УЧЕБНЫЕ ИЗДАНИЯ ДЛЯ БАКАЛАВРОВ

А. И. Новиков

ЭКОНОМЕТРИКА

учебное пособие

$$\begin{cases} a + b\overline{x} = \overline{y} \\ a\overline{x} + b\overline{x^2} = \overline{xy}, \end{cases}$$

$$\begin{cases} b = \dfrac{cov(x,y)}{var(x)} = \dfrac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - (\overline{x})^2} \\ a = \overline{y} - b\overline{x}. \end{cases}$$

$$var(y) = \frac{1}{n}\sum(y_i - \overline{y})^2$$

$$F = S_б^2 / S_м^2$$

«Дашков и К°»
Издательско-торговая корпорация

Серия «Учебные издания для бакалавров»

А. И. Новиков

ЭКОНОМЕТРИКА

Учебное пособие

3-е издание

Рекомендовано
федеральным государственным бюджетным учреждением
«Федеральный институт развития образования» (ФГБУ «ФИРО»)
в качестве учебного пособия для использования в образовательном процессе образовательных организаций, реализующих программы высшего образования по направлениям подготовки
«Экономика», «Менеджмент»
(уровень бакалавриата)

Регистрационный номер рецензии 527 от 04 февраля 2018 г.

Москва
Издательско-торговая корпорация «Дашков и К°»
2021

330.115(07)
УДК 336.02
ББК 65.26
Н73 1

Автор:
А. И. Новиков — доктор физико-математических наук, профессор.
Рецензенты:
В. Е. Поляк — кандидат физико-математических наук, член-корреспондент Международной академии информатизации;
М. В. Дуброва — кандидат экономических наук, профессор.

Новиков А. И.

Эконометрика: Учебное пособие / А. И. Новиков. — 3-е изд. — М.: Издательско-торговая корпорация «Дашков и К°», 2021. — 224 с.

ISBN 978-5-394-04051-1

Учебное пособие содержит основные теоретические положения эконометрики, а также большое количество примеров и упражнений.

Для студентов бакалавриата, обучающихся по направлениям подготовки «Экономика», «Менеджмент».

ISBN 978-5-394-04051-1

СОДЕРЖАНИЕ

ВВЕДЕНИЕ В ЭКОНОМЕТРИКУ

Предмет эконометрики

Закономерности в экономике выражаются в виде зависимостей экономических показателей и математических моделей их поведения. Такие зависимости и модели могут быть получены только путем обработки реальных статистических данных, с учетом внутренних связей и случайных факторов.

Эконометрика — наука, изучающая количественные закономерности и взаимозависимости в экономике методами математической статистики.

Цель эконометрики — эмпирический вывод экономических законов.

Задачи эконометрики — построение экономических моделей и оценивания их параметров, проверка гипотез о свойствах экономических показателей и формах их связи.

Эконометрический анализ служит основой для экономического анализа и прогнозирования, создавая возможность для принятия обоснованных экономических решений.

Типы данных

При моделировании экономических процессов оперируют с двумя типами данных: пространственными и временными.

Пространственные данные — это данные по какому-либо экономическому показателю, полученные от разных однотипных объектов (фирм, регионов и т. п.), но относящиеся к одному и тому же моменту времени (пространственный срез). Например, данные об объеме производства, количестве работников, доходе разных фирм в один и тот же момент времени.

Временные ряды — это данные, характеризующие один и тот же объект в различные моменты времени (временной срез). Например, ежеквартальные данные об инфляции, средней заработанной плате, данные о национальном доходе за последние годы.

Отличительная черта временных данных — упорядоченность во времени. Кроме того, наблюдения в близкие моменты времени часто бывают зависимы.

Любые экономические данные представляют характеристики какого-либо экономического объекта. Они формируются под воздействием множества факторов, не все из которых доступны внешнему контролю. Неконтролируемые (неучтенные) факторы обусловливают случайность данных, которые они определяют.

Поскольку экономические данные имеют статистическую природу, для их анализа и обработки необходимо применять статистические методы.

Классы моделей

Можно выделить три основных класса моделей: модели временных рядов, регрессионные модели с одним уравнением и системы одновременных уравнений.

К моделям временных рядов относятся *модели тренда и сезонности*. Тренд представляет собой устойчивое изменение уровня показателя в течение длительного времени. Сезонность характеризует устойчивые внутригодовые колебания уровня показателя.

Кроме того, к этому классу относится множество более сложных моделей, таких, например, как модель с распределенным лагом, модель авторегрессии и модель адаптивных ожиданий.

Их общей чертой является то, что они объясняют поведение временного ряда исходя только из его предыдущих и будущих значений.

Модели временных данных подразделяются также на модели, построенные по стационарным и нестационарным временным рядам.

Стационарные временные ряды — это ряды, которые имеют постоянное среднее значение и колеблются вокруг него с постоянной дисперсией. Распределение показателей уровня ряда в них не зависит от времени, т. е. такой временной ряд не содержит трендовой или сезонной компоненты.

В *нестационарных* временных рядах распределение уровня ряда зависит от времени, т. е. ряд имеет трендовую или сезонную компоненту.

В ***регрессионных моделях с одним уравнением*** объясняемая переменная представляется в виде функции от объясняющих переменных. Например, модель спроса на некоторый товар в зависимости от его цены и дохода.

По виду функции регрессионные модели подразделяются на *линейные* и *нелинейные*. Существуют эффективные методы оценки и анализа линейных регрессионных моделей. Анализ линейных регрессионных моделей является базовым в прикладной эконометрике.

Область применения регрессионных моделей, даже линейных, значительно шире, чем моделей временных рядов.

Системы одновременных уравнений описываются системами уравнений, состоящими из тождеств и регрессионных уравнений, в каждом из которых аргументы помимо объясняющих переменных содержат объясняемые переменные из других уравнений системы. Например, модель формирования доходов.

Все три класса моделей могут использоваться при моделировании экономических процессов.

Обычно предполагают, что все факторы, не учтенные явно в экономической модели, оказывают на объект некое результирующее воздействие, величина которого задается случайной компонентой.

Введение случайной компоненты в экономическую модель делает ее доступной для эмпирической проверки на основе статистических данных.

Типы зависимостей

В экономических исследованиях одной из основных задач является анализ зависимостей между переменными. Зависимость может быть функциональной либо статистической.

Функциональная зависимость задается в виде точной формулы, в которой каждому значению одной переменной соответствует строго определенное значение другой, воздействием случайных факторов при этом пренебрегают.

В экономике функциональная зависимость между переменными проявляется редко.

Статистической зависимостью называется связь переменных, на которую накладывается воздействие случайных факторов. При этом изменение одной переменной приводит к изменению математического ожидания другой переменной.

По направлению выделяют прямые и обратные связи. Направление изменения результативного признака при *прямой связи* совпадает с направлением изменения факторного признака, а при *обратной связи* — противоположно направлению изменения факторного признака.

Уравнение регрессии — это формула статистической связи между переменными. Если эта формула линейна, то имеем линейную регрессию.

Формула статистической связи *двух* переменных называется *парной регрессией*, зависимость от *нескольких* переменных — *множественной регрессией.*

1. ЭЛЕМЕНТЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

1.1. Операция суммирования

Пусть величина X задается последовательностью данных x_1, $x_2,\ldots, x_n$, каждое из которых можно записать как x_i, $i = 1,n$.

Сумма этих чисел обозначается следующим образом:

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \ldots + x_n .$$

Если из контекста ясно, каковы начальный и конечный суммируемые члены, то часто используют сокращенные обозначения:

$$\sum_{i=1}^{n} x_i = \sum x_i .$$

Сумма квадратов этих чисел обозначается следующим образом:

$$\sum x_i^2 = x_1^2 + x_2^2 + \ldots + x_n^2 .$$

Обозначим средние значения величин X, X^2, XY соответственно как

$$\bar{x} = \frac{\sum x_i}{n}, \quad \overline{x^2} = \frac{\sum x_i^2}{n}, \quad \overline{xy} = \frac{\sum x_i y_i}{n} .$$

Имеет место неравенство: $(\bar{x})^2 \le \overline{x^2}$.

Правила суммирования (a, b – константы):

1. $\sum a = na$.
2. $\sum b x_i = b \sum x_i = b n \bar{x}$.
3. $\sum (a + b x_i) = na + bn\bar{x}$.
4. $\sum (x_i + y_i) = \sum x_i + \sum y_i = n(\bar{x} + \bar{y})$.

5. $\sum (x_i - \bar{x}) = 0$.

6. $\dfrac{\sum (x_i - \bar{x})^2}{n} = \overline{x^2} - (\bar{x})^2$.

7. $\dfrac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n} = \overline{x\,y} - \bar{x}\,\bar{y}$.

1.2. Случайные величины

Случайной величиной (переменной) называется величина, которая под воздействием случайных факторов может с определенными вероятностями принимать те или иные значения из некоторого множества чисел.

Случайные величины обозначаются заглавными буквами латинского алфавита, например X, а их возможные значения — строчными, x.

Универсальным способом задания случайной величины X является задание ее функции распределения.

Функцией распределения $F(x)$ случайной величины X называется вероятность того, что величина X принимает значение, меньшее x, т. е.

$$F(x) = P(X < x),\ x \in R.$$

Различают дискретные и непрерывные случайные величины.

Дискретной называется случайная величина, которая принимает отдельные, изолированные друг от друга значения.

Непрерывной называется случайная величина, множество значений которой непрерывно заполняет некоторый числовой промежуток

В основе математической статистики лежат понятия генеральной и выборочной совокупностей.

Генеральная совокупность — множество всех значений случайной величины, которые она может принять в процессе наблюдения. Например, данные о доходах всех жителей страны.

Выборочная совокупность (выборка) — это множество наблюдений, составляющих лишь часть генеральной совокупности.

Для любой случайной величины важную роль помимо функции распределения играют числовые характеристики ее распределения.

Числовые характеристики совокупности

I. Выборочная совокупность

Пусть из генеральной совокупности извлекается выборка объема *n*.

Выборочной средней называется среднее арифметическое наблюдаемых значений случайной величины в выборке, т. е.

$$\bar{x} = \frac{1}{n}\sum x_i .$$

Выборочной дисперсией (вариацией) называется среднее арифметическое квадратов отклонения наблюдаемых значений случайной величины в выборке от их среднего значения, т. е.

$$var(x) = \frac{1}{n}\sum (x_i - \bar{x})^2, \text{ или } var(x) = \overline{x^2} - (\bar{x})^2,$$

где $\overline{x^2} = \dfrac{\sum x_i^2}{n}$.

Значения $\bar{x}$, $var(x)$ являются числовыми характеристиками выборочной совокупности.

Для разных выборок, взятых из одной и той же генеральной совокупности, выборочные средние и дисперсии будут различны, т. е. выборочные характеристики являются случайными величинами.

