, наиболее важной переменной в информационном множестве I_t для будущих спотовых ставок является форвардная ставка f_t. Мы можем тестировать, является ли ошибка прогноза некоррелированной с форвардной ставкой, регрессируя ε_t на константу и f_t. Это эквивалентно оцениванию следующей регрессии:

$$s30_t = \beta_0 + \beta_1 f_t + \varepsilon_t \tag{6.8.8}$$

и тестированию равенств $\beta_0 = 0$ и $\beta_1 = 1$. Поскольку ошибка является здесь ошибкой прогноза, ортогональной всему, что известно на дату t, включая f_t, регрессоры гарантированно будут ортогональными ошибкам. Поэтому может показаться, что для получения корректных статистических выводов мы можем оценить (6.8.8) посредством OLS и произвести коррекцию на сериальную корреляцию в ошибках, как указано в параграфе 6.6.

Рис. 6.3. Спотовая ставка иена/доллар, январь 1975 г. — декабрь 1989 г.

Однако, как показывает рис. 6.3, спотовый обменный курс выглядит как процесс с возрастающей дисперсией. Как мы убедимся в главе 9, мы не можем отвергнуть гипотезу о том, что этот процесс имеет единичный корень. Форвардная ставка обладает таким же свойством. (Ее график не показан здесь, поскольку он неотличим от графика спотовой ставки.) Таким образом, предположение 3.2, которое требует, чтобы регрессор (f_t) и зависимая переменная ($s30_t$) были стационарными, не выполняется. Проблема, в действительности, является более сложной. Мы видели на рис. 6.1, что *разность* между $s30_t$ и f_t ведет себя как стационарный ряд. На языке, вводимом далее в главе 10, $s30_t$ и f_t «коинтегрированы» с коинтегрирующим вектором $(1, -1)$, в этом случае OLS-оценка для β_1

в (6.8.8) быстро сходится к 1. Это можно увидеть на рис. 6.4, где показана зависимость $s30_t$ от f_t. Если получить оцененную линию регрессии, то ее угловой коэффициент равен 0,99. С точки зрения тестирования эффективности рынка, проблема состоит в том, что OLS-оценка сходится к 1 *независимо* от того, верна эта гипотеза или нет. Следовательно, у теста нет мощности.

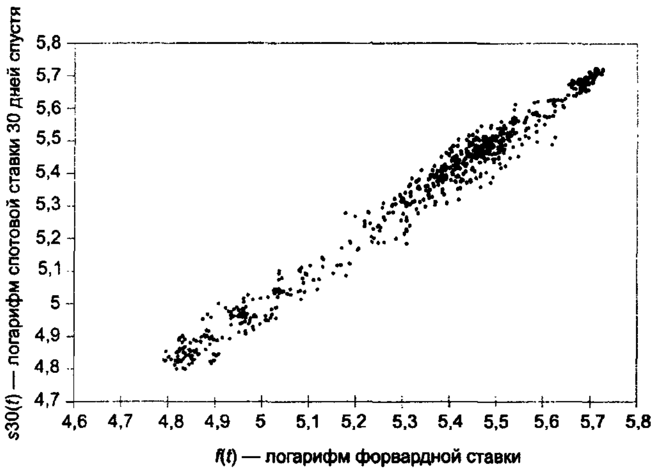

Рис. 6.4. График зависимости $s30$ от f, иена/доллар

Эта проблема может быть решена путем оценивания другой регрессии:

$$s30_t - s_t = \beta_0 + \beta_1(f_t - s_t) + \varepsilon_t. \tag{6.8.9}$$

Судя по графикам $s30_t - s_t$ и $f_t - s_t$ (не показаны), разумно предположить, что эти переменные являются стационарными, что соответствует предположению 3.2 (эргодическая стационарность). То, что при эффективности рынка это уравнение удовлетворяет другим предположениям GMM, можно проверить следующим образом. При $\beta_0 = 0$ и $\beta_1 = 1$ ошибка ε_t равна ошибке прогноза $s30_t - f_t$, и поэтому условия ортогональности $\mathrm{E}(\varepsilon_t)$ и $\mathrm{E}[(f_t - s_t)\cdot\varepsilon_t]$ (предположение 3.3 с $\boldsymbol{z} = \boldsymbol{x}$) выполняются. Условие идентификации заключается в том, что второй момент для $(1, f_t - s_t)$ является невырожденным, что верно тогда и только тогда, когда $\mathrm{Var}(f_t - s_t) > 0$. Это, очевидно, выполняется; если бы популяционная дисперсия была равна нулю, мы не наблюдали бы никакой вариации в $f_t - s_t$. Это оставляет нас с предположением $3.5'$, которое требует, чтобы в

$$\boldsymbol{g}_t \equiv \boldsymbol{x}_t \cdot \varepsilon_t = \begin{bmatrix} \varepsilon_t \\ (f_t - s_t)\varepsilon_t \end{bmatrix}$$

было не слишком много сериальной корреляции. Как уже было замечено, поскольку I_t включает $\{\varepsilon_{t-5}, \varepsilon_{t-6}, \dots\}$, ошибка прогноза удовлетворяет (6.8.5). В контрольном вопросе 3 вам будет предложено показать, что $\{\boldsymbol{g}_t\}$ наследует то же самое свойство, а именно:

$$\mathrm{E}(\boldsymbol{g}_t | \boldsymbol{g}_{t-5}, \boldsymbol{g}_{t-6}, \dots) = \mathbf{0}. \tag{6.8.10}$$

Поэтому, как и в безусловном тесте, условие Гордина выполняется, если соответствующие вторые моменты конечны. Поскольку $\mathrm{E}(\boldsymbol{g}_t \boldsymbol{g}'_{t-j}) = \mathbf{0}$ для $j \geq 5$, долговременную ковариационную матрицу $\boldsymbol{S}$ можно оценить, как в (6.6.4), с $q = 4$. (В эмпирическом упражнении вас попросят оценить $\widehat{\boldsymbol{S}}$ с помощью ядра Бартлетта, а также посредством VARHAC.)

Подводя итог, эффективной GMM-оценкой для $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ в (6.8.9) является OLS-оценка, и состоятельной оценкой ее асимптотической ковариационной матрицы является

$$\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}})} = \boldsymbol{S}_{\boldsymbol{xx}}^{-1}\, \widehat{\boldsymbol{S}}\, \boldsymbol{S}_{\boldsymbol{xx}}^{-1}, \tag{6.8.11}$$

где $\widehat{\boldsymbol{S}}$ задается в (6.6.4) с $q = 4$. В табл. 6.2 сообщаются результаты регрессионного теста для трех валют. Соответствующий график для обменного курса иена/доллар показан на рис. 6.5. Заметим, что для всех трех валют угловой коэффициент не только значимо отличается от 1,

Таблица 6.2

Регрессионные тесты на эффективность рынка: 1975–1989 годы

$s30_t - s_t = \beta_0 + \beta_1(f_t - s_t) + \varepsilon_t$

Валюта	**Регрессионные коэффициенты**			
	Константа	**Форвардная премия**	R^2	**Статистика Вальда для $H_0: \beta_0 = 0,\ \beta_1 = 1$**
иена/доллар	−12,8	−2,10	0,034	18,6
	(4,01)	(0,738)		(p = 0,009%)
немецкая марка/доллар	−13,6	−3,01	0,026	8,7
	(5,72)	(1,37)		(p = 1,312%)
фунт стерлингов/доллар	7,96	−2,02	0,033	12,9
	(3,54)	(0,85)		(p = 0,156%)

Примечание: Стандартные ошибки приведены в скобках. Они являются робастными к гетероскедастичности и допускают сериальную корреляцию до четырех лагов и вычислены по формуле (6.8.11). Объем выборки равен 778. Наши результаты несколько отличаются от результатов [Bekaert and Hodrick, 1993], поскольку эти авторы использовали для оценивания $\boldsymbol{S}$-оценку на основе ядра Бартлетта с $q(n) = 4$.

но и является отрицательным. Вытекающая из эффективности рынка нулевая гипотеза о том, что $\beta_0 = 0$ и $\beta_1 = 1$, может быть окончательно отвергнута. Это отражается в слишком большом количестве наблюдений в северо-западном квадранте на рис. 6.5; слишком часто ожидаемое ослабление доллара (отрицательные $f_t - s_t$) связано с фактическим *укреплением* доллара (положительные $s30_t - s_t$).

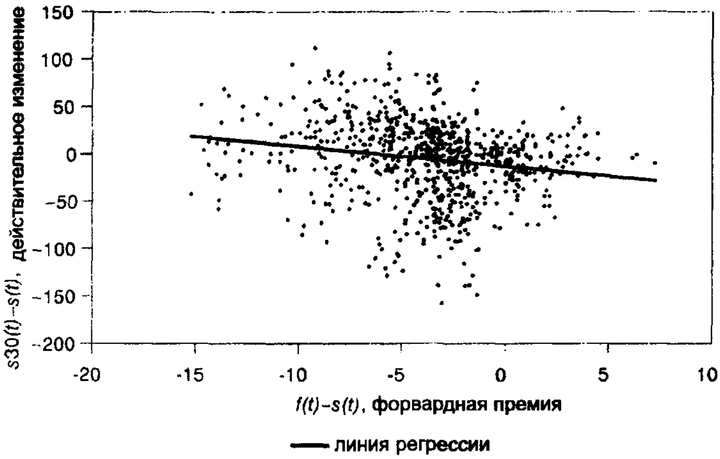

Рис. 6.5. График зависимости $s30 - s$ от $f - s$, иена/доллар

Контрольные вопросы

1. (Стандартная ошибка безусловного теста.) Положите $\varepsilon_t = s30_t - f_t$ и рассмотрите регрессию ε_t на константу. Для учета сериальной корреляции используйте (6.6.4) с $q = 4$ для $\widehat{S}$. Проверьте, что t-значение для гипотезы о том, что константа равна нулю, дается в (6.8.7).

2. (Отсутствие идентификации.) Для простоты предположим, что F_t является двухнедельным форвардным обменным курсом, так что $s30_t$ в (6.8.9) можно заменить на s_{t+2}. Предположим, что спотовый обменный курс является случайным блужданием и $I_t = \{s_t, s_{t-1}, \dots\}$. Каким должен быть f_t при нулевой гипотезе? Какое из требуемых предположений нарушается? [Ответ: Условие идентификации.] Предположим вместо этого, что s_t является AR(1)-процессом, удовлетворяющим $s_t = c + \phi s_{t-1} + \eta_t$, где $\{\eta_t\}$ является i.i.d., и $|\phi| < 1$. Проверьте, что $f_t = (1 + \phi)c + \phi^2 s_t$. **Указание:** $s_{t+2} = (1 + \phi)c + \phi^2 s_t + (\eta_{t+2} + \phi\eta_{t+1})$. *Проверьте, что* $\varepsilon_t = \eta_{t+2} + \phi\eta_{t+1}$.

3. Докажите (6.8.10). **Указание:** $\mathrm{E}(\varepsilon_t | I_t) = 0$. *Проверьте, что* I_t *включает* $s_t - f_t$ *и* $\{g_{t-5}, g_{t-6}, \dots\}$. *Используйте закон повторных математических ожиданий.*

Набор задач для главы 6

Аналитические упражнения

1. (Пределы в среднем квадратическом и их единственность.) В главе 2 мы определили сходимость в среднем квадратическом последовательности случайных величин $\{z_n\}$ к случайной величине z. В этом и в других вопросах мы будем иметь дело с последовательностями случайных величин с конечными вторыми моментами. Для таких последовательностей определение сходимости в среднем квадратическом таково:

   Определение. *Пусть $\{z_n\}$ является последовательностью случайных величин с $\mathrm{E}(z_n^2) < \infty$. Мы говорим, что $\{z_n\}$ сходится в среднем квадратическом к случайной величине z с $\mathrm{E}(z^2) < \infty$, записывая это как $z_n \to_{m.s.} z$, если $\mathrm{E}[(z_n - z)^2] \to 0$ при $n \to \infty$.*

   (a) Покажите, что если $z_n \to_{m.s.} z$ и $z_n \to_{m.s.} z'$, то $\mathrm{E}[(z - z')^2] = 0$. **Указание:** $z - z' = (z - z_n) + (z_n - z')$. *Определим* $\|x\| = \sqrt{\mathrm{E}(x^2)}$. *Тогда*

   $$\begin{aligned}\|z - z'\|^2 &= \|z - z_n\|^2 + 2\,\mathrm{E}[(z - z_n)(z_n - z')] + \|z_n - z'\|^2 \le \\ &\le \|z - z_n\|^2 + 2\sqrt{\mathrm{E}[(z - z_n)^2]}\sqrt{\mathrm{E}[(z_n - z')^2]} + \|z_n - z'\|^2 \le \\ &\text{(по неравенству Коши — Шварца } \mathrm{E}(xy) \le \sqrt{\mathrm{E}(x^2)}\sqrt{\mathrm{E}(y^2)}) = \\ &= (\|z - z_n\|^2 + \|z_n - z'\|^2)^2.\end{aligned}$$

   Поэтому $\|z - z'\| \le \|z - z_n\| + \|z_n - z'\|$. *Это называется неравенством треугольника.*

   (b) Покажите, что $\mathrm{Prob}(z = z') = 1$. **Указание:** *Используйте неравенство Чебышева:*

   $$\mathrm{Prob}\big(|z - z'| > \varepsilon\big) \le \frac{1}{\varepsilon^2}\|z - z'\|^2.$$

2. (Доказательство утверждения 6.1.) Хорошо известный результат о сходимости в среднем квадратическом (см., например: [Brockwell and Davis, 1991, Proposition 2.7.1] заключается в том, что

   (a) $\{z_n\}$ сходится в среднем квадратическом тогда и только тогда, когда $\mathrm{E}[(z_m - z_n)^2] \to 0$ при $m, n \to \infty$.

   (b) Если $x_n \to_{m.s.} x$ и $z_n \to_{m.s.} z$, то

   $$\lim_{n\to\infty} \mathrm{E}(x_n) = \mathrm{E}(x) \quad \text{и} \quad \lim_{n\to\infty} \mathrm{E}(x_n z_n) = \mathrm{E}(xz).$$

Отвечая, считайте этот результат доказанным.

(a) Докажите, что при выполнении предположений утверждения 6.1 (6.1.6) сходится в среднем квадратическом при $n \to \infty$. **Указание:** *Пусть*

$$y_{t,n} = \mu + \sum_{j=0}^{n} \psi_j \varepsilon_{t-j}.$$

Что должно быть доказано, так это то, что $\{y_{t,n}\}$ сходится в среднем квадратическом при $n \to \infty$ для каждого t. Учитывая (i) выше, достаточно доказать, что (предполагая $m > n$ без потери общности)

$$\mathrm{E}\Big[\Big(\sum_{j=n+1}^{m} \psi_j \varepsilon_{t-j}\Big)^2\Big] = \sigma^2 \sum_{j=n+1}^{m} \psi_j^2 \to 0 \text{ при } m, n \to \infty.$$

Используйте тот факт, что абсолютно суммируемая последовательность является ***квадратично суммируемой*** *(square summable), то есть*

$$\sum_{j=0}^{\infty} |\psi_j| < \infty \;\Rightarrow\; \sum_{j=0}^{\infty} \psi_j^2 < \infty,$$

и тот факт, что последовательность действительных чисел $\{\alpha_n\}$ сходится (к конечному пределу) тогда и только тогда, когда α_n является последовательностью Коши:

$$\alpha_n \to \alpha \;\Leftrightarrow\; |\alpha_m - \alpha_n| \to 0 \text{ при } m, n \to \infty.$$

Положите $\alpha_n = \sum_{j=0}^{n} \psi_j^2$.

(b) Докажите, что $\mathrm{E}(y_t) = \mu$. **Указание:** *Пусть $y_{t,n} = \mu + \sum_{j=0}^{n} \psi_j \varepsilon_{t-j}$, как в (a). В (a) было показано, что $y_{t,n} \to_{m.s.} y_t$.*

(c) (Доказательство части (b) утверждения 6.1.) Покажите, что

$$\mathrm{E}[(y_t - \mu)(y_{t-j} - \mu)] = \lim_{n \to \infty} \mathrm{E}[(y_{t,n} - \mu)(y_{t-j,n} - \mu)].$$

Показали ли мы, что $\{y_t\}$ является ковариационно стационарным? [Ответ: Да.]

(d) (Дополнительно.) Докажите часть (c) утверждения 6.1, принимая как доказанные следующие факты из математического анализа.

i) Если $\{a_j\}$ абсолютно суммируема, то $\{a_j\}$ суммируема (то есть $-\infty < \sum_{j=0}^{\infty} a_j < \infty$) и

$$\Big|\sum_{j=0}^{\infty} a_j\Big| \le \sum_{j=0}^{\infty} |a_j|.$$

ii) Рассмотрим последовательность с двумя индексами, $\{a_{jk}\}$ $(j, k = 0, 1, 2, \dots)$. Предположим, что $\sum_{j=0}^{\infty} |a_{jk}| < \infty$ для каждого k, и пусть $s_k \equiv \sum_{j=0}^{\infty} |a_{jk}|$. Предположим, что $\{s_k\}$ абсолютно суммируема. Тогда

$$\left| \sum_{j=0}^{\infty} \left(\sum_{k=0}^{\infty} a_{jk} \right) \right| < \infty \quad \text{и} \quad \sum_{j=0}^{\infty} \left(\sum_{k=0}^{\infty} a_{jk} \right) = \sum_{k=0}^{\infty} \left(\sum_{j=0}^{\infty} a_{jk} \right) < \infty.$$

Указание: *Получите*

$$\sum_{j=0}^{\infty} |\gamma_j| \le \sigma^2 \sum_{j=0}^{\infty} \left(\sum_{k=0}^{\infty} |\psi_{j+k}| \cdot |\psi_k| \right) < \infty.$$

3. (Автоковариации $h(L)x_t$.) Мы хотим получить автоковариации для $\{y_t\}$ в утверждении 6.2. Как и в утверждении 6.2, будем предполагать здесь, что $\{x_t\}$ является ковариационно стационарным. Сначала рассмотрим взвешенное среднее конечного количества слагаемых:

$$y_{t,n} = h_0 x_t + h_1 x_{t-1} + \dots + h_n x_{t-n}.$$

(a) Покажите, что для этого $\{y_{t,n}\}$

$$\gamma_j = \sum_{k=0}^{n} \sum_{l=0}^{n} h_k h_l \gamma_{j-k+l}^x,$$

где γ_j^x является автоковариацией $\{x_t\}$ j-го порядка.
Проверьте, что она редуцируется к (6.1.2), если $\{x_t\}$ является белым шумом. **Указание:** *Если вы найдете это сложным, положите сначала $n = 1$, а затем увеличивайте n, чтобы установить структуру.*

(b) Рассмотрим теперь взвешенное среднее возможно бесконечного числа слагаемых:

$$y_t = h_0 x_t + h_1 x_{t-1} + \dots.$$

Принимая утверждение 6.2(a) как доказанное, покажите, что

$$\gamma_j = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} h_k h_l \gamma_{j-k+l}^x, \quad \text{если } \{h_j\} \text{ абсолютно суммируема.}$$

Указание: *$y_{t,n} \to_{m.s.} y_t$ при $n \to \infty$ по утверждению 6.2. Продолжите, как в 2(c).*

4. (Однородные линейные разностные уравнения.) Рассмотрим однородное линейное разностное уравнение p-го порядка:

$$y_j - \phi_1 y_{j-1} - \phi_2 y_{j-2} - \cdots - \phi_p y_{j-p} = 0 \quad (j = p, p+1, \ldots), \tag{1}$$

где $(\phi_1, \phi_2, \ldots, \phi_p)$ являются вещественными константами с $\phi_2 \neq 0$ с начальным условием в виде заданных значений $(y_0, y_1, \ldots, y_{p-1})$. Соответствующее полиномиальное уравнение имеет вид:

$$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^{p-1} - \phi_p z^p = 0. \tag{2}$$

Пусть λ_k (k=1,2,. . . ,K) являются различными корнями этого уравнения (они могут быть комплексными) и пусть r_k является кратностью λ_k. Таким образом, $r_1 + r_2 + \cdots + r_K = p$. Указанное полиномиальное уравнение можно записать как

$$\prod_{k=1}^{K} (1 - \lambda_k^{-1} z)^{r_k} = 0. \tag{3}$$

В качестве примера рассмотрим полиномиальное уравнение второго порядка

$$1 - z + \frac{1}{4} z^2 = 0. \tag{4}$$

Оно имеет корень 2 кратности 2, поскольку

$$1 - z + \frac{1}{4} z^2 = (1 - \frac{1}{2} z)^2.$$

Поэтому в этом примере $K = 1$, $r_1 = 2$ и $\lambda_1 = 2$.

Как вы проверите ниже, решение однородного уравнения p-го порядка можно записать в виде:

$$y_j = \sum_{k=1}^{K} \sum_{n=0}^{r_k - 1} c_{kn} \cdot (j)^n \lambda_k^{-j}. \tag{5}$$

Коэффициенты c_{kn} (их ровно p, поскольку $r_1 + r_2 + \cdots + r_K = p$) можно определить из начальных условий, которые специфицируют значения для $(y_0, y_1, \ldots, y_{p-1})$. В примере выше, (уравнение (4)),

$$y_j = c_{10} \cdot 2^{-j} + c_{11} \cdot j \cdot 2^{-j}. \tag{6}$$

В важном частном случае, когда все корни различны, $K = p$ и $r_k = 1$ для всех k, и (5) принимает вид:

$$y_j = \sum_{k=1}^{p} c_{k0} \lambda_k^{-j}. \tag{7}$$

(a) Для частного случая $p = 2$ и различных корней, (7) имеет вид:

$$y_j = c_{10}\lambda_1^{-j} + c_{20}\lambda_2^{-j}. \tag{8}$$

Проверьте, что это является решением однородного разностного уравнения (1) с $p = 2$ для любого заданного (c_{10}, c_{20}). Чтобы определить (c_{10}, c_{20}), запишите (8) для $j = 0$ и $j = 1$ и решите его относительно значений c как функцию от $(y_0, y_1, \lambda_1, \lambda_2)$.

(b) Чтобы научиться (вспомнить), как работать с несколькими корнями, рассмотрим пример (4) выше. Проверьте, что (6) является решением однородного разностного уравнения второго порядка

$$y_j - y_{j-1} + \frac{1}{4}y_{j-2} = 0.$$

(c) (Дополнительно.) Докажите следующую лемму:

Лемма. *Пусть $0 \le \xi < 1$ и пусть n является неотрицательным целым числом. Тогда существуют два действительных числа A и b, такие что $\xi < b < 1$ и $(j)^n\xi^j < Ab^j$ для $j = 0, 1, 2, \ldots$.*

Указание: *Выберите любое b, такое, что $\xi < b < 1$. Поскольку при возрастании j, b^j в конце концов принимает значения, большие, чем $(j)^n\xi^j$, существует некоторое j, назовем его J, такое, что $(j)^n\xi^j < b^j$ для всех $j \ge J$.*

(d) Предположим, что выполняется условие стабильности для линейного однородного разностного уравнения p-го порядка. Докажите, что существуют два числа, A и b, такие что $0 < b < 1$ и

$$|y_j| < Ab^j \quad \text{для } j = 0, 1, 2, \ldots .$$

Указание:

$$\begin{aligned} |y_j| &= \left|\sum_{k=1}^{K}\sum_{n=0}^{r_k-1} c_{kn}\cdot(j)^n\lambda_k^{-j}\right| \le \\ &\le \sum_{k=1}^{K}\sum_{n=0}^{r_k-1} |c_{kn}|\cdot(j)^n\cdot|\lambda_k^{-1}|^j \le \\ &\le c\sum_{k=1}^{K}\sum_{n=0}^{r_k-1}(j)^n|\lambda_k^{-1}|^j, \end{aligned}$$

где $c = \max\{|c_{kn}|\}$. Поскольку $|\lambda_k^{-1}| < 1$ по условию стабильности, мы можем применить лемму с $\xi = |\lambda_k^{-1}|$ и утверждать, что

$$(j)^n|\lambda_k^{-1}|^j < A_k(b_k)^j \quad \text{для всех } j$$

для некоторого $A_k > 0$ *и* $|\lambda_k^{-1}| < b_k < 1$. *Положим* $A^* = \max\{A_k\}$ *и* $b = \max\{b_k\}$, *так что* $A_k(b_k)^j \le A^* b^j$ *для всех* k. *Поэтому*

$$|y_j| < c \sum_{k=1}^{K} \sum_{n=0}^{r_k-1} A^* b^j = cpA^* b^j.$$

(Напомним, что $p = r_1 + \cdots + r_K$.*) Тогда определим* A *как* cpA^*. *Вы можете найти такой способ доказательства, использующий неравенства, сложным. Если это так, положите* p, K *и* $(r_1, \ldots, r_K)$ *некоторым конкретным числам (например,* $p = 3$, $K = 2$, $r_1 = 1$, $r_2 = 2$*).*

5. (Уравнения Юла — Уокера.) В основном тексте автоковариацииAR(1)-процесса вычислялись на основании представления MA(∞). Здесь мы используем уравнения Юла — Уокера.

   (a) Из (6.2.1′) на с. 405 получите, что

   $$\gamma_j - \phi\gamma_{j-1} = \mathrm{E}[\varepsilon_t \cdot (y_{t-j} - \mu)] \quad (j = 0, 1, 2, \ldots).$$

   (b) (Легко.) Используя MA-представление, покажите, что

   $$\mathrm{E}[\varepsilon_t \cdot (y_{t-j} - \mu)] = \begin{cases} \sigma^2 & (j = 0) \\ 0 & (j = 1, 2, \ldots). \end{cases}$$

   (c) Выведите уравнения Юла — Уокера:

   $$\gamma_0 - \phi\gamma_1 = \sigma^2, \tag{9}$$

   $$\gamma_j - \phi\gamma_{j-1} = 0 \quad (j = 1, 2, \ldots). \tag{10}$$

   (d) Вычислите $(\gamma_0, \gamma_1, \ldots)$ из (9) и (10). **Указание:** *Сначала положите* $j = 1$ *в (10) для получения* $\gamma_1 = \phi\gamma_0$. *Это и (9) можно решить относительно* (γ_0, γ_1).

6. (AR(1) без использования фильтров.) Мы хотим доказать без помощи утверждения 6.2, что $\{y_t\}$, следующий AR(1)-уравнению (6.2.1) с условием стационарности $|\phi| < 1$, должен иметь MA(∞)-представление, если он является ковариационно стационарным.

   (a) Пусть $x_t \equiv y_t - \mu$, так что $\mathrm{E}(x_t) = 0$. Последовательной подстановкой из (6.2.1′) (на с. 405) получите

   $$x_t = x_{t,n} + \phi^n x_{t-n},$$

   где

   $$x_{t,n} = \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \cdots + \phi^{n-1}\varepsilon_{t-n+1}.$$

(b) Покажите, что $x_{t,n} \to_{m.s.} x_t$.
Указание: *Надо показать, что* $\mathrm{E}[(x_{t,n} - x_t)^2] \to 0$ *при* $n \to \infty$. *Поскольку* $(x_{t,n} - x_t)^2 = \phi^{2n} x_{t-n}^2$ *и* $\{y_t\}$ *предполагается ковариационно стационарным,* $\mathrm{E}(x_{t-n}^2) < \infty$.

(c) (Тривиально.) Вспомнив, что бесконечная сумма

$$\sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$$

определяется как предел в среднем квадратическом частичной суммы $x_{t,n}$, покажите, что y_t можно записать как

$$y_t = \mu + \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}.$$

7. (Сопровождающая форма.) Мы хотим обобщить доказательство, только что построенное в вопросе 6, на $\mathrm{AR}(p)$. Без потери общности, мы можем предположить, что $\mu = 0$ (или переопределить y_t как отклонение от среднего для y_t). Определим

$$\underset{(p\times 1)}{\boldsymbol{\xi}_t} = \begin{bmatrix} y_t \\ y_{t-1} \\ y_{T-2} \\ \vdots \\ y_{t-p+1} \end{bmatrix}, \quad \underset{(p\times 1)}{\boldsymbol{v}_t} = \begin{bmatrix} \varepsilon_t \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

$$\underset{(p\times p)}{\boldsymbol{F}} = \begin{bmatrix} \phi_1 & \phi_2 & \phi_3 & \dots & \phi_{p-1} & \phi_p \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix}.$$

Рассмотрим следующее векторное стохастическое разностное уравнение первого порядка:

$$\boldsymbol{\xi}_t = \boldsymbol{F}\boldsymbol{\xi}_{t-1} + \boldsymbol{v}_t. \tag{11}$$

В этой системе p уравнений, называемой **сопровождающей формой** (*companion form*), первое уравнение то же самое, как и $\mathrm{AR}(p)$-уравнение (6.2.6′) на с. 409, а остальные просто являются тождествами для y_{t-j} $(j = 1, 2, \dots, p-1)$.

(a) Последовательной подстановкой покажите, что

$$\begin{aligned} \boldsymbol{\xi}_t &= \boldsymbol{\xi}_{t,n} + (\boldsymbol{F})^n \boldsymbol{\xi}_{t-n}, \\ \boldsymbol{\xi}_{t,n} &\equiv \boldsymbol{v}_t + \boldsymbol{F}\boldsymbol{v}_{t-1} + \boldsymbol{F}^2\boldsymbol{v}_{t-2} + \dots + \boldsymbol{F}^{n-1}\boldsymbol{v}_{t-n+1}. \end{aligned} \tag{12}$$

Собственные значения (*eigenvalues*), или **характеристические корни** (*characteristic roots*) $p \times p$-матрицы $\boldsymbol{F}$ являются корнями детерминантного уравнения

$$|\boldsymbol{F} - \lambda \boldsymbol{I}_p| = 0. \tag{13}$$

Можно показать, что левая часть этого уравнения равна $\lambda^p - \phi_1\lambda^{p-1} - \cdots - \phi_{p-1}\lambda - \phi_p$. Поэтому это детерминантное уравнение принимает вид полиномиального уравнения p-го порядка:

$$\lambda^p - \phi_1\lambda^{p-1} - \cdots - \phi_{p-1}\lambda - \phi_p = 0. \tag{14}$$

(b) Проверьте это для $p = 2$.

В оставшейся части этого вопроса предположим, что все корни полиномиального уравнения p-го порядка (которые, как было только что видно, являются собственными значениями $\boldsymbol{F}$) различны. Пусть $(\lambda_1, \lambda_2, \ldots, \lambda_p)$ — эти различные собственные значения. Из матричной алгебры известно, что существует невырожденная $p \times p$-матрица $\boldsymbol{T}$, такая что

$$\boldsymbol{F} = \boldsymbol{T}\boldsymbol{\Lambda}\boldsymbol{T}^{-1}, \quad \underset{(p\times p)}{\boldsymbol{\Lambda}} = \begin{bmatrix} \lambda_1 & 0 & 0 & \ldots & 0 \\ 0 & \lambda_2 & 0 & \ldots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \ldots & 0 & \lambda_{p-1} & 0 \\ 0 & \ldots & 0 & 0 & \lambda_p \end{bmatrix}. \tag{15}$$

(Этот факт не зависит от особой формы матрицы $\boldsymbol{F}$; все, что нужно для (15), — это то, что значения λ являются различными собственными значениями матрицы $\boldsymbol{F}$.) Из (15) следует, что

$$(\boldsymbol{F})^n = \boldsymbol{T}(\boldsymbol{\Lambda})^n\boldsymbol{T}^{-1}.$$

(c) (Очень легко.) Докажите это для $n = 3$.

(d) Покажите, что $\boldsymbol{\xi}_{t,n} \rightarrow_{m.s.} \boldsymbol{\xi}$ при $n \rightarrow \infty$, если выполняется условие стационарности относительно $\mathrm{AR}(p)$-уравнения и если $\{y_t\}$ является ковариационно стационарным. **Указание:** *Условие стационарности требует, чтобы все корни в (13) были меньше 1 по абсолютной величине. В силу (12), то, что нужно показать, — это то, что* $(\boldsymbol{F})^n\boldsymbol{\xi}_{t-n} \rightarrow_{m.s.} \mathbf{0}$. *Вспомним, что* $\boldsymbol{z}_n \rightarrow_{m.s.} \boldsymbol{\alpha}$, *если сходимость в среднем квадратическом происходит покомпонентно.*

(e) Покажите, что y_t имеет $\mathrm{MA}(\infty)$-представление

$$y_t = \psi_0\varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + \ldots.$$

Как ψ_j связана с характеристическими корнями? **Указание:** *Из (d) немедленно следует, что*

$$\boldsymbol{\xi}_t = \boldsymbol{v}_t + \boldsymbol{F}\boldsymbol{v}_{t-1} + (\boldsymbol{F})^2\boldsymbol{v}_{t-2} + \dots .$$

Верхнее уравнение этого выражения является MA(∞)*-представлением для* y_t.

(Если не все собственные значения различны, то вместо (15) нам нужно использовать так называемое Жорданово разложение для $\boldsymbol{F}$.)

8. (AR(1), который начинается в момент $t = 0$.) В основном тексте MA(∞)-представление (6.2.2) AR(1)-процесса, следующего $y_t = c + \phi y_{t-1} + \varepsilon_t$, подразумевает, что процесс определен для $t = -1, -2, \dots$, так же как и для $t = 0, 1, 2, \dots$. Вместо этого рассмотрим AR(1)-процесс, который начинается в $t = 0$ с начальным значением y_0. Предположим, что $|\phi| < 1$ и что y_0 не коррелировано с последующим процессом белого шума $(\varepsilon_1, \varepsilon_2, \dots)$.

   (a) Запишите дисперсию для y_t как функцию от σ^2 ($\equiv \mathrm{Var}(\varepsilon_t)$), $\mathrm{Var}(y_0)$ и ϕ.**Указание:** *Последовательная замена дает:*

   $$y_t = (1+\phi+\phi^2+\dots+\phi^{t-1})c+\varepsilon_t+\phi\varepsilon_{t-1}+\phi^2\varepsilon_{t-2}+\dots+\phi^{t-1}\varepsilon_1+\phi^t y_0.$$

   Пусть μ и $\{\gamma_j\}$ являются средним и автоковариациями ковариационно стационарного AR(1)*-процесса, так что $\mu = c/(1-\phi)$, а $\{\gamma_j\}$ заданы в* (6.2.5). *Покажите, что*

   $$\lim_{t\to\infty} \mathrm{E}(y_t) = \mu, \quad \lim_{t\to\infty} \mathrm{Var}(y_t) = \gamma_0 \quad \text{и} \quad \lim_{t\to\infty} \mathrm{Cov}(y_t, y_{t-j}) = \gamma_j.$$

   (b) Покажите, что $\{y_t\}$ является ковариационно стационарным, если $\mathrm{E}(y_0) = \mu$ и $\mathrm{Var}(y_0) = \gamma_0$.

9. (Доказательство утверждения 6.8(a).) Мы хотим доказать утверждение 6.8(a). По неравенству Чебышева достаточно показать, что

$$\lim_{n\to\infty} \mathrm{Var}(\overline{y}) = 0.$$

Из (6.5.2):

$$\mathrm{Var}(\overline{y}) = \frac{\gamma_0}{n} + \frac{2}{n}\sum_{j=1}^{n-1}\left(1-\frac{j}{n}\right)\gamma_j =$$

$$= \frac{\gamma_0}{n} + \frac{2}{n}\sum_{j=1}^{n}\left(1-\frac{j}{n}\right)\gamma_j$$

$$\left(\text{поскольку } 1-\frac{j}{n} = 0 \text{ для } j = n\right)$$

$$\le \frac{\gamma_0}{n} + \frac{2}{n}\sum_{j=1}^{n}\left(1-\frac{j}{n}\right)|\gamma_j| \le$$

$$\leq \frac{\gamma_0}{n} + \frac{2}{n}\sum_{j=1}^{n}|\gamma_j|.$$

(a) (Дополнительно.) Докажите следующий результат:

$$\lim_{j\to\infty} a_j \;\Rightarrow\; \lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n}|a_j| = 0.$$

Указание: *$\{a_j\}$ сходится к 0. Поэтому (i) $|a_j| < M$ для всех j и (ii) для любого заданного $\varepsilon > 0$ существует положительное целое N, такое, что $|a_j| < \varepsilon/2$ для всех $j > N$. Поэтому*

$$\begin{aligned}\frac{1}{n}\sum_{j=1}^{n}|a_j| &= \frac{1}{n}\sum_{j=1}^{N}|a_j| + \frac{1}{n}\sum_{j=N+1}^{n}|a_j| < \frac{1}{n}\sum_{j=1}^{N}M + \frac{1}{n}\sum_{j=N+1}^{n}\frac{\varepsilon}{2} = \\ &= \frac{NM}{n} + \frac{(n-N)}{n}\frac{\varepsilon}{2} < \frac{NM}{n} + \frac{\varepsilon}{2}.\end{aligned}$$

Выбирая n достаточно большим, NM/n можно сделать меньше, чем $\varepsilon/2$.

(b) Принимая результат в (a) как доказанный, докажите, что $\mathrm{Var}(\overline{y}) \to 0$ при $n \to \infty$.

10. (Доказательство утверждения 6.8(b).)

(a) (Дополнительно.) Докажите следующий результат:

$$\sum_{j=1}^{\infty} a_j < \infty \Rightarrow \lim_{n\to\infty}\frac{1}{n}\sum_{j=1}^{n} j a_j = 0.$$

Указание:

$$\begin{aligned}\sum_{j=1}^{n} j a_j &= a_1 + 2a_2 + 3a_3 + \cdots + na_n = (a_1 + a_2 + \cdots + a_n) + \\ &\quad +(a_2 + a_3 + \cdots + a_n) + \cdots + (a_{n-1} + a_n) + a_n = \\ &= \sum_{j=1}^{n}\sum_{k=j}^{n} a_k \leq \sum_{j=1}^{n}\left|\sum_{k=j}^{n} a_k\right| = \sum_{j=1}^{N}\left|\sum_{k=j}^{n} a_k\right| + \sum_{j=N+1}^{n}\left|\sum_{k=j}^{n} a_k\right|.\end{aligned}$$

Поскольку $\{a_j\}$ суммируема, (i) $|\sum_{k=j}^{n} a_k| < M$ для всех j, n и (ii) существует положительное целое N, такое что $|\sum_{k=j}^{n} a_k| < \varepsilon/2$ для любого заданного $\varepsilon > 0$ для всех $j, n > N$ (это следует из того, что $\{\sum_{j=1}^{n} a_j\}$ является последовательностью Коши).

(b) Принимая этот результат как доказанный, докажите утверждение 6.8(b).

Эмпирические упражнения

1. Прежде чем отвечать, прочитайте по крайней мере p. 115–119 (за исключением последних двух параграфов на p. 119) в [Bekaert and Hodrick, 1993]. Файлы данных DM.ASC (для немецкой марки), POUND.ASC (британский фунт) и YEN.ASC (японская иена) содержат еженедельные данные по следующим пунктам:

   Столбец 1: дата наблюдения (например, «19850104» обозначает 4 января 1985 г.)

   Столбец 2: цена продажи доллара в единицах иностранной валюты на спотовом рынке в пятницу текущей недели (S_t)

   Столбец 3: цена продажи доллара в единицах иностранной валюты на 30-дневном форвардном рынке в пятницу текущей недели (F_t)

   Столбец 4: цена предложения доллара в единицах иностранной валюты на спотовом рынке на дату поставки на текущий форвардный контракт ($S30_t$).

   Период выборки — с первой недели 1975 г. по последнюю неделю 1989 г. Объем выборки равен 778. Как и в основном тексте, определим $s_t \equiv \log(S_t)$, $f_t \equiv \log(F_t)$, $s30_t \equiv \log(S30_t)$. Если I_t является информацией, доступной в пятницу недели t, то она включает в себя $\{s_t, s_{t-1}, \dots, f_t, f_{t-1}, \dots, s30_{t-5}, s30_{t-6}, \dots\}$. Заметим, что $s30_t$ не наблюдается до момента после пятницы недели $t+4$. Определим $\varepsilon_t \equiv s30_t - f_t$.

   Выберите предпочитаемую вами валюту для ответа на следующие вопросы:

   (a) (Работа в библиотеке/интернете.) Для вашего выбора иностранной валюты идентифицируйте неделю, когда абсолютное значение форвардной премии наибольшее. Для этой недели найдите некоторую меру внутренней одномесячной процентной ставки (например, одномесячной ставки депозитного сертификата) для США и отечественной валюты для проверки того, что дифференциал процентной ставки соответствует форвардной премии.

   (b) (Коррелограмма $\{\varepsilon_t\}$.) Изобразите выборочную коррелограмму для ε_t с 40 лагами. Исчезает ли автокорреляция после 4 лагов? (Вам самим решать, вычитать ли выборочное среднее при вычислении выборочных корреляций. Теория говорит о том, что популяционное среднее равно нулю, что вы, возможно, захотите использовать в расчетах. При вычислении коррелограммы для обменного курса иена/доллар, показанной на рис. 6.2, выборочное среднее было вычтено.)

(c) (Является ли логарифм спотовой ставки случайным блужданием?) Изобразите выборочную коррелограмму $s_{t+1} - s_t$ с 40 лагами. Для этих 40 автокорреляций используйте статистику Льюнга — Бокса для тестирования на сериальную корреляцию. Можете ли вы отвергнуть гипотезу о том, что $\{s_t\}$ является случайным блужданием со сносом?

(d) (Безусловный тест.) Проведите безусловный тест. Можете ли вы повторить результаты в табл. 6.1 для выбранной валюты?

(e) (Дополнительно, регрессионный тест с усеченным ядром.) Проведите регрессионный тест. Можете ли вы повторить результаты в табл. 6.2 для выбранной валюты?

Советы для пакета RATS: Используйте LINREG с опцией ROBUSTERRORS. Для включения автокорреляций до четвертой включительно положите LAGS = 4 в LINREG. Опция DAMP = 0 указывает LINREG использовать усеченное ядро.

Советы для пакета TSP: В процедуре GMM в TSP есть опция KERNEL, но выбор ограничен ядром Бартлетта и ядром Парзена. Вы можете использовать матричные команды в TSP чтобы выполнить оценивание с усеченным ядром.

(f) (Ядро Бартлетта.) Для проведения регрессионного теста используйте оценку для S на основе ядра Бартлетта. [Newey and West, 1994] предлагают процедуру автоматического выбора ширины окна на основе данных. Примите на веру, что глубина запаздываний автоковариаций, определенная этой процедурой, равна 13 для иены/доллар (поэтому при вычислении $\widehat{S}$ используются автоковариации до двенадцатого порядка), 9 для DM/доллар и 17 для фунта/доллар.

Советы для пакета RATS: Опция ROBUSTERRORS, DAMP = 1 указывает LINREG использовать ядро Бартлетта. Для иены/доллар, например, положите LAGS = 12.

Советы для пакета TSP: Положите `het, kernel = bartlett` в процедуре GMM. Положите NMA = 12 для иены/доллар.

Стандартная ошибка для коэффициента при $f - s$ для иены/доллар должна быть равной 0,6815.

(g) (Дополнительно, VARHAC.) В регрессионном тесте используйте оценку VARHAC для $\widehat{S}$. Положите $p_{\max}$ равным целой части $n^{1/3}$. На шаге 1 VARHAC вы можете оценить двумерную VAR и для каждого из двух уравнений VAR можете выбрать глубину запаздываний p, используя BIC на фиксированной выборке $t = p_{\max} + 1, \ldots, T$. Информационный критерий может быть взят в виде (6.6.11) или, альтернативно,

$$\log\left(SSR_p^(k)/(n - p_{\max})\right) + pK \cdot \log(n - p_{\max})/(n - p_{\max}). \quad (15)$$

Только для того, чтобы иметь четкое представление о том, как вы выбрали p в вашем ответе, используйте последнюю целевую функцию для выбора p посредством BIC. Для случая иена/доллар выбранная глубина запаздываний должна быть равна 4 для первого уравнения VAR и 6 для второго уравнения. Для иена/доллар стандартная ошибка для коэффициента при $f - s$ должна быть равной 0,8027.

2. (Продолжение упражнения по работе Фамы в главе 2.) Теперь мы знаем, как использовать трехмесячную ставку по казначейским векселям для тестирования эффективности рынка казначейских векселей США, используя месячные данные. Пусть

π_t = темп инфляции (в процентах, годовой темп), по трехмесячному периоду, заканчивающемуся в месяц $t-1$,

R_t = трехмесячная ставка по казначейским векселям (в процентах, годовая ставка) на начало месяца t.

Рассмотрим регрессию Фамы для проверки эффективности рынка:

$$\pi_{t+3} = \beta_0 + \beta_1 R_t + \varepsilon_t.$$

(a) (Тривиально.) Почему зависимой переменной является π_{t+3}, а не, скажем, π_t?

(b) Покажите, что при гипотезе эффективности рынка $\beta_1 = 1$ и автоковариации $\{\varepsilon_t\}$ исчезают после второго лага (то есть $\gamma_j = 0$ для $j > 2$).

(c) Пусть $r_{t+3} \equiv R_t - \pi_{t+3}$ является ex post процентной ставкой по трехмесячным казначейским векселям. Используйте $PAI3$ в MISHKIN.ASC как меру трехмесячного темпа инфляции для вычисления ex post реальной процентной ставки. (Датировка переменных в MISHKIN.ASC такова, что наблюдение процентной ставки за январь использует данные по казначейской ставке, истекающей в декабре, и январское наблюдение для трехмесячного темпа инфляции состоит из данных CPI с декабря по март. Поэтому в регрессии можно сопоставлять $PAI3$ и $TB3$ (трехмесячную ставку по казначейским векселям) для одного и того же месяца.) Изобразите выборочную коррелограмму для периода выборки 1/53–7/71. Что можно предположить о гипотезе эффективности рынка по этой коррелограмме?

(d) Проверьте эффективность рынка, оценивая регрессию Фамы для периода выборки 1/53–7/71. (Объем выборки должен быть равен 223; не выбрасывайте две трети наблюдений, как это сделал Фама в своих табл. 6–8.) Используйте (6.6.4) с $q = 2$ для $\widehat{S}$.

Ответы на избранные вопросы

Аналитические упражнения

2d. Поскольку $\{\psi_j\}$ абсолютно суммируема, $\psi_j \to 0$ при $j \to \infty$. Поэтому для любого j существует $A > 0$, такое что $|\psi_{j+k}| \le A$ для всех j, k. Следовательно, $|\psi_{j+k} \cdot \psi_k| \le A|\psi_k|$. Поскольку $\{\psi_k\}$ (и, следовательно, $\{A\psi_k\}$) абсолютно суммируема, это верно и для $\{\psi_{j+k} \cdot \psi_k\}$ $(k = 0, 1, 2, \dots)$ для любого заданного j. Таким образом, в силу (i),

$$|\gamma_j| = \sigma^2 \left| \sum_{k=0}^{\infty} \psi_{j+k}\psi_k \right| \le \sigma^2 \sum_{k=0}^{\infty} |\psi_{j+k}\psi_k| = \sigma^2 \sum_{k=0}^{\infty} |\psi_{j+k}||\psi_k| < \infty.$$

Теперь положим a_{jk} в (ii) равным $|\psi_{j+k}| \cdot |\psi_k|$. Тогда

$$\sum_{j=0}^{\infty} |a_{jk}| = \sum_{j=0}^{\infty} |\psi_k||\psi_{j+k}| \le |\psi_k| \sum_{j=0}^{\infty} |\psi_j| \le \infty.$$

Пусть

$$M \equiv \sum_{j=0}^{\infty} |\psi_j| \quad \text{и} \quad s_k \equiv |\psi_k| \sum_{j=0}^{\infty} |\psi_{j+k}|.$$

Тогда $\{s_k\}$ суммируема, поскольку $|s_k| \le |\psi_k| \cdot M$ и $\{\psi_k\}$ абсолютно суммируема. Следовательно, из (ii):

$$\sum_{j=0}^{\infty} \left(\sum_{k=0}^{\infty} |\psi_{j+k}| \cdot |\psi_k| \right) < \infty.$$

Эта двойная сумма, умноженная на σ^2, больше или равна $\sum_{j=0}^{\infty} |\gamma_j|$. Поэтому $\{\gamma_j\}$ абсолютно суммируема.

7е. ψ_j является элементом (1,1) в $\boldsymbol{T}(\boldsymbol{\Lambda})^j\boldsymbol{T}^{-1}$.

8а.

$$\begin{aligned}
\mathrm{E}(y_t) &= \frac{1-\phi^t}{1-\phi}c + \phi^t\,\mathrm{E}(y_0) = \\
&= (1-\phi^t)\mu + \phi^t\,\mathrm{E}(y_0) \\
&\text{(поскольку } \mu = c/(1-\phi)) \\
&= \mu + \phi^t \cdot [\mathrm{E}(y_0) - \mu], \\
\mathrm{Var}(y_t) &= (1+\phi^2+\phi^4+\cdots+\phi^{2t-2})\sigma^2 + \phi^{2t}\,\mathrm{Var}(y_0) \\
&\quad\text{(поскольку } \mathrm{Cov}(\varepsilon_t, y_0) = 0) \\
&= \frac{1-\phi^{2t}}{1-\phi^2}\sigma^2 + \phi^{2t}\,\mathrm{Var}(y_0) = \\
&= (1-\phi^{2t})\gamma_0 + \phi^{2t}\,\mathrm{Var}(y_0) \\
&\text{(поскольку } \gamma_0 \equiv \sigma^2/(1-\phi^2)), \\
&= \gamma_0 + \phi^{2t}\cdot[\mathrm{Var}(y_0) - \gamma_0], \\
\mathrm{Cov}(y_t, y_{t-j}) &= (\phi^j + \phi^{j+2} + \cdots + \phi^{2t-j-2})\sigma^2 + \phi^{2t-j}\,\mathrm{Var}(y_0) = \\
&= \phi^j\cdot(1+\phi^2+\cdots+\phi^{2t-2j-2})\sigma^2 + \phi^{2t-j}\,\mathrm{Var}(y_0) = \\
&= \phi^j\cdot\frac{1-\phi^{2(t-j)}}{1-\phi^2}\sigma^2 + \phi^{2t-j}\,\mathrm{Var}(y_0) = \\
&= \phi^j\cdot(1-\phi^{2(t-j)})\gamma_0 + \phi^{2t-j}\,\mathrm{Var}(y_0) = \\
&= \gamma_0\phi^j + \phi^{2t-j}\cdot[\mathrm{Var}(y_0) - \gamma_0] = \\
&= \gamma_j + \phi^{2t-j}\cdot[\mathrm{Var}(y_0) - \gamma_0].
\end{aligned}$$

Эмпирические упражнения

1с. Статистика Льюнга — Бокса с 40 лагами равна 84,0 с p-значением 0,0058%. Поэтому гипотеза о том, что спотовая ставка является случайным блужданием со сносом, может быть отвергнута. См. рис. 6.6 для коррелограммы иена/доллар.

1f. Для обменного курса иена/доллар стандартные ошибки равны 3,72 для константы и 0,68 для коэффициента при $f - s$. Статистика Вальда для гипотезы о том, что $\beta_0 = 0$ и $\beta_1 = 1$, равна 21,85 ($p = 0.002\%$).

2с. Коррелограмма для трехмесячной ex post реальной ставки показана на рис. 6.7. Теория предсказывает, что сериальная корреляция исчезает после второго лага. Коррелограмма в основном соответствует этому предсказанию.

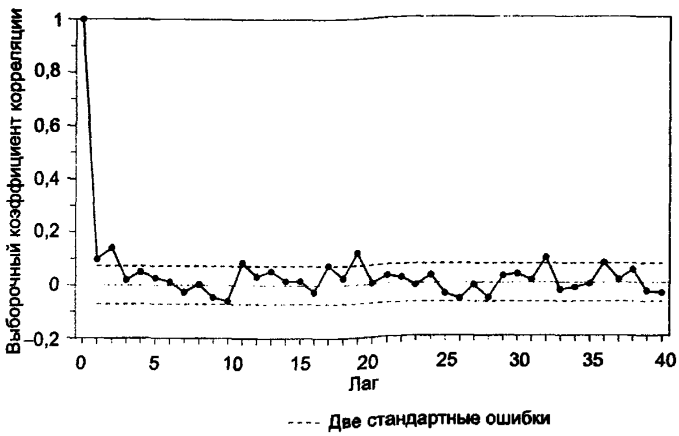

Рис. 6.6. Коррелограмма $s_{t+1} - s_t$, иена/доллар

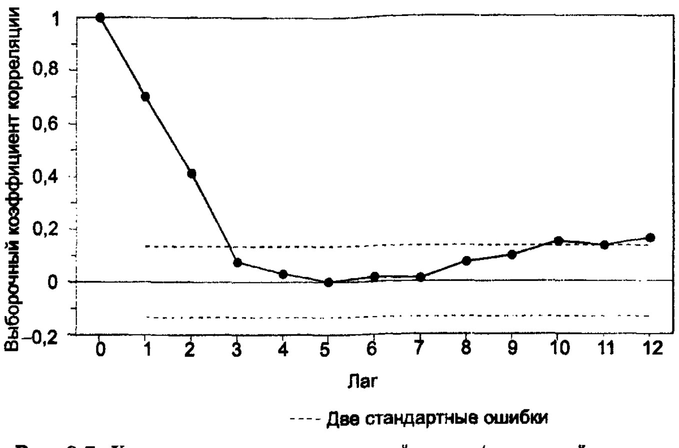

Рис. 6.7. Коррелограмма трехмесячной ex post реальной ставки

Литература

Andrews, D., 1991, "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

Bekaert, G., and R. Hodrick, 1993, "On Biases in the Measurement of Foreign Exchange Risk Premiums," *Journal of International Money and Finance*, 12, 115–138.

Brockwell, P., and R. Davis, 1991, *Time Series: Theory and Methods* (2d ed.), Berlin: Springer — Verlag.

Den Haan, W., and A. Levin, 1996a, "Inferences from Parametric and Non-Parametric Covariance Matrix Estimation Procedures," Technical Working Paper Series 195, NBER.

——, 1996b, "A Practitioner's Guide to Robust Covariance Matrix Estimation," Technical Working Paper Series 197, NBER.

Fuller, W., 1996, *Introduction to Statistical Time Series* (2d ed.), New York: Wiley.

Gordin, M. I., 1969, "The Central Limit Theorem for Stationary Processes," *Soviet Math. Dokl.*, 10, 1174–1176.

Gourieroux, C., and A. Monfort, 1997, *Time Series and Dynamic Models*, Cambridge: Cambridge University Press.

Hamilton, J., 1994, *Time Series Analysis*, Princeton: Princeton University Press.

Hannan, E. J., 1970, *Multiple Time Series*, New York: Wiley.

Hannan, E. J., and M. Deistler, 1988, *The Statistical Theory of Linear Systems*, New York: Wiley.

Hansen, L., 1982, "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

Lütkepohl, H., 1993, *Introduction to Multiple Time Series* Analysis (2d ed.), New York: Springer — Verlag.

Newey, W., and K. West, 1987, "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

——, 1994, "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61, 631–653.

Ng, S., and P. Perron, 2000, "A Note on Model Selection in *Time Series Analysis*," mimeo., Boston: Boston College.

White, H., 1984, *Asymptotic Theory for Econometricians*, New York: Academic.

Глава 7. Экстремальная оценка

В предыдущих главах для оценивания различных рассматриваемых нами моделей мы выбирали обобщенный метод моментов (GMM). Мы также вкратце описали метод максимального правдоподобия (ML). В этой главе мы расширим наш арсенал методов оценивания, исследуя класс оценок, называемых «экстремальными оценками», который включает в себя метод наименьших квадратов (линейный и нелинейный), обобщенный метод моментов (линейный и нелинейный) и метод максимального правдоподобия в качестве частных случаев. Как было подчеркнуто [Amemiya, 1985] и другими, этот унифицированный подход полезен тем, что он выявляет общую структуру, лежащую в основе этих внешне различных методов оценивания.

Параграф 7.1 вводит в рассмотрение экстремальные оценки. Два следующих параграфа посвящены асимптотическим свойствам экстремальных оценок. Параграф 7.4 определяет триаду тестовых статистик — статистики Вальда, множителей Лагранжа и отношения правдоподобий — для экстремальных оценок в общем случае и показывает, как эти статистики конкретизируются в частных случаях. Параграф 7.5 посвящен краткому обсуждению вычислительных аспектов экстремальных оценок.

Эта глава может показаться чрезвычайно сложной, так как в ней сочетается интенсивное использование асимптотической теории с математическим анализом. Основная идея, однако, довольно проста. Если вы можете читать вузовские учебники по микро- и макроэкономике, не испытывая особых трудностей, то будете способны понять большую часть текста при повторном (если не при первом) его прочтении.

Если вас не интересует асимптотика экстремальных оценок *сама по себе*, но вы собираетесь изучить главу 8, вы можете не читать эту главу целиком. Просто прочитайте параграф 7.1, а затем постарайтесь понять, о чем идет речь в теоремах 7.5, 7.6, 7.8, 7.9 и 7.11.

Примечание, касающееся обозначений: В отличие от предыдущих глав мы будем обозначать истинное значение векторного параметра вектором с нижним (подстрочным) индексом 0. Таким образом, $\boldsymbol{\theta}_0$ — это истинное значение параметра, тогда как $\boldsymbol{\theta}$ обозначает гипотетическое значение параметра. Такие обозначения стандартны в «литературе для избранных», посвященной нелинейному оцениванию. Кроме того, мы будем использовать «t» вместо «i» для обозначения номера наблюдения.

7.1. Экстремальные оценки

Оценка $\widehat{\boldsymbol{\theta}}$ называется **экстремальной** (*extremum*), если существует такая целевая функция $Q_n(\boldsymbol{\theta})$, что

$$\widehat{\boldsymbol{\theta}} \text{ максимизирует } Q_n(\boldsymbol{\theta}) \text{ при условии, что } \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p, \tag{7.1.1}$$

где Θ, называемое **пространством параметров** (*parameter space*), — множество всех возможных значений параметра. В этой книге мы ограничиваем наше внимание случаем, когда Θ является подмножеством конечномерного Евклидова пространства $\mathbb{R}^p$. Целевая функция $Q_n(\boldsymbol{\theta})$ зависит не только от $\boldsymbol{\theta}$, но также от выборки или от наблюдений $(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n)$, где $\boldsymbol{w}_t$ — t-е наблюдение, а n — длина выборки. В наших обозначениях на зависимость целевой функции от выборки размера n указывает нижний индекс n. Метод максимального правдоподобия (ML) и линейный и нелинейный обобщенный метод моментов (GMM) являются частными случаями экстремальных оценок. В этом параграфе представлены эти и другие экстремальные оценки и показывается, как они связаны друг с другом.

«Измеримость» $\widehat{\theta}$

Строго говоря, задача максимизации (7.1.1) необязательно имеет решение. Но вспомним следующий факт из математического анализа:

Пусть отображение $h\colon \mathbb{R}^p \to \mathbb{R}$ непрерывно и $A \subset \mathbb{R}^p$ — компакт (замкнутое и ограниченное множество). Тогда h имеет точку максимума на множестве A. То есть существует такой $x^* \in A$, что $h(\boldsymbol{x}^*) \geqslant h(\boldsymbol{x})$ для любого $\boldsymbol{x}$ в A.

Таким образом, если $Q_n(\boldsymbol{\theta})$ непрерывна по $\boldsymbol{\theta}$ для любых данных $(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n)$ и Θ — компакт, то тогда существует $\boldsymbol{\theta}$ — решение задачи (7.1.1) для любой заданной выборки $(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n)$. В случае когда решение неединственно, мы выберем одно из них. Таким образом, $\widehat{\boldsymbol{\theta}}$, однозначно определенная для любого набора данных $(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n)$, есть функция от данных. Однако, строго говоря, того, что $\widehat{\boldsymbol{\theta}}$ является функцией от вектора случайных величин $(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n)$, недостаточно для того, чтобы $\widehat{\boldsymbol{\theta}}$ была определена как случайная величина; необходимо, чтобы $\widehat{\boldsymbol{\theta}}$ была «измеримой» функцией от $(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n)$. (Если функция непрерывна, то она измерима.) Следующая лемма показывает, что $\widehat{\boldsymbol{\theta}}$ измерима, если таковой является функция $Q_n(\boldsymbol{\theta})$.

Лемма 7.1 (существование экстремальных оценок): *Предположим, что (i) пространство параметров* Θ *— компактное подмножество* $\mathbb{R}^p$, *(ii)* $Q_n(\boldsymbol{\theta})$ *непрерывна по* $\boldsymbol{\theta}$ *для любых данных* $(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n)$ *и (iii)* $Q_n(\boldsymbol{\theta})$ *— измеримая функция от данных для всех* $\boldsymbol{\theta}$ *из* Θ. *Тогда существует измеримая функция* $\widehat{\boldsymbol{\theta}}$, *зависящая от наблюдений, которая является решением* (7.1.1).

По поводу доказательства см., например: [Gourieroux and Monfort, 1995, Property 24.1]. Во всех примерах, приведенных в этой главе, $Q_n(\boldsymbol{\theta})$ измерима.

В большинстве прикладных задач мы не знаем верхнюю или нижнюю границу для истинного вектора параметров. Даже если мы знаем эти граничные значения, они не обязательно включены в пространство параметров, так что оно не будет замкнутым (см. пример CES далее). Таким образом, введение предположения об ограниченности и замкнутости Θ — это то, чего мы хотим избежать. Для некоторых асимптотических результатов в данной главе мы будем заменять предположение компактности некоторыми другими условиями, которые выполняются во многих прикладных задачах.

Два класса экстремальных оценок

В следующих двух параграфах мы получим асимптотические результаты для экстремальных оценок, а потом конкретизируем эти результаты для следующих классов двух экстремальных оценок.

1. **М-оценки.** Экстремальная оценка является **М-оценкой** (*M-estimator*), если целевая функция является выборочным средним:

$$Q_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{t=1}^{n} m(\boldsymbol{w}_t; \boldsymbol{\theta}), \tag{7.1.2}$$

где m — функция $(\boldsymbol{w}_t; \boldsymbol{\theta})$, принимающая действительные значения. Два изучаемых нами примера, которые дают М-оценки, — это **метод максимального правдоподобия (ML)** (*maximum likelihood*) и **нелинейный метод наименьших квадратов (NLS)** (*nonlinear least squares*).

2. **GMM.** Экстремальная оценка является **GMM-оценкой** (*GMM estimator*), если целевая функция может быть записана в виде:

$$Q_n(\boldsymbol{\theta}) = -\frac{1}{2}\boldsymbol{g}_n(\boldsymbol{\theta})'\widehat{\boldsymbol{W}}\boldsymbol{g}_n(\boldsymbol{\theta}) \quad \text{с} \quad \underset{(K\times 1)}{\boldsymbol{g}_n(\boldsymbol{\theta})} \equiv \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{g}(\boldsymbol{w}_t;\boldsymbol{\theta}), \tag{7.1.3}$$

где $\widehat{\boldsymbol{W}}$ — симметричная положительно определенная матрица размера $K \times K$, определяющая расстояние до $\boldsymbol{g}_n(\boldsymbol{\theta})$ от нуля. Она может зависеть от данных. Максимизация целевой функции равносильна минимизации расстояния $\boldsymbol{g}_n(\boldsymbol{\theta})'\widehat{\boldsymbol{W}}\boldsymbol{g}_n(\boldsymbol{\theta})$, что соответствует нашему определению GMM в главе 3. Как вскоре станет понятно, деление расстояния на два упрощает выражение для производной целевой функции. Мы вводили GMM-оценивание для линейных моделей, в которых $\boldsymbol{g}(\boldsymbol{w}_t;\boldsymbol{\theta})$ линейна по $\boldsymbol{\theta}$. Ниже мы покажем, что GMM-оценивание может быть распространено и на нелинейные модели..

Существуют также экстремальные оценки, которые не попадают ни в один из этих двух классов. Самый известный пример таких методов оценивания — **классические оценки минимального расстояния** (*classical minimum distance estimators*). Целевая функция в этом случае может быть представлена как $-\boldsymbol{g}_n(\boldsymbol{\theta})'\widehat{\boldsymbol{W}}\boldsymbol{g}_n(\boldsymbol{\theta})$, однако $\boldsymbol{g}_n(\boldsymbol{\theta})$ необязательно является выборочным средним. Теорема о состоятельности в следующем параграфе достаточно обща, чтобы охватить и такого вида оценки.

В оставшейся части этого параграфа мы представим примеры M-оценок и GMM-оценок. Каждый пример является комбинацией *модели* (множества процессов, порождающих данные, [DGP]) и *метода оценивания*.

Метод максимального правдоподобия (ML)

Основной пример M-оценки — это метод максимального правдоподобия в случае, когда $\{\boldsymbol{w}_t\}$ — i.i.d. (независимые, одинаково распределенные случайные величины). Моделью здесь является множество i.i.d.-последовательностей $\{\boldsymbol{w}_t\}$, где функция плотности вероятности $\boldsymbol{w}_t$ принадлежит семейству плотностей, описываемых конечномерным вектором $\boldsymbol{\theta}$: $f(\boldsymbol{w}_t;\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$. (Поскольку $\{\boldsymbol{w}_t\}$ одинаково распределены, функциональная форма $f(\,\cdot\,;\,\cdot\,)$ не зависит от t.) Функциональная форма $f(\,\cdot\,;\,\cdot\,)$ известна. Модель **параметрическая** (*parametric*), так как вектор параметров $\boldsymbol{\theta}$ конечномерный. При истинном значении параметра $\boldsymbol{\theta}_0$, функция плотности вероятности истинного DGP (того DGP, который на самом деле породил данные) есть $f(\boldsymbol{w}_t;\boldsymbol{\theta}_0)$[1]. Мы говорим, что модель **правильно специфицирована** (*correctly specified*), если $\boldsymbol{\theta}_0 \in \Theta$.

[1]Если бы мы следовали обозначениям, введенным в предыдущих параграфах, то истинное значение параметра записывалось бы как $\boldsymbol{\theta}$, а гипотетическое значение — как $\tilde{\boldsymbol{\theta}}$. В этой главе мы используем $\boldsymbol{\theta}_0$ для обозначения истинного значения и $\boldsymbol{\theta}$ — для обозначения гипотетического значения вектора параметров.

Так как w_t распределены независимо, совместная плотность вероятности данных $(w_1, w_2, \ldots, w_n)$ равна:

$$f(w_1, w_2, \ldots, w_n; \theta_0) = \prod_{t=1}^{n} f(w_t; \theta_0). \tag{7.1.4}$$

Поскольку распределение данных специфицировано полностью, естественным методом оценивания является метод максимального правдоподобия, который проводится следующим образом. После замены вектора истинных параметров θ_0 на вектор гипотетических значений θ эта функция плотности, рассматриваемая как функция от θ, называется **функцией правдоподобия** (*likelihood function*). Оценкой параметра θ_0 по методу максимального правдоподобия (ML) является значение θ, максимизирующее функцию правдоподобия. Так как логарифмирование — монотонное преобразование, максимизация функции правдоподобия равносильна максимизации **логарифма функции правдоподобия** (*log likelihood function*)[1]:

$$\ln f(w_1, w_2, \ldots, w_n; \theta) = \ln \left[\prod_{t=1}^{n} f(w_t; \theta) \right] = \sum_{t=1}^{n} \ln f(w_t; \theta). \tag{7.1.5}$$

Оценка ML для θ_0, таким образом, является M-оценкой с

$$m(w_t; \theta) = \ln f(w_t; \theta), \quad \text{то есть} \quad Q_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \ln f(w_t; \theta). \tag{7.1.6}$$

Прежде чем проиллюстрировать ML примером, полезно сделать три замечания:

- (Автокоррелированные наблюдения.) Среднее значение логарифма функции правдоподобия не будет иметь такую простую форму, если $\{w_t\}$ будут автокоррелированы. Примеры ML для автокоррелированных наблюдений будут изучены в следующей главе.
- (Альтернативные оценки.) Метод максимального правдоподобия — не единственный способ оценить вектор параметров. Например, пусть $\mu(\theta)$ — математическое ожидание w_t, соответствующее функции плотности вероятности $f(w_t; \theta)$. Это известная функция от θ.

[1] Для θ, такого, что $f(w_t; \theta) = 0$, операция логарифмирования не определена. Это не является проблемой, потому что такое значение θ никогда не максимизирует функцию правдоподобия (в уровнях) $f(w_1, w_2, \ldots, w_n; \theta)$. Для таких значений θ мы можем присвоить $\ln f(w_t; \theta)$ сколь угодно малое значение таким образом, что значение, максимизирующее функцию правдоподобия, будет также максимизировать и логарифм функции правдоподобия. См.: [White, 1994, p. 15] для более формального рассмотрения этого вопроса.

По построению, должно выполняться следующее условие равенства нулю математического ожидания:

$$\mathrm{E}[\boldsymbol{w}_t - \boldsymbol{\mu}(\boldsymbol{\theta}_0)] = \mathbf{0}. \tag{7.1.7}$$

Вектор параметров $\boldsymbol{\theta}_0$ может быть оценен посредством GMM с $\boldsymbol{g}(\boldsymbol{w}_t; \boldsymbol{\theta}) = \boldsymbol{w}_t - \boldsymbol{\mu}(\boldsymbol{\theta})$ в качестве целевой функции GMM (7.1.3).

- (Эффективность ML.) Как будет показано в двух следующих параграфах, при выполнении соответствующих условий ML состоятельна и асимптотически нормальна. Достаточно распространенной была вера в то, что оценка ML эффективна (то есть обладает минимальной асимптотической дисперсией) в классе всех состоятельных и асимптотически нормальных оценок. С помощью контрпримера (описанного, например, в работе [Ameniya, 1985, Example 4.2.4] было показано, что эта вера является ложной. Тем не менее оценка ML является эффективной в достаточно общих классах асимптотически нормальных оценок. Одним из таких общих классов являются GMM-оценки. В параграфе 7.3 мы отметим, что асимптотическая дисперсия любой оценки $\boldsymbol{\theta}_0$, полученной GMM, не ниже, чем у оценки ML.

Пример 7.1 (оценивание математического ожидания нормального распределения): Пусть данные $(w_1, \ldots, w_n)$ — последовательность скалярных i.i.d. случайных величин и распределение w_t задается $N(\mu, \sigma^2)$. Таким образом, $\boldsymbol{\theta} = (\mu, \sigma^2)'$ и

$$f(w_t; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(w_t - \mu)^2}{2\sigma^2}\right].$$

Среднее значение логарифма функции правдоподобия для данных $(w_1, \ldots, w_n)$

$$\frac{1}{n}\sum_{t=1}^{n} \ln f(w_t; \mu, \sigma^2) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma^2) - \frac{1}{n}\sum_{t=1}^{n}\left[\frac{(w_t - \mu)^2}{2\sigma^2}\right]. \tag{7.1.8}$$

Оценка ML для (μ_0, σ_0^2) — это экстремальная оценка, где в качестве $Q_n(\boldsymbol{\theta})$ выступает это среднее значение логарифма функции правдоподобия. Предположим, что дисперсия предполагается положительной, но никаких ограничений на μ_0 *априори* не вводится; так что пространство параметров Θ — это $\mathbb{R} \times \mathbb{R}_{++}$, где $\mathbb{R}_{++}$ — множество положительных действительных чисел. Нетрудно проверить, что ML-оценка для μ_0 — это выборочное среднее w_t. GMM-оценка для μ_0, основанная на условии равенства нулю математического ожидания $\mathrm{E}(w_t - \mu_0) = 0$, также является выборочным средним w_t.

Метод условного максимального правдоподобия

В большинстве приложений вектор $\boldsymbol{w}_t$ разбивается на две группы — y_t и $\boldsymbol{x}_t$, и интерес для исследователя представляет изучение того, как $\boldsymbol{x}_t$ влияет на условное распределение y_t при заданном $\boldsymbol{x}_t$. Мы называли переменную y_t зависимой переменной, а $\boldsymbol{x}_t$ — регрессорами. Никакие результаты в этой главе не зависят от того, является ли y_t скаляром; y_t может быть вектором, и все равно все результаты останутся верными, если вместо скаляра «y_t» написать вектор «$\boldsymbol{y}_t$». Тем не менее мы будем использовать скалярное обозначение просто потому, что во всех примерах, рассматриваемых в этой главе, будет всего одна зависимая переменная.

Пусть $f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}_0)$ — условная функция плотности y_t при заданном $\boldsymbol{x}_t$, а $f(\boldsymbol{x}_t;\boldsymbol{\psi}_0)$ — безусловная (частная) функция плотности $\boldsymbol{x}_t$. Тогда

$$f(y_t, \boldsymbol{x}_t;\boldsymbol{\theta}_0, \boldsymbol{\psi}_0) = f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}_0)f(\boldsymbol{x}_t;\boldsymbol{\psi}_0) \tag{7.1.9}$$

будет совместной функцией плотности вероятности $\boldsymbol{w}_t = (y_t, \boldsymbol{x}_t')'$. Пока что предположим, что $\boldsymbol{\theta}_0$ и $\boldsymbol{\psi}_0$ не связаны функциональной зависимостью (см. следующий абзац для примера функциональной связи). Среднее значение логарифма функции правдоподобия для данных $(\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_n)$ равно:

$$\frac{1}{n}\sum_{t=1}^{n} \ln f(\boldsymbol{w}_t;\boldsymbol{\theta}, \boldsymbol{\psi}) = \frac{1}{n}\sum_{t=1}^{n} \ln f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}) + \frac{1}{n}\sum_{t=1}^{n} \ln f(\boldsymbol{x}_t;\boldsymbol{\psi}). \tag{7.1.10}$$

Первый элемент, стоящий справа от знака равенства, — это среднее значение **логарифма условной функции правдоподобия** (*log conditional likelihood*). **Оценка условного ML** (*conditional ML estimator*) для $\boldsymbol{\theta}_0$ максимизирует этот первый элемент, игнорируя, таким образом, второй элемент. Это M-оценка с

$$m(\boldsymbol{w}_t;\boldsymbol{\theta}) = \ln f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}), \quad \text{то есть} \quad Q_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n} \ln f(y_t|\boldsymbol{x}_t, \boldsymbol{\theta}). \tag{7.1.11}$$

Второй элемент, стоящий справа от знака равенства в (7.1.10), — среднее значение логарифма частной функции правдоподобия. Оно не зависит от $\boldsymbol{\theta}$, так что оценка условного ML для $\boldsymbol{\theta}_0$ численно совпадает с (совместной или полной) оценкой ML, которая максимизирует среднее значение совместной функции правдоподобия (7.1.10).

Предположим теперь, что $\boldsymbol{\theta}_0$ и $\boldsymbol{\psi}_0$ связаны друг с другом функциональной зависимостью. Например, векторы $\boldsymbol{\theta}_0$ и $\boldsymbol{\psi}_0$ можно разбить как $\boldsymbol{\theta}_0 = (\boldsymbol{\alpha}_0', \boldsymbol{\beta}_0')'$, $\boldsymbol{\psi}_0 = (\boldsymbol{\beta}_0', \boldsymbol{\gamma}_0')'$. В этом случае оценки полного (безусловного) ML и условного ML перестанут быть численно одинаковыми.

Интуитивно понятно, что условному ML недостает информации, которая могла бы быть получена из частной функции правдоподобия. На самом деле можно показать, что оценка условного ML для $\boldsymbol{\theta}_0$ менее эффективна, чем оценка полного ML, полученная из задачи максимизации логарифма совместной функции правдоподобия (7.1.10) (см., например: [Gourieroux and Monfort, 1995, Section 7.5.3]. В большинстве прикладных исследований эта потеря эффективности неизбежна, так как мы не знаем или не хотим специфицировать параметрическую форму $f(\boldsymbol{x}_t; \psi)$ частного распределения.

Вот два примера условного ML (другие примеры будут в следующей главе).

Пример 7.2 (модель линейной регрессии с нормальными ошибками): В главе 2 мы рассматривали модель линейной регрессии с условной гомоскедастичностью. Предположим далее, что ошибка имеет нормальное распределение, так что

$$\{y_t, \boldsymbol{x}_t\} \text{ i.i.d.}, \quad y_t = \boldsymbol{x}_t'\boldsymbol{\beta}_0 + \varepsilon_t, \quad \varepsilon_t | \boldsymbol{x}_t \sim N(0, \sigma_0^2). \tag{7.1.12}$$

Условная функция правдоподобия y_t при условии $\boldsymbol{x}_t$, $f(y_t|\boldsymbol{x}_t; \boldsymbol{\theta})$, — это функция плотности для $N(\boldsymbol{x}_t'\boldsymbol{\beta}, \sigma^2)$. Таким образом, для $\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma^2)'$ и $\boldsymbol{w}_t = (y_t, \boldsymbol{x}_t)$ функция m будет выглядеть как

$$\begin{aligned} m(\boldsymbol{w}_t; \boldsymbol{\theta}) &= \ln f(y_t|\boldsymbol{x}_t; \boldsymbol{\beta}, \sigma^2) = \\ &= \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_t - \boldsymbol{x}_t'\boldsymbol{\beta})^2}{2\sigma^2}\right]\right) = \\ &= -\frac{1}{2}\ln(2\pi) - \ln(\sigma) - \frac{1}{2}\left(\frac{y_t - \boldsymbol{x}_t'\boldsymbol{\beta}}{\sigma}\right)^2. \end{aligned} \tag{7.1.13}$$

Пространством параметров Θ будет $\mathbb{R}^K \times \mathbb{R}_{++}$, где K — размерность $\boldsymbol{\beta}$, а $\mathbb{R}_{++}$ — множество положительных действительных чисел, что отражает *априорное* ограничение на дисперсию ошибок: $\sigma_0^2 > 0$. Как было проверено в главе 1, оценкой ML для $\boldsymbol{\beta}_0$ является оценка МНК, а оценкой ML для σ_0^2 будет сумма квадратов остатков, деленная на размер выборки n.

Пример 7.3 (модель пробит): В модели пробит (*probit model*) скалярная зависимая переменная y_t бинарная: $y_t \in \{0, 1\}$. Например, $y_t = 1$, если жена решает выйти на рынок труда, и $y_t = 0$ в противном случае; а $\boldsymbol{x}_t$ может включать зарплату мужа. Условное распределение y_t при заданном векторе регрессоров $\boldsymbol{x}_t$ записывается как

$$\begin{cases} f(y_t = 1|\boldsymbol{x}_t; \boldsymbol{\theta}_0) = \Phi(\boldsymbol{x}_t'\boldsymbol{\theta}_0), \\ f(y_t = 0|\boldsymbol{x}_t; \boldsymbol{\theta}_0) = 1 - \Phi(\boldsymbol{x}_t'\boldsymbol{\theta}_0). \end{cases} \tag{7.1.14}$$

где $\Phi(\cdot)$ — функция распределения стандартного нормального распределения. Более компактно это можно записать как

$$f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}_0) = \Phi(\boldsymbol{x}_t'\boldsymbol{\theta}_0)^{y_t}[1-\Phi(\boldsymbol{x}_t'\boldsymbol{\theta}_0)]^{(1-y_t)}.$$

Если не задаются никакие *априорные* ограничения на $\boldsymbol{\theta}_0$, множество параметров Θ представляет собой $\mathbb{R}^p$. Оценка ML для $\boldsymbol{\theta}_0$ будет M-оценкой с функцией m вида:

$$m(\boldsymbol{w}_t;\boldsymbol{\theta}) = \ln f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}) = y_t \ln\Phi(\boldsymbol{x}_t'\boldsymbol{\theta}) + (1-y_t)\ln[1-\Phi(\boldsymbol{x}_t'\boldsymbol{\theta})]. \quad (7.1.15)$$

Инвариантность оценок ML

Оценки безусловного и условного ML обладают некоторыми желаемыми свойствами. Одно из них, которое выполняется в конечных выборках, — это **инвариантность** (*invariance*). Чтобы описать свойство инвариантности для экстремальной оценки в общем случае, рассмотрим **репараметризацию** (*reparameterizing*) модели с помощью отображения или функции $\boldsymbol{\lambda} = \boldsymbol{\tau}(\boldsymbol{\theta})$, определенной на Θ. Пусть Λ будет множеством значений отображения:

$$\Lambda \equiv \boldsymbol{\tau}(\Theta) \equiv \{\boldsymbol{\lambda}|\boldsymbol{\lambda} = \boldsymbol{\tau}(\boldsymbol{\theta}) \text{ для некоторого } \boldsymbol{\theta}\in\Theta\}. \quad (7.1.16)$$

По определению, для любого $\boldsymbol{\lambda}$ из Λ существует хотя бы один вектор $\boldsymbol{\theta}$, такой, что $\boldsymbol{\lambda} = \boldsymbol{\tau}(\boldsymbol{\theta})$. Отображение $\boldsymbol{\tau}\colon \Theta \to \Lambda$ называется репараметризацией, если оно однозначное, то есть если для каждого $\boldsymbol{\lambda}$ из Λ существует ровно один $\boldsymbol{\theta}$, такой, что $\boldsymbol{\lambda} = \boldsymbol{\tau}(\boldsymbol{\theta})$. Этот единственный $\boldsymbol{\theta}$ является, таким образом, функцией от $\boldsymbol{\lambda}$. Это отображение Λ на Θ называется **обратным** (*inverse*) к $\boldsymbol{\tau}$ и обозначается как $\boldsymbol{\tau}^{-1}$. Мы будем говорить, что экстремальная оценка $\widehat{\boldsymbol{\theta}}$ **инвариантна** (*invariant*) к репараметризации $\boldsymbol{\tau}$, если экстремальной оценкой для репараметризованной модели будет $\boldsymbol{\tau}(\widehat{\boldsymbol{\theta}})$. Пусть $\widetilde{Q}_n(\boldsymbol{\lambda})$ — целевая функция, соответствующая репараметризованной модели. Экстремальная оценка будет инвариантной, если и только если

$$\widetilde{Q}_n(\boldsymbol{\lambda}) = Q_n(\boldsymbol{\tau}^{-1}(\boldsymbol{\lambda})) \quad \text{для всех} \quad \boldsymbol{\lambda}\in\Lambda. \quad (7.1.17)$$

Чтобы показать это, положим $\widehat{\boldsymbol{\lambda}} \equiv \boldsymbol{\tau}(\widehat{\boldsymbol{\theta}})$. Для любого $\boldsymbol{\lambda}\in\Lambda$ мы имеем $\widetilde{Q}_n(\boldsymbol{\lambda}) = Q_n(\boldsymbol{\tau}^{-1}(\boldsymbol{\lambda})) \leqslant Q_n(\widehat{\boldsymbol{\theta}})$, поскольку $\widehat{\boldsymbol{\theta}}$ максимизирует $Q_n(\boldsymbol{\theta})$ на множестве Θ, а $\boldsymbol{\tau}^{-1}(\boldsymbol{\lambda})$ принадлежит Θ. Но $Q_n(\widehat{\boldsymbol{\theta}}) = Q_n(\boldsymbol{\tau}^{-1}(\widehat{\boldsymbol{\lambda}})) = \widetilde{Q}_n(\widehat{\boldsymbol{\lambda}})$. Так что $\widetilde{Q}_n(\boldsymbol{\lambda}) \leqslant \widetilde{Q}_n(\widehat{\boldsymbol{\lambda}})$ для всех $\boldsymbol{\lambda}\in\Lambda$.

Метод максимального правдоподобия (условный и безусловный) инвариантен к репараметризации, поскольку функция правдоподобия после репараметризации принимает вид $\widetilde{f}(\cdot;\boldsymbol{\lambda}) = f(\cdot;\boldsymbol{\tau}^{-1}(\boldsymbol{\lambda}))$, что дает целевую функцию, которая удовлетворяет (7.1.17). Существует ли *не*инвариантная оценка? Ответ: да, может; в вопросе 5 вас попросят проверить, что оценка GMM не является инвариантной.

Нелинейный метод наименьших квадратов (NLS)

В NLS, как и в условном ML, вектор наблюдений $\boldsymbol{w}_t$ разбивается на две группы: y_t и $\boldsymbol{x}_t$ ($\boldsymbol{w}_t = (y_t, \boldsymbol{x}_t')'$). Моделью в NLS является набор случайных процессов $\{y_t, \boldsymbol{x}_t\}$, таких, что условное математическое ожидание $E(y_t|\boldsymbol{x}_t)$, являющееся функцией от $\boldsymbol{x}_t$, принадлежит семейству функций $\varphi(\boldsymbol{x}_t;\boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta$. Функциональная форма $\varphi(\cdot;\cdot)$ известна. Если $\boldsymbol{\theta}_0$ — истинное значение вектора параметров, то $E(y_t|\boldsymbol{x}_t) = \varphi(\boldsymbol{x}_t;\boldsymbol{\theta}_0)$ для истинного процесса порождения данных $\{y_t, \boldsymbol{x}_t\}$. Если мы определим $\varepsilon_t \equiv y_t - E(y_t|\boldsymbol{x}_t)$, то правильно специфицированная модель может быть записана как

$$y_t = \varphi(\boldsymbol{x}_t;\boldsymbol{\theta}_0) + \varepsilon_t, \ \mathrm{E}(\varepsilon_t|\boldsymbol{x}_t) = 0, \ \boldsymbol{\theta}_0 \in \Theta. \tag{7.1.18}$$

Наиболее часто использующийся метод для оценивания $\boldsymbol{\theta}_0$ — это метод **наименьших квадратов** (*least squares*), который минимизирует сумму квадратов остатков. Таким образом, оценивание NLS, которое является методом наименьших квадратов, примененным к вышеописанной модели, является M-оценкой с

$$m(\boldsymbol{w}_t;\boldsymbol{\theta}) = -[y_t - \varphi(\boldsymbol{x}_t;\boldsymbol{\theta})]^2, \quad \text{то есть} \quad Q_n(\boldsymbol{\theta}) = -\frac{1}{n}\sum_{t=1}^{n}[y_t - \varphi(\boldsymbol{x}_t;\boldsymbol{\theta})]^2. \tag{7.1.19}$$

Здесь максимизация $Q_n(\boldsymbol{\theta})$ идентична *минимизации* суммы квадратов остатков.

Пример 7.4 (производственная функция CES с аддитивными ошибками): В качестве иллюстрации рассмотрим производственную функцию CES[1]

$$Q_t = A_0 \cdot \left[\delta_0 K_t^{-\rho_0} + (1-\delta_0)L_t^{-\rho_0}\right]^{-1/\rho_0} + \varepsilon_t. \tag{7.1.20}$$

где Q_t — выпуск в период времени t, K_t — затраты капитала, L_t — затраты труда, а ε_t — аддитивный к производственной функции шок с $E(\varepsilon_t|K_t, L_t) = 0$. Это модель условного математического ожидания (7.1.18) с $y_t = Q_t$, $\boldsymbol{x}_t = (K_t, L_t)'$, $\boldsymbol{\theta}_0 = (A_0, \delta_0, \rho_0)'$, $\varphi(\boldsymbol{x}_t;\boldsymbol{\theta}) = A \cdot \left[\delta K_t^{-\rho} + (1-\delta)L_t^{-\rho}\right]^{-1/\rho}$. Стандартные свойства производственных функций (монотонность и вогнутость) выполняются, если $A > 0$, $0 < \delta < 1$, $-1 < \rho$, и это определяет пространство параметров Θ.

[1]CES — Constant elasticity of substitution — производственная функция с постоянной эластичностью замещения. — *Прим. науч. ред. перевода.*

Линейный и нелинейный GMM

В главе 3 мы применяли GMM для оценивания линейных уравнений. Оцениваемым линейным уравнением было

$$y_t = \boldsymbol{z}_t'\boldsymbol{\theta}_0 + \varepsilon_t. \tag{7.1.21}$$

Вектором инструментов был $\boldsymbol{x}_t$, а условиями ортогональности (равенства нулю математического ожидания) были уравнения:

$$\mathrm{E}[\boldsymbol{x}_t \cdot (y_t - \boldsymbol{z}_t'\boldsymbol{\theta}_0)] = \mathbf{0}. \tag{7.1.22}$$

Правильно специфицированной моделью здесь является набор эргодических стационарных процессов $\boldsymbol{w}_t = (y_t, \boldsymbol{z}_t', \boldsymbol{x}_t')$, таких, что эти условия равенства нулю математических ожиданий выполняются для $\boldsymbol{\theta}_0$, принадлежащего Θ. Оценкой линейного GMM для $\boldsymbol{\theta}_0$ является GMM-оценка с функцией $\boldsymbol{g}$ в целевой функции GMM (7.1.13), заданной следующим образом:

$$\boldsymbol{g}(\boldsymbol{w}_t; \boldsymbol{\theta}) \equiv \boldsymbol{x}_t \cdot (y_t - \boldsymbol{z}_t'\boldsymbol{\theta}) = \boldsymbol{x}_t \cdot y_t - \boldsymbol{x}_t\boldsymbol{z}_t'\boldsymbol{\theta}. \tag{7.1.23}$$

GMM может быть легко применен и к нелинейным уравнениям. Предположим, что оцениваемое уравнение нелинейно, а также что y_t входит в это уравнение неявно:

$$a(y_t, \boldsymbol{z}_t; \boldsymbol{\theta}_0) = \varepsilon_t. \tag{7.1.24}$$

Просто возьмем в качестве функции $\boldsymbol{g}$ следующую: $\boldsymbol{g}(\boldsymbol{w}_t; \boldsymbol{\theta}) = \boldsymbol{x}_t \cdot a(y_t, \boldsymbol{z}_t; \boldsymbol{\theta})$. Оценка такого вида называется **обобщенной оценкой нелинейного метода инструментальных переменных** (*generalized nonlinear instrumental variables estimator*)[1]. Это все еще особый вид GMM, потому что функция $\boldsymbol{g}$ здесь может быть записана как произведение вектора инструментов и ошибки. Рассматриваемые в данной главе свойства GMM-оценок не опираются на эту особую структуру.

В качестве примера нелинейного GMM рассмотрим нелинейное уравнение Эйлера [Hansen and Singleton, 1982].

Пример 7.5 (нелинейное уравнение Эйлера для потребления): Уравнение Эйлера для оптимизационной задачи домохозяйства, являющееся стандартным для макроэкономики, имеет вид:

$$\mathrm{E}\left[R_{t+1}\frac{\beta_0 u'(c_{t+1})}{u'(c_t)} \;\middle|\; I_t\right] = 1, \tag{7.1.25}$$

где R_{t+1} — валовая ex post норма отдачи (1 плюс норма отдачи) от рассматриваемого актива, c_t — потребление, β_0 — дисконт-фактор, $u'(c)$ —

[1] Эта оценка более общая, нежели обычная оценка нелинейного метода инструментальных переменных, потому что здесь не предполагается условная гомоскедастичность.

предельная полезность, а I_t — информация, доступная в момент времени t. Возьмем в качестве функции полезности $u(c) = c^{1-\alpha_0}/(1-\alpha_0)$. Тогда $u'(c) = c^{-\alpha_0}$, и уравнение Эйлера приобретает вид:

$$\mathrm{E}\left[a\left(\frac{c_{t+1}}{c_t}, R_{t+1}; \alpha_0, \beta_0\right) \;\middle|\; I_t\right] = 0,$$

$$\text{где} \quad a\left(\frac{c_{t+1}}{c_t}, R_{t+1}; \alpha_0, \beta_0\right) \equiv R_{t+1} \cdot \beta_0 \cdot \left(\frac{c_{t+1}}{c_t}\right)^{-\alpha_0} - 1. \quad (7.1.26)$$

Если $\boldsymbol{x}_t$ — вектор переменных, значения которых известны в момент времени t, то $\boldsymbol{x}_t \in I_t$, и, в соответствии с (7.1.26),

$$\mathrm{E}\left[\boldsymbol{x}_t \cdot a\left(\frac{c_{t+1}}{c_t}, R_{t+1}; \alpha_0, \beta_0\right) \;\middle|\; I_t\right] = \mathbf{0}. \quad (7.1.27)$$

Применяя оператор безусловного математического ожидания к обеим сторонам и используя закон полных математических ожиданий ($E[E(x|I_t)] = E(x)$), мы получим условия ортогональности (равенства нулю математического ожидания) $E[\boldsymbol{g}(\boldsymbol{w}_t; \boldsymbol{\theta}_0)] = \mathbf{0}$, где

$$\boldsymbol{g}(\boldsymbol{w}_t; \boldsymbol{\theta}_0) = \boldsymbol{x}_t \cdot a\left(\frac{c_{t+1}}{c_t}, R_{t+1}; \alpha_0, \beta_0\right) \quad \text{с} \quad \boldsymbol{w}_t = \begin{bmatrix} \boldsymbol{x}_t \\ c_{t+1}/c_t \\ R_{t+1} \end{bmatrix}, \; \boldsymbol{\theta}_0 = \begin{bmatrix} \beta_0 \\ \alpha_0 \end{bmatrix}. \quad (7.1.28)$$

Контрольные вопросы

1. (Оценивание модели пробит посредством NLS.) Проверьте, что для модели пробит $E(y_t|\boldsymbol{x}_t) = \Phi(\boldsymbol{x}_t'\boldsymbol{\theta}_0)$. Рассмотрите применение метода наименьших квадратов к этой модели. Специфицируйте функцию m для целевой функции M-оценки (7.1.2).

2. (Оценивание модели пробит посредством GMM.) Проверьте, что для модели пробит выполняются условия ортогональности $E(\boldsymbol{x}_t \cdot (y_t - \Phi(\boldsymbol{x}_t'\boldsymbol{\theta}_0))) = \mathbf{0}$. Специфицируйте функцию $\boldsymbol{g}$ для целевой функции оценки GMM (7.1.13).

3. (Оценивание модели GMM посредством ML?) Предположим, что величины $\{y_t, \boldsymbol{z}_t, \boldsymbol{x}_t\}$ i.i.d. в правильно специфицированной линейной модели, описанной в последнем подпараграфе. Сможете ли вы оценить $\boldsymbol{\theta}_0$ с помощью ML? [Ответ: Нет. Потому что эта модель не специфицирует параметрическую форму функции плотности $\boldsymbol{w}_t$.]

4. (Отличие $E(\varepsilon_t|\boldsymbol{x}_t) = 0$ от $E(\varepsilon_t \cdot \boldsymbol{x}_t) = \mathbf{0}$ в NLS.) Рассмотрите квадратичную модель NLS, где $\varphi(x_t; \boldsymbol{\theta}) = \theta_0 + \theta_1 x_t + \theta_2 x_t^2$. Предположим, что $E(\varepsilon_t x_t) = 0$, но $E(\varepsilon_t x_t^2) \neq 0$. Выполняется ли условие $E(\varepsilon_t|x_t) = 0$? [Ответ: Нет. Если $E(\varepsilon_t|x_t) = 0$, то $E(\varepsilon_t h(x_t)) = 0$ для любой (измеримой) функции h.] Будет ли оценка NLS для $\boldsymbol{\theta}_0$ состоятельной? [Ответ: Нет.]

5. (Отсутствие свойства инвариантности для GMM.) Рассмотрим линейную модель $y_t = \theta_0 z_t + \varepsilon_t$, где θ_0 и z_t — скаляры. Условия ортогональности: $E[\boldsymbol{x}_t \cdot (y_t - \theta_0 z_t)] = \mathbf{0}$. Запишите целевую функцию GMM $Q_n(\theta)$. Предположите, что $\Theta = \mathbb{R}_{++}$ (то есть $\theta_0 > 0$), и рассмотрите репараметризацию $\lambda = 1/\theta$. Используя эту репараметризацию, линейное уравнение можно переписать как $z_t = \lambda_0 y_t - \lambda_0 \varepsilon_t$, а условия ортогональности можно переписать как $E[\boldsymbol{x}_t \cdot (z_t - \lambda_0 y_t)] = \mathbf{0}$. Пусть $\widetilde{Q}_n(\lambda)$ будет целевой функцией GMM, ассоциируемой с этим набором условий ортогональности. Будут ли в данном случае выполняться равенства $Q_n(\theta) = \widetilde{Q}_n(1/\theta)$ или $\widetilde{Q}_n(\lambda) = Q_n(1/\lambda)$, если $\widehat{\boldsymbol{W}}$ одинакова в обеих целевых функциях? [Ответ: Нет.]

7.2. Состоятельность

В этом параграфе мы сначала представим набор достаточных условий, при выполнении которых экстремальная оценка состоятельна. Эти условия будут потом адаптированы для NLS, ML и GMM.

Две теоремы о состоятельности экстремальных оценок

Целевая функция $Q_n(\cdot)$ — случайная функция, потому что для каждого значения $\boldsymbol{\theta}$ ее значение $Q_n(\boldsymbol{\theta})$, будучи зависимым от данных $(\boldsymbol{w}_1, \dots, \boldsymbol{w}_n)$, является случайной величиной. Основополагающая идея состоятельности экстремальной оценки состоит в том, что если $Q_n(\boldsymbol{\theta})$ сходится по вероятности к $Q_0(\boldsymbol{\theta})$ и истинный параметр $\boldsymbol{\theta}_0$ решает «предельную задачу» максимизации предельной функции $Q_0(\boldsymbol{\theta})$, то пределом максимизирующего значения $\widehat{\boldsymbol{\theta}}$ должен быть вектор $\boldsymbol{\theta}_0$. Достаточными условиями того, что максимизирующее значение предельной функции будет пределом максимизирующего значения, являются требование к сходимости по вероятности быть «равномерной» (мы проясним смысл этого очень скоро) и требование к пространству параметров Θ быть компактным множеством. Как уже говорилось, пространство параметров не является компактом в большинстве прикладных задач. Поэтому мы приведем альтернативный набор условий состоятельности, которые не требуют, чтобы Θ было компактом.

Чтобы представить первую теорему о состоятельности экстремальных оценок, надо уточнить, что же мы имеем в виду, когда говорим, что сходимость «равномерна». Если вы уже знакомы из математического анализа с понятием равномерной сходимости последовательности *функций*, поймете, что определение, которое вы сейчас увидите, — это расширение определения на последовательность случайных функций, $Q_n(\cdot)$ $(n = 1, 2, \dots)$. **Точечная сходимость по вероятности** (*Pointwise convergence in probability*) $Q_n(\cdot)$ к некоторой неслучайной функции $Q_0(\cdot)$ просто означает, что $\operatorname{plim}_{n\to\infty} Q_n(\boldsymbol{\theta}) = Q_0(\boldsymbol{\theta})$ для всех $\boldsymbol{\theta}$, а именно что последовательность случайных величин $|Q_n(\boldsymbol{\theta}) - Q_0(\boldsymbol{\theta})|$ $(n = 1, 2, \dots)$

сходится по вероятности к 0 для любого $\boldsymbol{\theta}$. **Равномерная сходимость по вероятности** (*uniform convergence in probability*) сильнее. Сходимость должна происходить «равномерно» на пространстве параметров Θ в следующем смысле:

$$\sup_{\boldsymbol{\theta}\in\Theta} |Q_n(\boldsymbol{\theta}) - Q_0(\boldsymbol{\theta})| \underset{\mathrm{p}}{\to} 0 \quad \text{при} \quad n \to \infty. \tag{7.2.1}$$

Как и в случае сходимости по вероятности, можно просто расширить определение равномерной сходимости по вероятности на последовательности *векторных* случайных функций или матричных случайных функций (рассматривая матрицы как векторы, элементы которых были переставлены), требуя равномерной сходимости каждого элемента: последовательность векторных случайных функций $\{\boldsymbol{h}_n(\cdot)\}$ равномерно сходится по вероятности к неслучайной функции $\boldsymbol{h}_0(\cdot)$, если каждый элемент сходится равномерно. Эта поэлементная сходимость эквивалентна сходимости по норме:

$$\sup_{\boldsymbol{\theta}\in\Theta} \|\boldsymbol{h}_n(\boldsymbol{\theta}) - \boldsymbol{h}_0(\boldsymbol{\theta})\| \underset{\mathrm{p}}{\to} 0 \quad \text{при} \quad n \to \infty. \tag{7.2.2}$$

где $\|.\|$ — обычная Евклидова норма[1].

В следующем утверждении нашей первой теоремы о состоятельности условия, обеспечивающие однозначную определенность $\widehat{\boldsymbol{\theta}}$, обозначены как (i)–(iii), а те условия, которые обеспечивают состоятельность, — это (a) и (b).

Утверждение 7.1 (состоятельность для компактного пространства параметров): *Пусть (i) Θ — компактное подмножество $\mathbb{R}^p$, (ii) $Q_n(\boldsymbol{\theta})$ непрерывна по $\boldsymbol{\theta}$ для любого набора данных $(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n)$ и (iii) $Q_n(\boldsymbol{\theta})$ — измеримая функция от данных для любых значений $\boldsymbol{\theta}$ из Θ (так что по лемме 7.1 экстремальная оценка $\widehat{\boldsymbol{\theta}}$, определенная в (7.1.1), — это корректно определенная случайная величина). Если существует функция $Q_0(\boldsymbol{\theta})$, такая, что*

(a) (Идентифицируемость.) $Q_0(\boldsymbol{\theta})$ достигает своего максимального значения на Θ исключительно в $\boldsymbol{\theta}_0 \in \Theta$,

(b) (Равномерная сходимость.) $Q_n(\cdot)$ сходится равномерно по вероятности к $Q_0(\cdot)$,

то тогда $\widehat{\boldsymbol{\theta}} \to_p \boldsymbol{\theta}_0$.

[1] Евклидова норма p-мерного вектора $\boldsymbol{x} = (x_1, x_2, \ldots, x_p)'$, обозначаемая как $\|\boldsymbol{x}\|$, определяется как $\sqrt{x_1^2 + x_2^2 + \cdots + x_p^2}$.

Мы не будем доказывать этот результат; по поводу доказательства смотрите, например: [Amemiya, 1985, Theorem 4.1.1]. Для оценок, таких как оценки ML, NLS и GMM, мы позднее сформулируем условие идентификации (a) в более простом виде, чтобы его можно было проще интерпретировать интуитивно.

Теорема, которая не требует компактности Θ, выглядит следующим образом:

Утверждение 7.2 (состоятельность без требования компактности): *Пусть (i) — истинный вектор параметров* $\boldsymbol{\theta}_0$ *— является внутренним элементом выпуклого пространства параметров* Θ ($\subset \mathbb{R}^p$), *(ii)* $Q_n(\boldsymbol{\theta})$ *вогнута*[1] *на пространстве параметров для любого набора данных* $(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n)$ *и (iii)* $Q_n(\boldsymbol{\theta})$ *— измеримая функция от данных для любого* $\boldsymbol{\theta}$ *из* Θ. *Пусть* $\widehat{\boldsymbol{\theta}}$ *— экстремальная оценка, определенная в (7.1.1) (подождите немного, чтобы узнать о ее существовании). Если существует функция* $Q_0(\boldsymbol{\theta})$, *такая, что*

(a) *(Идентифицируемость.)* $Q_0(\boldsymbol{\theta})$ *достигает своего максимального значения на* Θ *исключительно в* $\boldsymbol{\theta}_0 \in \Theta$,

(b) *(Точечная сходимость.)* $Q_n(\boldsymbol{\theta})$ *сходится по вероятности к* $Q_0(\boldsymbol{\theta})$ *для всех* $\boldsymbol{\theta} \in \Theta$,

то тогда, при $n \to \infty$, $\widehat{\boldsymbol{\theta}}$ *существует с вероятностью, приближающейся к 1, и* $\widehat{\boldsymbol{\theta}} \to_p \boldsymbol{\theta}_0$.

По поводу доказательства см.: [Newey, McFadden 1994, pp. 2133–2144]. В большинстве прикладных задач Θ — открытое выпуклое множество (как в примерах 7.1–7.4), так что условие (i) выполняется. Требуется только точечная сходимость $Q_n(\boldsymbol{\theta})$ к $Q_0(\boldsymbol{\theta})$, что проще проверять в прикладных задачах. Однако за это приходится платить требованием вогнутости $Q_n(\boldsymbol{\theta})$ для любого набора данных (при вогнутости этой функции равномерная сходимость будет следовать из точечной сходимости), но во многих прикладных задачах это условие выполняется. Ниже мы проверим, что условие вогнутости выполняется для ML-оценивания модели линейной регрессии (после репараметризации), ML-оценивания модели пробит и линейного GMM.

[1] Скалярная функция $h(\boldsymbol{x})$ называется **вогнутой**, если $\alpha h(\boldsymbol{x}_1) + (1 - \alpha) h(\boldsymbol{x}_2) \leqslant h(\alpha \boldsymbol{x}_1 + (1 - \alpha)\boldsymbol{x}_2)$ для всех $\boldsymbol{x}_1, \boldsymbol{x}_2$, $0 \leqslant \alpha \leqslant 1$. Так, линейная функция вогнута. Некоторые авторы используют термин «глобальная вогнутость», чтобы отличить вышеописанный случай от концепции «локальной вогнутости», которая требует, чтобы гессиан, вычисленный для рассматриваемого вектора параметров, был неотрицательно определен. Если функция вогнута, она непрерывна. Так что условие (ii) в этом утверждении сильнее, чем условие (ii) в предыдущей теореме.

Состоятельность M-оценок

Для M-оценки целевая функция задается соотношением (7.1.2). Если $\{\boldsymbol{w}_t\}$ эргодический и стационарный, то из теоремы об эргодичности вытекает точечная сходимость $Q_n(\boldsymbol{\theta})$: $Q_n(\boldsymbol{\theta})$ точечно сходится по вероятности для любого значения $\boldsymbol{\theta} \in \Theta$ к $Q_0(\boldsymbol{\theta})$, задаваемой

$$Q_0(\boldsymbol{\theta}) = \mathrm{E}[m(\boldsymbol{w}_t;\boldsymbol{\theta})]. \tag{7.2.3}$$

(естественно, при условии, что $\mathrm{E}[m(\boldsymbol{w}_t;\boldsymbol{\theta})]$ существует и конечно). Чтобы использовать первый полученный результат о состоятельности (утверждение 7.1) для M-оценок, нам необходимо показать, что сходимость $\frac{1}{n}\sum_{t=1}^{n} m(\boldsymbol{w}_t;.)$ к $\mathrm{E}[m(\boldsymbol{w}_t;\boldsymbol{\theta})]$ является равномерной. Стандартный способ доказать равномерную сходимость по вероятности, который использует тот факт, что выражение является выборочным средним, — это **равномерный закон больших чисел** (*uniform law of large numbers.*). Версия этого закона для эргодических стационарных процессов (приведена как Lemma 2.4. в [Newey, McFadden, 1994] и как Theorem A.2.2 в [White, 1994]) выглядит следующим образом:

Лемма 7.2 (равномерный закон больших чисел): *Пусть $\{\boldsymbol{w}_t\}$ — эргодический стационарный процесс. Предположим, что (i) множество Θ компактно, (ii) $m(\boldsymbol{w}_t;\boldsymbol{\theta})$ непрерывна по $\boldsymbol{\theta}$ для всех $\boldsymbol{w}_t$, и (iii) $m(\boldsymbol{w}_t;\boldsymbol{\theta})$ измерима по $\boldsymbol{w}_t$ для всех $\boldsymbol{\theta}$ из Θ. Предположим также, что выполняется*

> **«условие доминирования»** *(dominance condition): существует функция $d(\boldsymbol{w}_t)$ (иногда называемая доминирующей функцией), такая, что $|m(\boldsymbol{w}_t;\boldsymbol{\theta})| \leqslant d(\boldsymbol{w}_t)$ для всех $\boldsymbol{\theta} \in \Theta$ и $\mathrm{E}[d(\boldsymbol{w}_t)] < \infty$.*

Тогда $\frac{1}{n}\sum_{t=1}^{n} m(\boldsymbol{w}_t;.)$ сходится равномерно по вероятности к $\mathrm{E}[m(\boldsymbol{w}_t;.)]$ на Θ. Более того, $\mathrm{E}[m(\boldsymbol{w}_t;\boldsymbol{\theta})]$ — непрерывная функция от $\boldsymbol{\theta}$.

Так как, по построению, $|m(\boldsymbol{w}_t;\boldsymbol{\theta})| \leqslant \sup_{\boldsymbol{\theta}\in\Theta} |m(\boldsymbol{w}_t;\boldsymbol{\theta})|$ для всех $\boldsymbol{\theta} \in \Theta$, мы можем использовать $\sup_{\boldsymbol{\theta}\in\Theta} |m(\boldsymbol{w}_t;\boldsymbol{\theta})|$ для $d(\boldsymbol{w}_t)$, так что представленное выше условие доминирования может быть записано просто как $E[\sup_{\boldsymbol{\theta}\in\Theta} |m(\boldsymbol{w}_t;\boldsymbol{\theta})|] < \infty$. Равномерный закон больших чисел может быть сразу же расширен на векторные случайные функции:

Пусть $\{\boldsymbol{w}_t\}$ — эргодический стационарный процесс. Предположим, что (i) пространство параметров Θ компактно, (ii) $\boldsymbol{h}(\boldsymbol{w}_t;\boldsymbol{\theta})$ непрерывна по $\boldsymbol{\theta}$ для всех $\boldsymbol{w}_t$ и (iii) $\boldsymbol{h}(\boldsymbol{w}_t;\boldsymbol{\theta})$ измерима по $\boldsymbol{w}_t$ для всех $\boldsymbol{\theta}$ из Θ. Предположим также, что выполняется

> **«условие доминирования»** *(dominance condition):*
>
> $$E[\sup_{\boldsymbol{\theta}\in\Theta} \|\boldsymbol{h}(\boldsymbol{w}_t;\boldsymbol{\theta})\|] < \infty.$$

Тогда $E[\boldsymbol{h}(\boldsymbol{w}_t;\boldsymbol{\theta})]$ — непрерывная функция от $\boldsymbol{\theta}$ и $\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{h}(\boldsymbol{w}_t;.)$ сходится равномерно по вероятности к $\mathrm{E}[\boldsymbol{h}(\boldsymbol{w}_t;.)]$ на Θ.