Пример 1.1. Вычислить выборочные характеристики по данным наблюдения:

№	1	2	3	4	5
X	2	6	10	14	18

▼ Исходные данные и расчетные показатели представлены в виде двух расчетных таблиц (два способа расчета $\bar{x}$, $var(x)$):

№	x	$(x-\bar{x})^2$
1	2	64
2	6	16
3	10	0
4	14	16
5	18	64
Итого	50	160
Среднее	10	32
	$\bar{x}$	var(x)
var(x) = 160/5 = 32		

№	x	x^2
1	2	4
2	6	36
3	10	100
4	14	196
5	18	324
Итого	50	660
Среднее	10	132
	$\bar{x}$	$\overline{x^2}$
var(x) = $\overline{x^2} - (\bar{x})^2$=132–100 = 32		

Для вычисления выборочной средней и выборочной дисперсии в Excel имеются функции:

$$\bar{x} = \text{СРЗНАЧ(массив\_}x), \; var(x) = \text{ДИСПР(массив\_}x).$$

II. Генеральная совокупность

Генеральной средней называется математическое ожидание случайной величины X, т. е.

$$\mu_x = M(X).$$

Она характеризует среднее значение случайной величины в генеральной совокупности.

Генеральной дисперсией называется математическое ожидание квадрата отклонения случайной величины X относительно ее средней, т. е.

$$\sigma_x^2 = D(X) = M(X - \mu_x)^2.$$

Она характеризует *меру рассеяния* случайной величины в генеральной совокупности относительно ее средней.

Стандартным отклонением случайной величины X называется корень квадратный из ее дисперсии, т. е.

$$\sigma_x = \sqrt{D(X)}.$$

Оно показывает, насколько в среднем отклоняется случайная величина в генеральной совокупности относительно ее средней.

Значения μ_x, σ_x^2 – это числовые характеристики генеральной совокупности (числа), которые в отличии от выборочных характеристик $\bar{x}$, $var(x)$ являются неслучайными величинами (истинными).

Две случайные величины называются *независимыми*, если закон распределения одной из них не меняется от того, какие возможные значения приняла другая величина.

Если случайные величины X, Y независимы, то

$$M(XY) = M(X)M(Y);$$

$$D(X+Y) = D(X) + D(Y).$$

Из условия, что выборочные наблюдения $X_1, X_2, \ldots, X_n$ независимы и имеют одинаковые распределения, вытекают следующие соотношения:

$$\mu_{\bar{x}} = M(\bar{X}) = \mu_x, \ \sigma_{\bar{x}}^2 = D(\bar{X}) = \sigma_x^2 / n, \sigma_{\bar{x}} = \sigma_x / \sqrt{n}.$$

Для задания непрерывной случайной величины обычно используется плотность распределения вероятностей $f(x) = F'(x)$.

Площадь S под кривой распределения $f(x)$ в интервале $(a;\ b)$ равна вероятности попадания случайной величине X в этот интервал:

$$S = \int_a^b f(x)dx = P(a < X < b), \text{ причем } \int_{-\infty}^{\infty} f(x)dx = 1.$$

Наблюдаемое распределение $f(x)$ сводят к одному из хорошо исследованных видов распределения (нормальное распределение или специальные распределения, получаемые из нормального).

В эконометрических исследованиях чаще всего используются нормальное распределение, распределение Стьюдента и распределение Фишера. На основе этих распределений построены статистические критерии, служащие для проверки гипотез.

Нормальное распределение случайной величины X характеризуется лишь двумя параметрами: средним значением μ и дисперсией σ^2.

Это обозначается как $X \sim N(\mu;\ \sigma^2)$.

График плотности нормального распределения $f(x)$ имеет колоколообразный вид, симметричный относительно центра распределения (рис. 1.1).

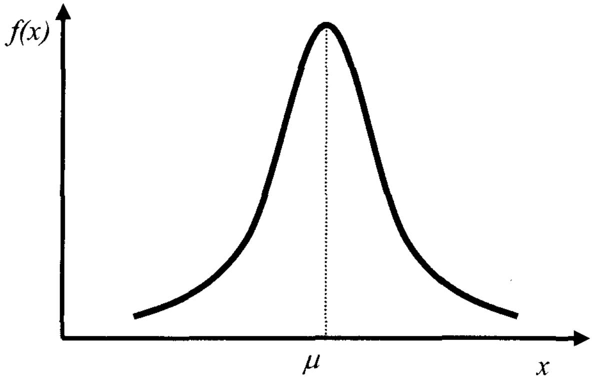

Рис. 1.1

Максимум этой функции находится в точке $x = \mu$, а разброс относительно этой точки определяется параметром σ^2. Чем меньше значение σ^2, тем более острый и высокий максимум $f(x)$.

Если от точки среднего значения (или точки максимума плотности нормального распределения) отложить вправо и влево три величины стандартного отклонения, то площадь под графиком плотности распределения, подсчитанная по этому промежутку, будет равна 99,73% всей площади под графиком (правило трех сигм).

Другими словами, 99,73% всех независимых наблюдений из нормального распределения лежат в радиусе трех стандартных отклонений от среднего значения.

Распределением Стьюдента (или *t*-распределением) называется распределение случайной величины

$$t_к = \frac{\xi_0}{\sqrt{\frac{1}{k}\sum_{i=1}^{k}\xi_i^2}},\ \xi_i \sim N(0;\ 1),\ i = 0,1,\ldots,k.$$

Распределение случайной по Стьюденту величины t характеризуется одним параметром k — числом степеней свободы.

Распределение Стьюдента используется для оценки среднего при неизвестной дисперсии выборки, t-распределение возникает в таблицах вывода регрессионного анализа, анализа временных рядов.

Распределением Фишера (или F-распределением) называется распределение случайной величины

$$F_{m,n} = \frac{\frac{1}{m}(\eta_1^2 + \eta_2^2 + \ldots + \eta_m^2)}{\frac{1}{n}(\xi_1^2 + \xi_2^2 + \ldots + \xi_n^2)}, \ \eta_1, \ldots, \eta_m, \xi_1, \ldots, \xi_n \sim N(0;\ 1).$$

Распределение Фишера имеет два параметра: m, n, которые принято называть числом степеней свободы.

Распределение Фишера возникает в регрессионном и дисперсионном анализах, а также в других видах многомерного анализа данных.

1.3. Статистические оценки и их свойства

Характеристики генеральной совокупности обычно неизвестны. Задача заключается в их оценке по характеристикам выборочной совокупности.

Характеристики генеральной совокупности принято называть *параметрами*, а выборочной совокупности — *оценками*.

Пусть искомый параметр генеральной совокупности есть θ_0, а на основе выборки объема n определяется оценка θ.

Для того чтобы выборочная оценка давала хорошее приближение оцениваемого параметра, она должна удовлетворять определенным требованиям (несмещенности, эффективности и состоятельности).

Несмещенность оценок. Оценка θ является *несмещенной*, если ее математическое ожидание равно оцениваемому параметру θ_0 при любом объеме выборки, т. е. $M(\theta) = \theta_0$. Если это не так, то оценка называется *смещенной*.

Требование несмещенности гарантирует отсутствие систематических ошибок при оценивании.

Выборочная средняя $\bar{x}$ является *несмещенной* оценкой генеральной средней μ, так как $M(\bar{x}) = \mu$.

Выборочная дисперсия $var(x)$ является *смещенной* оценкой генеральной дисперсии σ^2, т. е. $M[var(x)] \neq \sigma^2$.

В качестве *несмещенной оценки* генеральной дисперсии используется величина (исправленная дисперсия):

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}, \text{ для которой } M(S^2) = \sigma^2.$$

Величина S называется *стандартным отклонением* случайной величины X в выборке.

Для вычисления исправленной дисперсии и стандартного отклонения в Excel имеются функции:

S^2=ДИСП(массив_x), S = СТАНДОТКЛОН(массив_x).

Эффективность оценок. Несмещенная оценка θ называется *эффективной*, если она имеет минимальную дисперсию по сравнению с другими выборочными оценками, т. е. $min\sigma^2(\theta)$.

Предположим, что имеются две оценки параметра θ_0, рассчитанные на основе одной и той же информации (рис. 1.2). Оценка A является более эффективной, чем оценка B, так как имеет меньшую дисперсию, $\sigma_A^2(\theta) < \sigma_B^2(\theta)$.

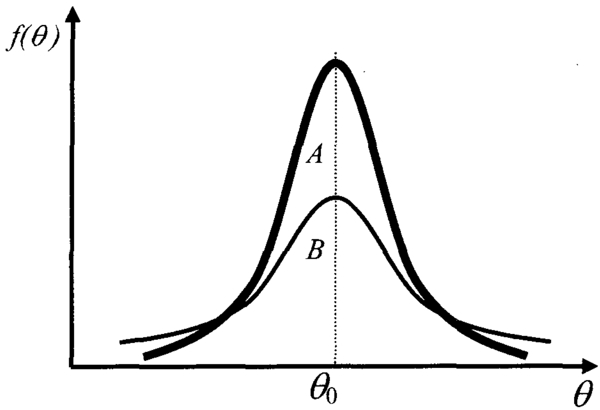

Рис. 1.2

Выборочная средняя $\bar{x}$ является *эффективной оценкой* генеральной средней μ, т. е. имеет наименьшую дисперсию в классе несмещенных оценок.

Покажем это для выборки из двух наблюдений X_1, X_2. Обобщенная оценка: $Z = \lambda_1 X_1 + \lambda_2 X_2$, где $\lambda_1 + \lambda_2 = 1$ является несмещенной, так как

$$M(Z) = M(\lambda_1 X_1 + \lambda_2 X_2) = \lambda_1 M(X_1) + \lambda_2 M(X_2) = (\lambda_1 + \lambda_2)\mu = \mu.$$

Теоретическая дисперсия обобщенной оценки равна:

$$D(Z) = D(\lambda_1 X_1 + \lambda_2 X_2) = \lambda_1^2 D(X_1) + \lambda_2^2 D(X_2) = (\lambda_1^2 + \lambda_2^2)\, \sigma_X^2.$$

Минимизация дисперсии $D(Z)$ эквивалентна минимизации выражения:

$$\lambda_1^2 + \lambda_2^2 = \lambda_1^2 + (1 - \lambda_1)^2 = 2\lambda_1^2 - 2\lambda_1 + 1 \to \min.$$

Минимум этого выражения достигается при $\lambda_1 = \lambda_2 = 1/2$, следовательно, выборочная средняя $(X_1 + X_2)/2$ имеет наименьшую дисперсию.

Этот вывод можно обобщить для выборок любого размера, если наблюдения независимы друг от друга.

Состоятельность оценок. Оценка θ называется *состоятельной*, если при $n \to \infty$ она стремится по вероятности к оцениваемому параметру θ_0, т. е.

$$\lim_{n\to\infty} P(|\theta_n - \theta_0| < \varepsilon) = 1.$$

Иначе говоря, состоятельной является такая оценка, которая дает точное значение для большой выборки.

На рис. 1.3 показано, как при различном объеме выборки может выглядеть распределение вероятностей (состоятельная оценка, смещенная на малой выборке).