Комбинируя эту лемму с первой теоремой состоятельности экстремальных оценок и учитывая то, что $Q_n(\boldsymbol{\theta})$ непрерывна, если $m(\boldsymbol{w}_t; \boldsymbol{\theta})$ непрерывна по $\boldsymbol{\theta}$, и что $Q_n(\boldsymbol{\theta})$ — измеримая функция данных, если $m(\boldsymbol{w}_t; \boldsymbol{\theta})$ измерима по $\boldsymbol{w}_t$, мы получаем

Утверждение 7.3 (состоятельность М-оценок с компактным пространством параметров): *Пусть $\{\boldsymbol{w}_t\}$ — эргодический стационарный процесс. Предположим, что (i) пространство параметров Θ — компактное подмножество $\mathbb{R}^p$, (ii) $m(\boldsymbol{w}_t; \boldsymbol{\theta})$ непрерывна по $\boldsymbol{\theta}$ для всех $\boldsymbol{w}_t$, и (iii) $m(\boldsymbol{w}_t; \boldsymbol{\theta})$ измерима по $\boldsymbol{w}_t$ для всех $\boldsymbol{\theta}$ в Θ. Пусть $\widehat{\boldsymbol{\theta}}$ — М-оценка, определенная в (7.1.1) и (7.1.2). Предположим также, что*

(a) (Идентификация.) $\mathrm{E}[m(\boldsymbol{w}_t; \boldsymbol{\theta})]$ принимает максимальное значение на множестве Θ исключительно в $\boldsymbol{\theta}_0 \in \Theta$,

(b) (Доминирование.) $\mathrm{E}[\sup_{\boldsymbol{\theta}\in\Theta} |m(\boldsymbol{w}_t; \boldsymbol{\theta})|] < \infty$.

Тогда $\widehat{\boldsymbol{\theta}} \to_p \boldsymbol{\theta}_0$.

Описать вторую теорему о состоятельности экстремальных оценок для М-оценок оказывается даже проще. Целевая функция $Q_n(\boldsymbol{\theta})$ вогнута, если $m(\boldsymbol{w}_t; \boldsymbol{\theta})$ является вогнутой по $\boldsymbol{\theta}$[1]. Как сказано выше, точечная сходимость $Q_n(\boldsymbol{\theta})$ к $Q_0(\boldsymbol{\theta})(= E[m(\boldsymbol{w}_t; \boldsymbol{\theta})])$ обеспечивается эргодической теоремой, если $E[m(\boldsymbol{w}_t; \boldsymbol{\theta})]$ существует и ограничена для всех $\boldsymbol{\theta}$. Таким образом,

Утверждение 7.4 (состоятельность М-оценок без требования компактности): *Пусть $\{\boldsymbol{w}_t\}$ — эргодический стационарный случайный процесс. Предположим, что (i) истинный вектор параметров $\boldsymbol{\theta}_0$ является внутренним элементом выпуклого пространства параметров Θ ($\subset \mathbb{R}^p$), (ii) $m(\boldsymbol{w}_t; \boldsymbol{\theta})$ вогнута на пространстве параметров для всех $\boldsymbol{w}_t$ и (iii) $m(\boldsymbol{w}_t; \boldsymbol{\theta})$ измерима по $\boldsymbol{w}_t$ для всех $\boldsymbol{\theta}$ из Θ. Пусть $\widehat{\boldsymbol{\theta}}$ — М-оценка, определенная в (7.1.1) и (7.1.2). Предположим также, что*

(a) (Идентификация.) $\mathrm{E}[m(\boldsymbol{w}_t; \boldsymbol{\theta})]$ принимает максимальное значение на множестве Θ исключительно в $\boldsymbol{\theta}_0 \in \Theta$,

(b) $\mathrm{E}[|m(\boldsymbol{w}_t; \boldsymbol{\theta})|] < \infty$ (то есть $\mathrm{E}[m(\boldsymbol{w}_t; \boldsymbol{\theta})]$ существует и конечно) для всех $\boldsymbol{\theta} \in \Theta$.

Тогда при $n \to \infty$ $\widehat{\boldsymbol{\theta}}$ существует с вероятностью, приближающейся к 1, и $\widehat{\boldsymbol{\theta}} \to_p \boldsymbol{\theta}_0$.

---

[1]Это факт из математического анализа: если $f(\boldsymbol{x})$ и $g(\boldsymbol{x})$ вогнутые функции, то вогнутой функцией будет и их сумма $f(\boldsymbol{x}) + g(\boldsymbol{x})$.

Следующий пример показывает, что оценка условного ML для модели пробит удовлетворяет условию вогнутости.

Пример 7.6 (вогнутость функции правдоподобия для модели пробит): Функция m для модели пробит дается в (7.1.15), где $\Phi(\cdot)$ — функция распределения вероятности для стандартного нормального распределения. Так как нормальное распределение симметрично, $\Phi(-v) = 1 - \Phi(v)$ и функция m может быть переписана как

$$m(\boldsymbol{w}_t; \boldsymbol{\theta}) = y_t \ln \Phi(\boldsymbol{x}_t'\boldsymbol{\theta}) + (1 - y_t) \ln \Phi(-\boldsymbol{x}_t'\boldsymbol{\theta}). \tag{7.2.4}$$

Вогнутость m следует из того, что и $\ln \Phi(\boldsymbol{x}_t'\boldsymbol{\theta})$ и $\ln \Phi(-\boldsymbol{x}_t'\boldsymbol{\theta})$ вогнуты по $\boldsymbol{\theta}$. Так как $\boldsymbol{x}_t'\boldsymbol{\theta}$ линейно по $\boldsymbol{\theta}$, достаточно показать, что $\ln \Phi(v)$ вогнута. Первая производная $\ln \Phi(v)$ равна:

$$\frac{\mathrm{d} \ln \Phi(v)}{\mathrm{d}v} = \frac{\phi(v)}{\Phi(v)}, \tag{7.2.5}$$

где $\phi(v)$ $(= \Phi'(v))$ — функция плотности для $N(0, 1)$. Известно, что $\phi(v)/\Phi(v)$ — убывающая функция. Таким образом, $\ln \Phi(v)$ — (строго) вогнутая функция.

Вогнутость после репараметризации

Утверждение 7.4 может быть легко расширено на оценки, целевые функции которых будут вогнутыми после репараметризации. Предположим, что все условия утверждения, кроме вогнутости, выполнены и что существует непрерывное взаимно однозначное отображение $\boldsymbol{\tau}(\boldsymbol{\theta})\colon \Theta \to \Lambda \equiv \boldsymbol{\tau}(\Theta)$, такое, что

$$\tilde{m}(\boldsymbol{w}_t; \boldsymbol{\lambda}) \equiv m(\boldsymbol{w}_t; \boldsymbol{\tau}^{-1}(\boldsymbol{\lambda}))$$

вогнута по $\boldsymbol{\lambda}$ и $\Lambda = \boldsymbol{\tau}(\Theta)$ — выпуклое множество. Пусть

$$\tilde{Q}_n(\boldsymbol{\lambda}) \equiv \frac{1}{n} \sum_{t=1}^{n} \tilde{m}(\boldsymbol{w}_t; \boldsymbol{\lambda}) -$$

целевая функция после данной репараметризации. Ясно, что все наши предпосылки утверждения 7.4 выполнены для $\boldsymbol{\lambda}$, Λ и $\tilde{m}(\boldsymbol{w}_t; \boldsymbol{\lambda})$: $\boldsymbol{\lambda}_0 \equiv \boldsymbol{\tau}(\boldsymbol{\theta}_0)$ — внутренняя точка Λ, $\mathrm{E}[\tilde{m}(\boldsymbol{w}_t; \boldsymbol{\lambda})]$ принимает максимальное значение исключительно в $\boldsymbol{\lambda}_0$ и $\mathrm{E}[|\tilde{m}(\boldsymbol{w}_t; \boldsymbol{\lambda})|] < \infty$ для всех $\boldsymbol{\lambda}$ в Λ. Тогда для достаточно большого n существует M-оценка $\widehat{\boldsymbol{\lambda}}$, такая, что $\operatorname{plim}_{n\to\infty} \widehat{\boldsymbol{\lambda}} = \boldsymbol{\lambda}_0$. Так как $\tilde{Q}_n(\boldsymbol{\lambda})$ удовлетворяет (7.1.17)для взаимно однозначного отображения $\boldsymbol{\tau}$, $\widehat{\boldsymbol{\theta}} \equiv \boldsymbol{\tau}^{-1}(\widehat{\boldsymbol{\lambda}}))$ максимизирует исходную целевую функцию $Q_n(\boldsymbol{\theta})$. Оценка $\widehat{\boldsymbol{\theta}}$ состоятельная, потому что

$$\begin{aligned}\operatorname*{plim}_{n\to\infty} \widehat{\boldsymbol{\theta}} = &= \operatorname*{plim}_{n\to\infty} \boldsymbol{\tau}^{-1}(\widehat{\boldsymbol{\lambda}}) = \\ &= \boldsymbol{\tau}^{-1}(\operatorname*{plim}_{n\to\infty} \widehat{\boldsymbol{\lambda}}) = \quad \text{(так как } \boldsymbol{\tau}^{-1} \text{ непрерывно)} \\ &= \boldsymbol{\tau}^{-1}(\boldsymbol{\lambda}_0) = \boldsymbol{\theta}_0.\end{aligned} \tag{7.2.6}$$

Пример 7.7 (Вогнутость функции правдоподобия для линейной регрессии): В условном ML-оценивании линейной модели регрессии $m(\boldsymbol{w}_t; \boldsymbol{\theta})$ задана в (7.1.13), где $\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma^2)'$ и $\boldsymbol{w}_t = (y_t, \boldsymbol{x}_t')'$. Эта функция необязательно вогнутая[1], но можно рассмотреть репараметризацию: $\gamma = 1/\sigma$ и $\boldsymbol{\delta} = \boldsymbol{\beta}/\sigma$. Это непрерывное взаимно однозначное отображение между $\Theta = \mathbb{R}^K \times \mathbb{R}_{++}$ и $\Lambda = \mathbb{R}^K \times \mathbb{R}_{++}$. Новое пространство параметров Λ выпукло. После репараметризации, в которой $\boldsymbol{\lambda} = (\boldsymbol{\delta}', \gamma)'$, функция m приобретает вид:

$$\widetilde{m}(\boldsymbol{w}_t; \boldsymbol{\lambda}) = -\frac{1}{2}\ln(2\pi) + \ln(\gamma) - \frac{1}{2}(\gamma y_t - \boldsymbol{x}_t'\boldsymbol{\delta})^2. \tag{7.2.7}$$

Так как $\ln(\gamma)$ вогнут по γ, он вогнут по $\boldsymbol{\lambda}$. Так как $v \equiv \gamma y_t - \boldsymbol{x}_t'\boldsymbol{\delta}$ линейна по $\boldsymbol{\lambda}$ и $-\frac{1}{2}v^2$ вогнута, $-\frac{1}{2}(\gamma y_t - \boldsymbol{x}_t'\boldsymbol{\delta})^2$ вогнута по $\boldsymbol{\lambda}$. Так что функция $\widetilde{m}$, приведенная выше, которая является суммой двух вогнутых функций (плюс константа), вогнута по $\boldsymbol{\lambda}$ для всех $\boldsymbol{w}_t$. Таким образом, учитывая то, что все условия утверждения 7.4, кроме вогнутости, выполняются для $\boldsymbol{\theta}$, Θ и $m(\boldsymbol{w}_t; \boldsymbol{\theta})$, оценка условного ML $\widehat{\boldsymbol{\theta}}$ будет состоятельной.

Идентификация в NLS и ML

Условия идентификации (а) (о том, что $\mathrm{E}[m(\boldsymbol{w}_t; \boldsymbol{\theta})]$ должна принимать максимальное значение исключительно в $\boldsymbol{\theta}_0$) в теоремах о состоятельности M-оценок могут быть сформулированы для NLS и ML более простым образом.

NLS

Рассмотрим сначала NLS, где $m(\boldsymbol{w}_t; \boldsymbol{\theta}) = -(y_t - \varphi(\boldsymbol{x}_t; \boldsymbol{\theta}))^2$ и $\mathrm{E}(y_t|\boldsymbol{x}_t) = \varphi(\boldsymbol{x}_t; \boldsymbol{\theta}_0)$. Мы показали в параграфе 2.9, что условное математическое ожидание минимизирует среднеквадратическую ошибку, то есть для лю-

[1] Например, вторая частная производная m по σ^2 может быть или не быть неположительной в зависимости от значения $y_t - \boldsymbol{x}_t'\boldsymbol{\beta}$. Вспомним, что необходимым условием вогнутости для дважды дифференцируемой функции $h(x_1, \ldots, x_p)$ является неположительность $\partial^2 h/(\partial x_j)$ для всех $j = 1, 2, \ldots, p$.

бой (измеримой) функции $h(\boldsymbol{x}_t)$

$$\begin{aligned}\text{среднеквадратическая ошибка } &\equiv \mathrm{E}\big[\{y_t - h(\boldsymbol{x}_t)\}^2\big] = \\ &= \mathrm{E}\big[\{y_t - \mathrm{E}(y_t|\boldsymbol{x}_t)\}^2\big] + \\ &+ \mathrm{E}\big[\{\mathrm{E}(y_t|\boldsymbol{x}_t) - h(\boldsymbol{x}_t)\}^2\big] \geq \\ &\geq \mathrm{E}\big[\{y_t - \mathrm{E}(y_t|\boldsymbol{x}_t)\}^2\big]. \qquad (7.2.8)\end{aligned}$$

Здесь мы всего лишь воспроизвели (2.9.4) в текущих обозначениях. Более того, условное математическое ожидание является по существу единственным решением задачи минимизации в следующем смысле.

Будучи функциями от $\boldsymbol{x}_t$, $h(\boldsymbol{x}_t)$ и $\mathrm{E}(y_t|\boldsymbol{x}_t)$, являются случайными величинами. Для двух случайных величин X и Y пусть $X \neq Y$ означает $\mathrm{Prob}(X \neq Y) > 0$ (то есть $\mathrm{Prob}(X = Y) < 1$), а $X = Y$ означает $\mathrm{Prob}(X \neq Y) = 0$. (Если X и Y отличаются друг от друга только для тех событий, вероятность которых равна нулю, нам не надо заботиться об этом отличии.) Таким образом, $h(\boldsymbol{x}_t) \neq \mathrm{E}(y_t|\boldsymbol{x}_t)$ означает, что $\mathrm{Prob}(h(\boldsymbol{x}_t) \neq \mathrm{E}(y_t|\boldsymbol{x}_t)) > 0$. Если $h(\boldsymbol{x}_t) \neq \mathrm{E}(y_t|\boldsymbol{x}_t)$ в этом смысле, то $\{\mathrm{E}(y_t|\boldsymbol{x}_t) - h(\boldsymbol{x}_t)\}^2 > 0$ с положительной вероятностью. Следовательно, $\mathrm{E}\big[\{\mathrm{E}(y_t|\boldsymbol{x}_t) - h(\boldsymbol{x}_t)\}^2\big]$ в (7.2.8) положительное и среднеквадратическая ошибка строго больше, чем $\mathrm{E}\big[\{y_t - \mathrm{E}(y_t|\boldsymbol{x}_t)\}^2\big]$.

Просто задав функцию $h(\boldsymbol{x}_t) = \varphi(\boldsymbol{x}_t;\boldsymbol{\theta})$ и заметив, что $\varphi(\boldsymbol{x}_t;\boldsymbol{\theta}_0) = \mathrm{E}(y_t|\boldsymbol{x}_t)$, мы можем сделать вывод, что

$$\begin{cases}\mathrm{E}\big[\{y_t - \varphi(\boldsymbol{x}_t;\boldsymbol{\theta})\}^2\big] > \mathrm{E}\big[\{y_t - \varphi(\boldsymbol{x}_t;\boldsymbol{\theta}_0)\}^2\big], & \text{если } \varphi(\boldsymbol{x}_t;\boldsymbol{\theta}) \neq \varphi(\boldsymbol{x}_t;\boldsymbol{\theta}_0), \\ \mathrm{E}\big[\{y_t - \varphi(\boldsymbol{x}_t;\boldsymbol{\theta})\}^2\big] = \mathrm{E}\big[\{y_t - \varphi(\boldsymbol{x}_t;\boldsymbol{\theta}_0)\}^2\big], & \text{если } \varphi(\boldsymbol{x}_t;\boldsymbol{\theta}) = \varphi(\boldsymbol{x}_t;\boldsymbol{\theta}_0).\end{cases} \qquad (7.2.9)$$

Таким образом, условие идентификации — что $\mathrm{E}[m(\boldsymbol{w}_t;\boldsymbol{\theta})] = -\mathrm{E}\big[\{y_t - \varphi(\boldsymbol{x}_t;\boldsymbol{\theta})\}^2\big]$ максимизируется (то есть $\mathrm{E}\big[\{y_t - \varphi(\boldsymbol{x}_t;\boldsymbol{\theta})\}^2\big]$ минимизируется) исключительно в $\boldsymbol{\theta}_0$ — выполняется, если

«идентификация условного математического ожидания:»

(*conditional mean identification*)

$$\varphi(\boldsymbol{x}_t;\boldsymbol{\theta}) \neq \varphi(\boldsymbol{x}_t;\boldsymbol{\theta}_0) \quad \text{для всех} \quad \boldsymbol{\theta} \neq \boldsymbol{\theta}_0. \qquad (7.2.10)$$

ML

В случае ML роль, которую играло выражение (7.2.8) для NLS, будет играть **информационное неравенство Кульбака — Лейблера** (*Kullback — Leibler information inequality*). Здесь мы исследуем идентификацию для условного ML (сделать то же самое для безусловного или совместного ML проще). Гипотетическая функция условной плотности $f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta})$, будучи функцией от $(y_t, \boldsymbol{x}_t)$, является случайной величиной для любого значения $\boldsymbol{\theta}$. Математическое ожидание случайной величины $\ln f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta})$ может быть записано как[1]

$$\mathrm{E}[\ln f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta})] = \int \ln f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}) f(y_t,\boldsymbol{x}_t;\boldsymbol{\theta}_0,\boldsymbol{\psi}_0)\mathrm{d}y_t\mathrm{d}\boldsymbol{x}_t. \tag{7.2.11}$$

Обратите особое внимание на то, что математическое ожидание здесь берется по отношению к *истинной* совместной функции плотности $f(y_t,\boldsymbol{x}_t;\boldsymbol{\theta}_0,\boldsymbol{\psi}_0)$. Информационное неравенство Кульбака — Лейблера (адаптированное к условным плотностям) утверждает, что

$$\begin{cases} \mathrm{E}[\ln f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta})] < \mathrm{E}[\ln f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}_0)], & \text{если} \quad f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}) \neq f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}_0), \\ \mathrm{E}[\ln f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta})] = \mathrm{E}[\ln f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}_0)], & \text{если} \quad f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}) = f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}_0). \end{cases} \tag{7.2.12}$$

(Конечно, при условии того, что $\mathrm{E}[\ln f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta})]$ существует и что оно конечное). (По поводу вывода этого неравенства см. аналитическое упражнение 1.) Отсюда немедленно следует, что условие идентификации (что $\mathrm{E}[\ln f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta})]$ максимизируется исключительно в $\boldsymbol{\theta}_0$) выполнено, если

«идентификация условной функции плотности»

(conditional density identification) :

$$f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}) \neq f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}_0) \quad \text{для всех} \quad \boldsymbol{\theta} \neq \boldsymbol{\theta}_0. \tag{7.2.13}$$

Мы можем превратить теоремы о состоятельности M-оценок (утверждения 7.3 и 7.4) в теоремы о состоятельности оценок условного ML с $\boldsymbol{w}_t = (y_t, \boldsymbol{x}_t')'$ и $m(\boldsymbol{w}_t;\boldsymbol{\theta}) = \ln f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta})$. Заменив условия идентификации для M-оценок на условие идентификации условной функции плотности, мы получим два следующих самодостаточных утверждения:

Утверждение 7.5 (Состоятельность оценки условного ML с компактным пространством параметров): *Пусть $\{y_t, \boldsymbol{x}_t\}$ — эргодический стационарный случайный процесс с условной функцией плотности $f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}_0)$ и пусть $\widehat{\boldsymbol{\theta}}$ — оценка условного ML, которая максимизирует среднее значение логарифма условной функции правдоподобия (выведенной в предположении, что $\{y_t, \boldsymbol{x}_t\}$ i.i.d.):*

$$\widehat{\boldsymbol{\theta}} = \underset{\theta\in\Theta}{\operatorname{argmax}} \frac{1}{n}\sum_{t=1}^{n} \ln f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}).$$

[1]Если $f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}) = 0$, то тогда нельзя определить логарифм этой функции. Если вас это беспокоит, предположите, что $f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta})$ положительно для всех y_t, $\boldsymbol{x}_t$ и $\boldsymbol{\theta}$. См.: [White, 1994, p. 9] относительно доказательства, которое избегает такой упрощающей предпосылки.

Предположим, что модель правильно специфицирована, так что $\boldsymbol{\theta}_0$ принадлежит Θ. Предположим, что (i) пространство параметров Θ — компактное подмножество $\mathbb{R}^p$, (ii) $f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta})$ непрерывна по $\boldsymbol{\theta}$ для всех $(y_t, \boldsymbol{x}_t)$ и (iii) $f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta})$ измерима по $(y_t, \boldsymbol{x}_t)$ для всех $\boldsymbol{\theta} \in \Theta$ (так что по лемме 7.1 $\widehat{\boldsymbol{\theta}}$ определена как случайная величина). Предположим также, что

(a) *(Идентификация.)* $\text{Prob}[f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}) \neq f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}_0)] > 0$ *для всех* $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ *в* Θ,

(b) *(Доминирование.)* $\text{E}[\sup_{\boldsymbol{\theta}\in\Theta} |\ln f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta})|] < \infty$ *(заметьте: математическое ожидание берется по* y_t *и по* $\boldsymbol{x}_t$*).*

Тогда $\widehat{\boldsymbol{\theta}} \to_p \boldsymbol{\theta}_0$.

Утверждение 7.6 (состоятельность оценки условного ML без компактности): *Пусть* $\{y_t, \boldsymbol{x}_t\}$ *эргодический стационарный с условной функцией плотности* $f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}_0)$ *и пусть* $\widehat{\boldsymbol{\theta}}$ *— оценка условного ML, которая максимизирует среднее значение логарифма условной функции правдоподобия (выведенной в предпосылках, что* $\{y_t, \boldsymbol{x}_t\}$ *i.i.d.):*

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\Theta}{\text{argmax}} \frac{1}{n}\sum_{t=1}^{n} \ln f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}).$$

Предположим, что модель правильно специфицирована, так что $\boldsymbol{\theta}_0$ принадлежит Θ. Предположим, что (i) вектор истинных параметров $\boldsymbol{\theta}_0$ является внутренним элементом выпуклого множества параметров Θ ($\subset \mathbb{R}^p$), (ii) $\ln f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta})$ вогнута по $\boldsymbol{\theta}$ для всех $(y_t, \boldsymbol{x}_t)$ и (iii) $\ln f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta})$ измерима по $(y_t, \boldsymbol{x}_t)$ для всех $\boldsymbol{\theta}$ из Θ. (Для достаточно большого n $\widehat{\boldsymbol{\theta}}$ однозначно определена; см. несколькими строками ниже.) Предположим, также, что

(a) *(Идентификация.)* $\text{Prob}[f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}) \neq f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}_0)] > 0$ *для всех* $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ *из* Θ,

(b) $\text{E}[|\ln f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta})|] < \infty$ *(то есть* $\text{E}[\ln f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta})]$ *существует и конечно) для всех* $\boldsymbol{\theta} \in \Theta$ *(заметьте: математическое ожидание берется по* y_t *и по* $\boldsymbol{x}_t$*).*

Тогда, при $n \to \infty$, $\widehat{\boldsymbol{\theta}}$ *существует с вероятностью, стремящейся к 1, и* $\widehat{\boldsymbol{\theta}} \to_p \boldsymbol{\theta}_0$.

Следует сделать два замечания:

- Информационное неравенство Кульбака — Лейблера (7.2.12) выполняется также для безусловных функций плотности (доказательство для безусловных плотностей проще). Таким образом, утверждения 7.5 и 7.6 верны и для безусловного ML; надо просто заменить $f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta})$ на $f(\boldsymbol{w}_t;\boldsymbol{\theta})$.

- Важной особенностью этих двух теорем о состоятельности оценки ML является то, что, *несмотря* на то что логарифм функции правдоподобия выводится при предположении о независимости и одинаковом распределении $\{y_t, \boldsymbol{x}_t\}$, оценки будут состоятельными даже в случае, когда $\{y_t, \boldsymbol{x}_t\}$ не i.i.d., а просто является стационарным и эргодическим процессом. Это так, потому что оценка ML является М-оценкой. Оценка ML, которая максимизирует функцию правдоподобия, отличную от функции правдоподобия модели, называется **оценкой квази-ML** (*quasi-ML estimator*), или **оценкой QML** (*QML estimator*). Другими словами, оценка QML максимизирует функцию правдоподобия в неверно специфицированной модели. Таким образом, оценки метода «максимального правдоподобия» в утверждениях 7.5 и 7.6 будут оценками QML, если $\{y_t, \boldsymbol{x}_t\}$ стационарный и эргодический, но не i.i.d. Эти теоремы говорят о том, что данные конкретные оценки QML состоятельны.

В частных случаях условие идентификации для условного ML может быть записано в еще более простом виде.

Пример 7.8 (идентификация в модели линейной регрессии): В случае линейной регрессии с нормальными ошибками логарифм условной функции правдоподобия для наблюдения t равен:

$$\ln f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}(y_t - \boldsymbol{x}_t'\boldsymbol{\beta})^2.$$

Так как логарифм — строго монотонная функция, идентификация условной плотности эквивалентна условию о том, что

$$\ln f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}) \neq \ln f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}_0)$$

с положительной вероятностью, если $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. Это условие выполняется, если $\mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}_t')$ невырождена (и, следовательно, положительно определена). Чтобы понять, почему это так, обозначим $\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma^2)'$ и $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0', \sigma_0^2)'$. Ясно, что $\ln f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}) \neq \ln f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}_0)$ с положительной вероятностью, если $\sigma^2 \neq \sigma_0^2$. Так что рассмотрим случай, когда $\sigma^2 = \sigma_0^2$, но $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$. Тогда

$$\mathrm{E}[\{\boldsymbol{x}_t'\boldsymbol{\beta} - \boldsymbol{x}_t'\boldsymbol{\beta}_0\}^2] = \mathrm{E}[\{\boldsymbol{x}_t'(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\}^2] = (\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\,\mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}_t')(\boldsymbol{\beta} - \boldsymbol{\beta}_0) > 0.$$

Таким образом, $x_t'\beta \neq x_t'\beta_0$ с положительной вероятностью; если $x_t'\beta = x_t'\beta_0$ с вероятностью 1 (то есть если $x_t'\beta \neq x_t'\beta_0$ с нулевой вероятностью), то тогда $\mathrm{E}[\{x_t'\beta - x_t'\beta_0\}^2]$ будет равняться нулю. Значит, $y_t - x_t'\beta \neq y_t - x_t'\beta_0$ с положительной вероятностью, из чего следует, что условие идентификации выполнено. Из того же условия относительно $\mathrm{E}(x_t x_t')$ следует также и условие (b) в утверждении 7.6, потому что

$$\begin{aligned}\mathrm{E}[(y_t - x_t'\beta)^2] &= \mathrm{E}[\{\varepsilon_t + x_t'(\beta_0 - \beta)\}^2] \text{ (поскольку } y_t = x_t'\beta_0 + \varepsilon_t) = \\ &= \mathrm{E}(\varepsilon_t^2) + (\beta_0 - \beta)'\,\mathrm{E}(x_t x_t')(\beta_0 - \beta),\\ &\text{так как } \mathrm{E}(\varepsilon_t \cdot x_t) = \mathbf{0}.\end{aligned}$$

Последнее слагаемое будет конечным, если конечно $\mathrm{E}(x_t x_t')$. Из этих результатов и отмеченного ранее факта, что логарифм функции правдоподобия вогнут после репараметризации, следует, что оценка условного ML-модели линейной регрессии с нормальными ошибками (выведенная в предположении, что $\{y_t, x_t\}$ i.i.d.) состоятельна, если $\{y_t, x_t\}$ эргодический и стационарный и если $\mathrm{E}(x_t x_t')$ невырождена.

Пример 7.9 (идентификация в моделях пробит): Тот же результат, который был выведен в предыдущем примере для линейной модели, также верен и в случае модели пробит, для которой условная функция правдоподобия для наблюдения t равна

$$f(y_t|x_t;\theta) = \Phi(x_t'\theta)^{y_t}\Phi(-x_t'\theta)^{1-y_t}.$$

Та же самая аргументация, которая использовалась в примере 7.8, показывает, что, при условии невырожденности $\mathrm{E}(x_t x_t')$, $x_t'\theta \neq x_t'\theta_0$ с положительной вероятностью, если $\theta \neq \theta_0$. Но так как $\Phi(v)$ строго монотонна, из $x_t'\theta \neq x_t'\theta_0$ с положительной вероятностью следует, что $\Phi(x_t'\theta) \neq \Phi(x_t'\theta_0)$ и $\Phi(-x_t'\theta) \neq \Phi(-x_t'\theta_0)$ с положительной вероятностью. Значит, условие идентификации (a) в утверждении 7.6 выполняется, если $\mathrm{E}(x_t x_t')$ невырождена.

Из невырожденности $\mathrm{E}(x_t x_t')$ следует также и то, что условие (b) утверждения 7.6 выполняется для модели пробит. Очень просто проверить, что

$$|\ln\Phi(v)| \leq |\ln\Phi(0)| + |v| + |v|^2 \quad \text{для всех} \quad v. \tag{7.2.14}$$

Комбинируя это ограничение на $|\ln\Phi(v)|$ с тем фактом, что y_t и $1 - y_t$ по абсолютной величине меньше или равны единице, легко показать, что $\mathrm{E}[|\ln f(y_t|x_t;\theta)|] < \infty$, если $\mathrm{E}(x_t x_t')$ существует и конечно (условие выполняется из-за невырожденности $\mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}_t'))$[1]. Мы, таким образом, делаем вывод, что оценка модели пробит ML (логарифм функции правдоподобия для которой выводился в предположении i.i.d. $\{y_t, \boldsymbol{x}_t\}$) состоятельна, если $\{y_t, \boldsymbol{x}_t\}$ стационарный и эргодический, и $\mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}_t')$ невырождена.

Состоятельность GMM

Перейдем теперь к GMM и адаптируем утверждение 7.1 к GMM, проверяя выполнение условий этого утверждения при выполнении набора соответствующих достаточных условий. Целевая функция GMM задается (7.1.13). Ясно, что непрерывность $Q_0(\boldsymbol{\theta})$ по $\boldsymbol{\theta}$ будет обеспечена, если $\boldsymbol{g}(\boldsymbol{w}_t; \boldsymbol{\theta})$ непрерывна по $\boldsymbol{\theta}$ для всех $\boldsymbol{w}_t$. Если $\{\boldsymbol{w}_t\}$ стационарный и эргодический, то $\boldsymbol{g}_n(\boldsymbol{\theta})(\equiv \sum_{t=1}^{n} \boldsymbol{g}(\boldsymbol{w}_t; \boldsymbol{\theta})) \to_p \mathrm{E}[\boldsymbol{g}(\boldsymbol{w}_t; \boldsymbol{\theta})]$. Следовательно, предельная функция $Q_0(\boldsymbol{\theta})$ имеет вид:

$$Q_0(\boldsymbol{\theta}) = -\frac{1}{2}\,\mathrm{E}[\boldsymbol{g}(\boldsymbol{w}_t; \boldsymbol{\theta})]'\,\boldsymbol{W}\,\mathrm{E}[\boldsymbol{g}(\boldsymbol{w}_t; \boldsymbol{\theta})]. \tag{7.2.15}$$

Эта функция неположительна, если $\boldsymbol{W}(\equiv \operatorname{plim} \widehat{\boldsymbol{W}})$ положительно определена. Функция достигает максимального значения, равного нулю, в точке $\boldsymbol{\theta}_0$, потому что $\mathrm{E}[\boldsymbol{g}(\boldsymbol{w}_t; \boldsymbol{\theta}_0)] = \mathbf{0}$ по условиям ортогональности. Таким образом, условие идентификации (что $Q_0(\boldsymbol{\theta})$ принимает максимальное значение исключительно в $\boldsymbol{\theta}_0$) утверждения 7.1 будет выполняться, если $\mathrm{E}[\boldsymbol{g}(\boldsymbol{w}_t; \boldsymbol{\theta})] \neq \mathbf{0}$ для $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. Учитывая равномерную сходимость $Q_n(\boldsymbol{\theta})$ к $Q_0(\boldsymbol{\theta})$, несложно показать, что это условие будет выполнено, если $\boldsymbol{g}_n(\cdot)$ сходится равномерно к $\mathrm{E}[\boldsymbol{g}(\boldsymbol{w}_t; \cdot)]$ (см.: [Newey and McFadden, 1994, p. 2132]. Многомерная версия равномерного закона больших чисел, приведенная прямо под леммой 7.2, в свою очередь, дает достаточное условие равномерной сходимости. Так что мы доказали

Утверждение 7.7 (состоятельность GMM с компактным пространством параметров): *Пусть $\{\boldsymbol{w}_t\}$ — стационарный и эргодический и пусть $\widehat{\boldsymbol{\theta}}$ —*

---

[1] Только для заинтересованных читателей здесь приведено доказательство:

$$\begin{aligned}
|\ln f(y_t|\boldsymbol{x}_t; \boldsymbol{\theta})| &\leqslant |y_t||\ln\Phi(\boldsymbol{x}_t'\boldsymbol{\theta})| + |1 - y_t||\ln\Phi(-\boldsymbol{x}_t'\boldsymbol{\theta})| \leqslant \\
\leqslant |\ln\Phi(\boldsymbol{x}_t'\boldsymbol{\theta})| &+ |\ln\Phi(-\boldsymbol{x}_t'\boldsymbol{\theta})| \ (\text{так как } |y_t| \leqslant 1 \text{ и } |1 - y_t| \leqslant 1) \leqslant \\
&\leqslant 2 \times [|\ln\Phi(0)| + |\boldsymbol{x}_t'\boldsymbol{\theta}| + |\boldsymbol{x}_t'\boldsymbol{\theta}|^2] \ (\text{по (7.2.14)}) \leqslant \\
&\leqslant 2 \times [|\ln\Phi(0)| + ||\boldsymbol{x}_t|| \cdot ||\boldsymbol{\theta}|| + ||\boldsymbol{x}_t||^2 \cdot ||\boldsymbol{\theta}||^2].
\end{aligned}$$

Последнее неравенство следует из неравенства Коши — Шварца, утверждающего, что $|\boldsymbol{x}'\boldsymbol{y}| \leqslant ||\boldsymbol{x}||||\boldsymbol{y}||$ для двух любых векторов $\boldsymbol{x}$ и $\boldsymbol{y}$ соответствующей размерности. Кроме того, $\mathrm{E}(||\boldsymbol{x}_t||^2)$, являющееся суммой $\mathrm{E}(x_{it}^2)$ по i, конечно, потому что из невырожденности $\mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}_t')$ следует, что $\mathrm{E}(x_{it}^2) < \infty$ для всех i. Если $\mathrm{E}(||\boldsymbol{x}_t||^2) < \infty$, то и $\mathrm{E}(||\boldsymbol{x}_t||) < \infty$.

GMM-оценка, определенная как

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmin}} \left[\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{g}(\boldsymbol{w}_t;\boldsymbol{\theta})\right]' \widehat{\boldsymbol{W}} \left[\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{g}(\boldsymbol{w}_t;\boldsymbol{\theta})\right],$$

где симметричная матрица $\widehat{\boldsymbol{W}}$ по предположению сходится по вероятности к некоторой симметричной положительно определенной матрице $\boldsymbol{W}$. Предположим, что модель специфицирована правильно, в том смысле, что $\mathrm{E}[\boldsymbol{g}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)] = \mathbf{0}$ (то есть выполняются условия ортогональности) для $\boldsymbol{\theta}_0 \in \Theta$. Предположим, что (i) пространство параметров Θ — компактное подмножество $\mathbb{R}^p$, (ii) $\boldsymbol{g}(\boldsymbol{w}_t;\boldsymbol{\theta})$ непрерывна по $\boldsymbol{\theta}$ для всех $\boldsymbol{w}_t$ и (iii) $\boldsymbol{g}(\boldsymbol{w}_t;\boldsymbol{\theta})$ измерима по $\boldsymbol{w}_t$ для всех $\boldsymbol{\theta}$ из Θ (так что $\widehat{\boldsymbol{\theta}}$ — корректно определенная случайная величина по лемме 7.1). Предположим также, что

(a) (Идентификация.) $\mathrm{E}[\boldsymbol{g}(\boldsymbol{w}_t;\boldsymbol{\theta})] \neq \mathbf{0}$ *для всех* $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ *из* Θ,

(b) (Доминирование.) $\mathrm{E}[\sup_{\boldsymbol{\theta}\in\Theta} \|\boldsymbol{g}(\boldsymbol{w}_t;\boldsymbol{\theta})\|\,] < \infty$.

Тогда $\widehat{\boldsymbol{\theta}} \to_p \boldsymbol{\theta}_0$.

Если $Q_n(\boldsymbol{\theta})$ вогнута, то требования компактности Θ, непрерывности $\boldsymbol{g}(\boldsymbol{w}_t;\boldsymbol{\theta})$ и условие доминирования могут быть заменены на условие, что $\mathrm{E}[\boldsymbol{g}(\boldsymbol{w}_t;\boldsymbol{\theta})]$ существует и конечно для всех $\boldsymbol{\theta}$. Мы не будем формулировать этот результат в качестве теоремы, потому что существует очень мало прикладных задач, в которых $Q_n(\boldsymbol{\theta})$ вогнута. Например, целевая функция GMM для уравнения Эйлера в примере 7.5 не является вогнутой. Чуть ли не единственным известным случаем является тот, в котором $\boldsymbol{g}(\boldsymbol{w}_t;\boldsymbol{\theta})$ линейна по $\boldsymbol{\theta}$, но этот линейный случай был довольно подробно рассмотрен в главе 3. Когда $\boldsymbol{g}(\boldsymbol{w}_t;\boldsymbol{\theta})$ нелинейна по $\boldsymbol{\theta}$, специфицировать простые условия для условий (a) (идентификация) и (b) (доминирование) в утверждении 7.7 в общем случае довольно сложно. Так что в большинстве прикладных задач выполнение этих условий просто предполагается.

Контрольные вопросы

1. (Необходимость идентификации функции плотности.) Мы показали, для условного ML, что идентификации условной функции плотности (7.2.13) достаточно для выполнения условия (a) (что $\mathrm{E}(\ln f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}))$ принимает наибольшее значение исключительно в $\boldsymbol{\theta}_0$). То, что это является и необходимым условием, ясно из информационного неравенства Кульбака — Лейблера. Не используя это неравенство, покажите необходимость идентификации условной плотности. **Указание:** *Покажите, что если бы (7.1.13) не выполнялось, то (a) в этом случае не выполнялось бы. Предположите, что существует $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_0$, такое, что $\boldsymbol{\theta}_1 \in \Theta$ и $f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}_1) = f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}_0)$. Тогда* $\mathrm{E}(\ln f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}_1)) = \mathrm{E}(\ln f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}_0))$.

2. (Необходимость идентификации условного математического ожидания.) Не используя (7.2.8), покажите, что условие идентификации условного математического ожидания является как необходимым, так и достаточным для идентификации в NLS.

3. (Идентификация в NLS.) Предположим, что функция φ в NLS линейна: $\varphi(\boldsymbol{x}_t; \boldsymbol{\theta}) = \boldsymbol{x}_t'\boldsymbol{\theta}$. Проверьте, что условие идентификации выполняется, если $\mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}_t')$ невырождена.

4. (Идентификация в NLS-оценивании модели пробит.) В модели пробит $\mathrm{E}(y_t|\boldsymbol{x}_t) = \Phi(\boldsymbol{x}_t'\boldsymbol{\theta}_0)$. Пусть $\boldsymbol{\theta}_0$ оценивается NLS. Проверьте, что условие идентификации условного математического ожидания выполняется, если $\mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}_t')$ невырождена.

5. (Квази-ML.) Предположим, что в утверждении 7.5 $\{y_t, \boldsymbol{x}_t\}$ i.i.d, так что функция правдоподобия сама по себе специфицирована правильно. Однако предположим, что истинное значение вектора параметров $\boldsymbol{\theta}_0$ *не* включено в предполагаемое пространство параметров Θ. Так что в этом смысле оценка $\widehat{\boldsymbol{\theta}}$ будет являться оценкой квази-ML. Опустив условие (a), предположите вместо него, что существует единственное значение $\boldsymbol{\theta}^*$ из Θ, которое максимизирует $\mathrm{E}[\ln f(y_t|\boldsymbol{x}_t; \boldsymbol{\theta})]$. Покажите, что оценка квази-ML $\widehat{\boldsymbol{\theta}}$ будет состоятельной оценкой для $\boldsymbol{\theta}^*$. **Указание:** *Замените $\boldsymbol{\theta}_0$ в утверждении 7.3 на $\boldsymbol{\theta}^*$. В утверждениях 7.1–7.4 $\boldsymbol{\theta}_0$ не обязан быть вектором «истинных» значений параметров.*

6. (Идентификация в линейном GMM.) Рассмотрим модель линейного GMM из главы 3. Проверьте, что ранговое условие идентификации, выведенное в главе 3, эквивалентно условию идентификации в утверждении 7.7.

7. (Вогнутость целевой функции в случае линейного GMM.) Проверьте, что целевая функция GMM в линейном GMM вогнута $\boldsymbol{\theta}$. **Указание:** *Гессиан (матрица вторых производных) функции $Q_n(\boldsymbol{\theta})$ равен $-\boldsymbol{S}_{xz}'\widehat{\boldsymbol{W}}\boldsymbol{S}_{xz}$, где $\boldsymbol{S}_{xz} \equiv \sum_{t=1}^{n} \boldsymbol{x}_t\boldsymbol{z}_t'$. Дважды непрерывно дифференцируемая функция вогнута, если и только если ее гессиан отрицательно полуопределен во всех точках.*

8. (Идентификация GMM при вырожденной $\boldsymbol{W}$.) Если бы мы требовали, что $\boldsymbol{W}$ должна быть хотя бы положительно полуопределена, а не обязательно строго положительно определена, то каким бы было условие идентификации в утверждении 7.7? **Указание:** *Предположите, что $\boldsymbol{W}$ положительно полуопределена. Покажите, что если $\mathrm{E}[\boldsymbol{g}(\boldsymbol{w}_t; \boldsymbol{\theta}_0)] = \boldsymbol{0}$ и $\boldsymbol{W}\,\mathrm{E}[\boldsymbol{g}(\boldsymbol{w}_t; \boldsymbol{\theta})] \neq \boldsymbol{0}$ для $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, то $Q_0(\boldsymbol{\theta}) = -\frac{1}{2}\mathrm{E}[\boldsymbol{g}(\boldsymbol{w}_t; \boldsymbol{\theta})]'\boldsymbol{W}\,\mathrm{E}[\boldsymbol{g}(\boldsymbol{w}_t; \boldsymbol{\theta})]$ имеет единственную точку максимума в $\boldsymbol{\theta}_0$.*

7.3. Асимптотическая нормальность

Базовым инструментом для вывода асимптотического распределения экстремальной оценки является теорема о среднем значении из математического анализа. Как она используется, зависит от того, является ли рассматриваемая оценка M-оценкой или оценкой GMM. Таким образом, доказательство асимптотической нормальности будет приведено отдельно для M-оценок (включая оценку ML как частного случая M-оценок) и для оценок GMM. За этим последуют короткие замечания об относительной эффективности GMM и ML. Подводя итоги, мы покажем в конце параграфа, что разложение в ряд Тейлора для отклонения выборочной оценки от истинного параметра имеет общую структуру для M-оценки и GMM.

Асимптотическая нормальность M-оценок

Приведенная в (7.1.2) целевая функция для M-оценки имеет вид:

$$Q_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n} m(\boldsymbol{w}_t;\boldsymbol{\theta}). \tag{7.3.1}$$

Удобно обозначить градиент (вектор первых производных) и гессиан (матрицу вторых производных) функции m как

$$\underset{(p\times 1)}{\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta})} = \frac{\partial m(\boldsymbol{w}_t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}. \tag{7.3.2}$$

$$\underset{(p\times p)}{\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta})} = \frac{\partial \boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \frac{\partial^2 m(\boldsymbol{w}_t;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}. \tag{7.3.3}$$

По аналогии с ML $\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta})$ будет называться **вектором вклада для наблюдения t** (*score vector for observation t*). Этот $\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta})$ не следует путать с вкладом в привычном понимании, который является градиентом целевой функции $Q_n(\boldsymbol{\theta})$. Функцию вклада во втором случае мы будем обозначать далее в этой главе как $s_n(\boldsymbol{\theta})$. То же самое относится к гессиану: $\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta})$ будет рассматриваться как **гессиан для наблюдения t** (*Hessian for observation t*), а гессиан $Q_n(\boldsymbol{\theta})$ будет обозначаться $\boldsymbol{H}_n(\boldsymbol{\theta})$.

Целью этого параграфа является асимптотическая нормальность $\widehat{\boldsymbol{\theta}}$, описанная ниже в (7.3.10). В процессе вывода этого свойства мы сделаем несколько предположений, которые будут собраны ниже в утверждении 7.8. Предположим, что $m(\boldsymbol{w}_t;\boldsymbol{\theta})$ дифференцируема по $\boldsymbol{\theta}$ и что $\widehat{\boldsymbol{\theta}}$ — внутренняя точка Θ. Тогда $\widehat{\boldsymbol{\theta}}$, будучи внутренним решением задачи максимизации $Q_n(\boldsymbol{\theta})$, удовлетворяет условиям первого порядка

$$\begin{aligned}\underset{(p\times 1)}{\mathbf{0}} &= \frac{\partial Q_n(\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \\ &= \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{s}(\boldsymbol{w}_t;\widehat{\boldsymbol{\theta}}) \quad \text{(по (7.3.1) и (7.3.2))}.\end{aligned} \tag{7.3.4}$$

Сейчас мы будем использовать следующий результат из математического анализа:

Теорема о среднем значении: *Пусть* $\boldsymbol{h} : \mathbb{R}^p \to \mathbb{R}^q$ *непрерывно дифференцируемо. Тогда* $\boldsymbol{h}(\boldsymbol{x})$ *можно представить в виде* ***разложения в среднем значении*** *(mean value expansion):*

$$\underset{(q\times 1)}{\boldsymbol{h}(\boldsymbol{x})} = \underset{(q\times 1)}{\boldsymbol{h}(\boldsymbol{x}_0)} + \underset{(q\times p)}{\frac{\partial \boldsymbol{h}(\overline{\boldsymbol{x}})}{\partial \boldsymbol{x}'}} \underset{(p\times 1)}{(\boldsymbol{x} - \boldsymbol{x}_0)}.$$

где $\overline{\boldsymbol{x}}$ *— среднее значение, лежащее между* $\boldsymbol{x}$ *и* $\boldsymbol{x}_0$[1].

Полагая $q = p$, $\boldsymbol{x} = \widehat{\boldsymbol{\theta}}$, $\boldsymbol{x}_0 = \boldsymbol{\theta}_0$ и $\boldsymbol{h}(\cdot) = \frac{\partial Q_n(\cdot)}{\partial \boldsymbol{\theta}}$ в теореме о среднем значении, мы получим следующее разложение:

$$\begin{aligned}
\underset{(p\times 1)}{\frac{\partial Q_n(\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}} &= \underset{(p\times 1)}{\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}} + \underset{(p\times p)}{\frac{\partial^2 Q_n(\overline{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}} \underset{(p\times 1)}{(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)} = \\
&= \underset{(p\times 1)}{\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{s}(\boldsymbol{w}_t; \boldsymbol{\theta}_0)} + \underset{(p\times p)}{\left[\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{H}(\boldsymbol{w}_t; \overline{\boldsymbol{\theta}})\right]} \underset{(p\times 1)}{(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)} \qquad (7.3.5)\\
&\quad (\text{по } (7.3.1)\text{–}(7.3.3)),
\end{aligned}$$

где $\overline{\boldsymbol{\theta}}$ — среднее значение, лежащее между $\widehat{\boldsymbol{\theta}}$ и $\boldsymbol{\theta}_0$. Требование теоремы о среднем значении относительно непрерывной дифференцируемости будет выполнено, если $m(\boldsymbol{w}_t; \boldsymbol{\theta})$ *дважды* непрерывно дифференцируема по $\boldsymbol{\theta}$. Объединив это уравнение с вышеупомянутым условием первого порядка, мы получим:

$$\underset{(p\times 1)}{\mathbf{0}} = \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{s}(\boldsymbol{w}_t; \boldsymbol{\theta}_0) + \left[\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{H}(\boldsymbol{w}_t; \overline{\boldsymbol{\theta}})\right](\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0). \qquad (7.3.6)$$

Предполагая, что $\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{H}(\boldsymbol{w}_t; \overline{\boldsymbol{\theta}})$ невырождена, это уравнение может быть решено относительно $(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$, что даст:

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -\left[\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{H}(\boldsymbol{w}_t; \overline{\boldsymbol{\theta}})\right]^{-1} \frac{1}{\sqrt{n}}\sum_{t=1}^{n} \boldsymbol{s}(\boldsymbol{w}_t; \boldsymbol{\theta}_0). \qquad (7.3.7)$$

Это выражение для (умноженной на $\sqrt{n}$) ошибки оценки будет называться **разложением в среднем значении ошибки оценки** (*mean value expansion for the sampling error*). Заметим, что вектор вклада $\boldsymbol{s}(\boldsymbol{w}_t; \boldsymbol{\theta})$ рассчитывается при истинном значении $\boldsymbol{\theta}_0$.

[1]Теорема о среднем значении применяется только к отдельным элементам $\boldsymbol{h}$, так что на самом деле $\overline{\boldsymbol{x}}$ изменяется от элемента к элементу векторного уравнения. Это усложнение не влияет на приведенный в тексте анализ.

Теперь, поскольку $\overline{\boldsymbol{\theta}}$ лежит между $\boldsymbol{\theta}_0$ и $\widehat{\boldsymbol{\theta}}$, $\overline{\boldsymbol{\theta}}$ — состоятельная оценка вектора $\boldsymbol{\theta}_0$, если таковой является $\widehat{\boldsymbol{\theta}}$. Если $\{\boldsymbol{w}_t\}$ — эргодический стационарный процесс, естественно будет предположить, что

$$\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{H}(\boldsymbol{w}_t;\overline{\boldsymbol{\theta}}) \underset{\mathrm{p}}{\to} \mathrm{E}\big[\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)\big]. \tag{7.3.8}$$

Стационарности и эргодичности $\boldsymbol{w}_t$ и состоятельности $\overline{\boldsymbol{\theta}}$ самих по себе недостаточно, чтобы обеспечить это; необходимо предположить выполнение некоторых технических условий. Одним из таких технических условий является равномерная сходимость $\frac{1}{n}\boldsymbol{H}(\boldsymbol{w}_t;\cdot)$ к $\mathrm{E}[\boldsymbol{H}(\boldsymbol{w}_t;\cdot)]$ в окрестности $\boldsymbol{\theta}_0$[1]. Но, по теореме о равномерной сходимости, равномерная сходимость $\frac{1}{n}\boldsymbol{H}(\boldsymbol{w}_t;\cdot)$ будет выполняться, если для гессиана выполняется следующее условие доминирования: для некоторой окрестности $\mathcal{N}$ точки $\boldsymbol{\theta}_0$ $\mathrm{E}[\sup_{\boldsymbol{\theta}\in\mathcal{N}}||\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta})||] < \infty$ (это условие (4) ниже)[2]. Это достаточное условие для (7.3.8)); если возможно напрямую проверить (7.3.8)), то тогда нет необходимости проверять условие доминирования для асимптотической нормальности.

Наконец, если выполняется (7.3.8)) и если

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0) \underset{\mathrm{d}}{\to} N(\mathbf{0},\boldsymbol{\Sigma}), \tag{7.3.9}$$

то, по теореме Слуцкого (см. лемму 2.4(c)), мы имеем:

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) \underset{\mathrm{d}}{\to} N\Big(\mathbf{0},\big(\mathrm{E}\big[\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)\big]\big)^{-1}\boldsymbol{\Sigma}\big(\mathrm{E}\big[\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)\big]\big)^{-1}\Big). \tag{7.3.10}$$

Собрав предположения, которые мы сделали ранее, мы получаем:

Утверждение 7.8 (асимптотическая нормальность М-оценок): *Предположим, что выполняются предположения утверждения 7.3 или утверждения 7.4, так что $\{\boldsymbol{w}_t\}$ — эргодический стационарный процесс*

---

[1]Это следствие приведенного ниже результата (см., например, Theorem 4.1.5 в [Amemiya, 1985]).

Предположим, что последовательность случайных функций $h_n(\boldsymbol{\theta})$ сходится равномерно по $\boldsymbol{\theta}$ к неслучайной функции $h_0(\boldsymbol{\theta})$ в открытой окрестности $\boldsymbol{\theta}_0$. Тогда $\mathrm{plim}_{n\to\infty} h_n(\widetilde{\boldsymbol{\theta}}) = h_0(\boldsymbol{\theta}_0)$, *если* $\mathrm{plim}\,\widetilde{\boldsymbol{\theta}} = \boldsymbol{\theta}_0$ *и $h_0(\boldsymbol{\theta})$ непрерывна в $\boldsymbol{\theta}_0$.*

Пусть роль $h_n(\boldsymbol{\theta})$ выполняет $\frac{1}{n}\boldsymbol{H}(\boldsymbol{w}_t;\cdot)$, а роль $h_0(\boldsymbol{\theta})$ — $\mathrm{E}[\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta})]$. Непрерывность $\mathrm{E}[\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta})]$ обеспечивается теоремой о равномерной сходимости. Равномерная сходимость необходима только в окрестности $\boldsymbol{\theta}_0$, а не во всем пространстве параметров, потому что $\widetilde{\boldsymbol{\theta}}$ близко к $\boldsymbol{\theta}_0$ для достаточно больших n.

[2]Матрицу можно рассматривать как вектор, элементы которого были переставлены. Так что Евклидова норма матрицы, такая как $||\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta})||$, есть квадратный корень из суммы квадратов всех ее элементов.

и M-оценка $\widehat{\theta}$, определенная в (7.1.1) и (7.1.2), состоятельна. Предположим также, что

(1) θ_0 *— внутренний элемент* Θ,

(2) $m(w_t;\theta)$ *дважды непрерывно дифференцируема по* θ *для любого* w_t,

(3) $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} s(w_t;\theta_0) \to_d N(\mathbf{0},\Sigma)$, Σ *положительно определенная матрица, где* $s(w_t;\theta)$ *определена в (7.3.2),*

(4) (условие локального доминирования на гессиан) для некоторой окрестности $\mathcal{N}$ *точки* θ_0,

$$\mathrm{E}[\sup_{\theta\in\mathcal{N}} \|H(w_t;\theta)\|\,] < \infty,$$

так что для любой состоятельной оценки $\widetilde{\theta}$ $\frac{1}{n}\sum_{t=1}^{n} H(w_t;\widetilde{\theta}) \to_p$ $\to_p \mathrm{E}\big[H(w_t;\theta_0)\big]$, *где* $H(w_t;\theta)$ *определена в (7.3.3),*

(5) $\mathrm{E}\big[H(w_t,\theta_0)\big]$ *невырождена.*

Тогда $\widehat{\theta}$ *асимптотически нормальна с*

$$\mathrm{Avar}(\widehat{\theta}) = \big(\mathrm{E}\big[H(w_t;\theta_0)\big]\big)^{-1}\Sigma\big(\mathrm{E}\big[H(w_t;\theta_0)\big]\big)^{-1}.$$

(Это Theorem 4.1.3 в [Amemiya, 1985], примененная к M-оценкам.) Необходимо сделать два замечания:

- Из предположений, которые мы сделали в процессе вывода асимптотической нормальности, следующие не перечислены в утверждении: (i) $\widehat{\theta}$ — внутренняя точка и (ii) $\frac{1}{n}\sum_{t=1}^{n} H(w_t;\overline{\theta})$ невырождена. Интуитивно понятно, что эти условия выполняются, потому что $\widehat{\theta}$ сходится по вероятности к внутренней точке θ_0 и $\frac{1}{n}\sum_{t=1}^{n} H(w_t;\overline{\theta})$ сходится по вероятности к невырожденной матрице $\mathrm{E}\big[H(w_t;\theta_0)\big]$). См.: [Newey and McFadden, 1994, p. 2152] для строгого доказательства. Технические моменты такого рода будут игнорироваться в оставшейся части главы.

- Если w_t стационарный и эргодический, то таким же является вектор вклада $s(w_t;\theta_0)$, а матрица Σ — долговременная ковариационная матрица $\{s(w_t;\theta_0)\}$[1]. Достаточным условием для (3) является условие Гордина, введенное в параграфе 6.5. Так что условие (3) в утверждении может быть заменено условием Гордина на $\{s(w_t;\theta_0)\}$. Оно будет выполняться, если, например, w_t i.i.d. и $\mathrm{E}[s(w_t;\theta_0)] = \mathbf{0}$. Предположение о том, что Σ положительно

[1] Долгосрочная ковариационная матрица была введена в параграфе 7.4.

определена, на самом деле не нужно для вывода утверждения, но мы можем также его сделать, потому что практически во всех прикладных задачах оно выполняется (или предполагается) и также потому, что оно потребуется для проверки гипотез далее в этой главе.

Состоятельное оценивание асимптотической ковариационной матрицы

Чтобы использовать асимптотические результаты для проверки гипотез, нам необходима состоятельная оценка для

$$\mathrm{Avar}(\widehat{\boldsymbol{\theta}}) = \left(\mathrm{E}\left[\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)\right]\right)^{-1}\boldsymbol{\Sigma}\left(\mathrm{E}\left[\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)\right]\right)^{-1}.$$

Так как $\widehat{\boldsymbol{\theta}} \to_p \boldsymbol{\theta}_0$, из условия (4) утверждения 7.8 следует, что

$$\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{H}(\boldsymbol{w}_t;\widehat{\boldsymbol{\theta}}) \underset{\mathrm{p}}{\to} \mathrm{E}\left[\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)\right].$$

Таким образом, при условии того, что нам доступна состоятельная оценка $\widehat{\boldsymbol{\Sigma}}$ матрицы $\boldsymbol{\Sigma}$,

$$\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\theta}})} = \left\{\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{H}(\boldsymbol{w}_t;\widehat{\boldsymbol{\theta}})\right\}^{-1}\widehat{\boldsymbol{\Sigma}}\left\{\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{H}(\boldsymbol{w}_t;\widehat{\boldsymbol{\theta}})\right\}^{-1} \tag{7.3.11}$$

является состоятельной оценкой асимптотической ковариационной матрицы. Чтобы получить $\widehat{\boldsymbol{\Sigma}}$, можно применить предложенные в параграфе 6.6 методы, такие, как VARHAC, к оцененным рядам $\{\boldsymbol{s}(\boldsymbol{w}_t;\widehat{\boldsymbol{\theta}})\}$ при условии выполнения некоторых надлежащих технических условий.

Асимптотическая нормальность условного ML

Мы применим сейчас этот результат об асимптотической нормальности к ML для случая i.i.d. данных. В ML, где функцией m является логарифм (условной) функции правдоподобия, условия первого порядка (7.3.4) называются **уравнениями правдоподобия** (*likelihood equations*). То, что функцией m является логарифм (условной) функции правдоподобия, приводит к упрощению асимптотической ковариационной матрицы $\widehat{\boldsymbol{\theta}}$. Мы покажем это для условного ML; сделать то же самое для безусловного или совместного ML проще.

Рассмотрим ниже два равенства в условии (3) утверждения 7.9. Второе равенство называется **равенством для информационных матриц** (*information matrix equality*) (не путать с информационным *не*равенством Кульбака — Лейблера). Оно называется так, потому что матрица $\mathrm{E}[\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)']$ — это **информационная матрица для наблюдения** $\boldsymbol{t}$ (*information matrix for observation t*). Вывод этих двух равенств при выполнении некоторых технических условий, позволяющих производить перестановку дифференцирования и интегрирования, остается в качестве аналитического упражнения 2. Здесь мы просто отметим применимость этих двух равенств в утверждении 7.8. Если $\boldsymbol{w}_t$ i.i.d., то из ЦПТ в формулировке Линдеберга — Леви и первого равенства ($\mathrm{E}[\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)] = \mathbf{0}$) следует, что

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0) \to_d N(\mathbf{0},\boldsymbol{\Sigma}), \text{ где } \boldsymbol{\Sigma} = \mathrm{E}[\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)']. \tag{7.3.12}$$

Из этого и из равенства для информационных матриц следует, что $\mathrm{Avar}(\widehat{\boldsymbol{\theta}})$ в утверждении 7.8 упрощается двумя способами как

$$\mathrm{Avar}(\widehat{\boldsymbol{\theta}}) = -\big\{\mathrm{E}\big[\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)\big]\big\}^{-1} = \big\{\mathrm{E}\big[\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)\,\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)'\big]\big\}^{-1}. \tag{7.3.13}$$

Таким образом, мы доказали

Утверждение 7.9 (асимптотическая нормальность условного ML): *Пусть $\boldsymbol{w}_t$ ($\equiv (y_t,\boldsymbol{x}_t')'$) i.i.d. Предположим, что выполняются предпосылки утверждения 7.5 или утверждения 7.6, так что $\widehat{\boldsymbol{\theta}} \to_p \boldsymbol{\theta}_0$. Предположим также, что*

(1) $\boldsymbol{\theta}_0$ — внутренняя точка Θ,

(2) $f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta})$ дважды непрерывно дифференцируема по $\boldsymbol{\theta}$ для всех $(y_t,\boldsymbol{x}_t)$,

(3) $\mathrm{E}[\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)] = \mathbf{0}$ и $-\mathrm{E}\big[\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)\big] = \mathrm{E}\big[\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)\,\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)'\big]$, где функции $\boldsymbol{s}$ и $\boldsymbol{H}$ определены в (7.3.2) и (7.3.3),

(4) (условие локального доминирования на гессиан) для некоторой окрестности $\mathcal{N}$ точки $\boldsymbol{\theta}_0$,

$$\mathrm{E}[\sup_{\boldsymbol{\theta}\in\mathcal{N}} \|\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta})\|\,] < \infty,$$

так что для любой состоятельной оценки $\widetilde{\boldsymbol{\theta}}$,

$$\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{H}(\boldsymbol{w}_t;\widetilde{\boldsymbol{\theta}}) \underset{p}{\to} \mathrm{E}\big[\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)\big].$$

(5) $\mathrm{E}\big[\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)\big]$ невырождена.

Тогда $\widehat{\boldsymbol{\theta}}$ асимптотически нормальна с $\mathrm{Avar}(\widehat{\boldsymbol{\theta}})$, заданной в (7.3.13).

Два замечания относительно этого утверждения:

- Условие (3) формулируется здесь как предположение, потому что его вывод требует перестановочности дифференцирования и интегрирования (см. аналитическое упражнение 2). Техническое условие, при котором такая перестановочность возможна, доступно из математического анализа (см., например, Lemma 3.6 в [Newey and McFadden, 1994]). Так что условие (3) может быть заменено таким техническим условием на $f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta})$. Мы не делаем этого здесь, потому что в большинстве прикладных задач условие (3) может быть проверено напрямую.

- Из обсуждения, проведенного выше, должно быть понятно, как можно применить это утверждение к безусловному ML: просто замените $f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta})$ на $f(\boldsymbol{w}_t;\boldsymbol{\theta})$.

Чтобы использовать этот асимптотический результат для проверки гипотез, нам нужна состоятельная оценка $\mathrm{Avar}(\widehat{\boldsymbol{\theta}})$. Учитывая, что она может быть записана как $-\left\{\mathrm{E}\left[\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)\right]\right\}^{-1}$, естественной оценкой является

$$\text{первая оценка } \mathrm{Avar}(\widehat{\boldsymbol{\theta}}) = -\left\{\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{H}(\boldsymbol{w}_t;\widehat{\boldsymbol{\theta}})\right\}^{-1}. \tag{7.3.14}$$

Поскольку $\widehat{\boldsymbol{\theta}} \to_p \boldsymbol{\theta}_0$, эта оценка состоятельна по условию (4) утверждения 7.9. Другая оценка, основанная на соотношении

$$\mathrm{Avar}(\widehat{\boldsymbol{\theta}}) = \left\{\mathrm{E}\left[\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)\,\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)'\right]\right\}^{-1}, \text{ — это}$$

$$\text{вторая оценка } \mathrm{Avar}(\widehat{\boldsymbol{\theta}}) = \left\{\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{s}(\boldsymbol{w}_t;\widehat{\boldsymbol{\theta}})\,\boldsymbol{s}(\boldsymbol{w}_t;\widehat{\boldsymbol{\theta}})'\right\}^{-1}. \tag{7.3.15}$$

Чтобы обеспечить состоятельность этой оценки, нам необходимо сделать некоторое предположение в дополнение к гипотезе утверждения 7.9 (см., например, Theorem 4.4 в [Newey and McFadden, 1994]). Мы не будем приводить его здесь, потому что во многих прикладных задачах состоятельность этой второй оценки может быть проверена напрямую.

Не существует никаких неопровержимых доводов для предпочтения одной оценки $\mathrm{Avar}(\widehat{\boldsymbol{\theta}})$ другой. С одной стороны, вторую оценку, которая требует только первых производных логарифма функции правдоподобия, вычислить проще, чем первую. Это может быть важным доводом, когда невозможно вычислить производные функции плотности аналитически и приходится использовать различные численные методы. С другой стороны, по крайней мере в некоторых случаях, первая оценка дает лучшую аппроксимацию асимптотического распределения (см., например: [Davidson и McKinnon, 1993]).

Два примера

Чтобы лучше понять результаты, касающиеся асимптотической нормальности условного ML, полезно посмотреть, как эти результаты могут быть применены к известным примерам. Мы рассмотрим два из них.

Пример 7.10 (асимптотическая нормальность в модели линейной регрессии): Запишем логарифм условной функции плотности для наблюдения t для модели линейной регрессии с нормальными ошибками:

$$\ln f(y_t|\boldsymbol{x}_t;\boldsymbol{\beta},\sigma^2) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma^2) - \frac{(y_t - \boldsymbol{x}_t'\boldsymbol{\beta})^2}{2\sigma^2}. \qquad (7.3.16)$$

Пространство $\Theta = \mathbb{R}^K \times \mathbb{R}_{++}$ (где K — размерность $\boldsymbol{\beta}$), и условие (1) утверждения 7.9 выполняется. Условие (2) выполняется очевидным образом. Чтобы проверить условие (3), проведем рутинные расчеты и получим, что[1]

$$\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}) = \begin{bmatrix} \frac{1}{\sigma^2}\boldsymbol{x}_t\cdot\widehat{\varepsilon}_t \\ -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}\widehat{\varepsilon}_t^2 \end{bmatrix}, \quad \boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta}) = \begin{bmatrix} -\frac{1}{\sigma^2}\boldsymbol{x}_t\boldsymbol{x}_t' & -\frac{1}{\sigma^4}\boldsymbol{x}_t\cdot\widehat{\varepsilon}_t \\ -\frac{1}{\sigma^4}\boldsymbol{x}_t'\cdot\widehat{\varepsilon}_t & \frac{1}{2\sigma^4} - \frac{1}{\sigma^6}\widehat{\varepsilon}_t^2 \end{bmatrix},$$

$$\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta})\,\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta})' = \begin{bmatrix} \frac{1}{\sigma^4}\boldsymbol{x}_t\boldsymbol{x}_t'\widehat{\varepsilon}_t^2 & -\frac{1}{2\sigma^4}\boldsymbol{x}_t\cdot\widehat{\varepsilon}_t + \frac{1}{2\sigma^6}\boldsymbol{x}_t\cdot\widehat{\varepsilon}_t^3 \\ -\frac{1}{2\sigma^4}\boldsymbol{x}_t'\cdot\widehat{\varepsilon}_t + \frac{1}{2\sigma^6}\boldsymbol{x}_t'\cdot\widehat{\varepsilon}_t^3 & \frac{1}{4\sigma^4} - \frac{1}{2\sigma^6}\widehat{\varepsilon}_t^2 + \frac{1}{4\sigma^8}\widehat{\varepsilon}_t^4 \end{bmatrix}, \qquad (7.3.17)$$

где $\boldsymbol{w}_t = (y_t, \boldsymbol{x}_t')'$, $\boldsymbol{\theta} = (\boldsymbol{\beta}', \sigma^2)'$, а $\widehat{\varepsilon}_t \equiv y_t - \boldsymbol{x}_t'\boldsymbol{\beta}$ (который следует отличать от $\varepsilon_t \equiv y_t - \boldsymbol{x}_t'\boldsymbol{\beta}_0$). Так что для $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ $\widehat{\varepsilon}_t$ в данной записи может быть заменен на ε_t. В модели линейной регрессии $\mathrm{E}(\varepsilon_t|\boldsymbol{x}_t) = 0$. Также, так как $\varepsilon_t \sim N(0, \sigma_0^2)$, мы имеем $E(\varepsilon_t^3) = 0$, $E(\varepsilon_t^4) = 3\sigma_0^4$. Используя эти соотношения, легко проверить (3). В частности,

$$-\mathrm{E}\big[\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)\big] = \mathrm{E}\big[\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)\,\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)'\big] = \begin{bmatrix} \frac{1}{\sigma_0^2}\mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}_t') & \boldsymbol{0} \\ \boldsymbol{0}' & \frac{1}{2\sigma_0^4}. \end{bmatrix}. \qquad (7.3.18)$$

Если $\mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}_t')$ невырождена, то $\mathrm{E}\big[\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)\big]$ невырождена и (5) выполняется. Для рассмотрения условия (4) пусть $\widetilde{\varepsilon}_t \equiv y_t - \boldsymbol{x}_t'\widetilde{\boldsymbol{\beta}}$ для некоторой состоятельной оценки $\widetilde{\boldsymbol{\beta}}$ и $\widetilde{\sigma}^2$. Условие (7.3.8)) в этом примере выглядит как

$$\begin{bmatrix} -\frac{1}{\widetilde{\sigma}^2}\frac{1}{n}\sum_{t=1}^n \boldsymbol{x}_t\boldsymbol{x}_t' & -\frac{1}{(\widetilde{\sigma}^2)^2}\frac{1}{n}\sum_{t=1}^n \boldsymbol{x}_t\cdot\widetilde{\varepsilon}_t \\ -\frac{1}{(\widetilde{\sigma}^2)^2}\frac{1}{n}\sum_{t=1}^n \boldsymbol{x}_t'\cdot\widetilde{\varepsilon}_t & \frac{1}{2(\widetilde{\sigma}^2)^2} - \frac{1}{(\widetilde{\sigma}^2)^3}\frac{1}{n}\sum_{t=1}^n \widetilde{\varepsilon}_t^2 \end{bmatrix} \underset{\mathrm{p}}{\to}$$

$$\underset{\mathrm{p}}{\to} \begin{bmatrix} -\frac{1}{\sigma_0^2}\mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}_t') & \boldsymbol{0} \\ \boldsymbol{0}' & -\frac{1}{2\sigma_0^4} \end{bmatrix}. \qquad (7.3.19)$$

[1]Для параметра σ^2 дифференцирование производится по самому параметру σ^2, а не по σ.

Показать это просто, учитывая, что $\tilde{\varepsilon}_t = \varepsilon_t - \boldsymbol{x}_t'(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$. Мы можем сделать вывод, что все условия утверждения 7.9 выполняются, если $\mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}_t')$ невырождена.

Пример 7.11 (асимптотическая нормальность ML-оценки модели пробит): Запишем логарифм условной функции правдоподобия для модели пробит:

$$\ln f(y_t|\boldsymbol{x}_t;\boldsymbol{\theta}) = y_t \ln \Phi(\boldsymbol{x}_t'\boldsymbol{\theta}) + (1-y_t)\ln[1-\Phi(\boldsymbol{x}_t'\boldsymbol{\theta})]. \tag{7.3.20}$$

Так как $\Theta = \mathbb{R}^p$, условие (1) утверждения 7.9 выполняется. Условие (2) очевидным образом выполняется. Для проверки условия (3) проведем скучные вычисления, которые дадут нам следующее:

$$\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}) = \frac{(y_t - \Phi(\boldsymbol{x}_t'\boldsymbol{\theta}))\phi(\boldsymbol{x}_t'\boldsymbol{\theta})}{(1-\Phi(\boldsymbol{x}_t'\boldsymbol{\theta}))\Phi(\boldsymbol{x}_t'\boldsymbol{\theta})}\boldsymbol{x}_t. \tag{7.3.21}$$

$$\begin{aligned}\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta}) = \Biggl\{ &- \left[\frac{y_t - \Phi(\boldsymbol{x}_t'\boldsymbol{\theta})}{\Phi(\boldsymbol{x}_t'\boldsymbol{\theta})(1-\Phi(\boldsymbol{x}_t'\boldsymbol{\theta}))}\right]^2 [\phi(\boldsymbol{x}_t'\boldsymbol{\theta})]^2 + \\ &+ \left[\frac{y_t - \Phi(\boldsymbol{x}_t'\boldsymbol{\theta})}{\Phi(\boldsymbol{x}_t'\boldsymbol{\theta})(1-\Phi(\boldsymbol{x}_t'\boldsymbol{\theta}))}\right] \phi'(\boldsymbol{x}_t'\boldsymbol{\theta})\Biggr\}\boldsymbol{x}_t\boldsymbol{x}_t'.\end{aligned} \tag{7.3.22}$$

(Чтобы вывести (7.3.22), используйте тот факт, что $y_t^2 = y_t$.) Здесь $\Phi(\cdot)$ — это (кумулятивная) функция распределения вероятности для $N(0,1)$, а $\phi(\cdot) \equiv \Phi'(\cdot)$ — его функция плотности. Тогда легко доказать условный вариант (3):

$$\begin{aligned}\mathrm{E}[\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)|\boldsymbol{x}_t] &= \boldsymbol{0} \quad \text{и} \\ -\mathrm{E}\bigl[\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)|\boldsymbol{x}_t\bigr] &= \mathrm{E}\bigl[\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)\,\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)'|\boldsymbol{x}_t\bigr].\end{aligned} \tag{7.3.23}$$

Так что (3) выполняется по закону полных математических ожиданий. В частности, для модели пробит просто показать, что

$$-\mathrm{E}\bigl[\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)\bigr] = \mathrm{E}\bigl[\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)\,\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)'\bigr] = \mathrm{E}\bigl[\lambda(\boldsymbol{x}_t'\boldsymbol{\theta}_0)\lambda(-\boldsymbol{x}_t'\boldsymbol{\theta}_0)\boldsymbol{x}_t\boldsymbol{x}_t'\bigr], \tag{7.3.24}$$

где

$$\lambda(v) \equiv \frac{\phi(v)}{1-\Phi(v)} \tag{7.3.25}$$

называется **обратным отношением Миллса** (*inverse Mill's ratio*), или **мерой риска** (*hazard*), для $N(0,1)$. Можно показать, что элемент в скобках в (7.3.22) лежит между 0 и 2 (см.: [Newey and McFadden, 1994, p. 2147]). Поэтому

$$\|\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta})\| \le 2\|\boldsymbol{x}_t\boldsymbol{x}_t'\|. \tag{7.3.26}$$

Евклидова норма $\|\boldsymbol{x}_t\boldsymbol{x}_t'\|$ есть квадратный корень из суммы квадратов элементов $\boldsymbol{x}_t\boldsymbol{x}_t'$. Поэтому математическое ожидание $\|\boldsymbol{x}_t\boldsymbol{x}_t'\|^2$ и, следовательно, математическое ожидание $\|\boldsymbol{x}_t\boldsymbol{x}_t'\|$ конечны, если $\mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}_t')$ (существует и) конечно. Тогда условие локального доминирования на гессиан (условие (4)) выполняется, если $\mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}_t')$ невырождена. Можно показать, что условие (5) выполняется, если $\mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}_t')$ невырождена (см. сноску на p. 2147 в [Newey and McFadden, 1994]. Мы снова можем сделать вывод о том, что условия утверждения 7.9 выполняются, если $\mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}_t')$ невырождена.

Асимптотическая нормальность GMM

Рассмотрим теперь GMM. Целевая функция GMM выглядит следующим образом:

$$Q_n(\boldsymbol{\theta}) = -\frac{1}{2}\boldsymbol{g}_n(\boldsymbol{\theta})'\,\widehat{\boldsymbol{W}}\boldsymbol{g}_n(\boldsymbol{\theta}), \text{ где } \underset{(K\times 1)}{\boldsymbol{g}_n(\boldsymbol{\theta})} \equiv \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{g}(\boldsymbol{w}_t;\boldsymbol{\theta}). \tag{7.3.27}$$

Как и в случае M-оценок, мы применяем теорему о среднем значении к условиям первого порядка задачи максимизации, но теорема будет применена к $\boldsymbol{g}_n(\boldsymbol{\theta})$, а не к ее первой производной. Таким образом, в отличие от случая M-оценок, целевая функция должна быть непрерывно дифференцируемой только один раз, а не дважды. Это отражает тот факт, что выборочное среднее входит в целевую функцию GMM по-другому.

В предположении, что $\widehat{\boldsymbol{\theta}}$ — внутренняя точка, условия первого порядка максимизации целевой функции таковы:

$$\underset{(p\times 1)}{\boldsymbol{0}} = \underset{(p\times 1)}{\frac{\partial Q_n(\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}}} = -\underset{(p\times K)}{\boldsymbol{G}_n(\widehat{\boldsymbol{\theta}})'}\ \underset{(K\times K)}{\widehat{\boldsymbol{W}}}\ \underset{(K\times 1)}{\boldsymbol{g}_n(\widehat{\boldsymbol{\theta}})}. \tag{7.3.28}$$

где $\boldsymbol{G}_n(\boldsymbol{\theta})$ — матрица Якоби для $\boldsymbol{g}_n(\boldsymbol{\theta})$:

$$\underset{(K\times p)}{\boldsymbol{G}_n(\boldsymbol{\theta})} \equiv \frac{\partial\boldsymbol{g}_n(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}. \tag{7.3.29}$$

Теперь применим теорему о среднем значении к $\boldsymbol{g}_n(\boldsymbol{\theta})$, а не к $\partial Q_n/\boldsymbol{\theta}'$, как в случае M-оценки, чтобы получить разложение в среднем значении:

$$\underset{(K\times 1)}{\boldsymbol{g}_n(\widehat{\boldsymbol{\theta}})} = \underset{(K\times 1)}{\boldsymbol{g}_n(\boldsymbol{\theta}_0)} + \underset{(K\times p)}{\boldsymbol{G}_n(\overline{\boldsymbol{\theta}})}\ \underset{(p\times 1)}{(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)}. \tag{7.3.30}$$

Подставляя это в записанное выше условие первого порядка, мы получаем:

$$\underset{(p\times 1)}{\boldsymbol{0}} = -\underset{(p\times K)}{\boldsymbol{G}_n(\widehat{\boldsymbol{\theta}})'}\ \underset{(K\times K)}{\widehat{\boldsymbol{W}}}\ \underset{(K\times 1)}{\boldsymbol{g}_n(\boldsymbol{\theta}_0)} - \underset{(p\times K)}{\boldsymbol{G}_n(\widehat{\boldsymbol{\theta}})'}\ \underset{(K\times K)}{\widehat{\boldsymbol{W}}}\ \underset{(K\times p)}{\boldsymbol{G}_n(\overline{\boldsymbol{\theta}})}\ \underset{(p\times 1)}{(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)}. \tag{7.3.31}$$

Решая это уравнение относительно $(\hat{\theta} - \theta_0)$ и учитывая, $= \frac{1}{n}\sum_{t=1}^{n} g(w_t, \theta)$, мы получаем:

$$\underset{(p\times 1)}{\sqrt{n}(\hat{\theta} - \theta_0)} = -\left[\underset{(p\times K)}{G_n(\hat{\theta})'} \underset{(K\times K)}{\widehat{W}} \underset{(K\times p)}{G_n(\bar{\theta})}\right]^{-1} \times$$

$$\times \underset{(p\times K)}{G_n(\hat{\theta})'} \underset{(K\times K)}{\widehat{W}} \underset{(K\times 1)}{\frac{1}{\sqrt{n}}\sum_{t=1}^{n} g(w_t; \theta_0)}.$$

В этом выражении якобиан $G_n(\theta)$ функции $g_n(\theta)$ вычисляется в разных точках — $\hat{\theta}$ и $\bar{\theta}$. Так как $g_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} g(w_t, \theta)$, якобиан любой оценки $\tilde{\theta}$ может быть записан как

$$G_n(\tilde{\theta}) = \frac{1}{n}\sum_{t=1}^{n} \frac{\partial g(w_t; \tilde{\theta})}{\partial \theta'}. \qquad (7.3.3)$$

Если w_t — стационарный и эргодический процесс и $\tilde{\theta}$ состоятельна, естественно предположить, что это выражение сходится к $\mathrm{E}[\partial g(w_t; \theta_0)/\partial \theta']$. Так же, как в случае M-оценок, это предположение будет верным, если $\partial g(w_t; \theta_0)/\partial \theta'$ удовлетворяет соответствующим условиям доминирования, сформулированным в условии (4) ниже. Теперь должно быть понятно, что эквивалентом утверждения 7.8 для GMM будет

Утверждение 7.10 (асимптотическая нормальность GMM): *Предположим, что выполняются предположения утверждения 7.7, так что $\{w_t\}$ стационарный и эргодический процесс, $\widehat{W}$ $(K \times K)$, сходится по вероятности к симметричной положительно определенной матрице W, и p-мерная GMM-оценка $\hat{\theta}$ состоятельна. Предположим также, что*

(1) θ_0 — внутренняя точка Θ,

(2) $\underset{(K\times 1)}{g(w_t; \theta)}$ непрерывно дифференцируема по θ для любых w_t,

(3) $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} g(w_t; \theta_0) \to_d N(0, \underset{(K\times K)}{S})$, S положительно определенная,

(4) (Условие локального доминирования $\frac{\partial g(w_t; \theta)}{\partial \theta'}$.) для некоторой окрестности $\mathcal{N}$ точки θ_0,

$$\mathrm{E}\left[\sup_{\theta \in \mathcal{N}} \left\| \frac{\partial g(w_t; \theta)}{\partial \theta'} \right\|\right] < \infty.$$

так что для любой состоятельной оценки $\tilde{\theta}$,

$$\frac{1}{n}\sum_{t=1}^{n} \underset{(K\times p)}{\frac{\partial g(w_t; \tilde{\theta})}{\partial \theta'}} \underset{p}{\to} \mathrm{E}\left[\frac{\partial g(w_t; \theta_0)}{\partial \theta'}\right].$$

(5) $\underset{(K\times p)}{\mathrm{E}\left[\frac{\partial \boldsymbol{g}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\right]}$ — *матрица полного столбцового ранга.*

Тогда будут верны следующие результаты:

(a) *(Асимптотическая нормальность.)* $\widehat{\boldsymbol{\theta}}$ *асимптотически нормальна с*

$$\underset{(p\times p)}{\mathrm{Avar}(\widehat{\boldsymbol{\theta}})} = (\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G})^{-1}\boldsymbol{G}'\boldsymbol{W}\boldsymbol{S}\boldsymbol{W}\boldsymbol{G}(\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G})^{-1},$$

где $\underset{(K\times p)}{\boldsymbol{G}} \equiv \mathrm{E}\left[\frac{\partial \boldsymbol{g}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\right]$,

(b) *(Состоятельная оценка асимптотической ковариационной матрицы.) Предположим, что имеется состоятельная оценка* $\widehat{\boldsymbol{S}}$ *для* $\boldsymbol{S}$. *Асимптотическая ковариационная матрица состоятельна и оценивается как*

$$\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\theta}})} = (\widehat{\boldsymbol{G}}'\widehat{\boldsymbol{W}}\widehat{\boldsymbol{G}})^{-1}\widehat{\boldsymbol{G}}'\widehat{\boldsymbol{W}}\widehat{\boldsymbol{S}}\widehat{\boldsymbol{W}}\widehat{\boldsymbol{G}}(\widehat{\boldsymbol{G}}'\widehat{\boldsymbol{W}}\widehat{\boldsymbol{G}})^{-1}, \tag{7.3.34}$$

где $\underset{(K\times p)}{\widehat{\boldsymbol{G}}} \equiv \boldsymbol{G}_n(\widehat{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial \boldsymbol{g}(\boldsymbol{w}_t;\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'}$.

Предположение о том, что $\boldsymbol{S}$ положительно определена, на самом деле не нужно для доказательства утверждения, однако мы можем его сделать, потому что практически во всех прикладных задачах это предположение выполняется (или предполагается), а также это нам будет необходимо позднее, при обсуждении проверки гипотез в данной главе.

Полезно отметить аналогию между линейным и нелинейным GMM, сравнив это утверждение с утверждением 3.1. В случае линейного GMM, $\boldsymbol{g}_n(\boldsymbol{\theta})$ задается как

$$\boldsymbol{g}_n(\boldsymbol{\theta}) = \left(\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{x}_t\cdot y_t\right) - \left(\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{x}_t\boldsymbol{z}_t'\right)\boldsymbol{\theta}. \tag{7.3.35}$$

Так что $\widehat{\boldsymbol{G}} \equiv \boldsymbol{G}_n(\widehat{\boldsymbol{\theta}})$ в утверждении 7.10 сводится к $(-\frac{1}{n}\sum \boldsymbol{x}_t\boldsymbol{z}_t')$, а $\boldsymbol{G}$ сводится к $-\mathrm{E}(\boldsymbol{x}_t\boldsymbol{z}_t')$. Суть утверждения 7.10 состоит в том, что асимптотическое распределение нелинейной оценки GMM может быть получено путем линейной аппроксимации $\boldsymbol{g}_n(\boldsymbol{\theta})$ в окрестности вектора истинных значений параметра.

Аналогия с линейным GMM наблюдается и далее:

- (Эффективность нелинейного GMM.) Используя текущие обозначения, запишем матричное неравенство (3.5.11) из параграфа 3.5:

$$(\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G})^{-1}\boldsymbol{G}'\boldsymbol{W}\boldsymbol{S}\boldsymbol{W}\boldsymbol{G}(\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G})^{-1} \geq (\boldsymbol{G}'\boldsymbol{S}^{-1}\boldsymbol{G})^{-1}, \tag{7.3.36}$$

  которое утверждает, что нижней границей для $\mathrm{Avar}(\widehat{\boldsymbol{\theta}})$ является $(\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G})^{-1}$. Нижняя граница достигается, если $\widehat{\boldsymbol{W}}$ удовлетворяет условию эффективности: $\mathrm{plim}_{n\to\infty}\widehat{\boldsymbol{W}} = \boldsymbol{S}^{-1}$, то есть $\widehat{\boldsymbol{W}} = \widehat{\boldsymbol{S}}^{-1}$ для некоторой состоятельной оценки $\boldsymbol{S}$.

- (J-статистика для проверки сверхидентифицирующих ограничений) Предположим, что условия утверждения 7.10 выполнены и что $\widehat{\boldsymbol{W}} = \widehat{\boldsymbol{S}}^{-1}$, так что $\widehat{\boldsymbol{W}}$ удовлетворяет условию эффективности:$\operatorname{plim} \widehat{\boldsymbol{W}} = \boldsymbol{S}^{-1}$. Тогда минимизированное расстояние $J \equiv n\boldsymbol{g}_n(\widehat{\boldsymbol{\theta}})'\widehat{\boldsymbol{S}}^{-1}\boldsymbol{g}_n(\widehat{\boldsymbol{\theta}})$ $(= -2nQ_n(\widehat{\boldsymbol{\theta}}))$ имеет асимптотическое распределение хи-квадрат с $K-p$ степенями свободы. Доказательство этого результата точно такое же, как и в линейном случае, см.: [Newey and McFadden, 1994, Section 9.5] для доказательства.

- (Оценивание $\boldsymbol{S}$.) $\boldsymbol{S}$ — это долговременная ковариационная матрица $\{\boldsymbol{g}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)\}$. При выполнении некоторых надлежащих технических ограничений методы, представленные в параграфе 6.6. — такие, как VARHAC, — могут применяться к оцененным рядам $\{\boldsymbol{g}(\boldsymbol{w}_t;\widehat{\boldsymbol{\theta}})\}$ для получения $\widehat{\boldsymbol{S}}$. В частности если $\boldsymbol{w}_t$ i.i.d., то тогда $\boldsymbol{S} = \mathrm{E}[\boldsymbol{g}(\boldsymbol{w}_t;\boldsymbol{\theta}_0) \times \boldsymbol{g}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)']$ и в роли $\widehat{\boldsymbol{S}}$ может выступать второй выборочный момент $\boldsymbol{g}(\boldsymbol{w}_t;\widehat{\boldsymbol{\theta}})$.

Сопоставление GMM с ML

Вопрос, на который мы только что ответили, таков: если принимать условия ортогональности как заданные, то каков оптимальный выбор взвешивающей матрицы $\boldsymbol{W}$? Связанный с предыдущим, но все же отдельный вопрос: каков оптимальный выбор условий ортогональности[1]? Этот вопрос также имеет определенный ответ при условии, что параметрическая форма функции плотности для данных $(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n)$ известна. Здесь мы сконцентрируемся на случае i.i.d., считая, что плотность вероятности $\boldsymbol{w}_t$ задана функцией $f(\boldsymbol{w}_t;\boldsymbol{\theta})$. Пусть $\widehat{\boldsymbol{\theta}}$ — оценка GMM, связанная с условиями ортогональности $\mathrm{E}[\boldsymbol{g}(\boldsymbol{w}_t,\boldsymbol{\theta}_0)] = \boldsymbol{0}$. Ее асимптотическая ковариационная матрица $\operatorname{Avar}(\widehat{\boldsymbol{\theta}})$ задается в утверждении 7.10. Несложно показать, что

$$\operatorname{Avar}(\widehat{\boldsymbol{\theta}}) \geq \mathrm{E}\big[\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)\,\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)'\big]^{-1}, \text{ где } \underset{(p\times 1)}{\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)} \equiv \frac{\partial \ln f(\boldsymbol{w}_t;\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}. \tag{7.3.37}$$

То есть матрица, обратная к информационной матрице $\mathrm{E}(\boldsymbol{s}\boldsymbol{s}')$, — нижняя граница для асимптотической ковариационной матрицы оценок GMM. Это матричное неравенство выполняется при выполнении предположений утверждения 7.10 и некоторых технических условий на $f(\boldsymbol{w}_t;\boldsymbol{\theta})$.

[1] Еще один вопрос об эффективности касается оптимального выбора инструментов в обобщенном нелинейном методе инструментальных переменных (частный случай GMM, в котором функция $\boldsymbol{g}$ — произведение вектора инструментов на вектор ошибок для оцениваемого уравнения). По поводу обсуждения оптимальных инструментов см.: [Newey, McFadden, 1994, Section 5.4].

которые допускают перестановочность дифференцирования и интегрирования. См.: [Newey and McFadden 1994, Theorem 5.1] для формулировки этих условий и доказательства. Тогда, исходя из этого результата, асимптотическая эффективность оценок ML выше, чем у GMM-оценок, потому что указанная нижняя граница $[\mathrm{E}(ss')]^{-1}$ — это асимптотическая ковариационная матрица оценки ML. Превосходство ML, однако, не очень удивительно, учитывая то, что ML использует информацию относительно параметрической формы $f(\boldsymbol{w}_t; \boldsymbol{\theta})$ функции плотности вероятности, в то время как GMM не делает этого.

Как вы покажете в контрольном вопросе 4 ниже, GMM достигает этой нижней границы, когда функция $\boldsymbol{g}$ в условиях ортогональности является функцией вклада для наблюдения t:

$$\underset{(K\times 1)}{\boldsymbol{g}(\boldsymbol{w}_t; \boldsymbol{\theta})} = \underset{(p\times 1)}{\boldsymbol{s}(\boldsymbol{w}_t; \boldsymbol{\theta})} \equiv \frac{\partial \ln f(\boldsymbol{w}_t; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}. \tag{7.3.38}$$

Таким образом, оценка GMM с оптимальными условиями ортогональности асимптотически эквивалентна ML. В действительности они численно эквивалентны, что можно показать следующим образом. Поскольку K (количество условий ортогональности) равно p (количеству параметров), когда $\boldsymbol{g}$ выбрано оптимально, как в (7.3.38), оценка GMM $\widehat{\boldsymbol{\theta}}$ должна удовлетворять условию $\boldsymbol{g}_n(\widehat{\boldsymbol{\theta}}) = \mathbf{0}$, что, в соответствии с (7.3.38), может быть записано как

$$\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{s}(\boldsymbol{w}_t; \widehat{\boldsymbol{\theta}}) = \mathbf{0}. \tag{7.3.39}$$

Это не что иное, как уравнения правдоподобия (то есть условия первого порядка для ML). Они имеют хотя бы одно решение, потому что оценка ML удовлетворяет им. Если у них ровно одно решение, то $\widehat{\boldsymbol{\theta}}$ является оценкой ML, равно как и оценкой GMM. Если у них более одного решения, то нам нужно выбрать одно из них в качестве GMM-оценки. Так как одним из решений является оценка ML, мы можем просто выбрать ее в качестве оценки GMM.

Выражение для ошибки оценки в едином формате

Чтобы подготовиться к изучению вопросов следующего параграфа, полезно привести несколько отличающееся доказательство асимптотической нормальности. Мы использовали теорему о среднем значении, чтобы выписать разложение в среднем значении для ошибки оценки, которое представлено в (7.3.7) для M-оценок и в (7.3.32) для GMM. В этом подпараграфе мы покажем ниже, что они могут быть записаны в одинаковом для обеих оценок виде (7.3.43).

Рассмотрим сначала M-оценку. Учитывая, что

$$\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} = \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{s}(\boldsymbol{w}_t; \boldsymbol{\theta}_0),$$

разложение в среднем значении (7.3.7) может быть записано как

$$\sqrt{n}\,(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -\left[\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{H}(\boldsymbol{w}_t; \overline{\boldsymbol{\theta}})\right]^{-1} \sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}. \tag{7.3.40}$$

По условию (4) утверждения 7.8, $\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{H}(\boldsymbol{w}_t; \overline{\boldsymbol{\theta}})$ сходится по вероятности к некоторой симметричной матрице $\boldsymbol{\Psi}$ размера $p \times p$, заданной как

$$\underset{(p\times p)}{\boldsymbol{\Psi}} = \mathrm{E}\big[\boldsymbol{H}(\boldsymbol{w}_t; \boldsymbol{\theta}_0)\big]. \tag{7.3.41}$$

Перепишем (7.3.40) в виде:

$$\begin{aligned} \sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = &- \boldsymbol{\Psi}^{-1}\sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} - \\ &- \left\{\left[\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{H}(\boldsymbol{w}_t; \overline{\boldsymbol{\theta}})\right]^{-1} - \boldsymbol{\Psi}^{-1}\right\}\sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}. \end{aligned} \tag{7.3.42}$$

По построению, выражение, заключенное в скобки, сходится по вероятности к нулю. По условию (3) утверждения 7.8,

$$\sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \ \left(= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} \boldsymbol{s}(\boldsymbol{w}_t; \boldsymbol{\theta})\right)$$

сходится к случайной величине. Так что последняя составляющая, которая является произведением выражения, заключенного в скобки, на $\sqrt{n} \times \times \frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}$, сходится по вероятности к нулю, по лемме 2.4(b). Этот факт можно компактно записать в виде **разложения Тейлора** (*Taylor expansion*)[1]:

$$\underset{(p\times 1)}{\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)} = -\underset{(p\times p)}{\boldsymbol{\Psi}^{-1}}\ \underset{(p\times 1)}{\sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}} + \underset{(p\times 1)}{o_p}, \tag{7.3.43}$$

где слагаемое «o_p» обозначает «некоторую случайную величину, которая сходится к нулю по вероятности»[2]. Точное выражение для слагаемого o_p будет зависеть от контекста. Здесь оно равно последней составляющей

[1] Для этого представления более подходит термин «разложение Тейлора», а «не разложение среднего значения», потому что матрица, которая перемножается с $\sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}$, вычисляется в $\boldsymbol{\theta}_0$, а не в $\overline{\boldsymbol{\theta}}$, как это делается в разложении в среднем значении.

[2] Обозначение «o_p» было введено в параграфе 2.1.

в (7.3.42). Как вы увидите, все, что важно для нашего анализа, — это чтобы слагаемое o_p исчезло (чтобы оно сходилось к нулю по вероятности), а не его выражение. Так как разность между $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ и $-\boldsymbol{\Psi}^{-1}\sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}$ исчезает, асимптотическое распределение $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ точно такое же, как асимптотическое распределение $-\boldsymbol{\Psi}^{-1}\sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}$ по лемме 2.4(а). Тогда из этого следует, что $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ сходится к нормальному распределению с нулевым математическим ожиданием и асимптотической ковариационной матрицей, заданной соотношением:

$$\mathrm{Avar}(\widehat{\boldsymbol{\theta}}) = \boldsymbol{\Psi}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Psi}^{-1} \text{ где } \underset{(p\times p)}{\boldsymbol{\Sigma}} \equiv \mathrm{Avar}\Big(\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\Big). \tag{7.3.44}$$

Для M-оценок $\sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta})$ и $\boldsymbol{\Sigma}$ является долговременной ковариационной матрицей $\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta})$. Если положить $\boldsymbol{\Psi} = \mathrm{E}[\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta})]$, то мы получим выражение для $\mathrm{Avar}\,\widehat{\boldsymbol{\theta}}$ из утверждения 7.8.

Теперь обратимся к GMM. Выражение для $\sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}$ выглядит следующим образом:

$$\sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} = -[\boldsymbol{G}_n(\boldsymbol{\theta}_0)]'\,\widehat{\boldsymbol{W}}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{g}(\boldsymbol{w}_t;\boldsymbol{\theta}_0). \tag{7.3.45}$$

(Вспомним, что $\boldsymbol{G}_n(\boldsymbol{\theta}) \equiv \frac{\partial \boldsymbol{g}_n(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}$.) Так как $\boldsymbol{G}_n(\boldsymbol{\theta}_0) = \frac{1}{n}\sum\frac{\partial \boldsymbol{g}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'} \to_p \boldsymbol{G}(\equiv \mathrm{E}[\frac{\partial \boldsymbol{g}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}'}])$, $\widehat{\boldsymbol{W}} \to_p \boldsymbol{W}$, и $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{g}(\boldsymbol{w}_t;\boldsymbol{\theta}_0) \to_d N(\boldsymbol{0},\boldsymbol{S})$ при выполнении условий утверждения 7.10, $\sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}$ будет сходиться к нормальному распределению с нулевым математическим ожиданием и

$$\underset{(p\times p)}{\boldsymbol{\Sigma}} \equiv \mathrm{Avar}\Big(\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}}\Big) = \underset{(p\times K)}{\boldsymbol{G}'}\underset{(K\times K)}{\boldsymbol{W}}\underset{(K\times K)}{\boldsymbol{S}}\underset{(K\times K)}{\boldsymbol{W}}\underset{(K\times p)}{\boldsymbol{G}}. \tag{7.3.46}$$

Теперь перепишем разложение в среднем значении (7.3.32) для ошибки GMM-оценки как

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -\boldsymbol{B}^{-1}\boldsymbol{c},\quad \boldsymbol{B} \equiv -\boldsymbol{G}_n(\widehat{\boldsymbol{\theta}})'\,\widehat{\boldsymbol{W}}\boldsymbol{G}_n(\overline{\boldsymbol{\theta}}),$$
$$\boldsymbol{c} \equiv -\boldsymbol{G}_n(\widehat{\boldsymbol{\theta}})'\,\widehat{\boldsymbol{W}}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{g}(\boldsymbol{w}_t;\boldsymbol{\theta}_0). \tag{7.3.47}$$

Далее, в контрольном вопросе 5, будет предложено показать, что это можно записать как разложение Тейлора (7.3.43) с симметричной матрицей $\boldsymbol{\Psi}$, заданной соотношением:

$$\underset{(p\times p)}{\boldsymbol{\Psi}} = -\underset{(p\times K)}{\boldsymbol{G}'}\underset{(K\times K)}{\boldsymbol{W}}\underset{(K\times p)}{\boldsymbol{G}}. \tag{7.3.48}$$

Подставляя (7.3.48) и (7.3.46) в (7.3.44), мы получим асимптотическую ковариационную матрицу GMM-оценки, описанную в утверждении 7.10. Табл. 7.1 подытоживает анализ, проведенный в этом подпараграфе, показывая, как стоит проводить подстановку, чтобы сделать общие формулы применимыми для M-оценок и GMM.

Таблица 7.1

Разложение Тейлора для отклонения выборочной оценки от истинного значения параметра

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -\boldsymbol{\Psi}^{-1}\sqrt{n}\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + o_p,\ \sqrt{n}\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \underset{d}{\to} N(\mathbf{0}, \boldsymbol{\Sigma}),\ \text{Avar}(\hat{\boldsymbol{\theta}}) = \boldsymbol{\Psi}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Psi}^{-1}$$

Элементы для замены	M-оценки	GMM
$Q_n(\boldsymbol{\theta})$	$\frac{1}{n}\sum_{t=1}^{n} m(\boldsymbol{w}_t, \boldsymbol{\theta})$	$-\frac{1}{2}\boldsymbol{g}_n(\boldsymbol{\theta})'\widehat{\boldsymbol{W}}\boldsymbol{g}_n(\boldsymbol{\theta})$
$\sqrt{n}\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}$	$\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{s}(\boldsymbol{w}_t; \boldsymbol{\theta}_0)$	$-[\boldsymbol{G}_n(\boldsymbol{\theta}_0)]'\widehat{\boldsymbol{W}}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \boldsymbol{g}(\boldsymbol{w}_t; \boldsymbol{\theta}_0)$
$\boldsymbol{\Psi}$	$E[\boldsymbol{H}(\boldsymbol{w}_t; \boldsymbol{\theta}_0)]$	$-\boldsymbol{G}'\mathbf{W}\mathbf{G}$
$\boldsymbol{\Sigma}$	долговременная ковариационная матрица для $\boldsymbol{s}(\boldsymbol{w}_t; \boldsymbol{\theta}_0)$	$\boldsymbol{G}'\mathbf{W}\mathbf{S}\mathbf{W}\mathbf{G}$

Примечание: Относительно определения $\boldsymbol{s}(\boldsymbol{w}_t; \boldsymbol{\theta})$ и $\boldsymbol{H}(\boldsymbol{w}_t; \boldsymbol{\theta})$ см. (7.3.2) и (7.3.3). Кроме того,

$$\boldsymbol{g}_n(\boldsymbol{\theta}) \equiv \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{g}(\boldsymbol{w}_t; \boldsymbol{\theta}) \text{ и } \boldsymbol{G} \equiv \mathrm{E}\left[\frac{\partial \boldsymbol{g}(\boldsymbol{w}_t; \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'}\right].$$

Между прочим, для анализа, проводимого в следующем параграфе, будет полезно заметить, что $\boldsymbol{\Sigma} = -\boldsymbol{\Psi}$ для ML с i.i.d. наблюдениями и для эффективного GMM (GMM с $\boldsymbol{W} = \boldsymbol{S}^{-1}$). Для ML $\boldsymbol{\Sigma} = \mathrm{E}[\boldsymbol{s}(\boldsymbol{w}_t; \boldsymbol{\theta}_0)\boldsymbol{s}(\boldsymbol{w}_t; \boldsymbol{\theta}_0)']$, что равняется матрице $\boldsymbol{\Psi}$ в (7.3.41), взятой со знаком минус, благодаря равенству для информационных матриц. Для эффективного GMM, задав $\boldsymbol{W} = \boldsymbol{S}^{-1}$ в (7.3.46), мы получим:

$$\boldsymbol{\Sigma} = \boldsymbol{G}'\boldsymbol{S}^{-1}\boldsymbol{G}. \tag{7.3.49}$$

что есть $\boldsymbol{\Psi}$ в (7.3.48) со знаком минус.

Контрольные вопросы

1. (Вклад и гессиан целевой функции.) Для M-оценок и, в частности, ML мы определили вклад наблюдения t, $\boldsymbol{s}(\boldsymbol{w}_t; \boldsymbol{\theta})$ как градиент (вектор первых производных) и гессиан для наблюдения t, $\boldsymbol{H}(\boldsymbol{w}_t; \boldsymbol{\theta})$, как гессиан функции $m(\boldsymbol{w}_t; \boldsymbol{\theta})$. Пусть вклад (без добавления слов «для наблюдения t») — это градиент целевой функции $Q_n(\boldsymbol{\theta})$, обозначаемый как $\boldsymbol{s}_n(\boldsymbol{\theta})$. Пусть гессиан (без добавления слов «для наблюдения t») — это гессиан целевой

функции $Q_n(\boldsymbol{\theta})$, обозначающийся как $\boldsymbol{H}_n(\boldsymbol{\theta})$. При выполнении условий утверждения 7.9 покажите, что

$$\mathrm{E}[\boldsymbol{s}_n(\boldsymbol{\theta}_0)] = \mathbf{0}, \quad -\frac{1}{n}\,\mathrm{E}[\boldsymbol{H}_n(\boldsymbol{\theta}_0)] = \mathrm{E}[\boldsymbol{s}_n(\boldsymbol{\theta}_0)\boldsymbol{s}_n(\boldsymbol{\theta}_0)'].$$

Указание: $\{\boldsymbol{s}_n(\boldsymbol{w}_t;\boldsymbol{\theta}_0)\}$ *i.i.d.*

2. (Равенство для условных информационных матриц в модели линейной регрессии.) Для линейной регрессии в примере 7.10 проверьте, что

$$\mathrm{E}[\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)|\boldsymbol{x}_t] = \mathbf{0} \text{ и } -\mathrm{E}[\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)|\boldsymbol{x}_t] = \mathrm{E}[\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)'|\boldsymbol{x}_t].$$

3. (Avar для модели линейной регрессии.) Для модели линейной регрессии из примера 7.10 пусть $(\widehat{\boldsymbol{\beta}},\widehat{\sigma}^2)$ — ML-оценка $(\boldsymbol{\beta},\sigma^2)$. Запишите первую оценку $\mathrm{Avar}(\widehat{\boldsymbol{\theta}})$, определенную в (7.3.14). Верно ли равенство:

$$\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\beta}})} = \widehat{\sigma}^2\left(\frac{1}{n}\sum \boldsymbol{x}_t\boldsymbol{x}_t'\right)^{-1}?$$

[Ответ: Да.]

4. (GMM с оптимальными условиями ортогональности.) Предположим, что соответствующие технические условия на гипотетическую функцию плотности $f(\boldsymbol{w}_t;\boldsymbol{\theta})$ выполняются, так что равенство для информационных матриц будет верным. Используя утверждение 7.10, покажите, что, когда функция $\boldsymbol{g}$ в моментных условиях задается как в (7.3.38), асимптотическая ковариационная матрица оценки GMM равняется матрице, обратной к информационной матрице для наблюдения t (являющейся нижней границей в (7.3.37)). **Указание:** *Так как K равняется p, $\boldsymbol{G}$ — квадратная матрица и выражение для $\mathrm{Avar}(\widehat{\boldsymbol{\theta}})$ в утверждении 7.10 редуцируется к $\mathrm{Avar}(\widehat{\boldsymbol{\theta}}) = \boldsymbol{G}^{-1}\boldsymbol{S}\boldsymbol{G}'^{-1}$. Из выбора $\boldsymbol{g}$ следует, что $\boldsymbol{G} = \mathrm{E}[\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)]$ и что $\boldsymbol{S} = \mathrm{E}[\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)']$.*

5. (Разложение Тейлора для ошибки GMM-оценки.) Покажите, что (7.3.47) может быть записано как разложение в ряд Тейлора (7.3.43) при выполнении условий утверждения 7.10, выполняя следующие этапы:

   (a) Покажите, что $\boldsymbol{B}^{-1} = \boldsymbol{\Psi}^{-1} + \boldsymbol{Y}_n$, где $\boldsymbol{Y}_n \to_p \mathbf{0}$.