Теорема Чебышева закона больших чисел утверждает, что

$$\lim_{n\to\infty} P(|\bar{x} - \mu| < \varepsilon) = 1.$$

т. е. выборочная средняя $\bar{x}$ является *состоятельной оценкой* генеральной средней μ.

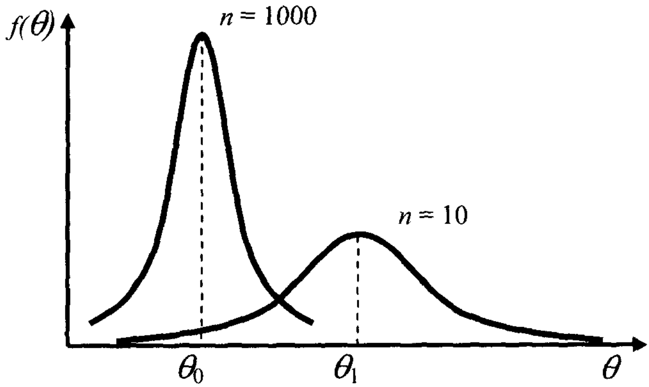

Рис. 1.3

1.4. Ковариация и корреляция

Различают выборочную и теоретическую ковариацию.

Выборочной ковариацией двух переменных x, y называется средняя величина произведения отклонений этих переменных от своих средних, т. е.

$$cov(x,y)=\frac{1}{n}\sum(x_i-\bar{x})(y_i-\bar{y}),\ \text{или}\ cov(x,y)=\overline{xy}-\bar{x}\,\bar{y},$$

где $\bar{x}$, $\bar{y}$ — выборочные средние переменных x, y;

$$\overline{xy}=\frac{\sum x_i y_i}{n}.$$

Заметим, что $cov(x,x)=\frac{1}{n}\sum(x_i-\bar{x})^2=var(x)$.

Правила расчета ковариации (a, b – константы)**:**

1) $cov(x,u+v)=cov(x,u)+cov(x,v)$;
2) $cov(x,a)=0$;
3) $cov(x,bu)=b\,cov(x,u)$;
4) $cov(u,v)=cov(v,u)$;
5) $var(u+v)=var(u)+var(v)+2cov(u,v)$.

Доказательство вытекает из определения ковариации. Например,

правило 2: $cov(x,a)=\frac{1}{n}\sum(x_i-\bar{x})(a-a)=0$;

правило 5: $var(u+v) = cov(u+v, u+v) = cov(u, u) + cov(v, v) + 2cov(u, v) = var(u) + var(v) + 2cov(u,v)$.

Теоретической ковариацией случайных величин X, Y называется математическое ожидание произведения отклонений этих величин от своих средних значений, т. е.

$$cov(X, Y) = M[(X-\mu_X)(Y-\mu_Y)].$$

Заметим, что $cov(X, X) = M(X-\mu_X)^2 = \sigma_x^2$.

Ковариация $cov(X, Y)$ есть *мера связи* двух случайных величин.

Если случайные величины X, Y независимы, то $cov(X, Y) = 0$, а если $cov(X, Y) \neq 0$, то X, Y — зависимы.

Более точной мерой зависимости между величинами является коэффициент корреляции.

Различают теоретический и выборочный коэффициенты корреляции.

Теоретический коэффициент корреляции определяется выражением:

$$\rho=\frac{cov(X,Y)}{\sigma_X\sigma_Y},$$

где σ_x, σ_y — стандартные отклонения случайных величин X, Y.

Коэффициент корреляции является безразмерной величиной, изменяющийся в пределах $-1 \leq \rho \leq 1$ и показывает *степень линейной связи* двух случайных величин. Коэффициент корреляции ρ является неслучайной величиной (истинной).

Случайные величины X, Y называются *некоррелированными*, если $\rho = 0$, и *коррелированными*, если $\rho \neq 0$.

Независимость случайных величин X, Y означает отсутствие любой связи (линейной и нелинейной), а *некоррелированность* – отсутствие только линейной связи.

Если случайные величины X, $У$ независимы, то они некоррелированы ($\rho = 0$), но из некоррелированности не следует их независимость, т. е. равенство $\rho = 0$ указывает лишь на отсутствие линейной связи между переменными, но не на отсутствие связи между ними вообще.

Выборочный коэффициент корреляции определяется выражением:

$$r = \frac{cov(x,y)}{\sqrt{var(x)var(y)}}.$$

При каждом конкретном значении ρ выборочный коэффициент корреляции является случайной величиной, изменяющийся в пределах $-1 \le r \le 1$.

На рис. 1.4 отражен геометрический смысл коэффициента корреляции.

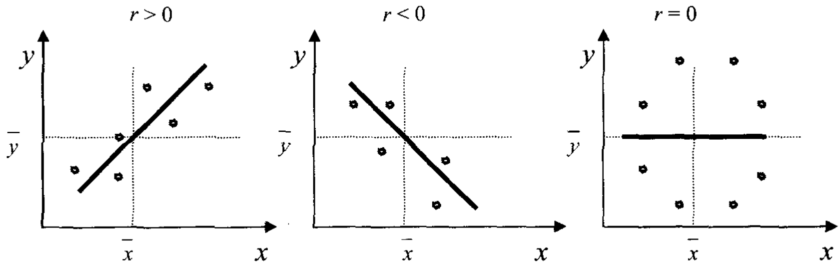

Рис. 1.4

Если $\rho = 0$ для генеральной совокупности, то это необязательно означает, что $r = 0$ для выборочной совокупности.

Пример 1.2. По исходным данным вычислить ковариацию и коэффициент корреляции между переменными x, y.

▼Представим исходные данные (x, y) и расчетные показатели в виде расчетной таблицы.

№	x	y	x^2	xy	y^2
1	2	1	4	2	1
2	6	2	36	12	4

№	x	y	x^2	xy	y^2
3	10	4	100	40	16
4	14	11	196	154	121
5	18	12	324	216	144
Итого	50	30	660	424	286
Среднее	10	6	132	84,8	57,2
	$\bar{x}$	$\bar{y}$	$\overline{x^2}$	$\overline{xy}$	$\overline{y^2}$

Окончательно имеем:

$$var(x) = \overline{x^2} - (\bar{x})^2 = 132 - 100 = 32;$$
$$var(y) = \overline{y^2} - (\bar{y})^2 = 57{,}2 - 36 = 21{,}2;$$
$$cov(x,y) = \overline{xy} - \bar{x}\,\bar{y} = 84{,}8 - 60 = 24{,}8.$$

$$r = \frac{cov(x,y)}{\sqrt{var(x)var(y)}} = \frac{24{,}8}{\sqrt{21{,}2 \cdot 32}} = 0{,}952. \blacktriangle$$

Для вычисления коэффициента корреляции в Excel имеется функция

r = коррел(массив_x; массив_y).

1.5. Проверка статистических гипотез

Статистической гипотезой H называется предположение относительно параметра или вида распределения случайной величины.

Нулевой гипотезой H_0 называют выдвигаемую гипотезу. Обычно считают, что H_0 — гипотеза об *отсутствии* различий.

Конкурирующей гипотезой H_1 называют гипотезу, которая противоречит нулевой. Таким образом, H_1 — гипотеза о *значимости* различий.

Статистическим критерием называется случайная величина, которая служит для проверки нулевой гипотезы. В качестве статистического критерия выбирается такая случайная величина, точное или приближенное распределение которой известно.

Если в качестве критерия проверки нулевой гипотезы используется случайная величина, подчиненная распределению Стьюдента, то ее обозначают через t (t-статистика), а подчиненная распределению Фишера – через F (F-статистика).

Проверку статистической гипотезы выполняют на основании результатов выборки. Поскольку выборка производится случайным образом, то появляется возможность принятия ошибочного решения.

Уровнем значимости α называется вероятность того, что будет отвергнута правильная нулевая гипотеза, т. е. $P_{Ho}(\overline{H}_0) = \alpha$.

Уровень значимости α устанавливается заранее.

В экономических исследованиях проверку гипотез осуществляют при 5%-ном и 1%-ном уровнях значимости.

Выбор, например 5%-ного уровня значимости означает, что в пяти случаях применения критерия из ста верная гипотеза будет отвергнута.

Расчетным значением критерия называют значение критерия, вычисленное по данным выборки.

Для проверки гипотезы используют критические точки, вычисляемые по известному распределению критерия.

Из сравнения расчетного значения критерия с критическими точками делается вывод о принятии или отклонении нулевой гипотезы.

Правила проверки нулевой гипотезы

***t*-статистика**. Определяется расчетное значение критерия t_p. Критическая точка $t_{кр}$ при заданном уровне значимости α и числе степеней свободы ν определяется по таблице t-распределения Стьюдента для двусторонней критической области или с помощью функции Excel:

$$t_{кр} = \text{СТЬЮДРАСПОБР}(\alpha;\ \nu).$$

Из сравнения t_p и $t_{кр}$, получаем:

- если $|t_p| < t_{кр}$, то H_0 принимается;
- если $|t_p| > t_{кр}$, то H_0 отклоняется.

***F*-статистика.** *Определяется расчетное значение критерия* F_p. Критическая точка $F_{кр}$ при заданном уровне значимости α и числу степеней свободы ν_1, ν_2 определяется по таблице F-распределения Фишера для правосторонней критической области или с помощью функции Excel:

$$F_{кр} = \text{FРАСПОБР}(\alpha;\ \nu_1;\ \nu_2).$$

Из сравнения расчетного значения F_p с критическим $F_{кр}$ получаем:

- если $F_p < F_{кр}$, то H_0 принимается;
- если $F_p > F_{кр}$, то H_0 отклоняется.

В статистических пакетах вместо критических точек используются:

- ***Р*-значение** — это значение уровня значимости, соответствующее вычисленной t-статистике. Определяется с помощью функции Excel:

$$P\text{-значение} = \text{СТЬЮДРАСП}(t_p;\ \nu;\ 2);$$

- **Значимость *F*** — значение уровня значимости, соответствующее вычисленной F-статистике. Определяется с помощью функции Excel:

$$\text{Значимость } F = \text{FРАСП}(F_p;\ \nu_1;\ \nu_2).$$

Из их сравнения с заданным уровнем α получаем:

- если P-значение (или Значимость F) $< \alpha$, то H_0 отклоняется;
- если P-значение (или Значимость F) $> \alpha$, то H_0 принимается.

Проверка гипотезы о корреляции случайных величин

Пусть по данным выборки объема n получен выборочный коэффициент корреляции $r \neq 0$. Требуется проверить гипотезу о равенстве нулю истинного значения коэффициента корреляции.

Выдвигаются гипотезы:

$$\begin{cases} H_0 : \rho = 0; \\ H_1 : \rho \neq 0. \end{cases}$$

В качестве критерия проверки нулевой гипотезы принимается случайная величина

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}},$$

которая имеет t-распределение Стьюдента с $\nu = n - 2$ степенями свободы.