   (b) Покажите, что $c = \sqrt{n}\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} + \boldsymbol{y}_n$, где $\boldsymbol{y}_n \to_p \mathbf{0}$. **Указание:**

$$\boldsymbol{G}_n(\widehat{\boldsymbol{\theta}}) - \boldsymbol{G}_n(\boldsymbol{\theta}_0) = (\boldsymbol{G}_n(\widehat{\boldsymbol{\theta}}) - \boldsymbol{G}) + (\boldsymbol{G} - \boldsymbol{G}_n(\boldsymbol{\theta}_0)) \to_p \mathbf{0} \text{ и}$$

$$\widehat{\boldsymbol{W}}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\boldsymbol{g}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)$$

   сходится по распределению к случайной величине.

   (c) Покажите, что $-\boldsymbol{B}^{-1}\boldsymbol{c} = -\boldsymbol{\Psi}^{-1}\sqrt{n}\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} + \boldsymbol{x}_n$, где

$$\boldsymbol{x}_n \equiv -\left[\boldsymbol{Y}_n\sqrt{n}\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} + \boldsymbol{\Psi}^{-1}\boldsymbol{y}_n + \boldsymbol{Y}_n\boldsymbol{y}_n\right].$$

   (d) Покажите, что $\boldsymbol{x}_n \to_p \mathbf{0}$.

6. (Почему расстояние в GMM делится на 2.) Проверьте, что если бы минус $g_n(\boldsymbol{\theta})'\widehat{S}^{-1}g_n(\boldsymbol{\theta})$ не делилось на 2 в определении целевой функции эффективного GMM, то мы бы не имели $\boldsymbol{\Sigma} = -\boldsymbol{\Psi}$.

7.4. Проверка гипотез

Как хорошо известно для ML, существует триада статистик — статистика Вальда, статистика множителей Лагранжа (LM) и статистика отношения правдоподобий (LR), которые могут быть использованы для проверки нулевой гипотезы. Эти три статистики асимптотически эквивалентны, в том смысле, что у них одинаковое асимптотическое распределение (χ^2). В этом параграфе мы покажем, что асимптотическая эквивалентность этой триады может быть распространена на GMM, выстраивая доказательство, применимое как к ML, так и к GMM. Краеугольным камнем этого доказательства будет общий формат ошибки оценки, который был получен в конце предыдущего параграфа. Читая этот параграф впервые, вы, возможно, захотите прочесть следующий подпараграф, а потом перейти к последнему подпараграфу.

Нулевая гипотеза

Мы уже рассматривали в контексте модели линейной регрессии проблему тестирования набора r возможно нелинейных ограничений (см. параграф 2.4). Кратко повторим: пусть $\boldsymbol{\theta}_0$ — p-мерный параметр модели. Нулевая гипотеза может быть выражена как

$$H_0: \underset{(r\times 1)}{\boldsymbol{a}(\boldsymbol{\theta}_0)} = \mathbf{0}. \tag{7.4.1}$$

Мы предполагаем, что $\boldsymbol{a}(\cdot)$ непрерывно дифференцируема. Также обозначим

$$\underset{(r\times p)}{\boldsymbol{A}(\boldsymbol{\theta})} \equiv \frac{\partial \boldsymbol{a}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} - \tag{7.4.2}$$

якобиан $\boldsymbol{a}(\boldsymbol{\theta})$. Мы предполагаем, что

$$\underset{(r\times p)}{\boldsymbol{A}_0} \equiv \boldsymbol{A}(\boldsymbol{\theta}_0) \text{ имеет полный столбцовый ранг.} \tag{7.4.3}$$

Это гарантирует, что указанные r ограничений не дублируют друг друга. Это условие на ранг дает нам $r \leqslant p$.

Пусть $\widehat{\boldsymbol{\theta}}$ — рассматриваемая экстремальная оценка. Это либо ML-оценка, либо GMM-оценка. Она решает задачу безусловной оптимизации (7.1.1). Статистика Вальда для проверки нулевой гипотезы использует оценку без ограничений. LM-статистика, напротив, использует **оценку с ограничениями** (*constrained estimator*), обозначаемую как $\tilde{\boldsymbol{\theta}}$, которая является решением задачи

$$\max_{\boldsymbol{\theta} \in \Theta} Q_n(\boldsymbol{\theta}) \quad \text{при ограничениях} \quad \boldsymbol{a}(\boldsymbol{\theta}) = \mathbf{0}. \tag{7.4.4}$$

Как было показано в параграфе 7.2, истинное значение параметра $\boldsymbol{\theta}_0$ является решением «предельной задачи безусловной оптимизации», где Q_n в задаче безусловной оптимизации (7.1.1) заменяется предельной функцией Q_0. Она также является решением «предельной задачи условной оптимизации», где Q_n в вышеописанной задаче условной оптимизации заменяется предельной функцией Q_0, потому что $\boldsymbol{\theta}_0$ удовлетворяет ограничениям. Оказывается, что равномерная сходимость по вероятности $Q_n(\cdot)$ к $Q_0(\cdot)$ обеспечивает то, что предел оценки с ограничениями $\tilde{\boldsymbol{\theta}}$ является решением предельной задачи условной оптимизации, что есть $\boldsymbol{\theta}_0$. По поводу доказательства см.: [Newey and McFadden, 1994, Theorem 9.1] для GMM и [Gallant and White, 1988, Lemma 7.3(b)] для общего случая экстремальных оценок. Также возможно показать, что если все условия, обеспечивающие асимптотическую нормальность для оценки модели без ограничений, выполняются, то тогда $\tilde{\boldsymbol{\theta}}$ будет также асимптотически нормальной при выполнении нулевой гипотезы. Для доказательства этого свойства для GMM-оценки см.: [Newey and McFadden, 1994, Theorem 9.2]; для ML, см., например: [Amemiya, 1985, Section 4.5][1].

Рабочие предпосылки

Доказательство того, что статистики Вальда, LM и LR имеют асимптотическое распределение $\chi^2(r)$, базируется на выполнении следующих условий:

(A) $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ допускает разложение в ряд Тейлора (7.3.43), которое имеет вид:
$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -\boldsymbol{\Psi}^{-1}\sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + o_p. \tag{7.4.5}$$

(B) $\sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \to_d N(\mathbf{0}, \boldsymbol{\Sigma})$, Σ положительно определена.

(C) $\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ сходится по распределению к некоторой случайной величине.

(D) $\boldsymbol{\Sigma} = -\boldsymbol{\Psi}$.

[1] Если вам интересно доказательство, то оно представляет собой всего лишь вариант разложения в среднем значении из рассуждений подпараграфа, посвященного LM-статистике, который базируется на разложении Тейлора.

Мы показали в предыдущем параграфе, что первые два условия выполняются для условного ML при выполнении условий утверждения 7.9, для безусловного ML — при выполнении этих условий, надлежащим образом модифицированных, а для GMM — при выполнении условий утверждения 7.10. Как только что было замечено, при выполнении тех же условий, оценка модели с ограничениями $\widetilde{\boldsymbol{\theta}}$ состоятельна и асимптотически нормальна, что обеспечивает выполнение условия (C). Таким образом, условия (A)–(C) вытекают из условий, которые обеспечивают асимптотическую нормальность экстремальной оценки.

Статистики Вальда и LM, если их определить должным образом, так, чтобы они не зависели от выполнения условия (D), асимптотически распределены по закону $\chi^2(r)$. Но если условие (D) выполняется, можно упростить выражения для этих статистик. Более того, без этого условия статистика LR *не будет* распределена асимптотически по $\chi^2(r)$. Как было отмечено в конце предыдущего параграфа, условие (D) выполняется для ML и для эффективного GMM с $\boldsymbol{W} = \boldsymbol{S}^{-1}$. Таким образом, та часть анализа, которая базируется на условии (D), неприменима к неэффективному GMM.

Статистика Вальда

В этой книге мы несколько раз получали статистику Вальда, используя «дельта-метод» из леммы 2.5. Здесь мы приведем несколько отличающуюся процедуру вывода этой статистики, основанную на разложении Тейлора. Применяя теорему о среднем значении к $\boldsymbol{a}(\widehat{\boldsymbol{\theta}})$ в окрестности $\boldsymbol{\theta}_0$ и учитывая, что $\boldsymbol{a}(\boldsymbol{\theta}_0) = \boldsymbol{0}$ при выполнении нулевой гипотезы, мы получим разложение в среднем значении для $\sqrt{n}\,\boldsymbol{a}(\widehat{\boldsymbol{\theta}})$:

$$\underset{(r\times 1)}{\sqrt{n}\,\boldsymbol{a}(\widehat{\boldsymbol{\theta}})} = \underset{(r\times p)}{\boldsymbol{A}(\overline{\boldsymbol{\theta}})}\ \underset{(p\times 1)}{\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)}. \tag{7.4.6}$$

Получение разложения Тейлора через разложение в среднем значении более прозрачно, чем то, которое было проведено в предыдущем параграфе. Так как $\overline{\boldsymbol{\theta}}$, лежащая между $\boldsymbol{\theta}_0$ и $\widehat{\boldsymbol{\theta}}$, состоятельна и так как $\boldsymbol{A}(\cdot)$ непрерывна, $\boldsymbol{A}(\overline{\boldsymbol{\theta}})$ сходится по вероятности к $\boldsymbol{A}_0(\equiv \boldsymbol{A}(\boldsymbol{\theta}_0))$. Более того, элемент, на который умножается $\boldsymbol{A}(\overline{\boldsymbol{\theta}})$, $\sqrt{n}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)$, сходится к некоторой случайной величине. Так что $(\boldsymbol{A}(\overline{\boldsymbol{\theta}}) - \boldsymbol{A}_0)\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ исчезает (сходится по вероятности к нулю), по лемме 2.4(b). Обозначая это слагаемое как o_p, перепишем разложение в среднем значении в виде разложения Тейлора:

$$\underset{(r\times 1)}{\sqrt{n}\,\boldsymbol{a}(\widehat{\boldsymbol{\theta}})} = \underset{(r\times p)}{\boldsymbol{A}_0}\ \underset{(p\times 1)}{\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)} + \underset{(r\times 1)}{o_p}. \tag{7.4.7}$$

(Учтите, что такая аргументация была бы неверна, если бы $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ не сходилось к некоторой случайной величине.)

Подставляя разложение Тейлора ((7.4.5) в это равенство, мы получаем уравнение, связывающее $\sqrt{n}\,\boldsymbol{a}(\widehat{\boldsymbol{\theta}})$ и $\sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}$:

$$\underset{(r\times 1)}{\sqrt{n}\,\boldsymbol{a}(\widehat{\boldsymbol{\theta}})} = -\underset{(r\times p)}{\boldsymbol{A}_0}\ \underset{(p\times p)}{\boldsymbol{\Psi}^{-1}}\ \underset{(p\times 1)}{\sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}} + o_p. \tag{7.4.8}$$

(Здесь o_p — это умноженное на $\boldsymbol{A}_0$ o_p из (7.4.7) плюс o_p из (7.4.7), но оно все равно сходится к нулю по вероятности.) Из этого выражения и асимптотической нормальности $\sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}$ (условие (B)) сразу же следует, что $\sqrt{n}\,\boldsymbol{a}(\widehat{\boldsymbol{\theta}})$ сходится к нормальному распределению с нулевым математическим ожиданием и

$$\begin{aligned} \underset{(r\times r)}{\operatorname{Avar}(\boldsymbol{a}(\widehat{\boldsymbol{\theta}}))} &= \boldsymbol{A}_0\boldsymbol{\Psi}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Psi}^{-1}\boldsymbol{A}_0' = \\ &= \underset{(r\times p)}{\boldsymbol{A}_0}\ \underset{(p\times p)}{\boldsymbol{\Sigma}^{-1}}\ \underset{(p\times r)}{\boldsymbol{A}_0'} \quad \text{(если выполняется условие (D)).} \end{aligned} \tag{7.4.9}$$

Поскольку матрица $\boldsymbol{A}_0$ размера $r \times p$ — матрица полного строкового ранга и $\boldsymbol{\Sigma}$ — положительно определенная матрица (по условию (B)), эта матрица размера $r \times r$ положительно определена. Таким образом, соответствующая квадратичная форма

$$n\boldsymbol{a}(\widehat{\boldsymbol{\theta}})'\left[\boldsymbol{A}_0\boldsymbol{\Sigma}^{-1}\boldsymbol{A}_0'\right]^{-1}\boldsymbol{a}(\widehat{\boldsymbol{\theta}}) \tag{7.4.10}$$

имеет асимптотическое распределение $\chi^2(r)$ при нулевой гипотезе. Как было замечено выше, $\boldsymbol{A}(\widehat{\boldsymbol{\theta}}) \to_p \boldsymbol{A}_0$. Так что если имеется состоятельная оценка $\widehat{\boldsymbol{\Sigma}}$ для $\boldsymbol{\Sigma}$, то $\boldsymbol{A}(\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{A}(\widehat{\boldsymbol{\theta}})'$ — состоятельная оценка для $\boldsymbol{A}_0\boldsymbol{\Sigma}^{-1}\boldsymbol{A}_0'$, из чего, по лемме 2.4(d), следует, что статистика Вальда, определенная как

$$W \equiv n\boldsymbol{a}(\widehat{\boldsymbol{\theta}})'\left[\boldsymbol{A}(\widehat{\boldsymbol{\theta}})\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{A}(\widehat{\boldsymbol{\theta}})'\right]^{-1}\boldsymbol{a}(\widehat{\boldsymbol{\theta}}), \tag{7.4.11}$$

также имеет асимптотическое $\chi^2(r)$ распределение при нулевой гипотезе.

Как получить состоятельную оценку $\widehat{\boldsymbol{\Sigma}}$ для $\boldsymbol{\Sigma}(= -\boldsymbol{\Psi})$, зависит от того, является ли экстремальная оценка ML-оценкой или оценкой эффективного GMM. Для ML она задается эффективной оценкой $-\boldsymbol{\Psi} = -\operatorname{E}[\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)]$:

$$-\frac{\partial^2 Q_n(\widehat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} = -\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{H}(\boldsymbol{w}_t;\widehat{\boldsymbol{\theta}}) \tag{7.4.12}$$

или состоятельной оценкой $\boldsymbol{\Sigma} = \operatorname{E}[\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)']$:

$$\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{s}(\boldsymbol{w}_t;\widehat{\boldsymbol{\theta}})\,\boldsymbol{s}(\boldsymbol{w}_t;\widehat{\boldsymbol{\theta}})'. \tag{7.4.13}$$

Для эффективного GMM, $\boldsymbol{\Sigma} = \boldsymbol{G}'\boldsymbol{S}^{-1}\boldsymbol{G}$ (см. (7.3.49)), что можно состоятельно оценить как

$$\widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{G}}'\,\widehat{\boldsymbol{S}}^{-1}\widehat{\boldsymbol{G}} \quad \text{с} \quad \underset{(K\times p)}{\widehat{\boldsymbol{G}}} \equiv \boldsymbol{G}_n(\widehat{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial \boldsymbol{g}(\boldsymbol{w}_t;\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'}. \tag{7.4.14}$$

Как уже отмечалось, $\widehat{\boldsymbol{S}}$ — оценка долговременной ковариационной матрицы, полученная из оцененного ряда $\{\boldsymbol{g}(\boldsymbol{w}_t;\widehat{\boldsymbol{\theta}})\}$.

Статистика множителей Лагранжа (LM-статистика)

Пусть γ_n $(r\times 1)$ — множитель Лагранжа для задачи с ограничениями (7.4.4). Оценка модели с ограничениями $\widetilde{\boldsymbol{\theta}}$ удовлетворяет условиям первого порядка:

$$\sqrt{n}\,\frac{\partial Q_n(\widetilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} + \boldsymbol{A}(\widetilde{\boldsymbol{\theta}})'\sqrt{n}\,\gamma_n = \underset{(p\times 1)}{\mathbf{0}}, \tag{7.4.15}$$

$$\sqrt{n}\,\boldsymbol{a}(\widetilde{\boldsymbol{\theta}}) = \underset{(r\times 1)}{\mathbf{0}}. \tag{7.4.16}$$

Мы хотим построить разложение Тейлора для этих условий первого порядка. По условию (C), $\sqrt{n}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ сходится к некоторой случайной величине. Тогда $\sqrt{n}\,\boldsymbol{a}(\widetilde{\boldsymbol{\theta}})$ допускает разложение Тейлора:

$$\sqrt{n}\,\boldsymbol{a}(\widetilde{\boldsymbol{\theta}}) = \boldsymbol{A}_0\,\sqrt{n}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + o_p. \tag{7.4.17}$$

В контрольном вопросе 2 будет предложено показать, что следующее разложение Тейлора верно как для ML, так и для GMM:

$$\sqrt{n}\,\frac{\partial Q_n(\widetilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + \boldsymbol{\Psi}\sqrt{n}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + o_p. \tag{7.4.18}$$

Следствием этого равенства является то, что $\sqrt{n}\,\frac{\partial Q_n(\widetilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}$ сходится к некоторой случайной величине. Тогда $\sqrt{n}\gamma_n$, удовлетворяющее (7.4.15), сходится к случайной величине. Следовательно, поскольку $\boldsymbol{A}(\widetilde{\boldsymbol{\theta}}) \to_p \boldsymbol{A}_0$, мы получаем:

$$\boldsymbol{A}(\widetilde{\boldsymbol{\theta}})'\sqrt{n}\,\gamma_n = \boldsymbol{A}_0'\,\sqrt{n}\,\gamma_n + (\boldsymbol{A}(\widetilde{\boldsymbol{\theta}}) - \boldsymbol{A}_0)'\sqrt{n}\,\gamma_n = \boldsymbol{A}_0'\,\sqrt{n}\,\gamma_n + o_p. \tag{7.4.19}$$

Подставляя эти три уравнения в условия первого порядка, мы получаем:

$$\begin{bmatrix} \underset{(p\times p)}{\boldsymbol{\Psi}} & \underset{(p\times r)}{\boldsymbol{A}_0'} \\ \underset{(r\times p)}{\boldsymbol{A}_0} & \underset{(r\times r)}{\mathbf{0}} \end{bmatrix} \begin{bmatrix} \underset{(p\times 1)}{\sqrt{n}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)} \\ \underset{(r\times 1)}{\sqrt{n}\,\gamma_n} \end{bmatrix} = \begin{bmatrix} \underset{(p\times 1)}{-\sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}} \\ \underset{(r\times 1)}{\mathbf{0}} \end{bmatrix} + o_p. \tag{7.4.20}$$

Используя формулу обращения для блочных матриц, можно решить это уравнение и получить в итоге:

$$\sqrt{n}\,\gamma_n = -(\boldsymbol{A}_0\boldsymbol{\Psi}^{-1}\boldsymbol{A}_0')^{-1}\boldsymbol{A}_0\boldsymbol{\Psi}^{-1}\sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} + o_p. \tag{7.4.21}$$

$$\sqrt{n}(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -\left[\boldsymbol{\Psi}^{-1} - \boldsymbol{\Psi}^{-1}\boldsymbol{A}_0'(\boldsymbol{A}_0\boldsymbol{\Psi}^{-1}\boldsymbol{A}_0')^{-1}\boldsymbol{A}_0\boldsymbol{\Psi}^{-1}\right]\sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} + o_p. \tag{7.4.22}$$

Дальнейший вывод LM-статистики такой же, как и для статистики Вальда. Согласно (7.4.21) $\sqrt{n}\,\gamma_n$ сходится к нормальному распределению с нулевым математическим ожиданием и асимптотической ковариационной матрицей, заданной соотношением:

$$\begin{aligned}\underset{(r\times r)}{\operatorname{Avar}(\gamma_n)} &= (\boldsymbol{A}_0\boldsymbol{\Psi}^{-1}\boldsymbol{A}_0')^{-1}\boldsymbol{A}_0\boldsymbol{\Psi}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Psi}^{-1}\boldsymbol{A}_0'(\boldsymbol{A}_0\boldsymbol{\Psi}^{-1}\boldsymbol{A}_0')^{-1} = \\ &= (\boldsymbol{A}_0\boldsymbol{\Sigma}^{-1}\boldsymbol{A}_0')^{-1} \quad \text{(если выполняется условие (D))}.\end{aligned} \tag{7.4.23}$$

Так как эта матрица размера $r \times r$ положительно определена, соответствующая ей квадратичная форма

$$n\gamma_n'(\boldsymbol{A}_0\boldsymbol{\Sigma}^{-1}\boldsymbol{A}_0')\gamma_n \tag{7.4.24}$$

при выполнении нулевой гипотезы имеет асимптотическое распределение $\chi^2(r)$. Если существует состоятельная оценка матрицы $\boldsymbol{\Sigma}$, обозначаемая как $\widetilde{\boldsymbol{\Sigma}}$, то тогда $\boldsymbol{A}(\widetilde{\boldsymbol{\theta}})\widetilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{A}(\widetilde{\boldsymbol{\theta}})'$ будет состоятельной оценкой для $\boldsymbol{A}_0\boldsymbol{\Sigma}^{-1}\boldsymbol{A}_0'$. Так что LM-статистика, определенная первой строкой выражения, приведенного ниже, при выполнении нулевой гипотезы имеет асимптотическое распределение $\chi^2(r)$. Эту статистику можно привести к более красивому виду, использовав условия первого порядка (7.4.15):

$$\begin{aligned}LM &\equiv n \underset{(1\times r)}{\gamma_n'} \underbrace{\left[\boldsymbol{A}(\widetilde{\boldsymbol{\theta}})\widetilde{\boldsymbol{\Sigma}}^{-1}\boldsymbol{A}(\widetilde{\boldsymbol{\theta}})'\right]}_{(r\times r)} \underset{(r\times 1)}{\gamma_n} = \\ &= n\big(\boldsymbol{A}(\widetilde{\boldsymbol{\theta}})'\gamma_n\big)'\widetilde{\boldsymbol{\Sigma}}^{-1}\big(\boldsymbol{A}(\widetilde{\boldsymbol{\theta}})'\gamma_n\big) = \\ &= n \underset{(1\times p)}{\left(\frac{\partial Q_n(\widetilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}}\right)'} \underset{(p\times p)}{\widetilde{\boldsymbol{\Sigma}}}^{-1} \underset{(p\times 1)}{\left(\frac{\partial Q_n(\widetilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}}\right)} \\ &\text{(так как } \boldsymbol{A}(\widetilde{\boldsymbol{\theta}})'\gamma_n = -\tfrac{\partial Q_n(\widetilde{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}} \text{ по (7.4.15))}.\end{aligned} \tag{7.4.25}$$

Последнее выражение для LM-статистики объясняет, почему ее также называют **тестовой статистикой вклада** (*score test statistic*). Она представляет собой расстояние от нуля до вклада, рассчитанного для оценки с ограничениями $\widetilde{\boldsymbol{\theta}}$. Однако вы не должны обманываться красотой этой статистики: несмотря на то что это квадратичная форма, соответствующая p-мерному вектору, количество степеней свободы ее асимптотического распределения χ^2 равно r, а не p, что ясно из приведенного процесса вывода этой статистики.

Для ML $\widetilde{\boldsymbol{\Sigma}}$ задается посредством (7.4.12) или (7.4.13), вычисленных в $\widetilde{\boldsymbol{\theta}}$. Для GMM состоятельной оценкой $\boldsymbol{\Sigma}$ будет:

$$\widetilde{\boldsymbol{\Sigma}} = \widetilde{\boldsymbol{G}}'\,\widetilde{\boldsymbol{S}}^{-1}\widetilde{\boldsymbol{G}} \quad \text{с} \quad \underset{(K\times p)}{\widetilde{\boldsymbol{G}}} \equiv \boldsymbol{G}_n(\widetilde{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{t=1}^{n}\frac{\partial \boldsymbol{g}(\boldsymbol{w}_t;\widetilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'}. \tag{7.4.26}$$

Здесь $\widetilde{\boldsymbol{S}}$ — оценка долговременной ковариационной матрицы, построенной для оцененных рядов $\{\boldsymbol{g}(\boldsymbol{w}_t;\widetilde{\boldsymbol{\theta}})\}$.

Статистика отношения правдоподобий (LR-статистика)

LR-статистика определяется как умноженная на $2n$ разность $Q_n(\widehat{\boldsymbol{\theta}}) - Q_n(\widetilde{\boldsymbol{\theta}})$. (Для эффективного GMM называть это статистикой отношения правдоподобия не совсем корректно, потому что Q_n для GMM — это не функция правдоподобия, но мы все же будем использовать этот термин как для ML, так и для GMM.) Для ML и GMM несложно показать (см. контрольные вопросы 6 и 7), что

$$LR \equiv 2n \cdot [Q_n(\widehat{\boldsymbol{\theta}}) - Q_n(\widetilde{\boldsymbol{\theta}})] = -n(\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}})'\boldsymbol{\Psi}(\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}}) + o_p. \tag{7.4.27}$$

Справедливость этого выражения не зависит от условия (D).

Из этого выражения легко вывести асимптотическое распределение LR-статистики. Вычитая (7.4.22) из (7.4.5), мы получаем:

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}}) = -\boldsymbol{\Psi}^{-1}\boldsymbol{A}_0'(\boldsymbol{A}_0\boldsymbol{\Psi}^{-1}\boldsymbol{A}_0')^{-1}\boldsymbol{A}_0\boldsymbol{\Psi}^{-1}\sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + o_p. \tag{7.4.28}$$

Подставляя это в (7.4.27), можно записать LR-статистику как

$$LR = -\left(\sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right)'\boldsymbol{\Psi}^{-1}\boldsymbol{A}_0'(\boldsymbol{A}_0\boldsymbol{\Psi}^{-1}\boldsymbol{A}_0')^{-1}\boldsymbol{A}_0\boldsymbol{\Psi}^{-1}\left(\sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right) + o_p. \tag{7.4.29}$$

Теперь воспользуемся предположением (D). LR-статистика превращается в

$$LR = \left(\sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right)'\boldsymbol{\Sigma}^{-1}\boldsymbol{A}_0'\left[\boldsymbol{A}_0\boldsymbol{\Sigma}^{-1}\boldsymbol{A}_0'\right]^{-1}\boldsymbol{A}_0\boldsymbol{\Sigma}^{-1}\left(\sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right) + o_p. \tag{7.4.30}$$

Это выражение распределено асимптотически по закону $\chi^2(r)$, потому что ковариационная матрица предельного распределения величины $\boldsymbol{A}_0\boldsymbol{\Sigma}^{-1}\left(\sqrt{n}\,\frac{\partial Q_n(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right)$ — положительно определенная матрица в скобках.

Это завершает доказательство асимптотической эквивалентности триады статистик. Но из этого доказательства должно быть понятно, что выполняется нечто более сильное, чем просто асимптотическая эквивалентность (тот факт, что три статистики имеют одинаковое асимптотическое распределение). Мы показали в доказательстве, что разность между статистикой Вальда и (7.4.10) есть o_p. Но, по (7.4.8), разность между вторым выражением и квадратичной формой в выражении для LR-статистики (7.4.30) — это o_p, в случае когда $\mathbf{\Sigma} = -\mathbf{\Psi}$. То есть численным отличием между статистикой Вальда и LR-статистикой будет o_p, если $\mathbf{\Sigma} = -\mathbf{\Psi}$. То же самое верно и для LM-статистики. Разностью между нею и (7.4.24) снова является o_p, когда $\mathbf{\Sigma} = -\mathbf{\Psi}$.

Итоговые выводы для триады статистик

Чтобы подытожить довольно длинное обсуждение в этом параграфе, приведем

Утверждение 7.11 (триада статистик): *Пусть $\widehat{\boldsymbol{\theta}}$ — экстремальная оценка, определенная в (7.1.1). Рассмотрим нулевую гипотезу (7.4.1), где $\boldsymbol{a}(\cdot)$ непрерывно дифференцируема (так что якобиан $\boldsymbol{A}(\boldsymbol{\theta}) \equiv \frac{\partial \boldsymbol{a}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}$ непрерывен) и выполняется ранговое условие (7.4.3). Пусть $\widetilde{\boldsymbol{\theta}}$ — экстремальная оценка модели с ограничениями, определенная в (7.4.4). Предположим:*

- *выполнение предположений утверждения 7.9, если экстремальная оценка — это оценка условного ML,*
- *выполнение предположений, соответствующим образом модифицированных, как показано прямо под утверждением 7.9, в случае оценки безусловного ML,*
- *выполнение предположений утверждения 7.10 с $\boldsymbol{W} = \boldsymbol{S}^{-1}$ и наличие состоятельной оценки $\widehat{\boldsymbol{S}}$ для $\boldsymbol{S}$, построенной из $\widehat{\boldsymbol{\theta}}$, и $\widetilde{\boldsymbol{S}}$ построенной из $\widetilde{\boldsymbol{\theta}}$, в случае эффективного GMM.*

Определим статистики Вальда, LM, и LR, как это сделано в табл. 7.2. Тогда

(a) экстремальная оценка с ограничениями $\widetilde{\boldsymbol{\theta}}$ состоятельна и асимптотически нормальна,

(b) при выполнении нулевой гипотезы все три статистики сходятся по распределению к $\chi^2(r)$, где r — количество ограничений в нулевой гипотезе,

(c) более того, при выполнении нулевой гипотезы численное различие между указанными тремя статистиками сходится к нулю по распределению.

Таблица 7.

Триада статистик

Статистика Вальда: $\underset{(1\times r)}{n a(\widehat{\theta})'} \big[\underset{(r\times p)}{A(\widehat{\theta})} \underset{(p\times p)}{\widehat{\Sigma}^{-1}} \underset{(p\times r)}{A(\widehat{\theta})'}\big]^{-1} \underset{(r\times 1)}{a(\widehat{\theta})}$

LM-статистика: $n \left(\underset{(1\times p)}{\frac{\partial Q_n(\widetilde{\theta})}{\partial \theta}} \right)' \underset{(p\times p)}{\widetilde{\Sigma}^{-1}} \left(\underset{(p\times 1)}{\frac{\partial Q_n(\widetilde{\theta})}{\partial \theta}} \right)$

LR-статистика: $2n \cdot [Q_n(\widehat{\theta}) - Q_n(\widetilde{\theta})]$

Элементы для подстановки	**Условный ML**	**Эффективный GMM**
$Q_n(\theta)$	$\frac{1}{n}\sum_{t=1}^{n} \log f(y_t \mid x_t; \theta)$	$-\frac{1}{2} g_n(\theta)' \widehat{S}^{-1} g_n(\theta)$
$\widehat{\Sigma}$	$-\frac{\partial^2 Q_n(\widehat{\theta})}{\partial\theta\partial\theta'}$ или $\frac{1}{n}\sum_{t=1}^{n} s(w_t;\widehat{\theta}) s(w_t;\widehat{\theta})'$	$\widehat{G}'\widehat{S}^{-1}\widehat{G}$, $\underset{(K\times p)}{\widehat{G}} \equiv G_n(\widehat{\theta})$
$\widetilde{\Sigma}$	заменить $\widehat{\theta}$ на $\widetilde{\theta}$ в выражении выше	$\widehat{G}'\widehat{S}^{-1}\widehat{G}$, $\underset{(K\times p)}{\widehat{G}} \equiv G_n(\widetilde{\theta})$

Примечание: По поводу определения $s(w_t;\theta)$ и $H(w_t;\theta)$ см. (7.3.2) и (7.3.3). Кроме того,

$$g_n(\theta) \equiv \frac{1}{n}\sum_{t=1}^{n} g(w_t;\theta), \; G_n(\theta) \equiv \frac{\partial g_n(\theta)}{\partial \theta'}, \text{ и } G \equiv \mathrm{E}\left[\frac{\partial g(w_t;\theta_0)}{\partial \theta'}\right].$$

В $Q_n(\widehat{\theta})$ и $Q_n(\widetilde{\theta})$ для GMM используется одна и та же матрица, стоящая здесь посередине.

Контрольные вопросы

1. Проверьте, что LM-статистика для ML может быть записана как

$$n\left(\frac{1}{n}\sum_{t=1}^{n} s(w_t;\widetilde{\theta})\right)' \left\{\frac{1}{n}\sum_{t=1}^{n} s(w_t;\widetilde{\theta}) s(w_t;\widetilde{\theta})'\right\}^{-1} \left(\frac{1}{n}\sum_{t=1}^{n} s(w_t;\widetilde{\theta})\right).$$

2. Выведите (7.4.18). **Указание:** *Для ML замените $\widehat{\theta}$ на $\widetilde{\theta}$ в* (7.3.5). *Для GMM выведите*

$$\sqrt{n}\frac{\partial Q_n(\widetilde{\theta})}{\partial \theta} =$$
$$= -G_n(\widetilde{\theta})'\widehat{W}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} g(w_t;\theta_0) - G_n(\widetilde{\theta})'\widehat{W} G_n(\overline{\theta})\sqrt{n}(\widetilde{\theta} - \theta_0),$$

где $\overline{\theta}$ лежит между θ_0 и $\widetilde{\theta}$. Вывод этого должен быть столько же простым, как и вывод (7.3.11).

3. Объясните, почему LR-статистика не будет асимптотически распределена по закону χ^2, если не выполняется условие $\Sigma = -\Psi$.

4. Рассмотрите триаду статистик, рассчитанных так, как указано в табл. 7.2, для ML, но предположите, что данные w_t автокоррелированы. Какая из трех оценок все еще будет распределена асимптотически по закону χ^2? [Ответ: Никакая.]

5. Предположим, что условие $\boldsymbol{\Sigma} = -\boldsymbol{\Psi}$ не выполняется, но предположим, что у нас есть состоятельная оценка $\widetilde{\boldsymbol{\Psi}}$ для матрицы $\boldsymbol{\Psi}$, равно как и состоятельная оценка $\widetilde{\boldsymbol{\Sigma}}$ для $\boldsymbol{\Sigma}$. Предложите LM-статистику, которая имеет асимптотическое распределение $\chi^2(r)$. **Указание:** *Используйте первое равенство в (7.4.23), которое остается верным и без выполнения $\boldsymbol{\Sigma} = -\boldsymbol{\Psi}$. Ответ таков:*

$$n\Big(\frac{\partial Q_n(\widetilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\Big)'\widetilde{\boldsymbol{\Psi}}^{-1}\boldsymbol{A}(\widetilde{\boldsymbol{\theta}})'\big[\boldsymbol{A}(\widetilde{\boldsymbol{\theta}})\widetilde{\boldsymbol{\Psi}}^{-1}\widetilde{\boldsymbol{\Sigma}}\widetilde{\boldsymbol{\Psi}}^{-1}\boldsymbol{A}(\widetilde{\boldsymbol{\theta}})'\big]^{-1} \times$$

$$\times\, \boldsymbol{A}(\widetilde{\boldsymbol{\theta}})\widetilde{\boldsymbol{\Psi}}^{-1}\Big(\frac{\partial Q_n(\widetilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}\Big). \tag{7.4.31}$$

6. (По желанию доказательство (7.4.27).) Докажите (7.4.27) для M-оценок. **Указание:** *Воспользуйтесь разложением в среднем значении «второго порядка» в окрестности $\widehat{\boldsymbol{\theta}}$*

$$Q_n(\widetilde{\boldsymbol{\theta}}) = Q_n(\widehat{\boldsymbol{\theta}}) + \frac{\partial Q_n(\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'}(\widetilde{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}) + \frac{1}{2}(\widetilde{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}})'\frac{\partial^2 Q_n(\overline{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}(\widetilde{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}), \tag{7.4.32}$$

где $\overline{\boldsymbol{\theta}}$ лежит между $\widehat{\boldsymbol{\theta}}$ и $\widetilde{\boldsymbol{\theta}}$. По условию первого порядка $\frac{\partial Q_n(\widehat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'} = \mathbf{0}$. Также $\frac{\partial^2 Q_n(\overline{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} \to_p \mathrm{E}[\boldsymbol{H}(\boldsymbol{w}_t; \boldsymbol{\theta}_0)]$.

7. (По желанию доказательство (7.4.27) для GMM.) Докажите (7.4.27) для GMM. **Указание:** *Выведите разложение Тейлора из разложения в среднем значении $\boldsymbol{g}_n(\boldsymbol{\theta})$ в окрестности $\widehat{\boldsymbol{\theta}}$*

$$\boldsymbol{g}_n(\widetilde{\boldsymbol{\theta}}) = \boldsymbol{g}_n(\widehat{\boldsymbol{\theta}}) + \boldsymbol{G}_n(\widehat{\boldsymbol{\theta}})(\widetilde{\boldsymbol{\theta}} - \widehat{\boldsymbol{\theta}}) + o_p. \tag{7.4.33}$$

Подставьте это выражение в $Q_n(\widetilde{\boldsymbol{\theta}}) = -\frac{1}{2}\boldsymbol{g}_n(\widetilde{\boldsymbol{\theta}})'\widehat{\boldsymbol{W}}\boldsymbol{g}_n(\widetilde{\boldsymbol{\theta}})$. После этого используйте условие первого порядка: $\boldsymbol{G}_n(\widehat{\boldsymbol{\theta}})'\widehat{\boldsymbol{W}}\boldsymbol{g}_n(\widehat{\boldsymbol{\theta}}) = \mathbf{0}$.

7.5. Численная оптимизация

До этого момента нас не интересовали вычислительные аспекты экстремальной оценки. В ML-оценивании модели линейной регрессии, например, целевая функция $Q_n(\boldsymbol{\theta})$ квадратична по $\boldsymbol{\theta}$, так что существует явная формула для оптимального значения параметра — оценки МНК. В линейном GMM целевая функция будет также линейной, и оценка линейного GMM может быть выписана в явном виде. В большинстве других случаев, однако, нет таких явных формул и необходимо применять какой-нибудь численный алгоритм, чтобы найти максимум. Существует несколько алгоритмов, но в этом параграфе будут рассмотрены только два самых важных. Больше деталей можно найти, например, в: [Judge et al., 1985, Appendix B]. Первый алгоритм используется для расчета M-оценок, в то время как второй используется для GMM.

Алгоритм Ньютона — Рафсона

Сначала мы рассмотрим M-оценку, целевой функцией $Q_n(\boldsymbol{\theta})$ которой является (7.1.2). Так как $Q_n(\boldsymbol{\theta})$ дважды непрерывно дифференцируема для M-оценок, для нее существует разложение Тейлора второго порядка:

$$Q_n(\boldsymbol{\theta}) \cong Q_n(\widehat{\boldsymbol{\theta}}_j) + \boldsymbol{s}_n(\widehat{\boldsymbol{\theta}}_j)'(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_j) + \frac{1}{2}(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_j)'\boldsymbol{H}_n(\widehat{\boldsymbol{\theta}}_j)(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_j). \quad (7.5.1)$$

где $\widehat{\boldsymbol{\theta}}_j$ — оценка на j-м шаге итеративной процедуры, которая вскоре будет описана, а $\boldsymbol{s}_n$ и $\boldsymbol{H}_n$ — градиент и гессиан целевой функции:

$$\boldsymbol{s}_n(\boldsymbol{\theta}) \equiv \frac{\partial Q_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \quad \boldsymbol{H}_n(\boldsymbol{\theta}) \equiv \frac{\partial^2 Q_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}. \quad (7.5.2)$$

Оценка $(j{+}1)$-го шага $\widehat{\boldsymbol{\theta}}_{j+1}$ максимизирует квадратичную функцию в правой части (7.5.1). Она задается формулой:

$$\widehat{\boldsymbol{\theta}}_{j+1} = \widehat{\boldsymbol{\theta}}_j - \left[\boldsymbol{H}_n(\widehat{\boldsymbol{\theta}}_j)\right]^{-1}\boldsymbol{s}_n(\widehat{\boldsymbol{\theta}}_j). \quad (7.5.3)$$

Эта итеративная процедура называется **алгоритмом Ньютона — Рафсона** (*Newton — Raphson algorithm*). Если целевая функция вогнута, то этот алгоритм часто быстро сходится к глобальному максимуму.

Для оценок ML, когда глобальный максимум $\widehat{\boldsymbol{\theta}}$ достигается таким образом, оценка $\mathrm{Avar}(\widehat{\boldsymbol{\theta}})$ получается как $-\boldsymbol{H}_n(\widehat{\boldsymbol{\theta}})^{-1}$ (вспомните, что $\boldsymbol{H}_n(\boldsymbol{\theta}) \equiv \equiv \frac{1}{n}\sum \boldsymbol{H}(\boldsymbol{w}_t; \boldsymbol{\theta})$).

Алгоритм Гаусса — Ньютона

Теперь обратимся к GMM, у которого целевой функцией является $Q_n(\boldsymbol{\theta}) = -\frac{1}{2}\boldsymbol{g}_n(\boldsymbol{\theta})'\widehat{\boldsymbol{W}}\boldsymbol{g}_n(\boldsymbol{\theta})$ (см. (7.1.13)). Как и при выводе асимптотического распределения, мы будем работать с линеаризацией функции $\boldsymbol{g}_n(\boldsymbol{\theta})$. Разложение Тейлора первого порядка для $\boldsymbol{g}_n(\boldsymbol{\theta})$ около $\widehat{\boldsymbol{\theta}}_j$ выглядит так:

$$\begin{aligned}\underset{(K\times 1)}{\boldsymbol{g}_n(\boldsymbol{\theta})} &\cong \boldsymbol{g}_n(\widehat{\boldsymbol{\theta}}_j) + \underset{(K\times p)}{\boldsymbol{G}_n(\widehat{\boldsymbol{\theta}}_j)}\underset{(p\times 1)}{(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_j)} = \\ &= \left[\boldsymbol{g}_n(\widehat{\boldsymbol{\theta}}_j) - \boldsymbol{G}_n(\widehat{\boldsymbol{\theta}}_j)\widehat{\boldsymbol{\theta}}_j\right] - \left[-\boldsymbol{G}_n(\widehat{\boldsymbol{\theta}}_j)\right]\boldsymbol{\theta} = \\ &= \boldsymbol{v}_j - \boldsymbol{G}_j\boldsymbol{\theta}.\end{aligned} \quad (7.5.4)$$

где

$$\boldsymbol{v}_j \equiv \boldsymbol{g}_n(\widehat{\boldsymbol{\theta}}_j) - \boldsymbol{G}_n(\widehat{\boldsymbol{\theta}}_j)\widehat{\boldsymbol{\theta}}_j, \quad \boldsymbol{G}_j \equiv -\boldsymbol{G}_n(\widehat{\boldsymbol{\theta}}_j). \quad (7.5.5)$$

(Вспомните, что $\boldsymbol{G}_n(\boldsymbol{\theta}) \equiv \frac{\partial \boldsymbol{g}_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}$.) Если бы функция $\boldsymbol{g}_n(\boldsymbol{\theta})$ в выражении для целевой функции GMM была этой линейной функцией $\boldsymbol{v}_j -$

$G_j\boldsymbol{\theta}$, то тогда целевая функция была бы квадратичной по $\boldsymbol{\theta}$ и значение, максимизирующее целевую функцию (или минимизирующее расстояние GMM), было бы линейной оценкой GMM:

$$\widehat{\boldsymbol{\theta}}_{j+1} = (\boldsymbol{G}_j'\widehat{\boldsymbol{W}}\boldsymbol{G}_j)^{-1}\boldsymbol{G}_j'\widehat{\boldsymbol{W}}\boldsymbol{v}_j. \tag{7.5.6}$$

Это оценка $(j+1)$-го шага в **алгоритме Гаусса — Ньютона** (*Gauss — Newton algorithm*)[1]. В отличие от алгоритма Ньютона — Рафсона, здесь нет необходимости вычислять вторые производные.

Запись алгоритмов Ньютона — Рафсона и Гаусса — Ньютона в общем формате

В алгоритмах Гаусса — Ньютона и Ньютона — Рафсона имеются и общие черты. Подставляя (7.5.5) в (7.5.6) и проводя перестановки, мы получаем:

$$\widehat{\boldsymbol{\theta}}_{j+1} = \widehat{\boldsymbol{\theta}}_j - \left[-\boldsymbol{G}_n(\widehat{\boldsymbol{\theta}}_j)'\,\widehat{\boldsymbol{W}}\boldsymbol{G}_n(\widehat{\boldsymbol{\theta}}_j)\right]^{-1}\left[-\boldsymbol{G}_n(\widehat{\boldsymbol{\theta}}_j)'\,\widehat{\boldsymbol{W}}\boldsymbol{g}_n(\widehat{\boldsymbol{\theta}}_j)\right]. \tag{7.5.7}$$

Выражение во вторых скобках есть не что иное, как градиент, вычисленный в $\widehat{\boldsymbol{\theta}}_j$ для целевой функции GMM $Q_n(\boldsymbol{\theta}) = -\frac{1}{2}\boldsymbol{g}_n(\boldsymbol{\theta})'\widehat{\boldsymbol{W}}\boldsymbol{g}_n(\boldsymbol{\theta})$. Роль, которую играет гессиан в целевой функции в (7.5.3), здесь играет выражение в первых скобках, которое является оценкой $-\boldsymbol{G}'\boldsymbol{W}\boldsymbol{G}$ в табл. 7.1, вычисленной в $\widehat{\boldsymbol{\theta}}_j$. Таким образом, аналогия между M-оценками и оценками GMM, описанная в табл. 7.1, наблюдается также и здесь. Более того, если $\boldsymbol{g}_n$ линейна, то эквивалент гессиана (выражение в первых скобках в (7.5.7)) *и есть* гессиан целевой функции GMM. Так что аналогия точная: алгоритм Гаусса — Ньютона совпадает с алгоритмом Ньютона — Рафсона.

Уравнения, нелинейные только по параметрам

В алгоритме Гаусса — Ньютона (7.5.7), тогда, когда $\boldsymbol{g}_n(\boldsymbol{\theta})(\equiv \frac{1}{n}\sum \boldsymbol{g}(\boldsymbol{w}_t;\boldsymbol{\theta}))$ нелинейна по $\boldsymbol{\theta}$, в общем случае необходимо вычислять средние по наблюдениям, чтобы вычислить градиент и эквивалент гессиана на каждой итерации. (Это также верно в общем случае для алгоритма Ньютона — Рафсона, если целевая функция не квадратичная по $\boldsymbol{\theta}$.) Для больших массивов данных это операция, требующая больших затрат времени работы центрального процессора. Есть, однако, один случай, когда усреднение по наблюдениям не требуется. Это верно для класса обобщенных нелинейных оценок методом инструментальных переменных (частный случай нелинейного GMM, в котором $\boldsymbol{g}(y_t, \boldsymbol{z}_t, \boldsymbol{\theta})$ можно представить

[1] Алгоритм Гаусса — Ньютона был изначально разработан для нелинейного метода наименьших квадратов. По поводу того, как алгоритм Гаусса — Ньютона применяется к NLS, см. контрольный вопрос 2. Приведенный здесь алгоритм является, таким образом, адаптацией для GMM.

как $a(y_t, z_t; \boldsymbol{\theta})\boldsymbol{x}_t$). Предположим, что выражение для $a(y_t, \boldsymbol{z}_t; \boldsymbol{\theta})$ имеет следующий вид:

$$a(y_t, \boldsymbol{z}_t; \boldsymbol{\theta}) = a_0(y_t, \boldsymbol{z}_t) + \underset{(1\times q)}{\boldsymbol{a}_1(y_t, \boldsymbol{z}_t)'}\underset{(q\times 1)}{\boldsymbol{\alpha}(\boldsymbol{\theta})}, \tag{7.5.8}$$

где $a_0(\cdot)$ и $\boldsymbol{a}_1(\cdot)$ — известные функции от $(y_t, \boldsymbol{z}_t)$ (но не функции от $\boldsymbol{\theta}$). Функция такого вида называется **нелинейной только по параметрам** (*nonlinear in parameters only*), или **линейной по переменным** (*linear in variables*). Можно легко показать, что

$$\boldsymbol{g}_n(\boldsymbol{\theta}) \equiv \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{g}(y_t, \boldsymbol{z}_t; \boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n} a(y_t, \boldsymbol{z}_t; \boldsymbol{\theta}) \cdot \boldsymbol{x}_t =$$

$$= \underset{(K\times 1)}{\left(\frac{1}{n}\sum_{t=1}^{n} a_0(y_t, \boldsymbol{z}_t) \cdot \boldsymbol{x}_t\right)} + \underset{(K\times q)}{\left(\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{a}_1(y_t, \boldsymbol{z}_t)'\right)}\underset{(q\times 1)}{\boldsymbol{\alpha}(\boldsymbol{\theta})}, \tag{7.5.9}$$

$$\boldsymbol{G}_n(\boldsymbol{\theta}) = \frac{\partial \boldsymbol{g}_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \underset{(K\times q)}{\left(\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{a}_1(y_t, \boldsymbol{z}_t)'\right)}\underset{(q\times p)}{\frac{\partial \boldsymbol{\alpha}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}}. \tag{7.5.10}$$

Из этих уравнений становится понятно, что выборочные средние нужно рассчитывать только один раз до начала итераций.

Контрольные вопросы

1. (Возрастает ли Q_n во время итераций?) Положим $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_{j+1}$ в (7.5.1) и перепишем это выражение в виде:

$$Q_n(\widehat{\boldsymbol{\theta}}_{j+1}) - Q_n(\widehat{\boldsymbol{\theta}}_j) \cong \boldsymbol{s}_n(\widehat{\boldsymbol{\theta}}_j)'(\widehat{\boldsymbol{\theta}}_{j+1} - \widehat{\boldsymbol{\theta}}_j) + \\ + \frac{1}{2}(\widehat{\boldsymbol{\theta}}_{j+1} - \widehat{\boldsymbol{\theta}}_j)' \boldsymbol{H}_n(\widehat{\boldsymbol{\theta}}_j)(\widehat{\boldsymbol{\theta}}_{j+1} - \widehat{\boldsymbol{\theta}}_j).$$

Рассмотрим алгоритм Ньютона — Рафсона (7.5.3). Предположим, что $\widehat{\boldsymbol{\theta}}_{j+1}$ достаточно близка к $\widehat{\boldsymbol{\theta}}_j$, чтобы знак $Q_n(\widehat{\boldsymbol{\theta}}_{j+1}) - Q_n(\widehat{\boldsymbol{\theta}}_j)$ был таким же, как знак правой части записанного уравнения. Покажите, что $Q_n(\widehat{\boldsymbol{\theta}}_{j+1}) \geqslant \geqslant Q_n(\widehat{\boldsymbol{\theta}}_j)$, если $Q_n(\boldsymbol{\theta})$ вогнута. **Указание:** *Используя* (7.5.3), *можно переписать правую часть уравнения как*

$$-\frac{1}{2}(\widehat{\boldsymbol{\theta}}_{j+1} - \widehat{\boldsymbol{\theta}}_j)' \boldsymbol{H}_n(\widehat{\boldsymbol{\theta}}_j)(\widehat{\boldsymbol{\theta}}_{j+1} - \widehat{\boldsymbol{\theta}}_j).$$

2. (Алгоритм Гаусса — Ньютона для NLS.) В NLS (нелинейном методе наименьших квадратов) целевая функция задается соотношением (7.1.19).

Пусть $\widehat{\boldsymbol{\theta}}_j$ — оценка на j-м шаге. Рассмотрим линейную аппроксимацию:

$$\varphi(\boldsymbol{x}_t;\boldsymbol{\theta}) \cong \varphi(\boldsymbol{x}_t;\widehat{\boldsymbol{\theta}}_j) + \underset{(p\times 1)}{\left(\frac{\partial\varphi(\boldsymbol{x}_t;\widehat{\boldsymbol{\theta}}_j)}{\partial\boldsymbol{\theta}}\right)'}\underset{(p\times 1)}{(\boldsymbol{\theta}-\widehat{\boldsymbol{\theta}}_j)} =$$

$$= \left[\varphi(\boldsymbol{x}_t;\widehat{\boldsymbol{\theta}}_j) - \left(\frac{\partial\varphi(\boldsymbol{x}_t;\widehat{\boldsymbol{\theta}}_j)}{\partial\boldsymbol{\theta}}\right)'\widehat{\boldsymbol{\theta}}_j\right] + \left(\frac{\partial\varphi(\boldsymbol{x}_t;\widehat{\boldsymbol{\theta}}_j)}{\partial\boldsymbol{\theta}}\right)'\boldsymbol{\theta}. \quad (7.5.11)$$

Замените $\varphi(\boldsymbol{x}_t;\boldsymbol{\theta})$ в функции m на эту линейную по $\boldsymbol{\theta}$ функцию. Покажите, что максимум в этой линейной задаче достигается при

$$\widehat{\boldsymbol{\theta}}_{j+1} = \widehat{\boldsymbol{\theta}}_j + \left[\frac{1}{n}\sum_{t=1}^{n}\left(\frac{\partial\varphi(\boldsymbol{x}_t;\widehat{\boldsymbol{\theta}}_j)}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial\varphi(\boldsymbol{x}_t;\widehat{\boldsymbol{\theta}}_j)}{\partial\boldsymbol{\theta}}\right)'\right]^{-1} \times$$

$$\times\left\{\frac{1}{n}\sum_{t=1}^{n}\underset{(p\times 1)}{\left(\frac{\partial\varphi(\boldsymbol{x}_t;\widehat{\boldsymbol{\theta}}_j)}{\partial\boldsymbol{\theta}}\right)}\cdot[y_t-\varphi(\boldsymbol{x}_t;\widehat{\boldsymbol{\theta}}_j)]\right\}. \quad (7.5.12)$$

Набор задач для главы 7

Аналитические упражнения

. (Информационное неравенство Кульбака — Лейблера.)
Пусть $f(y|\boldsymbol{x};\boldsymbol{\theta})$ — параметрическое семейство гипотетических функций условной плотности вероятности, а истинной функцией плотности является $f(y|\boldsymbol{x};\boldsymbol{\theta}_0)$. Предположим, что $\mathrm{E}[\ln f(y|\boldsymbol{x};\boldsymbol{\theta})]$ существует и конечно для всех $\boldsymbol{\theta}$. Информационное неравенство Кульбака — Лейблера утверждает, что

$$\mathrm{Prob}[f(y|\boldsymbol{x};\boldsymbol{\theta}) \neq f(y|\boldsymbol{x};\boldsymbol{\theta}_0)] > 0 \Rightarrow$$
$$\Rightarrow \mathrm{E}[\ln f(y|\boldsymbol{x};\boldsymbol{\theta})] < \mathrm{E}[\ln f(y|\boldsymbol{x};\boldsymbol{\theta}_0)].$$

Докажите это, предприняв следующие шаги. В доказательстве, исключительно ради простоты, предположите, что $f(y|\boldsymbol{x};\boldsymbol{\theta}) > 0$ для всех $(y,\boldsymbol{x})$ и $\boldsymbol{\theta}$, так что можно логарифмировать $f(y|\boldsymbol{x};\boldsymbol{\theta})$.

(a) Предположите, что $\mathrm{Prob}[f(y|\boldsymbol{x};\boldsymbol{\theta}) \neq f(y|\boldsymbol{x};\boldsymbol{\theta}_0)] > 0$. Пусть $\boldsymbol{w} = (y,\boldsymbol{x}')'$. Определим

$$a(\boldsymbol{w}) = f(y|\boldsymbol{x};\boldsymbol{\theta})/f(y|\boldsymbol{x};\boldsymbol{\theta}_0).$$

Проверьте, что $a(\boldsymbol{w}) \neq 1$ с положительной вероятностью, так что $a(\boldsymbol{w})$ — случайная величина, не являющаяся константой.

(b) Строгая версия **неравенства Йенсена** (*Jensen's inequality*) утверждает, что

если $c(x)$ — строго вогнутая функция, а x — случайная величина, отличная от постоянной, то $\mathrm{E}[c(x)] < c(\mathrm{E}(x))$.

Используя этот факт, покажите, что

$$\mathrm{E}[\ln a(\boldsymbol{w})] < \ln\{\mathrm{E}[a(\boldsymbol{w})]\}.$$

Указание: $\ln(x)$ — *строго вогнутая функция.*

(c) Покажите, что $\mathrm{E}[a(\boldsymbol{w})] = 1$. **Указание:** *условное математическое ожидание $a(\boldsymbol{w})$ равно 1, потому что*

$$\begin{aligned}\mathrm{E}[a(\boldsymbol{w})|\boldsymbol{x}] &= \int a(\boldsymbol{w}) f(y|\boldsymbol{x};\boldsymbol{\theta}_0)\mathrm{d}y = \\ &= \int \frac{f(y|\boldsymbol{x};\boldsymbol{\theta})}{f(y|\boldsymbol{x};\boldsymbol{\theta}_0)} f(y|\boldsymbol{x};\boldsymbol{\theta}_0)\mathrm{d}y = \\ &= \int f(y|\boldsymbol{x};\boldsymbol{\theta})\mathrm{d}y = \\ &= 1.\end{aligned}$$

Последнее равенство выполняется, потому что $f(y|\boldsymbol{x};\boldsymbol{\theta})$ — функция плотности вероятности для любых $\boldsymbol{x}$ и $\boldsymbol{\theta}$. Заметьте, что условное математическое ожидание берется по отношению к истинной условной функции плотности $f(y|\boldsymbol{x};\boldsymbol{\theta}_0)$.

(d) Наконец, покажите искомый результат.

2. (Равенство для информационных матриц.) Для ML вектор вклада и гессиан для наблюдения $(y, \boldsymbol{x})$ могут быть записаны (с $\boldsymbol{w} \equiv \equiv (y, \boldsymbol{x}')'$) как

$$\underset{(p\times 1)}{\boldsymbol{s}(\boldsymbol{w};\boldsymbol{\theta})} = \frac{\partial \ln f(y|\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}},$$

$$\underset{(p\times p)}{\boldsymbol{H}(\boldsymbol{w};\boldsymbol{\theta})} = \frac{\partial \boldsymbol{s}(\boldsymbol{w};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \frac{\partial^2 \ln f(y|\boldsymbol{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}.$$

Как обычно, для простоты предположите, что $f(y|\boldsymbol{x};\boldsymbol{\theta}) > 0$ для всех $(y, \boldsymbol{x})$ и $\boldsymbol{\theta}$, так что можно логарифмировать функцию плотности.

(a) Предполагая, что интегрирование (то есть вычисление математических ожиданий) и дифференцирование можно менять местами, покажите, что $\mathrm{E}[\boldsymbol{s}(\boldsymbol{w};\boldsymbol{\theta}_0)] = \boldsymbol{0}$. **Указание:** *Так как*

$f(y|\boldsymbol{x};\boldsymbol{\theta})$ — гипотетическая функция плотности, интеграл от нее равен единице:

$$\int f(y|\boldsymbol{x};\boldsymbol{\theta})\mathrm{d}y = 1.$$

Продифференцируйте обе стороны этого уравнения по $\boldsymbol{\theta}$, а затем поменяйте порядок дифференцирования и интегрирования, чтобы получить тождество:

$$\int \boldsymbol{s}(\boldsymbol{w};\boldsymbol{\theta})f(y|\boldsymbol{x};\boldsymbol{\theta})\mathrm{d}y = \underset{(p\times 1)}{\mathbf{0}}.$$

Положите $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Получите $\mathrm{E}[\boldsymbol{s}(\boldsymbol{w};\boldsymbol{\theta}_0)|\boldsymbol{x}] = \mathbf{0}$.

(b) Докажите равенство для информационных матриц:

$$-\mathrm{E}[\boldsymbol{H}(\boldsymbol{w};\boldsymbol{\theta}_0)] = \mathrm{E}[\boldsymbol{s}(\boldsymbol{w};\boldsymbol{\theta}_0)\boldsymbol{s}(\boldsymbol{w};\boldsymbol{\theta}_0)'].$$

Указание: *Дифференцируя обе стороны тождества в предыдущем указании и предполагая, что дифференцирование и интегрирование можно поменять местами, мы получаем:*

$$\int \frac{\partial}{\partial\boldsymbol{\theta}'}\underbrace{[\boldsymbol{s}(\boldsymbol{w};\boldsymbol{\theta})f(y|\boldsymbol{x};\boldsymbol{\theta})]}_{(p\times 1)}\mathrm{d}y = \underset{(p\times p)}{\mathbf{0}}.$$

Покажите, что подынтегральное выражение может быть записано в виде:

$$\frac{\partial}{\partial\boldsymbol{\theta}'}[\boldsymbol{s}(\boldsymbol{w};\boldsymbol{\theta})f(y|\boldsymbol{x};\boldsymbol{\theta})] =$$
$$= \boldsymbol{H}(\boldsymbol{w};\boldsymbol{\theta})f(y|\boldsymbol{x};\boldsymbol{\theta}) + \boldsymbol{s}(\boldsymbol{w};\boldsymbol{\theta})\boldsymbol{s}(\boldsymbol{w};\boldsymbol{\theta})'f(y|\boldsymbol{x};\boldsymbol{\theta}).$$

3. (Триада статистик для модели линейной регрессии.) Рассмотрите модель линейной регрессии с нормальными ошибками, в которой условная плотность для наблюдения t задается соотношением:

$$\ln f(y_t|\boldsymbol{x}_t;\boldsymbol{\beta},\sigma^2) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\sigma^2) - \frac{(y_t - \boldsymbol{x}_t'\boldsymbol{\beta})}{2\sigma^2}.$$

Пусть $(\widehat{\boldsymbol{\beta}}',\widehat{\sigma}^2)'$ — ML-оценка параметра $\boldsymbol{\theta} = (\boldsymbol{\beta}',\sigma^2)'$ для модели без ограничений, а $(\widetilde{\boldsymbol{\beta}}',\widetilde{\sigma}^2)'$ — ML-оценка для модели с ограничением $\boldsymbol{R\beta} = \boldsymbol{c}$, где $\boldsymbol{R}$ — матрица размера $r \times K$, состоящая из известных констант. Предположим, что $\Theta = \mathbb{R}^K \times \mathbb{R}_{++}$ и что $\mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}_t')$ невырождена. Пусть также

$$\widehat{\boldsymbol{\Sigma}} = \begin{bmatrix} \frac{1}{\widehat{\sigma}^2}\sum_{t=1}^n \boldsymbol{x}_t\boldsymbol{x}_t' & \mathbf{0} \\ \mathbf{0}' & \frac{1}{(\widehat{\sigma}^2)^2}\end{bmatrix}, \widetilde{\boldsymbol{\Sigma}} = \begin{bmatrix} \frac{1}{\widetilde{\sigma}^2}\sum_{t=1}^n \boldsymbol{x}_t\boldsymbol{x}_t' & \mathbf{0} \\ \mathbf{0}' & \frac{1}{(\widetilde{\sigma}^2)^2}\end{bmatrix}.$$

(a) Проверьте, что $\widehat{\beta}$ минимизирует сумму квадратов остатков, то есть является оценкой МНК. Проверьте, что $\widetilde{\beta}$ минимизирует сумму квадратов остатков при условии выполнения ограничения $\boldsymbol{R}\beta = \boldsymbol{c}$, то есть является МНК-оценкой для модели с ограничениями.

(b) Пусть $Q_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n} \ln f(y_t|\boldsymbol{x}_t; \boldsymbol{\beta}, \sigma^2)$. Покажите, что

$$Q_n(\widehat{\boldsymbol{\theta}}) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2} - \frac{1}{2}\ln\left(\frac{SSR_U}{n}\right),$$
$$Q_n(\widetilde{\boldsymbol{\theta}}) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2} - \frac{1}{2}\ln\left(\frac{SSR_R}{n}\right),$$

где $SSR_U(\equiv \sum(y_t - \boldsymbol{x}_t'\widehat{\beta})^2))$ — сумма квадратов остатков в модели без ограничений, и $SSR_R(\equiv \sum(y_t - \boldsymbol{x}_t'\widetilde{\beta})^2))$ — сумма квадратов остатков в модели с ограничениями. **Указание:** *Покажите, что $\widehat{\sigma}^2 = SSR_U/n$ и $\widetilde{\sigma}^2 = SSR_R/n$.*

(c) Проверьте, что данная $\widehat{\boldsymbol{\Sigma}}$, хотя и не равняется $-\frac{1}{n}\boldsymbol{H}(\boldsymbol{w}_t; \widehat{\boldsymbol{\theta}})$, является состоятельной оценкой для $-\mathrm{E}[\boldsymbol{H}(\boldsymbol{w}_t; \boldsymbol{\theta}_0)]$. Проверьте, что $\widetilde{\boldsymbol{\Sigma}}$, хотя и не равняется $-\frac{1}{n}\boldsymbol{H}(\boldsymbol{w}_t; \widetilde{\boldsymbol{\theta}})$, является состоятельной оценкой для $-\mathrm{E}[\boldsymbol{H}(\boldsymbol{w}_t; \boldsymbol{\theta}_0)]$. **Указание:** *Из обсуждения в примере 7.8 вытекает, что $\widehat{\boldsymbol{\theta}}$ состоятельна. Как отмечалось в параграфе 7.4, $\widetilde{\boldsymbol{\theta}}$ также состоятельна при выполнении нулевой гипотезы.*

(d) Покажите, используя $\widehat{\boldsymbol{\Sigma}}$ и $\widetilde{\boldsymbol{\Sigma}}$, формулы для которых приведены в табл. 7.2, что статистики Вальда, LM и LR могут быть записаны как

$$W = n \cdot \frac{(\boldsymbol{R}\widehat{\beta} - \boldsymbol{c})'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\widehat{\beta} - \boldsymbol{c})}{SSR_U},$$
$$LM = n \cdot \frac{(\boldsymbol{y} - \boldsymbol{X}\widetilde{\beta})'\boldsymbol{P}(\boldsymbol{y} - \boldsymbol{X}\widetilde{\beta})}{SSR_R},$$
$$LR = n \cdot \left\{\ln\left(\frac{SSR_R}{n}\right) - \ln\left(\frac{SSR_U}{n}\right)\right\},$$

где $\boldsymbol{y}(n \times 1)$ и $\boldsymbol{X}(n \times K)$ — вектор наблюдений и матрица, ассоциируемые с y_t и $\boldsymbol{x}_t$, а $\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$.
Указание: $\boldsymbol{A}(\boldsymbol{\theta})(r \times (K+1))$ *в табл. 7.2 равна:*

$$\left(\underset{(r\times K)}{\boldsymbol{R}} \ \vdots \ \underset{(r\times 1)}{\boldsymbol{0}}\right).$$

(e) Покажите, что эти три статистики можно записать также следующим образом:

$$W = n \cdot \frac{SSR_R - SSR_U}{SSR_U},$$
$$LM = n \cdot \frac{SSR_R - SSR_U}{SSR_R},$$
$$LR = n \cdot \ln\left(\frac{SSR_R}{SSR_U}\right).$$

Указание: *Как мы показали в аналитическом упражнении к главе 1,*

$$\begin{aligned} SSR_R - SSR_U &= (\widetilde{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}})'(\boldsymbol{X}'\boldsymbol{X})(\widetilde{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}) = \\ &= (\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{c})'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{c}) = \\ &= (\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}})'\boldsymbol{P}(\boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}). \end{aligned}$$

(f) Покажите, что $W \geqslant LR \geqslant LM$. (Эти неравенства не всегда выполняются в моделях нелинейной регрессии.)

Ответы на избранные вопросы

2a. Так как $f(y|\boldsymbol{x};\boldsymbol{\theta})$ — гипотетическая функция плотности вероятности, интеграл от нее равен единице:

$$\int f(y|\boldsymbol{x};\boldsymbol{\theta})\mathrm{d}y = 1. \tag{1}$$

Это тождество, верное для любого значения $\boldsymbol{\theta} \in \Theta$. Дифференцируя обе части этого тождества по $\boldsymbol{\theta}$, мы получим:

$$\frac{\partial}{\partial\boldsymbol{\theta}} \int f(y|\boldsymbol{x};\boldsymbol{\theta})\mathrm{d}y = \underset{(p\times1)}{\mathbf{0}}. \tag{2}$$

Если операции дифференцирования и интегрирования можно менять местами, то

$$\frac{\partial}{\partial\boldsymbol{\theta}} \int f(y|\boldsymbol{x};\boldsymbol{\theta})\mathrm{d}y = \int \frac{\partial}{\partial\boldsymbol{\theta}} f(y|\boldsymbol{x};\boldsymbol{\theta})\mathrm{d}y. \tag{3}$$

По определению вклада, $\boldsymbol{s}(\boldsymbol{w};\boldsymbol{\theta})f(y|\boldsymbol{x};\boldsymbol{\theta}) = \frac{\partial}{\partial\boldsymbol{\theta}} f(y|\boldsymbol{x};\boldsymbol{\theta})$. Подставляя это в (3), получаем:

$$\int \boldsymbol{s}(\boldsymbol{w};\boldsymbol{\theta})f(y|\boldsymbol{x};\boldsymbol{\theta})\mathrm{d}y = \underset{(p\times1)}{\mathbf{0}}. \tag{4}$$

Это верно для любого значения $\boldsymbol{\theta} \in \Theta$, в частности для $\boldsymbol{\theta}_0$. Полагая $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, мы получаем:

$$\int \boldsymbol{s}(\boldsymbol{w}; \boldsymbol{\theta}_0) f(y|\boldsymbol{x}; \boldsymbol{\theta}_0)\mathrm{d}y = \mathrm{E}[\boldsymbol{s}(\boldsymbol{w}; \boldsymbol{\theta}_0)|\boldsymbol{x}] = \underset{(p\times 1)}{\mathbf{0}}. \quad (5)$$

Затем, используя закон повторных математических ожиданий, мы получаем желаемый результат.

Литература

Amemiya, T., 1985, *Advanced Econometrics*, Cambridge: Harvard University Press.

Davidson, R., and J. MacKinnon, 1993, *Estimation and Inference in Econometrics*, New York: Oxford University Press.

Gallant, R., and H. White, 1988, *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, New York: Basil Blackwell.

Gourieroux, C., and A. Monfort, 1995, *Statistics and Econometric Models*, New York: Cambridge University Press.

Hansen, L. P., and K. Singleton, 1982, "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269–1286.

Judge, G., W. Griffiths, R. Hill, H. Lütkepohl, and T, Lee, 1985, *The Theory and Practice of Econometrics* (2d ed.), New York: Wiley.

Newey, W., and D. McFadden, 1994, "Large Sample Estimation and Hypothesis Testing," Chapter 36 in R. Engle and D. McFadden (eds.), *Handbook of Econometrics*, Volume IV, New York: North-Holland.

White, H., 1994, *Estimation, Inference, and Specification Analysis*, New York: Cambridge University Press.

Глава 8. Примеры применения метода максимального правдоподобия

Метод максимального правдоподобия (ML) был изложен в предыдущей главе как частный случай экстремальных оценок. Вследствие его важности в эконометрике в этой главе рассматривается применение ML к некоторым известным моделям. Мы уже упоминали о некоторых из этих моделей в главах 3 и 4 в контексте оценивания обобщенным методом моментов.

При выводе асимптотических свойств ML-оценок моделей, рассматриваемых здесь, будут использованы общие результаты для ML, полученные в главе 7. Перед чтением этой главы вам следует просмотреть параграф 7.1, предположения 7.5, 7.6, 7.8, 7.9 и 7.11.

Замечания по обозначениям: Мы сохраняем обозначения, введенные в предыдущей главе: истинное значение параметра обозначается индексом 0. Так что θ — гипотетическое значение параметра, θ_0 — истинное значение параметра.

8.1. Модели с качественным откликом

Во многих ситуациях в экономике и в других науках зависимая переменная представляет собой категоризованную переменную. Например, человек может входить или не входить в состав рабочей силы; пассажир выбирает тот или иной вид транспорта; пациент может умереть или выжить и так далее. Регрессионные модели, в которых зависимая переменная принимает дискретные значения, называются **моделями с качественным откликом** (*qualitative response*) — **(QR)**. Модель с качественным откликом описывается параметризованной функцией плотности $f(y_t|\boldsymbol{x}_t; \theta)$, представляющей семейство *дискретных* распределений вероятностей y_t при условии $\boldsymbol{x}_t$, индексируемых параметром θ.

QR-модель называется **моделью с бинарным выбором** (*binary response model*), если зависимая переменная может принимать только два значения (их можно считать равными 0 и 1 без потери общности), и называется **моделью с множественным выбором** (*multinomial response model*), если зависимая переменная принимает более двух значений.

В этом подпараграфе рассматриваются только модели с бинарным выбором. По поводу более подробного обсуждения моделей с бинарным и множественным откликом см.: [Amemiya, 1985, Chapter 9]).

Наиболее популярной моделью с бинарным откликом является пробит-модель, которая уже была представлена в примере 7.3. Другой популярной бинарной моделью является **логит-модель** (*logit model*):

$$\begin{cases} f(y_t = 1|\boldsymbol{x}_t; \boldsymbol{\theta}_0) = \Lambda(\boldsymbol{x}_t'\boldsymbol{\theta}_0), \\ f(y_t = 0|\boldsymbol{x}_t; \boldsymbol{\theta}_0) = 1 - \Lambda(\boldsymbol{x}_t'\boldsymbol{\theta}_0), \end{cases} \tag{8.1.1}$$

где Λ — функция логистического распределения:

$$\Lambda(v) \equiv \frac{\exp(v)}{1 + \exp(v)}. \tag{8.1.2}$$

При этом функция плотности $f(y_t|\boldsymbol{x}_t; \boldsymbol{\theta})$ может быть записана как

$$f(y_t|\boldsymbol{x}_t; \boldsymbol{\theta}_0) = \Lambda(\boldsymbol{x}_t'\boldsymbol{\theta}_0)^{y_t} \times [1 - \Lambda(\boldsymbol{x}_t'\boldsymbol{\theta}_0)]^{1-y_t}. \tag{8.1.3}$$

Беря логарифм от обеих частей этого выражения и заменяя истинное значение параметра $\boldsymbol{\theta}_0$ его гипотетическим значением $\boldsymbol{\theta}$, мы получаем логарифмическое правдоподобие для наблюдения t:

$$\log f(y_t|\boldsymbol{x}_t; \boldsymbol{\theta}) = y_t \log \Lambda(\boldsymbol{x}_t'\boldsymbol{\theta}) + (1 - y_t) \log[1 - \Lambda(\boldsymbol{x}_t'\boldsymbol{\theta})]. \tag{8.1.4}$$

Целевая функция логит-модели $Q_n(\theta)$ равна умноженному на $1/n$ логарифмическому правдоподобию выборки $(y_1, \boldsymbol{x}_1, y_2, \boldsymbol{x}_2, \ldots, y_n, \boldsymbol{x}_n)$. В предположении, что $\{y_t, \boldsymbol{x}_t\}$ — i.i.d., логарифмическое правдоподобие выборки представляет собой сумму по t логарифмов правдоподобия для наблюдения t. Следовательно,

$$Q_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{t=1}^{n} \{y_t \log \Lambda(\boldsymbol{x}_t'\boldsymbol{\theta}) + (1 - y_t) \log[1 - \Lambda(\boldsymbol{x}_t'\boldsymbol{\theta})]\}. \tag{8.1.5}$$

Ранее мы доказали, что для пробит-модели функция правдоподобия вогнута (пример 7.6) и что ML-оценка состоятельна (пример 7.9) и асимптотически нормальна (пример 7.11), в предположении, что $\mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}_t')$ — невырожденная матрица. Вы вскоре увидите, что вывод этих свойств для логит-модели проще. (Однако при первом прочтении вы, возможно, захотите пропустить обсуждение состоятельности и асимптотической нормальности оценок.)

Вклад и гессиан для наблюдения t

Функция логистического распределения обладает следующим удобным свойством:

$$\Lambda'(v) = \Lambda(v)(1 - \Lambda(v)), \quad \Lambda''(v) = [1 - 2\Lambda(v)]\Lambda(v)[1 - \Lambda(v)]. \tag{8.1.6}$$

Используя его, несложно вывести (см. контрольный вопрос 1(a)) следующие выражения для функции вклада и гессиана[1] для наблюдения t:

$$\mathbf{s}(\mathbf{w}_t;\boldsymbol{\theta}) \left(\equiv \frac{\partial \log f(y_t|\mathbf{x}_t;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right) = [y_t - \Lambda(\mathbf{x}_t'\boldsymbol{\theta})]\mathbf{x}_t, \tag{8.1.7}$$

$$\mathbf{H}(\mathbf{w}_t;\boldsymbol{\theta}) \left(\equiv \frac{\partial \mathbf{s}(\mathbf{w}_t;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right) = -\Lambda(\mathbf{x}_t'\boldsymbol{\theta})[1-\Lambda(\mathbf{x}_t'\boldsymbol{\theta})]\mathbf{x}_t\mathbf{x}_t', \tag{8.1.8}$$

где $\mathbf{w}_t = (y_t, \mathbf{x}_t')'$.

Состоятельность

Поскольку $\mathbf{x}_t\mathbf{x}_t'$ — положительно полуопределенная матрица, из выражения для $\mathbf{H}(\mathbf{w}_t;\boldsymbol{\theta})$ немедленно вытекает отрицательная полуопределенность гессиана, и поэтому $Q_n(\boldsymbol{\theta})$ вогнута. Следовательно, подходящей теоремой о состоятельности является утверждение 7.6. Мы покажем здесь, что условия этого утверждения выполняются при невырожденности $\mathrm{E}(\mathbf{x}_t\mathbf{x}_t')$. Условие (a) этого утверждения представляет собой идентификацию (условной) функции плотности: $f(y_t|\mathbf{x}_t,\boldsymbol{\theta}) \neq f(y_t|\mathbf{x}_t,\boldsymbol{\theta}_0)$ с положительной вероятностью для $\theta \neq \theta_0$[2]. Поскольку взятие логарифма является строго монотонным преобразованием, это условие эквивалентно

$$\log f(y_t|\mathbf{x}_t;\boldsymbol{\theta}) \neq \log f(y_t|\mathbf{x}_t;\boldsymbol{\theta}_0)$$
$$\text{с положительной вероятностью для } \boldsymbol{\theta} \neq \boldsymbol{\theta}_0, \tag{8.1.9}$$

где $\boldsymbol{\theta}_0$ — истинное значение параметра, и функция плотности f дана в (8.1.3). Проверка этого условия аналогична случаю пробит-модели, рассмотренному в примере 7.8, и производится следующим образом. При обсуждении примера 7.8 было установлено, что

$$\mathrm{E}(\mathbf{x}_t\mathbf{x}_t') \text{ невырождена } \Rightarrow \mathbf{x}_t'\boldsymbol{\theta} \neq \mathbf{x}_t'\boldsymbol{\theta}_0$$
$$\text{с положительной вероятностью для } \boldsymbol{\theta} \neq \boldsymbol{\theta}_0. \tag{8.1.10}$$

Поскольку логистическая функция распределения $\Lambda(v)$ — строго монотонная функция от v, мы имеем $\Lambda(\mathbf{x}_t'\boldsymbol{\theta}) \neq \Lambda(\mathbf{x}_t'\boldsymbol{\theta}_0)$, когда $\mathbf{x}_t'\boldsymbol{\theta} \neq \mathbf{x}_t'\boldsymbol{\theta}_0$. Таким образом, условие (a) следует из невырожденности $\mathrm{E}(\mathbf{x}_t\mathbf{x}_t')$. Переходя к условию (b) (говорящему, что $\mathrm{E}[|\log f(y_t|\mathbf{x}_t;\boldsymbol{\theta})|] < \infty$ для всех $\boldsymbol{\theta}$), легко показать, что

$$|\log\Lambda(v)| \leq |\log\Lambda(0)| + |v|. \tag{8.1.11}$$

[1]Напомним, что гессианом автор называет *матрицу* вторых производных. — *Прим. науч. ред. перевода.*

[2]Другими словами, пусть $X \equiv f(y_t|\mathbf{x}_t,\theta)$ и $Y \equiv f(y_t|\mathbf{x}_t,\theta_0)$. X и Y — случайные величины, так как они являются функциями от $\mathbf{w}_t \equiv (y_t,\mathbf{x}_t')'$. Условие может быть переформулировано как $\mathrm{Prob}(X \neq Y) > 0$ для $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$.

Условие (b) можно проверить аналогично проверке такого же условия для пробита в примере 7.9. Таким образом, как и в пробит-модели, невырожденность $\mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}_t')$ обеспечивает состоятельность оценки ML в логит-модели.

Асимптотическая нормальность

Чтобы доказать асимптотическую нормальность при том же условии, мы проверим выполнение условий утверждения 7.9. Условие (1) этого утверждения выполняется, если в качестве пространства параметров Θ берется $\mathbb{R}^p$. Условие (2) выполняется очевидным образом. Условие (3) легко можно проверить следующим образом. Так как $\mathrm{E}[y_t|\boldsymbol{x}_t] = \Lambda(\boldsymbol{x}_t'\theta)$, то из (8.1.7) следует $\mathrm{E}[\boldsymbol{s}(\boldsymbol{w}_t;\theta_0)|\boldsymbol{x}_t] = 0$. А значит, $\mathrm{E}[\boldsymbol{s}(\boldsymbol{w}_t;\theta_0)] = 0$, по закону полных математических ожиданий. Читателю оставляется в качестве контрольного вопроса 1(b) вывод равенства для условных информационных матриц

$$\mathrm{E}\big[\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)\,\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)'|\boldsymbol{x}_t\big] = -\,\mathrm{E}\big[\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta}_0)|\boldsymbol{x}_t\big]. \tag{8.1.12}$$

Следовательно, (3) выполняется по закону полных математических ожиданий. Условие локального доминирования (условие (4)) для логит-модели проверить легко, так как поскольку $|\Lambda(\boldsymbol{x}_t'\theta)[1-\Lambda(\boldsymbol{x}_t'\theta)]| < 1$, то $\|\boldsymbol{H}_t(\boldsymbol{w}_t;\theta)\| \le \|\boldsymbol{x}_t\boldsymbol{x}_t'\|$ для всех θ. Математическое ожидание $\|\boldsymbol{x}_t\boldsymbol{x}_t'\|^2$ (суммы квадратов элементов $\widehat{\boldsymbol{x}}_t\widehat{\boldsymbol{x}}_t'$) конечно, если $\mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}_t')$ — невырожденная матрица. Поэтому конечно $\mathrm{E}(\|\boldsymbol{x}_t\boldsymbol{x}_t'\|)$. Что касается условия (5), то аргумент в сноске 29 Ньюи и Мак-Фаддена [Newey and McFadden, 1994]), использовавшийся для проверки того же условия в пробит-модели, также может быть использован и для логит-модели[1].

Таким образом, мы приходим к выводу, что если $\{y_t,\boldsymbol{x}_t\} \sim$ i.i.d. и $\mathrm{E}(\boldsymbol{x}\boldsymbol{x}')$ — невырожденная матрица, то тогда в логит-модели ML-оценка $\widehat{\boldsymbol{\theta}}$ истинного значения параметра $\boldsymbol{\theta}_0$ является состоятельной и асимптотически нормальной с асимптотической ковариационной матрицей, задаваемой выражением:

$$\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\theta}})} = -\left[\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{H}(\boldsymbol{w}_t;\widehat{\boldsymbol{\theta}})\right]^{-1}, \tag{8.1.13}$$

где $\boldsymbol{w}_t = (y_t,\boldsymbol{x}_t')'$, и $\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta})$ задается в (8.1.8)).

---

[1]Мы приведем доказательство только для заинтересованных читателей. Для любого $\overline{v} > 0$, существует C, такое, что $\Lambda(v)[1-\Lambda(v)] \ge C > 0$ для $|v| \le \overline{v}$. Поэтому

$$\begin{aligned}\mathrm{E}\left[\Lambda(\boldsymbol{x}'\theta)[1-\Lambda(\boldsymbol{x}'\theta)]\,\boldsymbol{x}\boldsymbol{x}'\right] &\ge \mathrm{E}\left[1(|\boldsymbol{x}'\theta| \le \overline{v})\Lambda(\boldsymbol{x}'\theta)[1-\Lambda(\boldsymbol{x}'\theta)]\boldsymbol{x}\boldsymbol{x}'\right] \ge \\ &\ge C\,\mathrm{E}\left[1(|\boldsymbol{x}'\theta| \le \overline{v})\boldsymbol{x}\boldsymbol{x}'\right].\end{aligned}$$

где все вышеописанное — матричные неравенства, $1(|c| \le \overline{v})$ — индикаторная функция. Последний член положительно определен (а не просто положительно полуопределен) для достаточно больших $\overline{v}$ вследствие невырожденности $\mathrm{E}(\boldsymbol{x}\boldsymbol{x}')$.

Контрольные вопросы

1. (Функция вклада и гессиан для наблюдения t для функции плотности общего вида.) Замените логистическую функцию распределения $\Lambda(\boldsymbol{x}_t'\boldsymbol{\theta}_0)$ в (8.1.1) на произвольную функцию распределения $F(\boldsymbol{x}_t'\boldsymbol{\theta}_0)$.

   (a) Проверьте, что функция вклада и гессиан для наблюдения t задаются следующим образом:

   $$\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\theta}) = \frac{y_t - F_t}{F_t \cdot (1 - F_t)} f_t \boldsymbol{x}_t,$$

   $$\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta}) = -\left[\frac{y_t - F_t}{F_t \cdot (1 - F_t)}\right]^2 f_t^2 \boldsymbol{x}_t \boldsymbol{x}_t' + \left[\frac{y_t - F_t}{F_t \cdot (1 - F_t)}\right] f_t' \boldsymbol{x}_t \boldsymbol{x}_t',$$

   где $F_t = F(\boldsymbol{x}_t'\boldsymbol{\theta}_0)$, $f_t = f(\boldsymbol{x}_t'\boldsymbol{\theta}_0)$ и $f(\cdot) = F'(\cdot)$ — функция плотности.

   (b) Проверьте равенство для условных информационных матриц (8.1.12) для функции распределения F общего вида. **Указание:** *Покажите, что математическое ожидание второго члена выражения для гессиана в предыдущем вопросе равно нулю.*

2. Проверьте (8.1.7) и (8.1.8).

3. (Логит-модель для сериально коррелированных наблюдений.) Предположим, что последовательность $\{y_t, \boldsymbol{x}_t\}$ является стационарной и эргодической, но необязательно i.i.d. Состоятельна ли ML-оценка для логит-модели, описанная в тексте? [Ответ: Да. Так как утверждение 7.6, применимое в данном случае, не требует случайности выборки. Однако выражение для $\mathrm{Avar}(\widehat{\theta})$ нельзя получить как взятую со знаком «минус» матрицу, обратную к выборочному гессиану.]

8.2. Усеченные регрессионные модели

Модель с ограниченной зависимой переменной (LDV) (*limited dependent variable model*) представляет собой регрессионную модель, в которой или зависимая переменная y_t ограничена каким-либо образом, или наблюдения, для которых y_t не удовлетворяет некоторому предустановленному критерию, исключаются из выборки. LDV-модель первого типа называется **цензурированной моделью регрессии** (*censored regression model*), а модель второго типа — **усеченной моделью регрессии** (*truncated regression model*). В этом параграфе представлены усеченные регрессионные модели; цензурированные модели регрессии рассматриваются в следующем параграфе.

Модель

Пусть $\{y_t, \boldsymbol{x}_t\}$ i.i.d., и

$$y_t = \boldsymbol{x}_t'\beta_0 + \varepsilon_t, \quad \varepsilon_t | \boldsymbol{x}_t \sim N(0, \sigma_0^2). \tag{8.2.1}$$

Эта модель была бы в точности стандартной линейной регрессионной моделью с нормально распределенными ошибками, если бы не следующая особенность: в выборку включены только те наблюдения, для которых y_t удовлетворяет некоторому заранее установленному критерию. Мы рассмотрим только простейшее правило усечения: $y_t > c$, где c — известная константа. Это правило часто называют «усечением снизу». После рассмотрения этого частного случая несложно обобщить его на более общие правила усечения.

Усеченные распределения

Для продолжения изложения нам необходимы два результата из теории вероятностей.

Функция плотности усеченной случайной величины: *Если непрерывная случайная величина y имеет функцию плотности $f(y)$ и c — некоторая константа, то функция плотности после усечения $y > c$ определена на интервале (c, ∞) и задается следующим образом:*

$$f(y|y > c) = \frac{f(y)}{\text{Prob}(y > c)}. \tag{8.2.2}$$

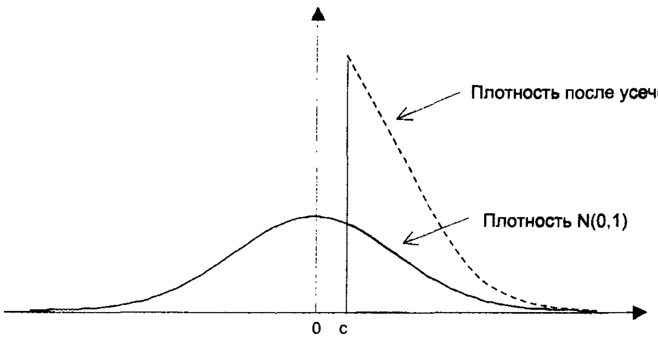

Рис. 8.1. Влияние усечения

На рис 8.1 показано, как усечение изменяет распределение для $N(0, 1)$. Сплошная кривая изображает плотность вероятности $N(0, 1)$. Поскольку это плотность вероятности, площадь под этой кривой равна единице. Пунктирная кривая изображает усеченную функцию плотности на интервале (c, ∞). Она находится выше сплошной кривой, так что площадь под пунктирной кривой также равна единице.

Далее приведен второй результат, который будет полезен в дальнейшем.

Моменты усеченного нормального распределения:
Если $y \sim N(\mu_0, \sigma_0^2)$ и c — константа, то математическое ожидание и дисперсия усеченного распределения равны соответственно

$$\mathrm{E}(y|y > c) = \mu_0 + \sigma_0\lambda(v), \tag{8.2.3}$$

$$\mathrm{Var}(y|y > c) = \sigma_0^2\{1 - \lambda(v)[\lambda(v) - v]\}, \tag{8.2.4}$$

где $v = (c - \mu_0)/\sigma_0$ и $\lambda(v) \equiv \frac{\phi(v)}{1-\Phi(v)}$.

Эта формула для математического ожидания ясно показывает, что выборочное среднее выборки из усеченного распределения не состоятельно в отношении μ_0. Это пример **смещения, обусловленного селективностью выборки** (*sample selection bias*). Это смещение равно $\sigma_0\lambda(v)$. Функция $\lambda(v)$ называется **обратным отношением Миллса** (*inverse Mill's ratio*), или **функцией риска** (*hazard function*). Функция $\lambda(v)$ выпукла и имеет асимптотами v при $v \to \infty$ и нуль при $v \to -\infty$. Поэтому ее производная $\lambda'(v)$ лежит в пределах от 0 до 1. Легко показать, что

$$\lambda'(v) = \lambda(v)(\lambda(v) - v). \tag{8.2.5}$$

Используя (8.2.3) и (8.2.4), мы можем продемонстрировать, как усечение изменяет форму регрессии y_t на $\boldsymbol{x}_t$. Из (8.2.1) следует, что распределением y_t при условии $\boldsymbol{x}_t$ *до* усечения выборки является $N(\boldsymbol{x}_t'\boldsymbol{\beta}_0, \sigma_0^2)$. Наблюдение t присутствует в выборке, если и только если $y_t > c$. Поэтому

$$\mathrm{E}(y_t|\boldsymbol{x}_t, t \text{ в выборке}) = \boldsymbol{x}_t'\boldsymbol{\beta}_0 + \sigma_0\lambda\left(\frac{c-\boldsymbol{x}_t'\boldsymbol{\beta}_0}{\sigma_0}\right), \tag{8.2.6}$$

$$\begin{aligned}&\mathrm{Var}(y_t|\boldsymbol{x}_t, t \text{ в выборке}) = \\ &= \sigma_0^2\left\{1 - \lambda\left(\frac{c-\boldsymbol{x}_t'\boldsymbol{\beta}_0}{\sigma_0}\right)\left[\lambda\left(\frac{c-\boldsymbol{x}_t'\boldsymbol{\beta}_0}{\sigma_0}\right) - \frac{c-\boldsymbol{x}_t'\boldsymbol{\beta}_0}{\sigma_0}\right]\right\}.\end{aligned} \tag{8.2.7}$$

Эта формула показывает, что OLS-оценка коэффициента при $\boldsymbol{x}_t$ из линейной регрессии y_t на $\boldsymbol{x}_t$ несостоятельна, так как дополнительная составляющая $\sigma_0\lambda\left(\frac{c-\boldsymbol{x}_t'\boldsymbol{\beta}_0}{\sigma_0}\right)$ (которая коррелирована с $\boldsymbol{x}_t$) будет включена в ошибку линейной регрессии. Поскольку функциональная форма функции риска $\lambda(v)$ известна (при предположении о нормальности ε_t), мы можем избежать смещения, обусловленного селективностью выборки, применяя нелинейный метод наименьших квадратов для оценивания параметров $(\boldsymbol{\beta}_0, \sigma_0^2)$, однако ML-оценка более предпочтительна, так как метод максимального правдоподобия является асимптотически более эффективным.

Заметим попутно, что проблема смещения, обусловленного селективностью выборки, не возникает, если отбор основан на значениях регрессоров, а не на значениях зависимой переменной. Предположим, что наблюдение t включается в выборку, если $\phi(\boldsymbol{x}_t) \geq 0$. Тогда

$$\begin{aligned}\mathrm{E}(y_t|\boldsymbol{x}_t, t \text{ в выборке}) &= \mathrm{E}[\boldsymbol{x}_t'\beta_0 + \varepsilon_t|\boldsymbol{x}_t, \phi(\boldsymbol{x}_t) \geq 0] = \\ &= \boldsymbol{x}_t'\beta_0 + \mathrm{E}\left[\varepsilon_t|\boldsymbol{x}_t, \phi(\boldsymbol{x}_t) \geq 0\right] = \\ &= \boldsymbol{x}_t'\beta_0. \end{aligned} \tag{8.2.8}$$

Последнее равенство выполнено, так как $\mathrm{E}[\varepsilon_t|\boldsymbol{x}_t, \phi(\boldsymbol{x}_t) \geq 0] = \mathrm{E}(\varepsilon_t|\boldsymbol{x}_t) = 0$, если $\phi(\boldsymbol{x}_t) \geq 0$.

Функция правдоподобия

При выводе функции правдоподобия используется уравнение (8.2.2). Как мы только что заметили, распределение $y_t|\boldsymbol{x}_t$ до усечения — это нормальное распределение $N(\boldsymbol{x}_t'\beta_0, \sigma_0^2)$ с функцией плотности

$$\frac{1}{\sqrt{2\pi\sigma_0^2}}\exp\left[-\frac{1}{2}\left(\frac{y_t - \boldsymbol{x}_t'\beta_0}{\sigma_0}\right)^2\right] = \frac{1}{\sigma_0}\phi\left(\frac{y_t - \boldsymbol{x}_t'\beta_0}{\sigma_0}\right), \tag{8.2.9}$$

где $\phi(\cdot)$ — функция плотности стандартного нормального распределения $N(0,1)$. Вероятность того, что наблюдение t представлено в выборке, равна

$$\begin{aligned}\mathrm{Prob}(y_t > c|\boldsymbol{x}_t) &= 1 - \mathrm{Prob}(y_t \leq c|\boldsymbol{x}_t) = \\ &= 1 - \mathrm{Prob}\left(\frac{y_t - \boldsymbol{x}_t'\beta_0}{\sigma_0} \leq \frac{c - \boldsymbol{x}_t'\beta_0}{\sigma_0} \;\Big|\; \boldsymbol{x}_t\right) = \\ &= 1 - \Phi\left(\frac{c - \boldsymbol{x}_t'\beta_0}{\sigma_0}\right) \quad \left(\text{поскольку } \tfrac{y_t - \boldsymbol{x}_t'\beta_0}{\sigma_0} \;\big|\; \boldsymbol{x}_t \sim N(0,1)\right). \end{aligned} \tag{8.2.10}$$

Поэтому из (8.2.2) следует, что функция плотности после усечения выборки, определенная на интервале (c, ∞), равна:

$$f(y_t|\boldsymbol{x}_t, t \text{ в выборке}: \beta_0, \sigma_0^2) = \frac{\frac{1}{\sigma_0}\phi\left(\frac{y_t - \boldsymbol{x}_t'\beta_0}{\sigma_0}\right)}{1 - \Phi\left(\frac{c - \boldsymbol{x}_t'\beta_0}{\sigma_0}\right)}. \tag{8.2.11}$$

Это функция плотности (условная по отношению к $\boldsymbol{x}_t$) для наблюдения t, *присутствующего в выборке*. Прологарифмировав и заменив (β_0, σ_0^2) на их гипотетические значения $(\boldsymbol{\beta}, \sigma^2)$, получим логарифмическое условное правдоподобие для наблюдения t:

$$\begin{aligned}&\log f(y_t|\boldsymbol{x}_t; \boldsymbol{\beta}, \sigma^2) = \\ &= \left\{-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{1}{2}\left(\frac{y_t - \boldsymbol{x}_t'\boldsymbol{\beta}}{\sigma}\right)^2\right\} - \\ &- \log\left[1 - \Phi\left(\frac{c - \boldsymbol{x}_t'\boldsymbol{\beta}}{\sigma}\right)\right]. \end{aligned} \tag{8.2.12}$$

Если бы не последняя составляющая, это выражение являлось бы логарифмической функцией правдоподобия (при условии $\boldsymbol{x}_t$) для обычной линейной регрессионной модели. Однако последняя составляющая в этом выражении необходима, так как значение зависимой переменной для наблюдения t, прошедшее тест $y_t > c$ и, следовательно, присутствующее в выборке, всегда больше c.

Репараметризация функции правдоподобия

Чтобы упростить анализ, логарифмическую функцию правдоподобия можно репараметризовать. Для этого введем обозначения:

$$\boldsymbol{\delta} = \boldsymbol{\beta}/\sigma, \quad \gamma = 1/\sigma. \tag{8.2.13}$$

(Мы уже использовали такую репараметризацию в примере 7.7 для линейной регрессионной модели.) Это взаимно однозначное соответствие между $(\boldsymbol{\beta}, \sigma^2)$ и $(\boldsymbol{\delta}, \gamma)$, обратное соответствие дается формулами: $\boldsymbol{\beta} = \boldsymbol{\delta}/\gamma$, $\sigma^2 = 1/\gamma^2$. После репараметризации логарифмическое условное правдоподобие принимает вид:

$$\begin{aligned}
&\log \widetilde{f}(y_t|\boldsymbol{x}_t; \boldsymbol{\lambda}) = \\
&= \Big[-\frac{1}{2}\log(2\pi) + \log(\gamma) - \\
&- \frac{1}{2}(\gamma y_t - \boldsymbol{x}_t'\boldsymbol{\delta})^2\Big] - \log[1 - \Phi(\gamma c - \boldsymbol{x}_t'\boldsymbol{\delta})].
\end{aligned} \tag{8.2.14}$$

Целевая функция ML представляет собой умноженное на $1/n$ логарифмическое правдоподобие выборки. Для случайной выборки логарифмическое правдоподобие выборки равно сумме по t логарифмических правдоподобий для наблюдений t. Таким образом, целевая функция в ML-оценивании репараметризованной усеченной регрессионной модели, обозначенная как $Q_n(\boldsymbol{\delta}, \gamma)$, представляет собой среднее по t выражений (8.2.14). Оценка максимального правдоподобия $(\widehat{\boldsymbol{\delta}}, \widehat{\gamma})$ истинных параметров $(\boldsymbol{\delta}_0, \gamma_0)$ — это такие значения $(\boldsymbol{\delta}, \gamma)$, которые максимизируют эту целевую функцию.

Проверка состоятельности и асимптотической нормальности

Следующая задача состоит в выводе вклада и гессиана и проверке выполнения условий состоятельности и асимптотической нормальности для репараметризованного логарифмического правдоподобия (8.2.14). (При первом прочтении вам, возможно, захочется пропустить эту часть.)

Вклад и гессиан для наблюдения t

Утомительные вычисления дают:

$$\underset{((K+1)\times 1)}{s(\boldsymbol{w}_t; \boldsymbol{\delta}, \gamma)} = \begin{bmatrix} (\gamma y_t - \boldsymbol{x}_t'\boldsymbol{\delta})\boldsymbol{x}_t \\ \frac{1}{\gamma} - (\gamma y_t - \boldsymbol{x}_t'\boldsymbol{\delta}) y_t \end{bmatrix} + \lambda(v_t) \begin{bmatrix} -\boldsymbol{x}_t \\ c \end{bmatrix}, \tag{8.2.15}$$

$$\underset{((K+1)\times(K+1))}{\boldsymbol{H}(\boldsymbol{w}_t; \boldsymbol{\delta}, \gamma)} = - \begin{bmatrix} \boldsymbol{x}_t\boldsymbol{x}_t' & -y_t\boldsymbol{x}_t \\ -y_t\boldsymbol{x}_t' & \frac{1}{\gamma^2} + y_t^2 \end{bmatrix} + \\ + \lambda(v_t)[\lambda(v_t) - v_t] \begin{bmatrix} \boldsymbol{x}_t\boldsymbol{x}_t' & -c\boldsymbol{x}_t \\ -c\boldsymbol{x}_t' & c^2 \end{bmatrix}, \tag{8.2.16}$$

где K — количество регрессоров, $\boldsymbol{w}_t = (y_t, \boldsymbol{x}_t')'$ и $v_t \equiv \gamma c - \boldsymbol{x}_t'\delta \left(= \frac{c - \boldsymbol{x}_t'\beta}{\sigma}\right)$. Гессиан отрицательно полуопределен *не на всей области определения* даже после репараметризации. Следовательно, целевая функция $\widetilde{Q}_n(\delta, \gamma)$ не является вогнутой[1]. Поэтому в данном случае подходящей теоремой о состоятельности является утверждение 7.6 (а не утверждение 7.5), тогда как подходящей теоремой об асимптотической нормальности остается утверждение 7.9.

Состоятельность

Условие (a) утверждения 7.5 суть идентифицируемость (условной) функции плотности:

$$\widetilde{f}(y_t|\boldsymbol{x}_t; \delta, \gamma) \neq \widetilde{f}(y_t|\boldsymbol{x}_t; \delta_0, \gamma_0)$$
$$\text{с положительной вероятностью для } (\delta, \gamma) \neq (\delta_0, \gamma_0),$$

где (δ_0, γ_0) — истинные значения параметров, а репараметризованный логарифм функции плотности $\widetilde{f}(y_t|\boldsymbol{x}_t; \delta, \gamma)$ представлен в (8.2.14). Очевидно, что при заданных $(y_t, \boldsymbol{x}_t)$ значение $\widetilde{f}(y_t|\boldsymbol{x}_t; \delta, \gamma)$ отличается для различных значений γ. Поэтому идентификация сводится к условию: $\boldsymbol{x}_t'\delta \neq \neq \boldsymbol{x}_t'\delta_0$ с положительной вероятностью для $\delta \neq \delta_0$. Но (8.1.10) указывает, что это условие выполнено, если матрица $\mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}_t')$ невырождена. Переходя к условию (b) утверждения (условие доминирования для

[1]Однако можно показать, что гессиан целевой функции $Q_n(\theta)$ отрицательно полуопределен в любой точке, представляющей решение уравнений правдоподобия (см.: [Orme, 1989]. Так как это исключает наличие седловых точек и локальных минимумов, ситуация, когда существует несколько локальных максимумов, невозможна. Поэтому (при условии, что локальный максимум $\widetilde{Q}_n(\delta, \gamma)$ достигается во внутренней точке пространства параметров) условия первого порядка являются необходимыми и достаточными для существования *глобального* максимума. Если некоторый алгоритм (такой как алгоритм Ньютона - Рафсона) сходится, найденное решение является глобальным максимумом.

$\widetilde{f}(y_t|\boldsymbol{x}_t;\boldsymbol{\delta},\gamma))$, легко показать, что это условие выполнено при предположении о невырожденности $\mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}_t')$[1]. Поэтому ML-оценка $(\widehat{\boldsymbol{\delta}},\widehat{\gamma})$ истинных значений параметров $(\boldsymbol{\delta},\gamma)$ является состоятельной при предположении о невырожденности $\mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}_t')$.

Асимптотическая нормальность

Используя (8.2.6), (8.2.7), (8.2.13), с помощью довольно громоздких вычислений можно показать, что $\mathrm{E}[\boldsymbol{s}|\boldsymbol{x}_t] = 0$. Чрезвычайно громоздкие вычисления приводят к равенству для условной информационной матрицы: $\mathrm{E}[\boldsymbol{s}\boldsymbol{s}'|\boldsymbol{x}_t] = -\mathrm{E}[\boldsymbol{H}|\boldsymbol{x}_t]$. Поэтому условие (3) утверждения 7.9 выполняется. Мы не будем обсуждать проверку условий (4) и (5) этого утверждения[2].

Итак, можно сделать следующий вывод: если $\{y_t,\boldsymbol{x}_t\}$ — последовательность независимых и одинаково распределенных случайных величин и $\mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}_t')$ — невырожденная матрица, то тогда ML-оценка параметров $(\widehat{\boldsymbol{\delta}},\widehat{\gamma})$ усеченной регрессионной модели является состоятельной и асимптотически нормально распределенной с асимптотической ковариационной матрицей, состоятельно оцениваемой посредством

$$\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\delta}},\widehat{\gamma})} = -\left[\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{H}(\boldsymbol{w}_t;\widehat{\boldsymbol{\delta}},\widehat{\gamma})\right]^{-1}, \tag{8.2.17}$$

где $\boldsymbol{w}_t = (y_t,\boldsymbol{x}_t')'$ и $\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\theta})$ определяется выражением (8.2.16).

Восстановление исходных параметров

Вследствие инвариантности оценок ML (см. параграф 7.1), ML-оценка для $(\boldsymbol{\beta}_0,\sigma_0^2)$ может быть посредством обратного отображения: $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\delta}}/\widehat{\gamma}$ и $\widehat{\sigma}^2 = 1/\widehat{\gamma^2}$. Если ML-оценка репараметризованной модели $(\widehat{\boldsymbol{\delta}},\widehat{\gamma})$ состоятельна, то таковой же будет и ML-оценка $(\widehat{\boldsymbol{\beta}},\widehat{\sigma}^2)$, восстановленная таким

---

[1]Только для заинтересованных читателей: необходимо показать, что

$$\mathrm{E}[\sup_{\theta\in\Theta}|\log\widetilde{f}(y_t|\boldsymbol{x}_t;\boldsymbol{\theta})|] < \infty,$$

где $\boldsymbol{\theta} = (\boldsymbol{\delta}',\gamma')'$ и Θ — компакт. Расширение аргументации примера 7.8 показывает, что существует доминирующая функция для выражения в скобках в уравнении (8.2.14). Чтобы показать существование доминирующей функции для $\log[1-\Phi(c\gamma-\boldsymbol{x}_t'\boldsymbol{\delta})]$, можно использовать неравенство (7.2.14).

[2]Похоже, что не существует опубликованных работ, проверяющих выполнение условий (4) и (5). Для случая, когда $\{\boldsymbol{x}_t\}$ — последовательность фиксированных констант, Сапра [Sapra, 1992] доказал асимптотическую нормальность в предположении, что $\boldsymbol{x}_t$ ограничена и $\lim_{n\to\infty}\sum_{t=1}^{n}\boldsymbol{x}_t\boldsymbol{x}_t'$ — невырожденная матрица. (Его доказательство дано для более общего случая наличия сериальной корреляции наблюдений.) Представляется разумным предположить, что для случая стохастических объясняющих переменных $\boldsymbol{x}_t$ (как здесь) достаточным условием состоятельности и асимптотической нормальности является невырожденность матрицы $\mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}_t')$.

образом. Асимптотическая ковариационная матрица $(\widehat{\beta}, \widehat{\sigma}^2)$ может быть получена дельта-методом (см. лемма 2.5) как $\widehat{J}\widehat{\mathrm{Avar}}(\widehat{\delta}, \widehat{\gamma})\widehat{J}'$, где $\widehat{J}$ — оценка матрицы Якоби для обратного отображения ($\beta = \delta/\gamma, \sigma^2 = 1/\gamma^2$), вычисленная в точке $(\widehat{\delta}, \widehat{\gamma})$:

$$\widehat{J} = \begin{bmatrix} \frac{1}{\widehat{\gamma}}I & -\frac{\widehat{\delta}}{\widehat{\gamma}^2} \\ 0' & -\frac{2}{\widehat{\gamma}^3} \end{bmatrix}. \tag{8.2.18}$$

Контрольные вопросы

1. (Второй момент y_t.) Из (8.2.6) и (8.2.7) выведите:

$$\mathrm{E}(y_t^2|x_t,\ t \text{ присутствует в выборке}) = \\ = \frac{1}{\gamma^2}\left[1 + \lambda(v_t)\gamma c + (x_t'\delta)^2 + \lambda(v_t)x_t'\delta\right],$$

где $v_t \equiv \gamma c - x_t'\delta$.

2. (Оценка усеченной регрессионной модели посредством GMM.) Усеченная модель регрессии также может быть оценена с помощью GMM. Рассмотрим $K + 1$ условий ортогональности. Первые K из них выглядят следующим образом:

$$\underset{(K\times 1)}{g_1(w_t; \delta, \gamma)} = \left(y_t - \frac{1}{\gamma}x_t'\delta - \frac{\lambda(v_t)}{\gamma}\right) \cdot x_t. \tag{8.2.19}$$

где $w_t = (y_t, x_t')'$ и $v_t \equiv \gamma c - x_t'\delta$. $(K + 1)$-е условие ортогональности имеет следующий вид:

$$g_2(w_t; \delta, \gamma) = y_t^2 - \frac{1}{\gamma^2}\left[1 + \lambda(v_t)\gamma c + (x_t'\delta)^2 + \lambda(v_t)x_t'\delta\right]. \tag{8.2.20}$$

(a) Проверьте, что $\mathrm{E}\left[g_1(w_t; \delta_0, \gamma_0)\right] = 0$ и $\mathrm{E}\left[g_2(w_t; \delta_0, \gamma_0)\right] = 0$. **Указание:** *Используйте закон полных математических ожиданий. v_t является функцией от x_t.*

(b) Покажите, что пара (δ^*, γ^*) удовлетворяет уравнениям правдоподобия тогда и только тогда, когда она удовлетворяет указанным $K + 1$ условиям ортогональности. **Указание:** *Пусть $s_2(w_t; \delta, \gamma)$ — $(K+1)$-й элемент $s(w_t; \delta, \gamma)$ в (8.2.15). Тогда*

$$s_2(w_t; \delta, \gamma) + \gamma g_2(w_t; \delta, \gamma) = (x_t'\delta)\left(y_t - \frac{1}{\gamma}x_t'\delta - \frac{\lambda(v_t)}{\gamma}\right).$$

8.3. Цензурированные регрессионные (тобит) модели

Цензурированная модель регрессии называется также **моделью тобит** (*Tobit model*) в честь Дж. Тобина [Tobin, 1958], который был первым, кто ввел понятие цензурирования в экономику. Простейшая тобит-модель может быть записана следующим образом:

$$y_t^* = \boldsymbol{x}_t'\beta_0 + \varepsilon_t, \quad t = 1, 2, \ldots, n, \tag{8.3.1}$$

$$y_t = \begin{cases} y_t^* & \text{если } y_t^* > c, \\ c & \text{если } y_t^* \leq c. \end{cases} \tag{8.3.2}$$

Здесь $\varepsilon_t|\boldsymbol{x}_t$ распределены нормально с параметрами $N(0, \sigma_0^2)$, и $\{y_t, \boldsymbol{x}_t\}$ $(t = 1, 2, \ldots, n)$ — i.i.d. Пороговое значение c известно. В отличие от усеченной модели регрессии, здесь усечения нет: наблюдения, для которых значение зависимой переменной y_t^* не удовлетворяет критерию $y_t^* > c$, *присутствуют* в выборке. Основное отличие тобит-модели от обычной линейной регрессионной модели состоит в том, что зависимая переменная цензурирована (то есть ограничена каким-то диапазоном значений), что приводит к необходимости различать наблюдаемую зависимую переменную y_t и ненаблюдаемую (латентную) переменную y_t^*; y_t — цензурированное значение y_t^*. Модель можно переписать в равносильном виде:

$$y_t = \max\{\boldsymbol{x}_t'\beta_0 + \varepsilon_t, c\}. \tag{8.3.3}$$

Функция правдоподобия для тобит-модели

Для наблюдений, не затронутых цензурированием Тобин [Tobin, 1958], назвал такие наблюдения **nonlimit observations**), функция плотности y_t (условная относительно $\boldsymbol{x}_t$) записывается как

$$\frac{1}{\sigma_0}\phi\left(\frac{y_t - \boldsymbol{x}_t'\beta_0}{\sigma_0}\right). \tag{8.3.4}$$

Эта функция плотности отличается от плотности для усеченной модели (8.2.11), поскольку в случае тобит-модели выборка не усечена. Для тех наблюдений в которых значение зависимой переменной изменяется из-за цензурирования (**limit observations**), все, что мы можем знать, — это то, что $y_t^* \leq c$, вероятность чего равна:

$$\begin{aligned} \text{Prob}(y_t^* \leq c|\boldsymbol{x}_t) &= \text{Prob}\left(\frac{y_t^* - \boldsymbol{x}_t'\beta_0}{\sigma_0} \leq \frac{c - \boldsymbol{x}_t'\beta_0}{\sigma_0} \mid \boldsymbol{x}_t\right) = \\ &= \Phi\left(\frac{c - \boldsymbol{x}_t'\beta_0}{\sigma_0}\right) \quad \left(\text{так как } \tfrac{y_t^* - \boldsymbol{x}_t'\beta_0}{\sigma_0}|\boldsymbol{x}_t \sim N(0, 1)\right). \end{aligned} \tag{8.3.5}$$

Поэтому функция плотности y_t, определенная на интервале $[c, \infty)$, описывается уравнением (8.3.4) для $y_t > c$, а в точке $y_t = c$ сосредоточена вероятностная масса $\Phi\left(\frac{c - x_t'\beta_0}{\sigma_0}\right)$. Эта функция плотности может быть записана как

$$\left[\frac{1}{\sigma_0}\phi\left(\frac{y_t - x_t'\beta_0}{\sigma_0}\right)\right]^{1-D_t} \times \left[\Phi\left(\frac{c - x_t'\beta_0}{\sigma_0}\right)\right]^{D_t}, \tag{8.3.6}$$

где дамми-переменная D_t является функцией от y_t и определяется как

$$D_t = \begin{cases} 0, & \text{если } y_t > c \text{ (то есть } y_t^* > c), \\ 1, & \text{если } y_t = c \text{ (то есть } y_t^* \leq c). \end{cases} \tag{8.3.7}$$

Переходя к логарифмам и заменяя (β_0, σ_0^2) на их гипотетические значения (β, σ^2), мы получаем логарифмическое условное правдоподобие для наблюдения t:

$$\log f(y_t|x_t; \beta, \sigma^2) = (1 - D_t)\log\left[\frac{1}{\sigma}\phi\left(\frac{y_t - x_t'\beta}{\sigma}\right)\right] + D_t \log\Phi\left(\frac{c - x_t'\beta}{\sigma}\right). \tag{8.3.8}$$

Таким образом, если выборка случайна, то среднее логарифмическое правдоподобие для тобит-модели имеет вид:

$$\begin{aligned} Q_n(\theta) &= \frac{1}{n}\sum_{t=1}^{n}\left\{(1 - D_t)\log\left[\frac{1}{\sigma}\phi\left(\frac{y_t - x_t'\beta}{\sigma}\right)\right] + D_t\log\Phi\left(\frac{c - x_t'\beta}{\sigma}\right)\right\} = \\ &= \frac{1}{n}\sum_{y_t > c}\left\{\log\left[\frac{1}{\sigma}\phi\left(\frac{y_t - x_t'\beta}{\sigma}\right)\right]\right\} + \frac{1}{n}\sum_{y_t = c}\left\{\log\Phi\left(\frac{c - x_t'\beta}{\sigma}\right)\right\}. \end{aligned} \tag{8.3.9}$$

Репараметризация

Так же, как и в случае усеченной модели регрессии, анализ становится проще, если провести репараметризацию (8.2.13). Репараметризованное логарифмическое условное правдоподобие имеет вид:

$$\begin{aligned} \log\widetilde{f}(y_t|x_t; \delta, \gamma) = (1 - D_t)\left\{-\frac{1}{2}\log(2\pi) + \log(\gamma) - \frac{1}{2}(\gamma y_t - x_t'\delta)^2\right\} + \\ + D_t\log\Phi(\gamma c - x_t'\delta). \end{aligned} \tag{8.3.10}$$

Мы уже видели, что выражение в фигурных скобках является вогнутым по (δ, γ) (см. пример 7.7). Мы также видели, что $\log\Phi(v)$ является вогнутой функцией (см. пример 7.6), из чего следует, что $\log\Phi(\gamma c - x_t'\delta)$ вогнута по (δ, γ). То есть репараметризованное логарифмическое правдоподобие для тобит-модели является вогнутой функцией, так как представляет собой неотрицательное взвешенное среднее двух вогнутых функций (это было обнаружено Олсеном [Olsen, 1978]). Как и в случае с пробит- и логит-моделями, подходящей теоремой о состоятельности оценок для тобит-модели является утверждение 7.6.

Чтобы проверить выполнение условий, обеспечивающих состоятельность и асимптотическую нормальность репараметризованного логарифмического правдоподобия, остается только получить выражения для вклада и гессиана.

Вклад и гессиан для наблюдения t

Используя тот факт, что $\lambda(-v) = \frac{\phi(-v)}{1-\Phi(-v)} = \frac{\phi(v)}{\Phi(v)}$, и (8.2.5), можно показать, что

$$\underset{((K+1)\times 1)}{\boldsymbol{s}(\boldsymbol{w}_t;\boldsymbol{\delta},\gamma)} = (1-D_t)\cdot\begin{bmatrix}(\gamma y_t - \boldsymbol{x}_t'\boldsymbol{\delta})\boldsymbol{x}_t \\ \frac{1}{\gamma} - (\gamma y_t - \boldsymbol{x}_t'\boldsymbol{\delta})y_t\end{bmatrix} + D_t\cdot\lambda(-v_t)\begin{bmatrix}-\boldsymbol{x}_t \\ c\end{bmatrix}, \quad (8.3.11)$$

$$\begin{aligned}\underset{((K+1)\times(K+1))}{\boldsymbol{H}(\boldsymbol{w}_t;\boldsymbol{\delta},\gamma)} &= -(1-D_t)\cdot\begin{bmatrix}\boldsymbol{x}_t\boldsymbol{x}_t' & -y_t\boldsymbol{x}_t \\ -y_t\boldsymbol{x}_t' & \frac{1}{\gamma^2}+y_t^2\end{bmatrix} - \\ &\quad - D_t\cdot\lambda(-v_t)[\lambda(-v_t)+v_t]\begin{bmatrix}\boldsymbol{x}_t\boldsymbol{x}_t' & -c\boldsymbol{x}_t \\ -c\boldsymbol{x}_t' & c^2\end{bmatrix},\end{aligned} \quad (8.3.12)$$

где $v_t \equiv \gamma c - \boldsymbol{x}_t'\boldsymbol{\delta}$ $(= \frac{c-\boldsymbol{x}_t'\boldsymbol{\beta}}{\sigma})$.

Состоятельность и асимптотическая нормальность

Для тобит-модели мы не будем проверять выполнение условий, обеспечивающих состоятельность и асимптотическую нормальность оценок. Для случая, когда $\{\boldsymbol{x}_t\}$ является последовательностью фиксированных констант, Амемия [Amemiya, 1973] доказал состоятельность и асимптотическую нормальность ML-оценок в тобит-модели в предположении, что $\boldsymbol{x}_t$ ограниченны и что матрица $\lim \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{x}_t\boldsymbol{x}_t'$ невырождена. Для случая, когда $\boldsymbol{x}_t$ стохастические (как здесь), разумным кажется предположение о том, что достаточным условием для состоятельности и асимптотической нормальности оценок является невырожденность матрицы $\mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}_t')$.

Восстановление исходных параметров

Так же, как и в случае усеченной модели регрессии, ML-оценка параметров (β_0, σ_0^2) может быть получена как $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\delta}}/\widehat{\gamma}$ и $\widehat{\sigma}^2 = 1/\widehat{\gamma}^2$. Асимптотическая ковариационная матрица $(\widehat{\boldsymbol{\beta}},\widehat{\sigma}^2)$ может быть получена дельта-методом как $\widehat{\boldsymbol{J}}\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\delta}},\widehat{\gamma})}\widehat{\boldsymbol{J}}'$, где $\widehat{\boldsymbol{J}}$ обозначает то же, что и в уравнении (8.2.18), и $\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\delta}},\widehat{\gamma})}$ равна $-\left[\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{H}(\boldsymbol{w}_t;\widehat{\boldsymbol{\delta}},\widehat{\gamma})\right]^{-1}$.

Контрольные вопросы

1. (Сравнение цензурированной и усеченной регрессий.) Одной из возможностей увидеть различия между усечением и цензурированием является вычисление $\mathrm{E}(y_t|\boldsymbol{x}_t)$ для цензурированной модели регрессии. Покажите, что

$$\mathrm{E}(y_t|\boldsymbol{x}_t) = \left[1 - \Phi\left(\frac{c - \boldsymbol{x}_t'\beta_0}{\sigma_0}\right)\right] \times \left[\boldsymbol{x}_t'\beta_0 + \sigma_0\lambda\left(\frac{c - \boldsymbol{x}_t'\beta_0}{\sigma_0}\right)\right] + \\ + \Phi\left(\frac{c - \boldsymbol{x}_t'\beta_0}{\sigma_0}\right)c.$$

Указание: *При $y_t > c$ распределение y_t (условное относительно $\boldsymbol{x}_t$) представляет собой не что иное, как усеченное распределение, которое упоминалось в предыдущем параграфе, с математическим ожиданием* (8.2.6). *Вероятность того, что $y_t > c$, дана в* (8.2.10).

2. Покажите, что гессиан в (8.3.12) отрицательно полуопределен. **Указание:**

$$\begin{bmatrix} \boldsymbol{x}_t\boldsymbol{x}_t' & -y_t\boldsymbol{x}_t \\ -y_t\boldsymbol{x}_t' & \frac{1}{\gamma^2} + y_t^2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_t \\ -y_t \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_t' & -y_t \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0}' & \frac{1}{\gamma^2} \end{bmatrix},$$

$$\begin{bmatrix} \boldsymbol{x}_t\boldsymbol{x}_t' & -c\boldsymbol{x}_t \\ -c\boldsymbol{x}_t' & c^2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_t \\ -c \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_t' & -c \end{bmatrix}.$$

Кроме того, $\lambda(v) \geq v$ для любого v (то есть $\lambda(-v) + v \geq 0$ для всех $-v$).

3. (Тобит при наличии сериальной корреляции наблюдений.) Пусть $\{y_t, \boldsymbol{x}_t\}$ — эргодический стационарный, но необязательно i.i.d. Является ли тобит ML-оценка, описанная в тексте, состоятельной? [Ответ: Да.]

8.4. Многомерные регрессии

В параграфе 1.5 и в примере 7.2 главы 7 мы уже отмечали, что OLS-оценка линейной регрессионной модели численно совпадает с ML-оценкой той же модели при нормально распределенных ошибках. Мы вывели GMM для моделей из одного уравнения с эндогенными регрессорами в главе 3 и для случая нескольких уравнений в главе 4, но мы не указывали, каковы ML-оценки для этих моделей. В этом параграфе мы рассмотрим свойства ML-оценки модели многомерной регрессии, которая является простейшим частным случаем модели, состоящей из нескольких уравнений.

Переформулирование модели многомерной регрессии

Модель многомерной регрессии описывается предположениями 4.1–4.5 и 4.7, где $\boldsymbol{z}_{tm} = \boldsymbol{x}_{tm} = \boldsymbol{x}_t$ для всех $m = 1, 2, \ldots, M$. Изменяя обозначения, введенные в главе 4, систему из M уравнений можно записать как

$$y_{tm} = \underset{(1\times K)}{\boldsymbol{x}_t'} \underset{(K\times 1)}{\boldsymbol{\pi}_{0m}} + v_{tm} \quad (m = 1, 2, \ldots, M;\ t = 1, 2, \ldots, n). \tag{8.4.1}$$

Здесь мы продолжаем придерживаться соглашения, согласно которому истинные значения параметра снабжаются нижним индексом «0». Если определить

$$\underset{(M\times 1)}{\boldsymbol{y}_t} = \begin{bmatrix} y_{t1} \\ y_{t2} \\ \vdots \\ y_{tM} \end{bmatrix}, \quad \underset{(M\times 1)}{\boldsymbol{v}_t} = \begin{bmatrix} v_{t1} \\ v_{t2} \\ \vdots \\ v_{tM} \end{bmatrix}, \quad \underset{(K\times M)}{\boldsymbol{\Pi}_0} = \begin{bmatrix} \boldsymbol{\pi}_{01} & \boldsymbol{\pi}_{02} & \cdots & \boldsymbol{\pi}_{0M} \end{bmatrix}, \tag{8.4.2}$$

то эта система M уравнений может быть записана более компактно:

$$\boldsymbol{y}_t = \boldsymbol{\Pi}_0'\boldsymbol{x}_t + \boldsymbol{v}_t \quad (t = 1, 2, \ldots, n). \tag{8.4.3}$$

Эта модель имеет следующие особенности:

- $\{\boldsymbol{y}_t, \boldsymbol{x}_t\}$ стационарный и эргодический (предположение 4.2).
- Общие регрессоры $\boldsymbol{x}_t$ являются предопределенными: $\mathrm{E}(\boldsymbol{x}_t \cdot v_{tm}) = \mathbf{0}$ для любого m (предположение 4.3).
- Ранговое условие идентификации (предположение 4.4) сводится к тому, что $\mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}_t')$ — невырожденная матрица.
- Вектор остатков условно гомоскедастичен: $\mathrm{E}(\boldsymbol{v}_t\boldsymbol{v}_t'|\boldsymbol{x}_t) = \boldsymbol{\Omega}_0$, и $\boldsymbol{\Omega}_0$ положительно определена (предположение 4.7).

GMM-оценка $\boldsymbol{\pi}_{0m}$ совпадает с OLS-оценкой:

$$\widehat{\boldsymbol{\pi}}_m = \left(\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{x}_t\boldsymbol{x}_t'\right)^{-1} \left(\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{x}_t \cdot y_{tm}\right). \tag{8.4.4}$$

Поэтому оценка GMM (OLS) для $\boldsymbol{\Pi}_0$ может быть найдена как

$$\underset{(K\times M)}{\widehat{\boldsymbol{\Pi}}} = \underset{(K\times K)}{\left(\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{x}_t\boldsymbol{x}_t'\right)^{-1}} \underset{(K\times M)}{\left(\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{x}_t\boldsymbol{y}_t'\right)}. \tag{8.4.5}$$

Функция правдоподобия

Мы еще недостаточно специфицировали модель, чтобы было возможным найти ее ML-оценку. Введем дополнительные предположения о том, что 1) $\boldsymbol{v}_t|\boldsymbol{x}_t \sim N(\mathbf{0}, \boldsymbol{\Omega}_0)$ и 2) $\{\boldsymbol{y}_t, \boldsymbol{x}_t\}$ — последовательность независимых одинаково распределенных случайных величин (это предположение усиливает предложения 4.2 и 4.5).

Многомерная регрессионная модель с предположением о нормальности (1) означает, что $\boldsymbol{y}_t|\boldsymbol{x}_t \sim N(\boldsymbol{\Pi}_0', \boldsymbol{\Omega}_0)$. Функция плотности этого многомерного нормального распределения имеет вид:

$$(2\pi)^{-M/2}|\boldsymbol{\Omega}_0|^{-1/2}\exp\left[-\frac{1}{2}(\boldsymbol{y}_t - \boldsymbol{\Pi}_0'\boldsymbol{x}_t)'\boldsymbol{\Omega}_0^{-1}(\boldsymbol{y}_t - \boldsymbol{\Pi}_0'\boldsymbol{x}_t)\right]. \tag{8.4.6}$$

Заменяя истинные значения параметров $(\boldsymbol{\Pi}_0, \boldsymbol{\Omega}_0)$ их гипотетическими значениями $(\boldsymbol{\Pi}, \boldsymbol{\Omega})$ и беря логарифм, мы получаем логарифмическое условное правдоподобие для наблюдения t:

$$\begin{aligned}\log f(\boldsymbol{y}_t|\boldsymbol{x}_t; \boldsymbol{\Pi}, \boldsymbol{\Omega}) = -\frac{M}{2}\log(2\pi) + \frac{1}{2}\log(|\boldsymbol{\Omega}^{-1}|) - \\ -\frac{1}{2}(\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t)'\boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t).\end{aligned} \tag{8.4.7}$$

где был использован следующий факт: $-\log(|\boldsymbol{\Omega}|) = \log(|\boldsymbol{\Omega}^{-1}|)$.

Для случайной выборки умноженное на $\frac{1}{n}$ логарифмическое условное правдоподобие выборки равно среднему по t логарифмических условных правдоподобий для наблюдения t. То есть целевая функция $Q_n(\boldsymbol{\Pi}, \boldsymbol{\Omega})$ записывается как

$$\begin{aligned}Q_n(\boldsymbol{\Pi}, \boldsymbol{\Omega}) = -\frac{M}{2}\log(2\pi) + \frac{1}{2}\log(|\boldsymbol{\Omega}^{-1}|) - \\ -\frac{1}{2n}\sum_{t=1}^{n}(\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t)'\boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t).\end{aligned} \tag{8.4.8}$$

Доказательство того, что последняя составляющая может быть переписана как

$$\frac{1}{2n}\sum_{t=1}^{n}(\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t)'\boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t) = \frac{1}{2}\operatorname{trace}[\boldsymbol{\Omega}^{-1}\widehat{\boldsymbol{\Omega}}(\boldsymbol{\Pi})]. \tag{8.4.9}$$

где $\widehat{\boldsymbol{\Omega}}(\boldsymbol{\Pi})$ определяется как

$$\underset{(M\times M)}{\widehat{\boldsymbol{\Omega}}(\boldsymbol{\Pi})} \equiv \frac{1}{n}\sum_{t=1}^{n}(\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t)(\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t)'. \tag{8.4.10}$$

остается в качестве контрольного вопроса 1. Тогда целевая функция может быть переписана как

$$Q_n(\boldsymbol{\Pi}, \boldsymbol{\Omega}) = -\frac{M}{2}\log(2\pi) + \frac{1}{2}\log(|\boldsymbol{\Omega}^{-1}|) - \frac{1}{2}\operatorname{trace}[\boldsymbol{\Omega}^{-1}\widehat{\boldsymbol{\Omega}}(\boldsymbol{\Pi})]. \tag{8.4.11}$$

ML-оценка для $(\boldsymbol{\Pi}_0, \boldsymbol{\Omega}_0)$ — это такие значения $(\boldsymbol{\Pi}, \boldsymbol{\Omega})$, которые максимизируют целевую функцию. Пространство параметров для $(\boldsymbol{\Pi}_0, \boldsymbol{\Omega}_0)$

определяется как

$\{(\mathbf{\Pi}, \mathbf{\Omega}) | \mathbf{\Omega}$ — симметричная и положительно определенная матрица$\}$.

Максимизация функции правдоподобия

Сейчас мы покажем, что решение указанной задачи максимизации численно совпадает с GMM-оценкой. То есть ML-оценка $\mathbf{\Pi}_0$ — это GMM/OLS-оценка $\widehat{\mathbf{\Pi}}$ в (8.4.5), а ML-оценка $\mathbf{\Omega}_0$ дается соотношением:

$$\widehat{\mathbf{\Omega}} = \frac{1}{n} \sum_{t=1}^{n} \widehat{\boldsymbol{v}}_t \widehat{\boldsymbol{v}}_t' = \widehat{\mathbf{\Omega}}(\widehat{\mathbf{\Pi}}), \tag{8.4.12}$$

где $\widehat{\boldsymbol{v}}_t \equiv \boldsymbol{y}_t - \widehat{\mathbf{\Pi}}' \boldsymbol{x}_t$.

При предположениях многомерной регрессии, упомянутых выше, $\widehat{\mathbf{\Omega}}(\mathbf{\Pi})$ положительно определена с вероятностью, равной единице, для любой заданной $\mathbf{\Pi}$ при достаточно больших n (доказательство этого утверждения остается читателю как аналитическое упражнение 2). Поэтому мы можем предположить, что $\widehat{\mathbf{\Omega}}(\mathbf{\Pi})$ положительно определена, а не просто положительно полуопределена. Доказательство того, что GMM-оценка численно совпадает с оценкой максимального правдоподобия, основано на следующей двухшаговой процедуре максимизации целевой функции.

Шаг 1: Первый шаг состоит в максимизации $Q_n(\mathbf{\Pi}, \mathbf{\Omega})$ по $\mathbf{\Omega}$ в предположении, что $\mathbf{\Pi}$ задана. При этом полезно воспользоваться следующим результатом из матричной алгебры:

Неравенство, включающее в себя след и определитель матрицы[1]: *Пусть $\boldsymbol{A}$ и $\boldsymbol{B}$ — симметричные и положительно определенные матрицы одинакового размера. Тогда функция*

$$f(\boldsymbol{A}) \equiv \log(|\boldsymbol{A}|) - \mathrm{trace}(\boldsymbol{AB})$$

достигает единственного максимума в точке $\boldsymbol{A} = \boldsymbol{B}^{-1}$.

Этот результат с $\boldsymbol{A} = \mathbf{\Omega}^{-1}$ и $\boldsymbol{B} = \widehat{\mathbf{\Omega}}(\mathbf{\Pi})$ немедленно приводит к выводу о том, что целевая функция (8.4.11) достигает своего единственного максимума в точке $\mathbf{\Omega} = \widehat{\mathbf{\Omega}}(\mathbf{\Pi})$ при заданной $\mathbf{\Pi}$. Подстановка этой $\mathbf{\Omega}$ в (8.4.11) дает нам (умноженную на $\frac{1}{n}$)

[1]Этот результат вытекает из Lemma 6A [Johansen, 1995]. Единственность точки максимума достигается благодаря линейности оператора следа и строгой вогнутости $\log(|A|)$, отмеченной в [Magnus and Neudecker, 1988, p. 222].

концентрированную логарифмическую функцию правдоподобия *(concentrated log likelihood function)* (концентрированную по отношению к $\mathbf{\Omega}$):

$$\begin{aligned}
&Q_n(\mathbf{\Pi}) \equiv Q_n(\mathbf{\Pi}, \widehat{\mathbf{\Omega}}(\mathbf{\Pi})) = \\
&= -\frac{M}{2}\log(2\pi) + \frac{1}{2}\log(|\widehat{\mathbf{\Omega}}(\mathbf{\Pi})^{-1}|) - \frac{1}{2}\operatorname{trace}[\widehat{\mathbf{\Omega}}(\mathbf{\Pi})^{-1}\widehat{\mathbf{\Omega}}(\mathbf{\Pi})] = \\
&= -\frac{M}{2}\log(2\pi) - \frac{1}{2}\log(|\widehat{\mathbf{\Omega}}(\mathbf{\Pi})|) - \frac{M}{2}.
\end{aligned} \tag{8.4.13}$$

Шаг 2: Внимательно глядя на концентрированную логарифмическую функцию правдоподобия $Q_n(\mathbf{\Pi})$, можно заметить, что ML-оценка $\mathbf{\Pi}_0$ должна доставлять минимум

$$|\widehat{\mathbf{\Omega}}(\mathbf{\Pi})| = \left|\frac{1}{n}\sum_{t=1}^{n}(\boldsymbol{y}_t - \mathbf{\Pi}'\boldsymbol{x}_t)(\boldsymbol{y}_t - \mathbf{\Pi}'\boldsymbol{x}_t)'\right|. \tag{8.4.14}$$

Доказательство того факта, что OLS-оценка $\widehat{\mathbf{\Pi}}$ из (8.4.5) доставляет минимум этому выражению, оставлено читателю в качестве аналитического упражнения 1. Наконец, подстановка $\mathbf{\Pi} = \widehat{\mathbf{\Pi}}$ в (8.4.10) дает (8.4.12).

Состоятельность и асимптотическая нормальность

В главе 4 было показано, что без предположения о нормальности остатков ($\varepsilon_t|\boldsymbol{x}_t$ распределены нормально) или предположения о том, что $\{\boldsymbol{y}_t, \boldsymbol{x}_t\}$ — i.i.d., GMM/OLS-оценка $\widehat{\mathbf{\Pi}}$ из (8.4.5) является состоятельной и асимптотически нормальной и что $\widehat{\mathbf{\Omega}}$ из (8.4.12) состоятельна. Из этого тривиально следует, что ML-оценка $\mathbf{\Pi}_0$ также является состоятельной и асимптотически нормальной и что ML-оценка $\mathbf{\Omega}_0$ состоятельна. Также следует отметить, что ML-оценка $\mathbf{\Pi}_0$, основанная на функции правдоподобия, которая предполагает нормальность, является состоятельной и асимптотически нормальной, даже если ошибки не распределены по нормальному закону.

Мы не будем заниматься асимптотической нормальностью ML-оценки $\mathbf{\Omega}_0$, $\widehat{\mathbf{\Omega}}(\widehat{\mathbf{\Pi}})$. Показать асимптотическую нормальность ML-оценки можно, например, проверив условия соответствующей теоремы об асимптотической нормальности из предыдущей главы.

Контрольные вопросы

1. Докажите (8.4.9). **Указание:** *Воспользуйтесь следующими свойствами оператора следа. 1)* $\operatorname{trace}(x) = x$, *если* x *— скаляр, 2)* $\operatorname{trace}(\boldsymbol{A} + \boldsymbol{B}) = \operatorname{trace}(\boldsymbol{A}) + \operatorname{trace}(\boldsymbol{A})$ *и 3)* $\operatorname{trace}(\boldsymbol{AB}) = \operatorname{trace}(\boldsymbol{BA})$ *при условии, что*

матричные произведения $\boldsymbol{AB}$ и $\boldsymbol{BA}$ определены. Ответ:

$$\frac{1}{n}\sum_{t=1}^{n}(\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t)'\boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t) =$$

$$= \operatorname{trace}\left[\frac{1}{n}\sum_{t=1}^{n}(\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t)'\boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t)\right] =$$

$$= \frac{1}{n}\sum_{t=1}^{n}\operatorname{trace}\left[(\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t)'\boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t)\right] =$$

$$= \frac{1}{n}\sum_{t=1}^{n}\operatorname{trace}\left[\boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t)(\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t)'\right]$$

$$= \operatorname{trace}\left[\frac{1}{n}\sum_{t=1}^{n}\boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t)(\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t)'\right]$$

$$= \operatorname{trace}\left[\boldsymbol{\Omega}^{-1}\frac{1}{n}\sum_{t=1}^{n}(\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t)(\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t)'\right].$$

8.5. FIML

Как было показано в главе 4, многомерная модель регрессии является частным случаем модели кажущихся несвязанными регрессий (SUR), которая, в свою очередь, является частным случаем системы одновременных уравнений, для которой применим трехшаговый метод наименьших квадратов (3SLS). Двухшаговый OLS (2SLS) совпадает с GMM, когда каждое уравнение системы оценивается отдельно, в то время как 3SLS совпадает с GMM когда оценивается вся система в целом. В этом параграфе представлены ML аналоги для 2SLS и 3SLS.

Система одновременных уравнений с общими инструментами: новые обозначения

Напомним, что эта модель представляет собой **модель из M уравнений** *M-equation system*, описываемую предположениями 4.1–4.5 и 4.7, где $\boldsymbol{x}_{tm} = \boldsymbol{x}_t$ для всех $m = 1, 2, \ldots, M$ (то есть набор инструментов одинаков для всех уравнений). Сохраняя договоренность снабжать истинные значения параметров нижним индексом 0, мы можем записать M уравнений системы следующим образом:

$$y_{tm} = \underset{(1\times L_m)}{\boldsymbol{z}'_{tm}}\ \underset{(L_m\times 1)}{\boldsymbol{\delta}_{0m}} + \varepsilon_{tm} \quad (m = 1, 2, \ldots, M;\ t = 1, 2, \ldots, n). \tag{8.5.1}$$

Отметим следующие особенности этой модели:

- В отличие от модели многомерной регрессии, в каждом из m уравнений регрессоры $\boldsymbol{z}_{tm}$ могут не быть ортогональны ошибкам, однако имеется K доступных предопределенных переменных $\boldsymbol{x}_t$, для которых условие ортогональности выполняется: $\mathrm{E}(\boldsymbol{x}_t \cdot \varepsilon_{tm}) = \mathbf{0}$ для всех m (это предположение 4.3). Определив вектор $\boldsymbol{\varepsilon}_t$ размерности $M \times 1$ как

$$\underset{(M\times 1)}{\boldsymbol{\varepsilon}_t} = \begin{bmatrix} \varepsilon_{t1} \\ \varepsilon_{t2} \\ \vdots \\ \varepsilon_{tM} \end{bmatrix}, \tag{8.5.2}$$

условия ортогональности можно выразить как

$$\mathrm{E}(\boldsymbol{x}_t \boldsymbol{\varepsilon}_t') = \underset{(K\times M)}{\mathbf{0}}. \tag{8.5.3}$$

Эти K предопределенных переменных используются как общий набор инструментов при GMM-оценивании.

- Переменные, входящие в систему, но не включенные в $\boldsymbol{x}_t$ (кроме ошибок), называются **эндогенными переменными** (*endogenous variables*).
- Ранговое условие идентификации (предположение 4.4) состоит в том, что матрица $\mathrm{E}(\boldsymbol{x}_t \boldsymbol{z}_{tm}')$ размерности $K \times L_m$ имеет полный столбцовый ранг для всех m. Уравнение m называется **сверхидентифицированным** (*overidentified*), если условие ранга выполнено и $L_m < K$.
- Вектор ошибок условно гомоскедастичен: $\mathrm{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t' | \boldsymbol{x}_t) = \boldsymbol{\Sigma}_0$ и $\boldsymbol{\Sigma}_0$ положительно определена (предположение 4.7).
- $\mathrm{E}(\boldsymbol{x}_t \boldsymbol{x}_t')$ — невырожденная матрица (следствие условной гомоскедастичности и требования невырожденности в предположении 4.5).

GMM-оценка системы в целом является 3SLS-оценкой, приведенной в (4.5.12), в то время как оценивание каждого уравнения в отдельности с помощью GMM представляет собой 2SLS, заданный соотношением (3.8.3).

Следующие примеры кратко иллюстрируют эти понятия.

Пример 8.1 (модель рыночного равновесия): Расширенная версия примера Уоркинга, рассмотренная в параграфе 3.1, представляет собой систему из двух уравнений:

$$q_t = \gamma_{11} p_t + \beta_{11} + \beta_{12} a_t + \varepsilon_{t1} \quad \text{(спрос)}. \tag{8.5.4}$$
$$q_t = \gamma_{21} p_t + \beta_{21} + \beta_{22} w_t + \varepsilon_{t2} \quad \text{(предложение)}. \tag{8.5.5}$$

В этой системе q_t — количество кофе, p_t — цена кофе. Переменная a_t, фигурирующая в уравнении спроса, отражает изменение предпочтений потребителей. Переменная w_t представляет собой количество осадков, которое влияет на предложение кофе. Эта система может быть приведена к общей форме введением следующих обозначений:

$$y_{t1} = q_t, \quad y_{t2} = q_t, \quad \boldsymbol{z}_{t1} = \begin{bmatrix} p_t \\ 1 \\ a_t \end{bmatrix}, \quad \boldsymbol{z}_{t2} = \begin{bmatrix} p_t \\ 1 \\ w_t \end{bmatrix},$$

$$\boldsymbol{\delta}_{01} = \begin{bmatrix} \gamma_{11} \\ \beta_{11} \\ \beta_{12} \end{bmatrix}, \quad \boldsymbol{\delta}_{02} = \begin{bmatrix} \gamma_{21} \\ \beta_{21} \\ \beta_{22} \end{bmatrix}.$$

(Заметим, что y_{t1} и y_{t2} являются одной и той же переменной.) Если $\mathrm{E}(\varepsilon_{tj}) = 0, \mathrm{E}(a_t\varepsilon_{tj}) = 0$ и $\mathrm{E}(w_t\varepsilon_{tj}) = 0$ для $j = 1, 2$, то тогда $(1, a_t, w_t)$ могут быть включены в общий набор инструментов, так что

$$\boldsymbol{x}_t = \begin{bmatrix} 1 \\ a_t \\ w_t \end{bmatrix}.$$

Остальные переменные (кроме ошибок), входящие в систему (то есть в данном примере p_t и q_t) являются эндогенными переменными. При условии, что ранговое условие идентификации выполнено, каждое уравнение идентифицировано точно, поскольку количество переменных в правых частях уравнений равно количеству инструментов.

Пример 8.2 (уравнение заработной платы): В главе 3 мы оценили стандартное уравнение заработной платы. Дополним это уравнение еще одним, объясняющим результат теста *Knowledge of the World of Work* (KWW).

$$LW_t = \gamma_{11}S_t + \beta_{11} + \beta_{12}IQ_t + \varepsilon_{t1}, \tag{8.5.6}$$

$$KWW_t = \gamma_{21}S_t + \beta_{21} + \varepsilon_{t2}. \tag{8.5.7}$$

где S_t — количество лет обучения индивида t. Предположим, что $\mathrm{E}(\varepsilon_{tj}) = 0$ и что $\mathrm{E}(IQ_t\varepsilon_{tj}) = 0$ для $j = 1, 2$. Пусть в нашем распоряжении имеется переменная MED (образование матери), которая не фигурирует ни в одном из уравнений, но которая является предопределенной, в том смысле, что $\mathrm{E}(MED_t\varepsilon_{tj}) = 0$ для $j = 1, 2$. То есть общий набор инструментов есть $\boldsymbol{x}_t = (1, IQ_t, MED_t)'$. Остальные переменные (кроме ошибок), входящие в систему, являются эндогенными. В этой системе из двух уравнений имеются три эндогенные переменные: LW, S и KWW. При условии выполнения рангового условия идентифицируемости первое уравнение точно идентифицировано, поскольку количество переменных в правой части уравнения равно количеству инструментов. Второе уравнение сверхидентифицировано.

Полная система одновременных уравнений

ML аналог 3SLS называется **оценкой максимального правдоподобия с полной информацией** (*full-information maximum likelihood*) — **FIML**. Перед применением FIML на модель одновременных уравнений необходимо наложить несколько дополнительных условий *до* введения предположений о нормальном распределении и независимости и одинаковой распределенности. Эти дополнительные требования состоят в том, чтобы модель из M уравнений была «полной» системой одновременных уравнений. Полнота требует выполнения двух условий:

1. Количество эндогенных переменных в системе должно равняться числу уравнений. Это условие подразумевает, что если $\boldsymbol{y}_t$ — вектор длины M, содержащий эти M эндогенных переменных, то $(y_{t1}, \ldots, y_{tM}, \boldsymbol{z}_1, \ldots, z_{tM})$ — это все элементы $(\boldsymbol{y}_t, \boldsymbol{x}_t)$, которые позволяют записать систему из M уравнений (8.5.1) в виде:

$$\underset{(M\times M)}{\boldsymbol{\Gamma}_0}\underset{(M\times 1)}{\boldsymbol{y}_t} + \underset{(M\times K)}{\boldsymbol{B}_0}\underset{(K\times 1)}{\boldsymbol{x}_t} = \underset{(M\times 1)}{\boldsymbol{\varepsilon}_t} \quad (t = 1, 2, \ldots, n). \tag{8.5.8}$$

где m-е уравнение имеет вид: $y_{tm} = \boldsymbol{z}_{tm}\boldsymbol{\delta}_{0m} + \varepsilon_{tm}$. (Это продемонстрировано для примера 8.1 выше.) Такая запись называется **структурной формой**, а $(\boldsymbol{\Gamma}_0, \boldsymbol{B}_0)$ называются **параметрами структурной формы** (*structural form parameters*). Как будет показано ниже, каждый из структурных параметров является функцией от $(\boldsymbol{\delta}_{01}, \ldots, \boldsymbol{\delta}_{0M})$.

Система уравнений из примера 8.1 удовлетворяет этому требованию. Система уравнений из примера 8.2 не является полной системой, поскольку количество эндогенных переменных превышает количество наблюдений. С помощью FIML невозможно оценить неполную систему (за исключением случая, когда мы дополним систему соответствующими уравнениями, см. обсуждение LIML далее). Напротив, неполную систему можно оценить с помощью 3SLS (или с помощью GMM для системы уравнений, если не вводится предположение об условной гомоскедастичности), если система удовлетворяет ранговому условию идентифицируемости.

2. Квадратная матрица $\boldsymbol{\Gamma}_0$ невырождена. Из этого следует, что структурная форма может быть решена относительно эндогенной переменной $\boldsymbol{y}_t$ следующим образом:

$$\boldsymbol{y}_t = -\boldsymbol{\Gamma}_0^{-1}\boldsymbol{B}_0\boldsymbol{x}_t + \boldsymbol{\Gamma}_0^{-1}\boldsymbol{\varepsilon}_t \equiv \boldsymbol{\Pi}_0'\boldsymbol{x}_t + \boldsymbol{v}_t. \tag{8.5.9}$$

где

$$\underset{(M\times K)}{\Pi_0'} \equiv -\underset{(M\times M)}{\Gamma_0^{-1}}\underset{(M\times K)}{B_0}, \tag{8.5.10}$$

$$\underset{(M\times 1)}{v_t} \equiv \underset{(M\times M)}{\Gamma_0^{-1}}\underset{(M\times 1)}{\varepsilon_t}. \tag{8.5.11}$$

Выражение (8.5.9) называется **приведенной формой** (*reduced-form representation)* структурной формы. Элементы Π_0 называются **коэффициентами приведенной формы**. В каждом уравнении приведенной формы регрессоры представляют собой один и тот же набор переменных x_t. Из (8.5.3) и (8.5.11) следует, что $\mathrm{E}(x_t v_t') = \mathbf{0}$, то есть все регрессоры являются предопределенными. Следовательно, система уравнений в приведенной форме является многомерной регрессионной моделью.

Взаимосвязь между $(\Gamma_0, \mathbf{B}_0)$ *и* δ_0

Пусть δ_0 — штабелированный вектор, включающий в себя все коэффициенты системы из M уравнений (8.5.1):

$$\underset{(\sum_{m=1}^{M} L_m\times 1)}{\delta_0} = \begin{bmatrix} \delta_{01} \\ \delta_{02} \\ \vdots \\ \delta_{0M} \end{bmatrix}. \tag{8.5.12}$$

Для понимания механизма FIML важно посмотреть, как матрицы коэффициентов (Γ_0, B_0) зависят от δ_0. В качестве иллюстрации рассмотрим пример 8.1. В этом примере вектор δ_0 определяется как

$$\underset{(6\times 1)}{\delta_0} = \begin{bmatrix} \gamma_{11} \\ \beta_{11} \\ \beta_{12} \\ \gamma_{21} \\ \beta_{21} \\ \beta_{22} \end{bmatrix}. \tag{8.5.13}$$

Две эндогенные переменные в примере — это (p_t, q_t). Порядок включения этих переменных не важен. Для определенности пусть p_t первой входит в вектор y_t: $y_t = (p_t, q_t)'$. Тогда параметры структурной формы могут быть записаны как

$$\underset{(2\times 2)}{\Gamma_0} = \begin{bmatrix} -\gamma_{11} & 1 \\ -\gamma_{21} & 1 \end{bmatrix}, \quad \underset{(2\times 3)}{B_0} = \begin{bmatrix} -\beta_{11} & -\beta_{12} & 0 \\ -\beta_{21} & 0 & -\beta_{22} \end{bmatrix}. \tag{8.5.14}$$

Мы привели этот пример для иллюстрации трех аспектов. Во-первых, каждая строка матрицы $\mathbf{\Gamma}_0$ содержит единицу, что отражает тот факт, что коэффициент при зависимой переменной в каждом из M уравнений (8.5.1) равен единице. В этом смысле $\mathbf{\Gamma}_0$ уже нормализована. Во-вторых, некоторые элементы $\mathbf{\Gamma}_0$ и $\boldsymbol{B}_0$ равны нулю, что отражает тот факт, что некоторые из эндогенных или предопределенных переменных не входят в отдельные уравнения системы. Эта особенность матрицы коэффициентов структурной формы называется **исключающими ограничениями** *(exclusion restrictions)*. (В примере 8.1 ни один из элементов $\mathbf{\Gamma}_0$ не равен нулю, так как обе эндогенные переменные входят в оба уравнения системы.) В-третьих, в системе нет ограничений, затрагивающих сразу несколько уравнений, поэтому каждый элемент $\boldsymbol{\delta}_0$ появляется в $(\mathbf{\Gamma}_0, \boldsymbol{B}_0)$ только один раз.

Функция правдоподобия для FIML

Чтобы полная система одновременных уравнений могла оцениваться посредством FIML, мы предполагаем, что 1) вектор структурных ошибок $\boldsymbol{\varepsilon}_t$ имеет совместное нормальное распределение, условное относительно $\boldsymbol{x}_t$ (то есть $\boldsymbol{\varepsilon}_t|\boldsymbol{x}_t \sim N(\mathbf{0}, \Sigma_0)$) и 2) $\{\boldsymbol{y}_t, \boldsymbol{x}_t\}$ — последовательность независимых одинаково распределенных случайных величин, а не просто эргодические стационарные мартингал-разности (это предположение — более жесткий аналог предположений 4.2 и 4.5).

Вследствие предположения о нормальном распределении структурных ошибок $\boldsymbol{\varepsilon}_t$ (1) и (8.5.11) мы имеем $\boldsymbol{v}_t|\boldsymbol{x}_t \sim N(\mathbf{0}, \mathbf{\Gamma}_0^{-1}\Sigma_0(\mathbf{\Gamma}_0^{-1})')$. Совмещая это с приведенной формой $\boldsymbol{y}_t = \mathbf{\Pi}_0'\boldsymbol{x}_t + \boldsymbol{v}_t$ и (8.5.10), получаем:

$$\boldsymbol{y}_t|\boldsymbol{x}_t \sim N(-\mathbf{\Gamma}_0^{-1}\boldsymbol{B}_0\boldsymbol{x}_t, \mathbf{\Gamma}_0^{-1}\Sigma_0(\mathbf{\Gamma}_0^{-1})'). \tag{8.5.15}$$

Поэтому логарифмическое правдоподобие для наблюдения t принимает вид:

$$\begin{aligned}\log f(\boldsymbol{y}_t|\boldsymbol{x}_t; \boldsymbol{\delta}, \mathbf{\Sigma}) = &-\frac{M}{2}\log(2\pi) - \frac{1}{2}\log(|\mathbf{\Gamma}^{-1}\Sigma(\mathbf{\Gamma}^{-1})'|) - \\ &-\frac{1}{2}[\boldsymbol{y}_t + \mathbf{\Gamma}^{-1}\boldsymbol{B}\boldsymbol{x}_t]'[\mathbf{\Gamma}^{-1}\Sigma(\mathbf{\Gamma}^{-1})']^{-1}[\boldsymbol{y}_t + \mathbf{\Gamma}^{-1}\boldsymbol{B}\boldsymbol{x}_t].\end{aligned} \tag{8.5.16}$$

Это правдоподобие является функцией от $(\boldsymbol{\delta}, \mathbf{\Sigma})$, поскольку, как было показано в (8.5.14), коэффициенты структурной формы $(\mathbf{\Gamma}, \boldsymbol{B})$ являются функциями от $\boldsymbol{\delta}$. Теперь

$$\begin{aligned}&[\boldsymbol{y}_t + \mathbf{\Gamma}^{-1}\boldsymbol{B}\boldsymbol{x}_t]'[\mathbf{\Gamma}^{-1}\Sigma(\mathbf{\Gamma}^{-1})']^{-1}[\boldsymbol{y}_t + \mathbf{\Gamma}^{-1}\boldsymbol{B}\boldsymbol{x}_t] = \\ &= [\boldsymbol{y}_t + \mathbf{\Gamma}^{-1}\boldsymbol{B}\boldsymbol{x}_t]'[\mathbf{\Gamma}'\mathbf{\Sigma}^{-1}\mathbf{\Gamma}][\boldsymbol{y}_t + \mathbf{\Gamma}^{-1}\boldsymbol{B}\boldsymbol{x}_t] = \\ &= [\mathbf{\Gamma}\boldsymbol{y}_t + \boldsymbol{B}\boldsymbol{x}_t]'\mathbf{\Sigma}^{-1}[\mathbf{\Gamma}\boldsymbol{y}_t + \boldsymbol{B}\boldsymbol{x}_t]\end{aligned} \tag{8.5.17}$$

и

$$|\mathbf{\Gamma}^{-1}\mathbf{\Sigma}(\mathbf{\Gamma}^{-1})'| = |\mathbf{\Gamma}^{-1}||\mathbf{\Sigma}||(\mathbf{\Gamma}^{-1})'| = |\mathbf{\Sigma}|/|\mathbf{\Gamma}|^2. \tag{8.5.18}$$

Подставляя эти выражения в (8.5.16), получаем:

$$\begin{aligned}\log f(\boldsymbol{y}_t|\boldsymbol{x}_t;\boldsymbol{\delta},\mathbf{\Sigma}) = -\frac{M}{2}\log(2\pi) + \frac{1}{2}\log(|\mathbf{\Gamma}|^2) - \frac{1}{2}\log(|\mathbf{\Sigma}|)\\ - \frac{1}{2}\big[\mathbf{\Gamma}\boldsymbol{y}_t + \boldsymbol{B}\boldsymbol{x}_t\big]'\mathbf{\Sigma}^{-1}\big[\mathbf{\Gamma}\boldsymbol{y}_t + \boldsymbol{B}\boldsymbol{x}_t\big].\end{aligned} \tag{8.5.19}$$

Усредняя по t, получаем целевую функцию FIML для случайной выборки:

$$\begin{aligned}Q_n(\boldsymbol{\delta},\mathbf{\Sigma}) = -\frac{M}{2}\log(2\pi) + \frac{1}{2}\log(|\mathbf{\Gamma}|^2) - \frac{1}{2}\log(|\mathbf{\Sigma}|)-\\ - \frac{1}{2n}\sum_{t=1}^{n}\big[\mathbf{\Gamma}\boldsymbol{y}_t + \boldsymbol{B}\boldsymbol{x}_t\big]'\mathbf{\Sigma}^{-1}\big[\mathbf{\Gamma}\boldsymbol{y}_t + \boldsymbol{B}\boldsymbol{x}_t\big].\end{aligned} \tag{8.5.20}$$

Оценка FIML истинных параметров $(\boldsymbol{\delta}_0, \mathbf{\Sigma}_0)$ — это $(\boldsymbol{\delta}, \mathbf{\Sigma})$, доставляющие максимум целевой функции.

Концентрированная функция правдоподобия для FIML

Как и в случае с максимизацией целевой функции для многомерной регрессионной модели, мы можем решить вопрос в два этапа. Сначала можно найти максимум $Q_n(\boldsymbol{\delta}, \mathbf{\Sigma})$ по $\mathbf{\Sigma}$, считая $\boldsymbol{\delta}$ заданным. В результате получаем:

$$\underset{(M\times M)}{\widehat{\mathbf{\Sigma}}(\boldsymbol{\delta})} = \frac{1}{n}\sum_{t=1}^{n}(\mathbf{\Gamma}\boldsymbol{y}_t + \boldsymbol{B}\boldsymbol{x}_t)(\mathbf{\Gamma}\boldsymbol{y}_t + \boldsymbol{B}\boldsymbol{x}_t)'. \tag{8.5.21}$$

Элемент (m, h) может быть записан как

$$\widehat{\sigma}_{mh} = \frac{1}{n}\sum_{t=1}^{n}(y_{tm} - \boldsymbol{z}_{tm}'\boldsymbol{\delta}_m)(y_{th} - \boldsymbol{z}_{th}'\boldsymbol{\delta}_h). \tag{8.5.22}$$

Подставляя (8.5.2) обратно в целевую функцию (8.5.20), умноженную на $1/n$, концентрированную функцию правдоподобия для FIML можно записать в виде:

$$\begin{aligned}Q_n(\boldsymbol{\delta}) &\equiv Q_n(\boldsymbol{\delta}, \widehat{\mathbf{\Sigma}}(\boldsymbol{\delta})) =\\ &= -\frac{M}{2}\log(2\pi) - \frac{M}{2} + \frac{1}{2}\log(|\mathbf{\Gamma}|^2) - \frac{1}{2}\log(|\widehat{\mathbf{\Sigma}}(\boldsymbol{\delta})|) =\\ &= -\frac{M}{2}\log(2\pi) - \frac{M}{2} - \frac{1}{2}\log\left|\frac{1}{n}\sum_{t=1}^{n}(\boldsymbol{y}_t + \mathbf{\Gamma}^{-1}\boldsymbol{B}\boldsymbol{x}_t)(\boldsymbol{y}_t + \mathbf{\Gamma}^{-1}\boldsymbol{B}\boldsymbol{x}_t)'\right|.\end{aligned} \tag{8.5.23}$$

Второй шаг состоит в максимизации этой концентрированной функции правдоподобия по $\boldsymbol{\delta}$. FIML-оценка $\widehat{\boldsymbol{\delta}}$ истинного значения параметра $\boldsymbol{\delta}_0$ — это такое значение $\boldsymbol{\delta}$, которое доставляет максимум этой функции[1]. FIML-оценка $\boldsymbol{\Sigma}_0$ — это $\widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\delta}})$.

Тестирование сверхидентифицирующих ограничений

Сравним теперь концентрированную функцию правдоподобия для FIML $Q_n(\boldsymbol{\delta})$ из (8.5.23) с концентрированной функцией правдоподобия для модели многомерной регрессии $Q_n(\boldsymbol{\Pi})$ из (8.4.13). Концентрированную функцию правдоподобия для FIML $Q_n(\boldsymbol{\delta})$ можно получить из $Q_n(\boldsymbol{\Pi})$ путем введения ограничений. Эти ограничения соответствуют следующей нулевой гипотезе:

$$H_0: \ \boldsymbol{\Pi}_0' = -\boldsymbol{\Gamma}_0^{-1}\boldsymbol{B}_0 \quad \text{или} \quad \boldsymbol{\Gamma}_0\boldsymbol{\Pi}_0' + \boldsymbol{B}_0 = \mathbf{0}. \tag{8.5.24}$$

Иными словами, FIML-оценка $\boldsymbol{\delta}_0$ — это оценка многомерной регрессии с ограничениями. Поэтому истинность нулевой гипотезы можно проверить, используя принцип отношения правдоподобий. Поскольку OLS-оценка $\widehat{\boldsymbol{\Pi}}$ из (8.4.5) максимизирует концентрированную функцию правдоподобия для многомерной регрессии $Q_n(\boldsymbol{\Pi})$ из (8.4.13), а FIML-оценка $\widehat{\boldsymbol{\delta}}$ максимизирует концентрированную функцию правдоподобия для FIML $Q_n(\boldsymbol{\delta})$ из (8.5.23), статистика теста отношения правдоподобий принимает следующий вид:

$$\begin{aligned} LR &= 2n \cdot (Q_n^*(\widehat{\boldsymbol{\Pi}}) - Q_n^*(\widehat{\boldsymbol{\delta}})) = \\ &= n \times \left(\log\left|\widehat{\boldsymbol{\Omega}}\left(-(\widehat{\boldsymbol{\Gamma}}^{-1}\widehat{\boldsymbol{B}})'\right)\right| - \log\left|\widehat{\boldsymbol{\Omega}}(\widehat{\boldsymbol{\Pi}})\right|\right), \end{aligned} \tag{8.5.25}$$

где $\widehat{\boldsymbol{\Omega}}(\boldsymbol{\Pi})$ определена в (8.4.10) и $(\widehat{\boldsymbol{\Gamma}}, \widehat{\boldsymbol{B}})$ — значение $(\boldsymbol{\Gamma}, \boldsymbol{B})$ при $\widehat{\boldsymbol{\delta}}$. Так как размерность $\boldsymbol{\delta}$ равна $\sum_{m=1}^{M} L_m$, количество ограничений, накладываемых нулевой гипотезой H_0 выше, равно:

$$KM - \sum_{m=1}^{M} L_m = \sum_{m=1}^{M} (K - L_m), \tag{8.5.26}$$

что равно общему числу сверхидентифицирующих ограничений в системе. Поэтому тест отношения правдоподобия, основанный на LR статистике из (8.5.25), называется **тестом на сверхидентифицирующие**

---

[1]То есть FIML-оценка $\widehat{\boldsymbol{\delta}}$ является экстремальной оценкой, которая максимизирует

$$-\frac{1}{2}\log\left|\frac{1}{n}\sum_{t=1}^{n}(\boldsymbol{y}_t + \boldsymbol{\Gamma}^{-1}\boldsymbol{B}\boldsymbol{x}_t)(\boldsymbol{y}_t + \boldsymbol{\Gamma}^{-1}\boldsymbol{B}\boldsymbol{x}_t)'\right|.$$

Некоторые свойства этой экстремальной оценки рассматриваются в дополнительном аналитическом упражнении 3.

ограничения (*test of overidentifying restrictions*). Он является тестом на спецификацию модели, так как тестируемые ограничения — это условие, предполагаемое в модели.

Свойства FIML-оценки

Для FIML-оценки не существует аналитического решения. Значит, для проверки состоятельности и асимптотической нормальности нам будет необходимо проверять условия соответствующих теорем из предыдущего параграфа. Далее мы приведем без доказательства основные свойства FIML-оценок.

Идентифицируемость

Формулирование условий идентифицируемости в различных формах для полной системы одновременных уравнений — одна из важнейших тем большинства учебников. Здесь же мы предлагаем в качестве дополнительного аналитического упражнения 3 показать, что условие идентифицируемости FIML как экстремальной оценки равносильно ранговому условию идентификации (то есть что матрица $\mathrm{E}(\boldsymbol{x}_t \boldsymbol{z}'_{tm})$ имеет полный столбцовый ранг для всех $m = 1, 2, \ldots, M$).

Асимптотические свойства

Хотя целевая функция FIML (8.5.20) выводилась в предположении нормального распределения ошибок $\boldsymbol{\varepsilon}_t|\boldsymbol{x}_t$, FIML-оценка $\boldsymbol{\delta}_0$ является состоятельной и асимптотически нормально распределенной и без этого предположения. Доказательство состоятельности приведено, например, в: [Amemiya, 1985, pp. 232–233]. Элегантное доказательство асимптотической нормальности распределения $\widehat{\boldsymbol{\delta}}$ принадлежит Хаусману [Hausman, 1975], который показывает, что итерационный метод инструментальных переменных может рассматриваться как алгоритм для решения условий первого порядка максимизации $Q_n(\boldsymbol{\delta}, \boldsymbol{\Sigma})$[1]. Это доказательство показывает, что FIML-оценка $\boldsymbol{\delta}_0$ асимптотически эквивалентна 3SLS-оценке[2]. Поэтому ее асимптотическая дисперсия дается выражением (4.5.14) и состоятельной оценкой асимптотической дисперсии является (4.5.17).

---

[1]Мы не будем приводить условия первого порядка (уравнения правдоподобия) для FIML-оценок, так как их вывод требует громоздких матричных обозначений. Детали можно найти в: [Amemiya, 1985, pp. 233–234] или [Davidson and MacKinnon, 1993, pp. 658–660]. Для SUR-модели уравнения правдоподобия и итерационную процедуру описать намного проще, см. следующий параграф.

[2]Эта асимптотическая эквивалентность не сохраняется при переходе к системе нелинейных уравнений; нелинейный FIML является более эффективным, чем нелинейный 3SLS. См.: [Amemiya, 1985, Section 8.2].

Инвариантность

Так как FIML — это ML-оценка, она обладает свойством инвариантности, которое обсуждалось в параграфе 7.1. Чтобы проиллюстрировать это свойство, вернемся к примеру 8.1. Предположим, мы хотим репараметризовать уравнение предложения следующим образом:

$$p_t = \widetilde{\gamma}_{21} q_t + \widetilde{\beta}_{21} + \widetilde{\beta}_{22} w_t + \widetilde{\varepsilon}_{t2} \quad \text{(предложение)}. \tag{8.5.27}$$

Параметры старого уравнения, описывающего предложение $(\gamma_{21}, \beta_{21}, \beta_{22})$, и параметры нового уравнения предложения $(\widetilde{\gamma}_{21}, \widetilde{\beta}_{21}, \widetilde{\beta}_{22})$ связаны взаимно однозначным соответствием:

$$\widetilde{\gamma}_{21} = 1/\gamma_{21}, \widetilde{\beta}_{21} = -\beta_{21}/\gamma_{21}, \widetilde{\beta}_{22} = -\beta_{22}/\gamma_{21}. \tag{8.5.28}$$

Если $(\widehat{\gamma}_{21}, \widehat{\beta}_{21}, \widehat{\beta}_{22})$ — это FIML-оценка параметров $(\gamma_{21}, \beta_{21}, \beta_{22})$ для системы уравнений, состоящей из уравнения спроса (8.5.4) и старого уравнения предложения (8.5.5), то тогда FIML-оценка, основанная на системе, состоящей из (8.5.4) и нового уравнения предложения, численно совпадает со значением вышеупомянутого отображения в точке $(\widehat{\gamma}_{21}, \widehat{\beta}_{21}, \widehat{\beta}_{22})$. 3SLS-оценка (и 2SLS по той же причине) не обладает таким свойством. Вышеизложенная дискуссия о FIML-оценках может быть подытожена следующим образом:

Утверждение 8.1 (асимптотические свойства FIML): *Рассмотрим систему из M уравнений* (8.5.1). *Пусть $\boldsymbol{\delta}_0$ — штабелированный вектор, включающий все коэффициенты этой системы. Пусть выполнены следующие предположения:*

- *выполнено ранговое условие идентификации: матрица* $\mathrm{E}(\boldsymbol{x}_t \boldsymbol{z}'_{tm})$ *имеет полный столбцовый ранг для всех* $m = 1, 2, \ldots, M$;
- $\mathrm{E}(\boldsymbol{x}_t \boldsymbol{x}'_t)$ *— невырожденная матрица;*
- *указанная система из M уравнений может быть представлена в форме полной системы одновременных уравнений* (8.5.8) *с невырожденной матрицей* $\boldsymbol{\Gamma}_0$;
- $\boldsymbol{\varepsilon}_t | \boldsymbol{x}_t \sim N(\boldsymbol{0}, \boldsymbol{\Sigma}_0)$, $\boldsymbol{\Sigma}_0$ *положительно определена,*
- $\{\boldsymbol{y}_t, \boldsymbol{x}_t\}$ *— i.i.d.;*
- *пространство параметров для* $(\boldsymbol{\delta}_0, \boldsymbol{\Sigma}_0)$ *является компактом, причем вектор истинных параметров* $(\boldsymbol{\delta}_0, \boldsymbol{\Sigma}_0)$ *является его внутренней точкой.*

Тогда

(a) *FIML-оценка* $(\widehat{\boldsymbol{\delta}}, \widehat{\boldsymbol{\Sigma}})$, *доставляющая максимум* (8.5.20), *состоятельна и асимптотически нормальна,*

(b) *асимптотическая дисперсия $\widehat{\boldsymbol{\delta}}$ равна* (4.5.15), *а состоятельная оценка асимптотической дисперсии задается* (4.5.17), *где* $\widehat{\sigma}^{mh}$ — (m,h)-*й элемент* $\widehat{\boldsymbol{\Sigma}}^{-1}$,

(c) *статистика отношения правдоподобий* (8.5.25) *для тестирования сверхидентифицирующих ограничений асимптотически распределена как* χ^2 *с* $KM - \sum_{m=1}^{M} L_m$ *степенями свободы.*

Кроме того, эти асимптотические свойства $\widehat{\boldsymbol{\delta}}$ сохраняются, даже если распределение $\boldsymbol{\varepsilon}_t|\boldsymbol{x}_t$ не является нормальным.

ML-оценивание SUR-модели

Модель SUR является частным случаем FIML, где $\boldsymbol{z}_{tm}$ является подвектором $\boldsymbol{x}_t$. Следовательно, это модель из M уравнений (8.5.1) с предопределенными регрессорами $\boldsymbol{z}_{tm}$. В отличие от многомерной модели регрессии, набор регрессоров может быть разным для разных уравнений системы. Легко показать (см. контрольный вопрос 3), что никакие из двух зависимых переменных не могут быть одной и той же переменной. То есть $\boldsymbol{y}_t$ — $(M\times1)$ в структурной форме (8.5.8) просто содержит M зависимых переменных. Кроме того, поскольку $\boldsymbol{z}_{tm}$ не включает в себя эндогенных переменных, $\boldsymbol{\Gamma}_0$ в структурной форме представляет собой единичную матрицу. Таким образом, целевая функция FIML (8.5.20) принимает вид:

$$Q_n(\boldsymbol{\delta},\boldsymbol{\Sigma}) = -\frac{M}{2}\log(2\pi)-\frac{1}{2}\log(|\boldsymbol{\Sigma}|)-\frac{1}{2n}\sum_{t=1}^{n}\left[\boldsymbol{\Gamma}\boldsymbol{y}_t+\boldsymbol{B}\boldsymbol{x}_t\right]'\boldsymbol{\Sigma}^{-1}\left[\boldsymbol{\Gamma}\boldsymbol{y}_t+\boldsymbol{B}\boldsymbol{x}_t\right], \tag{8.5.29}$$

где $\boldsymbol{\Gamma}$ — единичная матрица. (Мы продолжаем писать $\boldsymbol{\Gamma}$, чтобы упростить сравнение с целевой функцией LIML, которая будет рассмотрена в следующем параграфе.) Как и в FIML, $\boldsymbol{\Sigma}$, которая максимизирует целевую функцию при заданном $\boldsymbol{\delta}$, — это $\widehat{\boldsymbol{\Sigma}}(\boldsymbol{\delta})$, выражение для которой приведено в (8.5.21).

Для SUR-модели вклад по отношению к $\boldsymbol{\delta}$ может быть записан довольно просто. m-я строка матрицы $\boldsymbol{\Gamma}\boldsymbol{y}_t + \boldsymbol{b}\boldsymbol{x}_t$ представляет собой $y_{tm} - \boldsymbol{z}'_{tm}\boldsymbol{\delta}$. Определим $\boldsymbol{y}$ размера $(Mn \times 1)$ и $\boldsymbol{Z}$ размера $(MN \times \sum_{m=1}^{M} L_m)$, как в первом аналитическом упражнении к главе 4. Тогда

$$\sum_{t=1}^{n}\left[\boldsymbol{\Gamma}\boldsymbol{y}_t+\boldsymbol{B}\boldsymbol{x}_t\right]'\boldsymbol{\Sigma}^{-1}\left[\boldsymbol{\Gamma}\boldsymbol{y}_t+\boldsymbol{B}\boldsymbol{x}_t\right] = (\boldsymbol{y}-\boldsymbol{Z}\boldsymbol{\delta})'(\boldsymbol{\Sigma}^{-1}\otimes \boldsymbol{I}_n)(\boldsymbol{y}-\boldsymbol{Z}\boldsymbol{\delta}). \tag{8.5.30}$$

Поэтому

$$\frac{\partial Q_n(\boldsymbol{\delta},\boldsymbol{\Sigma})}{\partial\boldsymbol{\delta}} = \frac{1}{n}\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1}\otimes \boldsymbol{I}_n)(\boldsymbol{y}-\boldsymbol{Z}\boldsymbol{\delta}). \tag{8.5.31}$$

(Проделать то же самое для более общего случая FIML не так просто, так как, хотя (8.5.30) выполнено для FIML также, вклад $\frac{\partial Q_n}{\partial \delta}$ представляет собой более сложное выражение из-за наличия члена $\log(|\mathbf{\Gamma}|^2)$ в FIML функции правдоподобия. Иными словами, если $|\mathbf{\Gamma}|$ не зависит от $\boldsymbol{\delta}$, то тогда $\frac{\partial Q_n}{\partial \delta}$ для FIML совпадает с (8.5.31).) Приравнивая это к нулю и решая относительно $\boldsymbol{\delta}$, получаем:

$$\widehat{\boldsymbol{\delta}}(\boldsymbol{\Sigma}) \equiv \left[\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{I}_n)\boldsymbol{Z}\right]^{-1}\boldsymbol{Z}'(\boldsymbol{\Sigma}^{-1} \times \boldsymbol{I}_n)\boldsymbol{y}. \tag{8.5.32}$$

При условии, что $\widehat{\boldsymbol{\Sigma}}$ — состоятельная оценка $\boldsymbol{\Sigma}_0$, $\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{\Sigma}})$ представляет собой SUR-оценку $\boldsymbol{\delta}_0$, выведенную в главе 4.

Эти две функции $(\widehat{\boldsymbol{\delta}}(\boldsymbol{\Sigma}), \widehat{\boldsymbol{\Sigma}}(\boldsymbol{\delta}))$ определяют отображение из пространства параметров для $(\boldsymbol{\delta}, \boldsymbol{\Sigma})$ в себя, и решение условий первого порядка для максимизации целевой функции $Q_n(\boldsymbol{\delta}, \boldsymbol{\Sigma})$ является неподвижной точкой этого отображения. Эта неподвижная точка может быть вычислена с помощью **итерационного SUR** *(iterative SUR)*. Пусть $(\widehat{\boldsymbol{\delta}}^{(j)}, \widehat{\boldsymbol{\Sigma}}^{(j)})$ — оценка $(\boldsymbol{\delta}_0, \boldsymbol{\Sigma}_0)$ на шаге j. Оценка на следующем шаге вычисляется по формуле:

$$\widehat{\boldsymbol{\delta}}^{(j+1)} = \widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{\Sigma}}^{(j)}), \quad \widehat{\boldsymbol{\Sigma}}^{(j+1)} = \widehat{\boldsymbol{\Sigma}}(\widehat{\boldsymbol{\delta}}^{(j+1)}). \tag{8.5.33}$$

Первая часть этой итерации состоит в получении SUR-оценки при заданной $\widehat{\boldsymbol{\Sigma}}^{(j)}$, в то время как вторая часть обновляет полученную ранее оценку $\boldsymbol{\Sigma}_0$ с использованием текущей оценки $\boldsymbol{\delta}_0$. Если этот процесс сходится, то его результат — решение условий первого порядка.

Контрольные вопросы

1. (Вывод $f(\boldsymbol{y}_t|\boldsymbol{x}_t)$ из $f(\boldsymbol{\varepsilon}_t|\boldsymbol{x}_t)$.) В этом упражнении мы хотим вывести функцию правдоподобия для наблюдения t для FIML-оценки модели. При этом мы будем использовать следующий результат из теории вероятностей:

   Формула замены переменной: *Пусть $\boldsymbol{\varepsilon}$ — непрерывный M-мерный случайный вектор с функцией плотности $f_\varepsilon(\boldsymbol{\varepsilon})$ и $\boldsymbol{g} : \mathbb{R}^M \to \to \mathbb{R}^M$ — взаимно однозначное дифференцируемое отображение на открытом множестве S, при этом обратное отображение обозначается как $\boldsymbol{g}^{-1}(\cdot)$. Пусть $M \times M$ матрица $\boldsymbol{J}(\varepsilon)$ — матрица Якоби $\frac{\partial \boldsymbol{g}(\varepsilon)}{\partial \boldsymbol{\varepsilon}'}$ ($M \times M$ матрица частных производных, в которой элемент (i, j) равен $\frac{\partial g_i(\varepsilon)}{\partial \varepsilon_j}$). Пусть $|\boldsymbol{J}(\varepsilon)|$ не равен нулю в любой точке S. Тогда функция плотности $\boldsymbol{y} = \boldsymbol{g}(\varepsilon)$ задается как*

   $$f(\boldsymbol{y}) = \frac{f_\varepsilon(\boldsymbol{g}^{-1}(\boldsymbol{y}))}{\mathrm{abs}(|\boldsymbol{J}(\boldsymbol{g}^{-1}(\boldsymbol{y}))|)}.$$

   (Заметьте, что $|\boldsymbol{J}(\boldsymbol{g}^{-1}(\boldsymbol{y}))|$ — это определитель, который может как быть, так и не быть положительным.) Используя этот факт, выведите функцию плотности $f(\boldsymbol{y}_t|\boldsymbol{x}_t)$ из $f(\boldsymbol{\varepsilon}_t|\boldsymbol{x}_t)$. Также проверьте, что соответствующий логарифм функции правдоподобия задается (8.5.16). **Указание:** *Обратное отображение* $g^{-1}(y_t)$ *представляет собой левую часть* (8.5.8). *Следовательно,* $J(\varepsilon_t) = \Gamma_0^{-1}$. *Кроме того,* $\log(\mathrm{abs}(|\Gamma|)) = \frac{1}{2}\log(|\Gamma|^2)$.

2. (Инструменты за пределами системы.) Рассмотрим полную систему одновременных уравнений (8.5.8) с $|\Gamma_0| \neq 0$. Пусть одна из K предопределенных переменных, скажем x_{tK}, не входит ни в одно из M уравнений. Пусть $\widehat{E}^*(y|x)$ — проекция OLS y на x. Покажите, что $\widehat{E}^*(y_{tm}|x_t) = \widehat{E}^*(y_{tm}|\widetilde{x}_t)$, где $\widetilde{x}_t = (x_{t1}, \ldots, x_{t,K-1})'$ (так что $x_t = (\widetilde{x}_t{}'. x_{tK})'$). (Как будет показано в аналитическом упражнении 4, включение x_{tk} в набор инструментов в дополнение к $\widetilde{x}_t$ не влияет на асимптотическую дисперсию FIML-оценки δ_0; хотя при этом свойства в малых выборках, вероятно, ухудшатся.)

3. Вспомните, что в примере 8.1 y_{t1} и y_{t2} являются одной и той же переменной. В SUR-модели такого произойти не может: никакие два элемента из $(y_{t1}, y_{t2}, \ldots, y_{tM})$ не являются одной и той же переменной, если выполнены предположения модели. Докажите это. **Указание:** *(Доказательство от противного.) Предположите, что* y_{t1} *и* y_{t2} *являются одной и той же переменной. Тогда*
$$\varepsilon_{t1} - \varepsilon_{t2} = z_{t1}'\delta_{0,1} - z_{t2}'\delta_{0,2}.$$
Так как z_{t1} *и* z_{t2} *— подвекторы* x_t, *правая часть может быть переписана как* $x_t'\alpha$ *для некоторого* K*-мерного вектора* α. *Используйте условие ортогональности* $\mathrm{E}(x_t\varepsilon_t') = 0$ *и невырожденность* $\mathrm{E}(x_t x_t')$, *чтобы показать, что* $\alpha = 0$.

8.6. LIML

Определение LIML

Преимуществом метода FIML является то, что он позволяет использовать всю информацию, содержащуюся в полной системе одновременных уравнений. С другой стороны, это же является его слабым местом, поскольку оценка становится несостоятельной, если какая-либо часть системы неправильно специфицирована (этим же недостатком страдают все системные или совместные методы оценивания). Если вы уверены, что интересующее вас уравнение специфицировано правильно, а относительно прочих уравнений системы такой уверенности нет, возможно, будет разумнее применить методы оценивания отдельного уравнения (такие, как 2SLS). Остальная часть этого параграфа посвящена выводу ML-оценки, которая называется **оценкой максимального правдоподобия с ограниченной информацией** (*limited information maximum likelihood*), (**LIML**) и которая представляет собой ML аналог 2SLS-оценки.

Пусть нас интересует m-е уравнение системы из M уравнений. L_m регрессоров z_{tm} являются либо эндогенными, либо предопределенными. Пусть $\widetilde{y}_t$ — вектор, состоящий из M_m эндогенных переменных, а $\widetilde{x}_t$ —

вектор, состоящий из K_m предопределенных переменных, где $M_m+K_m = L_m$. δ_{0m} разбивается аналогичным образом как $\delta_{0m} = (\gamma'_{0m}, \beta'_{0m})'$. Тогда m-е уравнение может быть записано как

$$y_{tm} = \underset{(1\times L_m)}{z'_{tm}} \underset{(L_m\times 1)}{\delta_{0m}} + \varepsilon_{tm} = \underset{(1\times M_m)}{\widetilde{y}'_t} \underset{(M_m\times 1)}{\gamma_{0m}} + \underset{(1\times K_m)}{\widetilde{x}'_t} \underset{(K_m\times 1)}{\beta_{0m}} + \varepsilon_{tm}. \quad (8.6.1)$$

Очевидно, что это уравнение, рассматриваемое в изоляции, представляет собой неполную систему из-за существования M_m включенных эндогенных переменных $\widetilde{y}_t$. Однако если это уравнение дополнить M_m уравнениями в приведенной форме для $\widetilde{y}_t$ из (8.5.9), то система из этих $1+M_m$ уравнений будет полной системой одновременных уравнений. Чтобы объяснить эту идею более подробно, рассмотрим $\mathbf{\Pi}_0$ — соответствующую матрицу коэффициентов приведенной формы системы уравнений. Возьмем из $\mathbf{\Pi}_0$ соответствующие M_m столбцов и запишем уравнения в приведенной форме для M_m включенных в уравнение эндогенных переменных:

$$\underset{(M_m\times 1)}{\widetilde{y}_t} = \underset{(M_m\times K)}{\widetilde{\mathbf{\Pi}}'_0} \underset{(K\times 1)}{x_t} + \underset{(M_m\times 1)}{\widetilde{v}_t}. \quad (8.6.2)$$

Совмещая (8.6.1) и (8.6.2), получим систему из $1+M_m$ уравнений:

$$\underset{((1+M_m)\times(1+M_m))}{\overline{\mathbf{\Gamma}}_0} \underset{((1+M_m)\times 1)}{\overline{y}_t} + \underset{((1+M_m)\times K)}{\overline{B}_0} \underset{(K\times 1)}{x_t} = \underset{((1+M_m)\times 1)}{\overline{\varepsilon}_t}, \quad (8.6.3)$$

где

$$\overline{y}_t \equiv \begin{bmatrix} y_{tm} \\ \underset{(M_m\times 1)}{\widetilde{y}_t} \end{bmatrix}, \quad \overline{\varepsilon}_t \equiv \begin{bmatrix} \varepsilon_{tm} \\ \underset{(M_m\times 1)}{\widetilde{v}_t} \end{bmatrix},$$

$$\overline{\mathbf{\Gamma}}_0 \equiv \begin{bmatrix} 1 & -\underset{(1\times M_m)}{\gamma'_{0m}} \\ \underset{(M_m\times 1)}{\mathbf{0}} & I_{M_m} \end{bmatrix}, \quad \overline{B}_0 \equiv \begin{bmatrix} -\underset{(1\times K_m)}{\beta'_{0m}} \quad \underset{(1\times(K-K_m))}{\mathbf{0}'} \\ -\underset{(M_m\times K)}{\widetilde{\mathbf{\Pi}}'_0} \end{bmatrix}. \quad (8.6.4)$$

Здесь мы без потери общности предположили, что включенные в уравнение предопределенные переменные $\widetilde{x}_t$ являются первыми K_m элементами x_t. Система уравнений (8.6.3) является полной системой $1+M_m$ одновременных уравнений, поскольку

$$|\overline{\mathbf{\Gamma}}_0| = 1 \neq 0. \quad (8.6.5)$$

В качестве примера рассмотрим пример 8.2. Включенная эндогенная переменная в уравнении заработой платы — это S_t. Добавим к уравнению заработной платы уравнение в приведенной форме для S_t: регрессию S_t на константу, IQ_t и MED_t. Матрица $\overline{\mathbf{\Gamma}}_0$ для системы из двух уравнений имеет вид:

$$\overline{\mathbf{\Gamma}}_0 = \begin{bmatrix} 1 & -\gamma_{11} \\ 0 & 1 \end{bmatrix}.$$

Гипотетическое значение $\overline{\mathbf{\Gamma}}$ для $\overline{\mathbf{\Gamma}}_0$ имеет такую же структуру, как и $\overline{\mathbf{\Gamma}}_0$ в (8.6.4). Поэтому $|\overline{\mathbf{\Gamma}}| = 1$. Заменяя в целевой функции FIML (8.5.20) $(\boldsymbol{\delta}, \boldsymbol{B})$ на $(\boldsymbol{\gamma}_m, \boldsymbol{\beta}_m, \overline{\mathbf{\Pi}})$, $\mathbf{\Gamma}$ на $\overline{\mathbf{\Gamma}}$ и $\mathbf{\Sigma}$ на $\overline{\mathbf{\Sigma}}$ и используя условие $|\overline{\mathbf{\Gamma}}| = 1$, мы получаем целевую функцию LIML:

$$Q_n(\boldsymbol{\gamma}_m, \boldsymbol{\beta}_m, \overline{\mathbf{\Pi}}, \overline{\mathbf{\Sigma}}) = -\frac{M_m}{2}\log(2\pi) - \frac{1}{2}\log(|\overline{\mathbf{\Sigma}}|) - \\ - \frac{1}{2n}\sum_{t=1}^{n}\left[\overline{\mathbf{\Gamma}}\overline{\boldsymbol{y}}_t + \overline{\boldsymbol{B}}\boldsymbol{x}_t\right]' \Sigma\overline{\mathbf{\Sigma}}^{-1}\left[\overline{\mathbf{\Gamma}}\overline{\boldsymbol{y}}_t + \overline{\boldsymbol{B}}\boldsymbol{x}_t\right], \quad (8.6.6)$$

где $\overline{\mathbf{\Sigma}}$ — ковариационная матрица $\overline{\boldsymbol{\varepsilon}}_t$ размерности $(1 + M_m) \times (1 + M_m)$. Пусть $(\widehat{\boldsymbol{\gamma}}_m, \widehat{\boldsymbol{\beta}}_m, \widehat{\overline{\mathbf{\Pi}}}, \widehat{\overline{\mathbf{\Sigma}}})$ — FIML-оценка этой системы. **LIML-оценка** (*LIML estimator*) для $\boldsymbol{\delta}_{0m} = (\boldsymbol{\gamma}_{0m}', \boldsymbol{\beta}_{0m}')'$ есть $(\widehat{\boldsymbol{\gamma}}_m', \widehat{\boldsymbol{\beta}}_m')'$. Представленный способ вывода LIML-оценки принадлежит Пагану [Pagan, 1979] и отличается от способа, предложенного Андерсоном и Рубином [Anderson and Rubin, 1949]. Поскольку LIML-оценивание является частным случаем FIML-оценивания, предложение 8.1 гарантирует состоятельность и асимптотическую нормальность LIML-оценки, даже в случае, когда не выполняется предположение о нормальности ошибок.

Вычисление LIML

Для случая SUR-модели мы показали, что ML-оценка может быть вычислена с помощью итерационной SUR-процедуры. Если внимательно взглянуть на доказательство, то можно заметить, что оно не зависит от того, является ли матрица $\mathbf{\Gamma}$ в целевой функции SUR (8.5.29) единичной матрицей. Единственным требованием является равенство определителя матрицы $\mathbf{\Gamma}$ константе. Поэтому та же аргументация может быть применена к целевой функции LIML (8.6.6). То есть мы можем получить LIML-оценку параметра $\boldsymbol{\delta}_{0m}$, применяя итерационную SUR-процедуру к системе (8.6.3), состоящей из $(1 + M_m)$ уравнений. В частности, если $\mathbf{\Sigma}_0$ — известная матрица, то LIML-оценка является SUR-оценкой. Наш вывод о том, что SUR может быть использован для оценки коэффициентов при включенных *эндогенных* переменных, является удивительным, но верным. Причина этого в следующем: ошибка оценки $\widehat{\boldsymbol{\delta}}_m - \boldsymbol{\delta}_{0m}$ зависит не только от корреляции между включенными эндогенными переменными $\widetilde{\boldsymbol{y}}_t$ и ошибкой ε_{tm}, но и от корреляции между $\widetilde{\boldsymbol{y}}_t$ и $\widetilde{\boldsymbol{v}}_t$. Эти две корреляции взаимно уничтожают друг друга. Этот факт проверяется в контрольном вопросе 1 для простого примера.

Однако эта итерационная SUR-процедура обычно не используется для вычисления LIML-оценки, так как в случае LIML решение может быть найдено аналитически. Система одновременных уравнений (8.6.3) обладает двумя отличительными чертами. Первая — это особая структура $\overline{\mathbf{\Gamma}}$, что мы уже отмечали. Вторая особенность состоит в том, что на строки матрицы $\overline{B}$ не наложены исключающие ограничения, соответствующие включенным эндогенным переменным $\widetilde{y}_t$. Благодаря этим особенностям LIML-оценка $\widehat{\delta}_m = (\widehat{\gamma}'_m, \widehat{\beta}'_m)'$ и статистика максимального правдоподобия (8.5.25) могут быть вычислены в явном виде.

Чтобы записать в явном виде LIML-оценку и статистику отношения правдоподобий, нам необходимо ввести новые обозначения. Пусть $Z_m, \widetilde{Y}, \widetilde{X}, X$ и y_m — матрицы и вектор данных, которые соотносятся с $z_{tm}, \widetilde{y}_t, \widetilde{x}_t, x_t$ и y_{tm} соответственно следующим образом:

$$\underset{(n\times L_m)}{Z_m} \equiv \begin{bmatrix} z'_{1m} \\ \vdots \\ z'_{nm} \end{bmatrix}, \quad \underset{(n\times M_m)}{\widetilde{Y}} \equiv \begin{bmatrix} \widetilde{y}'_1 \\ \vdots \\ \widetilde{y}'_n \end{bmatrix}, \quad \underset{(n\times K_m)}{\widetilde{X}} \equiv \begin{bmatrix} \widetilde{x}'_1 \\ \vdots \\ \widetilde{x}'_n \end{bmatrix},$$
$$\underset{(n\times K)}{X} \equiv \begin{bmatrix} x'_1 \\ \vdots \\ x'_n \end{bmatrix}, \quad \underset{(n\times 1)}{y_m} \equiv \begin{bmatrix} y_{1m} \\ \vdots \\ y_{nm} \end{bmatrix}, \tag{8.6.7}$$

и пусть M и $\widetilde{M}$ — аннуляторы, основанные на X и $\widetilde{X}$ соответственно:

$$\underset{(n\times n)}{M} \equiv I_n - X(X'X)^{-1}X', \quad \underset{(n\times n)}{\widetilde{M}} \equiv I_n - \widetilde{X}(\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'. \tag{8.6.8}$$

Тогда LIML-оценка $\widehat{\delta}_m$ может быть записана как

$$\widehat{\delta}_m \equiv \left[Z'_m(I_n - kM)Z_m\right]^{-1} Z'_m(I_n - kM)y_m. \tag{8.6.9}$$

где k — наименьший характеристический корень

$$\widetilde{W}W^{-1}, \tag{8.6.10}$$

где $\widetilde{W}$ и W^{-1} — матрицы размерности $(1+M_m)\times(1+M_m)$, определяемые как

$$\widetilde{W} \equiv [y_m \vdots \widetilde{Y}]'\widetilde{M}[y_m \vdots \widetilde{Y}], \quad W \equiv [y_m \vdots \widetilde{Y}]'M[y_m \vdots \widetilde{Y}]. \tag{8.6.11}$$

(Вывод можно найти, например, у Дэвидсона и Мак-Киннона [Davidson and MacKinnon, 1993, pp. 645–649]). Статистика отношения правдоподобий (8.5.25) для тестирования сверхидентифицирующих ограничений сводится к

$$LR = n \log k. \tag{8.6.12}$$

Так как в дополнительных M_m уравнениях (8.6.2) не содержится сверхидентифицирующих ограничений, представляющее интерес уравнение является единственным источником сверхидентифицирующих ограничений, количество которых равно $K - L_m$. Поэтому из утверждения 8.1 следует, что эта статистика для проверки сверхидентифицирующих ограничений асимптотически распределена как $\chi^2(K - L_m)$. Когда уравнение идентифицировано точно ($K = L_m$), LR статистика должна быть равна нулю. Кроме того, может быть показано, что в этом случае $k = 1$ (подробности см., например, в Дэвидсоне и Мак-Кинноне [Davidson and MacKinnon, 1993, p. 647]).

Если мы необязательно требуем, чтобы k было именно таким, как только что было определено, то оценка (8.6.9) называется **оценкой *k*-класса** *(k-class estimator)*. Таким образом, LIML является оценкой k-класса, с k, определенным выше. Изучение формулы для 2SLS-оценки в терминах матриц данных (см. (3.8.3′) на с. 259) немедленно показывает, что 2SLS-оценка является оценкой k-класса с $k = 1$, а OLS-оценка — оценка k-класса с $k = 0$. Следовательно, LIML и 2SLS численно совпадают, когда уравнение точно идентифицировано (так что $k = 1$).

Мы уже показали, что LIML-оценка является состоятельной и асимптотически нормальной. Кроме того, используя явную формулу (8.6.9), можно показать, что асимптотическое распределение LIML-оценки δ_{0m} совпадает с асимптотическим распределением оценки 2SLS (см.: [Amemiya, 1985, Section 7.3.4]). Поэтому формула для асимптотической дисперсии LIML дается в (3.8.4), а ее состоятельная оценка — в (3.8.5).

Сравнение LIML и 2SLS

Поскольку LIML и 2SLS имеют одинаковое асимптотическое распределение, выбор в пользу одного из этих методов нельзя сделать на основании их асимптотических свойств. В конечных выборках LIML обладает свойством инвариантности, тогда как 2SLS таким свойством не обладает. Кроме того, из обширной литературы, посвященной конечным выборкам (см., например, Джадж и др. [Judge et al., 1985, Section 15.4]), можно заключить, что LIML предпочтительнее 2SLS. Ниже перечислены основные выводы из этой литературы. Однако в них предполагается, что предопределенные переменные x_t — фиксированные константы.

- Пусть ошибки нормально распределены. p-й момент 2SLS-оценки существует, если и только если $p < K - L_m + 1$, где K — количество предопределенных переменных и L_m — количество регрессоров в m-м уравнении. Таким образом, если уравнение идентифицируемо точно (то есть если $K = L_m$), то у 2SLS-оценки даже не существует математическое ожидание. Это утверждение справедливо, даже если ошибки имеют нормальное распределение, у которого все моменты конечны.

- У LIML-оценки нет конечных моментов, даже при предположении о нормальности ошибок.

- Однако большинство симуляций методом Монте-Карло (см., например: [Anderson, Kanitomo, and Sawa, 1982]) демонстрирует, что LIML сходится к нормальному распределению быстрее, чем 2SLS.

Контрольные вопросы

1. (Почему SUR может быть использован для оценивания уравнений в структурной форме.) Чтобы понять, почему SUR можно использовать для оценивания коэффициентов при включенных эндогенных переменных, рассмотрим следующий простой пример:

$$\begin{cases} y_{t1} = \gamma_0 y_{t2} + \varepsilon_{t1}, \\ y_{t2} = \beta_0 x_t + v_t, \end{cases}$$

где x_t — предопределенная переменная. Для простоты предположим, что

$$\underset{(2\times 2)}{\mathbf{\Sigma}_0} \equiv \mathrm{E}\left(\begin{bmatrix} \varepsilon_{t1} \\ v_t \end{bmatrix} \begin{bmatrix} \varepsilon_{t1} & v_t \end{bmatrix}\right)$$

известна. Пусть $(\widehat{\gamma}, \widehat{\beta})$ — SUR-оценка истинных значений параметров (γ_0, β_0).

(a) Покажите, что

$$\begin{bmatrix} \widehat{\gamma} \\ \widehat{\beta} \end{bmatrix} - \begin{bmatrix} \gamma_0 \\ \beta_0 \end{bmatrix} = \begin{bmatrix} \sigma^{11}\frac{1}{n}\sum y_{t2}^2 & \sigma^{12}\frac{1}{n}\sum y_{t2}x_t \\ \sigma^{21}\frac{1}{n}\sum x_t y_{t2} & \sigma^{22}\frac{1}{n}\sum x_t^2 \end{bmatrix}^{-1} \times$$
$$\times \begin{bmatrix} \sigma^{11}\frac{1}{n}\sum y_{t2}\varepsilon_{t1} + \sigma^{12}\frac{1}{n}\sum y_{t2}v_t \\ \sigma^{21}\frac{1}{n}\sum x_t\varepsilon_{t1} + \sigma^{22}\frac{1}{n}\sum x_t v_t \end{bmatrix},$$

где σ^{mh} — элемент матрицы $\mathbf{\Sigma}_0^{-1}$ с номером (m, h). **Указание:** *Формула для SUR-оценки задана* (4.5.12), (4.5.13′) *и* (4.5.14′) *на* *с. 309.*

(b) Покажите, что

$$\sigma^{11}\frac{1}{n}\sum_{t=1}^{n} y_{t2}\varepsilon_{t1} + \sigma^{12}\frac{1}{n}\sum_{t=1}^{n} y_{t2}v_t$$

сходится по вероятности к нулю. **Указание:**

$$\mathrm{E}(y_{t2}\varepsilon_{t1}) = \mathrm{E}(\varepsilon_{t1}v_t) = \sigma_{12}, \mathrm{E}(y_{t2}v_t) = \mathrm{E}(v_t^2) = \sigma_{22}.$$

8.7. Сериально коррелированные наблюдения

При выводе асимптотической дисперсии ML-оценок мы до сих пор предполагали независимость и одинаковую распределенность наблюдений. Этот параграф посвящен ML-оцениванию в случае, когда наблюдения сериально коррелированы.

Два вопроса

Когда между наблюдениями имеется сериальная корреляция, возникают два важных вопроса. Первый: является ли ML-оценка, выведенная в предположении о независимости и одинаковой распределенности наблюдений, состоятельной и асимптотически нормальной, если наблюдения сериально коррелированы? Предположим, что вектор наблюдений $\boldsymbol{w}_t$ — стационарный и эргодический, но необязательно i.i.d. Рассмотрим ML-оценку, которая максимизирует

$$Q_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{t=1}^{n} \log f(\boldsymbol{w}_t; \boldsymbol{\theta}), \tag{8.7.1}$$

где $\log f(\boldsymbol{w}_t; \boldsymbol{\theta})$ — логарифмическое правдоподобие для наблюдения t. Если бы наблюдения были независимыми и одинаково распределенными, то эта целевая функция представляла бы собой умноженное на $1/n$ логарифмическое правдоподобие выборки $(\boldsymbol{w}_1, \ldots, \boldsymbol{w}_n)$. Даже при том что логарифмическая функция правдоподобия для наблюдения t, $\log f(\boldsymbol{w}_t; \boldsymbol{\theta})$, специфицирована правильно, эта целевая функция специфицирована неправильно, поскольку ошибочно предполагает, что $\{\boldsymbol{w}_t\}$ — i.i.d. Так как целевая функция специфицирована неправильно, ML-оценка, которая максимизирует эту целевую функцию, представляет собой квази-ML-оценку. Является ли эта квази-ML-оценка, базирующаяся на неверной предпосылке об отсутствии сериальной корреляции, состоятельной и асимптотически нормальной?

Мы в действительности уже дали ответ на этот первый вопрос в предыдущей главе. Из утверждений 7.5 и 7.6 следует, что квази-ML-оценка является состоятельной, так как она является частным случаем M-оценки с $m(\boldsymbol{w}_t, \boldsymbol{\theta})$, равной логарифмическому правдоподобию для наблюдения t. Для установления асимптотической нормальности здесь можно воспользоваться утверждением 7.8, в котором перечислены условия, при которых M-оценка асимптотически нормально распределена. Просто воспроизводя заключение утверждения 7.8, получаем, что выражение для асимптотической дисперсии задается как

$$\mathrm{Avar}(\widehat{\boldsymbol{\theta}}) = \left(\mathrm{E}\left[\boldsymbol{H}(\boldsymbol{w}_t; \boldsymbol{\theta}_0)\right]\right)^{-1}\Sigma\left(\mathrm{E}\left[\boldsymbol{H}(\boldsymbol{w}_t; \boldsymbol{\theta}_0)\right]\right)^{-1}. \tag{8.7.2}$$

где $\boldsymbol{\theta}_0$ — истинное значение параметра $\boldsymbol{\theta}$, $\boldsymbol{H}(\boldsymbol{w}_t, \boldsymbol{\theta})$ — гессиан для наблюдения t, и $\boldsymbol{\Sigma}$ — долговременная дисперсия вклада для наблюдения t, вычисленная в точке $\boldsymbol{\theta}_0$, $\boldsymbol{s}(\boldsymbol{w}_t, \boldsymbol{\theta}_0)$. Состоятельная оценка этой асимптотической дисперсии может быть найдена как

$$\widehat{\mathrm{Avar}(\widehat{\boldsymbol{\theta}})} = \left\{\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{H}(\boldsymbol{w}_t; \widehat{\boldsymbol{\theta}})\right\}^{-1} \widehat{\boldsymbol{\Sigma}}\left\{\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{H}(\boldsymbol{w}_t; \widehat{\boldsymbol{\theta}})\right\}^{-1}. \tag{8.7.3}$$

где $\widehat{\mathbf{\Sigma}}$ — состоятельная оценка $\mathbf{\Sigma}$. Любой «непараметрический» метод оценивания, обсуждавшийся в главе 6 (например, VARHAC), может быть использован для расчета $\widehat{\mathbf{\Sigma}}$ по оцененной последовательности $\{\boldsymbol{s}(\boldsymbol{w}_t, \widehat{\boldsymbol{\theta}}\}$. Кроме того, нет необходимости в параметрической спецификации сериальной корреляции в $\{\boldsymbol{s}(\boldsymbol{w}_t, \boldsymbol{\theta}_0\}$.

Второй вопрос: как выглядит функция правдоподобия, которая учитывает наличие сериальной корреляции, и каковы свойства «истинной» ML-оценки, которая максимизирует эту целевую функцию? Этот вопрос может быть адресован как условному, так и безусловному методу максимального правдоподобия. Говоря об условном ML, мы разбиваем вектор наблюдений $\boldsymbol{w}_t$ на две группы (зависимая переменная y_t и регрессоры $\boldsymbol{x}_t$) и исследуем условное распределение $(y_t, \ldots, y_n)$ при условии $(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$. Параметризация этого условного распределения довольно проста, когда $\{\boldsymbol{w}_t\}$ независимы и одинаково распределены, поскольку это распределение может быть записано как произведение по всем t условных распределений y_t при условии $\boldsymbol{x}_t$. На самом деле случай независимых и одинаково распределенных величин — это то, что мы рассматривали в предыдущих главах. В случае, когда $\{\boldsymbol{w}_t\}$ сериально коррелированы, исследователь не знает динамической взаимосвязи между y_t и $\boldsymbol{x}_t$ достаточно хорошо, чтобы записать условное распределение[1]. В оставшейся части этого параграфа мы будем рассматривать только примеры безусловного ML-оценивания с сериально коррелированными наблюдениями.

Безусловный ML для зависимых наблюдений

Мы начнем с простейшего одномерного случая. Пусть $(y_0, y_1, \ldots, y_n)$ — имеющаяся в распоряжении выборка. (Для удобства обозначений мы предполагаем, что в выборку включается наблюдение для $t = 0$.) Когда между наблюдениями существует сериальная корреляция, функция плотности выборки больше не может быть записана как произведение индивидуальных функций плотности. Однако она может быть записана в терминах функций условной плотности. Совместная функция плотности (y_0, y_1) может быть записана следующим образом:

$$f(y_0, y_1) = f(y_1|y_0)f(y_0). \tag{8.7.4}$$

Аналогично совместная функция плотности для (y_0, y_1, y_2):

$$f(y_0, y_1, y_2) = f(y_2|y_1, y_0)f(y_0, y_1). \tag{8.7.5}$$

[1]Обсуждение условного ML в случаях, когда полная спецификация динамической взаимосвязи невозможна или нежелательна, см. у Вулдриджа [Wooldridge, 1994, Section 5]. Обсуждение возможных ограничений, накладываемых на динамические взаимосвязи между переменными (таких как «причинность по Грейнджеру» и «слабая экзогенность») см., например, в работе Дэвидсона и Мак-Киннона [Davidson and MacKinnon, 1993, Section 18.2].

Подставляя (8.7.4) в (8.7.5), получаем:

$$f(y_0, y_1, y_2) = f(y_2|y_1, y_0)f(y_1|y_0)f(y_0). \tag{8.7.6}$$

Повторяя такую последовательную подстановку, в итоге получаем:

$$f(y_0, y_1, \dots, y_n) = \prod_{t=1}^{n} f(y_t|y_{t-1}, \dots, y_1, y_0)f(y_0). \tag{8.7.7}$$

Если условная функция плотности $f(y_t|y_{t-1}, \dots, y_1, y_0)$ и безусловная функция плотности $f(y_0)$ могут быть параметризованы с помощью конечного вектора параметра $\boldsymbol{\theta}$, то тогда умноженное на $1/n$ логарифмическое правдоподобие выборки $(y_0, y_1, \dots, y_n)$ может быть записано как

$$\frac{1}{n}\sum_{t=1}^{n} \log f(y_t|y_{t-1}, \dots, y_1, y_0; \boldsymbol{\theta}) + \frac{1}{n}\log f(y_0; \boldsymbol{\theta}). \tag{8.7.8}$$

Оценка максимального правдоподобия $\widetilde{\boldsymbol{\theta}}$ истинного вектора параметров $\boldsymbol{\theta}_0$ — это такое значение $\boldsymbol{\theta}$, которое максимизирует эту логарифмическую функцию правдоподобия. Такую оценку мы будем называть **точной ML-оценкой** (*exact ML estimator*).

ML-оценка процесса AR(1)

В параграфе 6.2 мы довольно подробно рассмотрели процесс авторегрессии первого порядка AR(1)). Если мы предположим, что ошибки распределены нормально, мы получим **гауссовский AR(1)-процесс** (*Gaussian AR(1) process*):

$$y_t = c_0 + \phi_0 y_{t-1} + \varepsilon_t, \tag{8.7.9}$$

где $\varepsilon_t \sim i.i.d.N(0, \sigma_0^2)$. Мы предполагаем, что условие стационарности процесса $|\phi_0| < 1$ выполнено. Поэтому $\{y_t\}$ — эргодический стационарный процесс, в соответствии с утверждением 6.1 (b). Из параграфа 6.2 мы знаем, что безусловное математическое ожидание y_t равно $c_0/(1-\phi_0)$, а безусловная дисперсия равна $\sigma_0^2/(1-\phi_0^2)$. Кроме того, поскольку y_t может быть представлен как взвешенное среднее $(\varepsilon_t, \varepsilon_{t-1}, \dots)$, его распределение нормально при предположении о нормальности ε_t. Поэтому

$$y_0 \sim N\left(\frac{c_0}{1-\phi_0}, \frac{\sigma_0^2}{1-\phi_0^2}\right). \tag{8.7.10}$$

Кроме того, из (8.7.9) и предположения о нормальности ошибок непосредственно следует, что

$$y_t|(y_{t-1}, \dots, y_1, y_0) \sim N(c_0 + \phi_0 y_{t-1}, \sigma_0^2). \tag{8.7.11}$$

Поэтому, используя формулу (8.7.7), совместную функцию плотности $(y_0, y_1, \dots, y_n)$ для гипотетического значения параметра $\boldsymbol{\theta} \equiv (c, \phi, \sigma^2)'$ можно записать как

$$f(y_0, y_1, \dots, y_n; \boldsymbol{\theta}) = \prod_{t=1}^{n} \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2} (y_t - c - \phi y_{t-1})^2 \right] \right\} \times$$
$$\times \frac{1}{\sqrt{2\pi \frac{\sigma^2}{1-\phi^2}}} \exp\left[\frac{-\left(y_0 - \frac{c}{1-\phi}\right)^2}{2\frac{\sigma^2}{1-\phi^2}} \right]. \quad (8.7.12)$$

Прологарифмировав и разделив обе части выражения на n, получаем целевую функцию для точной ML-оценки:

$$\widetilde{Q}_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{t=1}^{n} \left\{ -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} (y_t - c - \phi y_{t-1})^2 \right\} +$$
$$+ \frac{1}{n} \left\{ -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log\left(\frac{\sigma^2}{1-\phi^2} \right) - \left[\frac{\left(y_0 - \frac{c}{1-\phi}\right)^2}{2\frac{\sigma^2}{1-\phi^2}} \right] \right\}. \quad (8.7.13)$$

Точная ML-оценка $\widetilde{\boldsymbol{\theta}}$ параметров AR(1)-процесса — это такое значение $\boldsymbol{\theta}$, которое максимизирует эту целевую функцию. В предположении, что решение внутреннее, эта оценка является решением условий первого порядка (уравнений правдоподобия), полученных путем приравнивания $\frac{\partial \widetilde{Q}_n}{\partial \boldsymbol{\theta}}$ к нулю. Они представляют собой систему нелинейных уравнений по $\boldsymbol{\theta}$, так что поиск $\widetilde{\boldsymbol{\theta}}$ требует привлечения итерационных алгоритмов, таких как рассмотренные в параграфе 7.5.

Условное ML-оценивание AR(1)-процессов

Источником нелинейности уравнений правдоподобия является логарифм правдоподобия y_0. Как альтернатива, опустим эту составляющую и найдем максимум следующего выражения:

$$Q_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{t=1}^{n} \left\{ -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma^2) - \frac{1}{2\sigma^2} (y_t - c - \phi y_{t-1})^2 \right\}. \quad (8.7.14)$$

Используя соотношение $f(y_0, y_1, \dots, y_n) = f(y_1, \dots, y_n | y_0) f(y_0)$ и (8.7.12), мы видим, что эта целевая функция равна умноженному на $1/n$ логарифмическому правдоподобию, условному относительно первого наблюдения y_0:

$$f(y_1, y_2, \dots, y_n | y_0; \boldsymbol{\theta}) = \prod_{t=1}^{n} \left\{ \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2} (y_t - c - \phi y_{t-1})^2 \right] \right\}. \quad (8.7.15)$$

Пусть $\widehat{\boldsymbol{\theta}}$ — значение параметра $\boldsymbol{\theta}$, которое максимизирует условное логарифмическое правдоподобие (8.7.14) (при условии y_0). Эта ML-оценка будет поэтому называться **y_0-условной ML-оценкой** (*y_0-conditional ML estimator*). То, что здесь оценка ищется при условии y_0, концептуально отличает эту оценку от ML-оценок, рассмотренных в предыдущих параграфах, где оценка выводилась при условии $\boldsymbol{x}_t$.

Прежде чем анализировать связь между двумя этими условными ML-оценками (точная ML-оценка $\widetilde{\boldsymbol{\theta}}$ и y_0-условная ML-оценка $\widehat{\boldsymbol{\theta}}$), мы выведем два асимптотических свойства y_0-условной ML-оценки. Первый вывод основан на поиске выражения для оценки в явном виде. Вспомним из параграфа 1.5 и примера 7.2, что целевая функция ML линейной регрессионной модели $\boldsymbol{y}_t = \boldsymbol{x}_t\ \boldsymbol{\beta}_0 + \varepsilon_t$ для независимых одинаково распределенных наблюдений представляется как

$$\frac{1}{n}\sum_{t=1}^{n}\left\{-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(y_t - \boldsymbol{x}_t'\boldsymbol{\beta})^2\right\}. \tag{8.7.16}$$

В параграфе 1.5 было показано (и это очень легко показать), что ML-оценка $\boldsymbol{\beta}_0$ совпадает с OLS-оценкой, а ML-оценка σ_0^2 равна сумме квадратов остатков, деленной на размер выборки. Теперь, при $\boldsymbol{x}_t = (1, y_{t-1})'$ и $\boldsymbol{\beta}_0 = (c_0, \phi_0)'$, целевая функция линейной регрессионной модели упрощается до $Q_n(\boldsymbol{\theta})$ (8.7.14) — условного логарифмического правдоподобия для AR(1)-процесса. Поэтому алгебраически условная ML-оценка $(\widehat{c}, \widehat{\phi})$ для (c_0, ϕ_0) численно совпадает с OLS-оценкой коэффициентов регрессии y_t на y_{t-1}, и условная ML-оценка $\widehat{\sigma}^2$ для σ_0 равна сумме квадратов остатков, деленной на n. Что касается асимптотических свойств оценки, то мы показали в параграфе 6.4, что $(\widehat{c}, \widehat{\phi})$ — состоятельная и асимптотически нормальная оценка с асимптотической дисперсией, указанной в утверждении 6.7 для $p = 1$, и что $\widehat{\sigma}^2$ — состоятельная оценка σ_0^2. Поэтому для асимптотической нормальности условной ML-оценки $(\widehat{c}, \widehat{\phi})$, которая максимизирует y_0-условную функцию правдоподобия для нормально распределенных ошибок, предположение о нормальном распределении не требуется.