По установленным правилам производится процедура проверки нулевой гипотезы.

Пример 1.3. По выборке объема $n = 5$ получен выборочный коэффициент корреляции $r = 0{,}952$. На 5%-ном уровне установить статистическую значимость выборочного коэффициента кореляции.

▼Расчетное значение критерия:

$$t_p = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0{,}952\sqrt{5-2}}{\sqrt{1-0{,}952^2}} = 5{,}4.$$

По условию $\alpha = 0{,}05$. Число степени свободы $\nu = n - 2 = 3$.

Критическую точку находим с помощью функции $t_{кр}$ = = СТЬЮДРАСПОБР(0,05; 3) = 3,18.

Поскольку $|t_p| = 5{,}4 > t_{кр} = 3{,}18$, то H_0 отклоняется, т. е. $r = 0{,}952$ статистически значим при 5%-ном уровне. ▲

Проверка гипотезы о равенстве двух дисперсий нормальных генеральных совокупностей

Пусть генеральные совокупности X, $У$ распределены нормально.

По независимым выборкам с объемами, соответственно равными n_1, n_2, извлеченным из этих совокупностей, найдены исправленные выборочные дисперсии S_x^2, S_y^2. Требуется по исправленным дисперсиям при заданном уровне значимости проверить нулевую гипотезу, что генеральные дисперсии рассматриваемых совокупностей равны между собой:

$$H_0\colon\ M(S_x^2) = M(S_y^2).$$

В качестве критерия проверки нулевой гипотезы принимается отношение большей исправленной дисперсии к меньшей, т. е. случайная величина

$$F = S_б^2 / S_м^2 .$$

Величина F при условии справедливости нулевой гипотезы имеет распределение Фишера со степенями свободы

$$v_1 = n_1 - 1,\ v_2 = n_2 - 1,$$

где n_1 — объем выборки с большей исправленной дисперсией;

n_2 – объем выборки с меньшей исправленной дисперсией.

По установленным правилам производится процедура проверки нулевой гипотезы.

Пример 1.4. По выборочным данным о расходах сырья при производстве продукции по старой и новой технологиям получены выборочные исправленные дисперсии $S_1^2 = 2{,}192$, $n_1 = 13$ и $S_2^2 = 1{,}611$, $n_2 = 9$.

Можно ли при уровне значимости 0,05 считать статистически значимым различие между исправленными дисперсиями..

▼Расчетное значение критерия $F_p = S_1^2 / S_2^2 = 2{,}192/1{,}611 = 1{,}36$.

Число степеней свободы $v_1 = n_1 - 1 = 12$, $v_2 = n_2 - 1 = 8$. Критическая точка $F_{кр}$ = FРАСПОБР (0,05; 12; 8) = 3,28.

Поскольку $F_p = 1{,}36 < F_{кр} = 3{,}28$, то H_0 принимается, т. е. выборочные исправленные дисперсии не различаются. ▲

2. МОДЕЛЬ ПАРНОЙ РЕГРЕССИИ

В модели парной линейной регрессии зависимость между переменными в генеральной совокупности представляется в виде

$$Y = \alpha + \beta X + \varepsilon,$$

где X — неслучайная величина;

Y; ε — случайные величины.

Величина Y называется *объясняемой* (зависимой) переменной, или результативным признаком, а X — *объясняющей* (независимой) переменной, или факторным признаком. Постоянные α, β — параметры уравнения.

Наличие случайного члена ε (ошибки регрессии) связано с воздействием на зависимую переменную других неучтенных в уравнении факторов.

На основе выборочного наблюдения оценивается выборочное уравнение регрессии (линия регрессии):

$$\hat{y} = a + bx,$$

где a, b — оценки параметров α, β.

2.1. Метод наименьших квадратов

Рассмотрим задачу «наилучшей» аппроксимации набора наблюдений (x_1, y_1), (x_2, y_2),... линейным уравнением (линией регрессии).

На рис. 2.1 приведены диаграмма рассеяния наблюдений и линия регрессии.

Расчетные значения $\hat{y}_i$ зависимой переменной в каждом наблюдении лежат на линии регрессии (прямой) и не совпадают в целом с соответствующими наблюдаемыми значениями y_i.

Определим остаток e_i в i-м наблюдении как разность между наблюдаемым и расчетным значениями зависимой переменной, т. е.

$$e_i = y_i - \hat{y}_i.$$

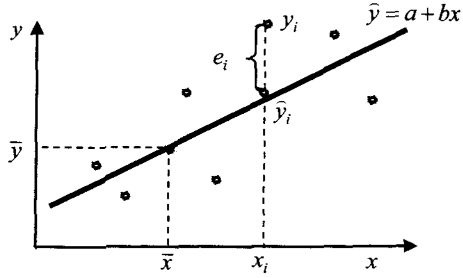

Рис. 2.1

Неизвестные значения a, b определяются **методом наименьших квадратов** (МНК).

Сущность МНК заключается в минимизации суммы квадратов остатков:

$$Q = \sum e_i^2 = \sum (y_i - \hat{y}_i)^2 = \sum (y_i - a - bx_i)^2 \to min,$$

где x_i, y_i — известные значения (числа);

a, b — неизвестные.

Запишем необходимые условия экстремума:

$$\begin{cases} Q_a^I = -2\sum (y_i - a - bx_i) = 0 \\ Q_b^I = -2\sum (y_i - a - bx_i)x_i = 0. \end{cases}$$

После преобразования получим следующую систему нормальных уравнений:

$$\begin{cases} a + b\bar{x} = \bar{y} \\ a\bar{x} + b\overline{x^2} = \overline{xy}, \end{cases}$$

решение которой:

$$\begin{cases} b = \dfrac{cov(x,y)}{var(x)} = \dfrac{\overline{xy} - \bar{x}\,\bar{y}}{\overline{x^2} - (\bar{x})^2} \\ a = \bar{y} - b\bar{x}. \end{cases}$$

Таким образом, *линия регрессии* (расчетное значение зависимой переменной):

$$\hat{y} = a + bx \text{, или } \hat{y} - \bar{y} = b(x - \bar{x}).$$

Линия регрессии проходит через точку $(\bar{x}, \bar{y})$ и выполняются равенства

$$\bar{y} = \bar{\hat{y}}, \bar{e} = 0.$$

Коэффициент b есть угловой коэффициент линии регрессии. Он показывает на сколько единиц в среднем изменяется переменная y при увеличении независимой переменной x на единицу.

Постоянная a дает прогнозируемое значение зависимой переменной при $x = 0$.

После построения уравнения регрессии, наблюдаемые значения y можно представить в виде

$$y_i = \hat{y}_i + e_i.$$

Остатки e_i, как и ошибки ε_i являются случайными величинами, однако они, в отличие от ошибок ε_i, являются наблюдаемыми.

Можно показать, что $cov(\hat{y}, e) = 0$.

Действительно, используя равенства

$$\hat{y} = a + bx, e = y - a - bx, cov(x, y) - b\,var(x) = 0,$$

получим

$$cov(\hat{y}, e) = cov(a + bx, y - a - bx) = b\,cov(x, y) - b^2\,var(x) =$$

$$= b\left[cov(x, y) - b\,var(x)\right] = 0.$$

2.2. Анализ вариации зависимой переменной

Определим *выборочные дисперсии* величин $y, \hat{y}, e$:

$var(y) = \dfrac{1}{n}\sum(y_i - \bar{y})^2$ — дисперсия *наблюдаемых* значений y;

$var(\hat{y}) = \frac{1}{n}\sum(\hat{y}_i - \bar{y})^2$ — дисперсия расчетных значений y;

$var(e) = \frac{1}{n}\sum(y_i - \hat{y}_i)^2$ — дисперсия остатков e.

Разброс наблюдаемых значений зависимой переменной характеризуется выборочной дисперсией $var(y)$.

Разложим дисперсию $var(y)$ на две части:

$$var(y) = var(\hat{y} + e) = var(\hat{y}) + var(e),$$

где $var(\hat{y})$ — часть, объясненная регрессионным уравнением;

$var(e)$ — необъясненная часть.

Коэффициентом детерминации R^2 называется отношение

$$R^2 = \frac{var(\hat{y})}{var(y)}, \quad 0 \le R^2 \le 1,$$

характеризующее долю вариации (разброса) зависимой переменной, объясненную уравнением регрессии.

Если $R^2 = 1$, то это означает точную подгонку, т. е. все точки наблюдения лежат на регрессионной прямой; если $R^2 = 0$, то регрессия ничего не дает, т. е. переменная x не улучшает качество предсказания у по сравнению с горизонтальной прямой $\hat{y} = \bar{y}$.

Чем ближе к единице R^2, тем лучше качество подгонки, т. е. $\hat{y}$ более точно аппроксимирует y.

Замечание. Вычисление R^2 корректно, если константа а включена в уравнение регрессии.

F-тест на качество оценивания

Для определения статистической значимости коэффициента детерминации R^2 проверяется гипотеза H_0: $F = 0$ для F-статистики, рассчитываемой по формуле

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

Величина F имеет распределение Фишера с $\nu_1 = 1$, $\nu_2 = n - 2$ степенями свободы.

По установленным правилам производится процедура проверки нулевой гипотезы.

Пример 2.1. Построить регрессионные зависимости: а) расходов на питание (y) и личных доходов (x); б) расходов на питание (y) и времени (t) по следующим данным (усл. ед.):

Год	2008	2009	2010	2011	2012
X	2	6	10	14	18
Y	1	2	4	11	12

и оценить качество подгонки.

▼а) Пусть истинная модель описывается выражением $y = \alpha + \beta x + \varepsilon$, а по выборочным наблюдениям определим оценки a, b.

Исходные данные и расчетные показатели представлены в виде расчетной таблицы:

Год	x	y	x^2	xy	$\hat{y}$	$(y-\bar{y})^2$	$(\hat{y}-\bar{y})^2$	$(y-\hat{y})^2$
2008	2	1	4	2	-0,2	25	38,44	1,44
2009	6	2	36	12	2,9	16	9,61	0,81
2010	10	4	100	40	6	4	0	4
2011	14	11	196	154	9,1	25	9,61	3,61
2012	18	12	324	216	12,2	36	38,44	0,04
Итого	50	30	660	424	30	106	96,1	9,9
Сред.	10	6	132	84,8	6	21,2	19,22	1,98
	$\bar{x}$	$\bar{y}$	$\overline{x^2}$	$\overline{xy}$	$\bar{\hat{y}}$	$var(y)$	$var(\hat{y})$	$var(e)$

Окончательно имеем:

$$var(x) = \overline{x^2} - (\bar{x})^2 = 132 - 100 = 32;$$

$$cov(x,y) = \overline{xy} - \bar{x}\cdot\bar{y} = 84{,}8 - 10\cdot 6 = 24{,}8;$$

$$b = \frac{cov(x,y)}{var(x)} = 24{,}8/32 = 0{,}775;$$

$$a = \bar{y} - b\bar{x} = 6 - 0{,}775\cdot 10 = -1{,}75.$$

Следовательно, $\hat{y} = -1{,}75 + 0{,}775x$ — линия регрессии.