Второй вывод менее прямой, но, несмотря на это, более поучителен. Как только что было замечено, целевая функция $Q_n(\boldsymbol{\theta})$ совпадает с (8.7.16) при $\boldsymbol{x}_t = (1, y_{t-1})'$ и $\boldsymbol{\beta}_0 = (c_0, \phi_0)'$. Поскольку $\{y_t, \boldsymbol{x}_t\}$ только стационарны и эргодичны, но необязательно независимо и одинаково распределены, условная ML-оценка параметров $(\widehat{c}, \widehat{\phi}, \widehat{\sigma}^2)$ может рассматриваться как квази-ML-оценка, рассмотренная в утверждении 7.6. В примере 7.8 было проверено, что эта целевая функция удовлетворяет условиям утверждения 7.6, если $\mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}_t')$ — невырожденная матрица ($\boldsymbol{x}_t = (1, y_{t-1})'$). В параграфе 6.4 мы показали, что это условие невырожденности выполнено для процесса AR(1), если $\sigma_0^2 > 0$. Таким образом, оценка условного ML состоятельна. Для доказательства асимптотической нормальности рассмотрим условную ML-оценку как M-оценку с $\boldsymbol{w}_t = (y_t, y_{t-1})'$ и

$$m(\boldsymbol{w}_t; \boldsymbol{\theta}) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(y_t - c - \phi y_{t-1})^2. \quad (8.7.17)$$

Следовательно, подходящей теоремой здесь будет утверждение 7.8. Обозначим как $\boldsymbol{s}(\boldsymbol{w}_t; \boldsymbol{\theta})$ вклад наблюдения t, и $\boldsymbol{H}(\boldsymbol{w}_t; \boldsymbol{\theta})$ — гессиан для наблюдения t, связанный с этой m функцией, и примем во внимание условия (1)–(5) утверждения 7.8. Из обсуждения в примере 7.10 следует, что все условия, кроме (3), выполнены, если $\mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}_t')$ — невырожденная матрица (где $\boldsymbol{x}_t = (1, y_{t-1})'$). Как мы только что отметили, это условие выполнено при $\sigma_0^2 > 0$. Нам осталось рассмотреть только выполнение условия (3): $\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \boldsymbol{s}(\boldsymbol{w}_t; \boldsymbol{\theta}_0) \to_d N(\boldsymbol{0}, \boldsymbol{\Sigma})$. В модели AR(1) $f(y_t|y_{t-1}, \ldots, y_1, y_0) = f(y_t|y_{t-1})$. Как вы сможете показать в контрольном вопросе 1, это специфическое свойство совместной функции плотности указывает на то, что последовательность $\{\boldsymbol{s}(\boldsymbol{w}_t; \boldsymbol{\theta}_0)\}$ является мартингал-разностью. Так как $\boldsymbol{s}(\boldsymbol{w}_t; \boldsymbol{\theta}_0)$ является функцией от $\boldsymbol{w}_t = (y_t, y_{t-1})'$ и $\{y_t\}$ стационарна и эргодична, последовательность $\{\boldsymbol{s}(\boldsymbol{w}_t; \boldsymbol{\theta}_0)\}$ также стационарна и эргодична. Используя CLT для эргодических стационарных мартингал-разностей из главы 2, можно заключить, что условие (3) выполняется с

$$\boldsymbol{\Sigma} = \mathrm{E}\big[\boldsymbol{s}(\boldsymbol{w}_t; \boldsymbol{\theta}_0)\, \boldsymbol{s}(\boldsymbol{w}_t; \boldsymbol{\theta}_0)'\big]. \quad (8.7.18)$$

Наконец, легко показать, что для наблюдения t выполняется равенство для информационных матриц $-\mathrm{E}\,[\boldsymbol{H}(\boldsymbol{w}_t; \boldsymbol{\theta}_0)] = \mathrm{E}\,[\boldsymbol{s}(\boldsymbol{w}_t; \boldsymbol{\theta}_0)\boldsymbol{s}(\boldsymbol{w}_t; \boldsymbol{\theta}_0)']$ (доказательство было приведено в примере 7.10). Подставляя это выражение и (8.7.18) в (8.7.2), мы заключаем, что асимптотическая дисперсия y_0-условной ML-оценки $\widehat{\boldsymbol{\theta}}$ истинного значения параметра $\boldsymbol{\theta}_0$ дается выражением:

$$\mathrm{Avar}(\widehat{\boldsymbol{\theta}}) = -\big(\mathrm{E}\big[\boldsymbol{H}(\boldsymbol{w}_t; \boldsymbol{\theta}_0)\big]\big)^{-1} = \big(\mathrm{E}\big[\boldsymbol{s}(\boldsymbol{w}_t; \boldsymbol{\theta}_0)\, \boldsymbol{s}(\boldsymbol{w}_t; \boldsymbol{\theta}_0)'\big]\big)^{-1}. \quad (8.7.19)$$

То есть обычные выводы для правильно специфицированной ML-оценки для независимых и одинаково распределенных наблюдений применимы и для y_0-условной ML-оценки, которая максимизирует правильно специфицированную функцию правдоподобия (8.7.14) для зависимых наблюдений[1].

Разность между (8.7.13) и (8.7.14) представляет логарифмическое правдоподобие y_0, деленное на n. Если размер выборки n достаточно велик, то первое наблюдение y_0 вносит пренебрежимо малый вклад

[1]Это доказательство можно применить и к более общему типу моделей, чем авторегрессионные процессы. См.: Lemma 5.2 в: [Wooldridge, 1994].

в функцию правдоподобия выборки. Точная ML-оценка и y_0-условная ML-оценка имеют одинаковое асимптотическое распределение при условии, что $|\phi_0| < 1$ (см., например: [Fuller, 1996, Sections 8.1 и 8.4]). По этой причине в большинстве случаев параметры авторегрессии оцениваются OLS (y_0-условным ML), а не точным ML.

Условное ML-оценивание процессов AR(p) и VAR(p)

Гауссовский AR(p)-процесс (*Gaussian AR(p) process*) может быть записан как

$$y_t = c_0 + \phi_{01}y_{t-1} + \phi_{02}y_{t-2} + \cdots + \phi_{0p}y_{t-p} + \varepsilon_t, \tag{8.7.20}$$

где $\varepsilon_t \sim i.i.d.N(0, \sigma_0^2)$. Условное ML-оценивание параметров процесса AR(p) абсолютно аналогично случаю AR(1). Если имеется выборка $(y_{-p+1}, y_{-p+2}, \ldots, y_0, y_1, \ldots, y_n)$, то тогда умноженное на $1/n$ логарифмическое правдоподобие этой выборки при условии $(y_{-p+1}, y_{-p+2}, \ldots, y_0)$ задается соотношением (8.7.16), где

$$\boldsymbol{x}_t = (1, y_{t-1}, \ldots, y_{t-p})' \text{ и } \boldsymbol{\beta}_0 = (c_0, \phi_{01}, \phi_{02}, \ldots, \phi_{0p})'.$$

Поэтому условная ML-оценка (условная относительно $(y_{-p+1}, \ldots, y_0)$) параметров $(c_0, \phi_{01}, \phi_{02}, \ldots, \phi_{0p})$ численно совпадает с OLS-оценкой коэффициентов регрессии y_t на константу и p лаговых значений y_t. Условная ML-оценка σ_0^2 равна сумме квадратов остатков, деленной на размер выборки. Из утверждения 6.7 следует, что условная ML-оценка коэффициентов является состоятельной и асимптотически нормальной, если выполняются условия стационарности (то есть когда корни характеристического полинома $1 - \phi_{01}z - \cdots - \phi_{0p}z^p = 0$ превосходят 1 по абсолютному значению). Как и в случае процесса AR(1), для выполнения свойств состоятельности и асимптотической нормальности не требуется предположения о нормальном распределении ошибок. Точная ML-оценка и условная ML-оценка имеют одинаковое асимптотическое распределение, если условие стационарности выполнено.

Эти результаты можно обобщить на случай векторного процесса. **Гауссовская VAR(p)** (*Gaussian VAR(p)*) (авторегрессия порядка p) может быть записана как

$$\underset{(M\times 1)}{\boldsymbol{y}_t} = \boldsymbol{c}_0 + \boldsymbol{\Phi}_{01}\boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_{0p}\boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t, \tag{8.7.21}$$

где $\boldsymbol{\varepsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{\Omega}_0)$. Целевая функция в условном ML-оценивании равна умноженному на $1/n$ условному логарифмическому правдоподобию:

$$Q_n(\boldsymbol{\theta}) = -\frac{M}{2}\log(2\pi) + \frac{1}{2}\log(|\boldsymbol{\Omega}^{-1}|) - \\ - \frac{1}{2n}\sum_{t=1}^{n}\{(\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t)'\boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t)\}, \tag{8.7.22}$$

где

$$\boldsymbol{\Pi}' \equiv \begin{bmatrix} \boldsymbol{c} & \boldsymbol{\Phi}_1 & \boldsymbol{\Phi}_2 & \cdots & \boldsymbol{\Phi}_p \end{bmatrix}, \quad \boldsymbol{x}_t \equiv \begin{bmatrix} 1 \\ \boldsymbol{y}_{t-1} \\ \boldsymbol{y}_{t-2} \\ \vdots \\ \boldsymbol{y}_{t-p} \end{bmatrix}. \tag{8.7.23}$$

Это совпадает с целевой функцией для ML-оценивания многомерной регрессионной модели, рассмотренной в параграфе 8.4. Поэтому условная ML-оценка параметров $(\boldsymbol{c}_0, \boldsymbol{\Pi}_0)$ численно совпадает с оценкой, получаемой применением OLS отдельно к каждому уравнению. Она является состоятельной и асимптотически нормальной, если матрицы коэффициентов удовлетворяют условию стационарности: все корни характеристического уравнения $|\boldsymbol{I}_M - \boldsymbol{\Phi}_{01}z - \cdots - \boldsymbol{\Phi}_{0p}z^p| = 0$ превосходят 1 по абсолютному значению. Этот результат не требует нормальности распределения вектора $\boldsymbol{\varepsilon}_t$.

Контрольные вопросы

1. (Вклад является мартингал-разностью.) Рассмотрим гауссовский AR(1)-процесс (8.7.9). Пусть $\boldsymbol{s}(\boldsymbol{w}_t; \boldsymbol{\theta})$ — функция вклада наблюдения t (градиент (8.7.17)), где $\boldsymbol{w} = (y_t, y_{t-1})'$ и $\boldsymbol{\theta} = (c, \phi, \sigma^2)'$.

   (a) Покажите, что

   $$\boldsymbol{s}(\boldsymbol{w}_t; \boldsymbol{\theta}) = \begin{bmatrix} \frac{1}{\sigma^2}\boldsymbol{x}_t \cdot (y_t - \boldsymbol{x}_t'\boldsymbol{\beta}) \\ -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}(y_t - \boldsymbol{x}_t'\boldsymbol{\beta})^2 \end{bmatrix},$$

   где $\boldsymbol{\beta} = (c, \varphi)'$ и $\boldsymbol{x}_t = (1, y_{t-1})'$.

   (b) Покажите, что

   $$\mathrm{E}[\boldsymbol{s}(\boldsymbol{w}_t; \boldsymbol{\theta}) | y_{t-1}, y_{t-2}, \ldots] = 0.$$

   Указание: *При* $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, $y_t - \boldsymbol{x}_t'\boldsymbol{\beta} = \varepsilon_t$.

   (c) Покажите, что $\{\boldsymbol{s}(\boldsymbol{w}_t; \boldsymbol{\theta}_0)\}$ является мартингал-разностью. **Указание:** $\boldsymbol{s}(\boldsymbol{w}_t; \boldsymbol{\theta}_0)$ *является (измеримой) функцией от* (y_t, y_{t-1}). *Следовательно,* $\boldsymbol{s}(\boldsymbol{w}_{t-j}; \boldsymbol{\theta}_0)$ $(j \geq 1)$ *можно вычислить на основе* $(y_{t-1}, y_{t-2}, \ldots)$.

   (d) Можно ли перенести результат пункта (c) на случай гауссовского AR(p)-процесса? [Ответ: Да.]

Набор задач для главы 8

Аналитические упражнения

1. Покажите, что

$$\left|\frac{1}{n}\sum_{t=1}^{n}(\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t)(\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t)'\right| \geq \left|\frac{1}{n}\sum_{t=1}^{n}\widehat{\boldsymbol{v}}_t\widehat{\boldsymbol{v}}_t'\right|, \tag{1}$$

где $\widehat{\boldsymbol{v}}_t \equiv \boldsymbol{y}_t - \widehat{\boldsymbol{\Pi}}'\boldsymbol{x}_t$ и $\widehat{\boldsymbol{\Pi}}$ определена в (8.4.5). (Так как $\widehat{\boldsymbol{v}}_t$ не зависит от $\boldsymbol{\Pi}$, по построению, это неравенство показывает, что левая часть достигает минимума при $\boldsymbol{\Pi} = \widehat{\boldsymbol{\Pi}}$.) **Указание:** *Используйте стратегию «добавить и вычесть», чтобы вывести соотношение:*

$$\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t = \widehat{\boldsymbol{v}}_t + \left(\widehat{\boldsymbol{\Pi}} - \boldsymbol{\Pi}\right)'\boldsymbol{x}_t.$$

Тогда

$$\sum_{t=1}^{n}(\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t)(\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t)' =$$

$$= \sum_{t=1}^{n}\widehat{\boldsymbol{v}}_t\widehat{\boldsymbol{v}}_t' + \left(\widehat{\boldsymbol{\Pi}} - \boldsymbol{\Pi}\right)'\sum_{t=1}^{n}\boldsymbol{x}_t\widehat{\boldsymbol{v}}_t' + \sum_{t=1}^{n}\widehat{\boldsymbol{v}}\boldsymbol{x}_t'\left(\widehat{\boldsymbol{\Pi}} - \boldsymbol{\Pi}\right) +$$

$$+ \left(\widehat{\boldsymbol{\Pi}} - \boldsymbol{\Pi}\right)'\left(\sum_{t=1}^{n}\boldsymbol{x}_t\boldsymbol{x}_t'\right)\left(\widehat{\boldsymbol{\Pi}} - \boldsymbol{\Pi}\right) =$$

$$= \sum_{t=1}^{n}\widehat{\boldsymbol{v}}_t\widehat{\boldsymbol{v}}_t' + \left(\widehat{\boldsymbol{\Pi}} - \boldsymbol{\Pi}\right)'\left(\sum_{t=1}^{n}\boldsymbol{x}_t\boldsymbol{x}_t'\right)\left(\widehat{\boldsymbol{\Pi}} - \boldsymbol{\Pi}\right)$$

(так как $\sum_{t=1}^{n}\boldsymbol{x}_t\widehat{\boldsymbol{v}}_t' = 0$*).*

Далее воспользуйтесь следующим результатом из матричной алгебры:

Пусть $\boldsymbol{A}$ и $\boldsymbol{B}$ — положительно полуопределенные матрицы одинакового размера. Тогда $|\boldsymbol{A} + \boldsymbol{B}| \geq |\boldsymbol{A}|$.

2. ($\widehat{\boldsymbol{\Omega}}(\boldsymbol{\Pi})$ положительно определена.) В условиях многомерной регрессионной модели покажите, что $\widehat{\boldsymbol{\Omega}}(\boldsymbol{\Pi})$, определенная в (8.4.10), является положительно определенной для любой заданной $\boldsymbol{\Pi}$. **Указание:** *Выведите соотношение $\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t = \boldsymbol{v}_t + (\boldsymbol{\Pi}_0 - \boldsymbol{\Pi})'\boldsymbol{x}_t$. Так как $\{\boldsymbol{y}_t, \boldsymbol{x}_t\}$ — последовательность независимых одинаково распределенных величин, $\widehat{\boldsymbol{\Omega}}(\boldsymbol{\Pi})$ сходится почти наверное к $\mathrm{E}\left[(\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t)(\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t)'\right]$. Покажите, что*

$$\mathrm{E}\left[(\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t)(\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t)'\right] = \boldsymbol{\Omega}_0 + (\boldsymbol{\Pi}_0 - \boldsymbol{\Pi})'\,\mathrm{E}(\boldsymbol{x}_t\boldsymbol{x}_t')(\boldsymbol{\Pi}_0 - \boldsymbol{\Pi}).$$

Затем используйте матричное неравенство, упомянутое в предыдущем упражнении. $\boldsymbol{\Omega}_0$ положительно определена по предположению.

3. (Дополнительное, идентификация $\boldsymbol{\delta}$ в FIML.) Как было замечено в сноске на с. 566 параграфа 8.5, FIML-оценка $\boldsymbol{\delta}_0$ является экстремальной оценкой с целевой функцией

$$Q_n(\boldsymbol{\delta}) = -|\widehat{\boldsymbol{\Omega}}(\boldsymbol{\delta})|, \tag{2}$$

где

$$\widehat{\boldsymbol{\Omega}}(\boldsymbol{\delta}) \equiv \frac{1}{n}\sum_{t=1}^{n}(\boldsymbol{y}_t + \boldsymbol{\Gamma}^{-1}\boldsymbol{B}\boldsymbol{x}_t)(\boldsymbol{y}_t + \boldsymbol{\Gamma}^{-1}\boldsymbol{B}\boldsymbol{x}_t)'. \tag{3}$$

Пусть

$$Q_0(\boldsymbol{\delta}) \equiv \operatorname*{plim}_{n\to\infty} Q_n(\boldsymbol{\delta}). \tag{4}$$

Чтобы показать состоятельность этой экстремальной оценки, необходимо проверить два условия утверждения 7.1 из предыдущей главы. Воспроизведем эти условия, используя текущие обозначения:

идентифицируемость: $Q_0(\boldsymbol{\delta})$ достигает единственного максимума на компактном пространстве параметров при $\boldsymbol{\delta} = \boldsymbol{\delta}_0$.

равномерная сходимость: $Q_n(\cdot)$ равномерно сходится по вероятности к $Q_0(\cdot)$.

В этом упражнении мы доказываем, что ранговое условие идентифицируемости ($\mathrm{E}(\boldsymbol{x}_t\boldsymbol{z}'_{tm})$ имеет полный столбцовый ранг для всех m) эквивалентно условию идентифицируемости указанной выше экстремальной оценки. В процессе доказательства все предположения относительно FIML-оценок, которые мы делали в тексте, считайте выполненными.

Чтобы начать обсуждение, введем некоторые новые обозначения. Воспроизведем m-е уравнение структурной формы из текста:

$$y_{tm} = \underset{(1\times L_m)}{\boldsymbol{z}'_{tm}}\ \underset{(L_m\times 1)}{\boldsymbol{\delta}_{0m}} + \varepsilon_{tm}. \tag{5}$$

Регрессоры $\boldsymbol{z}_{tm}$ состоят из M_m эндогенных переменных и K_m предопределенных, где $M_m + K_m = L_m$. Обозначим

$\underset{(M_m\times M)}{\boldsymbol{S}_m}$: матрица, выбирающая M_m эндогенных переменных из $\boldsymbol{y}_t$,

$\underset{(K_m\times K)}{\boldsymbol{C}_m}$: матрица, выбирающая K_m предопределенных переменных из $\boldsymbol{x}_t$.

Для иллюстрации: $\boldsymbol{S}_1$ и $\boldsymbol{C}_1$ для примера 8.1 с $\boldsymbol{y}_t = (p_t, q_t)'$ и $\boldsymbol{x}_t = (1, a_t, w_t)'$ определяются как

$$\boldsymbol{S}_1 = \begin{bmatrix} 1 & 0 \end{bmatrix}, \boldsymbol{C}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}.$$

Таким образом, регрессоры m-го уравнения могут быть представлены как

$$\boldsymbol{z}'_{tm} = \left(\boldsymbol{y}'_t \boldsymbol{S}'_m \;\vdots\; \boldsymbol{x}'_t \boldsymbol{C}'_m\right).$$

(a) Покажите, что

$$\underset{(K\times L_m)}{\mathrm{E}(\boldsymbol{x}_t \boldsymbol{z}'_{tm})} = \underset{(K\times K)}{\mathrm{E}(\boldsymbol{x}_t \boldsymbol{x}'_t)} \underbrace{\left[\underset{(K\times M)}{\boldsymbol{\Pi}_0} \underset{(M\times M_m)}{\boldsymbol{S}'_m} \;\vdots\; \underset{(K\times K_m)}{\boldsymbol{C}'_m} \right]}_{(K\times L_m)}. \tag{6}$$

Указание: *Используйте приведенную форму* (8.5.9) *и тот факт, что* $\mathrm{E}(\boldsymbol{x}_t \boldsymbol{v}'_t) = 0$. Так как умножение на невырожденную матрицу (здесь $\mathrm{E}(\boldsymbol{x}_t \boldsymbol{x}'_t)$) не меняет ранга, из уравнения (6) следует, что ранговое условие идентификации эквивалентно тому, что ранг $\left[\boldsymbol{\Pi}_0 \boldsymbol{S}_m \;\vdots\; \boldsymbol{C}'_m\right]$ равен L_m.

(b) Покажите, что

$$\underset{n\to\infty}{\mathrm{plim}}\ \widehat{\boldsymbol{\Omega}}(\boldsymbol{\delta}) = \boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Sigma}_0(\boldsymbol{\Gamma}_0^{-1})' + \left[\boldsymbol{\Pi}'_0 + \boldsymbol{\Gamma}^{-1}\boldsymbol{B}\right] \mathrm{E}(\boldsymbol{x}_t \boldsymbol{x}'_t)\left[\boldsymbol{\Pi}'_0 + \boldsymbol{\Gamma}^{-1}\boldsymbol{B}\right]'. \tag{7}$$

где $\boldsymbol{\Pi}'_0 \equiv -\boldsymbol{\Gamma}_0^{-1}\boldsymbol{B}_0$. **Указание:** *Так как* $\{\boldsymbol{y}_t, \boldsymbol{x}_t\}$ *— i.i.d., указанный предел по вероятности равен* $\mathrm{E}\left[(\boldsymbol{y}_t + \boldsymbol{\Gamma}^{-1}\boldsymbol{B}\boldsymbol{x}_t)(\boldsymbol{y}_t + \boldsymbol{\Gamma}^{-1}\boldsymbol{B}\boldsymbol{x}_t)'\right]$. *Кроме того,*

$$\boldsymbol{y}_t + \boldsymbol{\Gamma}^{-1}\boldsymbol{B}\boldsymbol{x}_t = \boldsymbol{v}_t + (\boldsymbol{\Pi}'_0 + \boldsymbol{\Gamma}^{-1}\boldsymbol{B})\boldsymbol{x}_t.$$

(c) Покажите, что $\mathrm{plim}\,|\widehat{\boldsymbol{\Omega}}(\boldsymbol{\delta})|$ достигает минимума, только если $\boldsymbol{\Gamma}\boldsymbol{\Pi}'_0 + \boldsymbol{B} = 0$. **Указание:** *Так как* $\mathrm{E}(\boldsymbol{x}_t \boldsymbol{x}'_t)$ *положительно определена, второе слагаемое в правой части уравнения (7) равно нулю, только если* $\boldsymbol{\Pi}_0 + \boldsymbol{B}'(\boldsymbol{\Gamma}^{-1})' = 0$. *Также используйте результат из линейной алгебры, упомянутый в упражнении 1 выше.*

Поэтому условие идентифицируемости для экстремальной оценки (что $Q_0(\boldsymbol{\delta})$ достигает единственного максимума на пространстве параметров при $\boldsymbol{\delta}_0$) выполнено, даже если $(\boldsymbol{\Gamma}, \boldsymbol{B}) = (\boldsymbol{\Gamma}_0, \boldsymbol{B}_0)$ — единственное решение $\boldsymbol{\Gamma}\boldsymbol{\Pi}'_0 + \boldsymbol{B} = 0$, рассматриваемой как система одновременных уравнений для $(\boldsymbol{\Gamma}, \boldsymbol{B})$.

(d) Пусть α'_m размера $(1 \times (M+K))$ — m-я строка матрицы $[\boldsymbol{\Gamma} \vdots \boldsymbol{B}]$ и $\boldsymbol{e}'_m$ — вектор размера $(1 \times (M+K))$, элементы которого, соответствующие y_{tm}, равны единице, а все другие равны нулю. Для иллюстрации: $\boldsymbol{e}'_1$ для примера 8.1, где $\boldsymbol{y}_t = (p_t, q_t)'$ и $\boldsymbol{x}_t = (1, a_t, w_t)'$ равен $(0, 1, 0, 0, 0)$. Проверьте для примера 8.1 общий результат о том, что

$$\underset{(1\times(M+K))}{\boldsymbol{\alpha}'_m} = \boldsymbol{e}'_m - \underset{(1\times(M_m+K_m))}{\boldsymbol{\delta}'_m} \begin{bmatrix} \underset{(M_m\times M)}{\boldsymbol{S}_m} & \underset{(M_m\times K)}{\boldsymbol{0}} \\ \underset{(K_m\times M)}{\boldsymbol{0}} & \underset{(K_m\times K)}{\boldsymbol{C}_m} \end{bmatrix}. \tag{8}$$

(e) Перепишите $\boldsymbol{\Gamma}\boldsymbol{\Pi}'_0 + \boldsymbol{B} = 0$ как

$$[\boldsymbol{\Gamma} \vdots \boldsymbol{B}] \begin{bmatrix} \boldsymbol{\Pi}'_0 \\ \boldsymbol{I}_K \end{bmatrix} = \boldsymbol{0}. \tag{9}$$

Покажите, что m-я строка $\boldsymbol{\Gamma}\boldsymbol{\Pi}'_0 + \boldsymbol{B} = 0$ может быть записана в виде:

$$\boldsymbol{\delta}'_m \begin{bmatrix} \boldsymbol{S}_m\boldsymbol{\Pi}'_0 \\ \boldsymbol{C}_m \end{bmatrix} = \boldsymbol{\pi}'_{0m}, \tag{10}$$

где

$$\boldsymbol{\pi}_{0m} \equiv \left[\boldsymbol{\Pi}_0 \vdots \boldsymbol{I}_K\right]\boldsymbol{e}_m. \tag{11}$$

(f) Проверьте, что $\boldsymbol{\delta}_m = \boldsymbol{\delta}_{0m}$ является *одним из* решений (10). Покажите, что необходимым и достаточным условием того, чтобы $\boldsymbol{\delta}_m = \boldsymbol{\delta}_{0m}$ было единственным решением (10), является ранговое условие идентификации уравнения m. **Указание:** *Вспомните следующий факт из линейной алгебры: Пусть $\boldsymbol{Ax} = \boldsymbol{y}$ имеет решение x_0. Необходимым и достаточным условием того, чтобы это решение было единственным, является полный столбцовый ранг матрицы $\boldsymbol{A}$.*

(g) Объясните, почему мы можем сделать вывод о том, что единственным решением $\boldsymbol{\Gamma}\boldsymbol{\Pi}'_0 + \boldsymbol{B} = 0$ является $(\boldsymbol{\Gamma}, \boldsymbol{B}) = (\boldsymbol{\Gamma}_0, \boldsymbol{B}_0)$. **Указание:** *Каждый элемент δ может встречаться в $\boldsymbol{\Gamma}$ или $\boldsymbol{B}$ только один раз.*

4. (Дополнительное, инструменты, которые не появляются в системе.) Рассмотрим полную систему одновременных уравнений, обсуждавшуюся в параграфе 8.5. Пусть

$$\underset{(K\times 1)}{\boldsymbol{x}_t} = \begin{bmatrix} \underset{((K-1)\times 1)}{\widetilde{\boldsymbol{x}}_t} \\ x_{tK} \end{bmatrix}.$$

Предположим, что x_{tK} не входит ни в одно из M структурных уравнений системы. Покажите, что исключение x_{tK} из набора инструментов не влияет ни на выполнимость рангового условия идентификации, ни на асимптотическую дисперсию FIML-оценки параметров δ_0. **Указание:** *Определите матрицы S_m и C_m как в упражнении 3, так что $z'_{tm} = (y_t S'_m \vdots x'_t C'_m)$ и выполняется (6). FIML-оценка асимптотически эквивалентна 3SLS, следовательно, ее асимптотическая дисперсия задается в* (4.5.15). *Последняя строка C'_m представляет собой вектор из нулей. Пусть $\widetilde{\mathbf{\Pi}}_0$ (размера $((K-1)\times M)$) — матрица коэффициентов приведенной формы, когда x_{tK} исключены из системы. Тогда*

$$\mathbf{\Pi}_0 = \begin{bmatrix} \underset{((K-1)\times M)}{\widetilde{\mathbf{\Pi}}_0} \\ \underset{(1\times M)}{\mathbf{0}'} \end{bmatrix}.$$

Литература

Amemiya, T., 1973, "Regression Analysis When the Dependent Variable is Truncated Normal", *Econometrica*, 41, 997–1016.

———, 1985, *Advanced Econometrics*, Cambridge: Harvard University Press.

Anderson, T. W., and H. Rubin, 1949, "Estimator of the Parameters of a Single Equation in a Complete System of Stochastic Equations", *Annals of Mathematical Statistics*, 20, 46–63.

Anderson, T. W., N. Kunitomo, and T. Sawa, 1982, "Evaluation of the Distribution Function of the Limited Information Maximum Likelihood Estimator", *Econometrica*, 50, 1009–1027.

Davidson, R., and J. MacKinnon, 1993, *Estimation and Inference in Econometrics*, New York: Oxford University Press.

Fuller,W., 1996, *Introduction to Statistical Time Series* (2d ed.), New York: Wiley.

Hausman, J., 1975, "An Instrumental Variable Approach to Full Information Estimators for Linear and Certain Nonlinear Econometric Models", *Econometrica*, 43, 727–738.

Johansen, S., 1995, *Likelihood-Based Inference in Cointegrated Vector Auto-Regressive Models*, New York: Oxford University Press.

Judge, G., W. Griffiths, R. Hill, H. Lütkepohl, and T. Lee, 1985, *The Theory and Practice of Econometrics* (2d ed.), New York: Wiley.

Magnus, J., and H. Neudecker, 1988, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, New York: Wiley.

Newey, W., and D. McFadden, 1994, "Large Sample Estimation and Hypothesis Testing", Chapter 36 in R. Engle and D. McFadden (eds.), *Handbook of Econometrics*, Volume IV, New York: North-Holland.

Olsen, R. J., 1978, "Note on the Uniqueness of the Maximum Likelihood Estimator for the Tobit Model", *Econometrica*, 46, 1211–1215.

Orme, C., 1989, "On the Uniqueness of theMaximum Likelihood Estimator in Truncated Regression Models", *Econometric Reviews*, 8, 217–222.

Pagan, A., 1979, "Some Consequences of Viewing LIML As an Iterated Aitken Estimator", *Economics Letters*, 3, 369–372.

Sapra, S. K., 1992, "Asymptotic Properties of a Quasi-Maximum Likelihood Estimator in Truncated Regression Model with Serial Correlation", *Econometric Reviews*, 11, 253–260.

Tobin, J., 1958, "Estimation of Relationships for Limited Dependent Variables", *Econometrica*, 26, 24–36.

Wooldridge, J., 1994, "Estimation and Inference for Dependent Processes", Chapter 45 in R. Engle and D. McFadden (eds.), *Handbook of Econometrics*, Volume IV, New York: North-Holland.

Глава 9. Эконометрика единичного корня

До настоящего момента наш анализ был ограничен стационарными процессами. В параграфе 9.1 вводится два класса процессов с трендом: процессы, стационарные относительно тренда, и процессы с единичным корнем. Процесс с единичным корнем называется также стационарным в разностях или интегрированным порядка 1 (I(1)), поскольку его первая разность является стационарным, или I(0)-процессом. Технические средства для получения предельного распределения статистик, включающих I(0)- и I(1)-процессы, собраны в параграфе 9.2. Используя эти средства, мы получим некоторые тесты на единичный корень при нулевой гипотезе о том, что рассматриваемый процесс является I(1)-процессом. Эти тесты не только являются популярными, но и обладают лучшими свойствами на конечных выборках по сравнению со многими другими существующими тестами. В параграфе 9.3 мы рассматриваем тесты Дики — Фуллера, которые были разработаны для тестирования нулевой гипотезы о том, что процесс является случайным блужданием, прототипичным I(1)-процессом, первая разность которого сериально некоррелирована. Параграф 9.4 показывает, что эти тесты можно обобщить таким образом, чтобы охватить I(1)-процессы с сериально коррелированными первыми разностями. Эти и другие тесты на единичный корень кратко сравниваются в параграфе 9.5. В прикладном параграфе 9.6 этой главы тесты на единичный корень используются для проверки паритета покупательной способности (PPP), фундаментального утверждения в международной экономике о том, что обменные курсы подгоняются к национальным уровням цен.

9.1. Моделирование трендов

Нет недостатка в примерах таких временных рядов, которые можно было бы разумно описать как тренды в экономике. На рис. 9.1 показано поведение логарифма реального ВВП США. Логарифм ВВП очевидно имеет возрастающий тренд, в том смысле, что его среднее, вместо того чтобы быть константой, как у стационарных процессов, неуклонно возрастает с течением времени. Тренд среднего называется **детерминированным трендом** (*deterministic trend*) или **временным трендом** (*time trend*). В случае логарифма ВВП США детерминированный тренд выглядит линейным. Другой вид тренда можно увидеть на рис. 6.3 на с. 454, где показано поведение логарифма обменного курса иена/доллар. Логарифм этого обменного курса не имеет тренда среднего, но кажется, что каждое его изменение оказывает перманентное воздействие на будущие значения ряда, так что наилучшим предиктором его будущих значений является его текущее значение. Процесс с этим свойством, которые не присуще стационарным процессам, называется **стохастическим трендом** (*stochastic trend*). Если вы вспомните определение мартингала, то оно соответствует этому описанию стохастического тренда. Стохастические тренды и мартингалы являются синонимами.

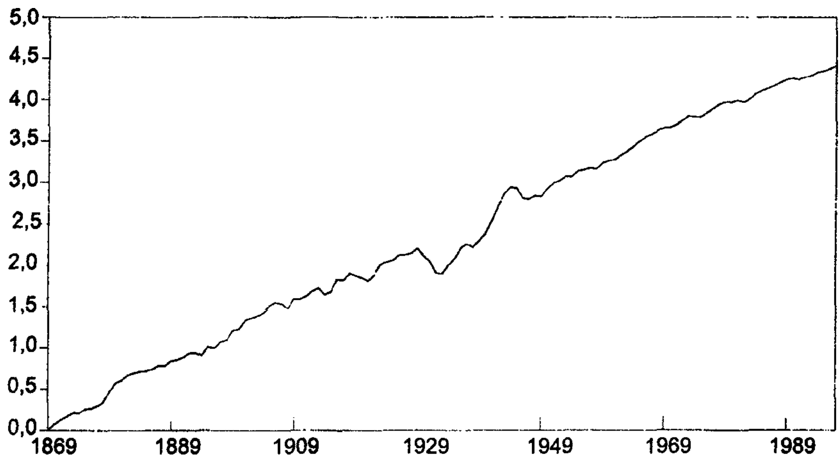

Рис. 9.1. Логарифм реального ВВП США, значение для 1869 года приравнено 0

Основная предпосылка этой главы и следующей заключается в том, что экономический временной ряд может быть представлен в виде суммы линейного временного тренда, стохастического тренда и стационарного процесса[1].

[1]Замечание о семантике. Процессы с трендами часто называются «нестационарными процессами». Мы не будем использовать этот термин в оставшейся части этой книги, поскольку процесс может быть нестационарным и без наличия тренда. Например, пусть ε_t является i.i.d.-процессом с единичной дисперсией и пусть d_t принимает значение 1 для нечетных t и 2 для четных t. Тогда процесс $\{u_t\}$, определенный как $u_t = d_t \cdot \varepsilon_t$, не является стационарным, поскольку его дисперсия зависит от t. Тем не менее этот процесс не может быть разумно описан как процесс с трендом.

Интегрированные процессы

Случайное блуждание является примером класса трендовых процессов, известных как **интегрированные процессы** (*integrated processes*). Чтобы дать точное определение интегрированных процессов, мы рассмотрим сначала I(0)-процессы. **I(0)-процесс** (*I(0) process*) является (строго) стационарным процессом, **долговременная дисперсия** (*long-run variance*) которого (определенная в обсуждении, следующем за утверждением 6.8 параграфа 6.5) конечна и положительна. (Через мгновение мы объясним, почему долговременная дисперсия должна быть положительной.) Следуя, например, [Hamilton 1994, p. 435], мы допускаем, что I(0)-процессы могут иметь ненулевые средние (некоторые авторы, например [Stock, 1994], требуют, чтобы I(0)-процессы имели нулевое среднее). Следовательно, I(0)-процесс может быть записан как

$$\delta + u_t, \tag{9.1.1}$$

где $\{u_t\}$ является стационарным процессом с нулевым средним и положительной долговременной дисперсией.

Определение интегрированных процессов вытекает из определения I(0)-процессов. Пусть Δ является оператором разности, так что для последовательности $\{\xi_t\}$

$$\begin{aligned} \Delta\xi_t \equiv (1-L)\xi_t &= \xi_t - \xi_{t-1}, \\ \Delta^2\xi_t = (1-L)^2\xi_t &= (\xi_t - \xi_{t-1}) - (\xi_{t-1} - \xi_{t-2}). \quad \text{и т. д.} \end{aligned} \tag{9.1.2}$$

Определение 9.1 (I(d)-процессы): Процесс называется интегрированным порядка d («**I**(d)») ($d = 1, 2, \ldots$), если d-я разность, $\Delta^d\xi_t$, является I(0)-процессом. В частности, процесс $\{\xi_t\}$ является интегрированным порядка 1 («**I**(1)»), если его первая разность, $\Delta\xi_t$, является I(0)-процессом.

Причина требования того, чтобы долговременная дисперсия I(0)-процесса была положительной, заключается в том, чтобы исключить следующую аномалию определения. Рассмотрим процесс v_t, определенный как

$$v_t \equiv \varepsilon_t - \varepsilon_{t-1}, \tag{9.1.3}$$

где $\{\varepsilon_t\}$ является независимым белым шумом. Как мы выяснили в контрольном вопросе 5 в параграфе 6.5, долговременная дисперсия $\{v_t\}$ равна нулю. Если бы указанного требования к долговременной дисперсии не было, то $\{v_t\}$ был бы I(0)-процессом. Но тогда, так как $\Delta\varepsilon_t = v_t$, процесс независимого белого шума $\{\varepsilon_t\}$ оказался бы I(1)-процессом! Если процесс $\{v_t\}$ записывается в виде первой разности I(0)-процесса, то он называется I(−1)-процессом. Долгосрочная дисперсия I(−1)-процесса равна нулю (доказательство этого составляет контрольный вопрос 1).

В оставшейся части этой главы интегрированные процессы, с которыми мы имеем дело, имеют порядок 1. Ниже приведено несколько замечаний относительно I(1)-процессов.

- (Когда он стартовал?) Как это ясно для случайных блужданий, дисперсия I(1)-процесса линейно возрастает по времени. Таким образом, если бы процесс начался в бесконечном прошлом, то дисперсия была бы бесконечной. Чтобы сосредоточиться на I(1)-процессах с конечной дисперсией, мы предполагаем, что процесс начался в конечном прошлом, и, без потери общности, мы можем предположить, что начальной датой является $t = 0$. Поскольку I(0)-процесс можно записать как (9.1.1) и поскольку, по определению I(1)-процесса, $\{\xi_t\}$ удовлетворяет соотношению $\Delta\xi_t = \delta + u_t$, мы можем записать ξ_t в уровнях как

$$\xi_t = \xi_0 + \delta \cdot t + (u_1 + u_2 + \cdots + u_t), \qquad (9.1.4)$$

  где $\{u_t\}$ — I(0) с нулевым средним. Поэтому спецификация процесса $\{\xi_t\}$ в уровнях должна включать предположение о начальном значении. Если не указано обратное, мы всюду предполагаем, что $E(\xi_0^2) < \infty$, так что ξ_0 может быть случайным.

- (Среднее процесса I(0) является трендом в I(1).) Как ясно из (9.1.4), I(1)-процесс может иметь линейный тренд, $\delta \cdot t$. Это является следствием допущения того, что I(0)-процессы имеют ненулевое среднее δ. Если $\delta = 0$, то I(1)-процесс не имеет тренда и называется **I(1)-процессом без сноса** (*driftless I(1) process*), в то время как если $\delta \neq 0$, процесс называется **I(1)-процессом со сносом** (*I(1) process with drift*). Очевидно, что I(1)-процесс со сносом может быть записан как сумма линейного тренда и I(1)-процесса без сноса.

- (Два других названия I(1)-процессов.) У I(1)-процесса есть два других названия. Он называется **процессом, стационарным в разностях** (*difference-stationary process*), поскольку его первая разность стационарна. Он также называется **процессом с единичным корнем** (*unit-root process*). Чтобы понять, почему он так называется, рассмотрим модель:

$$(1 - \rho L)y_t = \delta + u_t, \qquad (9.1.5)$$

  где $\{u_t\}$ — I(0) с нулевым средним. Это авторегрессионная модель с возможно сериально коррелированными ошибками, представленными посредством u_t. Если авторегрессионный корень ρ равен единице, то первая разность y_t является I(0)-процессом, так что $\{y_t\}$ является I(1)-процессом.

Почему важно знать, является ли рассматриваемый процесс I(1)-процессом

В оставшейся части этой главы мы будем иметь дело с различением **процессов, стационарных относительно тренда** (*trend-stationary processes*), которые можно записать как сумму линейного временного тренда (если таковой имеется) и стационарного процесса с одной стороны и процессов, стационарных в разностях (то есть I(1)-процессов с или без сноса), с другой. Как будет показано в утверждении 9.1, I(1)-процесс можно записать как сумму линейного тренда (если таковой имеется), стационарного процесса и стохастического тренда. Таким образом, разница между указанными двумя классами процессов состоит в наличии стохастического тренда. Имеются по крайней мере две причины, почему такое различение является важным.

1. Первая: это имеет большое значение для прогнозирования. Чтобы показать это, рассмотрим следующий простой AR(1)-процесс с трендом:

$$
\begin{aligned}
y_t &= \alpha + \delta \cdot t + z_t, \\
z_t &= \rho z_{t-1} + \varepsilon_t,
\end{aligned} \tag{9.1.6}
$$

где $\{\varepsilon_t\}$ является независимым белым шумом. Прогноз на s шагов вперед y_{t+s}, условный относительно $(y_t, y_{t-1}, \ldots)$, можно записать как

$$
\begin{aligned}
&\mathrm{E}(y_{t+s}|y_t, y_{t-1}, \ldots) = \\
&\quad = \mathrm{E}[\alpha + \delta \cdot (t+s)|y_t, y_{t-1}, \ldots] + \mathrm{E}(z_{t+s}|y_t, y_{t-1}, \ldots) = \\
&\quad = \alpha + \delta \cdot (t+s) + \mathrm{E}(z_{t+s}|y_t, y_{t-1}, \ldots).
\end{aligned} \tag{9.1.7}
$$

И $(y_t, y_{t-1}, \ldots)$, и $(z_t, z_{t-1}, \ldots)$ содержат одну и ту же информацию, поскольку между ними существует взаимно однозначное соответствие. Поэтому последнее слагаемое можно записать как

$$
\mathrm{E}(z_{t+s}|y_t, y_{t-1}, \ldots) = \mathrm{E}(z_{t+s}|z_t, z_{t-1}, \ldots) = \rho^s z_t, \tag{9.1.8}
$$

поскольку $z_{t+s} = \varepsilon_{t+s} + \rho\varepsilon_{t+s-1} + \cdots + \rho^{s-1}\varepsilon_{t+1} + \rho^s z_t$. Подставляя (9.1.8) в (9.1.7), мы получаем следующее выражение для прогноза на s шагов вперед:

$$
\begin{aligned}
\mathrm{E}(y_{t+s}|y_t, y_{t-1}, \ldots) &= \alpha + \delta \cdot (t+s) + \rho^s z_t = \\
&= \alpha + \delta \cdot (t+s) + \rho^s \cdot (y_t - \alpha - \delta \cdot t).
\end{aligned} \tag{9.1.9}
$$

Следует рассмотреть два случая:

Случай $|\rho| < 1$ Если $|\rho| < 1$, то $\{z_t\}$ является стационарным AR(1)-процессом и, таким образом, является I(0)-процессом с нулевым средним. Поэтому $\{y_t\}$ является процессом, стационарным относительно тренда. В этом случае, поскольку $\mathrm{E}(z_t^2) < \infty$, мы имеем:

$$\mathrm{E}[(\rho^s z_t)^2] = \rho^{2s}\,\mathrm{E}(z_t^2) \to 0 \text{ при } s \to \infty.$$

Следовательно, прогноз на s периодов вперед $\mathrm{E}(y_{t+s}|y_t, y_{t-1}, \dots)$ при $s \to \infty$ сходится в среднем квадратическом к линейному временному тренду $\alpha + \delta \cdot (t+s)$. Более точно:

$$\mathrm{E}\big\{[\mathrm{E}(y_{t+s}|y_t, y_{t-1}, \dots) - \alpha - \delta \cdot (t+s)]^2\big\} \to 0 \text{ as } s \to \infty. \quad (9.1.10)$$

То есть текущее и прошлые значения y не влияют на прогноз, если горизонт прогноза s достаточно далекий. В частности, если $\delta = 0$, то

$$\mathrm{E}(y_{t+s}|y_t, y_{t-1}, \dots) \underset{\text{m.s.}}{\to} \mathrm{E}(y_t) \text{ при } s \to \infty. \quad (9.1.11)$$

То есть долговременным прогнозом является безусловное среднее. Это свойство, называемое **возвращением к среднему** (*mean reversion*), выполняется для любого линейного стационарного процесса, а не только для стационарного AR(1)-процесса[1]. По этой причине линейный стационарный процесс иногда называют **преходящей компонентой** (*transitory component*).

Случай $\rho = 1$ Предположим теперь, что $\rho = 1$. Тогда $\{z_t\}$ является случайным блужданием без сноса (частный случай стохастического тренда), и y_t можно записать в виде:

$$y_t = (\alpha + z_0) + \delta \cdot t + \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t, \quad (9.1.12)$$

показывающем, что $\{y_t\}$ является случайным блужданием со сносом с начальным значением $z_0 + \alpha$ (частный случай I(1) или стационарного в разностях процесса). Полагая $\rho = 1$ в (9.1.9), мы получаем:

$$\mathrm{E}(y_{t+s}|y_t, y_{t-1}, \dots) = \delta \cdot s + y_t. \quad (9.1.13)$$

То есть случайное блуждание со сносом δ, как ожидается, растет с постоянной скоростью δ за период, каким бы ни было текущее значение y_t. В отличие от случая стационарности относительно тренда текущее значение оказывает здесь перманентное влияние

[1] Доказательство см., например, в: [Hamilton, 1994, p. 439].

на прогноз для всех горизонтов прогноза. Это происходит из-за наличия стохастического тренда z_t. Стохастический тренд называют также **перманентной компонентой** (*permanent component*).

2. Вторая причина для различения стационарных относительно тренда и стационарных в разностях процессов возникает в контексте статистических выводов. Предположим, например, что вы заинтересованы в проверке некоторой гипотезы относительно параметра β в уравнении регрессии:

$$y_t = x_t\beta + u_t, \tag{9.1.14}$$

где u_t является стационарной ошибкой. С одной стороны, если регрессор x_t стационарен относительно тренда, то тогда, как было показано в параграфе 2.12, t-значение для OLS-оценки параметра β будет иметь асимптотически стандартное нормальное распределение. С другой стороны, если x_t стационарен в разностях, то есть если он имеет стохастический тренд, то это t-значение имеет нестандартное предельное распределение, так что делать статистические выводы относительно β обычным способом нельзя. Регрессия в рассматриваемом случае называется «коинтегрирующей регрессией». Статистические выводы относительно коинтегрирующей регрессии будут изучаться в следующей главе.

Что следует брать в качестве нулевой гипотезы, I(0) или I(1)?

Для различения стационарных относительно тренда и стационарных в разностях процессов, или I(1)-процессов, в литературе обычной практикой является проверка нулевой гипотезы о том, что рассматриваемый процесс является I(1)-процессом, а не проверка нулевой гипотезы о стационарности относительно тренда. Предположительно у этого есть две причины. Одна из них состоит в том, что исследователь убежден, что в представляющем интерес уравнении переменные являются I(1), и хочет контролировать ошибку I рода относительно нулевой гипотезы I(1). Однако это в конечном счете не является убедительным, поскольку обычно исследователь не обладает строгим априорным убеждением относительно того, какими являются переменные, I(0) или I(1). Краткое изложение проблемы, которую это создает для статистических выводов, имеется в [Watson, 1994, pp. 2867–2870]. Другая причина состоит просто в том, что литература по тестам для нулевой гипотезы стационарности относительно тренда против альтернативы I(1) все еще недостаточно разработана. Есть несколько тестов для проверки нулевой гипотезы о том, что процесс является I(0) или, в более общем случае, стационарным относительно тренда; см. статьи [Tanaka, 1990], [Kwiatkowski, et al., 1992] и [Leybourne and McCabe, 1994], а также обзоры [Stock, 1994, Section 4] и [Maddala and Kim, 1998, Section 4.5]. На момент написания этой книги нет единственного теста для проверки I(0)-гипотезы, использующегося большинством исследователей, который имел бы хорошие свойства в конечных выборках. Поэтому в этой книге мы концентрируемся на тестах для проверки нулевой гипотезы I(1).

Другие подходы к моделированию трендов

В дополнение к стохастическим и временным трендам имеются два других популярных подхода к моделированию трендов: дробная интегрированность и тренды с изломом с неизвестной датой излома. Они не будут рассматриваться в этой книге. В качестве обзора первого подхода см.: [Baillie, 1996]. Количество публикаций о сдвигах в тренде быстро возрастает. Обзоры можно найти в [Stock, 1994, Section 5] и [Maddala and Kim, 1998, Chapter 13].

Контрольные вопросы

1. (Долгосрочная дисперсия разностей I(0)-процессов.) Пусть $\{u_t\}$ есть I(0) с $\mathrm{Var}(u_t) < \infty$. Проверьте, что процесс $\{\Delta u_t\}$ является ковариационно стационарным и что долговременная дисперсия $\{\Delta u_t\}$ равна нулю.
   Указание: *Долгосрочная дисперсия* $\{v_t\}$ *(где* $v_t \equiv \Delta u_t$*) определяется как предел для* $\mathrm{Var}(\sqrt{T}\bar{v})$ *при* $t \to \infty$.
   $$\sqrt{T}\bar{v} = (v_1 + v_2 + \cdots + v_T)/\sqrt{T} = (u_T - u_0)/\sqrt{T}.$$

9.2. Инструментарий для эконометрики единичных корней

Основным инструментом эконометрики единичных корней является то, что называют **функциональной центральной предельной теоремой** (*functional central limit theorem, FCLT*). Для того чтобы FCLT была применима, мы должны сузить класс I(1)-процессов, накладывая ограничения на соответствующие I(0)-процессы. Решение о том, какие ограничения следует наложить, то есть на каком классе I(0)-процессов сосредоточиться, изменяется от автора к автору в зависимости от используемой версии FCLT. В данной работе, следуя сложившейся в литературе практике, мы фокусируемся на линейных I(0)-процессах. Вслед за определением линейных I(0)-процессов в этом параграфе вводятся некоторые понятия и результаты, которые будут неоднократно использоваться в оставшейся части данной главы.

Линейные I(0)-процессы

Наше определение линейных I(0)-процессов заключается в следующем:

Определение 9.2 (линейные I(0)-процессы): Линейный I(0)-процесс можно записать как константу плюс линейный процесс с нулевым средним $\{u_t\}$, такой что

$$u_t = \psi(L)\varepsilon_t, \ \psi(L) \equiv \psi_0 + \psi_1 L + \psi_2 L^2 + \cdots \text{ при } t = 0, \pm 1, \pm 2, \dots \ . \tag{9.2.1}$$

$$\{\varepsilon_t\} \text{ является независимым белым шумом}$$

$$(\text{i.i.d. со средним 0 и } \mathrm{E}(\varepsilon_t^2) \equiv \sigma^2 > 0), \tag{9.2.2}$$

$$\sum_{j=0}^{\infty} j|\psi_j| < \infty, \tag{9.2.3a}$$

$$\psi(1) \neq 0 \tag{9.2.3b}$$

Предполагается, что процесс инноваций $\{\varepsilon_t\}$ является белым шумом i.i.d. Это требование введено только для упрощения изложения, и его можно существенно ослабить[1]. В оставшейся части этой главы мы называем линейные I(0)-процессы просто I(0)-процессами, без уточнения «линейные». Также, если не указано иное, $\{\varepsilon_t\}$ будет независимым процессом белого шума, а $\{u_t\}$ — I(0)-процессом с нулевым средним.

Условие (9.2.3a) (иногда называемое условием **1-суммируемости** (*one-summability*)) является более сильным, чем более известное условие абсолютной суммируемости. Это более сильное предположение сделает более легким доказательство результатов для больших выборок в следующем параграфе с помощью разложения Бевериджа — Нельсона, которое будет введено в ближайшее время. Поскольку последовательность $\{\psi_j\}$ абсолютно суммируема, линейный процесс $\{u_t\}$ является строго стационарным и эргодическим в силу утверждения 6.1(d). Автоковариация u_t j-го порядка будет обозначаться как γ_j. Поскольку $\mathrm{E}(u_t) = 0$, $\gamma_j = \mathrm{E}(u_t u_{t-j})$. По утверждению 6.1(c), автоковариации абсолютно суммируемы.

Чтобы понять условие (9.2.3b), вспомним, что, по утверждению 6.8(b), долговременная дисперсия (которую мы обозначаем в этой главе λ^2) равна значению производящей функции автоковариаций при $z = 1$. По утверждению 6.6 она дается формулой:

$$\lambda^2 \equiv \text{долговременная дисперсия } \{u_t\} = \gamma_0 + 2\sum_{j=1}^{\infty} \gamma_j = \sigma^2 \cdot [\psi(1)]^2. \tag{9.2.4}$$

[1] Например, процесс инноваций может быть стационарной мартингал-разностью. См., например: Theorem 3.8 в [Tanaka, 1996, p. 80].

Условие (9.2.3b) гарантирует положительность долговременной дисперсии.

Аппроксимация I(1)-процесса случайным блужданием

Пусть $\{\xi_t\}$ является I(1)-процессом, так что $\Delta\xi_t = \delta + u_t$, где $u_t \equiv \equiv \psi(L)\varepsilon_t$ является I(0)-процессом с нулевым средним, удовлетворяющим (9.2.1)–(9.2.3) с $\mathrm{E}(\xi_0^2) < \infty$. Используя следующее тождество (его надо проверить в контрольном вопросе 1):

$$\psi(L) = \psi(1) + \Delta\alpha(L), \quad \Delta \equiv 1 - L.$$

$$\alpha(L) \equiv \sum_{j=0}^{\infty}\alpha_j L^j, \quad \alpha_j = -(\psi_{j+1} + \psi_{j+2} + \cdots) \quad (j = 0, 1, 2, \ldots), \tag{9.2.5}$$

мы можем записать u_t как

$$u_t \equiv \psi(L)\varepsilon_t = \psi(1)\cdot\varepsilon_t + \eta_t - \eta_{t-1} \quad \text{с } \eta_t \equiv \alpha(L)\varepsilon_t. \tag{9.2.6}$$

Можно показать (см. аналитическое упражнение 7), что $\alpha(L)$ является абсолютно суммируемым. Поэтому, по утверждению 6.1(a) $\{\eta_t\}$ — ковариационно стационарный процесс с нулевым средним (фактически он является эргодически стационарным по утверждению 6.1(d)). Подставляя (9.2.6) в (9.1.4), мы получаем то, что известно в эконометрике как **разложение Бевериджа — Нельсона** (*Beveridge — Nelson decomposition*):

$$\begin{aligned}
\xi_t &= \delta\cdot t + \sum_{s=1}^{t}[\psi(1)\cdot\varepsilon_s + \eta_s - \eta_{s-1}] + \xi_0 = \\
&= \delta\cdot t + \psi(1)\sum_{s=1}^{t}\varepsilon_s + \eta_t + (\xi_0 - \eta_0) \\
&\text{(поскольку } \sum_{s=1}^{t}(\eta_s - \eta_{s-1}) = \eta_t - \eta_0).
\end{aligned} \tag{9.2.7}$$

Таким образом, любой линейный I(1)-процесс можно записать как сумму линейного тренда ($\delta\cdot t$), случайного блуждания без сноса, или стохастического тренда ($\psi(1)(\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t)$), стационарного процесса (η_t) и начального значения ($\xi_0 - \eta_0$). Формулируя несколько отличным образом, имеем:

Утверждение 9.1 (разложение Бевериджа — Нельсона):
Пусть $\{u_t\}$ является I(0)-процессом с нулевым средним, удовлетворяющим (9.2.1)–(9.2.3). Тогда

$$u_1 + u_2 + \cdots + u_t = \psi(1)(\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t) + \eta_t - \eta_0.$$

где $\eta_t \equiv \alpha(L)\varepsilon_t$, $\alpha_j = -(\psi_{j+1} + \psi_{j+2} + \cdots)$. Процесс $\{\eta_t\}$ является эргодически стационарным с нулевым средним[1].

На мгновение положим в (9.2.7) $\delta = 0$, так что $\{\xi_t\}$ является I(1)-процессом без сноса. Важным следствием указанного разложения является то, что любой I(1)-процесс без сноса доминируется стохастическим трендом $\psi(1)\sum_{s=1}^{t}\varepsilon_s$ в следующем смысле. Разделим обе части разложения (9.2.7) (с $\delta = 0$) на $\sqrt{t}$ и получим:

$$\frac{\xi_t}{\sqrt{t}} = \psi(1) \cdot \frac{1}{\sqrt{t}}\sum_{s=1}^{t}\varepsilon_s + \left[\frac{\xi_0}{\sqrt{t}} + \frac{\eta_t}{\sqrt{t}} - \frac{\eta_0}{\sqrt{t}}\right]. \tag{9.2.8}$$

Поскольку, по предположению, $\mathrm{E}(\xi_0^2) < \infty$, $\mathrm{E}[(\xi_0/\sqrt{t})^2] \to 0$ при $t \to \infty$, поэтому $\xi_0/\sqrt{t}$ сходится в среднем квадратическом (и, следовательно, по вероятности) к нулю. То же самое применимо к $\eta_t/\sqrt{t}$ и $\eta_0/\sqrt{t}$. Поэтому слагаемые в квадратных скобках в правой части (9.2.8) можно игнорировать асимптотически. В отличие от этого первое слагаемое имеет предельное распределение $N(0, \sigma^2 \cdot [\psi(1)]^2)$ по CLT Линдеберга — Леви из параграфа 2.1. В этом смысле стохастический тренд растет со скоростью $\sqrt{t}$. Используя (9.2.4), стохастический тренд можно записать как

$$\psi(1)(\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t) = \lambda \cdot \left(\frac{\varepsilon_1}{\sigma} + \frac{\varepsilon_2}{\sigma} + \cdots + \frac{\varepsilon_t}{\sigma}\right), \tag{9.2.9}$$

что показывает: изменения в стохастическом тренде в ξ_t имеют дисперсию λ^2, долговременную дисперсию $\{\Delta\xi_t\}$.

Теперь предположим, что $\delta \neq 0$, так что $\{\xi_t\}$ является I(1)-процессом со сносом. Как ясно из деления обеих сторон (9.2.7) на t, а не на $\sqrt{t}$ и применения той же самой аргументации, которая была только что приведена для случая $\delta = 0$, стохастический тренд, так же как и стационарную компоненту, можно игнорировать асимптотически, с ξ_t/t, сходящимся по вероятности к δ. В этом смысле временной тренд доминирует I(1)-процесс на больших выборках.

Связь с моделями ARMA

В параграфе 6.2 мы определили стационарный ARMA-процесс с нулевым средним $\{u_t\}$ как единственный ковариационно стационарный процесс, удовлетворяющий стационарному ARMA(p, q) уравнению:

$$\phi(L)u_t = \theta(L)\varepsilon_t, \tag{9.2.10}$$

[1]Относительно более общего условия, при котором $u_1 + \cdots + u_t$ раскладывается, как показано здесь, и которое не требует, чтобы I(0)-процесс $\{u_t\}$ был линейным, см.: Theorem 5.4 в [Hall and Heyde, 1980]. В этой теореме последовательность $\{\varepsilon_t\}$ является стационарной мартингал-разностью, поэтому стохастический тренд является мартингалом.

где $\phi(L)$ удовлетворяет условию стационарности. Если $\{\varepsilon_t\}$ является независимым процессом белого шума и если $\theta(1) \neq 0$, то процесс ARMA(p, q) является I(0)-процессом с нулевым средним. Чтобы увидеть это, заметим, что, в соответствии с утверждением 6.5(a), u_t можно записать как (9.2.1) с

$$\psi(L) = \phi(L)^{-1}\theta(L). \tag{9.2.11}$$

Поскольку $\phi(1) \neq 0$ по условию стационарности и $\theta(1) \neq 0$ по предположению, условие (9.2.3b) (о том, что $\psi(1) \neq 0$) выполняется. По утверждению 6.4(a), последовательность коэффициентов $\{\psi_j\}$ ограничена по абсолютному значению геометрически убывающей последовательностью, что гарантирует выполнение (9.2.3a).

При заданном только что определенном стационарном ARMA(p, q)-процессе u_t с нулевым средним, соответствующий ему I(1)-процесс $\{\xi_t\}$, определенный соотношением $\Delta\xi_t = u_t$, называется **интегрированным процессом авторегрессии — скользящего среднего** (ARIMA$(p, 1, q)$) (*autoregressive integrated moving average*). Подставляя соотношение $u_t = (1 - L)\xi_t$ в (9.2.10), мы получаем:

$$\phi^*(L)\xi_t = \theta(L)\varepsilon_t, \tag{9.2.12}$$

где $\phi^*(L) \equiv \phi(L)(1 - L)$, $\phi(L)$ стационарный и $\theta(1) \neq 0$. Один из корней соответствующего авторегрессионного полиномиального уравнения $\phi^*(z) = 0$ равен единице, а все другие корни лежат за пределами единичного круга. В более общем случае класс I(d)-процессов можно представить как удовлетворяющий (9.2.12), где $\phi^*(L)$ определяется уже как

$$\phi^*(L) = \phi(L)(1 - L)^d. \tag{9.2.13}$$

Поэтому $\phi^*(z) = 0$ имеет теперь единичный корень кратности d (то есть d единичных корней), а все другие корни лежат вне единичного круга. Процессы I(d) называют интегрированными процессами авторегрессии — скользящего среднего (ARIMA(p, d, q)). Первый параметр (p) равен количеству авторегрессионных лагов (не считая единичных корней), второй параметр (d) относится к порядку интегрированности, а третий параметр (q) равен количеству лагов в скользящем среднем. Поскольку $\phi(L)$ удовлетворяет условию стационарности, взятие d-х разностей ARIMA(p, d, q)-процесса приводит к стационарному ARMA(p, q)-процессу с нулевым средним. Поэтому ARIMA(p, d, q)-процесс является I(d).

Винеровский процесс

В двух следующих параграфах будут представлены разнообразные тесты на единичный корень. Предельные распределения их тестовых статистик будут записаны в терминах **винеровского процесса** (*Wiener process*) (известного также как **процесс броуновского движения** (*Brownian motion process*)). Некоторые из вас, возможно, уже знакомы с этим процессом по финансовым моделям с непрерывным временем, но чтобы освежить вашу память, приведем его определение.

Определение 9.3 (стандартный винеровский процесс): Стандартный винеровский процесс (броуновское движение) $W(\cdot)$ — это случайный процесс с непрерывным временем, связывающий с каждым моментом $t \in [0, 1]$ скалярную случайную величину $W(t)$, такую, что

(1) $W(0) = 0$;

(2) для любых моментов $0 \le t_1 < t_2 < \cdots < t_k \le 1$ изменения

$$W(t_2) - W(t_1), W(t_3) - W(t_2), \ldots, W(t_k) - W(t_{k-1})$$

являются независимыми нормально распределенными случайными величинами $W(s) - W(t) \sim N(0, (s - t))$ (так что, в частности, $W(1) \sim N(0, 1)$);

(3) с вероятностью 1 реализации $W(t)$ непрерывны по t.

Грубо говоря, **функциональная центральная предельная теорема** (*functional central limit theorem*) (**FCLT**), называемая также **принципом инвариантности** (*invariance principle*), утверждает, что винеровский процесс является аналогом случайного блуждания без сноса в непрерывном времени. Представьте, что вы генерируете реализацию стандартного случайного блуждания (изменения которого имеют единичную дисперсию) длины T, изменяете масштаб графика по вертикали, деля все значения реализации на $\sqrt{T}$, а затем сжимаете этот нормализованный график по горизонтали так, чтобы он соответствовал единичному интервалу $[0, 1]$. На панели (a) рис. 9.2 это сделано для размера выборки $T = 10$. Размер выборки возрастает до 100 на панели (b), а затем — до 1000 — на панели (c). Как показано на рисунке, график становится все более и более плотным в единичном интервале. FCLT гарантирует нам, что существует определенный предел при $T \to \infty$ и что предельный процесс является винеровским процессом. Свойство (2) определения 9.3 является математической формулировкой того, что последовательность мгновенных изменений винеровского процесса является i.i.d. Аналог в непрерывном времени случайного блуждания без сноса, изменения которого имеют дисперсию σ^2, а не единичную дисперсию, можно записать как $\sigma W(r)$.

Чтобы провести аналогию между случайным блужданием без сноса и винеровским процессом чуть дальше, рассмотрим «центрированное стандартное случайное блуждание», которое строится из стандартного случайного блуждания без сноса путем вычитания выборочного средне го. То есть пусть $\{\xi_t\}$ является случайным блужданием без сноса

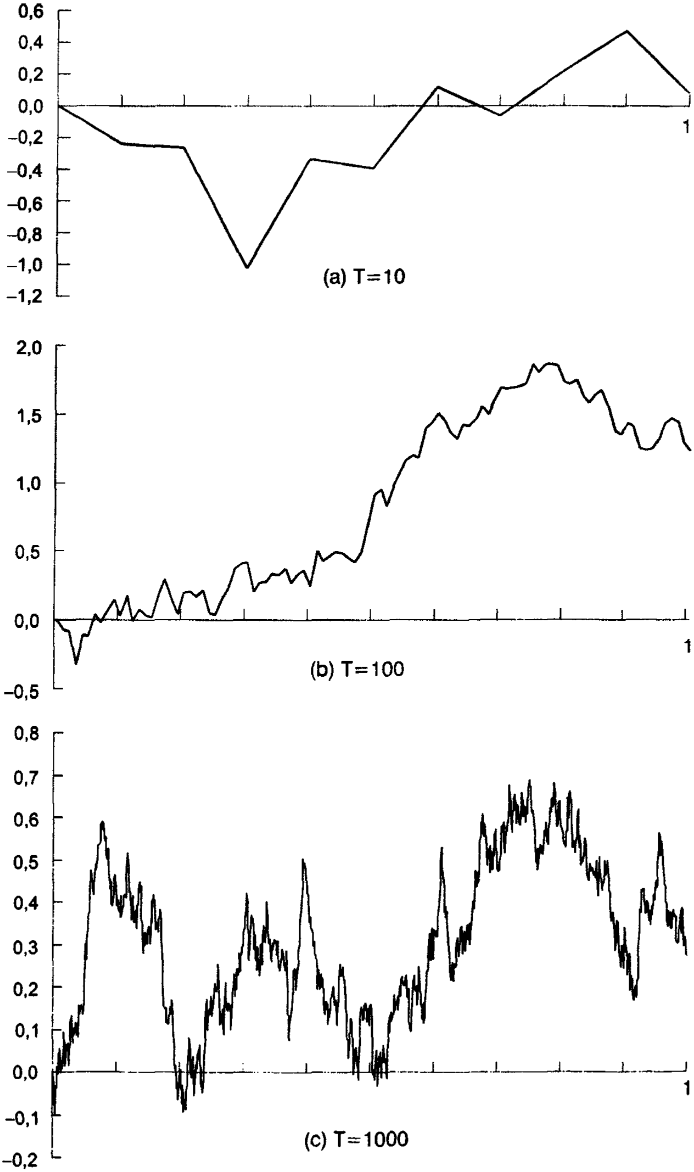

Рис. 9.2. Иллюстрация функциональной центральной предельной теоремы

с $\mathrm{Var}(\Delta\xi_t) = 1$, и определим

$$\xi_t^{\mu} \equiv \xi_t - \frac{\xi_0 + \xi_1 + \cdots + \xi_{T-1}}{T} \quad (t = 0, 1, \ldots, T-1). \tag{9.2.14}$$

Аналогом в непрерывном времени центрированного стандартного случайного блуждания является **центрированный стандартный винеровский процесс** (*demeaned standard Wiener process*), определяемый как

$$W^{\mu}(r) \equiv W(r) - \int_0^1 W(s)\,\mathrm{d}s. \tag{9.2.15}$$

(Мы определили центрированные ряды для $t = 0, 1, \ldots, T-1$, чтобы соответствовать соглашению о датировке утверждения 9.2 ниже. Если бы центрированный ряд был определен для $t = 1, 2, \ldots, T$, то его аналогом в непрерывном времени тоже был (9.2.15).)

Мы можем также создать из стандартного случайного блуждания $\{\xi_t\}$ детрендированный ряд:

$$\xi_t^{\tau} \equiv \xi_t - \widehat{\alpha} - \widehat{\delta}\cdot t \quad (t = 0, 1, \ldots, T-1), \tag{9.2.16}$$

где $\widehat{\alpha}$ и $\widehat{\delta}$ являются OLS-оценками константы и коэффициента при времени в регрессии ξ_t на $(1, t)$ $(t = 0, 1, \ldots, T-1)$. Аналогом в непрерывном времени детрендированного случайного блуждания является **детрендированный стандартный винеровский процесс** (*detrended standard Wiener process*), определяемый как

$$\begin{aligned} W^{\tau}(r) &\equiv W(r) - a - d\cdot r, \\ a &\equiv \int_0^1 (4 - 6s)W(s)\,\mathrm{d}s, \\ d &\equiv \int_0^1 (-6 + 12s)W(s)\,\mathrm{d}s. \end{aligned} \tag{9.2.17}$$

Коэффициенты a и d являются предельными случайными величинами при $T \to \infty$ для $\widehat{a}$ и $\widehat{d}$ соответственно (вывод см., например, в: [Phillips and Durlauf, 1986]). Следовательно, если линейный временной тренд подгоняется посредством OLS к случайному блужданию *без сноса*, оцененный линейный временной тренд $\widehat{\delta}$ сходится по распределению к случайной величине d; даже на больших выборках значение $\widehat{\delta}$ никогда не будет нулевым, разве только случайно. Этот феномен иногда называется **ложным детрендированием** (*spurious detrending*).

Полезная лемма

Тесты на единичный корень используют статистики, включающие I(0)- и I(1)-процессы. Ключевые результаты, собранные в следующем утверждении, будут использоваться для получения предельных распределений статистик тестов на единичный корень в параграфах 9.3–9.4.

Утверждение 9.2 (предельные распределения статистик, включающих I(0)- и I(1)-переменные): *Пусть $\{\xi_t\}$ является* I(1)*-процессом без сноса, так что $\Delta\xi_t$ является* I(0)*-процессом с нулевым средним, удовлетворяющим* (9.2.1)–(9.2.3), *и* $\mathrm{E}(\xi_0^2) < \infty$. *Пусть*

$$\lambda^2 \equiv \textit{долговременная дисперсия } \{\Delta\xi_t\} \quad \textit{и} \quad \gamma_0 \equiv \mathrm{Var}(\Delta\xi_t).$$

Тогда

(a) $\dfrac{1}{T^2}\sum_{t=1}^{T}(\xi_{t-1})^2 \to_d \lambda^2 \cdot \int_0^1 W(r)^2\,\mathrm{d}r,$

(b) $\dfrac{1}{T}\sum_{t=1}^{T}\Delta\xi_t\,\xi_{t-1} \to_d \frac{\lambda^2}{2}W(1)^2 - \frac{\gamma_0}{2}.$

Пусть $\{\xi_t^{\mu}\}$ является центрированным рядом, полученным из $\{\xi_t\}$ по формуле (9.2.14). *Тогда*

(c) $\dfrac{1}{T^2}\sum_{t=1}^{T}(\xi_{t-1}^{\mu})^2 \to_d \lambda^2 \cdot \int_0^1 [W^{\mu}(r)]^2\,\mathrm{d}r,$

(d) $\dfrac{1}{T}\sum_{t=1}^{T}\Delta\xi_t\,\xi_{t-1}^{\mu} \to_d \frac{\lambda^2}{2}\left\{[W^{\mu}(1)]^2 - [W^{\mu}(0)]^2\right\} - \frac{\gamma_0}{2}.$

Пусть $\{\xi_t^{\tau}\}$ является детрендированным рядом, полученным из $\{\xi_t\}$ по формуле (9.2.16). *Тогда*

(e) $\dfrac{1}{T^2}\sum_{t=1}^{T}(\xi_{t-1}^{\tau})^2 \to_d \lambda^2 \cdot \int_0^1 [W^{\tau}(r)]^2\,\mathrm{d}r,$

(f) $\dfrac{1}{T}\sum_{t=1}^{T}\Delta\xi_t\,\xi_{t-1}^{\tau} \to_d \frac{\lambda^2}{2}\left\{[W^{\tau}(1)]^2 - [W^{\tau}(0)]^2\right\} - \frac{\gamma_0}{2}.$

Сходимость здесь совместная. То есть вектор, состоящий из статистик, указанных в (a)–(f), сходится к случайному вектору, элементы которого являются соответствующими случайными величинами, также указанными в (a)–(f).

Часть (a), например, читается следующим образом: последовательность случайных величин $\{(1/T)^2\sum_{t=1}^{T}(\xi_{t-1})^2\}$, индексированная посредством T, сходится по распределению к случайной величине $\lambda^2\int W^2\mathrm{d}r$. Все предельные случайные величины в этом утверждении записываются в терминах стандартного винеровского процесса. Заметим, что *тот же самый* винеровский процесс $W(\cdot)$ появляется в (a) и (b) и что центрированный и детрендированный винеровские процессы в (c)–(f) получены из винеровского процесса, появляющегося в (a) и (b). Поэтому предельные случайные величины в (a)–(f) могут быть коррелированными.

Чтобы получить определенное понимание этого результата, временно предположим, что $\{\xi_t\}$ является случайным блужданием без сноса с $\mathrm{Var}(\Delta\xi_t) = \sigma^2$ и фиксированным ξ_0. Учитывая, что винеровский процесс является его аналогом в непрерывном времени, неудивительно, что «$\sum_{t=1}^{T}(\xi_{t-1})^2$», соответствующим образом нормализованная на некоторую степень T, становится «$\sigma^2 \int W^2$». Возможно, что вы не могли догадаться нормализовать ее посредством T^2. Один из способов понять, почему именно квадрат T дает правильную нормализацию, — это вспомнить, что $\mathrm{E}[(\xi_t)^2] = \mathrm{Var}(\xi_t) = \sigma^2 \cdot t$. Таким образом,

$$\mathrm{E}\left[\sum_{t=1}^{T}(\xi_{t-1})^2\right] = \sigma^2 \sum_{t=1}^{T}(t-1) = \sigma^2 \cdot (T-1)\frac{T}{2}. \tag{9.2.18}$$

Так что среднее $\sum_{t=1}^{T}(\xi_{t-1})^2$ растет со скоростью T^2. Чтобы построить случайную величину, которая имела бы предельное распределение, придется разделить $\sum_{t=1}^{T}(\xi_{t-1})^2$ на T^2.

Предположим теперь, что $\{\xi_t\}$ является общим I(1)-процессом без сноса, первая разность которого является процессом I(0) с нулевым средним, удовлетворяющим (9.2.1)–(9.2.3). Тогда сериальную корреляцию в $\Delta\xi_t$ можно принять во внимание путем замены σ^2 на λ^2 (долговременная дисперсия $\{\Delta\xi_t\}$). Это происходит благодаря следствию (9.2.9) разложения Бевериджа — Нельсона, говорящему, что I(1)-процесс без сноса на больших выборках доминируется случайным блужданием, изменения которого имеют дисперсию λ^2. Иначе говоря, предельное распределение для $\sum_{t=1}^{T}(\xi_{t-1})^2/T^2$ было бы тем же самым, если бы $\{\xi_t\}$, вместо того чтобы являться общим I(1)-процессом без сноса с сериально коррелированными первыми разностями, был случайным блужданием без сноса, изменения которого имеют дисперсию λ^2.

По поводу доказательства утверждения 9.2 см., например: [Stock, 1994, pp. 2751–2753]. Это, скорее, механическое применение FCLT и теоремы, называемой теоремой о непрерывном отображении. Поскольку $W(1) \sim N(0, 1)$, предельная случайная величина в (b) принимает вид:

$$\left(\frac{\lambda^2}{2}\right)X - \frac{\gamma_0}{2}, \tag{9.2.19}$$

где $X \sim \chi^2(1)$ (хи-квадрат с одной степенью свободы). Доказательство (b) можно выполнить без причудливого аппарата FCLT и теоремы о непрерывном отображении, и это остается в качестве аналитического упражнения 1.

Контрольные вопросы

1. (Проверка разложения Бевериджа — Нельсона для простого случая.) Проверьте (9.2.5) для $\psi(L) = 1 + \psi_1 L$. Сделайте то же самое для $\psi(L) = (1 - \phi L)^{-1}$, где $|\phi| < 1$.

2. (CLT для I(0).) Утверждение 6.9 касается асимптотики выборочного среднего линейного процесса. Проверьте, что абзац, расположенный непосредственно за утверждением 9.1, является доказательством утверждения 6.9 при дополнительном предположении о 1-суммируемости.

3. (Является ли стационарная компонента в разложении Бевериджа — Нельсона I(0)-процессом?) Рассмотрите процесс I(0) с нулевым средним:

$$u_t = \psi_0 \varepsilon_t - 2\varepsilon_{t-1} + \varepsilon_{t-2}.$$

Покажите, что долговременная дисперсия процесса $\{\eta_t\}$ в (9.2.6) равна нулю.

4. (Начальные значения и центрированные величины.) Пусть $\{\xi_t\}$ является I(1)-процессом без сноса, так что ξ_t можно записать как $\xi_t = \xi_0 + u_1 + u_2 + \cdots + u_t$. Влияет ли значение ξ_0 на значение ξ_t^{μ}, где $\{\xi_t^{\mu}\}$ является центрированным рядом, полученным по формуле (9.2.14)? [Ответ: Нет.]

5. (Линейные тренды и детрендированные величины.) Пусть $\{\xi_t\}$ — I(1)-процесс со сносом, так что его можно записать как сумму линейного временного тренда $\alpha + \delta \cdot t$ и I(1)-процесса без сноса. Влияют ли значения α и δ на значение ξ_t^{τ}, где $\{\xi_t^{\tau}\}$ является центрированным рядом, полученным по формуле (9.2.16)? [Ответ: Нет.]

9.3. Тесты Дики — Фуллера

В этом параграфе и в следующем мы представим несколько тестов для проверки нулевой гипотезы о том, что процесс порождения данных (DGP) является I(1)-процессом. В этом параграфе I(1)-процесс является случайным блужданием с или без сноса и тесты основаны на OLS-оценивании соответствующих AR(1)-уравнений. Эти тесты были разработаны в: [Dickey, 1976] и [Fuller, 1976] и называются **тестами Дики — Фуллера** (*Dickey — Fuller tests*), или **DF-тестами** (*DF tests*). Мы принимаем соглашение о том, что выборка содержит $T+1$ наблюдений $(y_0, y_1, \ldots, y_T)$, так что AR(1)-уравнение можно оценить для $t = 1, 2, \ldots, T$.

AR(1)-*модель*

Тестовые статистики Дики — Фуллера получаются из оценивания модели авторегрессии первого порядка:

$$y_t = \rho y_{t-1} + \varepsilon_t, \quad \{\varepsilon_t\} \text{ — независимый белый шум.} \tag{9.3.1}$$

(Случай с константой будет изучен через мгновение.) Эта модель включает и I(1)-, и I(0)-процессы: если $\rho = 1$, то есть если соответствующее полиномиальное уравнение $1 - \rho z = 0$ имеет единичный корень, то $\Delta y_t = \varepsilon_t$ и $\{y_t\}$ является случайным блужданием без тренда, а если $|\rho| < 1$, процесс является стационарным AR(1) с нулевым средним. Таким образом, гипотезу о том, что данные порождаются случайным блужданием без тренда, можно сформулировать как нулевую гипотезу о том, что $\rho = 1$. Если мы принимаем ту позицию, что DGP или I(0), или I(1), то $-1 < \rho \leq 1$. При условии этого *априорного* ограничения на ρ альтернативу I(0) можно выразить как $\rho < 1$. Поэтому для проверки нулевой гипотезы $\rho = 1$ против альтернативы I(0) будут использоваться односторонние тесты.

Соответствующие тесты будут основаны на $\widehat{\rho}$, OLS-оценке для ρ в AR(1)-уравнении (9.3.1). При нулевой гипотезе I(1), $\rho = 1$, ошибка оценки равна $\widehat{\rho} - 1$ и задается выражением:

$$\widehat{\rho} - 1 \equiv \frac{\sum_{t=1}^{T} y_t\, y_{t-1}}{\sum_{t=1}^{T}(y_{t-1})^2} - 1 = \frac{\sum_{t=1}^{T} \Delta y_t\, y_{t-1}}{\sum_{t=1}^{T}(y_{t-1})^2}. \tag{9.3.2}$$

Как будет проверено через мгновение, числитель имеет предельное распределение, если разделить его на T, а знаменатель имеет предельное распределение, если разделить его на T^2. Поэтому рассмотрим ошибку оценки, увеличенную в T раз:

$$T \cdot (\widehat{\rho} - 1) = T \cdot \frac{\sum_{t=1}^{T} \Delta y_t\, y_{t-1}}{\sum_{t=1}^{T}(y_{t-1})^2} = \frac{\frac{1}{T}\sum_{t=1}^{T} \Delta y_t\, y_{t-1}}{\frac{1}{T^2}\sum_{t=1}^{T}(y_{t-1})^2}. \tag{9.3.3}$$

Получение предельного распределения при нулевой гипотезе I(1)

Получить предельное распределение $T \cdot (\widehat{\rho} - 1)$ при нулевой гипотезе I(1) (случайное блуждание без тренда) очень просто. Поскольку $\{y_t\}$ является I(1)-процессом без сноса, применимы части (a) и (b) утверждения 9.2 с $\xi_t = y_t$. Кроме того, поскольку Δy_t является независимым белым шумом, λ^2 (долговременная дисперсия $\{\Delta y_t\}$) равна γ_0 (дисперсии Δy_t). Таким образом,

$$w_{1T} \equiv \frac{1}{T}\sum_{t=1}^{T} \Delta y_t\, y_{t-1} \underset{\mathrm{d}}{\to} \frac{\gamma_0}{2} W(1)^2 - \frac{\gamma_0}{2} \equiv w_1, \tag{9.3.4}$$

$$w_{2T} \equiv \frac{1}{T^2}\sum_{t=1}^{T}(y_{t-1})^2 \underset{\mathrm{d}}{\to} \gamma_0 \int_0^1 W(r)^2\,\mathrm{d}r \equiv w_2. \tag{9.3.5}$$

Как отмечалось в указанном утверждении, сходимость является *совместной*, так что $\boldsymbol{w}_T \equiv (w_{1T}, w_{2T})'$ как вектор сходится к 2×1 случайному вектору $\boldsymbol{w} \equiv (w_1, w_2)'$. Поскольку (9.3.3) является непрерывной функцией от $\boldsymbol{w}_T$ (оно равно w_{1T}/w_{2T}), из леммы 2.3(b) следует, что

$$\begin{aligned} T \cdot (\widehat{\rho} - 1) &= \frac{\frac{1}{T}\sum_{t=1}^{T} \Delta y_t\, y_{t-1}}{\frac{1}{T^2}\sum_{t=1}^{T}(y_{t-1})^2} = \frac{w_{1T}}{w_{2T}} \underset{\mathrm{d}}{\to} \\ &\underset{\mathrm{d}}{\to} \frac{w_1}{w_2} = \frac{\frac{\gamma_0}{2} W(1)^2 - \frac{\gamma_0}{2}}{\gamma_0 \cdot \int_0^1 W(r)^2\,\mathrm{d}r} = \frac{\frac{1}{2}(W(1)^2 - 1)}{\int_0^1 W(r)^2\,\mathrm{d}r} \equiv DF_\rho. \end{aligned} \tag{9.3.6}$$

Тестовая статистика $T \cdot (\widehat{\rho} - 1)$ называется **ρ-статистикой Дики — Фуллера** (*Dickey — Fuller, или DF, ρ statistic*). Стоит отметить несколько моментов:

- Невырожденное предельное распределение имеет $T \cdot (\widehat{\rho} - 1)$, а не обычное $\sqrt{T} \cdot (\widehat{\rho} - 1)$. Оценка $\widehat{\rho}$ называется **суперсостоятельной** (*superconsistent*), поскольку она сходится к истинному значению, равному единице, с более высокой скоростью (T).

- Нулевая гипотеза не специфицирует значения $\mathrm{Var}(\varepsilon_t)$ и y_0 (начальное значение) в DGP, однако они не влияют на предельное распределение, распределение случайной величины DF_ρ. То есть предельное распределение не включает эти **мешающие параметры** (*nuisance parameters*). Поэтому мы можем использовать $T \cdot (\widehat{\rho} - 1)$ в качестве тестовой статистики. Этот тест на случайное блуждание без сноса называется **ρ-тестом Дики — Фуллера** (*Dickey — Fuller ρ test*).

- Заметим, что и числитель, и знаменатель включают *один и тот же винеровский процесс*, так что они коррелированы. Поскольку $W(1)$ имеет стандартное нормальное распределение, числитель в выражении для DF_ρ можно было бы записать как $(\chi^2(1) - 1)/2$. Но при этом стало бы неясным сделанное здесь замечание.

Обычный t-тест для нулевой гипотезы $\rho = 1$ также имеет предельное распределение. Это всего лишь вопрос простой алгебры — показать, что t-значение $\rho = 1$ можно записать как

$$t = \frac{\widehat{\rho} - 1}{s \Big/ \sqrt{\sum_{t=1}^{T} (y_{t-1})^2}} = \frac{\frac{1}{T} \sum_{t=1}^{T} \Delta y_t \, y_{t-1}}{s \cdot \sqrt{\frac{1}{T^2} \sum_{t=1}^{T} (y_{t-1})^2}}, \tag{9.3.7}$$

где s является стандартной ошибкой регрессии:

$$s \equiv \sqrt{\frac{1}{T-1} \sum_{t=1}^{T} (y_t - \widehat{\rho} y_{t-1})^2}. \tag{9.3.8}$$

Легко доказать (см. контрольный вопрос 3), что s^2 является состоятельной оценкой для $\gamma_0 (\equiv \mathrm{Var}(\Delta y_t))$. Из этого, (9.3.4) и (9.3.5), тогда немедленно следует, что

$$t \underset{\mathrm{d}}{\to} \frac{\frac{1}{2}(W(1)^2 - 1)}{\sqrt{\int_0^1 W(r)^2 \, \mathrm{d}r}} \equiv DF_t. \tag{9.3.9}$$

Это предельное распределение также не зависит от мешающих параметров. Поэтому t-значение можно использовать для тестирования, однако при этом нельзя использовать критические значения стандартного нормального распределения; критические значения следует брать из (приведенных ниже) таблиц для DF_t. Этот тест называется **DF t-тестом** (*DF t test*).

Суммируем изложенное.

Утверждение 9.3 (тесты Дики — Фуллера на случайное блуждание без сноса, случай без константы): *Предположим, что $\{y_t\}$ является случайным блужданием без сноса (так что $\{\Delta y_t\}$ является независимым белым шумом) с $\mathrm{E}(y_0^2) < \infty$. Рассмотрим регрессию y_t на y_{t-1} (без константы) для $t = 1, 2, \ldots, T$. Тогда*

$$\text{DF } \rho\text{-статистика: } T \cdot (\widehat{\rho} - 1) \underset{d}{\to} DF_\rho,$$

$$\text{DF } t\text{-статистика: } t \underset{d}{\to} DF_t,$$

где $\widehat{\rho}$ является OLS-оценкой коэффициента при y_{t-1}, t является t-значением для гипотезы о том, что коэффициент при y_{t-1} равен 1, а DF_ρ и DF_t являются случайными величинами, определенными в (9.3.6) *и* (9.3.9).

Таким образом, у нас есть две тестовые статистики, $T \cdot (\widehat{\rho} - 1)$ и t-значение, для одной и той же нулевой гипотезы I(1). Сравнение (обобщений) этих DF-тестов будет обсуждаться в следующем параграфе.

В части (а) табл. 9.1 приведены критические значения для распределения на конечных выборках случайной величины $T \cdot (\widehat{\rho} - 1)$ для различных объемов выборки. Сплошная кривая на рис. 9.3 изображает распределение $\widehat{\rho}$ на конечных выборках для $T = 100$, показывающее, что OLS-оценка для ρ смещена вниз (то есть среднее $\widehat{\rho}$ меньше 1). Все это получено методом Монте-Карло в предположении, что распределение ε_t является гауссовским и $y_0 = 0$. По утверждению 9.3, при возрастании объема выборки T эти критические значения сходятся к критическим значениям для распределения DF_ρ, приведенным в строке для $T = \infty$. (Поскольку для выполнения асимптотических результатов в утверждении 9.3 y_0 необязательно должно быть нулевым, а распределение ε необязательно должно быть гауссовским, мы должны получить те же самые предельные критические значения, если критические значения для конечных выборок вычислять при различных предположениях о начальном значении y_0 и распределении ε_t.) По критическим значениям для $T = \infty$ мы можем видеть, что распределение DF_ρ также скошено влево с 5-процентным левосторонним критическим значением, равным $-8{,}1$ (так что $\mathrm{Prob}(DF_\rho < -8.1) = 5\%$), и 5-процентным правосторонним критическим значением, равным 1,28. Смещение в $\widehat{\rho}$ на конечных выборках отражается в скошенности предельного распределения $T \cdot (\widehat{\rho} - 1)$.

Таблица 9.1

Критические значения для теста ρ Дики — Фуллера

Объем выборки (T)	Вероятность того, что статистика меньше, чем указанное число							
	0,01	**0,025**	**0,05**	**0,10**	**0,90**	**0,95**	**0,975**	**0,99**
	Панель (a): $T \cdot (\widehat{\rho} - 1)$							
25	−11,8	−9,3	−7,3	−5,3	1,01	1,41	1,78	2,28
50	−12,8	−9,9	−7,7	−5,5	0,97	1,34	1,69	2,16
100	−13,3	−10,2	−7,9	−5,6	0,95	1,31	1,65	2,09
250	−13,6	−10,4	−8,0	−5,7	0,94	1,29	1,62	2,05
500	−13,7	−10,4	−8,0	−5,7	0,93	1,28	1,61	2,04
∞	−13,8	−10,5	−8,1	−5,7	0,93	1,28	1,60	2,03
	Панель (b): $T \cdot (\widehat{\rho}^{\mu} - 1)$							
25	−17,2	−14,6	−12,5	−10,2	−0,76	0,00	0,65	1,39
50	−18,9	−15,7	−13,3	−10,7	−0,81	−0,07	0,53	1,22
100	−19,8	−16,3	−13,7	−11,0	−0,83	−0,11	0,47	1,14
250	−20,3	−16,7	−13,9	−11,1	−0,84	−0,13	0,44	1,08
500	−20,5	−16,8	−14,0	−11,2	−0,85	−0,14	0,42	1,07
∞	−20,7	−16,9	−14,1	−11,3	−0,85	−0,14	0,41	1,05
	Панель (c): $T \cdot (\widehat{\rho}^{t} - 1)$							
25	−22,5	−20,0	−17,9	−15,6	−3,65	−2,51	−1,53	−0,46
50	−25,8	−22,4	−19,7	−16,8	−3,71	−2,60	−1,67	−0,67
100	−27,4	−23,7	−20,6	−17,5	−3,74	−2,63	−1,74	−0,76
250	−28,5	−24,4	−21,3	−17,9	−3,76	−2,65	−1,79	−0,83
500	−28,9	−24,7	−21,5	−18,1	−3,76	−2,66	−1,80	−0,86
∞	−29,4	−25,0	−21,7	−18,3	−3,77	−2,67	−1,81	−0,88

Источник: [Fuller, 1976, Table 8.5.1], исправленный в [Fuller, 1996, Table 10.A.1].

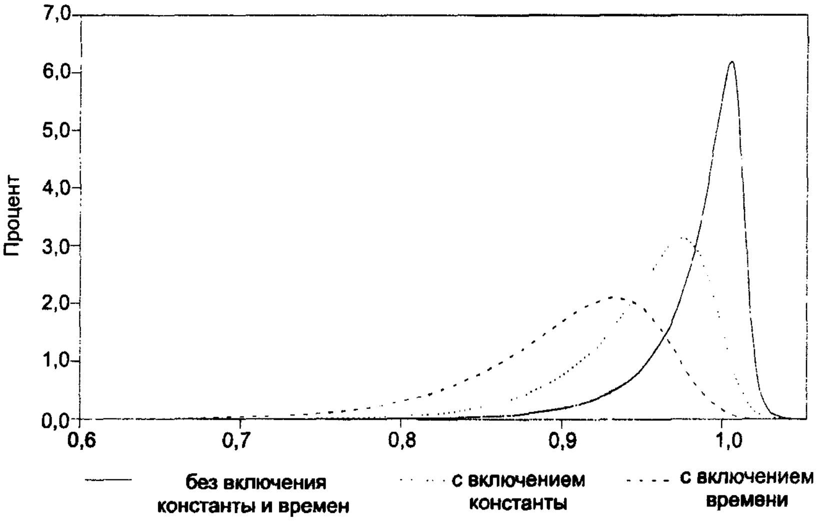

Рис. 9.3. Распределение на конечных выборках OLS-оценки для AR(1)-коэффициента, $T = 100$

Критические значения для DF_t (предельного распределения t-значения) находятся в строке для $T = \infty$ в табл. 9.2, панель (а). Как и DF_ρ, оно скошено влево. В этой таблице также приведены критические значения для конечных объемов выборки. И опять, они предполагают, что $y_0 = 0$ и ε_t имеет нормальное распределение.

Таблица 9.2

Критические значения для теста t Дики — Фуллера

Объем выборки (T)	**Вероятность того, что статистика меньше, чем указанное число**							
	0,01	**0,025**	**0,05**	**0,10**	**0,90**	**0,95**	**0,975**	**0,99**
				Панель (а): t				
25	−2,65	−2,26	−1,95	−1,60	0,92	1,33	1,70	2,15
50	−2,62	−2,25	−1,95	−1,61	0,91	1,31	1,66	2,08
100	−2,60	−2,24	−1,95	−1,61	0,90	1,29	1,64	2,04
250	−2,58	−2,24	−1,95	−1,62	0,89	1,28	1,63	2,02
500	−2,58	−2,23	−1,95	−1,62	0,89	1,28	1,62	2,01
∞	−2,58	−2,23	−1,95	−1,62	0,89	1,28	1,62	2,01
				Панель (b): t^μ				
25	−3,75	−3,33	−2,99	−2,64	−0,37	0,00	0,34	0,71
50	−3,59	−3,23	−2,93	−2,60	−0,41	−0,04	0,28	0,66
100	−3,50	−3,17	−2,90	−2,59	−0,42	−0,06	0,26	0,63
250	−3,45	−3,14	−2,88	−2,58	−0,42	−0,07	0,24	0,62
500	−3,44	−3,13	−2,87	−2,57	−0,44	−0,07	0,24	0,61
∞	−3,42	−3,12	−2,86	−2,57	−0,44	−0,08	0,23	0,60
				Панель (c): t^t				
25	−4,38	−3,95	−3,60	−3,24	−1,14	−0,81	−0,50	−0,15
50	−4,16	−3,80	−3,50	−3,18	−1,19	−0,87	−0,58	−0,24
100	−4,05	−3,73	−3,45	−3,15	−1,22	−0,90	−0,62	−0,28
250	−3,98	−3,69	−3,42	−3,13	−1,23	−0,92	−0,64	−0,31
500	−3,97	−3,67	−3,42	−3,13	−1,24	−0,93	−0,65	−0,32
∞	−3,96	−3,67	−3,41	−3,13	−1,25	−0,94	−0,66	−0,32

Источник: [Fuller, 1976, Table 8.5.2], исправленный в [Fuller, 1996, Table 10.A.2].

Включение постоянной составляющей

Недостатком тестов, основанных на AR(1)-уравнении без постоянной составляющей, является отсутствие инвариантности к прибавлению к ряду константы. Если тест проводится для ряда в логарифмах (как в примере 9.1 ниже), изменение единицы измерения ряда приводит к прибавлению к ряду в логарифмах некоторой константы, что влияет на значение тестовой статистики. Чтобы сделать тестовую статистику инвариантной к прибавлению константы, мы обобщаем модель, на основе которой получается тестовая статистика. Рассмотрим модель

$$y_t = \alpha + z_t,\ z_t = \rho z_{t-1} + \varepsilon_t,\ \{\varepsilon_t\} \text{ — независимый белый шум.} \quad (9.3.10)$$

Процесс $\{y_t\}$ получается прибавлением постоянной α к процессу, удовлетворяющему (9.3.1). При нулевой гипотезе $\rho = 1$ $\{z_t\}$ является случайным блужданием без сноса. Поскольку y_t можно записать в виде:

$$y_t = \alpha + z_0 + \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t,$$

$\{y_t\}$ также является случайным блужданием без сноса с начальным значением $y_0 \equiv \alpha + z_0$. При альтернативе $\rho < 1$ $\{y_t\}$ является стационарным AR(1) со средним α. Таким образом, класс I(0)-процессов, охватываемый моделью (9.3.10), шире, чем класс процессов, охватываемых моделью (9.3.1).

Исключая $\{z_t\}$ из (9.3.10), мы получаем AR1-уравнение с константой:

$$y_t = \alpha^* + \rho y_{t-1} + \varepsilon_t, \quad (9.3.11)$$

где

$$\alpha^* = (1-\rho)\alpha. \quad (9.3.12)$$

Поскольку $\alpha^* = 0$ при $\rho = 1$, нулевая гипотеза I(1) (о том, что DGP является случайным блужданием без сноса) является совместной гипотезой о том, что $\rho = 1$ *и* $\alpha^* = 0$ в терминах коэффициентов регрессии (9.3.1). Без ограничения $\alpha^* = 0$ $\{y_t\}$ было бы случайным блужданием *со* сносом. Мы разработаем тесты на случайное блуждание со сносом чуть позже. До этого мы продолжаем брать в качестве нулевой гипотезу случайного блуждания без сноса.

Пусть $\widehat{\rho}^{\mu}$ является здесь OLS-оценкой для ρ в (9.3.11), а t^{μ} — t-значение для нулевой гипотезы о том, что $\rho = 1$. Должно быть ясно, что α не будет влиять на значение $\widehat{\rho}^{\mu}$ или на ее OLS стандартной ошибки: прибавление одной и той же постоянной к y_t для всех t просто изменяет оцененную постоянную составляющую. Следовательно, распределение **DF ρ^{μ}-статистики** (*DF ρ^{μ} statistic*), $T\cdot(\widehat{\rho}^{\mu}-1)$ и **DF t^{μ}-статистики** (*DF t^{μ} statistic*), t^{μ}, как на конечных выборках, так и на больших выборках, не будет зависеть от α вне зависимости от значения ρ.

Как вам будет предложено доказать в аналитическом упражнении 2, $T\cdot(\widehat{\rho}^{\mu}-1)$ сходится по распределению к случайной величине, представляемой как

$$DF_{\rho}^{\mu} \equiv \frac{\frac{1}{2}\left([W^{\mu}(1)]^2 - [W^{\mu}(0)]^2 - 1\right)}{\int_0^1 [W^{\mu}(r)]^2\,\mathrm{d}r}, \quad (9.3.13)$$

а t^{μ} сходится по распределению к

$$DF_{t}^{\mu} \equiv \frac{\frac{1}{2}\left([W^{\mu}(1)]^2 - [W^{\mu}(0)]^2 - 1\right)}{\sqrt{\int_0^1 [W^{\mu}(r)]^2\,\mathrm{d}r}}. \quad (9.3.14)$$

Здесь $W^{\mu}(\cdot)$ является центрированным стандартным винеровским процессом, введенным в параграфе 9.2. Эти предельные распределения не зависят от мешающих параметров, таких как α. Основная идея доказательства заключается в следующем. Сначала получаем центрированный ряд, вычитая выборочное среднее из $\{y_t\}$ (или, равносильно, получая OLS-остатки от регрессии $\{y_t\}$ на константу). На втором шаге используем теорему Фриша — Во (приведенную в аналитическом упражнении 4 в главе 1), чтобы записать $T\cdot(\widehat{\rho}^{\mu}-1)$ и t-значение в виде формул, включающих центрированные ряды. Наконец, используем утверждение 9.2(c) и (d). Подведем итог.

Утверждение 9.4 (тесты Дики — Фуллера о случайном блуждании без сноса, случай с постоянной составляющей): *Предположим, что $\{y_t\}$ является случайным блужданием без сноса (так что $\{\Delta y_t\}$ является независимым белым шумом) с $\mathrm{E}(y_0^2) < \infty$. Рассмотрим регрессию y_t на $(1, y_{t-1})$ для $t = 1, 2, \ldots, T$. Тогда*

$$\text{DF } \rho^{\mu}\text{-статистика: } T\cdot(\widehat{\rho}^{\mu}-1) \underset{d}{\to} DF_{\rho}^{\mu}.$$

$$\text{DF } t^{\mu}\text{-статистика: } t^{\mu} \underset{d}{\to} DF_{t}^{\mu},$$

где $\widehat{\rho}^{\mu}$ — OLS-оценка коэффициента при y_{t-1}, t^{μ} — t-значение для гипотезы о том, что коэффициент при y_{t-1} равен 1, а DF_{ρ}^{μ} и DF_{t}^{μ} — две случайные величины, определенные в (9.3.13) *и* (9.3.14).

Критические значения для DF_{ρ}^{μ} приведены в табл. 9.1, панель (b), а для DF_{t}^{μ} — в табл. 9.2, панель (b), в строке для $T=\infty$. Распределение скошено влево сильнее, чем в случае без постоянной составляющей. В таблицах также есть критические значения для выборок конечного объема, основанные на предположении о том, что ε_t имеет нормальное распределение. Распределение $\widehat{\rho}^{\mu}$ на конечных выборках для $T = 100$ изображено на рис. 9.3. Оно показывает более сильное смещение для $\widehat{\rho}^{\mu}$, чем для $\widehat{\rho}$ (OLS-оценка без постоянной составляющей). В отличие от случая без включения постоянной составляющей критические значения на конечных выборках не зависят от y_0. Это происходит потому, что, когда $\rho = 1$, изменение начального значения y_0 просто прибавляет одну и ту же постоянную к y_t для всех t, так что значения DF ρ^{μ}- и t^{μ}-статистик инвариантны к y_0.

Мы начали этот подпараграф, сказав, что нулевая гипотеза I(1) является совместной гипотезой о том, что $\rho = 1$ *и* $\alpha^* = 0$. Тем не менее указанные тестовые статистики используются для проверки гипотезы о том, что $\rho = 1$. Для тестов на единичный корень это практика в профессии; тестовые статистики для совместной гипотезы используются редко.

Пример 9.1 (являются ли обменные курсы случайным блужданием?): Пусть y_t является спотовым обменным курсом иена/доллар. Чтобы сделать тесты инвариантными к выбору единиц (например, иена/доллар против иена/другая валюта), мы подбираем AR(1)-модель с постоянной составляющей, так что применимо утверждение 9.4. Для недельных данных по обменному курсу, использованных в эмпирическом приложении к в главе 6 и изображенных на рис. 6.3 на с. 454, указанное AR(1)-уравнение, оцененное посредством OLS, имеет вид:

$$\underset{(0.435)}{y_t = 0{,}162} + \underset{(0{,}0019285)}{0{,}9983376}\, y_{t-1}, \quad R^2 = 0.997. \quad SER = 2{,}824.$$

В скобках приведены стандартные ошибки. Из-за наличия лагированной переменной y_{t-1} фактический период выборки начинается со второй недели 1975 года, а не с первой. Поэтому $T = 777$ и

$$T \cdot (\widehat{\rho}^{\mu} - 1) = 777 \cdot (0{,}9983376 - 1) = -1{,}29,$$
$$t^{\mu} = \frac{0.9983376 - 1}{0{,}0019285} = -0{,}86.$$

Поскольку альтернативной является гипотеза о том, что $\rho < 1$, следует использовать односторонний тест. (Если вы полагаете, что процесс может быть взрывным AR(1) с $\rho > 1$, то вы должны применить двухсторонний тест.) Из табл. 9.1, панель (b) для DF_{ρ}^{μ} (асимптотическое) критическое значение, которое дает 5% нижнему хвосту, равно -14.1. То есть

$$\mathrm{P}(DF_{\rho}^{\mu} < -14{,}1) = 0.05.$$

Значение $-1{,}29$ статистики теста значительно выше этого критического значения, поэтому нулевая гипотеза без сомнения не отвергается на 5%-ном уровне значимости. Возвращаясь к DF t-тесту, из табл. 9.2, панель (b), находим, что 5-процентное критическое значение равно $-2{,}86$. t-значение -0.86 больше его, и поэтому нулевая гипотеза, без сомнения, не отвергается.

Включение временного тренда

Как было отмечено в параграфе 9.1, во многих экономических временных рядах проявляются временные тренды. Чтобы сделать DF-тест применимым к временным рядам с трендом, мы еще раз обобщаем модель, добавляя на этот раз линейный временной тренд:

$$y_t = \alpha + \delta \cdot t + z_t, \quad z_t = \rho z_{t-1} + \varepsilon_t, \quad \{\varepsilon_t\} \text{ — независимый белый шум.} \tag{9.3.15}$$

Здесь δ — угловой коэффициент линейного тренда, который может быть равным, а может быть и не равным нулю. При нулевой гипотезе $\rho = 1$ $\{z_t\}$ является случайным блужданием без сноса, и y_t можно записать как

$$\begin{aligned} y_t &= \alpha + \delta \cdot t + z_0 + \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t = \\ &= y_0 + \delta \cdot t + (\varepsilon_1 + \dots + \varepsilon_t) \end{aligned} \tag{9.3.16}$$

с $y_0 \equiv \alpha + z_0$. Поэтому $\{y_t\}$ — случайное блуждание со сносом, если $\delta \neq 0$, и без сноса, если $\delta = 0$. При альтернативе $\rho < 1$ процесс, являющийся суммой линейного тренда и стационарного AR(1)-процесса с нулевым средним, — стационарный относительно тренда.

Устраняя $\{z_t\}$ из (9.3.15), мы получаем:

$$y_t = \alpha^* + \delta^* \cdot t + \rho y_{t-1} + \varepsilon_t, \tag{9.3.17}$$

где

$$\alpha^* = (1-\rho)\alpha + \rho\delta, \ \delta^* = (1-\rho)\delta. \tag{9.3.18}$$

Поскольку $\delta^* = 0$, когда $\rho = 1$, нулевая гипотеза I(1) (о том, что DGP является случайным блужданием с или без сноса) является совместной гипотезой о том, что $\rho = 1$ и $\delta^* = 0$ в терминах регрессионных коэффициентов. На практике, однако, тесты на единичный корень обычно фокусируются на единственном ограничении $\rho = 1$.

Как в модели AR(1) с постоянной составляющей, формулировку утверждения 9.3 можно легко модифицировать для AR(1)-модели с трендом. Пусть $\widehat{\rho}^{\tau}$ является OLS-оценкой для ρ в (9.3.17), а t^{τ} является t-значением для нулевой гипотезы $\rho = 1$. Как вам будет предложено доказать в аналитическом упражнении 2, $T \cdot (\widehat{\rho}^{\tau} - 1)$ сходится по распределению к

$$DF_{\rho}^{\tau} \equiv \frac{\frac{1}{2}\left([W^{\tau}(1)]^2 - [W^{\tau}(0)]^2 - 1\right)}{\int_0^1 [W^{\tau}(r)]^2 \, \mathrm{d}r}, \tag{9.3.19}$$

а t^{τ} сходится по распределению к

$$DF_{t}^{\tau} \equiv \frac{\frac{1}{2}\left([W^{\tau}(1)]^2 - [W^{\tau}(0)]^2 - 1\right)}{\sqrt{\int_0^1 [W^{\tau}(r)]^2 \, \mathrm{d}r}}. \tag{9.3.20}$$

Здесь $W^{\tau}(\cdot)$ является детрендированным стандартным винеровским процессом, введенным в параграфе 9.2. Основная идея доказательства — использовать теорему Фриша — Во, чтобы записать $T \cdot (\widehat{\rho}^{\tau} - 1)$ и t^{τ} с помощью формул, включающих детрендированные ряды, и затем применить утверждение 9.2(e) и (f). Подытожим сказанное.