Коэффициент b = 0,775 показывает, что при увеличении дохода на 1 усл. ед. расходы на питание увеличиваются в среднем на 0,775 усл. ед.

Качество подгонки оцениваем коэффициентом детерминации:

$R^2 = \dfrac{var(\hat{y})}{var(y)} = \dfrac{19{,}22}{21{,}2} = 0{,}907$, т. е. 90,7% разброса зависимой переменной (расходы на питание) объясняется регрессией.

Значимость коэффициента R^2 проверяем по F-тесту.

Расчетное значение критерия (F_p):

$$F_p = \frac{R^2(n-2)}{1-R^2} = \frac{0{,}907 \cdot 3}{0{,}093} = 29{,}12$$

Зададим α = 0,05. Число степени свободы $\nu_1 = 1$, $\nu_2 = n - 2 = 3$.

Критическое значение критерия определяем с помощью функции

$F_{кр}$ = FРАСПОБР(0,05; 1; 3) = 10,13.

Поскольку $F_p = 29{,}12 > F_{кр} = 10{,}13$, то H_0 – отклоняется, т. е. $R^2 = 0{,}907$ статистически значим.

Этот же вывод следует, если использовать

Значимость F = FРАСП(29,12; 1; 3) = 0,0124.

Поскольку Значимость $F = 0{,}0124 < \alpha = 0{,}05$, то R^2 значим при уровне 5%.

Полученные значения показателей можно представить в виде компактной таблицы дисперсионного анализа:

Дисперсионный анализ					
	df	***SS***	***MS***	***F***	**Значимость *F***
Регрессия	1	96,1	96,1	29,1212	**0,01247**
Остаток	3	9,9	3,3		
Итого	4	106			

В таблице *df* — число степеней свободы; *SS* — сумма квадратов отклонений; *MS* — дисперсии, рассчитываемые по формуле *MS= SS/ df*; *F* — расчетное значение F-критерия; Значимость *F* —

значение уровня значимости, соответствующее расчетному значению критерия.

б) Пусть истинная модель у = α + β t + ε (модель временного ряда), выборочная регрессия $\hat{y} = a + bt$, где t — время, определяемое как t = 1 для 2008 г., t = 2 для 2009 г. и т. д.

Исходные данные и расчетные показатели представим в виде расчетной таблицы:

Год	t	y	t^2	ty	$\hat{y}$
2008	1	1	1	1	-0,2
2009	2	2	4	4	2,9
2010	3	4	9	12	6
2011	4	11	16	44	9,1
2012	5	12	25	60	12,2
Итого	15	30	55	121	30
Сред.	3	6	11	24,2	6
	$\bar{t}$	$\bar{y}$	$\overline{t^2}$	$\overline{ty}$	$\bar{\hat{y}}$

Окончательно имеем:

$$var(t) = \overline{t^2} - (\bar{t})^2 = 11 - 9 = 2.$$
$$cov(t, y) = \overline{ty} - \bar{t}\bar{y} = 24{,}2 - 18 = 6{,}2.$$
$$b = \frac{cov(t, y)}{var(t)} = \frac{6{,}2}{2} = 3{,}1.$$
$$a = \bar{y} - b\bar{t} = 6 - 3{,}1 \cdot 3 = -3{,}3.$$

Следовательно, $\hat{y} = -3{,}3 + 3{,}1t$ — линия регрессии.

Коэффициент детерминации R^2 = 0,952, поскольку расчетные значения $\hat{y}$ в случаях а) и б) совпадают

Пример 2.2. Покажем, что в модели регрессии **без свободного члена** $Y = \beta X + \varepsilon$ оценка МНК для β:

$$b = \overline{xy} / \overline{x^2}.$$

▼Выборочная регрессия для этой модели есть $\hat{y} = bx$. Наблюдаемые значения зависимой переменной связаны с расчетным уравнением $y_i = \hat{y}_i + e_i$. Оценку b найдем из минимизации величины:

$$Q=\sum e_i^2=\sum (y_i-bx_i)^2=\sum y_i^2-2b\sum x_iy_i+b^2\sum x_i^2 .$$

Запишем необходимые условия экстремума:

$$Q'_b=-2\sum x_iy_i+2b\sum x_i^2=0, \text{ откуда } b=\frac{\sum x_iy_i}{\sum x_i^2}=\overline{xy}/\overline{x^2} .$$

Вычисление R^2 при отсутствии свободного члена некорректно; при этом не выполняется условие $var(y)=var(\hat{y})+var(e)$.

Пример 2.3. Покажем, что в модели регрессии $Y=\alpha+\varepsilon$ оценка МНК для α есть: $a=\bar{y}$.

Выборочная регрессия для заданной модели есть $\hat{y}_i=a$. Наблюдаемые значения зависимой переменной связаны с расчетными значениями уравнением $y_i=\hat{y}_i+e_i=a+e_i$. Оценку а найдем из минимизации величины

$$Q=\sum e_i^2=\sum (y_i-a)^2=\sum y_i^2-2a\sum y_i+na^2.$$

Получаем

$$Q'=-2\sum y_i+2an=0, \text{ откуда } a=\frac{\sum y_i}{n}=\bar{y}.$$

Выборочная регрессия $\hat{y}=\bar{y}$.

Пример 2.4. Покажем, что $\sqrt{R^2}=r_{\hat{y},y}$, где $r_{\hat{y},y}$ — коэффициент корреляции между $\hat{y}$ и y.

Действительно, учитывая соотношение

$$cov(\hat{y},y)=cov(\hat{y},\hat{y}+e)=cov(\hat{y},\hat{y})+cov(\hat{y},e)=var(\hat{y}),$$

получим

$$r_{\hat{y},y}=\frac{cov(\hat{y},y)}{\sqrt{var(\hat{y})var(y)}}=\sqrt{\frac{var(\hat{y})}{var(y)}}=\sqrt{R^2} .$$

Пример 2.5. Покажем, что в случае парной регрессии $\hat{y}=a+bx$ $\quad r_{\hat{y},y}=r_{x,y}$.

Действительно, из соотношений

$$cov(\hat{y}, y) = cov(a + bx, y) = b\,cov(x, y);$$
$$var(\hat{y}) = var(a + bx) = b^2\,var(x)$$

имеем

$$r_{\hat{y},y} = \frac{cov(\hat{y}, y)}{\sqrt{var(\hat{y})var(y)}} = \frac{cov(x, y)}{\sqrt{var(x)var(y)}} = r_{x,y}\,.$$

Вывод. В случае парной регрессии коэффициент детерминации есть квадрат коэффициента корреляции переменных x и y, т. е. $R^2 = r^2_{x,y}$

Пример 2.6. Зависимость переменной в регрессии $y = \alpha + \beta x + \varepsilon$ разбивается на две компоненты: $y = y_1 + y_2$.

Рассмотрим две регрессии для компонент:

$$y_1 = \alpha_1 + \beta_1 x + \varepsilon_1\,,\ y_2 = \alpha_2 + \beta_2 x + \varepsilon_2\,.$$

Докажем следующие соотношения для МНК-оценок параметров трех регрессий: $a = a_1 + a_2$, $b = b_1 + b_2$.

Действительно,

$$b = \frac{cov(x, y)}{var(x)} = \frac{cov(x, y_1 + y_2)}{var(x)} = \frac{cov(x, y_1) + cov(x, y_2)}{var(x)} = b_1 + b_2\,;$$

$$a = \bar{y} - b\bar{x} = (\overline{y_1 + y_2}) - \bar{x}(b_1 + b_2) = a_1 + a_2\,.$$

Пример 2.7. Покажем, что если все значения переменных изменить на одно и то же число или в одно и то же число раз, то величина коэффициента b в парной регрессии не изменится.

Пусть $x' = x + c$, $y' = y + c$, тогда

$$b' = \frac{cov(x', y')}{var(x')} = \frac{cov(x + c, y + c)}{var(x + c)} = \frac{cov(x, y)}{var(x)} = b\,.$$

Пусть $x' = kx$, $y' = ky$, тогда

$$b' = \frac{cov(x', y')}{var(x')} = \frac{cov(kx, ky)}{var(kx)} = \frac{k^2\,cov(x, y)}{k^2\,var(x)} = b\,.$$

Пример 2.8. Покажем, что $b = r\frac{s_y}{s_x}$, где r — коэффициент корреляции между x, y, а s_x, s_y — их стандартные отклонения.

Действительно,

$$s_x = \sqrt{\frac{n}{n-1}var(x)}\ ,\ s_y = \sqrt{\frac{n}{n-1}var(y)}\ ,\ r = \frac{cov(x,y)}{\sqrt{var(x)var(y)}},$$

тогда

$$r\frac{s_y}{s_x} = \frac{cov(x,y)}{\sqrt{var(x)var(y)}} \cdot \frac{\sqrt{\frac{n}{n-1}var(y)}}{\sqrt{\frac{n}{n-1}var(x)}} = \frac{cov(x,y)}{var(x)} = b\ .$$

Случайные составляющие коэффициентов регрессии

Величина Y в модели регрессии $Y = \alpha + \beta X + \varepsilon$ имеет две составляющие: неслучайную ($\alpha + \beta X$) и случайную ε.

Оценки коэффициентов регрессии (a; b) являются линейными функциями Y, и теоретически их также можно представить в виде двух составляющих.

Воспользовавшись разложением показателей:

$$cov(x,y) = cov(x,\ \alpha + \beta x + \varepsilon) = \beta var(x) + cov(x,\varepsilon);$$

$$\overline{y} = \frac{\sum y_i}{n} = \frac{\sum(\alpha + \beta x_i + \varepsilon_i)}{n} = \alpha + \beta\overline{x} + \frac{\sum \varepsilon_i}{n}.$$

получим преобразованные соотношения для (a; b):

$$\begin{cases} b = \beta + \dfrac{cov(x,\varepsilon)}{var(x)} \\ a = \alpha\ + \left[\dfrac{\sum \varepsilon_i}{n} - \dfrac{\overline{x}\,cov(x,\varepsilon)}{var(x)}\right]. \end{cases}$$

Таким образом, коэффициенты (a, b) разложены на две составляющие: *неслучайную*, равную истинным значениям (α; β) и *случайную*, зависящую от ε.

На практике нельзя разложить коэффициенты регрессии на составляющие, так как значения (α; β) или фактические значения ε в выборке неизвестны.

2.4. Предпосылки регрессионного анализа

Линейная регрессионная модель с двумя переменными имеет вид

$$y_i = \alpha + \beta x_i + \varepsilon_i, \ (i = 1,n),$$

где Y — объясняемая переменная;

X — объясняющая переменная;

ε — случайный член.

Для того чтобы регрессионный анализ, основанный на МНК, давал наилучшие из всех возможных результаты, должны выполняться определенные условия **(условия Гаусса–Маркова)**.

1. Математическое ожидание случайного члена в любом наблюдении должно быть равно нулю, т. е.

$$M(\varepsilon_i) = 0, \ (i = 1,n).$$

2. Дисперсия случайного члена должна быть постоянной для всех наблюдений, т. е.

$$D(\varepsilon_i) = \sigma^2, \ (i = 1,n).$$

3. Случайные члены должны быть статистически независимы (некоррелированы) между собой, т. е.

$$M(\varepsilon_i \varepsilon_j) = 0, \ (i \neq j).$$

4. Объясняющая переменная X есть величина неслучайная.

При выполнении условий Гаусса–Маркова модель называется *классической нормальной линейной регрессионной моделью*.

Наряду с условиями Гаусса–Маркова обычно предполагается, что случайный член распределен нормально, т. е. $\varepsilon_i \sim N(0; \sigma^2)$.

Рассмотрим подробнее условия и предположения, лежащие в основе регрессионного анализа.

Первое условие означает, что случайный член не должен иметь систематического смещения. Если постоянный член α включен в уравнение регрессии, то это условие автоматически выполняется.

Второе условие означает, что дисперсия случайного члена в каждом наблюдении имеет только одно значение σ^2.

Условие независимости дисперсии случайного члена от номера наблюдения называется *гомоскедастичностью* (одинаковый разброс).

Зависимость дисперсии случайного члена от номера наблюдения называется *гетероскедастичностью.*

Таким образом, имеем:

- $D(\varepsilon_i) = \sigma^2$, $(i = 1,n)$ — гомоскедастичность;
- $D(\varepsilon_i) = \sigma_i^2$, $(i = 1,n)$ — гетероскедастичность.

Характерные диаграммы рассеяния для случаев гомоскедастичности и гетероскедастичности показаны на рис. 2.2, *а* и *б* соответственно.

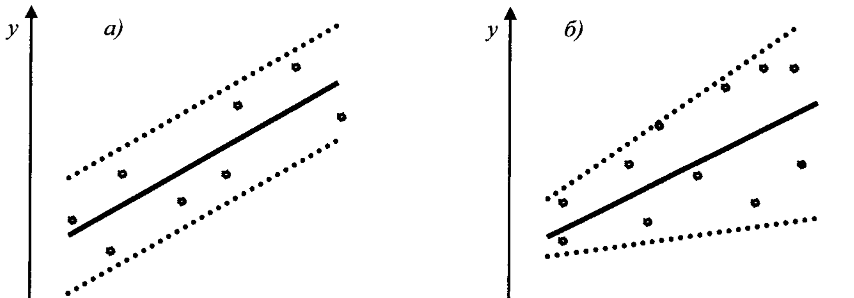

Рис. 2.2

Если условие гомоскедастичности не выполняется, то оценки коэффициентов регрессии будут *неэффективными*, хотя и *несмещенными.*

Существуют специальные методы диагностирования и устранения гетероскедастичности.

Третье условие указывает на некоррелированность случайных членов для каждых двух соседних наблюдений. Это условие часто нарушается, когда данные являются временными рядами. В случае когда третье условие не выполняется, говорят об *автокорреляции остатков*.

Типичный вид данных при наличии автокорреляции показан на рис. 2.3.

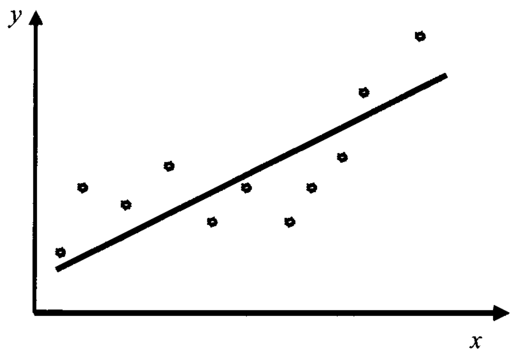

Рис. 2.3

Если условие независимости случайных членов не выполняется, то оценки коэффициентов регрессии, полученные по МНК, оказываются *неэффективными*, хотя и *несмещенными*.

Существуют методы диагностирования и устранения автокорреляции.

Четвертое условие о неслучайности объясняющей переменной является особенно важным. Если это условие нарушается, то оценки коэффициентов регрессии оказываются *смещенными* и *несостоятельными*.

Нарушение этого условие может быть связано с ошибками измерения объясняющих переменных или с использованием лаговых переменных.

Предположение о нормальности распределения случайного члена необходимо для проверки значимости параметров регрессии и для их интервального оценивания.

Теорема Гаусса-Маркова. Если предпосылки регрессионного анализа выполняются, то оценки (a, b), сделанные с помощью МНК, являются:

• *несмещенными*: $M(a) = \alpha$, $M(b) = \beta$, что означает отсутствие систематической ошибки в положении линии регрессии;

• *эффективными*: имеют наименьшую дисперсию в классе всех линейных несмещенных оценок, равную

$$D(a)=\frac{\sigma^2\,\overline{x^2}}{n\,var(x)},\quad D(b)=\frac{\sigma^2}{n\,var(x)};$$

• *состоятельными:* $\lim\limits_{n\to\infty} D(a)=0,\ \lim\limits_{n\to\infty} D(b)=0$,

что означает, что при достаточно большом n оценки (a, b) близки к (α, β).

Для проверки выводов теоремы воспользуемся оценками (a, b) в виде разложения:

$$\begin{cases} b=\beta+\dfrac{cov(x,\varepsilon)}{var(x)} \\ a=\alpha\ +\left[\dfrac{\sum\varepsilon_i}{n}-\dfrac{\bar{x}\,cov(x,\varepsilon)}{var(x)}\right] \end{cases}$$

и соотношением:

$$cov(x,\varepsilon)=\frac{1}{n}\sum(x_i-\bar{x})(\varepsilon_i-\bar{\varepsilon})=\frac{1}{n}(x_i-\bar{x})\varepsilon_i-\frac{\bar{\varepsilon}}{n}\sum(x_i-\bar{x})=\frac{1}{n}\sum(x_i-\bar{x})\varepsilon_i .$$

Пусть x неслучайная величина, тогда

$$M[cov(x,\varepsilon)]=M\left[\frac{1}{n}\sum(x_i-\bar{x})\varepsilon_i\right]=\frac{1}{n}\sum(x_i-\bar{x})M(\varepsilon_i)=0;$$

$$D[cov(x,\varepsilon)]=D\left[\frac{1}{n}\sum(x_i-\bar{x})\varepsilon_i\right]=\frac{1}{n^2}\sum(x_i-\bar{x})^2D(\varepsilon_i)=\frac{var(x)}{n}\sigma^2;$$

$$M\left[\frac{\sum\varepsilon_i}{n}\right]=\frac{\sum M(\varepsilon_i)}{n}=0,\ D\left[\frac{\sum\varepsilon_i}{n}\right]=\frac{\sum D(\varepsilon_i)}{n^2}=\frac{\sigma^2}{n};$$

$$M\left[\frac{\bar{x}\,cov(x,\varepsilon)}{var(x)}\right]=\frac{\bar{x}}{var(x)}M\left[cov(x,\varepsilon)\right]=0;$$

$$D\left[\frac{\bar{x}\,cov(x,\varepsilon)}{var(x)}\right]=\frac{(\bar{x})^2}{var^2(x)}D\left[cov(x,\varepsilon)\right]=\frac{(\bar{x})^2\sigma^2}{n\,var(x)}.$$

Вычислим математическое ожидание и дисперсию оценок b, a:

$$M(b)=\beta+M\left[\frac{cov(x,\varepsilon)}{var(x)}\right]=\beta+\frac{M\left[cov(x,\varepsilon)\right]}{var(x)}=\beta;$$

$$D(b)=D\left[\frac{cov(x,\varepsilon)}{var(x)}\right]=\frac{D\left[cov(x,\varepsilon)\right]}{var^2(x)}=\frac{\sigma^2}{n\,var(x)};$$

$$M(a)=\alpha+M\left[\frac{\sum\varepsilon_i}{n}\right]-M\left[\frac{\bar{x}\,cov(x,\varepsilon)}{var(x)}\right]=\alpha;$$

$$D(a)=D\left[\frac{\sum\varepsilon_i}{n}\right]+D\left[\frac{\bar{x}\,cov(x,\varepsilon)}{var(x)}\right]=\frac{\sigma^2}{n}\left[1+\frac{(\bar{x})^2}{var(x)}\right]=\frac{\sigma^2\,\overline{x^2}}{n\,var(x)}.$$

Расчет стандартной ошибки коэффициентов регрессии

Полученные теоретические дисперсии $D(a)$, $D(b)$ зависят от дисперсии σ^2 случайного члена. По данным выборки отклонения ε_i, а следовательно, и их дисперсии σ^2 неизвестны, поэтому они заменяются наблюдаемыми остатками e_i и их выборочной дисперсией $var(e)$. Но оценка $var(e)$ является *смещенной,* т. е. $M[var(e)] \neq \sigma^2$.

Несмещенной оценкой дисперсии σ^2 является величина (остаточная дисперсия):

$$S^2=\sum e_i^2/(n-2),$$

которая служит *мерой разброса* зависимой переменной вокруг линии регрессии.

Отметим, что в знаменателе остаточной дисперсии стоит число степеней свободы $(n-2)$, а не n, так как две степени свободы теряются при определении двух параметров $(a;\ b)$.

Величина S называется *стандартной ошибкой регрессии.*

Заменив в теоретических дисперсиях неизвестную σ^2 на оценку S^2, получим оценки дисперсий:

$$S_a^2 = \frac{\overline{x^2}\, S^2}{n\, var(x)} \quad , \quad S_b^2 = \frac{S^2}{n\, var(x)}$$

Величины S_a, S_b называется *стандартными ошибками* коэффициентов регрессии.

Статистические свойства МНК-оценок (*a*; *b*)

Пусть выполняется условие нормальности распределения случайного члена: $\varepsilon_i \sim N(0;\, \sigma^2)$. Тогда МНК-оценки коэффициентов регрессии также имеют нормальное распределение, поскольку являются линейными функциями от ε_i, т. е.

$$a \sim N\left(\alpha\, ; \frac{\sigma^2 \overline{x^2}}{n\, var(x)}\right); \qquad b \sim N\left(\beta ; \frac{\sigma^2}{n\, var(x)}\right).$$

Если условия нормальности распределения случайного члена не выполняются, то оценки (a; b) имеют асимптотически нормальное распределение.

Проверка гипотезы H_0: $\beta = \beta_0$

Пусть в теоретической зависимости $Y = \alpha + \beta X + \varepsilon$ случайный член ε распределен нормально с неизвестной дисперсией σ^2.

Величина β, хотя и неизвестна, но имеется основание предполагать, что она равна заданной величине β_0.

Выдвигаются гипотезы:

$$\begin{cases} H_0 : \beta = \beta_0 \\ H_1 : \beta \neq \beta_0. \end{cases}$$

Задача заключается в проверке нулевой гипотезы на основании выборочных данных.

Пусть по выборочным данным получена оценка b.