Утверждение 9.5 (тесты Дики — Фуллера о случайном блуждании с или без сноса): *Предположим, что* $\{y_t\}$ *является случайным блужданием с или без сноса с* $\mathrm{E}(y_0^2) < \infty$. *Рассмотрим регрессию* y_t *на* $(1, t, y_{t-1})$ *для* $t = 1, 2, \ldots, T$. *Тогда*

$$DF\ \rho^\tau\text{-статистика:}\quad T\cdot(\widehat{\rho}^\tau - 1) \underset{d}{\to} DF_\rho^\tau, \tag{9.3.21}$$

$$DF\ t^\tau\text{-статистика:}\quad t^\tau \underset{d}{\to} DF_t^\tau, \tag{9.3.22}$$

где $\widehat{\rho}^\tau$ *— OLS-оценка коэффициента при* y_{t-1}, t^τ *— t-значение для гипотезы о том, что коэффициент при* y_{t-1} *равен 1, а* DF_ρ^τ *и* DF_t^τ *— две случайные величины, определенные в* (9.3.19) *и* (9.3.20).

Критические значения для DF_ρ^τ приведены в табл. 9.1, панель (c), а для DF_ρ^τ — в табл. 9.2, панель (c). В таблицах также приведены критические значения для выборок конечного объема, основанные на предположении о том, что ε_t имеет нормальное распределение (начальное значение y_0, так же как и параметры тренда α и δ, не нужно специфицировать для вычисления распределений на конечных выборках, поскольку они не влияют на численные значения тестовых статистик при нулевой гипотезе $\rho = 1$). Для обеих DF-статистик ρ^τ- и t^τ-распределение даже более сильно скошено влево, чем для случая с константой. Для всех объемов выборок, отраженных в таблицах, 99%-ное критическое значение отрицательно, что означает, что более чем в 99% случаев $\widehat{\rho}^\tau$ меньше единицы, когда фактически истинное значение ρ равно единице! Это иллюстрируется штрихпунктирной кривой на рис. 9.3, которая представляет график распределения в конечных выборках $\widehat{\rho}^\tau$ для $T = 100$.

Дадим два комментария к утверждению 9.5:

- (Численная инвариантность.) Тестовые статистики $\widehat{\rho}^\tau$ и t^τ инвариантны к (α, δ) вне зависимости от значения ρ; прибавление $\alpha + \delta \cdot t$ к временному ряду $\{y_t\}$ просто изменяет OLS-оценки α^* и δ^*, но не $\widehat{\rho}^\tau$ и ее t-значение. Следовательно, распределения $T\cdot(\widehat{\rho}^\tau - 1)$ и t^τ в конечных выборках, так же как и на больших, не будет зависеть от (α, δ) вне зависимости от значения ρ.

- (Нужно ли включать в регрессию время?) Поскольку тестовые статистики инвариантны к значению δ в DGP, утверждение 9.5 применимо, даже когда $\delta = 0$. То есть статистики $\widehat{\rho}^\tau$ и t^τ можно использовать для проверки нулевой гипотезы о случайном блуждании *без тренда*. Однако если вы уверены в том, что процесс при нулевой гипотезе не имеет тренда, то следует использовать DF-тесты $\widehat{\rho}^\mu$ и t^μ в утверждении 9.4 с регрессией, не включающей время, поскольку мощность в конечных выборках против стационарной альтернативы в общем случае больше с $\widehat{\rho}^{\mu}$ и t^{μ}, а не с $\widehat{\rho}^{\tau}$ и t^{τ} (вас попросят проверить это в упражнении 1 на метод Монте-Карло). А если вы думаете, что процесс может иметь тренд, то должны использовать $\widehat{\rho}^{\tau}$ и t^{τ}, с компонентой времени в регрессии. Если вы не включаете время в регрессию и используете критические значения из табл. 9.1, панель (b), и табл. 9.2, панель (b), в то время как DGP действительно имеет тренд, то тест не будет иметь правильного размера на больших выборках (то есть вероятность отвергнуть нулевую гипотезу, когда она верна, не будет приближаться к предварительно специфицированному уровню значимости (номинальному размеру) при неограниченном увеличении объема выборки). Это происходит просто потому, что, когда DGP имеет тренд, предельные распределения $\widehat{\rho}^{\mu}$ и t^{μ} не есть DF_{ρ}^{μ} или DF_{t}^{μ}.

Контрольные вопросы

1. (Суперсостоятельность.) Пусть $\widehat{\rho}$ является OLS-оценкой коэффициента, как в (9.3.2). Проверьте, что $T^{1-\eta}\cdot(\widehat{\rho}-1)\to_{\mathrm{p}} 0$ для $0<\eta<1$, если $\rho=1$.

2. (Предельное распределение при общем I(1)-процессе без сноса.)
В AR(1)-модели (9.3.1) удалите предположение о том, что ошибки являются независимым белым шумом, заменяя ε_t на u_t и предполагая, что $\{u_t\}$ является общим I(0)-процессом с нулевым средним, удовлетворяющим (9.2.1)–(9.2.3). Следовательно, рассматриваемый процесс при нулевой гипотезе $\rho=1$ является общим I(1)-процессом без тренда. Покажите, что $T\cdot(\widehat{\rho}-1)$ сходится по распределению к
$$\frac{\frac{1}{2}\left(W(1)^2-\frac{\gamma_0}{\lambda^2}\right)}{\int_0^1 W(r)^2 dr},$$
где $\lambda^2\equiv$ долговременная дисперсия $\{\Delta y_t\}$ и $\gamma_0\equiv \mathrm{Var}(\Delta y_t)$.

3. (Состоятельность s^2.) Обычная стандартная ошибка OLS-регрессии для AR(1)-уравнения, s, определяется в (9.3.8). Если $|\rho|<1$, то мы знаем из параграфа 6.4, что s^2 является состоятельной оценкой для $\mathrm{Var}\,\varepsilon_t$. Покажите, что при нулевой гипотезе $\rho=1$ s^2 остается состоятельной оценкой для $\mathrm{Var}(\varepsilon_t)$. Поскольку $\Delta y_t=\varepsilon_t$, когда $\rho=1$, s^2 также является состоятельной для $\gamma_0(\equiv \mathrm{Var}(\Delta y_t))$. **Указание:** *Перепишите s^2 в виде:*
$$\begin{aligned} s^2 &\equiv \frac{1}{T-1}\sum_{t=1}^{T}(\Delta y_t-(\widehat{\rho}-1)y_{t-1})^2 = \\ &= \frac{1}{T-1}\sum_{t=1}^{T}(\Delta y_t)^2-\frac{2}{T-1}\cdot[T\cdot(\widehat{\rho}-1)]\cdot\frac{1}{T}\sum_{t=1}^{T}\Delta y_t y_{t-1}+ \\ &+ \frac{1}{T-1}\cdot[T\cdot(\widehat{\rho}-1)]^2\cdot\frac{1}{T^2}\sum_{t=1}^{T}(y_{t-1})^2. \end{aligned}$$
Также $\mathrm{E}[(\Delta y_t)^2]=\gamma_0$. *Используйте результат, который мы использовали постоянно: если $x_T\to_p 0$ и $y_T\to_d y$, то $x_T\cdot y_T\to_p 0$.*

4. (Два теста на случайное блуждание.) Рассмотрим две регрессии. Одна является регрессией Δy_t на Δy_{t-1}, а другая — регрессией Δy_t на y_{t-1}. Чтобы протестировать нулевую гипотезу о том, что $\{y_t\}$ является случайным блужданием без сноса, какое распределение следует использовать: t-значения из первой регрессии или t-значения из второй регрессии?

5. (Состоятельность DF-тестов.) Напомним, что тест называется **состоятельным** (*consistent*) против некоторого множества альтернатив, если вероятность отвержения неверной нулевой гипотезы, когда истинный DGP является одной из альтернатив приближается к единице при возрастании объема выборки. Пусть $T \cdot (\widehat{\rho} - 1)$ определяется как в (9.3.3).

   (a) Предположим, что $\{y_t\}$ является I(0)-процессом с нулевым средним и что его коэффициент автокорреляции первого порядка меньше 1 (так что $\mathrm{E}(y_t^2) > \mathrm{E}(y_t y_{t-1})$). Покажите, что
   $$\operatorname*{plim}_{T\to\infty} T \cdot (\widehat{\rho} - 1) = -\infty.$$
   Таким образом, DF ρ-тест является состоятельным против общих I(0)-альтернатив. **Указание:** *Покажите, что если $\{y_t\}$ является I(0)-процессом, то*
   $$\operatorname*{plim}_{T\to\infty} \widehat{\rho} = \frac{\mathrm{E}(y_t y_{t-1})}{\mathrm{E}(y_{t-1}^2)}.$$

   (b) Покажите, что DF t-тест является состоятельным против общих I(0)-альтернатив с нулевым средним. **Указание:** *Когда $\{y_t\}$ является I(0)-процессом с нулевым средним, s сходится по вероятности к некоторому положительному числу.*

6. (Влияние y_0 на тестовые статистики.) Проверьте, что, когда $\rho = 1$, значение y_0 не влияет на значения $\widehat{\rho}^{\mu}$ и t^{μ} на конечных выборках. Остается ли это верным, когда $\rho < 1$? [Ответ: Нет. Поскольку влияние изменения y_0 на y_t зависит от t.] Зависит ли мощность на конечных выборках против альтернативы $\rho < 1$ от y_0? [Ответ: Да.]

7. (Статистика Саргана — Бхаргавы [Sargan and Bhargava, 1983.] Использование тестовой статистики, основанной на OLS-оценке коэффициента — не единственный способ получения теста на единичный корень. Для выборки (y_0, $y_1, \dots, y_T$) **статистика Саргана — Бхаргавы** (*Sargan — Bhargava statstic*) определяется как
   $$SB = \frac{\frac{1}{T^2}\sum_{t=0}^{T}(y_t)^2}{\frac{1}{T}\sum_{t=1}^{T}(\Delta y_t)^2}.$$

   (a) Проверьте, что она является обратной к статистике Дарбина — Уотсона.

   (b) Покажите, что если $\{y_t\}$ является случайным блужданием без тренда, то
   $$SB \underset{d}{\to} \int_0^1 [W(r)]^2 \mathrm{d}r.$$
   Указание: $\sum_{t=1}^{T} y_t^2 = \sum_{t=1}^{T} y_{t-1}^2 + y_T^2$. *Также* $y_T^2/T^2 \to_p 0$.

(c) Проверьте, что при I(0)-альтернативе с $\mathrm{E}[(\Delta y_t)^2] \neq 0$ $SB \to_{\mathrm{p}} 0$. (Поэтому этот тест отвергает нулевую гипотезу I(1) при *малых* значениях SB.)

9.4. Расширенные тесты Дики — Фуллера

Процесс I(1) при нулевой гипотезе в предыдущем параграфе был случайным блужданием, которое не допускает сериальной корреляции в первых разностях. Тесты этого параграфа контролируют сериальную корреляцию путем добавления в авторегрессионное уравнение лагированных первых разностей. Они восходят к [Dickey and Fuller, 1979] и стали известны как **расширенные тесты Дики — Фуллера (ADF)** (*augmented Dickey — Fuller tests*).

Расширенная авторегрессия

Нулевая гипотеза I(1) в самом простом ADF-тесте состоит в том, что $\{y_t\}$ является ARIMA$(p, 1, 0)$-процессом, то есть $\{\Delta y_t\}$ является стационарным AR(p)-процессом с нулевым средним:

$$\Delta y_t = \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \cdots + \zeta_p \Delta y_{t-p} + \varepsilon_t, \quad \mathrm{E}(\varepsilon_t^2) \equiv \sigma^2, \tag{9.4.1}$$

где $\{\varepsilon_t\}$ является независимым белым шумом и все корни соответствующего полиномиального уравнения, $1 - \zeta_1 z - \zeta_2 z^2 - \cdots - \zeta_p z^p$, лежат за пределами единичного круга. (Позднее мы будем допускать, что $\{\Delta y_t\}$ является стационарным ARMA(p, q).) Таким образом, процесс в первых разностях $\{\Delta y_t\}$ является I(0)-процессом, поскольку его можно записать как $\Delta y_t = \psi(L)\varepsilon_t$, где

$$\psi(L) = (1 - \zeta_1 L - \zeta_2 L^2 - \cdots - \zeta_p L^p)^{-1}. \tag{9.4.2}$$

Модель, использованная для получения тестовых статистик, включает этот процесс в качестве частного случая:

$$y_t = \rho y_{t-1} + \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \cdots + \zeta_p \Delta y_{t-p} + \varepsilon_t. \tag{9.4.3}$$

Действительно, когда $\rho = 1$, (9.4.3) редуцируется к (9.4.1).

Уравнение (9.4.3) можно также записать как AR$(p+1)$:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_{p+1} y_{t-p-1} + \varepsilon_t. \tag{9.4.4}$$

где значения ϕ_j связаны с $(\rho, \zeta_1, \ldots, \zeta_p)$ соотношениями

$$\begin{aligned} &\rho \equiv \phi_1 + \phi_2 + \cdots + \phi_{p+1}, \\ &\zeta_j \equiv -(\phi_{j+1} + \phi_{j+2} + \cdots + \phi_{p+1}) \quad (j = 1, 2, \ldots, p). \end{aligned} \tag{9.4.5}$$

Обратно, любую авторегрессию $(p+1)$-го порядка можно записать в виде (9.4.3). Эта форма (9.4.3) была изначально предложена в [Fuller, 1976, p. 374]. Она называется **расширенной авторегрессией** (*augmented autoregression*), поскольку ее можно получить путем добавления лагированных приращений в правую часть $\mathrm{AR}(1)$-уравнения в уровнях (9.3.1).

Легко показать (контрольный вопрос 2), что полиномиальное уравнение, связанное с (9.4.4)), имеет корень внутри единичного круга, если $\rho > 1$. Таким образом, если мы принимаем условие, что DGP является или I(1)-, или I(0)-процессом, то ρ не может быть больше 1, и I(0)-альтернатива состоит в том, что $\rho < 1$.

Предельное распределение OLS-оценки

Создав модель, мы обращаемся теперь к задаче получения предельного распределения OLS-оценки для ρ в расширенной авторегрессии (9.4.3) при нулевой гипотезе I(1). Численно ту же самую оценку ρ можно получить из $\mathrm{AR}(p+1)$ уравнения (9.4.4)), вычисляя сумму OLS-оценок $(\phi_1, \phi_2, \dots, \phi_{p+1})$ (вспомним, что $\rho = \phi_1 + \dots + \phi_{p+1}$). Однако для цели получения предельного распределения расширенная регрессия более удобна. При нулевой гипотезе I(1) все $p+1$ регрессоров в (9.4.4)) являются I(1)-процессами без сноса, но это маскирует тот факт, который используется ниже, и из (9.4.3) ясно, что p их линейных комбинаций являются I(0) с нулевыми средними.

Для упрощения вычислений рассмотрим частный случай $p = 1$ (одно лагированное приращение в y_t, или AR(2)), так что расширенная авторегрессия имеет вид:

$$y_t = \rho y_{t-1} + \zeta_1 \Delta y_{t-1} + \varepsilon_t. \tag{9.4.6}$$

Мы предполагаем, что выборка есть $(y_{-1}, y_0, \dots, y_T)$, так что расширенную авторегрессию можно оценить для $t = 1, 2, \dots, T$ (или альтернативно выборка есть $(y_1, y_2, \dots, y_T)$, но суммирование в дальнейшем происходит с $t = 3$ по $t = T$, а не с $t = 1$ по $t = T$). Запишем это в виде:

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + \varepsilon_t$$

с

$$\underset{(2\times 1)}{\boldsymbol{x}_t} = \begin{bmatrix} y_{t-1} \\ \Delta y_{t-1} \end{bmatrix}, \qquad \underset{(2\times 1)}{\boldsymbol{\beta}} = \begin{bmatrix} \rho \\ \zeta_1 \end{bmatrix}. \tag{9.4.7}$$

Если $\boldsymbol{b} \equiv (\widehat{\rho}, \widehat{\zeta}_1)'$ является OLS-оценкой коэффициентов $\boldsymbol{\beta} (\equiv (\rho, \zeta_1)')$ расширенной авторегрессии, то ошибка оценки задается соотношением:

$$\boldsymbol{b} - \boldsymbol{\beta} = \Big(\sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t'\Big)^{-1} \sum_{t=1}^{T} \boldsymbol{x}_t \cdot \varepsilon_t. \tag{9.4.8}$$

Отдельные компоненты в (9.4.8) задаются посредством соотношения:

$$\underset{(2\times 2)}{X'X} \equiv \sum_{t=1}^{T} x_t x_t' = \begin{bmatrix} \sum_{t=1}^{T}(y_{t-1})^2 & \sum_{t=1}^{T} y_{t-1}\,\Delta y_{t-1} \\ \sum_{t=1}^{T} \Delta y_{t-1}\, y_{t-1} & \sum_{t=1}^{T}(\Delta y_{t-1})^2 \end{bmatrix},$$

$$\underset{(2\times 1)}{\sum_{t=1}^{T} x_t \cdot \varepsilon_t} = \begin{bmatrix} \sum_{t=1}^{T} y_{t-1}\,\varepsilon_t \\ \sum_{t=1}^{T} \Delta y_{t-1}\,\varepsilon_t \end{bmatrix}. \tag{9.4.9}$$

Мы можем применить теперь тот же самый трюк, который использовали в регрессии на время в параграфе 2.12. Для некоторой невырожденной матрицы Υ_T, которая специфицируется ниже, перепишем (9.4.8) в виде:

$$\Upsilon_T \cdot (b - \beta) = \Upsilon_T \begin{bmatrix} \widehat{\rho} - \rho \\ \widehat{\zeta}_1 - \zeta_1 \end{bmatrix} = A_T^{-1} c_T. \tag{9.4.10}$$

где

$$A_T \equiv \Upsilon_T^{-1} \left[\sum_{t=1}^{T} x_t\, x_t'\right] \Upsilon_T^{-1}, \quad c_T \equiv \Upsilon_T^{-1} \sum_{t=1}^{T} x_t \cdot \varepsilon_t. \tag{9.4.11}$$

Мы ищем такую матрицу Υ_T, чтобы $\Upsilon_T \cdot (b - \beta)$ имело невырожденное предельное распределение при нулевой гипотезе I(1). Это достигается, если положить:

$$\Upsilon_T = \begin{bmatrix} T & 0 \\ 0 & \sqrt{T} \end{bmatrix}. \tag{9.4.12}$$

При таком выборе шкалирующей матрицы Υ_T из (9.4.9) и (9.4.11) следует, что

$$A_T = \begin{bmatrix} \frac{1}{T^2} \sum_{t=1}^{T}(y_{t-1})^2 & \frac{1}{\sqrt{T}}\frac{1}{T} \sum_{t=1}^{T} y_{t-1}\,\Delta y_{t-1} \\ \frac{1}{\sqrt{T}}\frac{1}{T} \sum_{t=1}^{T} \Delta y_{t-1}\, y_{t-1} & \frac{1}{T} \sum_{t=1}^{T}(\Delta y_{t-1})^2 \end{bmatrix},$$

$$c_T = \begin{bmatrix} \frac{1}{T} \sum_{t=1}^{T} y_{t-1}\,\varepsilon_t \\ \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \Delta y_{t-1}\,\varepsilon_t \end{bmatrix}. \tag{9.4.13}$$

Рассмотрим элементы A_T и c_T и найдем их предельные распределения при нулевой гипотезе $\rho = 1$.

(1.1)**-й элемент** A_T**.** Так как $\{y_t\}$ является I(1)-процессом без сноса, этот элемент сходится по распределению к $\lambda^2 \cdot \int W^2 \mathrm{d}r$ по утверждению 9.2(a), где $\lambda^2 \equiv$ долговременная дисперсия $\{\Delta y_t\}$.

(2.2)**-й элемент** A_T**.** Он сходится по вероятности (и, следовательно, по распределению) к $\gamma_0 (\equiv \mathrm{Var}(\Delta y_t))$, поскольку $\{\Delta y_t\}$, будучи стационарным AR(1), является эргодически стационарным процессом с нулевым средним.

Внедиагональные элементы A_T. Они равны произведению $1/\sqrt{T}$ на

$$\frac{1}{T}\sum_{t=1}^{T}\Delta y_{t-1}\, y_{t-1}, \qquad (9.4.14)$$

что является средним произведения I(0)-переменной с нулевым средним на I(1)-переменную без сноса. По предположению 9.2(b), похожее произведение $(1/T)\sum_{t=1}^{T}\Delta y_t y_{t-1}$ сходится по распределению к некоторой случайной величине. Легко показать (контрольный вопрос 2), что (9.4.14) также сходится к случайной величине. Поэтому умноженное на $1/\sqrt{T}$ (9.4.14), являющееся внедиагональным элементом A_T, сходится по вероятности (а следовательно, по распределению) к 0.

Следовательно, A_T, надлежащим образом перешкалированная матрица $X'X$, является асимптотически диагональной:

$$A_T \underset{d}{\to} \begin{bmatrix} \lambda^2\cdot\int_0^1 W(r)^2\,\mathrm{d}r & 0 \\ 0 & \gamma_0 \end{bmatrix},$$

так что

$$(A_T)^{-1} \underset{d}{\to} \begin{bmatrix} \left(\lambda^2\cdot\int_0^1 W(r)^2\,\mathrm{d}r\right)^{-1} & 0 \\ 0 & \gamma_0^{-1} \end{bmatrix}. \qquad (9.4.15)$$

Мы переходим теперь к элементам c_T.

Первый элемент c_T. Используя разложение Бевериджа — Нельсона, можно показать (аналитическое упражнение 4), что

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\,\varepsilon_t \underset{d}{\to} c_1 \equiv \tfrac{1}{2}\sigma^2\cdot\psi(1)\cdot[W(1)^2-1], \qquad (9.4.16)$$

для общего I(0)-процесса с нулевым средним

$$\Delta y_t = \psi(L)\varepsilon_t. \qquad (9.4.17)$$

В данном случае при нулевой гипотезе $\Delta y_t = \zeta_1\Delta y_{t-1} + \varepsilon_t$, так что $\psi(L) = (1-\zeta_1 L)^{-1}$ и

$$\psi(1) = \frac{1}{1-\zeta_1}. \qquad (9.4.18)$$

Второй элемент c_T. Легко показать (контрольный вопрос 4), что $\{\Delta y_{t-1}\varepsilon_t\}$ является стационарной и эргодической мартингал-разностью, поэтому

$$\text{второй элемент } c_T = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Delta y_{t-1}\,\varepsilon_t \underset{d}{\to} c_2 \sim N(0, \gamma_0\cdot\sigma^2), \qquad (9.4.19)$$

где $\gamma_0 \equiv \mathrm{Var}(\Delta y_t)$ и $\sigma^2 \equiv \mathrm{Var}(\varepsilon_t)$.

Подставляя (9.4.15)–(9.4.19) в (9.4.10)) и замечая, что истинное значение ρ равно единице при нулевой гипотезе I(1), мы получаем, что для OLS-оценки расширенной авторегрессии[1]

$$\boldsymbol{\Upsilon}_T \cdot (\boldsymbol{b}-\boldsymbol{\beta}) = \begin{bmatrix} T\cdot(\widehat{\rho}-1) \\ \sqrt{T}(\widehat{\zeta}_1-\zeta_1)\end{bmatrix} \underset{\mathrm{d}}{\to} \begin{bmatrix} \left(\lambda^2 \cdot \int_0^1 W(r)^2\,\mathrm{d}r\right)^{-1} & 0 \\ 0 & \gamma_0^{-1}\end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} =$$

$$= \begin{bmatrix} \left(\lambda^2 \cdot \int_0^1 W(r)^2\,\mathrm{d}r\right)^{-1} c_1 \\ \gamma_0^{-1} c_2 \end{bmatrix}.$$

Таким образом,

$$T\cdot(\widehat{\rho}-1) \underset{\mathrm{d}}{\to} \frac{\sigma^2\psi(1)}{\lambda^2}\cdot\frac{\frac{1}{2}[W(1)^2-1]}{\int_0^1 W(r)^2\,\mathrm{d}r} \quad \text{или}$$

$$\frac{\lambda^2}{\sigma^2\psi(1)}\cdot T\cdot(\widehat{\rho}-1) \underset{\mathrm{d}}{\to} DF_\rho. \tag{9.4.20}$$

$$\sqrt{T}\cdot(\widehat{\zeta}_1-\zeta_1) \underset{\mathrm{d}}{\to} N\left(0,\frac{\sigma^2}{\gamma_0}\right), \tag{9.4.21}$$

где DF_ρ не что иное, как DF ρ-распределение, определенное в (9.3.6). Заметим, что оцененный коэффициент при I(0)-переменной с нулевым средним (Δy_{t-1}) имеет обычную $\sqrt{T}$ асимптотику, в то время как оцененный коэффициент при I(1)-переменной без сноса (y_{t-1}) имеет нестандартное предельное распределение.

Получение тестовых статистик

В предельное распределение статистики $\frac{\lambda^2}{\sigma^2\psi(1)}\cdot T\cdot(\widehat{\rho}-1)$ в (9.4.20) мешающие параметры не входят, но эта статистика сама по себе включает мешающие параметры через $\frac{\lambda^2}{\sigma^2\psi(1)}$. По формуле (9.2.4) для долговременной дисперсии λ^2 в $\{\Delta y_t\}$, это отношение равно $\psi(1)$, которое, в силу (9.4.18), оказывается равным $\frac{1}{1-\zeta_1}$ при нулевой гипотезе. Вернемся теперь к расширенной авторегрессии (9.4.6). Поскольку $\widehat{\zeta}_1$, OLS-оценка для ζ_1, состоятельна в силу (9.4.21), состоятельная оценка для $\frac{1}{1-\zeta_1}$ равна $(1-\widehat{\zeta}_1)^{-1}$. Тогда из (9.4.20) вытекает, что

$$\frac{T\cdot(\widehat{\rho}-1)}{1-\widehat{\zeta}_1} \underset{\mathrm{d}}{\to} DF_\rho. \tag{9.4.22}$$

[1] Здесь мы используем лемму 2.3(b) для утверждения о том, что $\boldsymbol{A}_T^{-1}\boldsymbol{c}_T$ сходится по распределению к $\boldsymbol{A}^{-1}\boldsymbol{c}$, где $\boldsymbol{A}$ и $\boldsymbol{c}$ являются предельными случайными величинами для $\boldsymbol{A}_T$ и $\boldsymbol{c}_T$ соответственно. Заметим, что в отличие от ситуаций, которые встречаются в стационарном случае, сходимость $\boldsymbol{A}_T$ к $\boldsymbol{A}$ есть сходимость по распределению, но не по вероятности. Мы доказали в учебнике, что $\boldsymbol{A}_T$ сходятся по распределению к $\boldsymbol{A}$, а $\boldsymbol{c}_T$ к $\boldsymbol{c}$, но мы не показывали, что $\boldsymbol{A}_T, \boldsymbol{c}_T$ сходится по распределению совместно к $\boldsymbol{A}, \boldsymbol{c}$, что нужно для использования Леммы 2.3(b). Однако, например, по Theorem 2.2 в: [Chan and Wei, 1988] сходимость является совместной.

То есть требуемая коррекция для $T \cdot (\widehat{\rho} - 1)$ выполняется через оценивание коэффициента при лагированном Δy в расширенной авторегрессии. После проведения такой коррекции предельное распределение статистики не зависит от мешающих параметров, контролирующих сериальную корреляцию $\{\Delta y_t\}$, таких как λ^2 и γ_0.

OLS t-значение для гипотезы $\rho = 1$ также имеет предельное распределение. Его можно записать как

$$t = \frac{\widehat{\rho} - 1}{s \cdot \sqrt{(1,1) \text{ элемент матрицы } \left(\sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t'\right)^{-1}}} = \\ = \frac{T \cdot (\widehat{\rho} - 1)}{s \cdot \sqrt{(1,1) \text{ элемент матрицы } (\boldsymbol{A}_T)^{-1}}}, \tag{9.4.23}$$

где s является обычной стандартной ошибкой OLS-регрессии. Из (9.4.15), (9.4.20), состоятельности s^2 для σ^2 и утверждения 9.2(a) и (b) для $\xi_t = y_t$, легко следует, что

$$t \underset{d}{\rightarrow} DF_t, \tag{9.4.24}$$

где DF_t является DF t-распределением, определенным в (9.3.9). Следовательно, для t-значения на самом деле нет необходимости производить коррекцию того факта, что лагированные Δy были включены в расширенную авторегрессию.

Все эти результаты для $p = 1$ можно легко обобщить:

Утверждение 9.6 (расширенный тест Дики — Фуллера на единичный корень без константы): *Предположим, что $\{y_t\}$ — ARIMA$(p, 1, 0)$-процесс, так что $\{\Delta y_t\}$ является стационарным AR(p)-процессом, следующим (9.4.1). Рассмотрим оценивание расширенной авторегрессии, (9.4.3), и пусть $(\widehat{\rho}, \widehat{\zeta}_1, \widehat{\zeta}_2, \dots, \widehat{\zeta}_p)$ являются OLS-оценками коэффициентов. Пусть также t является t-значением для гипотезы $\rho = 1$. Тогда*

$$\text{ADF } \rho\text{-статистика: } \frac{T \cdot (\widehat{\rho} - 1)}{1 - \widehat{\zeta}_1 - \widehat{\zeta}_2 - \dots - \widehat{\zeta}_p} \underset{d}{\rightarrow} DF_\rho, \tag{9.4.25}$$

$$\text{ADF } t\text{-статистика: } t \underset{d}{\rightarrow} DF_t, \tag{9.4.26}$$

где DF_ρ и DF_t являются двумя случайными величинами, определенными в (9.3.6) и (9.3.9).

Тестирование гипотез о ζ

Доказательство, приводящее к этому удивительно простому результату, является очень длинным, но оно имеет побочный результат. Очевидным обобщением (9.4.21) на авторегрессию p-го порядка является следующее:

$$\sqrt{T}\cdot\begin{bmatrix}\widehat{\zeta}_1-\zeta_1\\ \widehat{\zeta}_2-\zeta_2\\ \vdots\\ \widehat{\zeta}_p-\zeta_p\end{bmatrix}\underset{d}{\to} N\left(\begin{bmatrix}0\\0\\ \vdots\\0\end{bmatrix},\sigma^2\cdot\begin{bmatrix}\gamma_0 & \gamma_1 & \cdots & \gamma_{p-1}\\ \gamma_1 & \gamma_0 & \cdots & \gamma_{p-2}\\ \vdots & \vdots & \cdots & \vdots\\ \gamma_{p-1} & \gamma_{p-2} & \cdots & \gamma_0\end{bmatrix}^{-1}\right). \qquad (9.4.27)$$

где γ_j является автоковариацией $\{\Delta y_t\}$ j-го порядка. По утверждению 6.7 это то же самое асимптотическое распределение, которое вы бы получили, если бы расширенная авторегрессия (9.4.3) оценивалась посредством OLS с ограничением $\rho = 1$, то есть если (9.4.1) оценивается посредством OLS. Легко показать (контрольный вопрос 6), что гипотезы, затрагивающие только коэффициенты при I(0)-регрессорах с нулевым средним, $(\zeta_1, \zeta_2, \ldots, \zeta_p)$, можно тестировать обычным образом с обычными асимптотически оправданными t и F-статистиками.

Для получения предельных распределений регрессионных коэффициентов мы использовали тот факт, что один из регрессоров является I(0)-процессом с нулевым средним, тогда как другой является I(1)-процессом без сноса. Систематическое исследование более общих случаев можно найти в: [Sims, Stock and Watson, 1990] и [Watson, 1994, Section 2].

Что делать, когда p неизвестно

Утверждение 9.6 предполагает, что порядок авторегрессии p для Δy_t известен. Что мы можем сделать, когда порядок p неизвестен? Чтобы провести различие между истинным порядком p и количеством лагированных приращений, включенных в расширенную регрессию, мы в этом параграфе обозначаем последнее как $\widehat{p}$. Мы рассмотрели аналогичную проблему в параграфе 6.4. Разница состоит в том, что здесь DGP является стационарным процессом в первых разностях, а не в уровнях, как в параграфе 6.4, и что оцениваемое уравнение является расширенной авторегрессией со свободно оцениваемым ρ. Мы демонстрируем ниже без доказательства три результата для больших выборок относительно выбора $\widehat{p}$, при которых остается справедливым утверждение 9.6. Эти результаты применимы к более общему, чем предполагается в утверждении 9.6, классу процессов: $\{\Delta y_t\}$ является стационарным и обратимым

ARMA(p, q) с нулевым средним неизвестного порядка (хотя и с дополнительным предположением конечности четвертого момента ε_t)[1].

Таким образом, при записи процесса Δy_t в форме авторегрессии, порядок авторегрессии может быть бесконечным, как тогда, когда $q > 0$, или конечным, как тогда, когда $q = 0$.

Первый результат состоит в том, что заключение утверждения 9.6 продолжает выполняться, когда $\widehat{p}$, количество лагированных первых разностей в расширенной авторегрессии, увеличивается с объемом выборки с соответствующей скоростью.

(1) ([Said and Dickey, 1984], обобщение утверждения 9.6.) Предположим, что $\widehat{p}$ удовлетворяет

$$\widehat{p} \to \infty, \quad \text{но} \quad \frac{\widehat{p}}{T^{1/3}} \to 0 \quad \text{при} \quad T \to \infty. \tag{9.4.28}$$

(То есть $\widehat{p}$ стремится к бесконечности, но с более низкой скоростью, чем $T^{1/3}$.) Тогда две статистики, ADF ρ и ADF t, основанные на расширенной авторегрессии с $\widehat{p}$ лагированными изменениями, имеют те же самые предельные распределения, что и указанные в утверждении 9.6.

Этот результат, однако, не дает нам практического ориентира для выбора глубины запаздываний, поскольку существует бесконечно много правил для выбора $\widehat{p}$, удовлетворяющих (9.4.28). Естественный вопрос: может ли какое-либо из правил для выбора глубины запаздываний, рассмотренных в параграфе 6.4, последовательное t-правило от общего к частному или правила, основанные на информационных критериях, использоваться в данном контексте? Для освежения вашей памяти: правило, основанное на информационном критерии, полагает $\widehat{p}$ равным значению j, которое минимизирует

$$\log\left(\frac{SSR_j}{T}\right) + (j+1) \times \frac{C(T)}{T}, \tag{9.4.29}$$

где SSR_j является суммой квадратов остатков в модели авторегрессии с j лагами:

$$y_t = \rho y_{t-1} + \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \cdots + \zeta_j \Delta y_{t-j} + \varepsilon_t. \tag{9.4.30}$$

Компонента $C(T)/T$ умножается на $j+1$, поскольку уравнение имеет $j+1$ коэффициентов, включая коэффициент при лагированном y_t.

[1] См. параграф 6.2 относительно определения обратимых ARMA(p, q)-процессов. Если ARMA(p, q)-процесс является обратимым, то его можно записать в виде процесса авторегрессии бесконечного порядка.

Для информационного критерия Акаике (AIC) $C(T) = 2$, в то время как для байесовского информационного критерия (BIC), также называемого информационным критерием Шварца (SIC), $C(T) = \log(T)$. Для любого из этих правил $\widehat{p}$ выбирается из множества $j = 0, 1, 2, \ldots, p_{\max}$. В параграфе 6.4 эта верхняя граница $p_{\max}$ была положена равной некоторому известному целому числу, большему или равному истинному порядку процесса авторегрессии конечного порядка. Указанное $\widehat{p}$, выбранное одним из этих правил, является функцией от данных (не только функцией от T) и поэтому случайная величина. Конечно, когда $\{\Delta y_t\}$ является стационарным и обратимым $\mathrm{ARMA}(p, q)$-процессом, порядок модели авторегрессии бесконечен при $q > 0$, и верхняя граница очевидно не может быть положена равной истинному порядку. Однако ее можно сделать возрастающей вместе с объемом выборки T. Чтобы подчеркнуть зависимость $p_{\max}$ от T, запишем это как $p_{\max}(T)$. Если указанные правила модифицировать таким образом, то остается ли верным обобщение Said — Dickey утверждения 9.6, если $\widehat{p}$ выбирается одним из этих правил? Ответ, полученный в [Ng and Perron, 1995], утвердительный. Более конкретно:

(2) Предположим, что $\widehat{p}$ выбирается t-правилом от общего к частному с $p_{\max}(T)$, удовлетворяющим условию (9.4.28) (которое в обобщении Said — Dickey, приведенном выше, было условием для $\widehat{p}$, а не для $p_{\max}(T)$), и $p_{\max}(T) > c \cdot T^g$ для некоторых $c > 0$ и $0 < g < 1/3$. Тогда предельные распределения ADF ρ- и ADF t-статистик те же, что указаны в утверждении 9.6.[1]

(3) Предположим, что $\widehat{p}$ выбирается посредством AIC или BIC с $p_{\max}(T)$, удовлетворяющим условию (9.4.28). Тогда сохраняется то же самое заключение[2].

Таким образом, не имеет значения, какое правило, последовательное t, AIC или BIC, вы используете для выбора $\widehat{p}$ — распределения ADF-статистик на больших выборках те же самые, как и в утверждении 9.6.

Предложение для выбора $p_{\max}(T)$

Распределение в конечных выборках зависит, однако, от того, какое правило вы используете, а также от выбора верхней границы $p_{\max}(T)$.

[1] Это Theorem 5.3 в [Ng and Perron, 1995]. Их условие A1 есть (9.4.28). Их условие A2″ вытекает из второго условия, указанного здесь.

[2] Это Theorem 4.3 в [Ng and Perron, 1995]. Теорема не указывает верхнюю границу $p_{\max}(T)$, но (как указал Пьер Перрон в частной беседе с автором) из работы [Hannan and Deistler, 1988, Section 7.4 (ii)] вытекает, что верхней границей может быть (9.4.28).

Например, $p_{\max}(T) = [T^{1/4}]$ (целая часть $T^{1/4}$) удовлетворяет условиям в результате (2) для последовательного t-теста. То же относится и к $p_{\max}(T) = [100T^{3/10}]$. Важно поэтому, чтобы исследователи использовали одинаковое $p_{\max}(T)$ при решении вопроса о порядке расширенной авторегрессии. В этом контексте актуальной является литература по методу Монте-Карло, изучающая свойства различных тестов на единичный корень в конечных выборках. Симуляции, проведенные в [Schwert, 1989], показывают, что для малых и умеренно больших выборок (от $T = 25$ до $T = 1000$) включение достаточного количества лагов в авторегрессию является важным для минимизации искажений размера теста[1]. Выбор $\widehat{p}$ (количества лагов, включенных в авторегрессию), который был более или менее успешным в контролировании фактического размера в исследовании Шверта, равен $\left[12 \cdot \left(\frac{T}{100}\right)^{1/4}\right]$. Верхняя граница $p_{\max}(T)$ должна, следовательно, давать правилам выбора лага возможность выбора значения $\widehat{p}$ столь же большим. Поэтому в примерах и приложении в этой главе в любом из правил для выбора глубины запаздываний мы используем функцию:

$$p_{max}(T) = \left[12 \cdot \left(\frac{T}{100}\right)^{1/4}\right] \quad \text{(целая часть } 12 \cdot \left(\frac{T}{100}\right)^{1/4}\text{)}. \tag{9.4.31}$$

Эта функция удовлетворяет условиям результатов (2) и (3), приведенных выше.

Временной интервал выборки при выборе $\widehat{p}$ включает моменты $t = p_{\max}(T) + 2, p_{\max}(T) + 3, \ldots, T$. Первое t равно $p_{\max}(T) + 2$, поскольку $p_{\max}(T) + 1$ наблюдений необходимы для вычисления $p_{\max}(T)$ лагированных приращений в расширенной регрессии. Поскольку используется только $T - p_{\max}(T) - 1$ наблюдений для оценивания авторегрессии (9.4.30) для $j = 1, 2, \ldots, p_{\max}(T)$, объем выборки в целевой функции в (9.4.29) должен быть равным $T - p_{\max}(T) - 1$, а не T:

$$\log\left(\frac{SSR_j}{T - p_{max}(T) - 1}\right) + (j+1) \times \frac{C(T - p_{max}(T) - 1)}{T - p_{max}(T) - 1}. \tag{9.4.32}$$

Включение в регрессию постоянной составляющей

Как и в DF-тестах, мы можем модифицировать ADF-тесты так, чтобы они были инвариантными к прибавлению константы к ряду. Мы допускаем, чтобы $\{y_t\}$ отличался на неспецифицированную константу от процесса,

[1] Напомним, что искажением размера называют разность между фактическим, или точным, размером (вероятностью отвержения верной нулевой гипотезы в конечных выборках) и номинальным размером (вероятностью отвергнуть верную нулевую гипотезу, когда объем выборки бесконечен).

который следует расширенной авторегрессии без константы (9.4.3). Это сводится к замене y_t в (9.4.3) на $y_t - \alpha$, что приводит к

$$y_t = \alpha^* + \rho y_{t-1} + \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \cdots + \zeta_p \Delta y_{t-p} + \varepsilon_t, \tag{9.4.33}$$

где $\alpha^* = (1-\rho)\alpha$. Как и в AR(1)-модели без константы, нулевая гипотеза I(1) является совместной гипотезой о том, что $\rho = 1$ и $\alpha^* = 0$. Однако тесты на единичный корень обычно фокусируются на единственном ограничении $\rho = 1$.

В качестве аналитического упражнения оставляем доказать следующее:

Утверждение 9.7 (расширенный тест Дики — Фуллера на единичный корень с постоянной составляющей): *Предположим, что $\{y_t\}$ является ARIMA$(p,1,0)$-процессом, так что $\{\Delta y_t\}$ — стационарный AR(p)-процесс с нулевым средним, следующий (9.4.1). Рассмотрим оценивание расширенной авторегрессии с постоянной составляющей, (9.4.33), и пусть $(\widehat{\alpha}, \widehat{\rho}^\mu, \widehat{\zeta}_1, \widehat{\zeta}_2, \ldots, \widehat{\zeta}_p)$ являются OLS-оценками коэффициентов. Также пусть t^μ будет t-значением для гипотезы $\rho = 1$. Тогда*

$$\text{ADF } \rho^\mu\text{-статистика:} \quad \frac{T\cdot(\widehat{\rho}^\mu - 1)}{1 - \widehat{\zeta}_1 - \widehat{\zeta}_2 - \cdots - \widehat{\zeta}_p} \underset{d}{\to} DF_\rho^\mu, \tag{9.4.34}$$

$$\text{ADF } t^\mu\text{-статистика:} \quad t^\mu \underset{d}{\to} DF_t^\mu. \tag{9.4.35}$$

где случайные величины DF_ρ^μ и DF_t^μ определены в (9.3.13) и (9.3.14).

Основная идея доказательства заключается в использовании теоремы Фриша — Во, чтобы записать $\widehat{\rho}^\mu$ и t-значение посредством формул, содержащих центрированные ряды, созданные из $\{y_t\}$.

Прежде чем перейти к примеру, дадим два комментария по поводу утверждения 9.7.

- (Численная инвариантность.) Поскольку регрессия включает постоянную составляющую, тестовые статистики инвариантны к прибавлению константы, как следует из утверждения 9.4.

- (Обобщение Said — Dickey.) Обобщение Said — Dickey — Ng — Perron также применимо здесь: если $\{\Delta y_t\}$ является стационарным и обратимым ARMA(p,q)-процессом (так что Δy_t можно записать в форме авторегрессии возможно бесконечного порядка), то ADF ρ^μ- и t^μ-статистики имеют те же самые предельные распределения, которые указаны в утверждении 9.7, при условии, что $\widehat{p}$ (число лагированных приращений, включенных в авторегрессию) устанавливается либо последовательным t-правилом, либо AIC, либо BIC[1].

Пример 9.2 (ADF на ставку по казначейским обязательствам): В эмпирическом упражнении в главе 2 мы вскользь отметили, что номинальная процентная ставка R_t может иметь тренд для периода выборки, рассмотренного в [Fama, 1975]. Здесь мы проверяем, имеет ли этот процесс стохастический тренд (то есть является ли он процессом I(1) без сноса). Набор данных, который мы использовали в эмпирическом приложении в главе 2, состоял из месячных наблюдений одномесячной ставки по казначейским обязательствам США. Мы берем период выборки с января 1953 г. по июль 1971 г. ($T = 223$ наблюдения), период выборки в исследовании [Fama, 1975]. В процентной ставке временной тренд не проявляется. Поэтому мы включаем в авторегрессию постоянную составляющую, но не включаем время.

Максимальная глубина запаздываний $p_{\max}(T)$ полагается равной 14 согласно функции (9.4.31). В процессе выбора фактической глубины запаздываний $\widehat{p}$ мы фиксируем период выборки с $t = p_{\max} + 2 = 16$ (апрель 1954 г.) по $t = 223$ (июль 1971 г.) с объемом выборки, равным 208 ($= T - p_{\max}(T) - 1$). Последовательное t-правило выбирает $\widehat{p}$ равным 14. Целевой функцией правила, основанного на информационных критериях, когда объем выборки фиксированный, является (9.4.32). AIC выбирает $\widehat{p} = 14$, в то время как BIC выбирает $\widehat{p} = 1$. При заданном $\widehat{p} = 1$ мы оцениваем расширенную авторегрессию по максимальному периоду выборки, равному $t = \widehat{p} + 2, \ldots, T$ (с марта 1953 г. по июль 1971 г.), с 221 наблюдением. Оцененная авторегрессия имеет вид:

$$R_t = \underset{(0.057)}{0.088} + \underset{(0{,}0162)}{0{,}97705}\, R_{t-1} - \underset{(0.067)}{0.20}\ \Delta R_{t-1},$$

$$R^2 = 0.944, \text{ объем выборки} = 221.$$

На основе этой оцененной регрессии мы можем вычислить

$$\text{ADF } \widehat{\rho}^{\mu} = 221 \times \frac{0.97705 - 1}{1 + 0.20} = -4.22, \quad t^{\mu} = \frac{0.97705 - 1}{0.0162} = -1.42.$$

Из табл. 9.1, панель (b), 5-процентное критическое значение для ADF ρ^{μ}-статистики равно -14.1. 5-процентное критическое значение для ADF t^{μ}-статистики равно -2.86 из табл. 9.2, панель (b). Для каждой из статистик нулевая гипотеза I(1) не отвергается без сомнений.

[1] [Ng and Perron, 1995] доказали этот результат о выборе глубины запаздываний только для случая без постоянной составляющей. То, что это может быть обобщено на случай константы, а также на случай тренда, было подтверждено в частной беседе с одним из авторов статьи.

Включение линейного тренда

Как и в DF-тестах, здесь также можно допустить наличие линейного тренда. Заменяя y_t в расширенной авторегрессии (9.4.3) на $y_t - \alpha - \delta \cdot t$, мы получаем следующую расширенную авторегрессию с временным трендом:

$$y_t = \alpha^* + \delta^* \cdot t + \rho y_{t-1} + \zeta_1 \Delta y_{t-1} + \zeta_2 \Delta y_{t-2} + \cdots + \zeta_p \Delta y_{t-p} + \varepsilon_t, \tag{9.4.36}$$

где

$$\alpha^* = (1-\rho)\alpha + (\rho - \zeta_1 - \zeta_2 - \cdots - \zeta_p)\delta, \quad \delta^* = (1-\rho)\delta. \tag{9.4.37}$$

Поскольку $\delta^* = 0$, когда $\rho = 1$, нулевая гипотеза I(1) (о том, что процесс является I(1)-процессом с или без сноса) влечет выполнение $\rho = 1$ *и* $\delta^* = 0$ в (9.4.36), но мы, как обычно, фокусируемся на единственном ограничении $\rho = 1$.

Объединяя методы, использованные для доказательства утверждений 9.5 и 9.7, легко доказать следующее:

Утверждение 9.8 (расширенный тест Дики — Фуллера на единичный корень с линейным временным трендом): *Предположим, что $\{y_t\}$ является суммой линейного временного тренда и процесса* ARIMA$(p, 1, 0)$, *так что $\{\Delta y_t\}$ — стационарный* AR(p)*-процесс, среднее которого может быть, а может и не быть равным нулю. Рассмотрим оценивание расширенной авторегрессии с трендом,* (9.4.36), *и пусть $(\widehat{\alpha}, \widehat{\delta}, \widehat{\rho}^\tau, \widehat{\zeta}_1, \widehat{\zeta}_2, \ldots, \widehat{\zeta}_p)$ являются OLS-оценками коэффициентов. Также пусть t^τ будет t-значением для гипотезы $\rho = 1$. Тогда*

$$\textit{ADF } \rho^\tau\textit{-статистика:} \quad \frac{T \cdot (\widehat{\rho}^\tau - 1)}{1 - \widehat{\zeta}_1 - \widehat{\zeta}_2 - \cdots - \widehat{\zeta}_p} \underset{d}{\to} DF_\rho^\tau. \tag{9.4.38}$$

$$\textit{ADF } t^\tau\textit{-статистика:} \quad t^\tau \underset{d}{\to} DF_t^\tau, \tag{9.4.39}$$

где случайные величины DF_ρ^τ и DF_t^τ определены в (9.3.19) *и* (9.3.20).

Прежде чем перейти к примеру, сделаем три замечания:

- (Численная инвариантность к параметрам тренда.) Как и в утверждении 9.5, поскольку регрессия включает постоянную составляющую и время, ни α, ни δ не влияют на OLS-оценки $(\rho, \zeta_1, \ldots, \zeta_p)$ и их стандартные ошибки, так что (α, δ) не будут влиять на распределение ADF-статистик как в конечных, так и на больших выборках.

- (Обобщение Said — Dickey.) Обобщение Said — Dickey — Ng — Perron применимо и здесь: если $\{\Delta y_t\}$ является стационарным и обратимым процессом ARMA(p, q) с возможно ненулевым средним, то ADF ρ^{τ}- и t^{τ}-статистики имеют предельные распределения, указанные в утверждении 9.8, при условии, что $\widehat{p}$ выбирается последовательным t-правилом или правилами AIC или BIC.

- (Нужно ли включать время в регрессию?) Тот же самый комментарий, который мы приводили по поводу выбора между DF-тестами с или без тренда, применим и к ADF-тестам. С одной стороны, если вы уверены, что процесс при нулевой гипотезе не имеет тренда, то должны использовать ADF ρ^{μ}- и t^{μ}-тесты из утверждения 9.7 с расширенной авторегрессией без использования времени, поскольку на конечных выборках мощность ADF-тестов в общем случае больше в расширенной авторегрессии, не включающей время. С другой стороны, если вы полагаете, что процесс может иметь тренд, то должны использовать ADF ρ^{τ}- и t^{τ}-тесты из утверждения 9.8 с включением времени в расширенную авторегрессию.

Пример 9.3 (имеет ли ВВП единичный корень?): Как было показано на рис. 9.1, логарифм ВВП США имеет выраженный линейный тренд, поэтому мы включаем время в расширенную авторегрессию. Используя логарифм квартальных данных реального ВВП США с 1947:Q1 по 1998:Q1 (205 наблюдений), мы оцениваем расширенную авторегрессию (9.4.36). Как и в примере 9.2, мы полагаем $p_{\max}(T) = [12 \cdot (205/100)^{1/4}]$ (что равно 14). Так же, как и в примере 9.2, период выборки фиксирован ($t = p_{\max}(T) + 2, \ldots, T$) в процессе выбора количества лагированных приращений. Последовательное t-правило выбирает $\widehat{p} = 12$, в то время как и AIC, и BIC выбирают $\widehat{p} = 1$. При заданном $\widehat{p} = 1$ расширенная авторегрессия, оцененная по максимальной выборке $t = \widehat{p} + 2, \ldots, T$ (с 1947:Q3 по 1998:Q1), имеет вид:

$$\begin{array}{ccccccccc} y_t = & 0.22 & + & 0.00022\,t & + & 0.9707\,y_{t-1} & + & 0.348\,\Delta y_{t-1}. \\ & (0.10) & & (0.00011) & & (0.0137) & & (0.066) \end{array}$$

$$R^2 = 0.999,\ \text{объем выборки} = 203.$$

На основе этой оцененной регрессии мы можем вычислить

$$\text{ADF}\ \widehat{\rho}^{\tau} = 203 \times \frac{0.9707388 - 1}{1 - 0.348} = -9.11,\quad t^{\mu} = \frac{0.9707388 - 1}{0.0137104} = -2.13.$$

Из табл. 9.1, панель (с), 5-процентное критическое значение для ADF $\widehat{\rho}^{\tau}$-статистики равно -21.7, из табл. 9.2, панель (с), 5-процентное критическое значение для ADF t^{τ}-статистики равно -3.41. Для каждой из этих статистик нулевая гипотеза I(1) не отвергается без сомнений.

Данных, охватывающих чуть более 50 лет, может оказаться недостаточно для различения между I(1) и I(0)-альтернативами. (Как замечено в [Perron, 1991], среди прочего, именно протяженность данных, а не частота выборки, например квартальная против годовой, более важна с точки зрения мощности тестов на единичный корень, поскольку $1-\widehat{\rho}$ намного меньше для квартальных данных, чем для месячных.) Сейчас мы используем для ADF-тестов данные годового ВВП с 1869 по 1997 год[1]. Поэтому $T = 129$. Как и прежде, мы полагаем $p_{\max}(T) = [12{\cdot}(129/100)^{1/4}]$ (что равно 12) и фиксируем период выборки в процессе выбора глубины запаздываний. Последовательное t-правило выбирает 9 для $\widehat{p}$, в то время как значение, выбранное AIC и BIC, равно 1. Когда $\widehat{p} = 1$, регрессия, оцененная по максимальной выборке $t = \widehat{p}+2,\ldots,T$ (с 1871 по 1997, 127 наблюдений), имеет вид:

$$\begin{array}{rccccc} y_t = & 0{,}67 & + \ 0{,}0014\, t + & 0{,}85886\, y_{t-1} + & 0{,}32 & \Delta y_{t-1}, \\ & (0{,}18) & (0{,}0048) & (0{,}03989) & (0{,}086) & \end{array}$$

$$R^2 = 0{,}999, \text{ объем выборки} = 127.$$

ADF ρ^τ и t^τ-статистики равны $-26{,}2$ и $-3{,}53$ соответственно, что меньше (больше по абсолютному значению), чем соответствующие 5-процентные критические значения, равные $-21{,}7$ и $-3{,}41$. На годовых данных, покрывающих намного более длинный диапазон времени, мы можем отвергнуть нулевую гипотезу I(1) о том, что ВВП имеет стохастический тренд.

Примеры 9.1–9.3 создают впечатление легкости, с которой нулевая гипотеза I(1) не отвергается, и это приводит к подозрению, что DF- и ADF-тесты могут иметь низкую мощность на конечных выборках. Проблемы мощности будут изучаться в следующем параграфе.

Сводный обзор DF- и ADF-тестов и других тестов на единичный корень

Рассмотренные нами в этом и предыдущих параграфах различные I(1)-процессы удовлетворяют модели (множеству DGP):

$$y_t = d_t + z_t, \quad z_t = \rho z_{t-1} + u_t, \quad \{u_t\} \sim \text{процесс с нулевым средним.} \tag{9.4.40}$$

Ограничения на (9.4.40), помимо ограничения $\rho = 1$, которое характеризует нулевую гипотезу I(1) для каждого случая, таковы:

(1) Утверждение 9.3: $d_t = 0$, $\{u_t\}$ — независимый белый шум.

(2) Утверждение 9.4: $d_t = \alpha$, $\{u_t\}$ — независимый белый шум.

[1] Данные годового ВВП с 1929 года доступны на сайте бюро экономического анализа (*Bureau of Economic Analysis*). Данные о ВВП из [Balke and Gordon, 1989] используются для экстраполяции ВВП до 1869 года.

(3) Утверждение 9.5: $d_t = \alpha + \delta \cdot t$, $\{u_t\}$ — независимый белый шум.

(4) Обобщение Said — Dickey утверждения 9.6: $d_t = 0$, $\{u_t\}$ — $\mathrm{ARMA}(p, q)$ с нулевым средним.

(5) Обобщение Said — Dickey утверждения 9.7: $d_t = \alpha$, $\{u_t\}$ — $\mathrm{ARMA}(p, q)$ с нулевым средним.

(6) Обобщение Said — Dickey утверждения 9.8: $d_t = \alpha + \delta \cdot t$, $\{u_t\}$ — $\mathrm{ARMA}(p, q)$ с нулевым средним.

Контрольные вопросы

1. (Неединственность расщепления на I(0)-процесс с нулевым средним и I(1)-процесс без тренда.) Предположим, что в (9.4.4)) $p = 2$. Покажите, что авторегрессию третьего порядка можно эквивалентно записать как

$$y_t = a_1 \cdot y_{t-1} + a_2 \cdot (y_{t-1} - y_{t-2}) + a_3 \cdot (y_{t-1} - y_{t-3}) + \varepsilon_t.$$

Как (a_1, a_2, a_3) связаны с (ϕ_1, ϕ_2, ϕ_3)? Какой регрессор является I(0)-процессом с нулевым средним, а какой — I(1)-процессом без тренда? Проверьте, что OLS-оценка коэффициента при y_{t-1}, полученная при оценивании этого уравнения, численно равна OLS-оценке для ρ из (9.4.3) с $p = 2$.

2. Пусть $\phi(z) = 1 - \phi_1 z - \cdots - \phi_{p+1} z^{p+1}$ является полиномом, ассоциированным с $\mathrm{AR}(p+1)$-процессом (9.4.4)). Покажите, что уравнение $\phi(z) = 0$ имеет вещественный корень внутри единичного круга, если $\rho > 1$. **Указание:** *$\phi(1) = 1 - (\phi_1 + \cdots + \phi_{p+1}) = 1 - \rho < 0$, если $\rho > 1$. $\phi(0) = 1 > 0$.*

3. Докажите, что (9.4.14) $\to_d \frac{\lambda^2}{2} W(1)^2 + \frac{\gamma_0}{2}$. **Указание:** *Асимптотическое распределение $\frac{1}{T}\sum_{t=1}^{T} \Delta y_{t-1} y_{t-1}$ то же самое, что и для $\frac{1}{T}\sum_{t=1}^{T} \Delta y_t y_t$. К тому же $\Delta y_t y_t = \Delta y_t y_{t-1} + (\Delta y_t)^2$. Используйте утверждение 9.2(b).*

4. Пусть $\{\Delta y_t\}$ является I(0)-процессом с нулевым средним, удовлетворяющим (9.2.1)–(9.2.3). (Стационарный $\mathrm{AR}(p)$-процесс с нулевым средним для $\{\Delta y_t\}$, рассмотренный в утверждении 9.6, является его частным случаем.)

   (a) Покажите, что $\{\Delta y_{t-1}\varepsilon_t\}$ является стационарной и эргодической мартингал-разностью. **Указание:** *Поскольку $\{\varepsilon_t\}$ и $\{\Delta y_t\}$ совместно эргодически стационарны, $\{\Delta y_{t-1} \cdot \varepsilon_t\}$ является эргодически стационарным. Чтобы показать, что $\{\Delta y_{t-1}\varepsilon_t\}$ является мартингал-разностью, используйте закон повторных математических ожиданий. В $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\}$ информации больше, чем в $\{\Delta y_{t-2}\varepsilon_{t-1}, \Delta y_{t-3}\varepsilon_{t-2}, \ldots\}$.*

   (b) Докажите (9.4.19). **Указание:** *Вы должны показать, что $\mathrm{E}[(\Delta y_{t-1}\varepsilon_t)^2] = \gamma_0 \cdot \sigma^2$.*

5. Проверьте (9.4.24). **Указание:** *Из (9.4.20) и (9.2.4) вытекает, что*

$$T \cdot (\widehat{\rho} - 1) \underset{d}{\to} \frac{1}{\psi(1)} \frac{\frac{1}{2}(W(1)^2 - 1)}{\int_0^1 W^2 dr}.$$

Кроме того, по (9.4.15) *и* (9.2.4)

$$s \cdot \sqrt{(1,1) \text{ элемент } (\boldsymbol{A}_T)^{-1}} \underset{d}{\to} 1 \Big/ \sqrt{\psi(1)^2 \int_0^1 W^2 \mathrm{d}r}.$$

Поскольку из условия стационарности $\psi(1) > 0$, *то* $\sqrt{\psi(1)^2} = \psi(1)$.

6. (t-тест на ζ.) В тексте утверждается (сразу вслед за (9.4.27)), что обычное t-значение для каждого ζ имеет асимптотическое стандартное нормальное распределение. Проверьте это для случая $p = 1$. **Указание:** *Для* $p = 1$ (9.4.27) *имеет вид:*

$$\sqrt{T} \cdot (\widehat{\zeta}_1 - \zeta_1) \underset{d}{\to} N(0, \sigma^2/\gamma_0).$$

Вам нужно показать, что умноженная на $\sqrt{T}$ *стандартная ошибка* $\widehat{\zeta}_1$ *сходится по вероятности к квадратному корню из* σ^2/γ_0. *Умноженная на* $\sqrt{T}$ *стандартная ошибка* $\widehat{\zeta}_1$ *равна* $s \cdot \sqrt{(2,2) \text{ элемент } (\boldsymbol{A}_T)^{-1}}$.

9.5. Какой тест на единичный корень использовать

Кроме DF и ADF-тестов, доступен и ряд других тестов на единичный корень (перечень их см. в: [Maddala and Kim, 1998, Chapter 3]). Самый известный из них — тест Филлипса [Phillips, 1987] для случая (4) (упомянутого в конце предыдущего параграфа) и тест Филлипса — Перрона [Phillips and Perron, 1988], обобщение теста Филлипса, покрывающее случаи (5) и (6). (Тесты Филлипса выводятся в аналитическом упражнении 6.) Эти тесты основаны на OLS-оценке коэффициента при y_{t-1} в AR(1)-уравнении, а не в расширенной авторегрессии, с долговременной дисперсией $\{\Delta y_t\}$, оцененной по остаткам в AR(1)-уравнении. Мы не представляем их в учебнике, поскольку свойства этих тестов в конечных выборках довольно плохие (см. ниже). Существует новое поколение тестов на единичный корень с разумно низкими искажениями размера и хорошей мощностью. Они включают ADF-GLS-тест [Elliott, Rothenberg and Stock, 1996] и M-тесты [Perron and Ng, 1996].

Асимптотика локальности к единице

Эти и большинство других тестов на единичный корень состоятельны против всех фиксированных I(0)-альтернатив[1]. Какой состоятельный тест следует выбрать в сравнении с другим? Напомним, что в параграфе 2.4 мы определили асимптотическую мощность против последовательности локальных альтернатив, дрейфующих к нулевой гипотезе со скоростью $\sqrt{T}$. Оказывается, что в контексте единичных корней соответствующая скорость равна T, а не $\sqrt{T}$ (см., например: [Stock, 1994,

[1] То, что DF-тесты состоятельны против общих I(0)-альтернатив, проверялось в контрольном вопросе 5 к параграфу 9.3. По поводу состоятельности ADF-тестов см., например: [Stock, 1994, p. 2770].

pp. 2772–2773]). То есть при T, стремящемся к бесконечности, вероятность отвергнуть нулевую гипотезу $\rho = 1$, когда истинный DGP имеет вид:

$$\rho = 1 + \frac{c}{T}, \tag{9.5.1}$$

имеет предел, заключенный между 0 и 1. Этот предел и есть асимптотическая мощность. Последовательность локальных альтернатив, описанных в (9.5.1), называется **локальными к единице альтернативами** (*local-to-unity alternatives*). По определению, асимптотическая мощность равна номинальному размеру при $c = 0$. Можно показать, что DF и ADF ρ-тесты являются более мощными, чем DF и ADF t-тесты против локальных к единице альтернатив (см.: [Stock, 1994, pp. 2774–2776]). То есть для любого номинального размера и любого c асимптотическая мощность DF и ADF ρ-статистик больше, чем мощность DF и ADF t-статистик. На этом основании мы должны предпочесть основанный на ρ-тест тесту, основанному на t, однако этот вердикт отменяется, когда мы изучаем их свойства в конечных выборках.

Свойства в конечных выборках

Обращаясь к вопросу о тестировании нулевой гипотезы I(1) против фиксированной, а не дрейфующей, I(0)-альтернативы, двумя желательными свойствами теста в конечных выборках являются достаточно малые искажения размера и высокая мощность против I(0)-альтернатив. Существует большой объем исследований методом Монте-Карло свойств на конечных выборках DF, ADF и других тестов на единичный корень (в качестве ссылок см. статьи, цитированные на с. 2777 в [Stock, 1994], а также [Ng and Perron, 1998], по поводу сравнения ADF-GLS- и M-тестов). Основные выводы заключаются в следующем.

- Практически все тесты страдают от искажений размера, в частности, когда нулевая гипотеза в некотором смысле близка к I(0). M-тест в общем случае имеет наименьшее искажение размера, следующий за ним — ADF t. Искажения размера у ADF ρ-теста намного больше, чем у ADF t-теста. Например, нулевая гипотеза I(1), рассмотренная в [Schwert, 1989], имеет вид:

$$y_t - y_{t-1} = \varepsilon_t + \theta\varepsilon_{t-1}, \{\varepsilon_t\} \text{ — гауссовские i.i.d.} \tag{9.5.2}$$

Если θ близко к -1, то этот I(1)-процесс очень близок к процессу белого шума, поскольку AR-корень, равный единице, почти равен MA-корню $-\theta$. Симуляции в [Schwert, 1989] показали, что тест ADF ρ- и тесты Z_ρ и Z_t Филлипса — Перрона (см. аналитическое упражнение 6) слишком часто отвергают нулевую гипотезу I(1) в конечных выборках, когда θ близко к -1. В частности, для тестов Филлипса — Перрона в случае, когда $\theta = -0.8$ и номинальный размер равен 5%, фактический размер превышает 90% даже для выборок столь большого размера, как 1000.

- Как в случае ADF-тестов выбранная глубина запаздываний $\widehat{p}$ влияет на размер и мощность в конечных выборках? Для заданного правила выбора $\widehat{p}$ мощность в общем случае выше для ADF ρ-тестов, чем для ADF t-тестов.

- Мощность в общем случае намного больше для ADF-GLS и M-тестов, чем для ADF t- и ρ-тестов.

В отличие от M-статистик статистику ADF-GLS можно легко вычислить в стандартных регрессионных пакетах. В эмпирическом упражнении в этой главе вам будет предложено использовать ADF-GLS для тестирования нулевой гипотезы I(1).

9.6. Приложение: Паритет покупательной способности

Теория **паритета покупательной способности, ППС** (*purchasing power parity, PPP*) утверждает, что для любых двух стран обменный курс равен отношению уровней цен в этих странах. Иными словами, национальные уровни цен, конвертированные в общую валюту, должны быть равными. Пусть P_t — индекс цен для Соединенных Штатов, P_T^* — индекс цен для рассматриваемой зарубежной страны (скажем, Великобритании) и S_t — обменный курс в долларах за единицу иностранной валюты. ППС утверждает, что

$$P_t = S_t \cdot P_t^*. \tag{9.6.1}$$

или, переходя к логарифмам и используя для них строчные символы[1],

$$p_t = s_t + p_t^*. \tag{9.6.2}$$

С ППС связан, но отличается от него **закон одной цены** (*law of one price*), который утверждает, что (9.6.1) выполняется для любого товара, и при этом P_t представляет собой цену интересующего товара в долларах, P_t^* — цену этого товара в иностранной валюте. Если международный ценовой арбитраж работает без сбоев, закон единой цены должен выполняться для каждой даты t. Если закон единой цены имеет место для всех товаров и если корзина товаров для индекса цен одинакова для двух стран, то должен выполняться ППС.

[1]Уравнения (9.6.1) или (9.6.2) являются формулировкой «абсолютного ППС». То, что называется относительным ППС, требует только того, чтобы выполнялась первая разность от (9.6.2), то есть темпы роста обменного курса должны компенсировать разницу между темпами роста в отечественных и зарубежных индексах цен. Относительный ППС не следует путать с более слабой версией ППС, которая будет введена ниже.

В связи с целым рядом факторов, ограничивающих международный ценовой арбитраж, ППС не выполняется в точности для каждой даты[1] t. Более слабая версия ППС заключается в том, что отклонение от ППС

$$z_t \equiv s_t - p_t + p_t^*, \tag{9.6.3}$$

которое является реальным обменным курсом, стационарно. Даже если это не обеспечивает выполнение закона единой цены в краткосрочной перспективе, международный арбитраж должен сказываться в долговременной перспективе.

Смущающая жизнеспособность модели случайного блуждания?

В течение многих лет исследователи находили сложным отвергать гипотезу о том, что при режимах плавающего обменного курса реальные обменные курсы следуют случайному блужданию. Как отметил [Rogoff, 1996], это приводило к своего рода смущению, поскольку каждая разумная теоретическая модель предполагает выполнение слабой версии ППС. Тем не менее исследования с середины 1980-х годов, в которых привлекались данные за более протяженные интервалы времени, смогли отвергнуть нулевую гипотезу случайного блуждания. Хорошим примером является работа [Lothian and Taylor, 1996]. Чтобы обнаружить, что нулевую гипотезу случайного блуждания можно отвергнуть, эта работа использует охватывающие два столетия, с 1791 по 1990 год, ежегодные данные по реальным обменным курсам доллар/стерлинг и франк/стерлинг. В дальнейшем мы применяем ADF t-тест к тем же данным по обменному курсу доллар/стерлинг (долларов за фунт).

Реальный обменный курс доллар/стерлинг, вычисленный как (9.6.3) и изображенный на рис. 9.4, не обнаруживает временного тренда. Поэтому мы не включаем время в расширенную авторегрессию. Если бы мы имели дело со слабой версией закона единой цены, с P_t и P_t^*, обозначающими внутреннюю и зарубежную цены определенного товара, то тогда могли бы исключить из расширенной авторегрессии константу, поскольку изменение в единицах измерения товара не влияет на z_t. Но P_t и P_t^* здесь являются индексами цен и изменение в базовом году приводит к прибавлению константы к z_t. Чтобы сделать тест инвариантным к таким изменениям, в расширенную авторегрессию постоянная составляющая включается. Применяя ADF-, а не DF-тест, мы можем допустить сериальную корреляцию в первых разностях z_t при нулевой гипотезе, но глубина запаздываний, выбранная посредством BIC (с $p_{\max}(T) = [12 \cdot (200/100)^{1/4}] = 14$), оказывается равной 0, так что ADF

[1] См., например, Section 4 в [Rogoff, 1996] относительно перечня факторов, которые препятствуют выполнению ППС.

t-тест сводится к DF t-тесту, а расширенная авторегрессия редуцируется до AR(1)-уравнения. Оцененное AR(1)-уравнение имеет вид:

$$\underset{(0{,}052)}{z_t = 0{,}179} + \underset{(0{,}0326)}{0{,}8869}\, z_{t-1}, \quad SER = 0{,}071, \quad R^2 = 0{,}999, \quad t = 1792, \ldots, 1990.$$

DF t-статистика (t^{μ}) равна $-3{,}5$ ($= (0{,}8869 - 1)/0{,}0326$), что значимо на 1%-ном уровне. Таким образом, мы можем, несомненно, отвергнуть гипотезу о том, что реальный обменный курс стерлинг/доллар является случайным блужданием.

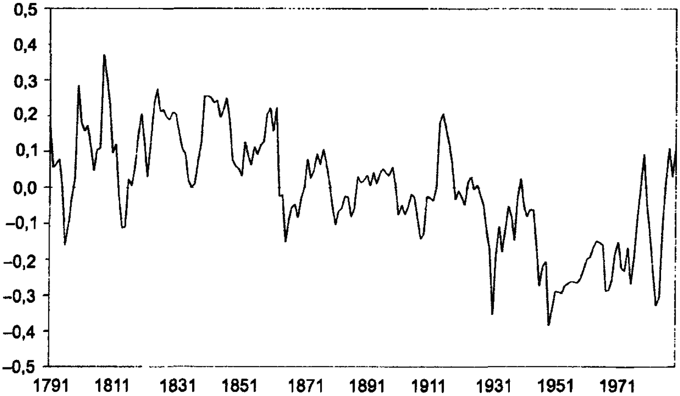

Рис. 9.4. Реальный обменный курс доллар/стерлинг, 1791–1990, значение 1914 года полагается равным 0

Очевидным недостатком этого результата является то, что период выборки 1791–1990 годы включает периоды и фиксированного, и плавающего курса. Вполне вероятно, что арбитраж мировой цены работает более эффективно при фиксированных курсовых режимах. Если это так, то на периодах фиксированного курса коэффициент при z_{t-1} должен быть более близким к 0. [Lothian and Taylor, 1996] сообщают, что если использовать простой тест Чоу, то нельзя отвергнуть гипотезу о том, что коэффициент при z_{t-1} один и тот же до и после введения плавающего курса в 1973 году. В эмпирическом упражнении в этой главе вам будет предложено проверить их результат, а также использовать ADF-GLS для проверки той же самой гипотезы.

Набор задач для главы 9

Аналитические упражнения

1. Докажите утверждение 9.2(b). **Указание:** *По определению* $\Delta\xi_t$, *мы имеем* $\xi_t = \xi_{t-1} + \Delta\xi_t$. *Поэтому*

$$(\xi_t)^2 = (\xi_{t-1} + \Delta\xi_t)^2.$$

Из этого выведите соотношение:

$$\Delta\xi_t \cdot \xi_{t-1} = \left(\frac{1}{2}\right)\left[(\xi_t)^2 - (\xi_{t-1})^2 - (\Delta\xi_t)^2\right].$$

Просуммируйте это по $t = 1, 2, \dots, T$, чтобы получить:

$$\sum_{t=1}^{T} \Delta\xi_t \cdot \xi_{t-1} = \left(\frac{1}{2}\right)\left[(\xi_T)^2 - (\xi_0)^2\right] - \left(\frac{1}{2}\right)\sum_{t=1}^{T}(\Delta\xi_t)^2.$$

разделите обе стороны на T, чтобы получить:

$$\frac{1}{T}\sum_{t=1}^{T} \Delta\xi_t \cdot \xi_{t-1} = \frac{1}{2}\left(\frac{\xi_T}{\sqrt{T}}\right)^2 - \frac{1}{2}\left(\frac{\xi_0}{\sqrt{T}}\right)^2 - \frac{1}{2T}\sum_{t=1}^{T}(\Delta\xi_t)^2. \qquad (*)$$

Примените утверждение 6.9 и Эргодическую теорему к слагаемым в правой части (*).

2. (Доказательство утверждения 9.4.) Пусть $\widehat{\rho}^{\mu}$ — OLS-оценка для ρ в AR(1)-уравнении с постоянной составляющей, (9.3.11), и определим центрированный ряд $\{y^{\mu}_{t-1}\}$ $(t = 1, 2, \dots, T)$ как

$$y^{\mu}_{t-1} \equiv y_{t-1} - \overline{y}, \quad \overline{y} \equiv \frac{y_0 + y_1 + \dots + y_{T-1}}{T} \quad (t = 1, 2, \dots, T). \qquad (1)$$

Это y^{μ}_{t-1} равно OLS-остатку от регрессии y_{t-1} на константу для $t = 1, 2, \dots, T$.

(a) Выведите формулу:

$$\widehat{\rho}^{\mu} - 1 = \frac{\sum_{t=1}^{T} \Delta y_t \, y^{\mu}_{t-1}}{\sum_{t=1}^{T}(y^{\mu}_{t-1})^2}. \qquad (2)$$

Указание: *По теореме Фриша — Во $\widehat{\rho}^{\mu}$ численно равна OLS-оценке коэффициента в регрессии y_t на y^{μ}_{t-1} (без постоянной составляющей). Таким образом,*

$$\widehat{\rho}^{\mu} = \frac{\sum_{t=1}^{T} y_t \, y^{\mu}_{t-1}}{\sum_{t=1}^{T}(y^{\mu}_{t-1})^2} \quad \text{или} \quad \widehat{\rho}^{\mu} - 1 = \frac{\sum_{t=1}^{T}(y_t - y^{\mu}_{t-1}) \cdot y^{\mu}_{t-1}}{\sum_{t=1}^{T}(y^{\mu}_{t-1})^2}. \qquad (3)$$

Используйте тот факт, что $\sum_{t=1}^{T} y^{\mu}_{t-1} = 0$, чтобы показать, что $\sum_{t=1}^{T}(y_t - y^{\mu}_{t-1}) \cdot y^{\mu}_{t-1} = \sum_{t=1}^{T}(y_t - y_{t-1}) \cdot y^{\mu}_{t-1}$.

(b) (Легко.) Покажите, что при нулевой гипотезе о том, что $\{y_t\}$ является I(1)-процессом без сноса (необязательно случайным блужданием без сноса),

$$T \cdot (\widehat{\rho}^{\mu} - 1) \underset{\mathrm{d}}{\to} \frac{\frac{\lambda^2}{2}\left([W^{\mu}(1)]^2 - [W^{\mu}(0)]^2\right) - \frac{\gamma_0}{2}}{\lambda^2 \cdot \int_0^1 [W^{\mu}(r)]^2\,\mathrm{d}r}, \tag{4}$$

где γ_0 является дисперсией Δy_t, λ^2 — долговременная дисперсия Δy_t и $W^{\mu}(\cdot)$ — центрированное стандартное броуновское движение, введенное в параграфе 9.2. **Указание:** *Используйте утверждение 9.2(c) и (d) с $\xi_t = y_t$.*

(c) (Тривиально.) Проверьте, что $T \cdot (\widehat{\rho}^{\mu} - 1)$ сходится к DF^{μ}_{ρ} (определенному в (9.3.13)), если $\{y_t\}$ является случайным блужданием без сноса. **Указание:** *Если y_t — случайное блуждание, то $\lambda^2 = \gamma_0$.*

(d) (Дополнительно.) Пусть s является стандартной ошибкой регрессии для AR(1) уравнения с постоянной составляющей, (9.3.11). Покажите, что s^2 является состоятельной оценкой для $\gamma_0 (\equiv \mathrm{Var}(\Delta y_t))$ при нулевой гипотезе о том, что $\{y_t\}$ является I(1)-процессом без сноса (так что Δy_t является I(0)-процессом с нулевым средним, удовлетворяющим (9.2.1)–(9.2.3)).

Указание: *Пусть $(\widehat{\alpha}^*, \widehat{\rho}^{\mu})$ — OLS-оценка для (α^*, ρ). Покажите, что $\widehat{\alpha}^*$ сходится к нулю по вероятности. Запишите s^2 как*

$$\begin{aligned} s^2 &= \frac{1}{T-1}\sum_{t=1}^{T}[(\Delta y_t - \widehat{\alpha}^*) - (\widehat{\rho}^{\mu} - 1)y_{t-1}]^2 = \\ &= \frac{1}{T-1}\sum_{t=1}^{T}(\Delta y_t - \widehat{\alpha}^*)^2 - \\ &\quad - \frac{2}{T-1}\cdot[T\cdot(\widehat{\rho}^{\mu} - 1)]\cdot\frac{1}{T}\sum_{t=1}^{T}(\Delta y_t - \widehat{\alpha}^*)\cdot y_{t-1} + \\ &\quad + \frac{1}{T-1}\cdot[T\cdot(\widehat{\rho}^{\mu} - 1)]^2\cdot\frac{1}{T^2}\sum_{t=1}^{T}(y_{t-1})^2. \end{aligned} \tag{5}$$

Поскольку $\operatorname{plim}\widehat{\alpha}^* = 0$, легко показать, что первое слагаемое в правой части (5) сходится по вероятности к $\gamma_0 = \mathrm{E}\big[(\Delta y_t)^2\big]$. Для доказательства того, что второе слагаемое сходится к нулю по вероятности, считайте доказанным, что

$$\frac{1}{\sqrt{T}}\,\frac{1}{T}\sum_{t=1}^{T} y_{t-1} \underset{d}{\to} \lambda\int_0^1 W(r)\,\mathrm{d}r. \tag{6}$$

(e) Покажите, что $t^{\mu} \to_{\mathrm{d}} DF_t^{\mu}$, если $\{y_t\}$ является случайным блужданием без сноса. **Указание:** *Поскольку* $\{y_t\}$ *является случайным блужданием без сноса,* $\lambda^2 = \gamma_0 = \sigma^2$($\equiv$ *дисперсия* ε_t). *Поэтому* s *является состоятельной оценкой для* σ.

$$t^{\mu} = \frac{\widehat{\rho}^{\mu} - 1}{s \times \sqrt{(2,2)\ \text{элемент}\ (\boldsymbol{X}'\boldsymbol{X})^{-1}}}. \tag{7}$$

Покажите, что квадратный корень из $(2,2)$ *элемента в* $(\sum\limits_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t')^{-1}$ *равен:*

$$\frac{1}{\sqrt{\sum\limits_{t=1}^{T} (y_{t-1}^{\mu})^2}},$$

где $\boldsymbol{x}_t \equiv [1, y_{t-1}]'$.

3. (Доказательство утверждения 9.5.) Пусть $\widehat{\rho}^{\tau}$ является OLS-оценкой для ρ в AR(1)-уравнении с временным трендом, (9.3.17), и определим детрендированный ряд $\{y_{t-1}^{\tau}\}$ ($t = 1, 2, \ldots, T$) как

$$y_{t-1}^{\tau} \equiv y_{t-1} - \widehat{\alpha} - \widehat{\delta} \cdot t, \tag{8}$$

где $(\widehat{\alpha}, \widehat{\delta})$ являются OLS-оценками коэффициентов для регрессии y_{t-1} на $(1, t)$ для $t = 1, 2, \ldots, T$.

(a) Выведите

$$\widehat{\rho}^{\tau} - 1 = \frac{\sum_{t=1}^{T} \Delta y_t\, y_{t-1}^{\tau}}{\sum_{t=1}^{T} (y_{t-1}^{\tau})^2}. \tag{9}$$

Указание: *По теореме Фриша — Во,* $\widehat{\rho}^{\tau}$ *численно равна OLS-оценке коэффициента при* y_{t-1}^{τ} *в регрессии* y_t *на* y_{t-1}^{τ} *(без постоянной составляющей или времени). Таким образом,*

$$\widehat{\rho}^{\tau} = \frac{\sum_{t=1}^{T} y_t\, y_{t-1}^{\tau}}{\sum_{t=1}^{T} (y_{t-1}^{\tau})^2} \quad \text{или} \quad \widehat{\rho}^{\tau} - 1 = \frac{\sum_{t=1}^{T} (y_t - y_{t-1}^{\tau}) \cdot y_{t-1}^{\tau}}{\sum_{t=1}^{T} (y_{t-1}^{\tau})^2}. \tag{10}$$

Покажите, что $\sum_{t=1}^{T} (y_t - y_{t-1}^{\tau}) \cdot y_{t-1}^{\tau} = \sum_{t=1}^{T} (y_t - y_{t-1}) \cdot y_{t-1}^{\tau}$. *(По построению,* $\sum_{t=1}^{T} y_{t-1}^{\tau} = 0$ *и* $\sum_{t=1}^{T} t \cdot y_{t-1}^{\tau} = 0$ *(это нормальные уравнения).)*

(b) (Легко.) Покажите, что при предположении о том, что $y_t = \alpha + \delta \cdot t + \xi_t$, где $\{\xi_t\}$ является I(1)-процессом без сноса,

$$T \cdot (\widehat{\rho}^{\tau} - 1) \underset{\mathrm{d}}{\to} \frac{\frac{\lambda^2}{2}\left([W^{\tau}(1)]^2 - [W^{\tau}(0)]^2\right) - \frac{\gamma_0}{2}}{\lambda^2 \int_0^1 [W^{\tau}(r)]^2\, \mathrm{d}r}, \tag{11}$$

где γ_0 является дисперсией Δy_t, λ^2 — долговременная дисперсия Δy_t и $W^\tau(\cdot)$ — детрендированное стандартное броуновское движение, введенное в параграфе 9.2. **Указание:** *Пусть ξ_{t-1}^τ — остатки от гипотетической регрессии ξ_{t-1} на $(1, t)$ для $t = 1, 2, \ldots, T$, где $\xi_t = y_t - \alpha - \delta \cdot t$ (эта регрессия гипотетическая, поскольку мы не наблюдаем $\{\xi_t\}$). Тогда $\xi_{t-1}^\tau = y_{t-1}^\tau$ для $t = 1, 2, \ldots, T$, поскольку ξ_t отличается от y_t только на линейный тренд $\alpha + \delta \cdot t$. Используйте утверждение 9.2, пункты (e) и (f). Поскольку $\Delta y_t = \delta + \Delta\xi_t$ при нулевой гипотезе, дисперсия $\Delta\xi_t$ равна дисперсии Δy_t, а долговременная дисперсия $\Delta\xi_t$ равна долговременной дисперсии Δy_t.*

(c) (Тривиально.) Проверьте, что $T \cdot (\widehat{\rho}^\tau - 1)$ сходится к DF_ρ^τ (определенной в (9.3.19)), если $\{y_t\}$ является случайным блужданием с или без сноса.

4. (Доказательство (9.4.16).) В тексте осталось недоказанным, что

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\varepsilon_t \underset{\mathrm{d}}{\to} c_1 \equiv \frac{1}{2}\cdot\psi(1)\cdot[W(1)^2 - 1] \qquad (9.4.16)$$

Докажите это при условии, что $\{y_t\}$ является I(1)-процессом без сноса, так что $\{\Delta y_t\}$ является I(0)-процессом с нулевым средним, удовлетворяющим (9.2.1)–(9.2.3). **Указание:** *ε_t в (9.4.16) является процессом инноваций в MA-представлении для $\{\Delta y_t\}$. Используя разложение Бевериджа — Нельсона, имеем:*

$$y_t = \psi(1)w_t + \eta_t + (y_0 - \eta_0), \quad w_t \equiv \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t.$$

Поэтому

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}\varepsilon_t = \psi(1)\frac{1}{T}\sum_{t=1}^{T} w_{t-1}\varepsilon_t + \frac{1}{T}\sum_{t=1}^{T}\eta_{t-1}\varepsilon_t + (y_0 - \eta_0)\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t. \quad (*)$$

Покажите, что второе и третье слагаемые в правой части уравнения сходятся по вероятности к нулю. После того как вы выполнили аналитическое упражнение 1, легко показать, что

$$\frac{1}{T}\sum_{t=1}^{T} w_{t-1}\varepsilon_t \underset{d}{\to} \left(\frac{\sigma^2}{2}\right)[W(1)^2 - 1].$$

5. (Дополнительно, доказательство утверждения 9.7.) Чтобы упростить вычисления, положим $p = 1$, так что расширенная авторегрессия имеет вид:

$$y_t = \alpha^* + \rho y_{t-1} + \zeta_1 \Delta y_{t-1} + \varepsilon_t.$$

Пусть $(\widehat{\alpha}^*, \widehat{\rho}^\mu, \widehat{\zeta}_1)$ — OLS-оценки коэффициентов. Покажите, что

$$\frac{T \cdot (\widehat{\rho} - 1)}{1 - \widehat{\zeta}_1} \underset{\mathrm{d}}{\to} DF_\rho^\mu.$$

Указание: *По теореме Фриша — Во, $(\widehat{\rho}^\mu, \widehat{\zeta}_1)$ численно те же самые, что и OLS-оценки коэффициентов в регрессии y_t на $(y_{t-1}^\mu, (\Delta y_{t-1})^{(\mu)})$, где y_{t-1}^μ — остатки от регрессии y_{t-1} на константу, а $(\Delta y_{t-1})^{(\mu)}$ — остатки от регрессии Δy_{t-1} на константу для $t = 1, 2, \ldots, T$. Следовательно, если $\boldsymbol{b} = (\widehat{\rho}^\mu, \widehat{\zeta}_1)'$, $\boldsymbol{\beta} = (\rho, \zeta_1)'$ и $\boldsymbol{x}_t = (y_{t-1}^\mu, (\Delta y_{t-1})^{(\mu)})'$, то они удовлетворяют* (9.4.8). *Поэтому*

$$\textit{(1,1)-й элемент в } \boldsymbol{A}_T = \frac{1}{T^2} \sum_{t=1}^T (y_{t-1}^\mu)^2,$$

$$\textit{1-й элемент в } \boldsymbol{c}_T = \frac{1}{T} \sum_{t=1}^T y_{t-1}^\mu \cdot \varepsilon_t.$$

Примените утверждение 9.2, пункты (c) и (d) с $\xi_t = y_t$ к этим выражениям.

6. (Тесты Филлипса.) Предположим, что $\{y_t\}$ является I(1)-процессом без тренда, так что Δy_t может быть сериально коррелированным. Тем не менее оценим AR(1)-уравнение для $\{y_t\}$ без постоянной составляющей, а не расширенную авторегрессию без константы. Пусть $\widehat{\rho}$ является OLS-оценкой коэффициента при y_{t-1}. Определим

$$Z_\rho \equiv T \cdot (\widehat{\rho} - 1) - \frac{1}{2} \cdot \left(\frac{T^2 \cdot \widehat{\sigma}_{\widehat{\rho}}^2}{s^2} \right) \cdot (\widehat{\lambda}^2 - \widehat{\gamma}_0).$$

$$Z_t \equiv \frac{s}{\widehat{\lambda}} \cdot t - \frac{1}{2} \cdot \left(\frac{T \cdot \widehat{\sigma}_{\widehat{\rho}}^2}{s \cdot \widehat{\lambda}} \right) \cdot (\widehat{\lambda}^2 - \widehat{\gamma}_0).$$

где $\widehat{\sigma}_{\widehat{\rho}}^2$ — стандартная ошибка для $\widehat{\rho}$, s — стандартная ошибка OLS-регрессии, $\widehat{\lambda}^2$ — состоятельная оценка для λ^2, $\widehat{\gamma}_0$ — состоятельная оценка для γ_0 и t — t-значение для гипотезы $\rho = 1$.

(a) Покажите, что Z_ρ и Z_t можно записать как

$$Z_\rho = \frac{\frac{1}{T} \sum_{t=1}^T \Delta y_t y_{t-1} - \frac{1}{2}(\widehat{\lambda}^2 - \widehat{\gamma}_0)}{\frac{1}{T^2} \sum_{t=1}^T (y_{t-1})^2}.$$

$$Z_t = \frac{s}{\widehat{\lambda}} \cdot t - \frac{\frac{1}{2}(\widehat{\lambda}^2 - \widehat{\gamma}_0)}{\widehat{\lambda} \times \sqrt{\frac{1}{T^2} \sum_{t=1}^T y_{t-1}^2}}.$$

Указание: *Из алгебраических преобразований метода наименьших квадратов,*

$$\frac{\widehat{\sigma}_{\widehat{\rho}}}{s} = \frac{1}{\sqrt{\sum_{t=1}^{T}(y_{t-1})^2}}.$$

(b) Покажите, что

$$Z_\rho \underset{\mathrm{d}}{\to} DF_\rho \text{ и } Z_t \underset{\mathrm{d}}{\to} DF_t.$$

Указание: *Используйте утверждение 9.2, пункты (a) и (b) с* $\xi_t = y_t$.

7. (Дополнительно.) Покажите, что $\alpha(L)$ в (9.2.5) абсолютно суммируем. **Указание:** *Доказательство в одну строку имеет вид:*

$$\sum_{j=0}^{\infty}|\alpha_j| = \sum_{j=0}^{\infty}\left|-\sum_{i=j+1}^{\infty}\psi_i\right| \le \sum_{j=0}^{\infty}\sum_{i=j+1}^{\infty}|\psi_i| = \sum_{i=0}^{\infty} i|\psi_i| < \infty.$$

Проверьте каждое из этих равенств и неравенств.

Упражнения на метод Монте-Карло

1. (При включении времени вы получаете меньшую мощность.) В этой симуляции мы хотим проверить, что в конечных выборках мощность против стационарных альтернатив больше у DF t^μ-теста, чем у DF t^τ-теста, то есть включение времени в AR(1)-уравнение уменьшает мощность. DGP, который мы берем в качестве стационарной альтернативы, является стационарным AR(1)-процессом:

$$y_t = \rho y_{t-1} + \varepsilon_t, \ \rho = 0.95, \ \varepsilon_t \sim \text{i.i.d.}N(0,1), \ y_0 = 0.$$

Возьмите объем выборки T равным 100 и номинальный размер равным 5%. Ваша компьютерная программа для вычисления мощности на конечных выборках должна иметь следующие шаги:

(1) Установите счетчики равными нулю. Должно быть два счетчика: один — для DF t^μ, а другой — для DF t^τ. Они будут записывать количество случаев, когда (неверная) нулевая гипотеза отвергается.

(2) Начните цикл с большим количеством повторений (скажем, 1 миллион). В каждом повторении проделайте следующее:

(i) Сгенерируйте вектор данных $(y_0, y_1, \ldots, y_T)$.
Совет для пакета GAUSS: Как было отмечено в упражнении на метод Монте-Карло для главы 1, чтобы сгенерировать $(y_1, y_2, \ldots, y_T)$, исключите создание цикла внутри

текущего цикла; используйте матричные операторы. Матричная формула имеет вид:

$$\underset{(T\times 1)}{\boldsymbol{y}} = \underset{(T\times 1)}{\boldsymbol{r}} \cdot y_0 + \underset{(T\times T)}{\boldsymbol{A}}\underset{(T\times 1)}{\boldsymbol{\varepsilon}},$$

где

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix}, \quad \boldsymbol{r} = \begin{bmatrix} \rho \\ \rho^2 \\ \vdots \\ \rho^T \end{bmatrix},$$

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ \rho & 1 & 0 & \cdots & 0 \\ \rho^2 & \rho & 1 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \cdots & \rho & 1 \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{bmatrix}.$$

Чтобы определить $\boldsymbol{r}$, используйте GAUSS-команду `seqm`. Чтобы определить $\boldsymbol{A}$, используйте `toeplitz` и `lowmat`. $\boldsymbol{r}$ и $\boldsymbol{A}$ должны определяться в шаге (1) вне текущего цикла.

(ii) Вычислите t^μ и t^τ по $(y_0, y_1, \ldots, y_T)$. (Поэтому объем выборки фактически равен $T+1$.)

(iii) Увеличьте счетчик для t^μ на 1, если $t^\mu < -2.86$ (5%-ное критическое значение для DF_t^μ). Увеличьте счетчик для t^τ на 1, если $t^\tau < -3{,}41$ (5%-ное критическое значение для DF_t^τ).

(3) После окончания цикла разделите счетчик для каждой статистики на число повторений для вычисления частоты отвержения нулевой гипотезы. Эта частота сходится к мощности в конечных выборках при возрастании числа Монте-Карло повторений к бесконечности. Мощность должна оказаться около 0,124 для DF t^μ-теста (так что только приблизительно в 12% случаев мы отвергаем неверную нулевую гипотезу!) и 0,092 для DF t^τ-теста. Поэтому действительно мощность оказывается меньше, когда в регрессию включается время, по крайней мере против этого частного DGP.

2. (Искажения размера.) В этих симуляциях Монте-Карло мы вычисляем искажения размера для ADF ρ^μ- и t^μ-тестов. Наш DGP при нулевой гипотезе — один из рассмотренных в [Schwert, 1989]:

$$y_t = \rho y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}, \ \rho = 1, \ \theta = -0.8,$$
$$\varepsilon_t \sim \text{i.i.d.}N(0,1), \ y_0 = 0, \ \varepsilon_0 = 0.$$

Для $T = 100$ и номинального размера 5% вычислите размер на конечных выборках, или точный размер, для ρ^{μ} и t^{μ}. Было бы полезно выбрать число лагированных приращений в расширенной регрессии одним из стандартных, зависящих от данных методов (последовательное t, AIC или BIC), но, чтобы оставить программу простой, а также чтобы сократить время вычислений, положите его равным 4. Таким образом, уравнение регресии, в отличие от DGP, имеет вид:

$$y_t = \text{const} + \rho y_{t-1} + \zeta_1 \cdot (y_{t-1} - y_{t-2}) + \zeta_2 \cdot (y_{t-2} - y_{t-3}) + \\ + \zeta_3 \cdot (y_{t-3} - y_{t-4}) + \zeta_4 \cdot (y_{t-4} - y_{t-5}) + \text{ошибка}.$$

Поскольку выборка включает y_0, период выборки, который может использовать четыре лагированных приращения, включает значения t с $t = 5$ (не $p+2 = 6$) по $t = T$, и фактический объем выборки равен $T - 4$. Проверьте, что искажение размера меньше для ADF t^{μ}-теста, чем для ADF ρ^{μ}-теста. (Размер должен быть около 0,496 для ADF ρ^{μ} и 0,290 для ADF t^{μ}.)

Эмпирические упражнения

Прежде чем отвечать, прочитайте [Lothian and Taylor, 1996, pp. 488–509] (но пропустите их обсуждение тестов Филлипса — Перрона). Файл LT.ASC содержит следующие годовые данные:

Столбец 1: год

Столбец 2: обменный курс доллар/стерлинг (назовем его S_t)

Столбец 3: U.S. WPI (wholesale price index, индекс оптовых цен), нормализованный к 100 для 1914 года (P_t)

Столбец 4: U.K. WPI, нормализованный к 100 для 1914 года (P_t^*)

Период выборки с 1791 по 1990 год (200 наблюдений). Это те же данные, которые использовали [Lothian and Taylor, 1996], сделанные доступными для нас. Относительно источников данных и того, как авторы объединили их для построения последовательных временных рядов, см. приложение в [Lothian and Taylor, 1996]. (Согласно приложению, обменный курс (и, возможно, WPI) являются годовыми средними. Это довольно неудачно, из-за проблемы агрегирования по времени нужно использовать данные на момент времени: если некоторая переменная следует случайному блужданию на основе дневных данных, то среднегодовые значения этой переменной не следуют случайному блужданию. См. об этом более подробно часть (f) ниже.)

Вычислите реальный обменный курс доллар/стерлинг как

$$z_t \equiv \ln(S_t) - \ln(P_t) + \ln(P_t^*).$$

Поскольку S_t выражается в долларах за фунт, его возрастание означает укрепление стерлинга. График реального обменного курса на рис. 9.4 не показывает явного временного тренда, поэтому мы применяем ADF t^μ-тест. Следовательно, оцениваемая расширенная авторегрессия имеет вид:

$$z_t = \text{const} + \rho z_{t-1} + \zeta_1 \cdot (z_{t-1} - z_{t-2}) + \cdots + \zeta_p \cdot (z_{t-p} - z_{t-p-1}) + \text{ошибка},$$

где p — число лагированных приращений, включенных в расширенную авторегрессию (которое обозначалось как $\widehat{p}$ в тексте).

(a) (ADF с автоматическим выбором лагов.) Для всего периода выборки с 1791 по 1990 год примените для выбора p (число лагированных приращений, которые нужно включить) последовательное t-правило, AIC и BIC. Возьмите $p_{\max}$ (максимальное число лагированных изменений в расширенной авторегрессии, обозначенное в тексте как $p_{\max}(T)$) согласно (9.4.31) (так что $p_{\max} = 14$). Для последовательного t-теста используйте 10-процентный уровень значимости. (Вы должны найти $p = 14$ последовательным t, 14 по AIC и 0 по BIC.) Проверьте, что значение t^μ-статистики для $p = 0$ (при оценивании по всей выборке) соответствует значению τ_μ, приведенному в Table 1 [Lothian and Taylor, 1996].

Советы для пакета RATS: RATS позволяет вам изменять количество регрессоров в цикле. Поскольку RATS OLS-процедура `LINREG` не позволяет получать стандартные ошибки, в этой программе нельзя вычислить DF t^μ. Поэтому вычисляйте DF t^μ-статистику как t-значение для коэффициента при z_{t-1} в регрессии $z_t - z_{t-1}$. В `LINREG` не вычисляются также ни AIC, ни BIC. Поэтому используйте COMPUTE. Ниже в кодах RATS `z` является логарифмом реального обменного курса, а `dz` — его первая разность. При применении правил последовательного t, AIC и BIC зафиксируйте период выборки с 1806 по 1990 год (что соответствует $p_{\max} + 2, \ldots, T$). Коды RATS для AIC и BIC таковы:

```
linreg dz 1806:1 1990:1;# constant z{1}
compute ssrr=%rss
compute akaike = log(%rss/%nobs)+%nreg*2.0/%nobs
compute schwarz = log(%rss/%nobs)+%nreg*log(%nobs)
                    /%nobs
```

```
display akaike schwarz
* run ADF with 1 through 14 lagged changes
do p=1,14
  linreg dz 1806:1 1990:1;# constant z{1} dz{1 to p}
  compute akaike = log(%rss/%nobs)+%nreg*2.0/%nobs
  compute schwarz = log(%rss/%nobs)+%nreg
  *log(%nobs)/%nobs
  display akaike schwarz
end do
```

Советы для пакета TSP: В отличие от RATS TSP не позволяет вам изменять период выборки и количество регрессоров в цикле. Процедура `OLSQ` распечатывает BIC, но не AIC, поэтому AIC можно вычислить, используя команду `SET`.

(b) (DF-тест на периоде плавающего курса.) Примените DF t^{μ}-тест к периоду плавающего обменного курса 1974–1990 годы. Из-за лагированной зависимой переменной z_{t-1} первое уравнение соответствует $t = 1975$:

$$z_{1975} = \text{const} + \rho z_{1974} + \text{ошибка}.$$

Поэтому объем выборки равен 16. Можете ли вы отвергнуть нулевую гипотезу случайного блуждания на 5%-ном уровне?

(c) (DF-тест на периоде золотого стандарта.) Примените DF t^{μ}-тест к периоду золотого стандарта 1870–1913 годы. Положите $t = 1$ для 1871 года, так что объем выборки равен 43. Можете ли вы отвергнуть нулевую гипотезу случайного блуждания на 5%-ном уровне? Находится ли оценка для ρ (коэффициент при z_{t-1} в регрессии z_t на константу и z_{t-1}) ближе к нулю, чем в (b)?

(d) (Тест Чоу.) Отвергнув нулевую гипотезу I(1), мы действуем в предположении, что логарифм реального обменного курса z_t следует стационарному AR(1)-процессу. Для тестирования нулевой гипотезы о том, что в AR(1)-уравнении нет структурного изменения, разделим выборку на два подпериода, 1791–1973 и 1974–1990 годы, и применим тест Чоу. Вы можете вспомнить, что тест Чоу изначально разработан для строго экзогенных регрессоров. Если K — количество регрессоров, включая константу (2 в данном случае), — это F-тест с $(K, T-2K)$ степенями свободы. Этот результат неприменим здесь непосредственно, поскольку в AR(1)-уравнении регрессором является лагированная зависимая переменная, которая не является строго экзогенной. Однако, как вы проверили в аналитическом упражнении 12 в главе 2, если переменные (регрессоры и зависимая переменная) эргодически стационарны (что выполняется здесь, поскольку $\{z_t\}$ является эргодически стационарным, когда $|\rho| < 1$), то на больших выборках с возрастающим объемом каждой из двух подвыборок $K \cdot F$ асимптотически имеет распределение $\chi^2(K)$. Этот результат не требует, чтобы регрессоры были строго экзогенными. Вычислите $K \cdot F$ и проверьте утверждение (см. р. 498–502 в [Lothian and Taylor, 1996]) о том, что «в течение периода плавающего курса нет признаков структурного сдвига в параметрах». В данном случае, с наличием лагированной зависимой переменной, возникает вопрос о том, куда включать уравнение для 1974 года,

$$z_{1974} = \text{const} + \rho z_{1973} + \text{ошибка},$$

в первую часть выборки или во вторую. На больших выборках это не имеет значения; включите это уравнение в первую подвыборку (так что объем первой подвыборки равен 183). (Замечание: $K \cdot F$ должно быть около 1,26.)

(e) (ADF-GLS.) Как было упомянуто в параграфе 9.5, ADF-GLS-тест достигает существенно более высокой мощности ценой только несколько большего искажения размера. ADF-GLS t^{μ}- и t^{τ}-статистики вычисляются следующим образом. Как в ADF t^{μ}- и t^{τ}-тестах, процесс при нулевой гипотезе имеет вид:

$$y_t = d_t + z_t,\ z_t = \rho z_{t-1} + u_t,\ \rho = 1,$$
$$\{u_t\} \text{ является I(0)-процессом с нулевым средним.} \quad (3)$$

Запишите $d_t = \boldsymbol{x}_t'\boldsymbol{\beta}$. В случае, когда d_t является константой, $\boldsymbol{x}_t = 1$ и $\boldsymbol{\beta} = \alpha$, в то время как в случае, когда d_t является линейным трендом, $\boldsymbol{x}_t = (1, t)'$ и $\boldsymbol{\beta} = (\alpha, \delta)'$. Мы предполагаем, что выборка имеет объем T и это $(y_1, y_2, \dots, y_T)$.

(i) (GLS-оценивание $\boldsymbol{\beta}$.) При оценивании регрессионного уравнения

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + \text{ошибка},\ (t = 1, 2, \dots, T). \quad (4)$$

примените GLS, предполагая, что процесс ошибок является AR(1)-процессом с авторегрессионным коэффициентом $\bar{\rho} = 1 + \bar{c}/T$. (Значение $\bar{c}$ будет специфицировано через мгновение.) То есть оцените $\boldsymbol{\beta}$ посредством OLS, регрессируя

$$y_1, y_2 - \bar{\rho}y_1, y_3 - \bar{\rho}y_2, \dots, y_T - \bar{\rho}y_{T-1}$$

на

$$\boldsymbol{x}_1, \boldsymbol{x}_2 - \bar{\rho}\boldsymbol{x}_1, \boldsymbol{x}_3 - \bar{\rho}\boldsymbol{x}_2, \dots, \boldsymbol{x}_T - \bar{\rho}\boldsymbol{x}_{T-1}.$$

(В отличие от истинного GLS, первые наблюдения, y_1 и $\boldsymbol{x}_1$, здесь не взвешиваются посредством $\sqrt{1-\overline{\rho}^2}$, но это не имеет значения на больших выборках.) Положите $\overline{c}$ равным

$$\overline{c} = -7$$

в центрированном случае (с $\boldsymbol{x}_t = 1$),

$$\overline{c} = -13{,}5$$

в детрендированном случае. (Грубо говоря, если $\rho = 1 + c/T$, то такой выбор $\overline{c}$ гарантирует, что асимптотическая мощность при $c = \overline{c}$ составляет примерно 50% асимптотической мощности, достигаемой идеальным тестом, построенным специально для этой альтернативы, $c = \overline{c}$.) Назовем полученную оценку $\widehat{\boldsymbol{\beta}}_{\text{GLS}}$ и определим $\widetilde{y}_t \equiv y_t - \boldsymbol{x}_t'\widehat{\boldsymbol{\beta}}_{\text{GLS}}$. (Это *не следует* путать с GLS-остатком, который равен $(y_t - \overline{\rho}y_{t-1}) - (\boldsymbol{x}_t - \overline{\rho}\boldsymbol{x}_{t-1})'\widehat{\boldsymbol{\beta}}_{\text{GLS}}$.)

(ii) (ADF без постоянной составляющей, основанный на остатках.) Оцените расширенную авторегрессию *без* постоянной составляющей, использующую $\widetilde{y}_t$ вместо y_t:

$$\widetilde{y}_t = \rho\widetilde{y}_{t-1} + \zeta_1 \cdot (\widetilde{y}_{t-1} - \widetilde{y}_{t-2}) + \dots + \zeta_p \cdot (\widetilde{y}_{t-p} - \widetilde{y}_{t-p-1}) + \text{ошибка}.$$

t-значение для $\rho = 1$ равно ADF-GLS t-статистике; назовем ее ADF-GLS t^{μ}, если $\boldsymbol{x}_t = 1$ и ADF-GLS t^{τ}, если $\boldsymbol{x}_t = (1, t)'$. Предельное распределение ADF-GLS t^{μ} такое же, как и у DF_t (оно табулировано в табл. 9.2, панель (a)), в то время как предельное распределение ADF-GLS t^{τ} табулировано в Table I.C. в [Elliott et al., 1996]. Воспроизведем их левосторонние критические значения:

$$-3{,}48 \text{ для } 1\%,\ -3{,}15 \text{ для } 2{,}5\%,$$
$$-2{,}89 \text{ для } 5\% \text{ и } -2{,}57 \text{ для } 10\%.$$

[Elliott et al., 1996] показали для ADF-GLS-теста, что его асимптотическая мощность выше, чем у ADF ρ- и t-тестов, и что, в сравнении с ADF-тестами, его мощность на конечных выборках существенно выше, лишь с незначительным ухудшением искажений размера.

Примените ADF-GLS t^{μ}-тест к подвыборке с золотым стандартом 1870–1913 годов. Выберите глубину запаздываний по BIC с $p_{\max}$ (максимальное число лагированных приращений в расширенной авторегрессии, обозначенное как $p_{\max}(T)$ в тексте) согласно (9.4.31) (так что $p_{\max} = 9$). Можете ли вы отвергнуть нулевую гипотезу о том, что реальный обменный курс является I(1)-процессом без сноса? (Замечание: BIC выберет 0 для p. Статистика должна быть равна $-2{,}24$. Вы должны быть в состоянии отвергнуть нулевую гипотезу I(1) на 5%-ном уровне.)

Советы для пакета TSP: Пример кодов TSP, предполагающих в этой части, что $p = 0$.

```
?
? (e) ADF-GLS
?
? Step 1: GLS demeaning
?
set rhobar=1-7/(1913-1870+1);
smpl 1870 1913;
const=1;
ct=1;
zt=z; ? z is the log real exchange rate
smpl 1871 1913;
zt=z-rhobar*z(-1);
ct=const-rhobar*const(-1);
smpl 1870 1913;
olsq zt ct;
z=z-@coef(1);
? @coef(1) is the GLS estimate of the mean
?
? Step 2: DF with demeaned variable
?
smpl 1871 1913;
olsq z z(-1);
set t=(@coef(1)-1)/@ses(1);print t;
```

(f) (Дополнительное аналитическое упражнение об агрегировании по времени.) Предположим, что y_t является случайным блужданием без сноса. Создайте на его основе усредненные по времени ряды по формуле:

$$\overline{y}_1 = \frac{y_1 + y_2 + \cdots + y_n}{n}, \quad \overline{y}_2 = \frac{y_{n+1} + y_{n+2} + \cdots + y_{2n}}{n}, \quad \ldots$$

(Думайте об y_t как об обменном курсе в день t, а о $\overline{y}_j$ — как о среднем дневных курсов за год j.) Покажите, что $\{\overline{y}_j\}$ не является случайным блужданием. Более конкретно, покажите, что

$$\mathrm{Corr}(\overline{y}_{j+1} - \overline{y}_j, \overline{y}_{j+2} - \overline{y}_{j+1}) \to 0{,}25 \text{ при } n \to \infty.$$

Ответы на избранные вопросы

Аналитические упражнения

1. Поскольку $\mathrm{E}(\xi_0^2) < \infty$, $\xi_0/\sqrt{T}$ сходится в среднем квадратическом (и, следовательно, по вероятности) к 0 при $T \to \infty$. Таким образом, второе слагаемое в правой части $(*)$ в указании можно игнорировать. Что касается первого слагаемого, то мы имеем:

$$\frac{\xi_T}{\sqrt{T}} = \frac{\xi_0}{\sqrt{T}} + \frac{\Delta\xi_1 + \Delta\xi_2 + \cdots + \Delta\xi_T}{\sqrt{T}}.$$

Первое слагаемое в правой части этого уравнения можно игнорировать. По утверждению 6.9, второе слагаемое сходится к $N(0, \lambda^2)$, поскольку $\Delta\xi_t$ является линейным процессом с абсолютно суммируемыми MA-коэффициентами. Поэтому первое слагаемое в правой части $(*)$ сходится по распределению к $(\lambda^2/2)X$, где $X \sim \chi^2(1)$. Поскольку $\{\Delta\xi_t\}$ является эргодически стационарным, последнее слагаемое в правой части $(*)$ сходится по вероятности к $\frac{1}{2}\,\mathrm{E}[(\Delta\xi_t)^2]$, по эргодической теореме.

5. (1, 1)-й элемент $\boldsymbol{A}_T$ сходится к случайной величине, указанной в утверждении 9.2(c). Что касается первого элемента, c_T, то, по разложению Бевериджа — Нельсона:

$$y_t = \psi(1)w_t + \eta_t + (y_0 - \eta_0), \quad w_t \equiv \varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t. \tag{1}$$

Следовательно,

$$y_{t-1}^{\mu} = \psi(1)w_{t-1}^{\mu} + \eta_{t-1}^{\mu}, \tag{2}$$

где

$$x_{t-1}^{\mu} = x_{t-1} - \overline{x}, \quad \overline{x} = \frac{x_0 + \cdots + x_{T-1}}{T} \text{ для } x = w, \eta$$

и

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^{\mu} \cdot \varepsilon_t = \psi(1)\frac{1}{T}\sum_{t=1}^{T} w_{t-1}^{\mu} \cdot \varepsilon_t + \frac{1}{T}\sum_{t=1}^{T} \eta_{t-1}^{\mu} \cdot \varepsilon_t. \tag{3}$$

Второе слагаемое в правой части этого уравнения можно игнорировать асимптотически. По утверждению 2.9(d) для $\xi_t = w_t$ (случайное блуждание без сноса), первое слагаемое в правой части сходится к

$$\psi(1)\left(\frac{\sigma^2}{2}\right)\left\{[W^\mu(1)]^2 - [W^\mu(0)]^2 - 1\right\}.$$

Отсюда мы имеем:

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}^\mu \cdot \varepsilon_t \underset{\mathrm{d}}{\to} \frac{1}{2}\cdot\sigma^2\cdot\psi(1)\left\{[W^\mu(1)]^2 - [W^\mu(0)]^2 - 1\right\}. \tag{4}$$

Матрица $\boldsymbol{A}_T$ асимптотически диагональна также и для центрированного случая. Внедиагональные элементы $\boldsymbol{A}_T$ равны $1/\sqrt{T}$, умноженному на

$$\frac{1}{T}\sum_{t=1}^{T}(\Delta y_{t-1})^{(\mu)} y_{t-1}^\mu. \tag{5}$$

Поскольку $\sum_{t=1}^{T} y_{t-1}^\mu = 0$, это равно:

$$\frac{1}{T}\sum_{t=1}^{T}\Delta y_{t-1} y_{t-1}^\mu. \tag{6}$$

Путем небольших алгебраических манипуляций, аналогичных применявшимся в контрольном вопросе 3 в параграфе 9.4 и утверждении 9.2(d), легко показать, что (6) имеет предельное распределение. Поэтому (6), умноженное на $1/\sqrt{T}$, что является внедиагональным элементом $\boldsymbol{A}_T$, сходится к нулю по вероятности. Тогда, используя те же аргументы, которые мы использовали в тексте для случая без константы, легко показать, что $\widehat{\zeta}_1$ состоятельна для ζ_1. Поэтому $\psi(1)$ состоятельно оценивается посредством $1/(1-\widehat{\zeta}_1)$. Собирая все вместе, получаем:

$$\frac{T\cdot(\widehat{\rho}^\mu - 1)}{1-\widehat{\zeta}_1} \underset{\mathrm{d}}{\to} DF_\rho^\mu.$$

Эмпирические упражнения

(a) При $p = 0$ оцененное AR(1)-уравнение

$$z_t = \text{const} + \underset{(0{,}0326)}{0{,}8869}\, z_{t-1},\quad SSR = 0{,}995135,\quad R^2 = 0{,}790,\quad T = 199.$$

Значение t^μ равно $-3{,}47$, что соответствует значению в Table 1 [Lothian and Taylor, 1996]. 5-процентное критическое значение равно $-2{,}86$ по табл. 9.2, панель (b).

(b) Оцененное AR(1)-уравнение для периода 1974–1990 годов имеет вид:

$$z_t = \text{const} + \underset{(0{,}1904)}{0{,}7704}\, z_{t-1}, \quad SSR = 0.150964, \quad R^2 = 0{,}539, \quad T = 16.$$

$t^{\mu} = -1.21 > -2.86$. Мы не можем отвергнуть нулевую гипотезу на 5%-ном уровне.

(c) Оцененное AR(1)-уравнение для периода 1870–1913 годов имеет вид:

$$z_t = \text{const} + \underset{(0{,}1045)}{0{,}7222}\, z_{t-1}, \quad SSR = 0{,}069429, \quad R^2 = 0{,}538, \quad T = 43.$$

$t^{\mu} = -2{,}66 > -2.86$. Мы не можем отвергнуть нулевую гипотезу на 5%-ном уровне.

(d) Оцененное AR(1)-уравнение для периода 1791–1974 годов имеет вид:

$$z_t = \text{const} + \underset{(0{,}0329)}{0.8993}\, z_{t-1}, \quad SSR = 0{,}837772, \quad R^2 = 0.805, \quad T = 183.$$

Из этого результата и результатов в (a) и (b)

$$K \cdot F = \frac{0.995135 - (0{,}837772 + 0{,}150964)}{\frac{0{,}837772+0{,}150964}{(199-4)}} = 1{,}26.$$

5-процентное критическое значение для $\chi^2(2)$ равно 5,99. Поэтому мы принимаем нулевую гипотезу о стабильности параметров.

(e) ADF-GLS $t^{\mu} = -2{,}24$. 5%-ное критическое значение в табл. 9.2, панель (a) равно $-1{,}95$. Поэтому мы *можем* отвергнуть нулевую гипотезу I(1).

Литература

Baillie, R., 1996, "Long Memory Processes and Fractional Integration in Econometrics," *Journal of Economics*, 73, 5–59.

Balke, N., and R. Gordon, 1989, "The Estimation of Prewar Gross National Product: Methodology and New Evidence," *Journal of Political Economy*, 97, 39–92.

Chan, N., and C.Wei, 1988, "Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes," *Annals of Statistics*, 16, 367–401.

Dickey, D., 1976, "Estimation and Hypothesis Testing in Nonstationary Time Series," Ph.D. dissertation, Iowa State University.

Dickey, D., and W. Fuller, 1979, "Distribtion of the Estimators for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association*, 74, 427–431.

Elliott, G., T. Rothenberg, and J. Stock, 1996, "Efficient Tests for an Autoregressive Unit Root," *Econometrica*, 64, 813–836.

Fama, E., 1975, "Short-Term Interest Rates as Predictors of Inflation," *American Economic Review*, 65, 269–282.

Fuller,W., 1976, *Introduction to Statistical Time Series*, New York: Wiley.

______,1996, *Introduction to Statistical Time Series* (2d ed.), New York: Wiley.

Hall, P., and C. Heyde, 1980, *Martingale Limit Theory and Its Application*, New York: Academic.

Hamilton, J., 1994, *Time Series Analysis*, Princeton: Princeton University Press.

Hannan, E. J., and M. Deistler, 1988, *The Statistical Theory of Linear Systems*, New York: Wiley.

Kwiatkowski, D., P. Phillips, P. Schmidt, and Y. Shin, 1992, "Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root," *Journal of Economics*, 54, 159–178.

Leybourne, S., and B. McCabe, 1994, "A Consistent Test for a Unit Root," *Journal of Business and Economic Statistics*, 12, 157–166.

Lothian, J., and M. Taylor, 1996, "Real Exchange Rate Behavior: The Recent Float from the Perspective of the Past Two Centuries," *Journal of Political Economy*, 104, 488–509.

Maddala, G. S., and In-Moo Kim, 1998, *Unit Roots, Cointegration, and Structural Change*, New York: Cambridge University Press.

Ng, S., and P. Perron, 1995, "Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag," *Journal of the American Statistical Association*, 90, 268–281.

______,1998, "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power," working paper, mimeo.

Perron, P., 1991, "Test Consistency with Varying Sampling Frequency," *Econometric Theory*, 7, 341–368.

Perron, P., and S. Ng, 1996, "Useful Modifications to Unit Root Tests with Dependent Errors and Their Local Asymptotic Properties," *Review of Economic Studies*, 63, 435–463.

Phillips, P., 1987, "Time Series Regression with a Unit Root," *Econometrica*, 55, 277–301.

Phillips, P., and S. Durlauf, 1986, "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies*, 53, 473–496.

Phillips, P., and P. Perron, 1988, "Testing for a Unit Root in Time Series Regression," *Biometrika*, 75, 335–346.

Rogoff, K., 1996, "The Purchasing Power Parity Puzzle," *Journal of Economic Literature*, 34, 647–668.

Said, S., and D. Dickey, 1984, "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order," *Biometrika*, 71, 599–607.

Sargan, D., and A. Bhargava, 1983, "Testing Residuals from Least Squares Regression for Being Generated by the Gaussian Random Walk," *Econometrica*, 51, 153–174.

Schwert, G., 1989, "Tests for Unit Roots: A Monte Carlo Investigation," *Journal of Business and Economic Statistics*, 7, 147–159.

Sims, C., J. Stock, and M. Watson, 1990, "Inference in Linear Time Series Models with Some Unit Roots," *Econometrica*, 58, 113–144.

Stock, J., 1994, "Unit Roots, Structural Breaks and Trends," Chapter 46 in R. Engle and D. McFadden (eds.), *Handbook of Econometrics*, Volume IV, New York: North Holland.

Tanaka, K., 1990, "Testing for a Moving Average Unit Root," *Econometric Theory*, 6, 433–444.

——, 1996, Time Series Analysis, New York: Wiley.

Watson, M., 1994, "Vector Autoregressions and Cointegration," Chapter 47 in R. Engle and D. Mc-Fadden (eds.), *Handbook of Econometrics*, Volume IV, New York: North Holland.

Глава 10. Коинтеграция

Как иллюстрируют примеры из предыдущего параграфа, многие экономические временные ряды могут характеризоваться как I(1). Но очень часто их линейные комбинации оказываются стационарными. Такие переменные называются **коинтегрированными** (*cointegrated*), а веса в линейной комбинации называются **коинтегрирующим вектором** (*cointegrating vector*). В этой главе изучаются коинтегрированные системы, а именно векторные I(1)-процессы, элементы которых коинтегрированы.

В предыдущей главе мы определили одномерные I(0)- и I(1)-процессы. Первый параграф данной главы представляет обобщение этих понятий на векторные процессы и определяет коинтеграцию для векторных I(1)-процессов. В параграфе 10.2 приводятся альтернативные представления коинтегрированной системы. В параграфе 10.3 более подробно изучается тест, проверяющий, является ли векторный I(1)-процесс коинтегрированным. Техника для оценивания коинтегрирующих векторов и получения статистических выводов о них разрабатывается в параграфе 10.4.

В качестве приложения в параграфе 10.5 рассматривается функция спроса на деньги. Логарифм реального предложения денег, номинальная процентная ставка и логарифм реального дохода оказываются I(1), но существование стабильной функции спроса на деньги подразумевает, что эти переменные коинтегрированы, с коэффициентами в функции спроса на деньги, формирующими коинтегрирующий вектор. Методы параграфа 10.4 используются для оценивания этого коинтегрирующего вектора.

В отличие от других глав мы не представляем результаты в виде ряда утверждений, поскольку уровень предмета таков, что формулировать предположения трудно без введения дополнительных технических терминов. Однако для читателей, которые интересуются техническими деталями, мы указываем в тексте ссылки на источники, в которых эти предположения сформулированы формально.

Замечание об обозначениях: В отличие от других глав символ «n» используется в качестве размерности векторных процессов, и, как и в предыдущих частях, символ «T» используется для объема выборки, а «t» — индекс наблюдения. Также, если y_t является t-м наблюдением n-мерного векторного процесса, его j-й элемент будет обозначаться как y_{jt} вместо y_{tj}. Мы производим эти изменения, чтобы соответствовать обозначениям, используемым в большинстве оригинальных работ по коинтеграции.

10.1. Коинтегрированные системы

На рис. 10.1(а) изображены логарифмы квартальных реального совокупного личного располагаемого дохода на душу населения и личных расходов на потребление для США с 1947:Q1 по 1998:Q1. Оба ряда име-

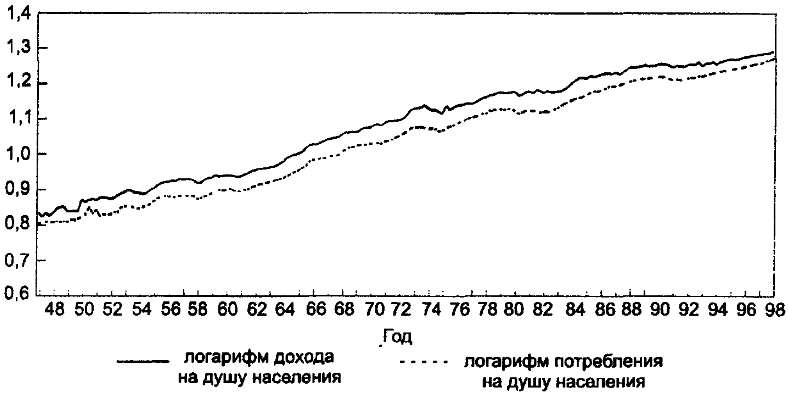

Рис. 10.1(а). Логарифмы дохода и потребления

ют линейные тренды и — как будет проверено в примере ниже — стохастические тренды. Однако эти два I(1)-ряда проявляют тенденцию двигаться совместно, и это наводит на мысль о том, что их разность стационарна. Это пример **коинтеграции** (*cointegration*). Коинтеграционных соотношений в экономике предостаточно. В самом деле, многие из переменных, которые мы изучали в этой книге, являются коинтегрированными: цены на один и тот же сырьевой товар в двух различных странах (разность должна быть стационарной при более слабой версии паритета покупательной способности), долговременная и краткосрочная процентные ставки (даже хотя они имеют тренды, дифференциал доходности может быть стационарным), форвардный и спотовый обменные курсы и так далее. Коинтеграция может определяться для более чем двух переменных. Например, существование стабильной функции спроса на деньги подразумевает, что линейная комбинация логарифма реальной денежной массы, номинальной процентной ставки и логарифма агрегированного дохода может быть стационарной, несмотря на то что каждая из трех переменных является I(1).

Мы начинаем этот параграф обобщая определения одномерных I(0)- и I(1)-процессов из предыдущей главы на векторные процессы. После этого будет формально введено понятие коинтеграции для многомерных I(1)-процессов.

Линейные векторные I(0)- и I(1)-процессы

Наше определение векторных I(0)- и I(1)-процессов следует [Hamilton, 1994] и [Johansen, 1995]. Для введения многомерного обобщения линейного I(0)-процесса с нулевым средним рассмотрим n-мерный VMA (векторное скользящее среднее)-процесс с нулевым средним:

$$\underset{(n\times 1)}{\boldsymbol{u}_t} = \underset{(n\times n)}{\boldsymbol{\Psi}(L)}\underset{(n\times 1)}{\boldsymbol{\varepsilon}_t}, \quad \boldsymbol{\Psi}(L) \equiv \boldsymbol{I}_n + \boldsymbol{\Psi}_1 L + \boldsymbol{\Psi}_2 L^2 + \cdots, \tag{10.1.1}$$

где $\boldsymbol{\varepsilon}_t$ является i.i.d. с

$$\mathrm{E}(\boldsymbol{\varepsilon}_t) = \mathbf{0}, \quad \mathrm{E}(\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') = \underset{(n\times n)}{\boldsymbol{\Omega}}, \quad \boldsymbol{\Omega} \text{ положительно определена.} \tag{10.1.2}$$

(Процесс ошибок $\boldsymbol{\varepsilon}_t$ может быть векторной мартингал-разностью, удовлетворяющей определенным свойствам, но мы не будем допускать такую степень общности.) Как и в одномерном случае, мы накладываем два ограничения. Первое является 1-суммируемостью[1].

$$\{\boldsymbol{\Psi}_j\} \text{ является 1-суммируемой.} \tag{10.1.3}$$

Поскольку $\{\boldsymbol{\Psi}_j\}$ абсолютно суммируема, когда выполняется 1-суммируемость, многомерная версия утверждения 6.1(d) означает, что $\boldsymbol{u}_t$ является (строго) стационарным и эргодическим. Второе ограничение на VMA-процесс с нулевым средним имеет вид:

$$\boldsymbol{\Psi}(1) \neq \underset{(n\times n)}{\mathbf{0}} \quad (n \times n \text{ матрица нулей}). \tag{10.1.4}$$

То есть по крайней мере один элемент $n \times n$ матрицы $\boldsymbol{\Psi}(1)$ ненулевой. Это естественное многомерное обобщение условия (9.2.3b) в определении одномерных I(0)-процессов на с. 599.

(Линейный) n-мерный векторный I(0)-процесс с нулевым средним (*(linear) zero-mean n-dimensional vector I(0) process*), или **(линейная)**

[1] Последовательность матриц $\{\boldsymbol{\Psi}_j\}$ называется 1-суммируемой, если $\{\Psi_{j,kl}\}$ 1-суммируема (то есть $\sum_{j=0}^{\infty} j|\Psi_{j,kl}| < \infty$) для всех $k,l = 1,2,\ldots,n$, где $\Psi_{j,kl}$ — (k,l)-й элемент $n \times n$ матрицы $\boldsymbol{\Psi}_j$.

I(0)-система с нулевым средним (*(linear) zero-mean I(0) system*) — это VMA-процесс, удовлетворяющий (10.1.1)–(10.1.4). При заданной I(0)-системе $\boldsymbol{u}_t$ **(линейный) векторный I(0)-процесс** (*(linear) vector I(0) process*), или **(линейную) I(0)-систему** (*(linear) I(0) system*), можно записать как

$$\boldsymbol{\delta} + \boldsymbol{u}_t,$$

где $\boldsymbol{\delta}$ — n-мерный вектор, представляющий среднее стационарного процесса. На основании формул (6.5.6) и (6.3.15), долговременная ковариационная матрица I(0)-системы имеет вид:

$$\boldsymbol{\Psi}(1)\boldsymbol{\Omega}\boldsymbol{\Psi}(1)'. \tag{10.1.5}$$

Поскольку $\boldsymbol{\Omega}$ положительно определена и $\boldsymbol{\Psi}(1)$ удовлетворяет (10.1.4), по крайней мере один из диагональных элементов этой долговременной ковариационной матрицы положительный, из чего вытекает, что по крайней мере один из элементов I(0)-системы сам по себе является I(0) (то есть I(0) как одномерный процесс). Мы не требуем, чтобы $\boldsymbol{\Psi}$ была обязательно невырожденной. По существу, допущение этой вырожденности составляет суть теории коинтеграции. Следовательно, долговременная ковариационная матрица также может быть вырожденной.

Пример 10.1 (двумерная I(0)-система): Рассмотрим следующий двумерный VMA-процесс первого порядка:

$$\begin{cases} u_{1t} = \varepsilon_{1t} - \varepsilon_{1,t-1} + \gamma\varepsilon_{2,t-1}, \\ u_{2t} = \varepsilon_{2t}, \end{cases} \quad \text{или} \quad \boldsymbol{u}_t = \boldsymbol{\varepsilon}_t + \boldsymbol{\Psi}_1\boldsymbol{\varepsilon}_{t-1}. \tag{10.1.6}$$

где

$$\boldsymbol{\varepsilon}_t = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}, \boldsymbol{\Psi}_1 = \begin{bmatrix} -1 & \gamma \\ 0 & 0 \end{bmatrix}. \tag{10.1.7}$$

Поскольку это VMA конечного порядка, условие 1-суммируемости выполняется тривиальным образом. Требование (10.1.4) выполняется, поскольку

$$\boldsymbol{\Psi}(1) = \boldsymbol{I}_2 + \boldsymbol{\Psi}_1 + \begin{bmatrix} 0 & \gamma \\ 0 & 1 \end{bmatrix} \neq \underset{(2\times 2)}{\boldsymbol{0}}. \tag{10.1.8}$$

Поэтому этот пример соответствует нашему определению I(0)-систем. Если $\gamma = 0$, то первый элемент, u_{1t}, фактически является одномерным I(−1)-процессом.

***n*-мерную I(1)-систему** (*n-dimensional I(1) system*) $\boldsymbol{y}_t$, связанную с I(0)-системой $\boldsymbol{u}_t$ с нулевым средним, можно определить через соотношение:

$$\Delta\boldsymbol{y}_t = \boldsymbol{\delta} + \boldsymbol{u}_t. \tag{10.1.9}$$

Поэтому среднее Δy_t равно δ. Поскольку не каждая компонента соответствующего I(0)-процесса с нулевым средним u_t должна сама по себе быть I(0), некоторые компоненты I(1)-системы y_t сами по себе могут не быть I(1)[1]. Подставляя (10.1.1) в этой выражение, мы получим:

$$\Delta y_t = \delta + \Psi(L)\varepsilon_t. \tag{10.1.10}$$

Это называется **представлением I(1)-системы в форме векторного скользящего среднего (VMA)** (*vector moving average representation*). В уровнях y_t можно записать как

$$y_t = y_0 + \delta \cdot t + u_1 + u_2 + \cdots + u_t, \tag{10.1.11}$$

или, записывая подробно,

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{nt} \end{bmatrix} = \begin{bmatrix} y_{1,0} \\ y_{2,0} \\ \vdots \\ y_{n,0} \end{bmatrix} + \begin{bmatrix} \delta_1 \cdot t \\ \delta_2 \cdot t \\ \vdots \\ \delta_n \cdot t \end{bmatrix} + \begin{bmatrix} u_{1,1} \\ u_{2,1} \\ \vdots \\ u_{n,1} \end{bmatrix} + \begin{bmatrix} u_{1,2} \\ u_{2,2} \\ \vdots \\ u_{n,2} \end{bmatrix} + \cdots + \begin{bmatrix} u_{1,t} \\ u_{2,t} \\ \vdots \\ u_{n,t} \end{bmatrix}.$$

Что касается начального значения y_0, мы предполагаем, что это вектор фиксированных констант или что он случайный, но не зависящий от ε_t для всех t.

Пример 10.2 (двумерная I(1)-система): I(1)-система, для которой соответствующий I(0)-процесс с нулевым средним является двумерным процессом из примера 10.1, имеет вид:

$$\begin{cases} \Delta y_{1t} = \delta_1 + \varepsilon_{1t} - \varepsilon_{1,t-1} + \gamma\varepsilon_{2,t-1}, \\ \Delta y_{2t} = \delta_2 + \varepsilon_{2t}, \end{cases} \quad \text{или} \quad \Delta y_t = \delta + \varepsilon_t + \Psi_1\varepsilon_{t-1}. \tag{10.1.12}$$

где Ψ_1 задается в (10.1.7). В уровнях:

$$\begin{cases} y_{1t} = y_{1,0} + \delta_1 \cdot t + (\varepsilon_{1t} - \varepsilon_{0t}) + \gamma \cdot (\varepsilon_{2,0} + \varepsilon_{2,1} + \cdots + \varepsilon_{2,t-1}), \\ y_{2t} = y_{2,0} + \delta_2 \cdot t + + \cdot (\varepsilon_{2,1} + \varepsilon_{2,2} + \cdots + \varepsilon_{2,t}). \end{cases} \tag{10.1.13}$$

Второй элемент в y_t есть I(1) как одномерный процесс. Если $\gamma = 0$, то первый элемент y_{1t} является стационарным относительно тренда, поскольку отклонение от функции тренда $(y_{1,0} - \varepsilon_{1,0}) + \delta_1 \cdot t$ является стационарным (фактически i.i.d.). В противном случае y_{1t} является I(1).

[1] Во многих учебниках и также в [Engle and Granger, 1987] все компоненты I(1)-системы сами по себе предполагаются I(1), но такое предположение не является необходимым для разработки теории коинтеграции. Этому вопросу придается особая роль, например в [Lütkepohl, 1993, pp. 352–353] и [Johansen, 1995, Chapter 3].

Разложение Бевериджа — Нельсона

Для I(1)-системы, связанным с которой I(0)-процессом с нулевым средним является u_t, удовлетворяющий (10.1.1)–(10.1.4), легко получить разложение Бевериджа — Нельсона (BN). Вспомним из параграфа 9.2, что одномерное BN-разложение основано на переписывании MA-фильтра как (9.2.5) на с. 600. Очевидная многомерная версия этого имеет вид:

$$\underset{(n\times n)}{\boldsymbol{\Psi}(L)} = \underset{(n\times n)}{\boldsymbol{\Psi}(1)} + (1-L)\underset{(n\times n)}{\boldsymbol{\alpha}(L)}, \quad \boldsymbol{\alpha}(L) = \sum_{j=0}^{\infty} \boldsymbol{\alpha}_j L^j,$$
$$\underset{(n\times n)}{\boldsymbol{\alpha}_j} = -(\boldsymbol{\Psi}_{j+1} + \boldsymbol{\Psi}_{j+2} + \cdots). \tag{10.1.14}$$

Поэтому многомерный аналог (9.2.6) на с. 600 имеет вид:

$$\boldsymbol{u}_t = \boldsymbol{\Psi}(1)\boldsymbol{\varepsilon}_t + \boldsymbol{\eta}_t - \boldsymbol{\eta}_{t-1}, \quad \underset{(n\times 1)}{\boldsymbol{\eta}_t} \equiv \boldsymbol{\alpha}(L)\boldsymbol{\varepsilon}_t. \tag{10.1.15}$$

Поскольку $\boldsymbol{\Psi}(L)$ является 1-суммируемым, то $\boldsymbol{\alpha}(L)$ абсолютно суммируем, и $\boldsymbol{\eta}_t$ является ковариационно стационарным процессом. Подставляя это в (10.1.11), мы получаем многомерную версию BN-разложения:

$$\boldsymbol{y}_t = \boldsymbol{\delta}\cdot t + \boldsymbol{\Psi}(1)(\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_2 + \cdots + \boldsymbol{\varepsilon}_t) + \boldsymbol{\eta}_t + (\boldsymbol{y}_0 - \boldsymbol{\eta}_0). \tag{10.1.16}$$

Как и в одномерном случае, I(1)-система $\boldsymbol{y}_t$ раскладывается на линейный тренд $\boldsymbol{\delta}\cdot t$, стохастический тренд $\boldsymbol{\Psi}(1)(\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_2 + \cdots + \boldsymbol{\varepsilon}_t)$, стационарную компоненту $\boldsymbol{\eta}_t$ и начальное условие $\boldsymbol{y}_0 - \boldsymbol{\eta}_0$. По построению $\boldsymbol{\eta}_0$ является случайной величиной. Поэтому, если $\boldsymbol{y}_0$ не полностью коррелировано с $\boldsymbol{\eta}_0$, начальное условие будет случайным.

Пример 10.2 (продолжение): Для двумерной I(1)-системы в примере 10.2 матричная версия $\boldsymbol{\alpha}(L)$ равна $-\boldsymbol{\Psi}_1$, так что стационарная компонента $\boldsymbol{\eta}_t$ в BN-разложении имеет вид:

$$\boldsymbol{\eta}_t = \begin{bmatrix} \varepsilon_{1t} - \gamma\varepsilon_{2t} \\ 0 \end{bmatrix}. \tag{10.1.17}$$

(Вам нетрудно будет проверить это исходя из матричной версии (9.2.5).) Двумерный стохастический тренд записывается как

$$\boldsymbol{\Psi}(1)(\boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_2 + \cdots + \boldsymbol{\varepsilon}_t) = \begin{bmatrix} 0 & \gamma \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \sum_{s=1}^t \varepsilon_{1s} \\ \sum_{s=1}^t \varepsilon_{2s} \end{bmatrix} = \begin{bmatrix} \gamma\cdot\sum_{s=1}^t \varepsilon_{2s} \\ \sum_{s=1}^t \varepsilon_{2s} \end{bmatrix}. \tag{10.1.18}$$

Таким образом, двумерный стохастический тренд приводится в движение одним общим стохастическим трендом, $\sum_{s=1}^t \varepsilon_{2s}$. Это пример представления «общих трендов» в [Stock and Watson, 1993] (см. контрольный вопрос 2 ниже).

Определение коинтеграции

I(1)-система y_t не является стационарной, поскольку содержит стохастический тренд $\Psi(1) \cdot (\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t)$, и, если $\delta \neq 0$, содержит детерминированный тренд $\delta \cdot t$. Детерминированные и стохастические тренды, однако, могут исчезать, если мы берем подходящую линейную комбинацию элементов y_t. Для осуществления этой идеи[1] умножим обе стороны BN-разложения (10.1.16) на некоторый подходящий вектор a, чтобы получить:

$$a'y_t = a'\delta \cdot t + a'\Psi(1)(\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t) + a'\eta_t + a'(y_0 - \eta_0). \quad (10.1.19)$$

Если a удовлетворяет условию

$$a'\Psi(1) = \underset{(1\times n)}{0'}, \quad (10.1.20)$$

то стохастический тренд удаляется и $a'y_t$ принимает вид:

$$a'y_t = a'\delta \cdot t + a'\eta_t + a'(y_0 - \eta_0). \quad (10.1.21)$$

Строго говоря, этот процесс не является стационарным относительно тренда, поскольку начальное условие $a'(y_0 - \eta_0)$ может коррелировать с последующими значениями $a'\eta_t$ (см. контрольный вопрос 3 ниже в качестве иллюстрации). Процесс будет стационарным относительно тренда, если, например, начальное значение y_0 будет таково, что $a'(y_0 - \eta_0) = 0$.

Теперь мы готовы определить коинтеграцию[2].

Определение 10.1 (коинтеграция): Пусть y_t — n-мерная I(1)-система, связанная с которой I(0)-система с нулевым средним u_t удовлетворяет (10.1.1)–(10.1.4). Мы говорим, что система y_t является **«коинтегрированной»**, с n-мерным **«коинтегрирующим вектором»** a, если $a \neq 0$ и $a'y_t$ можно сделать стационарным относительно тренда путем подходящего выбора его начального значения y_0.

Определение коинтеграции, таким образом, хотя и продиктовано логической последовательностью, *не*обязательно означает, что теория коинтеграции требует, чтобы начальное значение y_0 было выбрано так, как указано в определении. Процесс в (10.1.21) не является стационарным, поскольку η_t и η_0 коррелированы. Однако, так как $\eta_t = \alpha(L)\varepsilon_t$ и $\alpha(L)$ абсолютно суммируем, η_t и η_0 будут становиться асимптотически независимыми при росте t. В этом смысле процесс (10.1.21) является

[1] Идея изначально была предложена в: [Granger, 1981]. См. также: [Aoki, 1968] и [Box and Tiao, 1977].

[2] Наше определение то же самое, что и Definition 3.4 в [Johansen, 1995].

«асимптотически стационарным», и асимптотическая стационарность — это все, что нужно для оценивания и получения статистических выводов для коинтегрированных I(1)-систем[1].

С определенной таким образом коинтеграцией мы можем определить следующие связанные понятия:

- (Ранг коинтеграции.) **Ранг коинтеграции** (cointegrating rank) — это число линейно независимых коинтегрирующих векторов, а **коинтеграционное пространство** (*cointegrating space*) — это пространство, натянутое на коинтегрирующие векторы. Как показывает предшествующее обсуждение, n-мерный вектор $\boldsymbol{a}$ является коинтегрирующим вектором тогда и только тогда, когда выполняется (10.1.20). Поскольку если ранг коинтеграции равен h, существует h таких линейно независимых векторов, то из этого вытекает, что
$$\underset{(n\times n)}{\operatorname{rank}\,[\boldsymbol{\Psi}(1)]} = n - h. \qquad (10.1.22)$$
Иными словами, ранг коинтеграции h равен $n - \operatorname{rank}[\boldsymbol{\Psi}(1)]$.

- (Коинтеграция внутри подмножеств $\boldsymbol{y}_t$.) По крайней мере один из элементов коинтегрирующего вектора ненулевой. Предположим, без потери общности, что первый элемент коинтегрирующего вектора отличен от нуля. Тогда мы говорим, что y_{1t} (первый элемент $\boldsymbol{y}_t$) **коинтегрирован с** (*cointegrated with*) $\boldsymbol{y}_{2t}$ (оставшиеся $n-1$ элементов $\boldsymbol{y}_t$), или что y_{1t} является **частью коинтеграционного соотношения** (*part of a cointegrating relationship*). Мы можем также определить коинтеграцию для подмножества $\boldsymbol{y}_t$. Например, $n-1$ переменных в $\boldsymbol{y}_{2t}$ некоинтегрированы, если не существует $(n-1)$-мерного вектора $\boldsymbol{b} \neq \boldsymbol{0}$, такого что (10.1.20) выполняется с $\boldsymbol{a}' = (0, \boldsymbol{b}')$. Следовательно, $\boldsymbol{y}_{2t}$ некоинтегрированы тогда и только тогда, когда последние $n-1$ строк в $\boldsymbol{\Psi}(1)$ являются линейно независимыми.

- (Стохастическая коинтеграция.) Заметим, что детерминированный тренд $\boldsymbol{a}'\boldsymbol{\delta}\cdot t$ не удаляется из $\boldsymbol{a}'\boldsymbol{y}_t$, если только коинтегрирующий вектор не удовлетворяет соотношению:
$$\boldsymbol{a}'\boldsymbol{\delta} = 0. \qquad (10.1.23)$$

Во многих приложениях коинтегрирующий вектор удаляет и стохастический, и детерминированный тренды в (10.1.19). То есть

[1] Различие между стационарностью и асимптотической стационарностью обсуждается в [Lütkepohl, 1993, pp. 348–349].

вектор $\boldsymbol{a}$, который удовлетворяет (10.1.20), обязательно удовлетворяет (10.1.23), так что $\boldsymbol{a}'\boldsymbol{y}_t$, который можно теперь записать как

$$\boldsymbol{a}'\boldsymbol{y}_t = \boldsymbol{a}'\boldsymbol{\eta}_t + \boldsymbol{a}'(\boldsymbol{y}_0 - \boldsymbol{\eta}_0), \tag{10.1.24}$$

является стационарным (не только стационарным относительно тренда) для подходящего выбора $\boldsymbol{y}_0$. Из этого следует, что $\boldsymbol{\delta}$ является линейной комбинацией столбцов $\boldsymbol{\Psi}(1)$, так что

$$\operatorname{rank} \underset{(n\times(n+1))}{\left[\boldsymbol{\delta} \ \vdots \ \boldsymbol{\Psi}(1)\right]} = n - h. \tag{10.1.25}$$

Если не указано обратное, мы будем предполагать, что (10.1.25), так же как и (10.1.22), выполняются, когда ранг коинтеграции равен h. Если мы хотим описать случай, когда коинтегрирующий вектор удаляет стохастический тренд, но необязательно удаляет детерминированный тренд, то будем использовать термин **«стохастическая коинтеграция»** (*stochastic cointegration*). Конечно, если $\boldsymbol{y}_t$ не содержит детерминированный тренд (то есть если $\boldsymbol{\delta} = \boldsymbol{0}$), то отличия между коинтеграцией и стохастической коинтеграцией нет.

Пример 10.2 (продолжение): В двумерной I(1)-системе (10.1.12) в примере 10.2 матрица $\boldsymbol{\Psi}(1)$ задается в (10.1.8), поэтому

$$\left[\boldsymbol{\delta} \ \vdots \ \boldsymbol{\Psi}(1)\right] = \begin{bmatrix} \delta_1 & 0 & \gamma \\ \delta_2 & 0 & 1 \end{bmatrix}.$$

Ранг $\boldsymbol{\Psi}(1)$ равен 1, так что ранг коинтеграции равен 1 ($= 2 - 1$). Все коинтегрирующие векторы можно записать как $(c, -c\gamma)'$, $c \neq 0$. Предположение о том, что коинтегрирующий вектор также удаляет детерминированный тренд, можно записать как $c\delta_1 - c\gamma\delta_2 = 0$, или $\delta_1 = \gamma\delta_2$, из чего следует, что ранг матрицы $[\boldsymbol{\delta} \ \vdots \ \boldsymbol{\Psi}(1)]$, указанной выше, равен единице.

Из определения коинтеграции вытекает довольно много следствий в отношении I(1)-систем. Например:

- ($h < n$.) Для n-мерной I(1)-системы ранг коинтеграции не может быть равным n. Если бы он был равным n, то тогда бы $\operatorname{rank}[\boldsymbol{\Psi}(1)] = 0$ и $\boldsymbol{\Psi}(1)$ была бы матрицей из нулей, что исключено требованием (10.1.4).

- (Следствие положительной определенности долговременной дисперсии $\Delta\boldsymbol{y}_t$.) Долгосрочная ковариационная матрица $\Delta\boldsymbol{y}_t$ задается как $\boldsymbol{\Psi}(1)\boldsymbol{\Omega}\boldsymbol{\Psi}(1)'$ (см. (10.1.5)), которая положительно определена тогда и только тогда, когда $\boldsymbol{\Psi}(1)$ невырождена. Следовательно, $\boldsymbol{y}_t$

некоинтегрирован тогда и только тогда, когда долговременная ковариационная матрица $\Delta \boldsymbol{y}_t$ положительно определена[1]. Поскольку долговременная дисперсия каждого элемента $\Delta \boldsymbol{y}_t$ положительна, если долговременная ковариационная матрица $\Delta \boldsymbol{y}_t$ положительно определена, то из этого следует, что каждая компонента $\boldsymbol{y}_t$ сама по себе является I(1) (то есть является одномерным I(1)-процессом), если $\boldsymbol{y}_t$ некоинтегрирован. То же самое верно для подмножества $\boldsymbol{y}_t$. Например, рассмотрим $\boldsymbol{y}_{2t}$, последние $n-1$ компонент в $\boldsymbol{y}_t$. Долгосрочная ковариационная матрица $\Delta \boldsymbol{y}_{2t}$ задается как $\boldsymbol{\Psi}_2(1)\boldsymbol{\Omega}\boldsymbol{\Psi}_2(1)'$, где $\boldsymbol{\Psi}_2(1)$ — последние $n-1$ строк в $\boldsymbol{\Psi}(1)$. Она положительно определена тогда и только тогда, когда строки $\boldsymbol{\Psi}_2(1)$ линейно независимы или $\boldsymbol{y}_{2t}$ некоинтегрирован. В частности, каждая компонента $\boldsymbol{y}_{2t}$ сама по себе является I(1), если $\boldsymbol{y}_{2t}$ некоинтегрирован.

- Предположим, что y_{1t} коинтегрирован с $\boldsymbol{y}_{2t}$. Тогда $\boldsymbol{y}_{2t}$ некоинтегрирован, если $h = 1$, и коинтегрирован, если $h > 1$. Причина заключается в следующем. Из коинтеграции y_{1t} с $\boldsymbol{y}_{2t}$ следует то, что существует коинтегрирующий вектор, (назовем его $\boldsymbol{a}_1$), первый элемент которого ненулевой. Если $h = 1$, то не должно существовать $(n-1)$-мерного вектора $\boldsymbol{b} \neq \boldsymbol{0}$, такого что $\boldsymbol{a}_2' = (0, \boldsymbol{b}')$ являлся бы коинтегрирующим вектором, поскольку $\boldsymbol{a}_1$ и $\boldsymbol{a}_2$ линейно независимы. Такой вектор $\boldsymbol{a}_2$ можно найти, если $h > 1$.

- (VAR в первых разностях?) Если $\boldsymbol{y}_t$ является стационарным в разностях (без сноса), возникает искушение моделировать его как стационарную VAR конечного порядка $\boldsymbol{\Phi}(L)\Delta \boldsymbol{y}_t = \boldsymbol{\varepsilon}_t$, где $\boldsymbol{\Phi}(L)$ — матричный полином степени p, такой что все корни $|\boldsymbol{\Phi}(z)| = 0$ лежат вне единичного круга. Но тогда $\boldsymbol{y}_t$ не может быть коинтегрированным. Причина заключается в следующем. Если полином $\boldsymbol{\Phi}(L)$ удовлетворяет условию стационарности, то последовательность матриц коэффициентов $\{\boldsymbol{\Psi}_j\}$ обратного полинома $\boldsymbol{\Psi}(L) = \boldsymbol{\Phi}(L)^{-1}$ ограничена геометрически убывающей последовательностью (см. параграф 6.3). Поэтому $\boldsymbol{\Psi}(L)$ является 1-суммируемым, и $\Delta \boldsymbol{y}_t = \boldsymbol{\Psi}(L)\boldsymbol{\varepsilon}_t$ является VMA-процессом, удовлетворяющим (10.1.2) и (10.1.3). Кроме того,
$$|\boldsymbol{\Psi}(1)| = |\boldsymbol{\Phi}(1)^{-1}| = |\boldsymbol{\Phi}(1)|^{-1} \neq 0.$$
Поэтому $\boldsymbol{\Psi}(1)$ невырождена и долговременная ковариационная матрица положительно определена.

[1]Эта эквивалентность специфична для линейных процессов. В общем случае положительная определенность долговременной ковариационной матрицы $\Delta \boldsymbol{y}_t$ достаточна, но не всегда необходима, для того чтобы $\boldsymbol{y}_t$ не был коинтегрирован. См.: [Phillips, 1986, p. 321].

Контрольные вопросы

1. Пусть (y_{1t}, y_{2t}) будет как в примере 10.2 с $\gamma = 0$. Покажите, что $\{y_{1t} + y_{2t}\}$ является I(1). **Указание:** *Вы должны показать, что долговременная дисперсия* $\{\Delta y_{1t} + \Delta y_{2t}\}$ *положительна.*

2. (Представление общих трендов в [Stock and Watson, 1988].) Пусть $\boldsymbol{w}_t \equiv \boldsymbol{\varepsilon}_1 + \boldsymbol{\varepsilon}_2 + \cdots + \boldsymbol{\varepsilon}_t$, так что BN-разложение имеет вид:

$$\boldsymbol{y}_t = \boldsymbol{\delta} \cdot t + \boldsymbol{\Psi}(1)\boldsymbol{w}_t + \boldsymbol{\eta}_t + (\boldsymbol{y}_0 - \boldsymbol{\eta}_0).$$

Результат из линейной алгебры утверждает, что

если $\boldsymbol{C}$ является $n \times n$ матрицей ранга $n - h$, то существует $n \times n$ невырожденная матрица $\boldsymbol{G}$ и $n \times (n-h)$ матрица $\boldsymbol{F}$ полного ранга, такая что

$$\underset{(n\times n)}{\boldsymbol{C}}\underset{(n\times n)}{\boldsymbol{G}} = \Big[\underset{(n\times(n-h))}{\boldsymbol{F}} \;\vdots\; \underset{(n\times h)}{\boldsymbol{0}}\Big].$$

Покажите, что перманентную компоненту $\boldsymbol{\Psi}(1)\boldsymbol{w}_t$ можно записать как $\boldsymbol{F}\boldsymbol{\tau}_t$, где $\boldsymbol{F}$ — $n\times(n-h)$ матрица полного ранга по столбцам, и $\boldsymbol{\tau}_t$ — $(n-h)$-мерное случайное блуждание с положительно определенной $\mathrm{Var}(\Delta\boldsymbol{\tau}_t)$. **Указание:** $\boldsymbol{\Psi}(1)\boldsymbol{w}_t = \boldsymbol{\Psi}(1)\boldsymbol{G}\boldsymbol{G}^{-1}\boldsymbol{w}_t$. *Пусть* $\boldsymbol{\tau}_t$ *является первыми* $n - h$ *элементами в* $\boldsymbol{G}^{-1}\boldsymbol{w}_t$. Это представление делает понятным, что I(1)-система с рангом коинтеграции h имеет $n - h$ общих стохастических трендов.

3. ($\boldsymbol{a}'\boldsymbol{y}_t$ не вполне стационарный.) Чтобы показать, что процесс в (10.1.21) не вполне стационарный, рассмотрим простой случай, где $\boldsymbol{\eta}_t = \boldsymbol{\varepsilon}_t - \boldsymbol{\varepsilon}_{t-1}$, $\boldsymbol{\delta} = \boldsymbol{0}$ и $\boldsymbol{y}_0 = \boldsymbol{0}$. Проверьте, что

$$\mathrm{Cov}(\boldsymbol{\eta}_t, \boldsymbol{\eta}_0) = \begin{cases} 2\boldsymbol{\Omega} & \text{для } t = 0, \\ -\boldsymbol{\Omega} & \text{для } t = 1, \\ 0 & \text{для } t > 1. \end{cases}, \quad \mathrm{Var}(\boldsymbol{a}'\boldsymbol{y}_t) = \begin{cases} 0 & \text{для } t = 0, \\ 6\boldsymbol{a}'\boldsymbol{\Omega}\boldsymbol{a} & \text{для } t = 1, \\ 4\boldsymbol{a}'\boldsymbol{\Omega}\boldsymbol{a} & \text{для } t > 1. \end{cases}$$

где $\boldsymbol{\Omega} = \mathrm{Var}(\boldsymbol{\varepsilon})$. Проверьте, что $\{\boldsymbol{a}'\boldsymbol{y}_t\}$ является стационарным для $t = 2, 3, \ldots$.

4. (Что, если некоторые элементы стационарные?) Пусть $\boldsymbol{y}_t$ является I(1)-системой, и предположим, что первый элемент $\boldsymbol{y}_t$ является стационарным. Покажите, что в этом случае ранг коинтеграции как минимум равен 1.

5. (Матрица коинтегрирующих векторов.) Предположим, что ранг коинтеграции n-мерной I(1)-системы $\boldsymbol{y}_t$ равен h, и пусть $\boldsymbol{A}$ является $n \times h$ матрицей, содержащей h линейно независимых коинтегрирующих векторов. Пусть $\boldsymbol{F}$ является любой $h \times h$ невырожденной матрицей. Покажите, что столбцы $\boldsymbol{AF}$ есть h линейно независимых коинтегрирующих векторов. **Указание:** *Умножение на невырожденную матрицу не изменяет ранга.*

10.2. Альтернативные представления коинтегрированных систем

В дополнение к представлению общих трендов (см. контрольный вопрос 2 в предыдущей главе) существуют три других полезных представления коинтегрированных векторных I(1)-процессов: треугольное представление Филлипса [Phillips, 1991]), VAR-представление и VECM (векторная модель коррекции ошибок, *vector error-correction model*) в [Davidson et al., 1978]. Эти представления вводятся в данном параграфе.

Треугольная система Филлипса

Это представление удобно для оценивания коинтегрирующих векторов. Представление обоснованно для любого ранга коинтеграции, но мы изначально предположим, что ранг коинтеграции h равен 1. Пусть $\boldsymbol{a}$ является коинтегрирующим вектором, и предположим, без потери общности, что первый элемент в $\boldsymbol{a}$ ненулевой (если он равен нулю, то можно изменить порядок переменных в системе, так что первый элемент после переупорядочивания будет ненулевым). Разобьем $\boldsymbol{y}_t$ следующим образом:

$$\underset{(n\times 1)}{\boldsymbol{y}_t} \equiv \begin{bmatrix} y_{1t} \\ \underset{((n-1)\times 1)}{\boldsymbol{y}_{2t}} \end{bmatrix}. \tag{10.2.1}$$

Таким образом, y_{1t} коинтегрирован с $\boldsymbol{y}_{2t}$. Поскольку коинтегрирующий вектор, умноженный на скаляр, также является коинтегрирующим вектором, мы можем нормализовать коинтегрирующий вектор $\boldsymbol{a}$ так, чтобы его первый элемент был равен единице:

$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} 1 \\ \underset{((n-1)\times 1)}{-\boldsymbol{\gamma}} \end{bmatrix}. \tag{10.2.2}$$

Мы видели в предыдущем параграфе, что если коинтегрирующий вектор $\boldsymbol{a}$ удаляет не только стохастический тренд (то есть $\boldsymbol{a}'\boldsymbol{\Psi}(1) = \boldsymbol{0}'$), но также и детерминированный тренд (то есть $\boldsymbol{a}'\boldsymbol{\delta} = 0$), то тогда $\boldsymbol{a}'\boldsymbol{y}_t$ можно записать как (10.1.24). Полагая $\boldsymbol{a}$ в (10.1.24) равным коинтегрирующему вектору (10.2.2), мы получаем:

$$y_{1t} = \boldsymbol{\gamma}'\boldsymbol{y}_{2t} + z_t^* + \mu, \tag{10.2.3}$$

где

$$z_t^* \equiv (1, -\boldsymbol{\gamma}')\boldsymbol{\eta}_t, \quad \mu \equiv (1, -\boldsymbol{\gamma}')(\boldsymbol{y}_0 - \boldsymbol{\eta}_0). \tag{10.2.4}$$

Поскольку η_t совместно стационарный, z_t^* является стационарным. Это уравнение, где значения z_t^* рассматриваются как ошибки, а μ как константа, называют **коинтегрирующей регрессией** (*cointegrating regression*). Ее регрессионные коэффициенты γ, связывающие перманентную компоненту в y_{1t} с перманентными компонентами, входящими в $\boldsymbol{y}_{2t}$, можно интерпретировать как описывающие долговременное соотношение между y_{1t} и $\boldsymbol{y}_{2t}$. Коинтегрирующий вектор будет иметь компоненту тренда $\boldsymbol{a}'\boldsymbol{\delta} \cdot t$ в качестве дополнительного регрессора, если коинтегрирующий вектор не удаляет детерминированный тренд в $\boldsymbol{a}'\boldsymbol{y}_t$. **Треугольное представление** (*triangular representation*) — это система из n уравнений, состоящая из коинтегрирующей регрессии и последних $n-1$ строк в (10.1.9):

$$\Delta \boldsymbol{y}_{2t} = \boldsymbol{\delta}_2 + \boldsymbol{u}_{2t} = \underset{((n-1)\times 1)}{\boldsymbol{\delta}_2} + \underset{((n-1)\times n)}{\boldsymbol{\Psi}_2(L)} \underset{(n\times 1)}{\boldsymbol{\varepsilon}_t}. \tag{10.2.5}$$

где $\boldsymbol{\delta}_2$ и $\boldsymbol{u}_{2t}$ — последние $n-1$ элементов в n-мерном векторе $\boldsymbol{\delta}$ и $\boldsymbol{u}_t$ соответственно, а $\boldsymbol{\Psi}_2(L)$ — последние $n-1$ строк в $\boldsymbol{\Psi}(L)$ в VMA-представлении (10.1.10). Как было отмечено в предыдущем параграфе, следствием предположения о том, что $h = 1$, является то, что $\boldsymbol{y}_{2t}$ некоинтегрирован (он был бы коинтегрированным, если бы $h > 1$). В частности, каждый элемент $\boldsymbol{y}_{2t}$ является I(1).

Рассмотрим общий случай, когда ранг коинтеграции h необязательно равен 1. Изменяя порядок элементов в $\boldsymbol{y}_t$, если это необходимо, можно (см., например: [Hamilton, 1994, pp. 576–577] выбрать h линейно независимых коинтегрирующих векторов, $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_h$, таких что

$$\underset{(n\times h)}{\boldsymbol{A}} \equiv \begin{bmatrix} \boldsymbol{a}_1 & \cdots & \boldsymbol{a}_h \end{bmatrix} = \begin{bmatrix} \underset{(h\times h)}{\boldsymbol{I}_h} \\ \underset{((n-h)\times h)}{-\boldsymbol{\Gamma}} \end{bmatrix}. \tag{10.2.6}$$

Разобьем $\boldsymbol{y}_t$, соответственно, как

$$\underset{(n\times 1)}{\boldsymbol{y}_t} \equiv \begin{bmatrix} \underset{(h\times 1)}{\boldsymbol{y}_{1t}} \\ \underset{((n-h)\times 1)}{\boldsymbol{y}_{2t}} \end{bmatrix}. \tag{10.2.7}$$

Умножая обе стороны BN-разложения (10.1.16) слева на матрицу $\boldsymbol{A}'$ и замечая, что $\boldsymbol{A}'\boldsymbol{\Psi}(1) = \boldsymbol{0}$ (поскольку столбцы в $\boldsymbol{A}$ являются коинтегрирующими векторами), $\boldsymbol{A}'\boldsymbol{\delta} = \boldsymbol{0}$ (если эти коинтегрирующие векторы также удаляют детерминированный тренд) и $\boldsymbol{A}'\boldsymbol{y}_t = \boldsymbol{y}_{1t} - \boldsymbol{\Gamma}'\boldsymbol{y}_{2t}$, мы получим следующие h коинтегрирующих регрессий:

$$\underset{(h\times 1)}{\boldsymbol{y}_{1t}} = \underset{(h\times(n-h))}{\boldsymbol{\Gamma}'} \underset{((n-h)\times 1)}{\boldsymbol{y}_{2t}} + \underset{(h\times 1)}{\boldsymbol{\mu}} + \underset{(h\times 1)}{\boldsymbol{z}_t^*}. \tag{10.2.8}$$

где $\boldsymbol{\mu} \equiv \boldsymbol{A}'(\boldsymbol{y}_0 - \boldsymbol{\eta}_0)$ и $\boldsymbol{z}_t^* \equiv \boldsymbol{A}'\boldsymbol{\eta}$. Поскольку $\boldsymbol{\eta}_t$ совместно стационарный, это выполняется и для $\boldsymbol{z}_t^*$. Треугольное представление — это эти h коинтегрирующих регрессий, дополненные оставшейся частью VMA-представления:

$$\underset{((n-h)\times 1)}{\Delta \boldsymbol{y}_{2t}} = \underset{((n-h)\times 1)}{\boldsymbol{\delta}_2} + \underset{((n-h)\times 1)}{\boldsymbol{u}_{2t}} = \underset{((n-h)\times 1)}{\boldsymbol{\delta}_2} + \underset{((n-h)\times n)}{\boldsymbol{\Psi}_2(L)} \underset{(n\times 1)}{\boldsymbol{\varepsilon}_t}. \tag{10.2.9}$$

Здесь $\boldsymbol{\Psi}_2(L)$ — это последние $n-h$ строк в $\boldsymbol{\Psi}(L)$ в VMA-представлении (10.1.10). Легко показать (контрольный вопрос 1), что $\boldsymbol{y}_{2t}$ некоинтегрирован.

Чтобы проиллюстрировать треугольное представление и то, как его можно получить из VMA-представления, рассмотрим:

Пример 10.3 (из VMA в треугольное): В двумерной системе (10.1.12) примера 10.2 ранг коинтеграции равен 1. Коинтегрирующий вектор, первый элемент которого равен единице, имеет вид: $(1, -\gamma)'$. Так что z_t^* и μ в (10.2.3) имеют вид:

$$\begin{aligned} z_t^* &= (1, -\gamma)\eta_t = \varepsilon_{1t} - \gamma\varepsilon_{2t}, \\ \mu = (1, -\gamma)(y_0 - \eta_0) &= (y_{1,0} - \gamma y_{2,0}) - (\varepsilon_{1,0} - \gamma\varepsilon_{2,0}), \end{aligned} \tag{10.2.10}$$

и треугольное представление имеет вид:

$$\begin{cases} y_{1t} = \mu + \gamma y_{2t} + (\varepsilon_{1t} - \gamma\varepsilon_{2t}), \\ \Delta y_{2t} = \delta_2 + \varepsilon_{2t}. \end{cases} \tag{10.2.11}$$

VAR и коинтеграция

В стационарном случае мы нашли полезным и удобным моделировать векторный процесс как VAR конечного порядка. Хотя, как видно выше, коинтегрированную I(1)-систему нельзя представить в виде VAR в разностях конечного порядка, некоторые коинтегрированные системы могут допускать представление VAR конечного порядка в *уровнях*. Поэтому предположим, что коинтегрированную I(1)-систему $\boldsymbol{y}_t$ можно записать как

$$\begin{aligned} \underset{(n\times 1)}{\boldsymbol{y}_t} &= \boldsymbol{\alpha} + \boldsymbol{\delta}\cdot t + \boldsymbol{\xi}_t, \quad \boldsymbol{\Phi}(L)\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_t \\ \underset{(n\times n)}{\boldsymbol{\Phi}(L)} &\equiv \boldsymbol{I}_n - \boldsymbol{\Phi}_1 L - \boldsymbol{\Phi}_2 L^2 - \cdots - \boldsymbol{\Phi}_p L^p. \end{aligned} \tag{10.2.12}$$

Для дальнейшего использования мы удаляем отсюда $\boldsymbol{\xi}$, получая

$$\boldsymbol{y}_t = \boldsymbol{\alpha}^* + \boldsymbol{\delta}^*\cdot t + \boldsymbol{\Phi}_1\boldsymbol{y}_{t-1} + \boldsymbol{\Phi}_2\boldsymbol{y}_{t-2} + \cdots + \boldsymbol{\Phi}_p\boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t. \tag{10.2.13}$$

где

$$\underset{(n\times 1)}{\boldsymbol{\alpha}^*} \equiv (\boldsymbol{\Phi}_1 + 2\boldsymbol{\Phi}_2 + \cdots + p\boldsymbol{\Phi}_p)\boldsymbol{\delta} + \boldsymbol{\Phi}(1)\,\boldsymbol{\alpha}, \quad \underset{(n\times 1)}{\boldsymbol{\delta}^*} \equiv \boldsymbol{\Phi}(1)\,\boldsymbol{\delta}. \tag{10.2.14}$$

Как мы узнаем, что эта VAR в уровнях конечного порядка является коинтегрированной I(1)-системой? Очевидный способ узнать это — получить VMA-представление, $\Delta\boldsymbol{\xi}_t = \boldsymbol{\Psi}(L)\boldsymbol{\varepsilon}_t$ из VAR-представления, $\boldsymbol{\Phi}(L)\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_t$, и посмотреть, удовлетворяет ли $\boldsymbol{\Psi}(L)$ определению коинтегрированной системы. Получить его несколько сложнее, поскольку VMA-представление дается в первых разностях, но это можно сделать довольно легко следующим образом. Беря первую разность от обеих частей $\boldsymbol{\Phi}(L)\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_t$ и замечая, что $\Delta = 1 - L$ и $(1 - L)\boldsymbol{\Phi}(L) = \boldsymbol{\Phi}(L)(1 - L)$, мы получаем:

$$\boldsymbol{\Phi}(L)\Delta\boldsymbol{\xi}_t = (1 - L)\boldsymbol{\varepsilon}_t. \tag{10.2.15}$$

Подставляя $\Delta\boldsymbol{\xi}_t = \boldsymbol{\Psi}(L)\boldsymbol{\varepsilon}_t$ в это уравнение, получаем:

$$\boldsymbol{\Phi}(L)\boldsymbol{\Psi}(L)\boldsymbol{\varepsilon}_t = (1 - L)\boldsymbol{\varepsilon}_t. \tag{10.2.16}$$

Это уравнение должно выполняться для любой реализации $\boldsymbol{\varepsilon}_t$, так что

$$\boldsymbol{\Phi}(L)\boldsymbol{\Psi}(L) = (1 - L)\boldsymbol{I}_n. \tag{10.2.17}$$

Последнее уравнение можно решить относительно $\boldsymbol{\Psi}(L)$, умножая обе части слева на $\boldsymbol{\Phi}(L)^{-1}$, обращение $\boldsymbol{\Phi}(L)$[1]. Это дает: $\boldsymbol{\Psi}(L) = \boldsymbol{\Phi}(L)^{-1}(1 - L)$. Вопрос заключается в том, при каких условиях на $\boldsymbol{\Phi}(L)$ эта последовательность $\boldsymbol{\Psi}(L)$ 1-суммируема и $\operatorname{rank}[\boldsymbol{\Psi}(1)] = n - h$?

Мы можем легко вывести необходимое условие. Полагая $L = 1$ в равенстве (10.2.17), получаем

$$\underset{(n\times n)}{\boldsymbol{\Phi}(1)}\,\underset{(n\times n)}{\boldsymbol{\Psi}(1)} = \underset{(n\times n)}{\mathbf{0}}. \tag{10.2.18}$$

Поскольку ранг $\boldsymbol{\Psi}(1)$ равен $n - h$, когда ранг коинтеграции $\boldsymbol{\xi}_t$ равен h, ранг $\boldsymbol{\Phi}(1)$ не больше, чем h. Как показано ниже, этот ранг фактически равен h. Для формулирования необходимого и достаточного условия пусть $\boldsymbol{U}(L)$ и $\boldsymbol{V}(L)$ — $n \times n$ матричные лаговые полиномы, все корни которых лежат вне единичного круга, и пусть $\boldsymbol{M}(L)$ — матричный полином, имеющий вид:

$$\boldsymbol{M}(L) = \begin{bmatrix} (1 - L)\boldsymbol{I}_{n-h} & \underset{((n-h)\times h)}{\mathbf{0}} \\ \underset{(h\times(n-h))}{\mathbf{0}} & \boldsymbol{I}_h \end{bmatrix}. \tag{10.2.19}$$

[1]Обращение существует, поскольку $\boldsymbol{\Phi}_0 = \boldsymbol{I}_n$ невырождена; см. параграф 6.3. Поскольку мы не предполагаем выполненным условие стационарности $\boldsymbol{\Phi}(L)$, обратный фильтр может не быть абсолютно суммируемым.

То есть первые $n-h$ диагональных элементов диагональной матрицы $\boldsymbol{M}(L)$ равны $(1-L)$, а остальные h диагональных элементов равны единице. Необходимое и достаточное условие для того, чтобы VAR-процесс конечного порядка $\{\boldsymbol{\xi}_t\}$, следующий $\boldsymbol{\Phi}(L)\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_t$, был коинтегрированной I(1)-системой ранга h, состоит в том, что $\boldsymbol{\Phi}(L)$ можно факторизовать как $\boldsymbol{\Phi}(L) = \boldsymbol{U}(L)\boldsymbol{M}(L)\boldsymbol{V}(L)$[1]. Следовательно, все корни $|\boldsymbol{\Phi}(z)| = 0$ лежат на или за пределами единичной окружности, и все те, которые находятся на единичной окружности, являются единичными корнями ($z=1$). Того, что $\boldsymbol{\Phi}(L)$ имеет $n-h$ единичных корней, а остальные корни лежат за пределами единичной окружности, недостаточно; см. пример в контрольном вопросе 4, где $\boldsymbol{\Phi}(z)$ имеет два единичных корня и один корень вне единичной окружности, однако эта система не I(1). $n-h$ единичных корней должны быть расположены в системе особым образом, указываемым разложением $\boldsymbol{\Phi}(z) = \boldsymbol{U}(z)\boldsymbol{M}(z)\boldsymbol{V}(z)$.

Полагая $z=1$ в этой факторизации, мы получаем $\boldsymbol{\Phi}(1) = \boldsymbol{U}(1)\boldsymbol{M}(1)\times$ $\times\boldsymbol{V}(1)$. Поскольку все корни $\boldsymbol{U}(z)$ и $\boldsymbol{V}(z)$ лежат за пределами единичной окружности, $\boldsymbol{U}(1)$ и $\boldsymbol{V}(1)$ невырождены, и ранг $\boldsymbol{\Phi}(1)$ равен рангу $\boldsymbol{M}(1)$, который равен h. То есть

$$\operatorname{rank}[\boldsymbol{\Phi}(1)] = h. \tag{10.2.20}$$

Тогда из основ линейной алгебры следует, что существуют две $n\times h$ матрицы полного столбцового ранга, $\boldsymbol{B}$ и $\boldsymbol{A}$, такие что

$$\underset{(n\times n)}{\boldsymbol{\Phi}(1)} = \underset{(n\times h)}{\boldsymbol{B}}\ \underset{(h\times n)}{\boldsymbol{A}'}. \tag{10.2.21}$$

Это иногда называют **условием пониженного ранга** (*reduced rank condition*). Выбор $\boldsymbol{B}$ и $\boldsymbol{A}$ не единственный; если $\boldsymbol{F}$ — $h\times h$ невырожденная матрица, то $\boldsymbol{B}(\boldsymbol{F}')^{-1}$ и $\boldsymbol{AF}$ вместо $\boldsymbol{B}$ и $\boldsymbol{A}$ также удовлетворяют (10.2.21). Подставляя (10.2.21) в (10.2.18), мы получаем $\boldsymbol{BA}'\boldsymbol{\Psi}(1) = \boldsymbol{0}$. Поскольку $\boldsymbol{B}$ имеет полный столбцовый ранг, из этого уравнения следует, что $\boldsymbol{A}'\boldsymbol{\Psi}(1) = \boldsymbol{0}$. Поэтому столбцы $n\times h$ матрицы $\boldsymbol{A}$ являются коинтегрирующими векторами.

Векторная модель коррекции ошибок (VECM)

Для одномерного случая мы получили расширенную авторегрессию (9.4.3) на с. 621 из AR-уравнения. Ту же самую идею можно применить здесь к VAR. Это вопрос простой алгебры — показать, что VAR-представление $\boldsymbol{\Phi}(L)\boldsymbol{\xi}_t = \boldsymbol{\varepsilon}_t$ в (10.2.12) можно эквивалентно записать как

$$\boldsymbol{\xi}_t = \boldsymbol{\rho}\boldsymbol{\xi}_{t-1} + \boldsymbol{\zeta}_1\Delta\boldsymbol{\xi}_{t-1} + \boldsymbol{\zeta}_2\Delta\boldsymbol{\xi}_{t-2} + \cdots + \boldsymbol{\zeta}_{p-1}\Delta\boldsymbol{\xi}_{t-p+1} + \boldsymbol{\varepsilon}_t, \tag{10.2.22}$$

[1]Этот результат является следствием леммы Сэма Ю, цитированной в: [Engle and Yoo, 1991]. Доступное изложение есть в: [Watson, 1994, pp. 2870–2873].

где

$$\underset{(n\times n)}{\rho} \equiv \Phi_1 + \Phi_2 + \cdots + \Phi_p, \tag{10.2.23}$$

$$\underset{(n\times n)}{\zeta_s} \equiv -(\Phi_{s+1} + \Phi_{s+2} + \cdots + \Phi_p) \text{ для } s = 1, 2, \ldots, p-1. \tag{10.2.24}$$

Вычитая $\boldsymbol{\xi}_{t-1}$ из обеих частей (10.2.22) и замечая, что $\boldsymbol{\rho} - \boldsymbol{I}_n = -(\boldsymbol{I}_n - \boldsymbol{\Phi}_1 - \boldsymbol{\Phi}_2 - \cdots - \boldsymbol{\Phi}_p) = -\boldsymbol{\Phi}(1)$, мы можем переписать (10.2.22) как

$$\begin{aligned}\Delta\boldsymbol{\xi}_t &= -\boldsymbol{\Phi}(1)\boldsymbol{\xi}_{t-1} + \zeta_1\Delta\boldsymbol{\xi}_{t-1} + \zeta_2\Delta\boldsymbol{\xi}_{t-2} + \cdots + \zeta_{p-1}\Delta\boldsymbol{\xi}_{t-p+1} + \boldsymbol{\varepsilon}_t = \\ &= -\boldsymbol{B}\boldsymbol{A}'\boldsymbol{\xi}_{t-1} + \zeta_1\Delta\boldsymbol{\xi}_{t-1} + \zeta_2\Delta\boldsymbol{\xi}_{t-2} + \cdots + \zeta_{p-1}\Delta\boldsymbol{\xi}_{t-p+1} + \boldsymbol{\varepsilon}_t.\end{aligned} \tag{10.2.25}$$

Используя соотношение $\boldsymbol{y}_t = \boldsymbol{\alpha} + \boldsymbol{\delta}\cdot t + \boldsymbol{\xi}_t$, его можно перевести в уравнение для $\boldsymbol{y}_t$:

$$\begin{aligned}\Delta\boldsymbol{y}_t = \boldsymbol{\alpha}^* + \boldsymbol{\delta}^*\cdot t - \boldsymbol{\Phi}(1)\,\boldsymbol{y}_{t-1} + \\ + \zeta_1\Delta\boldsymbol{y}_{t-1} + \cdots + \zeta_{p-1}\Delta\boldsymbol{y}_{t-p+1} + \boldsymbol{\varepsilon}_t =\end{aligned} \tag{10.2.26}$$

$$\begin{aligned}= \boldsymbol{\alpha}^* + \boldsymbol{\delta}^*\cdot t - \underset{(n\times h)}{\boldsymbol{B}}\,\underset{(h\times 1)}{\boldsymbol{z}_{t-1}} + \\ + \zeta_1\Delta\boldsymbol{y}_{t-1} + \cdots + \zeta_{p-1}\Delta\boldsymbol{y}_{t-p+1} + \boldsymbol{\varepsilon}_t,\end{aligned} \tag{10.2.27}$$

где $\boldsymbol{\alpha}^*$ и $\boldsymbol{\delta}^*$ выражаются как в (10.2.14), а $\boldsymbol{z}_t$ задается как

$$\underset{(h\times 1)}{\boldsymbol{z}_t} \equiv \underset{(h\times n)}{\boldsymbol{A}'}\,\underset{(n\times 1)}{\boldsymbol{y}_t}. \tag{10.2.28}$$

Поскольку, как было отмечено выше, $n\times h$ матрица $\boldsymbol{A}$ содержит h коинтегрирующих векторов, $\boldsymbol{z}_t$ является стационарным относительно тренда (с подходящим выбором начального значения $\boldsymbol{y}_0$)[1]. Это представление называется **векторной моделью коррекции ошибок (VECM)** (*vector error-correction model*), или **представлением в форме коррекции ошибок** (*error-correction representation*). Если бы в VECM не было слагаемого $\boldsymbol{B}\boldsymbol{z}_{t-1}$, то процесс, выраженный как VAR в первых разностях,

[1] На тот случай, если вы задаетесь вопросом, почему $\boldsymbol{y}_0$ годится. Действительно, вы можете решить (10.2.27) относительно $\boldsymbol{z}_t$ как

$$\boldsymbol{z}_{t-1} = (\boldsymbol{B}'\boldsymbol{B})^{-1}\boldsymbol{B}'[\boldsymbol{\alpha}^* + \boldsymbol{\delta}^*\cdot t - \Delta\boldsymbol{y}_t + \zeta_1\Delta\boldsymbol{y}_{t-1} + \cdots + \zeta_{p-1}\Delta\boldsymbol{y}_{t-p+1} + \boldsymbol{\varepsilon}_t].$$

Поэтому вы можете подумать, что $\boldsymbol{z}_t$ является стационарным относительно тренда вне зависимости от выбора $\boldsymbol{y}_0$. Строго говоря, это не так, поскольку, в отличие от VMA-представления, $\Delta\boldsymbol{y}_t$ в VAR/VECM-представлении не определен для $t = -1, -2, \ldots$. Сделать $\Delta\boldsymbol{y}_t$ $(t = 1, 2, \ldots)$ стационарным можно только при разумном выборе $\boldsymbol{y}_0$. Этот вопрос упоминается в: [Johansen, 1995, Theorem 4.2].

не был бы коинтегрированным. VECM вмещает в себя h коинтегрирующих соотношений путем включения h линейных комбинаций уровней переменных. Если в коинтегрирующих соотношениях нет временного тренда, то есть если $\boldsymbol{A}'\boldsymbol{\delta} = \mathbf{0}$, то $\boldsymbol{\delta}^* = \boldsymbol{\Phi}(1)\boldsymbol{\delta} = \boldsymbol{B}\boldsymbol{A}'\boldsymbol{\delta} = \mathbf{0}$. Поэтому VAR- и VECM-представления не включают временной тренд, за исключением возможного существования временных трендов в элементах $\boldsymbol{y}_t$.

То, что один и тот же I(1)-процесс имеет VMA-, VAR- и VECM-представления, известно как *Теорема Грейнджера о представлении* (*Granger Representation Theorem*).

Пример 10.4 (из VMA в VAR/VECM): В предыдущем примере мы вывели треугольное представление из VMA-представления (10.1.12). В этом примере мы выведем VAR и VECM-представления из того же самого VMA-представления. Для $\boldsymbol{\Psi}(L)$ в (10.1.12) легко проверить, что (10.2.17) выполняется с

$$\boldsymbol{\Phi}(L) = \boldsymbol{I}_2 - \boldsymbol{\Phi}_1 L, \quad \boldsymbol{\Phi}_1 = \begin{bmatrix} 0 & \gamma \\ 0 & 1 \end{bmatrix}. \tag{10.2.29}$$

Поэтому VMA можно представить как VAR конечного порядка. Для этой VAR мы имеем:

$$\boldsymbol{\Phi}(1) = \boldsymbol{I}_2 - \boldsymbol{\Phi}_1 = \begin{bmatrix} 1 & -\gamma \\ 0 & 0 \end{bmatrix}. \tag{10.2.30}$$

что можно записать как (10.2.21) с

$$\boldsymbol{B} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad \text{и} \quad \boldsymbol{A} = \begin{bmatrix} 1 \\ -\gamma \end{bmatrix} \quad \text{(например).} \tag{10.2.31}$$

Из (10.2.27) и (10.2.28) вытекает, что VECM для такого выбора $\boldsymbol{B}$ и $\boldsymbol{A}$ имеет вид:

$$\begin{bmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{bmatrix} = \begin{bmatrix} \gamma\delta_2 + \alpha_1 - \gamma\alpha_2 \\ \delta_2 \end{bmatrix} + \begin{bmatrix} \delta_1 - \gamma\delta_2 \\ 0 \end{bmatrix} \cdot t - \begin{bmatrix} 1 \\ 0 \end{bmatrix} z_{t-1} + \boldsymbol{\varepsilon}_t.$$

$$z_t \equiv \boldsymbol{A}'\boldsymbol{y}_t = y_{1t} - \gamma y_{2t}. \tag{10.2.32}$$

Детерминированный тренд исчезает, если $\delta_1 = \gamma\delta_2$, то есть если коинтегрирующий вектор также удаляет детерминированный тренд.

ML-процедура Йохансена

В заключение этого параграфа, введя VECM-представление, мы даем основные положения предложенной в [Johansen, 1988] методики оценивания коинтегрированных систем методом максимального правдоподобия (ML). Вернемся к VAR-представлению (10.2.13). Для простоты будем считать, что в коинтегрирующих соотношениях тренд отсутствует, так что $\delta^* = \mathbf{0}$[1]. В параграфе 8.7 мы рассматривали условное ML-оценивание VAR (условное на начальные значения $\boldsymbol{y}$). Если вектор ошибок ε_t является совместно нормальным $N(\mathbf{0}, \boldsymbol{\Omega})$ и если у нас есть выборка из $T + p$ наблюдений $(\boldsymbol{y}_{-p+1}, \boldsymbol{y}_{-p+2}, \ldots, \boldsymbol{y}_T)$, тогда (среднее) логарифмическое правдоподобие для $(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T)$, условное на начальные значения $(\boldsymbol{y}_{-p+1}, \boldsymbol{y}_{-p+2}, \ldots, \boldsymbol{y}_0)$, задается как

$$Q_T(\boldsymbol{\theta}) = -\frac{n}{2}\log(2\pi) + \frac{1}{2}\log(|\boldsymbol{\Omega}^{-1}|) - \frac{1}{2T}\sum_{t=1}^{T}\{(\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t)'\boldsymbol{\Omega}^{-1}(\boldsymbol{y}_t - \boldsymbol{\Pi}'\boldsymbol{x}_t)\}. \tag{10.2.33}$$

где n является размерностью системы (не объемом выборки), $\boldsymbol{\theta} = (\boldsymbol{\Pi}, \boldsymbol{\Omega})$ и

$$\boldsymbol{\Pi}' \equiv \begin{bmatrix} \underset{(n\times 1)}{\boldsymbol{\alpha}^*} & \underset{(n\times n)}{\boldsymbol{\Phi}_1} & \underset{(n\times n)}{\boldsymbol{\Phi}_2} & \cdots & \underset{(n\times n)}{\boldsymbol{\Phi}_p} \end{bmatrix}, \quad \underset{((1+np)\times 1)}{\boldsymbol{x}_t} \equiv \begin{bmatrix} 1 \\ \boldsymbol{y}_{t-1} \\ \boldsymbol{y}_{t-2} \\ \vdots \\ \boldsymbol{y}_{t-p} \end{bmatrix}. \tag{10.2.34}$$

Пусть

$$\underset{(n\times n)}{\zeta_0} \equiv -\boldsymbol{\Phi}(1) = -\left[\boldsymbol{I}_n - (\boldsymbol{\Phi}_1 + \cdots + \boldsymbol{\Phi}_p)\right]. \tag{10.2.35}$$

Тогда условие пониженного ранга состоит в том, что $\boldsymbol{\xi}_0 = -\boldsymbol{B}\boldsymbol{A}'$, и VECM (10.2.26) можно записать как

$$\Delta\boldsymbol{y}_t = \boldsymbol{\alpha}^* + \zeta_0\boldsymbol{y}_{t-1} + \zeta_1\Delta\boldsymbol{y}_{t-1} + \cdots + \zeta_{p-1}\Delta\boldsymbol{y}_{t-p+1} + \varepsilon_t. \tag{10.2.36}$$

Уравнения (10.2.36) и (10.2.24) дают взаимно однозначное отображение между $(\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_p)$ и $(\xi_0, \ldots, \xi_{p-1})$. Таким образом, то же самое среднее логарифмическое правдоподобие $Q_T(\boldsymbol{\theta})$ можно записать в терминах параметров VECM как

$$-\frac{n}{2}\log(2\pi) + \frac{1}{2}\log(|\boldsymbol{\Omega}^{-1}|) - \frac{1}{2T}\sum_{t=1}^{T}\{(\Delta\boldsymbol{y}_t - \widetilde{\boldsymbol{\Pi}}'\widetilde{\boldsymbol{x}}_t)'\boldsymbol{\Omega}^{-1}(\Delta\boldsymbol{y}_t - \widetilde{\boldsymbol{\Pi}}'\widetilde{\boldsymbol{x}}_t)\}. \tag{10.2.37}$$

где

$$\widetilde{\boldsymbol{\Pi}}' \equiv \begin{bmatrix} \underset{(n\times 1)}{\boldsymbol{\alpha}^*} & \underset{(n\times n)}{\zeta_0} & \underset{(n\times n)}{\zeta_1} & \cdots & \underset{(n\times n)}{\zeta_{p-1}} \end{bmatrix}, \quad \underset{((1+np)\times 1)}{\widetilde{\boldsymbol{x}}_t} \equiv \begin{bmatrix} 1 \\ \boldsymbol{y}_{t-1} \\ \Delta\boldsymbol{y}_{t-1} \\ \vdots \\ \Delta\boldsymbol{y}_{t-p+1} \end{bmatrix}. \tag{10.2.38}$$

[1] В качестве подробного рассмотрения временного тренда в VECM см.: [Johansen, 1995, Sections 5.7 and 6.2].

ML-оценка параметров VECM представляет набор $(\alpha^*, \zeta_0, \ldots, \zeta_{p-1}, \mathbf{\Omega})$, который максимизирует эту целевую функцию при ограничениях, связанных с пониженным рангом и состоящих в том, что $\boldsymbol{\xi}_0 = -\boldsymbol{B}\boldsymbol{A}'$ для некоторых $n \times h$ матриц полного ранга $\boldsymbol{A}$ и $\boldsymbol{B}$. Это ограничение вмещает h коинтегрирующих соотношений в I(1)-системе. Очевидно, что максимизированное логарифмическое правдоподобие тем больше, чем больше предполагаемый ранг коинтеграции. Таким образом, мы можем использовать статистику отношения правдоподобий для тестирования нулевой гипотезы $h = h_0$ против альтернативной гипотезы о *более* высоком ранге коинтеграции. В отличие от случая стационарной VAR предельное распределение этой статистики отношения правдоподобий нестандартно. При определенном таким образом ранге коинтеграции h можно получить ML-оценку для h коинтегрирующих векторов как оценку для $\boldsymbol{A}$. Относительно подробного изложения этой процедуры см.: [Hamilton, 1994, Chapter 20], а также [Johansen, 1995, Chapter 6].

Контрольные вопросы

1. (Вектор ошибок в треугольном представлении.) Рассмотрим вектор ошибок $(\boldsymbol{z}_t^*, \boldsymbol{u}_{2t})$ в треугольном представлении (10.2.8) и (10.2.9). Его можно записать как

$$\begin{bmatrix} \underset{(h\times 1)}{\boldsymbol{z}_t^*} \\ \underset{((n-h)\times 1)}{\boldsymbol{u}_{2t}} \end{bmatrix} = \mathbf{\Psi}^*(L)\boldsymbol{\varepsilon}_t \equiv \begin{bmatrix} \underset{(h\times n)}{\mathbf{\Psi}_1^*(L)} \\ \underset{((n-h)\times n)}{\mathbf{\Psi}_2(L)} \end{bmatrix} \boldsymbol{\varepsilon}_t. \tag{10.2.39}$$

Здесь

$$\mathbf{\Psi}_1^*(L) = \underset{(h\times n)}{(\boldsymbol{I}, -\mathbf{\Gamma}')}\ \underset{(n\times n)}{\alpha(L)}, \tag{10.2.40}$$

где $\alpha(L)$ из BN-разложения, а $\mathbf{\Psi}_2(L)$ — это последние $n - h$ строк $\mathbf{\Psi}(L)$ в VMA-представлении.

(a) Покажите, что $\mathbf{\Psi}_2(L)$ имеет полный строковый ранг (то есть $n - h$ строк $\mathbf{\Psi}_2(1)$ линейно независимы). **Указание:** *Предположим, вопреки утверждению, что существует $(n-h)$-мерный вектор $\boldsymbol{b} \neq \mathbf{0}$, такой что $\boldsymbol{b}'\mathbf{\Psi}_2(1) = \mathbf{0}'$. Покажите, что тогда $(\mathbf{0}', \boldsymbol{b}')$ был бы (n-мерным) коинтегрирующим вектором, а ранг коинтеграции был бы не менее $h + 1$.*

(b) Проверьте, что $\mathbf{\Psi}^*(L)$ абсолютно суммируем.

(c) Запишите $\mathbf{\Psi}^*(L) = \mathbf{\Psi}_0^* + \mathbf{\Psi}_1^* L + \mathbf{\Psi}_2^* L^2 + \ldots$. Для двумерного процесса в примере 10.3 проверьте, что $\mathbf{\Psi}_0^*$ не является диагональной матрицей. (Следовательно, даже если элементы $\boldsymbol{\varepsilon}_t$ некоррелированны, $\boldsymbol{z}_t^*$ и $\boldsymbol{u}_{2t}$ могут быть коррелированными.)

2. (От треугольного представления к VMA.) В примере 10.3 начните с треугольного представления (10.2.11) и восстановите VMA-представление. **Указание:** *Возьмите первую разность коинтегрирующей регрессии.*

3. (Альтернативное разложение $\Phi(1)$.) Для двумерной системы в примере 10.4 проверьте, что
$$B = (\gamma, 0)', \quad A = (1/\gamma, -1)'$$
является альтернативным разложением $\Phi(1)$. Запишите VECM, соответствующую этому выбору B и A.

4. (Трехмерная VAR, являющаяся I(2).) Рассмотрим трехмерную VAR, заданную как
$$\begin{cases} y_{1t} = 2y_{1,t-1} - y_{1,t-2} + \varepsilon_{1t} & (\text{то есть } \Delta^2 y_{1t} = \varepsilon_{1t}), \\ y_{2t} = \rho y_{2,t-1} + \varepsilon_{2t} & \text{с } |\rho| < 1, \text{ и} \\ y_{3t} = \varepsilon_{3t}. \end{cases}$$
Проверьте, что это I(2)-система, записывая $\Delta^2 \boldsymbol{y}_t$ как векторный процесс скользящего среднего. Запишите $\Phi(L)$ для этой системы и покажите, что $\Phi(z)$ имеет три корня, два из которых единичные, а один лежит за пределами единичной окружности.

10.3. Тестирование нулевой гипотезы об отсутствии коинтеграции

После того как мы ввели понятие коинтеграции, нам нужно разобраться с двумя проблемами. Первая — как определить ранг коинтеграции, а вторая — как оценить коинтегрирующие векторы и делать статистические выводы на их основе. Первая проблема будет обсуждаться в текущем параграфе, а вторая будет темой следующего. Существует несколько процедур определения ранга коинтеграции. Среди них тест отношения правдоподобий Йохансена [Johansen, 1988]), полученный из процедуры оценивания методом максимального правдоподобия VECM (кратко рассмотренной в конце предыдущего параграфа), и процедура общих трендов Стока и Ватсона [Stock and Watson, 1988]). Эти процедуры позволяют нам тестировать нулевую гипотезу $h = h_0$, где h_0 — некоторое произвольное число между 0 и $n-1$. Мы не рассматриваем эти процедуры[1]. Здесь мы рассматриваем только простой тест, предложенный в [Engle and Granger, 1987] и обобщенный в [Phillips and Ouliaris, 1990]. В этом тесте нулевая гипотеза имеет вид: $h = 0$ (отсутствие коинтеграции), а альтернативная — $h \geq 1$.

Ложные регрессии

Тест, предложенный Энглом и Грейнджером [Engle and Granger, 1987], основывается на OLS-оценивании регрессии
$$y_{1t} = \mu + \boldsymbol{\gamma}' \boldsymbol{y}_{2t} + z_t^*. \tag{10.3.1}$$

[1] См.: [Maddala and Kim, 1998, Section 7] в качестве каталога доступных процедур.

где y_{1t} — первый элемент $\boldsymbol{y}_t$, $\boldsymbol{y}_{2t}$ — вектор остальных $n-1$ элементов, а z_t^* — ошибка.

Эта регрессия была бы коинтегрирующей регрессией, если бы $h=1$, и y_{1t} был частью коинтегрирующего соотношения. Однако при нулевой гипотезе $h=0$ (отсутствие коинтеграции) эта регрессия не представляет коинтегрирующее соотношение. Пусть $(\widehat{\mu}, \widehat{\gamma})$ — OLS-оценки коэффициентов для (μ, γ). Оказывается, что $\widehat{\gamma}$ не дает состоятельных оценок для любых популяционных параметров в системе! Например, даже если y_{1t} не связан с $\boldsymbol{y}_{2t}$ (в том смысле, что Δy_{1t} и $\Delta \boldsymbol{y}_{2s}$ независимы для всех s, t), t- и F-статистики, ассоциированные с OLS-оценками, становятся произвольно большими при росте объема выборки, создавая ложное впечатление, что между y_{1t} и $\boldsymbol{y}_{2t}$ существует тесная связь. Этот феномен, называемый **ложной регрессией** (*spurious regression*), был впервые обнаружен в исследованиях методом Монте-Карло, представленных в [Granger and Newbold, 1974]. [Phillips, 1986] теоретически получил распределение на больших выборках статистик для ложных регрессий. Например, распределение t-значения, деленного на $\sqrt{T}$, сходится к некоторому невырожденному распределению.

Тест на коинтеграцию, основанный на остатках

Тем не менее регрессия (10.3.1) предоставляет полезный метод тестирования нулевой гипотезы об отсутствии коинтеграции, поскольку OLS-остатки, $y_{1t}-\widehat{\mu}-\widehat{\gamma}'\boldsymbol{y}_{2t}$, должны иметь стохастический тренд, если $\boldsymbol{y}_t$ некоинтегрирован, и быть стационарными в противном случае. Энгл и Грейнджер [Engle and Granger, 1987] предложили применять для тестирования нулевой гипотезы об отсутствии коинтеграции ADF t-тест для остатков. Из-за использования остатков этот тест называется **тестом на коинтеграцию, основанным на остатках** (*residual-based test for cointegration*). Асимптотические распределения тестовой статистики для некоторых ведущих тестов на единичный корень были получены теоретически, и (асимптотические) критические значения для них были табулированы в [Phillips and Ouliaris, 1990] и [Hansen, 1992a].

В отличие от одномерных тестов на единичный корень эти асимптотические распределения (и, следовательно, асимптотические критические значения) зависят от размерности n системы. Это происходит потому, что остатки зависят от $(\widehat{\mu}, \widehat{\gamma})$, которые, будучи оценками на основе данных, являются случайными величинами. Здесь мы указываем соответствующие асимптотические критические значения для случая, когда тестом на единичный корень, применяемым к остаткам, является ADF t-тест из утверждения 9.6 для авторегрессий без константы или времени. То есть ADF t-статистика является t-значением коэффициента при x_{t-1} в следующей расширенной авторегрессии, оцениваемой по остаткам:

$$\Delta x_t = (\rho - 1)x_{t-1} + \zeta_1 \Delta x_{t-1} + \cdots + \zeta_p \Delta x_{t-p} + \text{ошибка}, \qquad (10.3.2)$$

где x_t здесь является остатком от регрессии (10.3.1). Включать константу в эту расширенную авторегрессию нет необходимости, поскольку при уже включенной в регрессию константе выборочное среднее остатков гарантированно равно нулю. Также нет необходимости включать время, поскольку переменные в регрессии (10.3.1) неявно или явно включают временные тренды (более подробно об этом см. ниже). Число лагированных изменений, p, должно возрастать к бесконечности с объемом выборки T, но со скоростью более низкой, чем $T^{1/3}$. Более точно[1],

$$p \to \infty, \quad \text{но} \quad \frac{p}{T^{1/3}} \underset{\mathrm{p}}{\to} 0 \quad \text{при} \quad T \to \infty. \qquad (10.3.3)$$

Критические значения те же самые, если вместо ADF t используется тест Филлипса Z_t (см. аналитическое упражнение 6 в предыдущей главе). Надо рассмотреть три случая.

1. $\mathrm{E}(\Delta \mathbf{y}_{2t}) = \mathbf{0}$ и $\mathrm{E}(\Delta y_{1t}) = 0$, так что ни один элемент I(1)-системы не имеет сноса. Соответствующие критические значения приведены в табл. 10.1(a), которая воспроизводит Table IIb в [Phillips and Ouliaris, 1990]. Относительно формулирования условий на VMA-представление, при которых получено асимптотическое распределение, см.: [Hamilton, 1994, Proposition 19.4].

2. $\mathrm{E}(\Delta \mathbf{y}_{2t}) \neq \mathbf{0}$, но $\mathrm{E}(\Delta y_{1t})$ может быть, а может и не быть нулевым. Этот случай обсуждался в [Hansen, 1992a]. Пусть $g(\equiv n - 1)$ является количеством регрессоров помимо константы в регрессии (10.3.1). Некоторые из g регрессоров имеют снос. Поскольку линейные тренды из различных регрессоров можно скомбинировать в один[2], регрессию (10.3.1) можно записать как регрессию y_{1t} на константу, $g - 1$ I(1)-регрессоров без сноса и один I(1)-регрессор со сносом. Теперь, поскольку линейные тренды доминируют над стохастическими трендами (в смысле, точно указанном при обсуждении BN-разложения в предыдущей главе), I(1)-регрессор с трендом ведет себя на больших выборках практически как время. Поэтому остатки получаются «асимптотически теми же самыми»,

---

[1] См.: [Phillips and Ouliaris, 1990, Theorem 4.2]. Глубина запаздываний должна быть случайной величиной, поскольку она может зависеть от данных. Условие то же самое, что и в обобщении Said — Dickey теста ADF в предыдущей главе (см. параграф 9.4), за исключением того, что p здесь может быть случайной величиной.

[2] Например, пусть $n = 3$, так что есть два регрессора, y_{2t} и y_{3t}, с коэффициентами γ_1 и γ_2 соответственно. Если δ_2 и δ_3 являются сносами в y_{2t} и y_{3t} соответственно, то линейные тренды в регрессии (10.3.1) равны $\gamma_1\delta_2 t$ и $\gamma_1\delta_3 t$ и их можно объединить в единственный временной тренд $(\gamma_1\delta_2 + \gamma_2\delta_3)t$.

что и остатки от регрессии с константой, $g - 1$ I(1)-переменными без сноса и времени как регрессоров, в том смысле, что предельное распределение статистики, основанной на остатках от первой регрессии, то же самое, что и для основанной на остатках от второй регрессии.

Таблица 10.1

Критические значения для ADF t-статистики, примененной к остаткам

Оцененная регрессия $y_{1t} = \mu + \gamma' y_{2t}$				
Количество регрессоров, исключая постоянную	**1%**	**2,5%**	**5%**	**10%**
(а) Регрессоры не имеют сноса				
1	−3,96	−3,64	−3,37	−3,07
2	−4,31	−4,02	−3,77	−3,45
3	−4,73	−4,37	−4,11	−3,83
4	−5,07	−4,71	−4,45	−4,16
5	−5,28	−4,98	−4,71	−4,43
(b) Некоторые регрессоры имеют снос				
1	−3,96	−3,67	−3,41	−3,13
2	−4,36	−4,07	−3,80	−3,52
3	−4,65	−4,39	−4,16	−3,84
4	−5,04	−4,77	−4,49	−4,20
5	−5,36	−5,02	−4,74	−4,46

Источник: Для панели (а) — [Phillips and Ouliaris, 1990, Table IIb]. Для панели (b) — первая строка из [Fuller, 1996, Table 10.A.2], остальные строки из [Phillips and Ouliaris, 1990, Table IIc].

Критические значения для ADF t-теста, основанные на последней регрессии, приведены в Table IIc в [Phillips and Ouliaris, 1990]. Следовательно, чтобы найти соответствующее критическое значение, когда регрессия (10.3.1) включает константу и g регрессоров, но не включает время, обратитесь к этой таблице для $g-1$ регрессоров. Если регрессия имеет только один регрессор помимо константы (то есть если $g = 1$), то эта регрессия асимптотически эквивалентна регрессии y_{1t} на константу и время. Для этого случая предельное распределение ADF t-статистики, вычисленной по остаткам, оказывается распределением Дики — Фуллера (DF_t^τ) в утверждении 9.8. В табл. 10.1(b) объединяются эти распределения: для случая единственного I(1)-регрессора ($g = 1$) она показывает ADF t^τ-распределение, а для случая g (>1)-регрессоров она показывает критические значения из Table IIc в [Phillips and Ouliaris, 1990] для $g - 1$ регрессоров. Например, если $g = 2$, то 5%-ное критическое значение равно -3.806, что равно 5%-ному критическому значению в Table IIc в [Phillips and Ouliaris, 1990] для одного регрессора.

3. Остается случай, когда $\mathrm{E}(\Delta \boldsymbol{y}_{2t}) = \mathbf{0}$ и $\mathrm{E}(\Delta y_{1t}) \neq 0$. Поскольку y_{1t} имеет снос, а $\boldsymbol{y}_{2t}$ не имеет, мы должны включить в регрессию (10.3.1) время, чтобы удалить линейный тренд из остатков. Обсуждение для предыдущего случая делает понятным, что ADF t-статистика, вычисленная при использовании остатков из регрессии y_{1t} на константу, g I(1)-регрессоров $\boldsymbol{y}_{2t}$ *и* время, имеет асимптотическое распределение, приведенное в табл. 10.1(b) для $g+1$ регрессоров (или Table IIc в [Phillips and Ouliaris, 1990] для g регрессоров). Например, если $g = 2$, регрессия (10.3.1) включает константу, два I(1)-регрессора и время, то критические значения можно найти из табл. 10.1(b) для *трех* регрессоров.

Несколько комментариев к тестам на коинтеграцию, основанным на остатках.

- (Состоятельность теста.) Альтернативная гипотеза состоит в том, что $\boldsymbol{y}_t$ коинтегрирован (то есть $h \geq 1$). Тест является состоятельным против этой альтернативы при условии, что y_{1t} коинтегрирован с $\boldsymbol{y}_{2t}$[1]. Причина (мы обсудим ее более подробно ниже для случая $h = 1$) заключается в том, что OLS-остатки в регрессии (10.3.1) будут сходиться к стационарному процессу. Однако если y_{1t} является I(1) и не является частью коинтеграционного соотношения, то тест может не иметь мощности против этой альтернативы о коинтеграции, поскольку OLS-остатки, $y_{1t} - \boldsymbol{\gamma}'\boldsymbol{y}_{2t}$, с ненулевым (единичным) коэффициентом при I(1)-переменной y_{1t} не будут сходиться к стационарному процессу. Таким образом, выбор нормализации (того, какую переменную следует использовать как зависимую переменную) имеет значение для состоятельности теста.

- (Нужно ли включать в регрессию время?) Если время включено в регрессию (10.3.1), то снос в y_{1t}, $\mathrm{E}(\Delta y_{1t})$, влияет только на коэффициент при времени, делая численные значения остатков (и, следовательно, ADF t-значения) инвариантными к $\mathrm{E}(\Delta y_{1t})$. Это означает, что случай 3 процедуры, в котором время включается в число регрессоров, можно использовать для случая 1, где $\mathrm{E}(\Delta y_{1t})$ равно нулю. То есть если вы включаете время в регрессию (10.3.1), то соответствующее критическое значение для случая 1 дается в табл. 10.1(b) для $g+1$ регрессоров. Эта процедура обоснованна также для случая 2, поскольку если время включено в дополнение к константе и g I(1)-регрессоров со сносом, то регрессию можно

[1]Для случая $h = 1$ это является следствием Theorem 5.1 в [Phillips and Ouliaris, 1990]. Их примечание (d) к этой теореме показывает, что тест является состоятельным, когда $h > 1$, и, следовательно, $\boldsymbol{y}_{2t}$ коинтегрирован.

переписать как регрессию с константой, g I(1)-регрессорами и временем, которое объединяют сносы в g I(1)-регрессорах. Эта регрессия попадает под случай 3, и применяются критические значения, заданные в табл. 10.1(b) для g+1 регрессоров. Следовательно, когда время включено в регрессию (10.3.1), можно использовать те же самые критические значения вне зависимости от местоположения сносов. Возможным недостатком является пониженная мощность на конечных выборках. Мощность на конечных выборках очевидно ниже, по крайней мере для DGP, изученных в [Hansen, 1992a] в его симуляциях.

- (Выбор глубины запаздываний.) Требование (10.3.3) не дает практического правила на конечных выборках для выбора глубины запаздываний p в расширенной авторегрессии, и ее нужно оценивать по остаткам. Представляется, что в контексте тестов на основе остатков нет работ, сравнимых с работой [Ng and Perron, 1995] для одномерных I(1)-процессов[1]. Обычная практика здесь — действовать так же, как и в одномерном контексте, который использует информационный критерий Акаике (AIC) или байесовский информационный критерий (BIC), называемый также информационным критерием Шварца (SIC), для определения числа лагированных изменений.

- (Рассмотрение для конечных выборок.) Помимо основанного на остатках ADF t-теста, имеется еще ряд тестов для проверки нулевой гипотезы об отсутствии коинтеграции. Таковыми являются тесты, предложенные в [Phillips and Ouliaris, 1990], [Stock and Watson, 1988] и [Johansen, 1988]. [Haug, 1996] приводит недавнее исследование методом Монте-Карло, изучающее свойства в конечных выборках этих и других тестов. Это исследование выявляет компромисс между мощностью и искажениями размера (то есть тесты с наименьшим искажением размера, как правило, имеют малую мощность). Основанный на остатках ADF t-тест с глубиной запаздываний, выбранной по AIC, хотя и менее мощный, чем некоторые другие тесты, имеет наименьшее искажение размера для рассмотренных DGP.

В следующем примере основанный на остатках тест с ADF t-статистикой применяется к отношению потребление-доход:

Пример 10.5 (коинтегрированы ли потребление и доход?): Как уже было отмечено в связи с рис. 10.1(a), логарифм дохода (y_t) и логарифм потребления (c_t), по всей видимости, коинтегрированы, с коинтегрирующим вектором $(1, -1)$. Этот рисунок, однако, несколько обманчив.

[1] Теперь такие работы есть. — *Прим. пер.*

График разности $y_t - c_t$ (которая равна $-\log(1 - s)$, где s — норма сбережения) на рис. 10.1(b) показывает тенденцию к повышению сразу после войны и понижательную тенденцию с середины 1980-х годов. (Последнее является получившим широкую огласку фактом, что норма личных сбережений в США снижается.) Как видно ниже, результаты теста зависят от того, включаются эти периоды или нет. Сначала мы сосредоточимся на периоде с 1950:Q1 по 1986:Q4 (148 наблюдений).

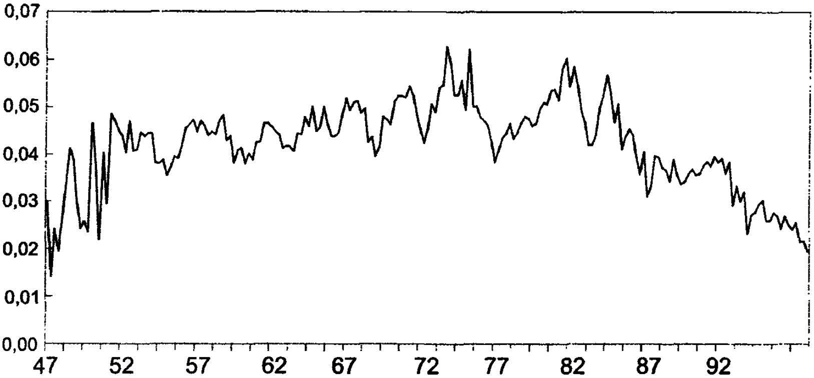

Рис. 10.1(b). Логарифм дохода минус логарифм потребления

Мы сначала проверим, являются ли указанные два ряда индивидуально I(1), проводя ADF t-тест с константой и временным трендом в расширенной авторегрессии. Применяя BIC для выбора количества лагированных изменений в расширенной авторегрессии, мы следуем той же самой практике, что и в предыдущей главе: максимальная глубина ($p_{\max}$) равна $[12 \cdot (T/100)^{1/4}]$ (целая часть $12 \cdot (T/100)^{1/4}$), в процессе выбора глубины запаздываний p период выборки фиксирован и включает моменты времени $t = p_{\max} + 2, p_{\max} + 3, \ldots, T$, а после задания p для оценивания расширенной авторегрессии с p лагированными изменениями используется максимальная выборка $t = p + 2, p + 3 \ldots, T$. Для текущего объема выборки $p_{\max}$ равна 13.

Для располагаемого дохода BIC выбирает глубину запаздываний, равную 0, и ADF t-статистика равна $-1{,}80$. 5%-ное критическое значение из табл. 9.2(с) равно $-3{,}41$, поэтому мы не отвергаем нулевую гипотезу о том, что y_t является I(1). Для потребления глубина запаздываний по BIC равна 1 и ADF t-статистика равна $-2{,}07$. Мы не отвергаем нулевую гипотезу I(1) на 5%-ном уровне. Таким образом, оба ряда можно описать как I(1) со сносом.

Обращаясь к тесту на коинтеграцию, основанному на остатках, находим, что OLS-оценивание статической регрессии дает следующий результат:

$$c_t = \underset{(0{,}009)}{-0{,}046} + \underset{(0{,}0037)}{0{,}975}\, y_t, \quad R^2 = 0{,}998, \quad t = 1950{:}\mathrm{Q}1 \text{ to } 1986{:}\mathrm{Q}4. \tag{10.3.4}$$

Чтобы провести основанный на остатках ADF t-тест, оценивается расширенная авторегрессия без константы или времени по остаткам с глубиной запаздываний, равной 0, согласно BIC. ADF t-статистика равна $-5{,}49$. Поскольку ряды имеют временные тренды, то, для того чтобы найти критические значения, мы обращаемся к табл. 10.1(b), а не к табл. 10.1(a). Для $g = 1$ 5%-ное критическое значение равно $-3{,}41$, поэтому мы можем отвергнуть нулевую гипотезу об отсутствии коинтеграции.

Если основанный на остатках тест проводится по всей выборке с 47:Q1 по 98:Q1, то ADF t-статистика с глубиной запаздываний 1, определенной согласно BIC, равна $-2{,}94$ и, таким образом, мы не можем отвергнуть гипотезу о том, что регрессия потребления на доход является ложной!

Тестирование нулевой гипотезы о коинтеграции

В тестах, рассмотренных выше, коинтеграция берется в качестве альтернативной, а не нулевой, гипотезы. Но очень часто в экономике гипотезой, представляющей экономический интерес, является коинтегрированность рассматриваемых переменных, так что было бы желательно разработать тесты, где нулевой является гипотеза $h = 1$, а не $h = 0$. Совсем недавно было предложено несколько тестов для проверки нулевой гипотезы о наличии коинтеграции. В качестве каталога таких тестов см.: [Maddala and Kim, 1998, Section 4.5]. Как было верно и в отношении тестирования нулевой гипотезы о том, что одномерный процесс является I(0)-процессом, все еще не существует единственного теста на наличие коинтеграции, используемого большинством исследователей[1].

Контрольные вопросы

1. Проверьте, что если рассматривается тест на коинтеграцию, основанный на регрессии (10.3.1), не включающей время в качестве регрессора, то критические значения для случаев 1 и 2 отличаются незначительно.

[1]Использовать для тестирования нулевой гипотезы о наличии коинтеграции против альтернативы об отсутствии коинтеграции процедуры [Johansen, 1988] и [Stock and Watson, 1988] нельзя, поскольку в их тестах альтернативная гипотеза специфицирует *более* высокий ранг коинтеграции, чем нулевая.

10.4. Статистические выводы о коинтегрирующих векторах

В предыдущем параграфе мы проверяли, является ли I(1)-система коинтегрированной. В данном параграфе мы предполагаем, что система коинтегрирована и что ранг коинтеграции и соответствующее треугольное представление известны[1]. Наша цель — оценить коинтегрирующие векторы в треугольном представлении и получить статистические выводы о них. По большей части мы акцентируемся на частном случае, где ранг коинтеграции равен 1; общий случай кратко обсуждается в конце параграфа.