В качестве критерия проверки нулевой гипотезы принимают случайную величину

$$t = \frac{b - \beta_0}{S_b},$$

которая имеет распределение Стьюдента с $\nu = n - 2$ степенями свободы.

По таблице критических точек распределения Стьюдента по заданному уровню значимости α и числу ν степеней свободы находят критическую точку $t_{кр}$. Сравнивая наблюдаемое значение критерия с критическим, можно принять или отвергнуть нулевую гипотезу.

Результаты оценивания регрессии совместимы не только с конкретной гипотезой H_0: $\beta = \beta_0$, но и с некоторым их множеством. Любое значение β, совместимое с оценкой b, удовлетворяет условию:

$$\left|\frac{b-\beta}{S_b}\right| < t_{kp}\text{, или } -t_{kp} < \frac{\beta - b}{S_b} < t_{kp}.$$

Разрешив это неравенство относительно β, получим:

$$b - t_{кр} S_b < \beta < b + t_{кр} S_b,$$

т. е. *доверительный интервал* для величины β.

Посредине интервала лежит величина b. Границы интервала одинаково отстоят от b, зависят от выбора уровня значимости и являются случайными числами.

Доверительный интервал покрывает значение параметра β с заданной вероятностью $(1 - \alpha)$, т. е. $P(b - t_{кр} S < \beta < b + t_{кр} S_b) = 1 - \alpha$.

Проверка гипотезы H_0: $\beta = 0$

Пусть по выборке получена оценка коэффициента регрессии b.

Для определения статистической значимости коэффициента регрессии b проверяется гипотеза H_0: $\beta = 0$ для t-статистики, рассчитываемой по формуле

$$t = b/S_b.$$

Величина t имеет распределение Стьюдента с $\nu = n - 2$ степенями свободы.

По установленным правилам производится процедура проверки нулевой гипотезы.

Пример 2.9. В примере 2.1 получена модель $\hat{y} = -1{,}75 + 0{,}775x$ зависимости расходов y на питание от личного дохода x. Рассчитаем стандартную ошибку, значимость и доверительный интервал коэффициентов модели.

▼Воспользуемся результатами примера 2.1.

$n = 5$; $\sum(y-\hat{y})^2 = 9{,}9$; $\overline{x^2} = 132$; $var(x) = 32$

$a = -1{,}75$; $b = 0{,}775$; $\alpha = 0{,}05$.

Стандартная ошибка регрессии

$$S = \sqrt{\frac{\sum(y-\hat{y})^2}{n-2}} = \sqrt{\frac{9{,}9}{3}} = 1{,}817\,.$$

Стандартная ошибка коэффициента b:

$$S_b = \frac{S}{\sqrt{n\,var(x)}} = \frac{1{,}817}{\sqrt{5\cdot 32}} = 0{,}144\,.$$

Расчетное значение критерия $t_p = b/S_b = 0{,}775/0{,}144 = 5{,}4$.

Число степени свободы $\nu = n - 2 = 3$.

Критическое значение $t_{кр}$= СТЬЮДРАСПОБР(0,05; 3) = 3,18.

Поскольку $t_p = 5{,}4 > t_{кр} = 3{,}18$, то $b = 0{,}775$ значим на уровне 5%.

Этот же вывод следует, если использовать

P-значение = СТЬЮДРАСП(5,4; 3; 2) = 0,0124.

Поскольку P-значение = 0,0124 < α = 0,05, то $b = 0{,}775$ значим.

Доверительный интервал для b есть $(b - t_{кр}\,S_b\,;\, b + t_{кр}\,S_b) =$
$= (0{,}775 - 3{,}18\cdot 0{,}144;\ 0{,}775 + 3{,}18\cdot 0{,}144) = (0{,}32;\ 1{,}23)$.

Проведем подобные расчеты для коэффициента $a = -1{,}75$.

$$S_a = S\cdot\sqrt{\frac{\overline{x^2}}{n\,var(x)}} = 1{,}816\cdot\sqrt{\frac{132}{5\cdot 32}} = 1{,}65\,;\ t_p = a/S_a = -1{,}75/1{,}65 =$$
$= -1{,}06$.

P-значение = СТЬЮДРАСП(1,06; 3; 2) = 0,367.

Поскольку P-значение = 0,367 > α = 0,05, то $a = -1{,}75$ незначим.

Доверительный интервал: $(a - t_{кр} S_a;\ a + t_{кр} S_a) = (-7{,}00;\ 3{,}50)$.

Полученные значения коэффициентов регрессии и их статистические оценки можно представить в виде компактной таблицы:

	Коэффициенты	**Стандартная ошибка**	***t*-статистика**	***P*-значение**	**Нижние 95%**	**Верхние 95%**
Y-пересеч.	–1,75	1,65	–1,06	0,36669	–7,001	3,501
X	0,775	0,144	5,40	0,01247	0,318	1,232

В таблице столбцы имеют следующую интерпретацию:

- Коэффициенты — значения коэффициентов регрессии;
- Стандартная ошибка — стандартные ошибки коэффициентов регрессии;
- t-статистика — расчетные значения t-критерия;
- P-значение — значения уровней значимости, соответствующие расчетным значениям t-критерия;
- Нижние 95% и Верхние 95% — соответственно нижние и верхние границы доверительных интервалов для коэффициентов регрессии.

Пример 2.10. Покажем, что в выборочной регрессии *без свободного члена* $\hat{y} = b\,x$ стандартная ошибка оценки b:

$$S_b = \frac{S}{\sqrt{n\,\overline{x^2}}},\ \text{где}\ S^2 = \frac{\sum (y_i - b\,x_i)^2}{n-1}.$$

▼Подставим в оценку для b выражение $y_i = \beta\,x_i + \varepsilon_i$, получим:

$$b = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{\sum x_i(\beta\,x_i + \varepsilon_i)}{\sum x_i^2} = \beta + \frac{\sum x_i \varepsilon_i}{\sum x_i^2}.$$

Оценка b является *несмещенной*, так как $M(b) = \beta$.

Дисперсия оценки b:

$$D(b) = \frac{\sum x_i^2 D(\varepsilon_i)}{(\sum x_i^2)^2} = \frac{\sigma^2}{\sum x_i^2} = \frac{\sigma^2}{n\overline{x^2}}.$$

В исходной модели оценивается один параметр, поэтому оценкой σ^2 является:

$$S^2 = \frac{\sum e_i^2}{n-1} = \frac{\sum(y_i - bx_i)^2}{n-1}.$$

Следовательно, $S_b = \dfrac{S}{\sqrt{n\overline{x^2}}}$ ▲

Пример 2.11. Покажем, что в выборочной регрессии $\hat{y} = a$ стандартная ошибка оценки a есть

$$S_a = \frac{S}{\sqrt{n}}, \text{ где } S^2 = \frac{\sum(y_i - \bar{y})^2}{n-1}.$$

▼Подставим в оценку для a выражение $y_i = \alpha + \varepsilon_i$, получим:

$$a = \frac{\sum y_i}{n} = \frac{\sum(\alpha + \varepsilon_i)}{n} = \alpha + \frac{\sum \varepsilon_i}{n}.$$

Оценка a является несмещенной, так как $M(a) = \alpha$.

Дисперсия оценки a есть:

$$D(a) = \frac{\sum D(\varepsilon_i)}{n^2} = \frac{\sigma^2}{n}.$$

В исходной модели оценивается один параметр, поэтому оценкой σ^2 является:

$$S^2 = \frac{\sum e_i^2}{n-1} = \frac{\sum(y_i - \bar{y})^2}{n-1}.$$

Следовательно, $S_a = \dfrac{S}{\sqrt{n}}$ ▲

Упражнение 2.1. По данным примера 2.1 постройте зависимость расходов на питание y от личного дохода x для модели регрессии *без свободного члена* и рассчитайте стандартную ошибку коэффициента регрессии.

Ответ: $\hat{y} = 0{,}642x$, $S_b = 0{,}061$.

Взаимозависимость критериев

В парном регрессионном анализе эквивалентны критерии:

$$t_b = \frac{b}{S_b},\quad t_r = r\sqrt{\frac{n-2}{1-r^2}},\quad F = \frac{R^2(n-2)}{1-R^2}.$$

Связь между критериями выражается равенством:

$$t_b = t_r = \sqrt{F},$$

причем, для критических значений критериев при любом уровне значимости имеем:

$$(t_b)_{кр} = (t_r)_{кр} = \sqrt{F_{кр}},$$

и эти критерии дают один и тот же результат.

Вывод. Проверки значимости коэффициента b в парной линейной регрессии, коэффициента корреляции r и коэффициента детерминации R^2 эквивалентны.

2.5. Пакет анализа Excel (программа «Регрессия»)

Расчет параметров уравнения линейной регрессии, проверку их статистической значимости и построение интервальных оценок можно выполнить значительно быстрее автоматически при использовании Пакета анализа Excel (программа «Регрессия»).

Пусть исходные данные примера 2.1 (расходы на питание — личный доход) представлены в Excel.

Выбираем команду **Анализ данных ⇒ Регрессия**.

В диалоговом окне режима **Регрессия** задаются следующие параметры:

• **Входной интервал У** — вводится ссылка на ячейки, содержащие данные по результативному признаку.

• **Входной интервал X** — вводится ссылка на ячейки, содержащие факторные признаки.

• **Метки** — установите флажок в активное состояние, если выделены и заголовки столбцов.

• **Константа-ноль** — установите флажок в активное состояние, если оцениваете регрессионное уравнение без свободного члена.

При необходимости задаются и другие параметры.

Результаты расчетов с использованием инструмента **Регрессия** выводятся под общим названием **Вывод итогов** в виде следующих таблиц:

Регрессионная статистика	
Множественный R	0,952
R-квадрат	**0,907**
Нормированный R-квадрат	0,875
Стандартная ошибка	1,817
Наблюдения	5

Дисперсионный анализ					
	df	***SS***	***MS***	***F***	**Значимость *F***
Регрессия	1	96,1	96,1	29,12	**0,01247**
Остаток	3	9,9	3,3		
Итого	4	106			

	Коэффициенты	**Стандартная ошибка**	***t*-статистика**	***P*-значение**	**Нижние 95%**	**Верхние 95%**
Y-пересеч.	**-1,75**	1,65	-1,06	0,36669	-7,001	3,501
X	**0,775**	0,14361	5,40	**0,01247**	0,318	1,232

Результаты работы программы «Регрессия» полностью совпадают с полученными ранее расчетами.

При необходимости выводятся предсказанные значения $\hat{y}_t$ результативного признака и значения остатков.

ВЫВОД ОСТАТКА		
Наблюдение	**Предсказанное y**	**Остатки**
1	–0,2	1,2
2	2,9	–0,9
3	6	–2
4	9,1	1,9
5	12,2	–0,2

Коэффициенты регрессии, их стандартные ошибки и коэффициент детерминации составляют:

$$a = -1{,}75;\ b = 0{,}775;\ S_a = 1{,}65;\ S_b = 0{,}143;\ R^2 = 0{,}907.$$

Результаты регрессионного анализа принято записывать в виде:

$$\underset{(1{,}65)\,(0{,}143)}{\hat{y} = -1{,}75 + 0{,}775x};\quad R^2 = 0{,}907,$$

где в скобках указаны стандартные ошибки коэффициентов регрессии.