Оценка SOLS

Для случая $h = 1$ треугольное представление имеет вид:

$$\begin{cases} y_{1t} = \mu + \boldsymbol{\gamma}' \boldsymbol{y}_{2t} + z_t^*, \\ \underset{((n-1)\times 1)}{\Delta \boldsymbol{y}_{2t}} = \boldsymbol{\delta}_2 + \boldsymbol{u}_{2t}, \end{cases} \qquad \begin{bmatrix} \underset{(1\times 1)}{z_t^*} \\ \underset{((n-1)\times 1)}{\boldsymbol{u}_{2t}} \end{bmatrix} = \underset{(n\times n)}{\boldsymbol{\Psi}^*(L)} \underset{(n\times 1)}{\boldsymbol{\varepsilon}_t}. \tag{10.4.1}$$

где $\boldsymbol{y}_{2t}$ некоинтегрирован. (Это просто воспроизведение (10.2.3) и (10.2.5) с (10.2.39) для $h = 1$.) Поэтому теперь регрессия (10.3.1) является коинтегрирующей. Следовательно, существует единственный $(n-1)$-мерный вектор $\boldsymbol{\gamma}$, такой что $y_{1t} - \widetilde{\boldsymbol{\gamma}}' \boldsymbol{y}_{2t}$ является стационарным процессом (z_t^*) плюс инвариантная к времени случайная величина (μ), когда $\widetilde{\boldsymbol{\gamma}} = \boldsymbol{\gamma}$, и имеет стохастический тренд, когда $\widetilde{\boldsymbol{\gamma}} \neq \boldsymbol{\gamma}$. Это говорит о том, что OLS-оценка $(\widehat{\mu}, \widehat{\boldsymbol{\gamma}})$ вектора коэффициентов, которая минимизирует сумму квадратов остатков, является состоятельной. Фактически, как показано в [Phillips and Durlauf, 1986] и [Stock, 1987] для случая, в котором $\boldsymbol{\delta}_2$ в (10.4.1) ($= \mathrm{E}(\Delta \boldsymbol{y}_{2t})$) равен нулю, и в [Hansen, 1992a, Theorem 1(b)] для случая $\boldsymbol{\delta}_2 \neq \mathbf{0}$, OLS-оценка $\widehat{\boldsymbol{\gamma}}$ является суперсостоятельной для коинтегрирующего вектора $\boldsymbol{\gamma}$, так что ошибка этой оценки сходится к 0 со скоростью, более высокой, чем обычная скорость $\sqrt{T}$[2]. При этом

[1] На практике мы редко обладаем таким знанием, решение рассматривать систему с h коинтегрирующими соотношениями обычно основывается на результатах предварительных тестов (таких как упомянутых в предыдущем параграфе), предназначенных для определения ранга коинтеграции. Это создает **проблему предварительного тестирования** (*pretest problem*), поскольку распределение оцененных коинтегрирующих векторов не принимает во внимание неопределенность относительно ранга коинтеграции. Эта проблема не изучалась обстоятельно.

[2] Скорость сходимости равна T, если $\boldsymbol{\delta}_2 = \mathbf{0}$, и $T^{3/2}$, если $\boldsymbol{\delta}_2 \neq \mathbf{0}$. Все цитируемые здесь исследования предполагают, что μ является фиксированной константой. То же самое заключение будет выполняться, даже если μ рассматривается как случайная величина. Как было отмечено в параграфе 10.1 (см. абзац прямо под (10.1.21)), строго говоря, процесс $\mu + z_t^*$ является нестационарным даже если z_t^* является стационарным.

R^2 сходится к единице[1]. OLS-оценку для γ из коинтегрирующей регрессии мы будем называть **статической OLS (SOLS)** («*static*» *OLS*) оценкой коинтегрирующего вектора. Поскольку SOLS-оценка состоятельна, остатки сходятся к стационарному процессу с нулевым средним. Таким образом, если одномерный тест на единичный корень, такой как ADF-тест, применяется к остаткам, тест будет отвергать нулевую гипотезу I(1) на больших выборках. Вот почему основанный на остатках тест из предыдущего параграфа является состоятельным против альтернативной гипотезы о коинтеграции.

Этот факт — то, что OLS-оценки коэффициентов являются состоятельными, когда регрессоры являются I(1)-процессами и некоинтегрированы является замечательным результатом, резко контрастирующим со случаем стационарных регрессоров. Чтобы оценить по достоинству это отличие, вспомните из главы 3, что нужно, чтобы получить состоятельную оценку γ, когда регрессоры $\mathbf{y}_{2t}$ являются стационарными: для того чтобы OLS-оценка была состоятельной, ошибки z_t^* не должны коррелировать с $\mathbf{y}_{2t}$; в противном случае нам нужны инструментальные переменные для $\mathbf{y}_{2t}$. В отличие от этого, если y_{1t} коинтегрирован с $\mathbf{y}_{2t}$ и если I(1)-регрессоры $\mathbf{y}_{2t}$ некоинтегрированы, как здесь, мы не должны беспокоиться о смещении из-за одновременности, по крайней мере на больших выборках, даже если ошибки z_t^* и I(1)-регрессоры коррелированы[2].

Однако в конечных выборках смещение SOLS-оценки (разность между математическим ожиданием оценки и истинным значением коэффициента) может быть большим, что было отмечено в [Banerjee et al., 1986] и [Stock, 1987]. Другой недостаток SOLS состоит в том, что асимптотическое распределение соответствующей t-статистики зависит от мешающих параметров (которыми являются коэффициенты в $\mathbf{\Psi}^*L$ в (10.4.1)), что затрудняет статистические выводы. Позднее в этом параграфе мы введем другую оценку (называемую DOLS-оценкой), которая является эффективной, и при этом связанные с ней тестовые статистики (такие как t-статистика и статистика Вальда) имеют обычные асимптотические распределения.

Однако, поскольку μ и z_t^* асимптотически независимы при $t \to \infty$, процесс $\mu + z_t^*$ асимптотически стационарный, и это все, что требуется здесь для асимптотических результатов.

[1] Относительно формулирования условий, при которых эти результаты выполняются, см.: [Hamilton 1994, Proposition 19.2]. Эти условия являются ограничениями на $\mathbf{\Psi}^*(L)\varepsilon_t$ в (10.4.1).

[2] Если $\mathbf{y}_{2t}$ коинтегрированы, то $h \geq 2$ и регрессия (10.3.1) — с $n-1$ регрессорами — не будет более коинтегрирующей регрессией (коинтегрирующая регрессия должна иметь $n-h$ регрессоров, см. треугольное представление (10.2.8)). В этом случае можно применить общую методологию, представленную в [Sims, Stock and Watson, 1990]. По поводу доступного изложения этой методологии см.: [Hamilton, 1994, Chapter 18] или [Watson, 1994, Section 2].

Двумерный пример

Чтобы прояснить источник и природу корреляции между ошибками и I(1)-регрессорами, а также для того, чтобы проложить путь к введению DOLS-оценки, рассмотрим двумерную версию (10.4.1). Предположим на время, что $\mathbf{\Psi}^*(L) = \mathbf{\Psi}_0^*$ (так что отсутствует сериальная корреляция в (z_t^*, u_{2t})), $\delta_2 = 0$ и $\mu = 0$. Треугольное представление можно записать как

$$\begin{cases} y_{1t} = \gamma y_{2t} + z_t^*, \\ \Delta y_{2t} = u_{2t}, \end{cases} \qquad \begin{bmatrix} z_t^* \\ u_{2t} \end{bmatrix} = \mathbf{\Psi}_0^* \boldsymbol{\varepsilon}_t. \tag{10.4.2}$$

Ошибка z_t^* и I(1)-регрессор y_{2t} в коинтегрирующей регрессии в (10.4.2) могут быть коррелированными, поскольку

$$\begin{aligned} \operatorname{Cov}(y_{2t}, z_t^*) &= \operatorname{Cov}(y_{2,0} + \Delta y_{2,1} + \Delta y_{2,2} + \cdots + \Delta y_{2,t}, z_t^*) = \\ &= \operatorname{Cov}(\Delta y_{2,1} + \Delta y_{2,2} + \cdots + \Delta y_{2t}, z_t^*) \\ &\quad (\text{так как } \operatorname{Cov}(y_{2,0}, z_t^*) = 0) \\ &= \operatorname{Cov}(u_{2,1} + u_{2,2} + \cdots + u_{2t}, z_t^*) \quad (\text{так как } \Delta y_{2t} = u_{2t}) = \\ &= \operatorname{Cov}(u_{2t}, z_t^*) \quad (\text{так как } (u_{2t}, z_t^*) \text{ являются i.i.d.}). \end{aligned} \tag{10.4.3}$$

Поскольку требования диагональности $\mathbf{\Psi}_0^*$ и $\mathbf{\Omega}$ не было, z_t^* и u_{2t} могут быть одномоментно коррелированными.

Чтобы изолировать эту возможную корреляцию, рассмотрим проекцию наименьших квадратов z_t^* на константу и u_{2t} ($= \Delta y_{2t}$). Вспомним из параграфа 2.9, что $\widehat{\mathrm{E}}^*(y|1, x) = [\mathrm{E}(y) - \beta_0 \mathrm{E}(x)] + \beta_0 x$, где $\beta_0 = \operatorname{Cov}(x, y)/\operatorname{Var}(x)$ и что ошибка проекции наименьших квадратов, $y - \widehat{\mathrm{E}}^*(y|1, x)$, имеет нулевое среднее и не коррелирована с x. Здесь среднее равно нулю и для z_t^*, и для u_{2t}. Поэтому $\widehat{\mathrm{E}}^*(z_t^*|1, u_{2t}) = \beta_0 u_{2t}$, и если мы обозначаем ошибку проекции метода наименьших квадратов как $v_t = z_t^* - \widehat{\mathrm{E}}^*(z_t^*|1, u_{2t})$, то имеем:

$$\begin{gathered} z_t^* = \beta_0 u_{2t} + v_t = \beta_0 \Delta y_{2t} + v_t, \\ \mathrm{E}(v_t) = 0, \quad \operatorname{Cov}(u_{2t}, v_t) = \operatorname{Cov}(\Delta y_{2t}, v_t) = 0. \end{gathered} \tag{10.4.4}$$

Подставляя это в коинтегрирующую регрессию, получаем:

$$y_{1t} = \gamma y_{2t} + \beta_0 \Delta y_{2t} + v_t. \tag{10.4.5}$$

Эта регрессия будет называться **расширенной коинтегрирующей регрессией** (*augmented cointegrating regression*). Сейчас мы покажем, что I(1)-регрессор y_{2t} является *строго экзогенным* в том смысле, что для всех t, s $\operatorname{Cov}(y_{2s}, v_t) = 0$. По построению, $\operatorname{Cov}(\Delta y_{2t}, v_t) = 0$. При $t \neq s$ для $\operatorname{Cov}(\Delta y_{2s}, v_t)$ имеем:

$$\begin{aligned} \operatorname{Cov}(\Delta y_{2s}, v_t) &= \operatorname{Cov}(u_{2s}, v_t) = \operatorname{Cov}(u_{2s}, z_t^* - \beta_0 u_{2t}) = \\ &= \operatorname{Cov}(u_{2s}, z_t^*) - \beta_0 \operatorname{Cov}(u_{2s}, u_{2t}) = 0. \end{aligned} \tag{10.4.6}$$

поскольку (z_t^*, u_{2t}) является i.i.d. Поэтому Δy_{2t} является строго экзогенным. Строгая экзогенность Δy_{2t} влечет то же самое и для y_{2t}, поскольку

$$\begin{aligned}\operatorname{Cov}(y_{2s}, v_t) &= \operatorname{Cov}(y_{2,0} + \Delta y_{2,1} + \Delta y_{2,2} + \cdots + \Delta y_{2,t}, v_t) = \\ &= \operatorname{Cov}(\Delta y_{2,1} + \Delta y_{2,2} + \cdots + \Delta y_{2s}, v_t) \\ &\text{(поскольку } \operatorname{Cov}(y_{2,0}, v_t) = 0) \\ &= 0 \quad \text{для всех } t, s \text{ по (10.4.6).}\end{aligned} \tag{10.4.7}$$

Продолжение двумерного примера

Резюмируя, мы показали, что в расширенной коинтегрирующей регрессии (10.4.5) I(1)-регрессор является строго экзогенным, если $\boldsymbol{\Psi}^*(L) = \boldsymbol{\Psi}_0^*$. Пусть $(\widehat{\gamma}, \widehat{\beta}_0)$ является OLS-оценкой для (γ, β_0) из расширенной коинтегрирующей регрессии. (Эту $\widehat{\gamma}$ следует отличать от SOLS-оценки для γ в коинтегрирующей регрессии в (10.4.2).) Эта регрессия, (10.4.5), очень похожа на расширенную авторегрессию (9.4.6) в том, что один из двух регрессоров является I(0) с нулевым средним, а другой — I(1) без сноса. Мы показали для расширенной авторегрессии, что матрица «$\boldsymbol{X}'\boldsymbol{X}$», если она правильно шкалирована на T и $\sqrt{T}$, является асимптотически диагональной, так что наличие I(0)-регрессоров можно игнорировать при выводе предельного распределения $\widehat{\gamma}$, OLS-оценки коэффициента при I(1)-регрессоре. То же самое верно и здесь, и та же самая аргументация, использующая асимптотическую диагональность шкалированной соответствующим образом матрицы «$\boldsymbol{X}'\boldsymbol{X}$», показывает, что обычное t-значение для тестирования гипотезы $\gamma = \gamma_0$ асимптотически эквивалентно:

$$\widetilde{t} = \frac{\dfrac{1}{T}\sum\limits_{t=1}^{T} y_{2t} v_t}{\sqrt{\dfrac{\sigma_v^2}{T^2}\sum\limits_{t=1}^{T}(y_{2t})^2}}, \tag{10.4.8}$$

где σ_v^2 является дисперсией v_t. То есть разность между обычным t-значением и $\widetilde{t}$ в (10.4.8) сходится к нулю по вероятности, и поэтому предельные распределения t и $\widetilde{t}$ одинаковы.

В расширенной авторегрессии, используемой для ADF-теста, предельное распределение ADF t-статистики было нестандартным (оно являлось распределением Дики — Фуллера). В отличие от этого асимптотическое распределение $\widetilde{t}$ (и, следовательно, обычного t-значения) является *стандартным нормальным*. Чтобы увидеть, почему это так, вспомним, что I(1)-регрессор y_{2t} является строго экзогенным в том смысле, что $\operatorname{Cov}(y_{2s}, v_t) = 0$ для всех t, s. Чтобы развить интуицию, *временно* предположим, что (z_t^*, u_{2t}) совместно нормальны. Тогда y_{2s} и v_t являются независимыми для всех (s, t), а не просто некоррелированными. Следовательно, распределение v_t, условное относительно $(y_{2,1}, y_{2,2}, \ldots, y_{2T})$, то же самое, что и безусловное распределение, которое есть $N(0, \sigma_v^2)$. Поэтому распределение числителя (10.4.8), условное относительно $(y_{2,1}, \ldots, y_{2T})$, имеет вид:

$$N\left(0, \frac{\sigma_v^2}{T^2}\sum_{t=1}^{T}(y_{2t})^2\right).$$

Но стандартное отклонение нормального распределения равно знаменателю $\tilde{t}$, и поэтому

$$\tilde{t} \mid (y_{2,1}, y_{2,2}, \ldots, y_{2T}) \sim N(0,1).$$

Поскольку это условное распределение $\tilde{t}$ не зависит от $(y_{2,1}, y_{2,2}, \ldots, y_{2T})$, *без*условным распределением $\tilde{t}$ также будет $N(0,1)$. (Этот способ доказательства не является новым для вас; см. абзац, содержащий (1.4.3), в главе 1). Вспомним из главы 2, что, даже если предположение нормальности отбросить, распределение обычного t-значения является стандартным нормальным, хотя и асимптотически, на больших выборках. Тот же самый вывод верен и здесь: предельное распределение $\tilde{t}$ есть $N(0,1)$, даже если (z_t^*, u_{2t}) не являются совместно нормальными. Доказательство этого требует аргументов, отличных от тех, которые были использованы в главе 2, поскольку y_{2t} здесь имеет стохастический тренд. Мы не будем давать здесь формальное доказательство: только краткое резюме. Доказательство состоит из двух частей. Первая заключается в том, чтобы получить предельное распределение числителя (10.4.8) (что можно сделать, используя результат, сформулированный, например, Proposition 18.1(e) в [Hamilton, 1994] или Lemma 2.3(c) в [Watson, 1994]. Это распределение является нестандартным. Вторая состоит в том, чтобы показать, что нормализация (10.4.8) преобразует нестандартное распределение в нормальное распределение (Lemma 5.1 в [Park and Phillips, 1988]).

Этот пример двумерной I(1)-системы является частным в нескольких отношениях: (a) нет сериальной корреляции в процессе ошибок (z_t^*, u_{2t}), (b) I(1)-регрессор y_{2t} является скаляром, (c) y_{2t} не имеет сноса, и (d) $\mu = 0$. Конечно, ослабление (a) требует некоторого размышления.

Допущение сериальной корреляции

Если предположение $\Psi^*(L) = \Psi_0^*$, сделанное в (10.4.2), не выполняется, то (z_t^*, u_{2t}) является сериально коррелированным. Следовательно, I(1)-регрессор y_{2t} в расширенной коинтегрирующей регрессии (10.4.5) больше не будет строго экзогенным, поскольку $\mathrm{Cov}(\Delta_{2s}, v_t)$, оставаясь нулевой для $t = s$, больше не является гарантированно нулевой для $t \neq s$. Чтобы устранить эту корреляцию, рассмотрим проекцию метода наименьших квадратов z_t^* на текущие, прошлые и будущие значения u_2, а не только на текущие значения u_2. Замечая, что $\Delta y_{2t} = u_{2t}$, эту проекцию можно записать как

$$z_t^* = \beta(L)\Delta y_{2t} + v_t, \quad \beta(L) \equiv \sum_{j=-\infty}^{\infty} \beta_j L^j. \tag{10.4.9}$$

(Это v_t отличается от v_t в (10.4.5).) По построению, для всех t, s имеем: $\mathrm{E}(v_t) = 0$ и $\mathrm{Cov}(\Delta y_{2s}, v_t) = 0$. Двухсторонний фильтр $\beta(L)$ может иметь бесконечный порядок, но предположим, на данный момент, что это не так и $\beta_j = 0$ для $|j| > p$. Таким образом,

$$\begin{aligned} z_t^* = {}&\beta_0\Delta y_{2t} + \beta_{-1}\Delta y_{2,t+1} + \dots + \beta_{-p}\Delta y_{2,t+p} + \\ &+ \beta_1\Delta y_{2,t-1} + \dots + \beta_p\Delta y_{2,t-p} + v_t. \end{aligned} \tag{10.4.10}$$

Подставляя это в коинтегрирующую регрессию в (10.4.2), мы получаем (значительно) расширенную коинтегрирующую регрессию:

$$\begin{aligned} y_{1t} = {}&\gamma y_{2t} + \beta_0\Delta y_{2t} + \beta_{-1}\Delta y_{2,t+1} + \dots + \beta_{-p}\Delta y_{2,t+p} + \\ &+ \beta_1\Delta y_{2,t-1} + \dots + \beta_p\Delta y_{2,t-p} + v_t. \end{aligned} \tag{10.4.11}$$

Поскольку $\mathrm{Cov}(\Delta y_{2s}, v_t) = 0$ для всех t и s, Δy_{2t} является строго экзогенным, что означает (как было показано выше), что регрессор в уровнях y_{2t} также является строго экзогенным. Таким образом, включая в расширенную коинтегрирующую регрессию не только текущее значение, но также прошлые и будущие значения I(1)-регрессора, мы в состоянии сохранить строгую экзогенность y_{2t}. OLS-оценку коинтегрирующего вектора γ, основанную на расширенной коинтегрирующей регрессии, называют **динамической OLS (DOLS)** (*«dynamic» OLS*) оценкой, чтобы отличить ее от SOLS-оценки, основанной на коинтегрирующей регрессии без приращений в y_2.

В расширенной регрессии имеется $2+2p$ регрессоров, первый из которых есть I(1) без сноса, а остальные — I(0) с нулевым средним. Как и выше, при надлежащем шкалировании на T и $\sqrt{T}$, I(1)-регрессор асимптотически не коррелирован с $2p+1$ I(0)-регрессорами с нулевым средним, так что соответствующим образом шкалированная матрица $\boldsymbol{X}'\boldsymbol{X}$ снова является блочно диагональной, и I(0)-регрессоры с нулевым средним можно игнорировать при выводе предельного распределения DOLS-оценки для γ. Следовательно, выражение для случайной величины, которая асимптотически эквивалентна t-значению для $\gamma = \gamma_0$, снова задается как (10.4.8), выражение, полученное для случая, в котором (z_t^*, u_{2t}) сериально некоррелированы.

Однако, когда (z_t^*, u_{2t}) сериально коррелированы, вывод асимптотического распределения (10.4.8) отличается от случая, когда (z_t^*, u_{2t}) сериально некоррелированы, поскольку ошибки v_t могут быть теперь сериально коррелированными; проекция (10.4.9), устраняя корреляцию между

Δy_{2s} и v_t для всех s, t, не устраняет собственную сериальную корреляцию в v_t. С целью изучения асимптотического распределения, пусть $\boldsymbol{V}$ $(T \times T)$ является автоковариационной матрицей T последовательных значений v_t, а λ_v^2 является долговременной дисперсией v_t. Также пусть $\boldsymbol{X}$ в оставшейся части этого подпараграфа является $T \times 1$ матрицей $(y_{2,1}, y_{2,2}, \ldots, y_{2T})'$. И опять, чтобы развить интуицию, временно предположим, что (z_t^*, u_{2t}) является совместно нормальным. Тогда, поскольку y_{2t} строго экзогенный, распределение числителя в (10.4.8), условное относительно $\boldsymbol{X}$, является нормальным с нулевым средним и условной дисперсией

$$\frac{1}{T^2}\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X}. \tag{10.4.12}$$

Чтобы сделать отношение стандартным нормальным на конечных выборках, пришлось бы заменить знаменатель в (10.4.8) на квадратный корень из этой величины. К счастью, оказывается (см.: Corollary 2.7 в [Phillips, 1988]), что на больших выборках распределение числителя является тем же самым, что и распределение, которое вы бы получили, если бы v_t был бы сериально некоррелированным, но с дисперсией λ_v^2, а не σ_v^2.

Следовательно, все, что требуется, чтобы модифицировать отношение (10.4.8) для учета сериальной корреляции в v_t, — это заменить σ_v^2 $(\equiv \mathrm{Var}(v_t))$ на λ_v^2 (долговременная дисперсия v_t), а именно если $\widetilde{t}$ задается как (10.4.8) с σ_v^2, все еще находящейся в знаменателе, то шкалированное значение

$$\left(\frac{\sigma_v}{\lambda_v}\right) \cdot \widetilde{t} \tag{10.4.13}$$

распределено асимптотически как $N(0, 1)$! Поскольку OLS t-значение для γ асимптотически эквивалентно $\widetilde{t}$, то из этого следует, что

$$\left(\frac{s}{\widehat{\lambda}_v}\right) \cdot t \underset{\mathrm{d}}{\rightarrow} N(0, 1), \tag{10.4.14}$$

где $\widehat{\lambda}_v$ является некоторой состоятельной оценкой для λ_v, а s — обычная стандартная ошибка OLS регрессии (легко показать, что s является состоятельной для σ_v). Иными словами, если мы шкалируем обычную стандартную ошибку для $\widehat{\gamma}$ (OLS-оценка коэффициента при y_{2t} в (10.4.11)) как

$$\text{шкалированная стандартная ошибка} =$$
$$= \left(\frac{\widehat{\lambda}_v}{s}\right) \times \text{обычная стандартная ошибка}, \tag{10.4.15}$$

то t-значение, основанное на шкалированной стандартной ошибке, является асимптотически $N(0, 1)$.

Приведенная аргументация применима только к коэффициенту при I(1)-регрессоре, поэтому t-значения для значений β в (10.4.11), даже когда их стандартные ошибки шкалированы так, как только что было описано, необязательно являются $N(0,1)$ на больших выборках.

Поскольку регрессоры строго экзогенны, обсуждение в параграфах 1.6 и 6.7 наводит на мысль, что можно применить GLS. Однако из отмеченного выше факта, что $\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X}$ ведет себя асимптотически как $\lambda_v\boldsymbol{X}'\boldsymbol{X}$, следует, что GLS-оценка для γ (называемая **DGLS**-оценкой) асимптотически эквивалентна DOLS-оценке (или более точно: T, умноженное на их разность, сходится к 0 по вероятности) (см.: [Phillips, 1988] и [Phillips and Park, 1988]). Следовательно, выигрыша в эффективности от коррекции на сериальную корреляцию в ошибках v_t посредством GLS нет.

Состоятельную оценку для λ_v получить несложно. Вспомним, что в параграфе 6.6 мы использовали для оценивания долговременной ковариационной матрицы векторного процесса процедуру VARHAC. Эту же самую процедуру можно применить в данном случае скалярного процесса ошибок. Рассмотрим подбор AR(p)-процесса к остаткам $\widehat{v}_t$ от расширенной коинтегрирующей регрессии

$$\widehat{v}_t = \phi_1\widehat{v}_{t-1} + \phi_2\widehat{v}_{t-2} + \cdots + \phi_p\widehat{v}_{t-p} + e_t \quad (t = p+1, \ldots, T). \quad (10.4.16)$$

(p здесь не следует путать с p в расширенной коинтегрирующей регрессии.) Для выбора глубины запаздываний используем BIC посредством поиска по всем возможным количествам лагов $0, 1, \ldots, T^{1/3}$. Используя соотношение (6.2.21) на с. 415 и замечая, что долговременная дисперсия равна значению производящей функции автоковариаций при $z = 1$, долговременную дисперсию λ_v^2 для v_t можно оценить как

$$\widehat{\lambda}_v^2 = \frac{\widehat{\sigma}_e^2}{(1 - \widehat{\phi}_1 - \cdots - \widehat{\phi}_p)^2}, \quad \widehat{\sigma}_e^2 = \frac{1}{T-p}\sum_{t=p+1}^{T}\widehat{e}_t^2. \quad (10.4.17)$$

где $\widehat{\phi}_j$ ($j = 1, 2, \ldots, p$) — оцененные AR(p) коэффициенты, а $\widehat{e}_t$ — остатки от AR(p) уравнения (10.4.16).

Общий случай

Обратимся теперь к общему случаю n-мерной коинтегрированной системы с возможно ненулевым сносом. Мы фокусируемся на случае $h = 1$, поскольку обобщение на случай, в котором $h > 1$, следует непосредственно. Так что треугольное представление имеет вид (10.4.1), а расширенная коинтегрирующая регрессия имеет вид:

$$\begin{aligned} y_{1t} = \mu + \boldsymbol{\gamma}'\boldsymbol{y}_{2t} + \boldsymbol{\beta}_0'\Delta\boldsymbol{y}_{2t} + \boldsymbol{\beta}_{-1}'\Delta\boldsymbol{y}_{2,t+1} + \cdots + \boldsymbol{\beta}_{-p}'\Delta\boldsymbol{y}_{2,t+p} + \\ + \boldsymbol{\beta}_1'\Delta\boldsymbol{y}_{2,t-1} + \cdots + \boldsymbol{\beta}_p'\Delta\boldsymbol{y}_{2,t-p} + v_t. \quad (10.4.18) \end{aligned}$$

Здесь β_j ($j = 0, 1, 2, \ldots, p, -1, -2, \ldots, -p$) являются коэффициентами проекции метода наименьших квадратов в проекции z_t^* (ошибка в коинтегрирующей регрессии) на текущие, прошлые и будущие значения $\Delta \boldsymbol{y}_2$. DOLS-оценка коинтегрирующего вектора γ — это просто OLS-оценка для γ из этой расширенной коинтегрирующей регрессии.

Следующие результаты доказаны в [Saikkonen, 1991] и [Stock and Watson, 1993][1].

1. Описанные выше результаты для двумерного случая остаются в силе. То есть DOLS-оценка для γ является суперсостоятельной, и надлежащим образом шкалированные, t-статистика и статистика Вальда для тестирования гипотез относительно γ имеют обычные асимптотические распределения (стандартное нормальное и хи-квадрат). Правильное шкалирование — умножить обычное t-значение на $(s/\widehat{\lambda}_v)$ и статистику Вальда на квадрат $(s/\widehat{\lambda}_v)$. Однако это простое шкалирование не работает для t-значения и статистики Вальда в случае тестирования гипотез, включающих μ или $\boldsymbol{\beta}_j$.

2. Если двухсторонний фильтр $\boldsymbol{\beta}(L)$ в проекции z_t^* на опережающие и запаздывающие значения $\Delta \boldsymbol{y}_{2t}$ имеет бесконечный порядок, то тогда в состав ошибки v_t входит остаток от усечения,

$$\sum_{j=p+1}^{\infty} \boldsymbol{\beta}'_{-j} \Delta \boldsymbol{y}_{2,t+j} + \sum_{j=p+1}^{\infty} \boldsymbol{\beta}'_j \Delta \boldsymbol{y}_{2,t-j}.$$

   Все результаты, только что упомянутые выше, выполняются при условии, что p в (10.4.18) полагается таким, чтобы его значение возрастало с увеличением объема выборки T со скоростью, меньшей, чем $T^{1/3}$. См.: [Saikkonen, 1991, Section 4] относительно формулирования требуемых условий регулярности.

3. Рассматриваемая оценка является эффективной в некотором точном смысле[2]. Она асимптотически эквивалентна другим эффективным оценкам, таким как процедура метода максимального правдоподобия Йохансена, основанная на VECM, **«полностью модифицированная»** оценка (*Fully Modified estimator*) Филлипса и Хансена

[1] Эти авторы предполагают, что μ является фиксированной константой. Однако те же самые выводы сохраняются, даже если μ является инвариантной ко времени случайной величиной.

[2] t-значение является асимптотически $N(0, 1)$, но сама DOLS-оценка не является асимптотически нормальной. Поэтому обычный критерий сравнения дисперсий асимптотических распределений в классе асимптотически нормальных оценок здесь неприменим. См.: [Saikkonen, 1991, Section 3] относительно подходящего определения эффективности.

[Phillips and Hansen, 1990]), оценка нелинейнго метода наименьших квадратов [Phillips and Loretan, 1991] (а также DGLS, как было отмечено выше). Для всех этих эффективных оценок t-статистика и статистика Вальда для тестирования гипотез относительно γ, правильным образом шкалированные, если это необходимо, имеют стандартные асимптотические распределения (более подробно см.: [Watson, 1994, Section 3.4.3]).

Другие оценки и свойства на конечных выборках

[Stock and Watson, 1993] изучали свойства только что упомянутых оценок (за исключением оценки Филлипса — Лоретана) в конечных выборках. Для различных DGP и объемов выборки ($T = 100$ и 300), которые они изучали, вырисовались следующие результаты. 1) Смещение является наименьшим для оценки Йохансена и наибольшим для SOLS. 2) Однако дисперсия оценки Йохансена намного больше, чем дисперсии других эффективных оценок, и DOLS имеет наименьшую среднеквадратичную ошибку (RMSE). 3) Для всех изученных оценок распределение t-значений (шкалированных правильным образом, если необходимо) проявляет тенденцию иметь более длинные хвосты, по сравнению с $N(0, 1)$, указывая, что нулевая гипотеза будет отвергаться слишком часто. Эти выводы в целом согласуются с выводами на основании исследований методом Монте-Карло в [Maddala and Kim, 1998, Section 5.7].

Контрольные вопросы

1. Как мы подчеркивали, имеется сходство между расширенной авторегрессией (9.4.6) на с. 622 и расширенной коинтегрирующей регрессией (10.4.5). Тогда почему t-значение при I(1)-регрессоре в расширенной коинтегрирующей регрессии (10.4.5) асимптотически распределено как $N(0, 1)$, в то время как в (9.4.6) это не так?

10.5. Приложение: Спрос на деньги в США

Литература по оцениванию функции спроса на деньги очень обширна (см.: [Goldfeld and Sichel, 1990], в качестве обзора). Уравнение спроса на деньги, обычно оцениваемое в литературе, имеет вид:

$$m_t - p_t = \mu + \gamma_y y_t + \gamma_R R_t + z_t^*, \qquad (10.5.1)$$

где m_t — логарифм денежной массы в период t, p_t — логарифм уровня цен, y_t — логарифм дохода, R_t — номинальная процентная ставка и z_t^* — некоторая ошибка. γ_y является эластичностью спроса на деньги по доходу, а γ_R — полуэластичностью спроса на деньги по процентной ставке (это полуэластичность, поскольку процентная ставка входит в уравнение спроса на деньги в уровнях, а не в логарифмах). Большинство эмпирических исследований, которые предшествовали литературе по единичным корням и коинтеграции, страдают от двух недостатков. Первый, как будет проверено ниже, заключается в том, что включенные переменные, $m_t - p_t$, y_t и R_t, по-видимому, содержат стохастические тренды. Если переменные имеют тренды, обычные статистические выводы, опирающиеся на предположение о стационарности, не будут обоснованными. Второй — регрессоры могут быть эндогенными. В случае спроса на деньги это может быть серьезной проблемой, поскольку практически во всех макроэкономических моделях сдвиг в спросе на деньги, представленный ошибками z_t^*, влияет на номинальную процентную ставку и, возможно, доход. Разрешение этих проблем обеспечивается эконометрической техникой оценивания коинтегрирующих регрессий, представленной в предыдущем параграфе.

Другая проблема, которая рассматривается в этом параграфе, — это проблема стабильности функции спроса на деньги. Консенсусом в литературе является то, что спрос на деньги нестабилен. Это оспаривается в [Lucas, 1988], который, используя годовые данные с 1900 года, утверждает, что полуэластичность процентной ставки стабильна, если наложить ограничение, что эластичность по доходу равна единице. Мы проверим утверждение Лукаса, применяя современные эконометрические методы, развитые в этой главе.

Данные

Наш анализ основан на годовых данных, исследованных в [Lucas, 1988] и расширенных в [Stock and Watson, 1993], чтобы покрыть период 1900–1989 годов. Их мерами денежной массы, дохода и номинальной процентной ставки служат M1, чистый национальный продукт (ВНП) и ставка по шестимесячным векселям (в процентах в годовом исчислении) соответственно. Использование *чистого* национального продукта, а не валового национального продукта, следует традиции, восходящей к [Friedman, 1959], переменная масштаба деятельности (y_t) в уравнении спроса на деньги должна быть мерой богатства, а не мерой объема сделок. Поскольку официальная статистика по M1 до 1900 года отсутствует, приходится производить некоторое склеивание рядов из нескольких источников. Для денежной массы и процентной ставки месячные данные усредняются, чтобы получить годовые наблюдения. Подробнее о построении данных см. Appendix B в [Stock and Watson, 1993].

($m - p$, y, R) как коинтегрированная система

На рис. 10.2 и 10.3 изображены графики рядов $m - p$, y и R. Все три ряда имеют явные тренды. Логарифм реальной денежной массы ($m - p$), с одной стороны, быстро возрастал в течение первой половины века, но не испытывал почти никакого роста в дальнейшем до 1981 года. Логарифм дохода (y), с другой стороны, неуклонно рос в течение всего периода выборки с серьезным нарушением этого роста с 1930-х до начала 1940-х, поэтому скорость обращения M1 ($y - (m - p)$), также изображенная на рис. 10.3, сократилась в течение этого периода, а затем неуклонно возрастала до 1981 года. За этим движением скорости обращения M1 довольно тесно следует номинальная процентная ставка, что наводит на мысль о том, что $m - p$, y и R являются коинтегрированными с коинтегрирующим вектором, компонента которого, соответствующая y, близка к 1.

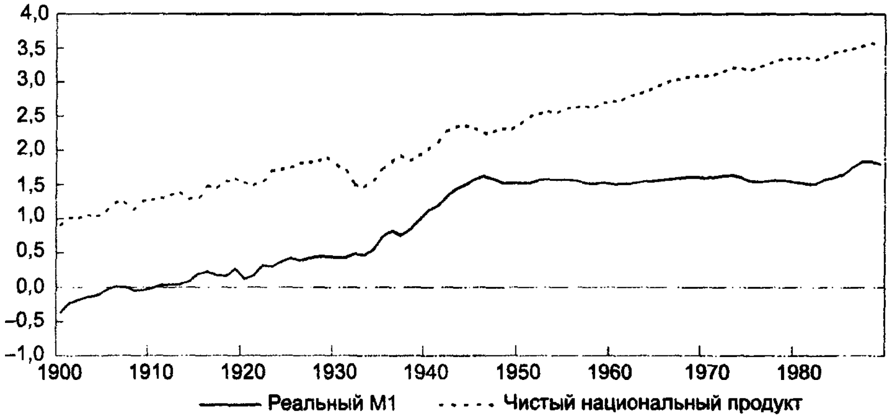

Рис. 10.2. Реальный ВНП США и реальная M1 в логарифмах

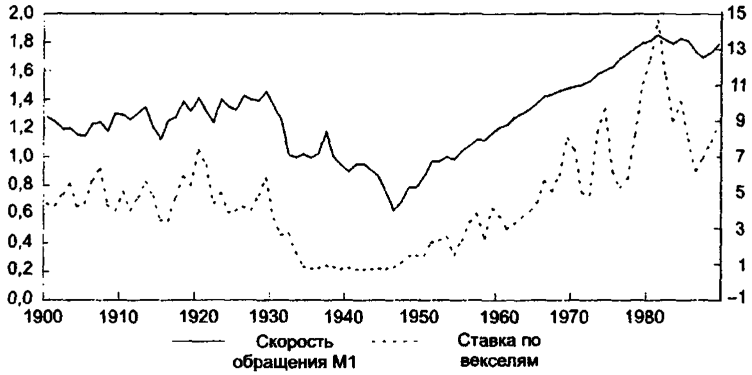

Рис. 10.3. Логарифм скорости обращения M1 (левая шкала) и краткосрочная ставка по векселям (правая шкала)

Анализ этих рисунков подсказывает, что эти три переменные, $m - p$, y и R, индивидуально можно хорошо описать I(1)-процессом. Действительно, [Stock and Watson, 1993] на основе ADF-тестов сообщают, что $m - p$ и y каждый являются I(1)-процессами со сносом, в то время как

R является I(1)-процессом без сноса. Они также сообщают, основываясь на тестировании общих трендов в [Stock and Watson, 1988], что эта трехмерная система является коинтегрированной с рангом коинтеграции, равным 1. В дальнейшем мы исходим из предположения, что $m - p$ коинтегрирована с y и R. (В эмпирическом упражнении вас попросят проверить справедливость этой предпосылки.)

DOLS

Первая строка табл. 10.2 показывает SOLS-оценку коинтегрирующей регрессии (10.5.1). (Период выборки — 1903–1987 годы, поскольку в расширенную коинтегрирующую регрессию DOLS строкой ниже будут включены два опережающих и два запаздывающих изменения.) Стандартные ошибки не приводятся, поскольку асимптотическое распределение

Таблица 10.2

SOLS- и DOLS-оценки, 1903–1987 годы

	γ_y	γ_R	R^2	**Стд. ошибка регрессии**
SOLS-оценки	0,943	−0,082	0,959	0,135
DOLS-оценки	0,970	−0,101	0,982	0,096
(Шкалированные стандартные ошибки)	(0,046)	(0,013)		

Источник: Точечные оценки и стандартные ошибки из Table III в [Stock and Watson, 1993]. R^2 и стандартная ошибка регрессии из авторских расчетов.

соответствующего t-отношения, зависящее от мешающих параметров, неизвестно. SOLS-оценка эластичности по доходу, равная 0,943, близка к единице, но мы не можем сказать, является ли она незначимо отличной от единицы. Чтобы быть в состоянии делать статистические выводы, мы обращаемся к DOLS-оцениванию коинтегрирующего вектора, которое должно дать оценки параметров $(\mu, \gamma_y, \gamma_R)$ коинтегрирующей регрессии (10.5.1) за счет добавления Δy_t, ΔR_t и их опережающих и запаздывающих значений. Следуя [Stock and Watson, 1993], число опережающих и запаздывающих значений (произвольно) выбрано равным 2, поэтому расширенная коинтегрирующая регрессия, связанная с (10.5.1), имеет вид:

$$\begin{aligned} m_t - p_t = \mu + \gamma_y y_t + \gamma_R R_t + \\ + \beta_{y0}\Delta y_t + \beta_{y,-1}\Delta y_{t+1} + \beta_{y,-2}\Delta y_{t+2} + \beta_{y1}\Delta y_{t-1} + \beta_{y2}\Delta y_{t-2} + \\ + \beta_{R0}\Delta R_t + \beta_{R,-1}\Delta R_{t+1} + \beta_{R,-2}\Delta R_{t+2} + \\ + \beta_{R1}\Delta R_{t-1} + \beta_{R2}\Delta R_{t-2} + v_t. \quad (10.5.2) \end{aligned}$$

Это обусловливает период выборки $t = 1903, \ldots, 1987$. Точечные DOLS-оценки (γ_y, γ_R), приведенные во второй строке табл. 10.2, получены в результате оценивания этого уравнения посредством OLS.

Чтобы вычислить соответствующие стандартные ошибки, необходимо оценить долговременную дисперсию ошибки v_t. С этой целью мы подбираем к DOLS-остаткам процесс авторегрессии. Порядок авторегрессии (снова произвольно) полагается равным 2. С DOLS-остатками, вычисленными для $t = 1903, \ldots, 1987$, периодом выборки для оценивания AR(2) становится $t = 1905, \ldots, 1987$ (объем выборки $= 83$). Оцененная авторегрессия имеет вид:

$$\widehat{v}_t = 0{,}93806\,\widehat{v}_{t-1} - 0{,}13341\,\widehat{v}_{t-2}, \quad SSR = 0{,}19843. \tag{10.5.3}$$

Затем мы используем формулу (10.4.17), но, чтобы иметь возможность воспроизвести DOLS-оценки в [Stock and Watson, 1993, Table III], при вычислении $\widehat{\sigma}_e^2$ мы делим SSR не на $T-p$, а на $T-p-K$, где K является количеством регрессоров в расширенной коинтегрирующей регрессии (которое равно здесь 13). Таким образом, $\widehat{\sigma}_e^2 = 0{,}19843/(83-2-13) = 0{,}0029181$ и

$$\widehat{\lambda}_v^2 = \frac{0{,}0029181}{(1 - 0{,}93806 + 0{,}13341)^2} = 0{,}07647. \tag{10.5.4}$$

Оценка $\widehat{\lambda}_v$ является квадратным корнем из этой величины, который равен 0,277. Поскольку стандартная ошибка DOLS регрессии равна 0,096, множитель в формуле (10.4.15) для шкалирования обычных стандартных ошибок OLS из расширенной коинтегрирующей авторегрессии равен $0{,}277/0{,}096 = 2{,}89$. Числа в скобках в табл. 10.2 являются скорректированными стандартными ошибками, вычисленными таким образом. Теперь мы можем видеть, что оцененная эластичность по доходу не является значимо отличной от единицы.

Нестабильный спрос на деньги?

В табл. 10.3 приводятся оценки параметров по SOLS и DOLS для двух подвыборок, 1903–1945 и 1946–1987 годы. (Следуя [Stock and Watson, 1993], дата сдвига 1946 года была выбрана и по причине естественного сдвига в конце Второй мировой войны, и по причине того, что она разделяет выборку почти пополам.) В резком контрасте с оценками, основанными на первой половине выборки, послевоенные оценки сильно отличаются от тех, которые основаны на полной выборке.

[Lucas, 1988] утверждает, что оценки, подобные этим, соответствуют стабильному спросу на деньги с единичной эластичностью по доходу. В поддержку этого суждения он указывает на рис. 10.4. На этом рисунке изображен график зависимости $m_t - p_t - y_t$ (логарифм обратной скорости обращения, которая была бы зависимой переменной в уравнении спроса на деньги (10.5.1), если бы эластичность по доходу γ_y полагалась равной единице) от процентной ставки. При этом послевоенные наблюдения указаны другим набором символов, нежели довоенные наблюдения. Отличительной чертой является то, что, хотя послевоенные процентные ставки значительно выше, послевоенные наблюдения лежат на линии, определяемой предвоенными наблюдениями[1].

Таблица 10.3

Спрос на деньги в двух подвыборках

	1903–1945		1946–1987	
	γ_y	γ_R	γ_y	γ_R
SOLS-оценки	0,919	−0,085	0,192	−0,016
DOLS-оценки	0,887	−0,104	0,269	−0,027
(Пересчет стандартных ошибок)	(0,197)	(0,038)	(0,213)	(0,025)

Источник: Table III в [Stock and Watson, 1993].

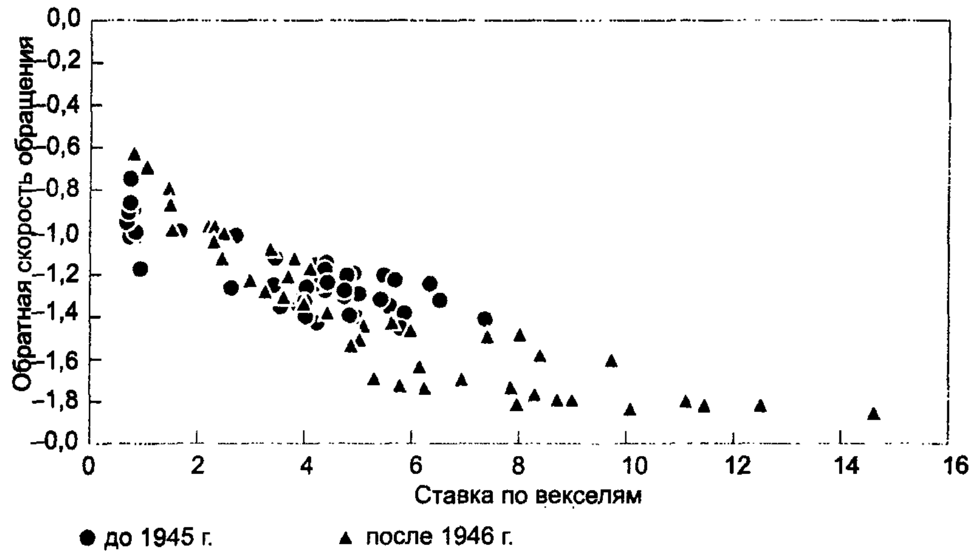

Рис. 10.4. График зависимости $m - p - y$ от R

Этот факт позволяет предположить, что послевоенные оценки страдают от проблемы сильной мультиколлинеарности: из-за схожих трендов в y и R в послевоенном периоде (показанном на рис. 10.2 и 10.3), γ_y и γ_R, оцененные по послевоенному периоду, сильно коррелированы. Хотя для каждого элемента послевоенная оценка (γ_y, γ_R) отличается от предвоенной, они могут не быть совместно значимо отличными. Эту возможность можно было бы легко проверить тестом Чоу на структурный сдвиг, по крайней мере если бы регрессоры являлись стационарными.

[1]В оригинальном графике в [Lucas, 1988] период выборки — 1900–1985 годы, а дата сдвига — 1957 год, но оба графика говорят, по существу, об одном и том же.

В тесте Чоу вы могли бы оценить

$$m_t - p_t = \mu + \gamma_y y_t + \gamma_R R_t + \delta_0 D_t + \delta_y y_t D_t + \delta_R R_t D_t + z_t^*, \qquad (10.5.5)$$

где D_t является дамми-переменной, значение которой равно 1, если $t \geq \geq 1946$ и 0 в противном случае. Таким образом, константа и коэффициенты при y_t и R_t равны $(\mu, \gamma_y, \gamma_R)$ до 1945 года и равны $(\mu + \delta_0, \gamma_y + +\delta_y, \gamma_R + \delta_R)$ после 1945 года. Если регрессоры являются стационарными, то статистика Вальда для тестирования нулевой гипотезы о том, что $\delta_0 = 0$, $\delta_y = 0$, $\delta_R = 0$ асимптотически распределена по $\chi^2(3)$, когда обе подвыборки растут с сохранением неизменным относительного объема (см. аналитическое упражнение 12 в главе 2). Теперь, когда регрессоры являются I(1)-процессами и некоинтегрированы, как в данном случае, в [Hansen, 1992b, Theorem 3(a)] показано, что статистика Вальда (правильным образом шкалированная, если необходимо), имеет то же самое асимптотическое распределение (χ^2), если параметры коинтегрирующей регрессии оценены эффективным методом, таким как DOLS (который добавляет Δy_t, ΔR_t и их опережающие и запаздывающие значения в коинтегрирующую регрессию (10.5.5)), **полностью модифицированный метод наименьших квадратов** (*FMOLS — Fully Modified estimation*) или ML Йохансена[1]. DOLS-оценки регрессии (10.5.5), пополненной двумя опережающими и двумя запаздывающими значениями Δy и ΔR, приведены в табл. 10.4. Незначимая статистика Вальда поддерживает точку зрения Лукаса о том, что спрос на деньги в Соединенных Штатах в XX веке был стабильным.

Таблица 10.4

Тест на структурный сдвиг

	γ_y	γ_R	δ_0	δ_y	δ_R	Статистика Вальда
DOLS-оценки	0,925	−0,090	1,36	−0,52	0,048	5,95
Шкалированные стандартные ошибки	(0,142)	(0,026)	(0,72)	(0,31)	(0,034)	(p-значение = 0,11)

Источник: Вычисления автора, которые нужно проверить в эмпирическом упражнении.

[1] Мы отмечали в предыдущем параграфе, что статистика Вальда не распределена асимптотически как хи-квадрат даже после шкалирования, если гипотеза включает в себя μ. Однако результат в [Hansen, 1992b] показывает, что статистика Вальда для тестирования гипотезы о *различии* в μ асимптотически распределена как хи-квадрат. [Hansen, 1992b] также разрабатывает тест на структурный сдвиг в неизвестную дату сдвига.

Набор задач для главы 10

Эмпирические упражнения

(Прежде чем выполнять это задание, просмотрите [Stock and Watson, 1993, Section 7 и Appendix B.] В файле MPYR.ASC находятся следующие годовые данные:

- Столбец 1: натуральный логарифм M1 (упоминаемый как m).
- Столбец 2: натуральный логарифм дефлятора ВНП (p).
- Столбец 3: натуральный логарифм ВНП (y).
- Столбец 4: ставка по векселям в процентах в годовом исчислении (R).

Период выборки — с 1900 по 1989 год. Это те же самые данные, которые использовались в [Stock and Watson, 1993] в их исследовании спроса на деньги США. См. их Appendix B в качестве источника данных.

Вопросы (a) и (b) заключаются в проверке презумпции того, что ряд $m-p$ коинтегрирован с (y, R). Остальные вопросы связаны с оцениванием коинтегрирующего вектора.

(a) (Одномерные тесты на единичный корень.) В [Stock and Watson, 1993, Appendix B] сообщается, что ADF t^{μ}- и t^{τ}-статистики, вычисленные с использованием 2 и 4 лагированных первых разностей, не в состоянии отвергнуть гипотезу единичного корня для каждого из рядов $m-p$ и R на 10%-ном уровне и что гипотеза единичного корня не отвергается для y с 4 лагами, но отвергается на 10%-ном (но не 5%-ном) уровне с 2 лагами. Проверьте эти выводы. Используйте t^{μ}-тест для R и t^{τ} для $m-p$ и y и используйте асимптотические критические значения (критические значения для $T=\infty$).

(b) (Тесты, основанные на остатках.) Проведите тест Энгла — Грейнджера для проверки нулевой гипотезы о том, что ряд $m-p$ не коинтегрирован с y и R. Положите $p=1$ (выбрано по BIC). (Если вы не включаете время в регрессию, то ADF t-статистика, полученная по остаткам, должна быть равной примерно -4.7. Следует применить случай 2, рассмотренный в параграфе 10.3. Поэтому 5%-ное критическое значение должно быть равным -3.80.)

(c) (DOLS.) Воспроизведите SOLS и DOLS-оценки коинтегрирующего вектора, приведенные в табл. 10.2.

(d) (Тест Чоу.) Воспроизведите DOLS-оценки, приведенные в табл. 10.4. Они основаны на следующей расширенной коинтегрирующей регрессии:

$$m_t - p_t = \mu + \gamma_y y_t + \gamma_R R_t + \delta_0 D_t + \delta_y y_t D_t + \delta_R R_t D_t +$$
$$+ \beta_{y0}\Delta y_t + \beta_{y,-1}\Delta y_{t+1} + \beta_{y,-2}\Delta y_{t+2} + \beta_{y1}\Delta y_{t-1} + \beta_{y2}\Delta y_{t-2} +$$
$$+ \beta_{R0}\Delta R_t + \beta_{R,-1}\Delta R_{t+1} + \beta_{R,-2}\Delta R_{t+2} + \beta_{R1}\Delta R_{t-1} + \beta_{R2}\Delta R_{t-2} + v_t.$$

Нулевая гипотеза состоит в том, что $\delta_0 = \delta_y = \delta_R = 0$. Положите p в (10.4.16) равным двум. Вычислите статистику Вальда для тестирования нулевой гипотезы об отсутствии структурного сдвига.

Литература

Aoki, Masanao, 1968, "Control of Large-Scale Dynamic Systems by Aggregation," IEEE Transactions on Automatic Control*IEEE Transactions on Automatic Control*, AC-13, 246–253.

Banerjee, A., J. Dolado, D. Hendry, and G. Smith, 1986, "Exploring Equilibrium Relationships in Econometrics through Static Models: Some Monte Carlo Evidence," *Oxford Bulletin of Economics and Statistics*, 48, 253–277.

Box, G., and G. Tiao, 1977, "A Canonical Analysis of Multiple Time Series," *Biometrika*, 64, 355–365.

Davidson, J., D. Hendry, F. Srba, and S. Yeo, 1978, "Econometric Modelling of the Aggregate Time- Series Relationships between Consumers' Expenditure and Income in the United Kingdom," *Economic Journal*, 88, 661–692.

Engle, R., and C. Granger, 1987, "Co-Integration and Error Correction: Representation, Estimation, and Testing," *Econometrica*, 55, 251–276.

Engle, R., and S. Yoo, 1991, "Cointegrated Economic Time Series: An Overview with New Results," in R. Engle and C. Granger (eds.), *Long-Run Economic Relations: Readings in Cointegration*, New York: Oxford University Press.

Friedman, M., 1959, "The Demand for Money: Some Theoretical and Empirical Results," *Journal of Political Economy*, 67, 327–351.

Fuller,W., 1996, *Introduction to Statistical Time Series* (2d ed.), New York: Wiley.

Goldfeld, S., and D. Sichel, 1990, "The Demand for Money," Chapter 8 in B. Friedman and F. Hahn (eds.), *Handbook of Monetary Economics*, Volume I, New York: North-Holland.

Granger, C., 1981, "Some Properties of Time Series Data and Their Use in Econometric Model Specification," *Journal of Economics*, 16, 121–130.

Granger, C., and P. Newbold, 1974, "Spurious Regressions in Econometrics," *Journal of Economics*, 2, 111–120.

Hamilton, J., 1994, *Time Series Analysis*, Princeton: Princeton University Press.

Hansen, B., 1992a, "Efficient Estimation and Testing of Cointegrating Vectors in the Presence of Deterministic Trends," *Journal of Economics*, 53, 87–121.

______,1992b, "Tests for Parameter Instability in Regressions with I(1) Processes," *Journal of Business and Economic Statistics*, 10, 321–335.

Haug, A., 1996, "Tests for Cointegration: A Monte Carlo Comparison," *Journal of Economics*, 71, 89–115.

Johansen, S., 1988, "Statistical Analysis of Cointegrated Vectors," *Journal of Economic Dyamics and Control*, 12, 231–254.

______,1995, *Likelihood-Based Inference in Cointegrated Vector Auto-Regressive Models*, New York: Oxford University Press.

Lucas, R., 1988, *Money Demand in the United States: A Quantitative Review, Carnegie-Rochester Conference Series on Public Policy*, Volume 29, 137–168.

Lütkepohl, H., 1993, *Introduction to Multiple Time Series Analysis* (2d ed.), New York: Springer-Verlag.

Maddala, G. S., and In-Moo Kim, 1998, *Unit Roots, Cointegration, and Structural Change*, New York: Cambridge University Press.

Ng, S., and P. Perron, 1995, "Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncated Lag," *Journal of the American Statistical Association*, 90, 268–281.

Park, J., and P. Phillips, 1988, "Statistical Inference in Regressions with Integrated Processes: Part 1," *Econometric Theory*, 4, 468–497.

Phillips, P., 1986, "Understanding Spurious Regressions in Econometrics," *Journal of Economics*, 33, 311–340.

______,1988, "Weak Convergence to the Matrix Stochastic Integral R 1 0 BdB0," *Journal of Multivariate Analysis*, 24, 252–264.

______,1991, "Optimal Inference in Cointegrated Systems," *Econometrica*, 59, 283–306.

Phillips, P., and S. Durlauf, 1986, "Multiple Time Series Regression with Integrated Processes," *Review of Economic Studies*, 53, 473–495.

Phillips, P., and B. Hansen, 1990, "Statistical Inference in Instrumental Variables Regression with I(1) Processes," *Review of Economic Studies*, 57, 99–125.

Phillips, P., and M. Loretan, 1991, "Estimating Long-Run Economic Equilibria," *Review of Economic Studies*, 58, 407–436.

Phillips, P., and S. Ouliaris, 1990, "Asymptotic Properties of Residual Based Tests for Cointegration," *Econometrica*, 58, 165–193.

Phillips, P., and J. Park, 1988, "Asymptotic Equivalence of OLS and GLS in Regressions with Integrated Regressors," *Journal of the American Statistical Association*, 83, 111–115.

Saikkonen, P., 1991, "Asymptotically Efficient Estimation of Cointegration Regressions," *Econometric Theory*, 7, 1–21.

Sims, C., J. Stock, and M. Watson, 1990, "Inference in Linear Time Series Models with Some Unit Roots," *Econometrica*, 58, 113–144.

Stock, J., 1987, "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors," *Econometrica*, 55, 1035–1056.

Stock, J., and M. Watson, 1988, "Testing for Common Trends," *Journal of the American Statistical Association*, 83, 1097–1107.

______, 1993, "A Simple Estimator of Cointegrating Vectors in Higher Order Integrated Systems," *Econometrica*, 61, 783–820.

Watson, M., 1994, "Vector Autoregressions and Cointegration," Chapter 47 in R. Engle, and D. Mc-Fadden (eds.), *Handbook of Econometrics*, Volume IV, New York: North Holland.

Приложение А. Блочные (клеточные) матрицы и произведения Кронекера

Блочные (клеточные) матрицы

Иногда удобно разбивать элементы матрицы на MN подматриц как

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & \cdots & \boldsymbol{A}_{1N} \\ \vdots & & \vdots \\ \boldsymbol{A}_{N1} & \cdots & \boldsymbol{A}_{NN} \end{bmatrix}.$$

Матрицу в таком виде называют **блочной (клеточной) матрицей** (*partitioned matrix*). Подстрочные индексы подматриц определяются тем же самым способом, что и элементы матрицы. Например, мы можем записать:

$$\boldsymbol{A} = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix},$$

где

$$\boldsymbol{A}_{11} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \end{bmatrix}, \quad \boldsymbol{A}_{12} = \begin{bmatrix} 7 \\ 8 \end{bmatrix}, \quad \boldsymbol{A}_{21} = \begin{bmatrix} 3 & 6 \end{bmatrix}, \quad \boldsymbol{A}_{22} = 9.$$

Наиболее известный частный случай — когда $M = N$ и внедиагональные блоки ($\boldsymbol{A}_{mn}$ для $m \neq n$) являются нулевыми матрицами:

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_{11} & & \\ & \ddots & \\ & & \boldsymbol{A}_{MM} \end{bmatrix}.$$

Это называется **блочно-диагональной матрицей** (*block diagonal matrix*).

Сложение и умножение блочных (клеточных) матриц

Сложение и умножение матриц обобщается на случай блочных (клеточных) матриц. Таким образом,

$$\boldsymbol{A} + \boldsymbol{B} = \begin{bmatrix} \boldsymbol{A}_{11} + \boldsymbol{B}_{11} & \cdots & \boldsymbol{A}_{1N} + \boldsymbol{B}_{1N} \\ \vdots & & \vdots \\ \boldsymbol{A}_{M1} + \boldsymbol{B}_{M1} & \cdots & \boldsymbol{A}_{MN} + \boldsymbol{B}_{MN} \end{bmatrix}. \tag{A.1}$$

$$\boldsymbol{AB} = \begin{bmatrix} \boldsymbol{A}_{11} & \cdots & \boldsymbol{A}_{1N} \\ \vdots & & \vdots \\ \boldsymbol{A}_{M1} & \cdots & \boldsymbol{A}_{MN} \end{bmatrix} \begin{bmatrix} \boldsymbol{B}_{11} & \cdots & \boldsymbol{B}_{1L} \\ \vdots & & \vdots \\ \boldsymbol{B}_{N1} & \cdots & \boldsymbol{B}_{NL} \end{bmatrix} =$$

$$= \begin{bmatrix} \sum_{n=1}^{N} \boldsymbol{A}_{1n}\boldsymbol{B}_{n1} & \cdots & \sum_{n=1}^{N} \boldsymbol{A}_{1n}\boldsymbol{B}_{nL} \\ \vdots & & \vdots \\ \sum_{n=1}^{N} \boldsymbol{A}_{Mn}\boldsymbol{B}_{n1} & \cdots & \sum_{n=1}^{N} \boldsymbol{A}_{Mn}\boldsymbol{B}_{nL} \end{bmatrix}. \tag{A.2}$$

Во всех этих выражениях матрицы должны иметь согласованные размеры для выполнения операций. Так, при сложении размеры $\boldsymbol{A}_{mn}$ и $\boldsymbol{B}_{mn}$ должны быть одинаковыми для всех $m(= 1, 2, \ldots, M)$ и $n(= 1, 2, \ldots, N)$. При умножении количество столбцов в $\boldsymbol{A}_{mn}$ должно быть равным количеству строк в $\boldsymbol{B}_{nl}$ для всех $m(= 1, 2, \ldots, M)$, $n(= 1, 2, \ldots, N)$, $l(= 1, 2, \ldots, L)$. Частный случай умножения, когда $\boldsymbol{B}$ является **штабелированным вектором** (*stacked vector*) $\boldsymbol{c}$:

$$\boldsymbol{Ac} = \begin{bmatrix} \boldsymbol{A}_{11} & \cdots & \boldsymbol{A}_{1N} \\ \vdots & & \vdots \\ \boldsymbol{A}_{M1} & \cdots & \boldsymbol{A}_{MN} \end{bmatrix} \begin{bmatrix} \boldsymbol{c}_1 \\ \vdots \\ \boldsymbol{c}_N \end{bmatrix} = \begin{bmatrix} \sum_{n=1}^{N} \boldsymbol{A}_{1n}\boldsymbol{c}_n \\ \vdots \\ \sum_{n=1}^{N} \boldsymbol{A}_{Mn}\boldsymbol{c}_n \end{bmatrix} \tag{A.3}$$

Несколько часто встречающихся случаев имеют вид:

$$\begin{bmatrix} \boldsymbol{A}_{11} & & \\ & \ddots & \\ & & \boldsymbol{A}_{MM} \end{bmatrix} \begin{bmatrix} \boldsymbol{c}_1 \\ \vdots \\ \boldsymbol{c}_M \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_{11}\boldsymbol{c}_1 \\ \vdots \\ \boldsymbol{A}_{MM}\boldsymbol{c}_M \end{bmatrix}. \tag{A.4}$$

$$\begin{bmatrix} \boldsymbol{A}_{11} & & \\ & \ddots & \\ & & \boldsymbol{A}_{MM} \end{bmatrix} \begin{bmatrix} \boldsymbol{B}_{11} & & \\ & \ddots & \\ & & \boldsymbol{B}_{MM} \end{bmatrix} =$$

$$= \begin{bmatrix} \boldsymbol{A}_{11}\boldsymbol{B}_{11} & & \\ & \ddots & \\ & & \boldsymbol{A}_{MM}\boldsymbol{B}_{MM} \end{bmatrix}. \tag{A.5}$$

$$\begin{bmatrix} \boldsymbol{A}'_{11} & & \\ & \ddots & \\ & & \boldsymbol{A}'_{MM} \end{bmatrix} \begin{bmatrix} \boldsymbol{B}_{11} & \cdots & \boldsymbol{B}_{1M} \\ \vdots & & \vdots \\ \boldsymbol{B}_{M1} & \cdots & \boldsymbol{B}_{MM} \end{bmatrix} \begin{bmatrix} \boldsymbol{A}_{11} & & \\ & \ddots & \\ & & \boldsymbol{A}_{MM} \end{bmatrix} =$$

$$= \begin{bmatrix} A'_{11}B_{11}A_{11} & \cdots & A'_{11}B_{1M}A_{MM} \\ \vdots & & \vdots \\ A'_{MM}B_{M1}A_{11} & \cdots & A'_{MM}B_{MM}A_{MM} \end{bmatrix}, \tag{A.6}$$

$$\begin{bmatrix} A'_{11} & & \\ & \ddots & \\ & & A'_{MM} \end{bmatrix} \begin{bmatrix} B_{11} & \cdots & B_{1M} \\ \vdots & & \vdots \\ B_{M1} & \cdots & M_{MM} \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_M \end{bmatrix} =$$

$$= \begin{bmatrix} A'_{11}B_{11}c_1 + & \cdots & +A'_{11}B_{1M}c_M \\ & \vdots & \\ A'_{MM}B_{M1}c_1 + & \cdots & +A'_{MM}B_{MM}c_M \end{bmatrix}, \tag{A.7}$$

$$\begin{bmatrix} A'_{11} & \cdots & A'_{MM} \end{bmatrix} \begin{bmatrix} B_{11} & \cdots & B_{1M} \\ \vdots & & \vdots \\ B_{M1} & \cdots & B_{MM} \end{bmatrix} \begin{bmatrix} A_{11} \\ \vdots \\ A_{MM} \end{bmatrix} =$$

$$= \sum_{m=1}^{M} \sum_{h=1}^{M} A'_{mm} B_{mh} A_{hh}, \tag{A.8}$$

где c_m $(m = 1, 2, \ldots, M)$ являются векторами-столбцами.

Обращение блочных (клеточных) матриц

Матрица, обратная к блочно-диагональной матрице, имеет вид:

$$\begin{bmatrix} A_{11} & & \\ & \ddots & \\ & & A_{MM} \end{bmatrix}^{-1} = \begin{bmatrix} A_{11}^{-1} & & \\ & \ddots & \\ & & A_{MM}^{-1} \end{bmatrix}, \tag{A.9}$$

при условии, что A_{mm} $(m = 1, 2, \ldots, M)$ обратимы. Это можно проверить непосредственным умножением.

Для блочной (клеточной) матрицы размерности 2×2 общего вида **блочная обратная** (*partitioned inverse*) матрица имеет вид:

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}^{-1} = \begin{bmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}F_2A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}F_2 \\ -F_2A_{21}A_{11}^{-1} & F_2 \end{bmatrix}, \tag{A.10}$$

где $F_2 = (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}$. Это можно легко проверить, умножая A на обратную ей матрицу. Вследствие симметрии вычисления верхний левый блок можно записать также как

$$F_1 = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}.$$

Произведения Кронекера

Для матриц $\boldsymbol{A}$ $(M \times N)$ и $\boldsymbol{B}$ $(K \times L)$ общего вида **произведение Кронекера** (*Kronecker product*) определяется как

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{11}\boldsymbol{B} & \cdots & a_{1N}\boldsymbol{B} \\ \vdots & & \vdots \\ a_{M1}\boldsymbol{B} & \cdots & a_{MN}\boldsymbol{B} \end{bmatrix}. \tag{A.11}$$

Это матрица размера $MK \times NL$. В частном случае, когда $\boldsymbol{A}$ и $\boldsymbol{B}$ являются векторами:

$$\boldsymbol{a} \otimes \boldsymbol{b} = \begin{bmatrix} a_1 \\ \vdots \\ a_M \end{bmatrix} \otimes \begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix} = \begin{bmatrix} a_1\boldsymbol{b} \\ \vdots \\ a_M\boldsymbol{b} \end{bmatrix}. \tag{A.12}$$

Очень просто, но громоздко проверить, что

$$(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = \boldsymbol{AC} \otimes \boldsymbol{BD}, \tag{A.13}$$

при условии, что $\boldsymbol{A}$ и $\boldsymbol{C}$ имеют согласованные размеры и $\boldsymbol{B}$ и $\boldsymbol{D}$ имеют согласованные размеры, и что

$$(\boldsymbol{A} \otimes \boldsymbol{B})' = \boldsymbol{A}' \otimes \boldsymbol{B}', \tag{A.14}$$

$$(\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}, \tag{A.15}$$

при условии обратимости $\boldsymbol{A}$ и $\boldsymbol{B}$.

Предметный указатель

А

Абсолютно суммируемые коэффициенты:
- $AR(p)$ как $MA(\infty)$ с 409–410
- $MA(\infty)$ с 397–398

Абсолютно суммируемые обращения лаговых полиномов 404
Абсолютно суммируемый фильтр 400
Автоковариационная матрица 127
Автоковариация 127
Автоковариация порядка j 127
Автокорреляции коэффициент 127
Автокорреляции коэффициент порядка j 127
Авторегрессии модель первого порядка 421
Авторегрессии модель первого порядка:
- автоковариации 408
- описание 31
- оценивание 423
  - – методом максимального правдоподобия 579–583
- распределение оценки наименьших квадратов для конечных выборок 612

Авторегрессии процесс первого порядка 405–408
Авторегрессии процесс порядка p 405–408
Авторегрессия: оценивание 429
Авторегрессия: оценивание:
- векторных авторегрессий 427–428
- глубина запаздываний для оценивания 424–425
- модели AR(1) 421–423
- модели $ARMA(p,q)$ 428–429
- последовательная процедура от общего к частному, использующая t-тесты 424–425

Анализ временных рядов:
- ключевые концепции 124
- ковариационно стационарные процессы 126–127
- мартингалы 129–130
- необходимость эргодической стационарности 124–125
- последовательность мартингал-разностей 131–132
- процессы белого шума 128
- процессы с авторегрессионной гетероскедастичностью 126–127
- случайные блуждания 130
- строгая экзогенность 31–32
- формулировки отсутствия сериальной зависимости 133–134
- эргодическая теорема 128–129

Анализ методом Монте-Карло 107
Аннулятор 40
Асимптотики:
- двухшагового метода наименьших квадратов 257
- для выборочных средних процессов с сериальной корреляцией 430–435
- локальные к единице 637
- мощность состоятельного теста 148–149
- оценки обобщенного метода моментов 244
- оценок максимального правдоподобия с полной информацией 568
- с фиксированными регрессорами 194

Асимптотическая дисперсия в теории для больших выборок 121
Асимптотическая мощность 148
Асимптотическая нормальность:
- М-оценок 502–503
- в линейной регрессионной модели 497–498
- выражение для ошибки оценки 515–518
- многомерных регрессий 558
- моделей QR 542
- обобщенного метода моментов 511–514
- обобщенного метода моментов для одного уравнения 232–233
- обобщенного метода моментов для нескольких уравнений 294
- оценки максимального правдоподобия для пробит-модели 498–499
- оценок максимального правдоподобия с полной информацией 567
- состоятельное оценивание асимптотической дисперсии 506
- сравнение с обобщенным методом моментов 514–515
- усеченных регрессионных моделей 547–550
- условного максимального правдоподобия 506–511
- экстремальных оценок 501–518

Асимптотическая теория 114
см. также асимптотическая диспер-

сия в теории для больших выборок
Асимптотически более эффективная 165
Асимптотически нормальные оценки 121
Асимптотическое распределение:
- оценки обобщенного метода моментов 238-239
- оценок наименьших квадратов 141-143, 190-191
- Тобит-модели 553

Б

Байесовский информационный критерий BIC 425-427
Бевериджа — Нельсона разложение BN 600-601, 665-666
Белого шума процесс 128, 394, 416
Белого шума процесс векторный 416
Бинарного отклика модель 539
Блочная матрица:
- обращение 710-711
- описание 708
- сложение и умножение 708-710
Бокса — Пирса Q-статистика 171-173, 202-203
Бройша — Годфри тест 178, 202, 209
Броуновского движения процессы 603

В

Вальда принцип 65
Вальда статистика 522-524, 527
Вектор вклада для наблюдения t 502
Вектор данных 28
Векторной авторегрессии коэффициентов оценивание 427-428
Векторной авторегрессии процесс VAR:
- оценивание методом условного максимального правдоподобия 583-584
- порядка p 419
Векторные процессы 416
Векторный процесс ARMA(p:q) 419-420
Векторный процесс скользящего среднего бесконечного порядка 417
Взвешенная регрессия 82
Взвешенного метода наименьших квадратов оценки:
- с известным α 163-164
- с оцененным α 164-165
- сравнение с обычными оценками наименьших квадратов 165
Взвешенный метод наименьших квадратов 81
Взвешенный метод наименьших квадратов или взвешенная регрессия 82
Взвешивающая матрица 236
Винеровский процесс 605
Влиятельности анализ 45
Влиятельные наблюдения 43-46
«Внутри» оценка 360
Вогнутости условие:
- для функции правдоподобия линейной регрессии 493
- для функции правдоподобия пробит-модели 492
- после репараметризации 492
Возвращение к среднему 596
Временной тренд:
- инкорпорирование в расширенный тест Дики — Фуллера 633-635
- описание 190
Временные ряды:
- графики сериально некоррелированных 134
- описание 124
Выборка 25
Выборка Блэкмана — Ньюмарка 280
Выборочная автоковариация j-го порядка 171
Выборочное среднее 122
Выборочный коэффициент корреляции i-го порядка 127
Выражение для ошибки оценки 237, 368

Г

Гаусса — Маркова теорема 48, 49, 80
Гаусса — Ньютона алгоритм 530
Гауссовская модель векторной авторегрессии 583
Гауссовский процесс авторегрессии первого порядка 579
Гауссовский процесс авторегрессии порядка p 583
Гессиан для наблюдения t:
- для М-оценок 502
- для моделей с качественным откликом 540-541
- для тобит-моделей 553
Гиперсостоятельные оценки 192
Гипотеза эффективности рынка 449, *см. также* форвардная премия рыночного равновесия модели
Глубина запаздываний:
- для оценивания авторегрессий 424-425
- использование информационных критериев Акаике, Шварца (байесовского) 425-427
Гомоскедастичность безусловная:
- описание 34
- сравнение с условной 155
Гомоскедастичность:
- варианты асимптотических тестов при условной гомоскедастичности 157-158
- обобщенный метод моментов для нескольких уравнений и условная гомоскедастичность 303-304

– обобщенный метод моментов и условная гомоскедастичность 255–264
– определение условной и безусловной гомоскедастичности 34–35
– оценивание с параметризованной условной гомоскедастичностью 166
– оценивание с параметризованной условной гомоскедастичностью 161
– оценки фиксированных эффектов и ослабление требования условной гомоскедастичности 365–366
– распределение на больших выборках и условная гомоскедастичность 156–157
– редукция к формулам для конечных выборок 155–156
– тестирование условной гомоскедастичности 159–161
– Уайта тест на условную гомоскедастичность 160–161
– условная и безусловная 155
Гордина условие:
– для эргодических стационарных процессов 432–435
– ограничение степени сериальной корреляции 436–437

Д

Дамми-переменные 94
Данные 25
Дарбина h-статистика 178
Дарбина — Уотсона статистика 68
Двумерная I(0)-система 663
Двумерная I(1)-система 664
Двумерная система 692
Двусторонние фильтры 696
Двусторонний t-тест 69
Двухшаговый метод наименьших квадратов:
– альтернативные способы вывода 259
– асимптотические свойства 257–258
– из эффективного GMM 256–257
– как две регрессии 260–261
– как оценка инструментальных переменных 260
– коррекция смещения с помощью 271–272
– описание 222–223
– свойства на конечных выборках 258–259
– сравнение с оценкой максимального правдоподобия с ограниченной информацией 575–576
«Дельта-метод» 120
Детерминированный тренд 592
skip Фуллера t-статистика в расширенной модели, критические значения 683
Дики — Фуллера тесты:
– инкорпорирование временных трендов 616–619
– модель авторегрессии первого порядка 608–619
– обзор 635–636
– расширенные 621–636
– случайных блужданий без сноса 611–615
Добавления переменной тест 253–255
Доверительный интервал 61
Долговременная дисперсия 431
Долговременная ковариационная матрица S:
– когда автоковариации зануляются и оцениваются 437–438
– представление через матрицу данных 444–445
– процедура VARHAC для оценивания 440–442
– условная гомоскедастичность и оценивание 443–447
– ядра для оценивания 438–440
Доллар/йена:
– график $s30$ против форвардной ставки 455
– коррелограмма 452
– ошибка прогноза 451
– спотовая ставка 1975–89, 454
Доллар/стерлинг реальные обменные курсы 641
Доминирования условие 490
Доступный обобщенный метод наименьших квадратов 83
Дэвидсона — Мак-Киннона коррекция 209

Е

Единичный корень ВВП 634–635, *см. также* доллар/иена

З

Зависимая переменная 25
Зависимая переменная лагированная 31
Закон больших чисел слабый 122
Закон больших чисел усиленный 122
Закон одной цены 639–640
Закон полных математических ожиданий 30, 196, 510
Законы больших чисел:
– Чебышева 122
– второй Колмогорова 122
– для ковариационно-стационарных вопросов 430–431
– описание 122
– усиленный и слабый 122

Запаздываний глубина:
- для оценивания авторегрессий 424-425
- использование информационных критериев Акаике, Шварца (байесовского) 425-427
Запаздывания оператор 399
Заработной платы уравнение полулогарифмическое 267
Заработной платы уравнение:
- IQ как мера способностей 268
- модель компонент ошибки 355
- обзор 561
- обнаружение влияния длительности пребывания в должности 33
- описание 26-27 217
- полулогарифмическое 267
- система уравнений 288-289
- смещение из-за пропуска переменной 267-268
- тестирование предопределенности образования 250-251
- упрощенная версия 229-230

И

Идентифицируемое точно 232
Иена/доллар:
- график $s30$ против форвардной ставки 455
- коррелограмма 452
- ошибка прогноза 451
- спотовая ставка 454
Инвариантности принцип 603
Инвариантность:
- максимального правдоподобия 483-484
- оценок максимального правдоподобия с полной информацией 568-569
Индивидуальный эффект (неоднородность индивидов или фиксированный эффект) 354-356
Инструмент релевантный 221
Инструментальная переменная или инструмент:
- описание 221
- требования предопределенности/корреляции 231
Инструментальных переменных оценка 221, 235
Инфляция:
- гипотеза эффективного рынка 181-182, 185-187
- реальные процентные ставки и инфляция 183
- сравнение теории для больших выборок и теории для конечных выборок 187-189
Информационная матрица 74
- для наблюдения 506
Информационное множество 129
Информационное неравенство Кульбака — Лейблера 494
Искажение размера 146
Исключающие ограничения 564

Й

Йенсена неравенство 534
Йохансена процедура максимального правдоподобия 677-679

К

Кажущиеся несвязанными регрессии, итерационная процедура 570
Кажущиеся несвязанными регрессии:
- асимптотические свойства 310
- как обобшенный метод моментов для одного уравнения 18
- как обобщенный метод моментов для нескольких уравнений 306-307
- как специализация оценки максимального правдоподобия с полной информацией 569-570
- сравнение с обобщенным методом моментов для нескольких уравнений и методом наименьших квадратов 310-312
Казначейство США — веб-сайт 206
Квадратично суммируемая 459-460
Квази- (или псевдо-) максимальное правдоподобие 76
Классическая линейная модель регрессии:
- Крамера — Рао граница 73-75
- для случайных выборок 34
- матричная запись 27-28
- обобщенная 78
- описание 25, 660
- предположение линейности 25-27
- предположение строгой экзогенности 28-32
- статистика отношения правдоподобий 253
- функция правдоподобия 71
Кобба — Дугласа технология:
- идентификация 88
- параметризованная функция издержек, выведенная на ее основе 86-88
- транслогарифмическая функция издержек как обобщение 94-95
Ковариационно-стационарные процессы
- закон больших чисел 430-431
- процесс белого шума 128, 394-395
- фильтрация 398-399
Коинтеграции ранг:
- описание 667
- определение 680-687
Коинтеграционная связь 667
Коинтеграционное пространство 667

Коинтеграция в векторной авторегрессии конечного порядка 673–675
Коинтегрированные системы:
- альтернативные представления 671–679
- Бевериджа — Нельсона разложение и коинтегрированные системы 600–601
- векторные авторегрессии и коинтегрированные системы 673–675
- Йохансена процедура максимального правдоподобия 677–679
- $m-p$, y и R как коинтегрированная система 698–700
- линейные векторные I(0)- и I(1)-процессы 662–664
- общий случай n-мерной коинтегрированной системы 695–697
- определение 666–670
- оценивание динамическим методом наименьших квадратов 689, 690, 693, 695, 696, 700–701, 703
- оценка SOLS 688–689, 700
- представление в форме векторной модели коррекции ошибок 675–677
- приложение к спросу на деньги в США 697–703
- проверка нулевой гипотезы об отсутствии коинтеграции 687
- свойства оценок в конечных выборках 697
- статистические выводы о векторах 688
- тест, основанный на остатках 681–687
- треугольное представление Филлипса 671–673

Коинтегрирующая регрессия 672
Коинтегрирующий вектор 667
Колмогорова второй усиленный закон больших чисел 122
Концентрированная логарифмическая функция правдоподобия 558
Концентрированная логарифмическая функция правдоподобия:
- для линейной модели регрессии 25
- для метода максимального правдоподобия с полной информацией 559
- для многомерной модели регрессии 557

Коррекция степеней свободы 154
Коррелограмма 127
Корреляция между уравнениями, или одновременная, корреляция 288
Коши последовательность 404
Коши — Шварца неравенство 198
Коэффициент автокорреляции j-го порядка 127
Коэффициент детерминации 42–43
Коэффициент детерминации нецентрированный 42–43
Коэффициент проекции метода наименьших квадратов 168
Коэффициенты проекции метода наименьших квадратов 168
Крамера — Рао:
- граница 73–75
- нижняя граница 70

Критерии выбора теста на единичный корень:
- асимптотики локальные к единице 637
- свойства на малых выборках 638–639

Кронекера дельта 196
Кронекера произведения 711

Л

Лаговые полиномы:
- абсолютно суммируемые обращения 404
- обращения 402–404

Лаговый полином степени p 399
Лагранжа множителей статистика 524–528
Линдеберга — Леви центральная предельная теорема 134
Линдеберга — Феллера центральная предельная теорема 194
Линейная по переменным 532
Линейная регрессионная модель:
- асимптотическая нормальность 509–510
- идентификация в методе максимального правдоподобия 497–498
- с нормальными ошибками 482

Линейной гипотезы проверка 145–146
Линейные векторные I(0)- и I(1)-процессы 662–664
Линейные оценки обобщенного метода моментов 485–486
Линейные процессы:
- линейные I(0)-процессы 599–600
- описание 394–395
- скользящего среднего бесконечного порядка 395–398
- скользящего среднего порядка q 395

Линейный процесс скользящего среднего бесконечного порядка 395–399, 409–410, 431
Линейный процесс скользящего среднего порядка q 395
Логарифм дохода минус логарифм потребления 686
Логарифмическая функцию правдоподобия 71, 72
Логарифмическое условное правдоподобие 481

Логит-модель 540
Логлинейное уравнение 87
Ложная регрессия 680–681
Ложное детрендирование 605
Локальные к единице асимптотики 638
Лонгитюдные данные 352
Люнга — Бокса Q-статистика 171–173

М

М-оценки:
- асимптотическая нормальность 502–506
- как экстремальная оценка 477–478
- описание 477
- состоятельность 490–492

Максимального правдоподобия принцип:
- для оценивания параметров модели 69
- основная идея 70
- связь с обобщенным методом моментов 18

Мартингал по отношению к $\{z_i\}$ 129
Мартингал-разность 131
Мартингалы 129–130
Математические обозначения 21
Матрица данных 28
Матричные обозначения 27–28
Мера риска для стандартного нормального распределения 510
Метод
- минимального расстояния 236
- моментов 235
- наименьших квадратов с ограничениями 66

Модели с качественным откликом:
- асимптотическая нормальность 542
- вклад и гессиан для наблюдения t 540–541
- описание 539–540
- состоятельность 541

Модели:
- определение 25
- правильно специфицированные 478

Модель компонент ошибок:
- индивидуальный эффект 354
- описание 352, 353
- отклонения от групповых средних и модель компонент ошибок 356
- оценка фиксированных эффектов и модель компонент ошибок 358–370
- репараметризация модели 357–358

Модель коррекции ошибок векторная 671
Модель с ограниченной зависимой переменной 543
Модель с множественным выбором 539
Мощность критерия 147

Н

Наблюденный уровень значимости, p-значение 61
Наилучшая линейная несмещенная оценка 49
Наилучшая несмещенная оценка 75
Неоднородность индивидов 354
Неравенство Крамера — Рао 73–74
Неравенство треугольника 458
Номинальный размер теста 146

О

Обменные курсы 616, *см. также* реальный обменный курс
Обнуление:
- пропущенные наблюдения 367
- против сжатия 368–369

Обобщенная нелинейная оценка инструментальных переменных 485
Обобщенного метода моментов для одного уравнения оценки:
- асимптотическая мощность 243–244
- асимптотическая теория 238–258
- асимптотические свойства 238–244
- асимптотическое распределение 238–239
- в случае преопределенных регрессоров 261
- для конечных выборок 244
- определение 236
- оценивание дисперсии ошибки 239–240
- оценивание долговременной ковариационной матрицы 241
- получение двухшаговой оценки наименьших квадратов из эффективного обобщенного метода моментов для одного уравнения 256–261
- последствия условной гомоскедастичности 255–264
- тестирование гипотез 240–245
- трехшаговый метод наименьших квадратов 18
- эффективная 242–243
- эффективная с ограничением 251

Обобщенного метода моментов для нескольких уравнений оценки:
- асимптотическая теория 296–299
- вывод 294–296
- инструментальная оценка с полной информацией 304–305
- кажущиеся несвязанными регрессии 308–310
- как оценки случайных эффектов 318–322

- общие коэффициенты 314–323
- объединенная OLS-оценка 319–320 322
- оценка трехшагового метода наименьших квадратов 305–308
- преимущества 287
- сравнение SUR и OLS 310–312
- сравнение со случаем одного уравнения 299–303
- условная гомоскедастичность 303–304
- условная гомоскедастичность эффективной оценки 317–320

Обобщенного метода моментов для одного уравнения оценки:
- сравнение с обобщенным методом моментов для нескольких уравнений 299–303

Обобщенный метод моментов:
- инкорпорирование сериальной корреляции 436–442
- инкорпорирование сериальной корреляции и связь с обобщенным методом наименьших квадратов 445–447
- общая формулировка 228–233
- описание 234–237
- приемы 18
- процедуры оценивания 18
- связь между методом максимального правдоподобия и обобщенным методом моментов 18
- смещение вследствие эндогенности 218–223
- *см. также* обобщенного метода моментов для нескольких уравнений оценки

Обобщенный метод наименьших квадратов:
- инкорпорирование сериальной корреляции в обобщенный метод моментов и связь с обобщенным методом наименьших квадратов 445–447
- ограничивающее свойство обобщенного метода наименьших квадратов 82–83
- описание 78
- последствия ослабления предположения 79
- свойства на конечных выборках 48–53
- эффективное оценивание с известной матрицей V 79–81

Образование:
- отдача от образования 265–273
- смещение из-за способностей 271–272
- тестирование предопределенности в уравнении заработной платы 250
- уравнение заработной платы 217
- эндогенность 272

Обратимости условие 412

Обратное отношение Миллса 510, 545

Обращение блочных матриц 710–711

Обращения:
- лаговых полиномов 402–404
- фильтра 401–402
- фильтров для векторных процессов 418

Общие коэффициенты:
- как неограничивающее условие в обобщенном методе моментов для нескольких уравнений 322–323
- обобщенный метод моментов для нескольких уравнений и общие коэффициенты 314–323

Объединенная оценка наименьших квадратов 319–320

Объясняющие переменные 25

Обычный метод наименьших квадратов 18

Односторонний t-тест 69

Ортогональная 30

Отклонения
- от групповых средних 356
- от среднего 169

Отношения правдоподобий принцип:
- F-отношение 66
- тестирование гипотез 251–255

Отношения правдоподобий статистика 526–527

Отношения правдоподобий тест 76

Оценивание методом максимального правдоподобия:
- асимптотическая нормальность условного 506–511
- идентификация в линейной модели регрессии 497–498
- идентификация условной плотности 494–495
- инвариантность 483–484
- Йохансена 677–679
- как M-оценка 477
- максимальное правдоподобие с ограниченной информацией 542, 571
- максимальное правдоподобие с полной информацией 559–570
- многомерные регрессии 539–542
- модели с качественным откликом 539–542
- процессов AR(p) и VAR(p) 583–584
- процессов AR(1) 579–583
- сериально коррелированные наблюдения 576–584
- состоятельность 495–497
- сравнение с асимптотической нормальностью обобщенного метода моментов 514–515

- тобит (цензурированная регрессия) модели 550–553
- усеченные модели регрессии 543–550
- условное 481–483

Оценивание спроса на деньги 697–703
Оценивания процедура 24
Оценка динамического метода наименьших квадратов 689, 690, 693, 695, 696, 700–701, 703
Оценка k-класса 575
Оценка ковариационная 360
Оценка максимального правдоподобия точная 579
Оценка максимального правдоподобия, условная относительно y_0 481
Оценка наименьших квадратов статическая 689–700
Оценки максимального правдоподобия с ограниченной информацией:
- вычисление 573–575
- описание 571–573
- сравнение с двухшаговым методом наименьших квадратов 575

Оценки максимального правдоподобия с полной информацией:
- исходная информация 559–563
- оценивание методом максимального правдоподобия 569–570
- полная система одновременных уравнений 562–563
- свойства 567–569
- связи между коэффициентами 563–564
- тестирование сверхидентифицирующих ограничений 566–567
- функция правдоподобия 564–565

Оценки минимального расстояния классические 478
Оценки с ограничениями 521
Ошибки в переменных 224–226

П

Панель несбалансированная:
- асимптотические свойства оценки фиксированных эффектов 369–370
- «обнуление» пропущенных наблюдений 367–368
- «обнуление» против сжатия 368–369
- описание 367
- смещение из-за отбора 367, 369

Панельное исследование динамики доходов 225
Панельные данные:
- модель компонент ошибки 359–370
- несбалансированные 367–370
- описание 279, 289, 352
- пространство параметров 476

Параметризация функции плотности 70
Паритет покупательной способности 639–640
Перекрестные ограничения 327
Переменная в левой части 25
Переменные в правой части 25
Перманентная компонента 597
Перманентного дохода гипотеза 224–226
Питменовская последовательность 149
Питменовский снос 149
Поддерживаемая гипотеза 55
Подобранное значение 40
Подход
- в частотной области 414
- во временной области 414

Покрытый процентный паритет, уравнение 449
Полный столбцовый ранг 32
Пониженного ранга условие 675
Последовательное t-правило от общего к частному 424–425
Последовательности случайных величин:
- взгляд на оценки как на последовательности случайных величин 121
- обзор предельных теорем 115–121

Последовательность локальных альтернатив 148
Постоянные реальные процентные ставки 181
Правдоподобие безусловное 70–71
Правильно специфицированная 478
Правильно специфицированная модель 302
Предел по вероятности 115
Предельная склонность к потреблению 225–226
Предельное распределение 117
Предопределенные регрессоры 137, 229, 261
Представление в виде скользящего среднего 405–408
Представление в форме векторного скользящего среднего 662–664
Представление в форме коррекции ошибок 675
Преходящая компонента 596
Приведенная форма 563
Приведенной формы коэффициенты 563
Пробит-модель:
- асимптотическая нормальность оценки максимального правдоподобия 510–511
- идентификация при оценивании методом максимального правдоподобия 498–499
- условное максимальное правдоподобие 482–483

Прогноза ошибка, иена/доллар 451
Проекционная матрица $\boldsymbol{P}$ 40–41

Проекция методом наименьших квадратов:
- наилучший линейный прогноз, использующий проекцию методом наименьших квадратов 167–169
- описание 166–167
Производящая функция автоковариаций 413–415
Проинтегрированный d раз процесс авторегрессии — скользящего среднего 602
Проинтегрированный процесс авторегрессии — скользящего среднего 602
Простая регрессионная модель 26
Пространство параметров 476
Процентные ставки:
- инфляция и процентные ставки 183
- номинальные 185–187
- *см. также* реальные процентные ставки
Процесс авторегрессии порядка p 583–584
Процесс независимого белого шума 128
Процесс порождения данных 137–584
Процесс с единичным корнем 594
Процессы авторегрессии — скользящего среднего:
- ARMA(pq) 410–411, 428–429
- автоковариации AR(1) 408–412
- как инструмент эконометрики единичных корней 601–602
- обратимость 412
- описание 405
- производящая функция автоковариаций и спектр 413–415
- процесс авторегрессии первого порядка 405–408
Процессы с авторегрессионной гетероскедастичностью 132
Процессы с авторегрессионной гетероскедастичностью 133

Р

Равенство для информационной матрицы 73
Равномерная сходимость по вероятности 488
Равномерный закон больших чисел 490
Разложение в среднем значении 503
Расширенная авторегрессия 621–622
Расширенная коинтегрирущая регрессия:
- допущение сериальной корреляции 692–695
- пример с двумя переменными 690–692
Расширенные тесты Дики — Фуллера:
- без постоянной составляющей 626–627
- включение в регрессию постоянной составляющей 630–632
- включение временного тренда 633–635
- вывод статистик тестов 625–627
- для казначейских обязательств 632
- предельное распределение оценки наименьших квадратов 622-625
- предложение для выбора $p_{\max}(T)$ 629–630
- при неизвестном значении p 627–635
- расширенная авторегрессия 621–622
- с постоянной составляющей 630–632
Реализация или выборочная траектория 124
Реальный обменный курс 641, *см. также* Обменные курсы
Регрессии коэффициенты 25
Регрессии на время:
- оценивание методом наименьших квадратов 192
- предположения 190
- тестирование гипотез 192–193
Регрессии функция 25
Регрессируемая переменная 25
Регрессия 25
Регрессия с ограничениями 66
Регрессоры:
- описание 25
- предопределенные 261, 229
- фиксированные 34, 194
- эндогенные 218–219, 229
Репараметризация:
- вогнутость после 492
- максимального правдоподобия 483–484
- тобит-модели 550
- усеченных моделей регрессии 547
Робастная стандартная ошибка 145
Робастное t-отношение 146–147
Рост вложений в оборудование и инвестиции в Ботсване 45

С

Саргана статистика 257–258
Саргана — Бхаргавы статистика 620
Сбалансированная панель 367
Свертка 400
Сверхидентифицировано 89, 232
Сдвиги предложения:
- смещение из-за эндогенности 223–227
- Уоркинга пример 229
Сериальная корреляция:

- Гордина условие ограничивающее степень сериальной корреляции 436–437
- асимптотики для выборочных средних 430–435
- векторные процессы 395–447
- инкорпорирование в обобщенный метод моментов 436–442
- инкорпорирование в обобщенный метод моментов и связь с обобщенным методом наименьших квадратов 445–447
- линейные процессы 394–404
- оценивание авторегрессий 421–429
- оценивание долговременной ковариационной матрицы 437–447
- процессы авторегрессии — скользящего среднего 405–415
- расширенная коинтегрирующая регрессия и 692–695
- форвардная премия 448–456

Сериальной корреляции тестирование:
- Бокса — Пирса и Люнга — Бокса 171–173
- Бройша — Годфри 178
- выборочные автокорреляции, вычисляемые по остаткам 173–175
- описание 170
- основанное на вспомогательной регрессии 176–178
- с предопределенными регрессорами 175–176

Скользящего среднего процесс бесконечного порядка 397

Скользящего среднего процесс второго порядка 201–202

Скользящего среднего процесс порядка q 395

Скорость конвергенции 371

Слабо (или ковариационно) стационарный случайный процесс 126

Слуцкого теорема 119

Случайные блуждания:
- аппроксимация I(1)-процесса 600–601
- жизнеспособность 640–641
- как пример мартингала 130
- обменные курсы как случайные блуждания 615–616
- описание 130
- со сносом 407
- тесты Дики — Фуллера на отсутствие сноса 608–616

Случайные выборки 34

Случайных эффектов метод:
- как обобщенный метод моментов для нескольких уравнений 318
- как частный случай обобщенного метода моментов 18
- сравнение случайных и фиксированных эффектов 363–365

Смещение вследствие одновременности 220

Смещение из-за отбора 367 369

Смещение из-за способности 271–272

Смещение из-за эндогенности:
- наблюдаемые сдвиги предложения 220–223
- описание 218–220
- ошибки в переменных как пример 224–226
- пример со склонностью к потреблению 224
- производственная функция как пример 226–227

Смещение, обусловленное селективностью выборки 545

Собственные значения 465

Совета управляющих федеральной резервной системы США веб-сайт 206

Сопровождающая форма 410, 464

Состоятельная при гетероскедастичности и автокоррелированности 437

Состоятельная при гетероскедастичности и автокоррелированности оценка векторной авторегрессии 440–442

Состоятельная при гетероскедастичности стандартная ошибка 145–146

Состоятельность для экстремальных оценок:
- для M-оценок 490–492
- теоремы 487–489

Состоятельные оценки 121

Состоятельные тесты 147–148

Спектральная плотность (функция) 414

Спецификационный тест 247

Спроса и предложения модель:
- наблюдаемые сдвиги предложения 220–223
- смещение вследствие эндогенности 218–219

Среднее по ансамблю 124

Среднеквадратичная ошибка 166

Стабильности условие 403

Ставки по казначейским обязательствам, расширенный тест Дики – Фуллера 621, 632

Стандартная ошибка
- оценки наименьших квадратов 58
- регрессии (SER) 41
- уравнения (SEE) 41

Стандартный винеровский процесс
- детрендированный 605
- центрированный 605

Статистический сборник Министерства финансов 206

Стационарности условие 406

Стационарность 126

Стационарные относительно тренда процессы 595–597

Стационарный в разностях 126
Стационарный в разностях процесс 594
Стационарный относительно тренда 126, 190
Степени свободы 154
Стерлинг/доллар реальные обменные курсы 641
Стохастическая коинтеграция 668
Стохастические процессы:
- ковариация 126–127
- описание 124
- различные классы 125–128
- *см. также* последовательность случайных величин
Стохастический тренд 592
Строгая мультиколлинеарность 32
Строгая экзогенность 28–32
Структурная форма 562
Структурной формы параметры 562
Сумма квадратов
- остатков 37
- остатков (SSR) 36
- остаточная (RSS) 36
- ошибок (ESS) 36
Суперсостоятельные оценки 192, 610
Сходимость
- в среднеквадратичном 116
- по вероятности точечная 487
- по распределению 116
- почти наверное 115
Сходится по вероятности:
- асимптотически эквивалентные последовательности 119
- описание 115
- почти наверное 115
- Слуцкого теорема 119
- сохранение для непрерывных преобразований 118
США:
- единичный корень в ВВП 634
- индексированные на инфляцию облигации, проданные с аукциона 1997 г. 188
- реальные обменные курсы стерлинг/доллар 641
- спрос на деньги 697–703
- электроэнергетическая отрасль в США 84–91, 334
- *см. также* доллар/иена

Т

Тейлора разложение 515–518, 521–522
Телескопическая сумма 433–434
Теорема Гаусса — Маркова, 48, 49, 80
Теорема Грейнджера о представлении 677
Теорема о среднем значении 501, 503, 511
Теория для больших выборок:
- для обобщенного метода моментов для одного уравнения 238–258
- инфляция и теория для конечных выборок 187–189
- описание 114
- оценивание с параметризованной условной гетероскедастичностью 161–166
- последствия условной гомоскедастичности 154–158
- предельные теоремы для последовательностей случайных величин 115–124
- проекция методом наименьших квадратов 166–167
- распределение оценки наименьших квадратов 136–144
- регрессии на время 190–197
- состоятельное оценивание 151–153
- тестирование
  - - линейных гипотез 145–149
  - - на сериальную корреляцию 170–178
  - - нелинейных гипотез 150–151
  - - условной гомоскедастичности 159–161
- эконометрика рациональных ожиданий 179–189
Теория для конечных выборок 187
Теория для больших выборок:
- для обобщенного метода моментов для нескольких уравнений 296–299
Тест Чоу 105, 641
Тест на коинтегрированность, основанный на остатках 681–685
Тестирование гипотез:
- в предположении нормальности 55–57
- в экстремальных оценках 520–528
- для регрессий на время 192–193
- добавление переменной 253–255
- линейных 61–63
- о предопределенности образования в уравнении заработной платы 250–251
- о подмножествах условий ортогональности 248–251, 262–263
- о сверхидентифицирующих ограничениях 246–251
- об отдельных коэффициентах регрессии 57–59
- при оценивании обобщенным методом моментов 240–245
- с использованием принципа отношения правдоподобий 251–255
- условной гомоскедастичности 159–162, 264
Тестирование подмножеств условий ортогональности 248–253, 264, 265

Тестирование сверхидентифицирующих ограничений 246
Тестовая статистика вклада 525
Тобит (цензурированная регрессия) модели 550–553
Точный размер теста 146
Транслогарифмическая функция издержек:
- доли факторов 326–327
- многомерная регрессия с перекрестными ограничениями 331–333
- описание 94, 325–326
- природа ограничений 330–331
- свойства 328–329
- стохастические спецификации 329–330
- уравнения для 333
- эластичности замещения 327–328
- электрические сети общего пользования в США 334

Треугольное представление 672
Трехшаговый метод наименьших квадратов:
- как обобщенный метод моментов для одного уравнения 18
- как обобщенный метод моментов для системы уравнений 305–308

Триады статистик резюме 527–528

У

Уайта стандартная ошибка 145–146
Уайта тест на условную гомоскедастичность 160
Уравнение недоидентифицированное 232
Уравнение сверхидентифицированное 89
Уравнение идентифицируемое точно 232
Уравнения правдоподобия 78
Уровни IQ:
- как мера способностей 268
- ошибки в переменных 224–271

Усечения эффект 544
Усеченное ядро 439
Усеченные модели регрессии:
- описание 543
- проверка состоятельности и асимптотической нормальности 547–549
- распределения 544–546
- репараметризация функции правдоподобия 547
- функция правдоподобия 546–547

Условие Линдеберга 194
Условие идентифицируемости
- порядковое 232
- ранговое 230–232

Условная гомоскедастичность:
- безусловная и условная 155
- варианты асимптотического теста при условной гомоскедастичности 157–158
- обобщенного метода наименьших квадратов для системы уравнений 303–304
- определение 34
- оценивание долговременной ковариационной матрицы при условной гомоскедастичности 443–447
- оценивание параметризованной условной гомоскедастичности 161–166
- оценка фиксированных эффектов и ослабление требования уловной гомоскедастичности 356
- оценки обобщенного метода моментов и последствия условной гомоскедастичности 256–264
- собственная 133
- тестирование 159–161, 264
- эффективная оценка обобщенного метода моментов для нескольких уравнений и условная гомоскедастичность 317–320
- *см. также* гомоскедастичностьядерное оценивание долговременной ковариационной матрицы при условной гомоскедастичности 438–439

Условное максимальное правдоподобие 481–483
Условное правдоподобие 70–71

Ф

Фактический размер теста 146
Федеральная энергетическая комиссия 85
Федерального резерва бюллетень (Federal Reserve Bulletin) 206
Фиксированные регрессоры:
- асимптотики с фиксированными регрессорами 194
- описание 34

Фиксированный эффект 354
Фиксированных эффектов оценка:
- несбалансированные панели 367–370
- описание 356
- ослабление требования условной гомоскедастичности 365–366
- свойства на больших выборках 360–362
- сравнение со случайными эффектами 363–365
- формула 359–360

Филлипса тест 637
Филлипса треугольное представление 671–673

Филлипса — Перрона тест 637
Фильтр коммутативный 400
Фильтрация:
- ковариационно-стационарных процессов 126
- описание 398
- производящая функция автоковариаций 413
Фильтры:
- абсолютно суммируемые 400
- векторные процессы 416–418
- коммутативные 400
- многомерные 416–417
- обращения 401–402
- описание 399
- произведение фильтров 400–401
Фирм гетерогенность 86
Форвардная премия:
- гипотеза эффективного рынка 449–450
- как оптимальные предикторы 448
- описание 448
- регрессионные тесты 453–457
- средние темпы изменения за период 1975–89, 453
- тестирование равенства нулю безусловного среднего 450–453
Формулировка отсутствия сериальной зависимости 133
Фриша — Во теорема 98–99
Функции правдоподобия:
- для многомерных регрессий 555–558
- для регрессионных моделей 71
- квази- (псевдо-) максимум 76
- максимизация через концентрацию 71–73
- метод максимального правдоподобия 478–480
- описание 70
- оценок максимального правдоподобия с полной информацией 564–565
- тобит-модель 551–552
- усеченной модели регрессии 546–547
- условные и безусловные, 70–71
Функциональная центральная предельная теорема 598, 604
Функция
- потребления 26
- риска 545

Х

Хансена тест на сверхидентифицирующие ограничения 247–248
Характеристические корни 465
Хаусмана статистика 263
Холла мартингальная гипотеза 130

Ц

Цензурированная модель регрессии 543
Центральные предельные теоремы:
- Гордина для эргодических случайных процессов с нулевым средним 434
- Линдеберга — Леви 122, 134–135, 432
- Линдеберга — Феллера 194
- для MA(∞) 431–432
- для нормально распределенных ошибок 55–57
- для эргодических стационарных последовательностей мартингал-разностей 134–135
- функциональная 603, 604

Ч

Частота 414
Чебышева слабый закон больших чисел 122, 431
Числами с плавающей запятой 46
Чоу тест 106
Чоу тест на структурное изменение 105

Ш

Шварца (информационный критерий) правило 425–426
Ширина окна 439

Э

Эйлера уравнение 485–486
Эконометрика единичных корней:
- выбор теста на единичный корень 637–639
- Дики — Фуллера тесты 608–619
- инструментарий 598–607
- использование винеровского процесса 602–605
- использование линейных I(0)-процессов 593
- использование моделей авторегрессии — скользящего среднего 601–602
- моделирование трендов 591–598
- приложение: паритет покупательной способности 639–641
- тренды для интегрированных процессов 593–595
- упражнения на метод Монте-Карло 647–649
Эконометрика рациональных ожиданий:
- гипотеза эффективного рынка 179–187
- номинальная процентная ставка как оптимальный предиктор 185, 187
- описание 179
- последующее развитие 188–189
- тестирование на сериальную корреляцию 183
Экстремальные оценки:

- М-оценки как экстремальные оценки 477–480
- асимптотическая нормальность 501–522
- измеримость 476–477
- классы 477–478
- нелинейный метод наименьших квадратов как экстремальная оценка 477
- описание 476
- оценка обобщенного метода моментов как экстремальная оценка 477–478
- проверка гипотез 520–528
- состоятельность 487–500
- теоремы о состоятельности 487–489
- численная оптимизация 529–532

Электроснабжающая отрасль (США):
- данные 84–85
- метод наименьших квадратов с ограничениями 89
- определение выпуска 86
- особенности 84
- применение технологии Кобба — Дугласа 86–88
- применение эконометрики 85–86
- результаты использования транслогарифмической функции издержек 334
- тестирование однородности функции издержек 90–91
- эндогенность выпуска 95
- *см. также* СШАЭндогенные переменные 560

Эндогенные регрессоры:
- вызываемые ими смещения 218
- описание 218–219, 229

Энергетический спектр 414

Эргодическая стационарность:
- анализ временных рядов и эргодическая стационарность 124–125
- Гордина условие для 431–432
- описание 128

Эргодическая теорема 128–129

Эффективная оценка инструментальных переменных с полной информацией:
- обобщенного метода моментов для нескольких уравнений 294–296
- свойства на больших выборках 305

Эффективная оценка обобщенного метода моментов с ограничениями 251

Эффективности рынка гипотеза 179–181, 185–187

Ю

Юла — Уокера уравнения 408, 463

Я

Ядерные оценки 438–440

Ядро спектральное квадратичное 440

ADF-GLS-тест 637

CAN 121

DGLS-оценка 695

F-отношение:
- вывод на основе принципа отношения правдоподобий 66
- распределение 63–65
- формулы, содержащие остаточные суммы квадратов в регрессиях без ограничений и с ограничениями 65–66

F-тест:
- в сравнении с t-тестом 66–67
- используемые принципы 63–66
- как тест отношения правдоподобий 76

I(0)-процесс 593

I(1)-процессы:
- аппроксимация случайным блужданием 600–601
- важность 595–597
- идентификация 597
- коинтегрированные системы и I(1)-процессы 661–664
- со сносом 594
- типы 593–595

I(1)-процесс без сноса 594

I(d)-процессы 593

Journal of Business and Economic Statistics 248

KWW-тест 266

Limit observations 551

Monthly Labor Review 206

OLS-дамми-оценка 360

QR-декомпозиция 46

t-значение 58

t-отношение:
- описание 58
- распределение 58–60
- робастное 146–147

t-тест

- двусторонний 69
- односторонний 69

t-тест:
- решающее правило 59–60
- сравнение с F-тестом 66–67

Young Men's Cohort of National Longitudinal Survey (NLS-Y) 266

Учебное издание

Серия «Академический учебник»

Фумио Хайяши
Эконометрика

Главный редактор *В.В. Анашвили*
Заведующая редакцией *Ю.В. Бандурина*
Выпускающий редактор *Е.В. Попова*
Редактор *О.В. Черкасова*
Художник *Е.Н. Спасская*
Оригинал-макет *О.З. Элоев*
Верстка *С.А. Кулешов*
Корректор *О.Г. Соколова*

Подписано в печать 16.01.2017. Формат 70x108 $^{1}/_{16}$
Гарнитура AntiquaPSCyr. Усл. печ. л. 63,7
Тираж 1000 экз. Изд. № 106. Заказ № 1417

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