Статистическая значимость коэффициента $R^2 = 0{,}907$ устанавливается по F-тесту. Поскольку Значимость $F = 0{,}0124 < \alpha = 0{,}05$, то $R^2 = 0{,}907$ значим при уровне 5%. Модель в целом значима.

Обычно проверка значимости коэффициента a не производится. Оценим статистическую значимость коэффициента b.

Поскольку P-значение $= 0{,}0124 < 0{,}05$, то коэффициент $b = 0{,}775$ значим на уровне 5%.

Результаты оценивания регрессии совместимы не только с полученным значением коэффициента регрессии $b = 0{,}775$, но и с некоторым его множеством (доверительным интервалом). С вероятностью 95% доверительный интервал коэффициента b есть (0,318 … 1,232).

Пример 2.12. Имеются данные (усл. ед.) о расходах на питание y и душевого дохода x для девяти групп семей:

x	63	158	260	370	480	593	728	935	1880
y	43	62	90	111	130	149	165	191	241

Используя результаты работы программы «Регрессия», проанализируйте зависимость расходов на питание от величины душевого дохода.

▼ Результаты регрессионного анализа записываем в виде:

$$\underset{(11{,}72)}{\hat{y}=66{,}04}+\underset{(0{,}015)}{0{,}107x}, \quad R^2=0{,}885,$$

где в скобках указаны стандартные ошибки коэффициентов регрессии.

Качество модели оценивается коэффициентом детерминации R^2.

Величина R^2 = 0,885 означает, что фактором душевого дохода можно объяснить 88,5% вариации (разброса) расходов на питание.

Установим статистическую значимость коэффициента R^2.

Поскольку Значимость F = 0, 000158 < α = 0,05, то R^2= 0,885 значим при уровне 5%.

Направление связи между у и х определяет знак коэффициента b = 0,107 > 0, т. е. связь является прямой (положительной).

Коэффициент b = 0,107 показывает, что при увеличении душевого дохода на 1 усл. ед. расходы на питание в среднем увеличиваются на 0,107 усл. ед.

Оценим статистическую значимость коэффициента b. Поскольку P-значение = 0,000158 < α =0,05, то коэффициент b = 0,107 значим на 5%-ном уровне. С вероятностью 95% доверительный интервал коэффициента b есть (0,073; 0,142). ▲

2.6. Нелинейные регрессии

Нелинейность регрессии проявляется как по переменным, так и по параметрам.

Нелинейность по переменным устраняется путем замены переменной. Например, нелинейное уравнение $y=\alpha+\beta\sqrt{x}+\varepsilon$ после замены переменной $z=\sqrt{x}$ становится линейным: $y = \alpha + \beta z + \varepsilon$, и для оценки ее параметров используется МНК.

Пример 2.13. Имеются данные между ежегодным потреблением бананов (y) и годовым доходом (x) 10 американских семей (усл. ед.):

x	1	2	3	4	5	6	7	8	9	10
y	2	7	9	12	10	12	11	12	13	12

Рассмотреть следующие варианты уравнения регрессии:

$$1)\ y = \alpha + \beta x + \varepsilon;$$
$$2)\ y = \alpha + \beta\sqrt{x} + \varepsilon;$$
$$3)\ y = \alpha + \beta / x + \varepsilon$$

и выбрать лучший.

▼Исходные и преобразованные данные для построения искомых уравнений регрессии представлены в таблице.

y	x	$\sqrt{x}$	$1/x$
2	1	1	1
7	2	1,414	0,5
9	3	1,732	0,333
12	4	2	0,25
10	5	2,236	0,2
12	6	2,449	0,166
11	7	2,645	0,142
12	8	2,828	0,125
13	9	3	0,111
12	10	3,162	0,1

1. Оценка первого уравнения по исходным данным (y, x):

$$\hat{y} = 5{,}13 + 0{,}88x,\ R^2 = 0{,}645,\ S = 2{,}10\,.$$

2. Оценка второго уравнения по преобразованным данным $(y, \sqrt{x})$:

$$\hat{y} = 0{,}774 + 4{,}106\sqrt{x},\ R^2 = 0{,}762,\ S = 1{,}72\,.$$

3. Оценка третьего уравнения по преобразованным данным $(y,\ 1/x)$:

$$\hat{y} = 13{,}42 - 11{,}67/x,\ R^2 = 0{,}942,\ S = 0{,}85\,.$$

Качество оценивания последнего варианта уравнения выше других. ▲

Нелинейность по параметру часто устраняется путем логарифмического преобразования уравнения. Например, следующие нелинейные уравнения после логарифмирования сводятся к линейным:

- степенная функция $y = \alpha x^{\beta}\varepsilon \sim ln y = ln\alpha + \beta ln x + ln\varepsilon$;
- экспоненциальная функция $y = \alpha e^{\beta x}\varepsilon \sim ln y = ln\alpha + \beta x + ln\varepsilon$.

Использование МНК для нахождения оценок параметров этих уравнений требует, чтобы $ln\varepsilon$ имел нормальное распределение.

Однако уравнение $y = \alpha x^{\beta} + \varepsilon$, в котором случайный член ε является аддитивным, уже никакими преобразованиями не приводится к линейному.

В экономике функции вида $y = \alpha\, x^{\beta}\varepsilon$ применяются при моделировании кривых спроса, а вида $y = \alpha\ e^{\beta x}\ \varepsilon$ — при моделировании временных трендов, при этом вместо x используется время t, а вместо β — постоянный темп прироста r, т. е. $y = \alpha\, e^{rt}\ \varepsilon$.

В экономическом анализе часто используется эластичность функции.

Эластичность функции $\boldsymbol{y = f(x)}$ рассчитывается как относительное изменение y к относительному изменению x, т. е.:

$$Э = (\frac{dy}{y})/(\frac{dx}{x}) = \frac{x}{y} f'(x) .$$

Эластичность функции показывает, на сколько процентов изменяется функция $y = f(x)$ при изменении независимой переменной на 1%.

Для степенной функции $y = ax^{b}$ эластичность представляет собой постоянную величину, равную b.

В силу того, что эластичность линейной функции $y = a + bx$ не является постоянной величиной, а зависит от x, т. е.

$$Э = b\frac{x}{y},$$

обычно вычисляется средний показатель эластичности по формуле

$$\overline{Э} = b\frac{\overline{x}}{\overline{y}},$$

где $\overline{x}, \overline{y}$ — средние значения переменных x, y в выборке.

Например, для зависимости расходов на питание от доходов $\hat{y} = -1{,}75 + 0{,}775x$,$(\overline{x} = 10, \overline{y} = 6)$ средний показатель эластичности есть 1,29 и показывает, что с увеличением дохода на 1% расходы на питание возрастут в среднем на 1,29%.

Пример 2.14. По данным примера 2.1 построить зависимость расходов на питание (y) от доходов (x) в виде: 1) степенной функции; 2) экспоненциального временного тренда.

▼Исходные и преобразованные данные для построения указанных уравнений представлены в таблице.

t	*x*	*y*	*lnx*	*lny*
1	2	1	0,69314	0
2	6	2	1,79175	0,69314
3	10	4	2,30258	1,38629
4	14	11	2,63905	2,39789
5	18	12	2,89037	2,48490

1. Оценив регрессию между (lny, lnx), получим преобразованное уравнение: $ln\,\hat{y} = -1{,}049 + 1{,}183\,ln\,x$.

Выполнив обратные преобразования, получим: $\hat{y} = 0{,}350x^{1{,}183}$.

Эластичность спроса на продукты питания по доходу составляет 1,183.

Это означает, что увеличение дохода на 1% приведет к увеличению расходов на питание на 1,183%.

2. Оценив регрессию между (lny, t), получим преобразованное уравнение $ln\,\hat{y} = -0{,}61 + 0{,}667 \cdot t$. Выполнив обратные преобразования, получим $\hat{y} = 0{,}543e^{0{,}667t}$.

Постоянный темп роста $r = 0{,}667$. Это означает, что расходы на продукты питания в течение выборочного периода росли с темпом 66,7% в год. ▲

Упражнение 2.2. Директор по маркетингу определил общий объем продаж (y) товара в зависимости от цены (x) (усл. ед.):

Цена x	**Спрос** y
10	700
20	350
30	300
40	200
50	110
60	50

Установите степенную зависимость между ценой и спросом, предскажите величину спроса для цены 45 усл. ед. и насколько уменьшится спрос при росте цены на 1%.

Ответ. Функция спроса: $y = 17662x^{-1,307}$. Прогнозное значение спроса на товар при цене 45 усл. ед. равно 122,07 усл. ед. При увеличении цены товара на 1% спрос уменьшится на 1,31%.

Упражнение 2.3. По данным (усл. ед.) об объеме продаж компании за 10 лет

Год t	**Объем продаж** y
1	70
2	183
3	340
4	649
5	1243
6	1979
7	4096
8	6440
9	8459
10	12154

постройте модель экспоненциального роста и сделайте прогноз объема продаж на два года вперед.

Ответ. Модель: $y = 58{,}55e^{0{,}569t}$. Прогноз $y(11) = 30729$; $y(12) =$ $= 54303$.

2.7. Прогнозирование в регрессионных моделях

Под прогнозированием в эконометрике понимается построение оценки зависимой переменной для некоторого набора независимых переменных, которых нет в исходных наблюдениях.

Различают точечное и интервальное прогнозирование. В первом случае оценка — некоторое число, во втором — интервал, в котором находится истинное значение зависимой переменной с заданным уровнем значимости.

Рассмотрим регрессионную модель:

$$y = \alpha + \beta x + \varepsilon.$$

Действительное значение зависимой переменной при $x = x_p$ есть:

$$y_p = \alpha + \beta x_p + \varepsilon_p ,$$

где значения α, β, ε_p неизвестны и $M(\varepsilon_p) = 0$, $D(\varepsilon_p) = \sigma^2$.

Предсказанным значением является оценка y_p (точечный прогноз):

$$\hat{y}_p = a + b x_p .$$

Ошибка предсказания равна разности между предсказанным и действительным значениями:

$$\Delta_p = \hat{y}_p - y_p .$$

Ошибка предсказания и дисперсия прогноза есть:

$$M(\Delta_p) = 0,\ D(\Delta_p) = \left[1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{n\,var(x)}\right]\sigma^2 .$$

Из формулы следует, что чем больше x_p отклоняется от выборочного среднего $\bar{x}$, тем больше дисперсия ошибки предсказания, и чем больше объем выборки n, тем меньше дисперсия.

Заменяя в дисперсии прогноза σ^2 на ее оценку S^2 (остаточная дисперсия) и извлекая квадратный корень, получим *стандартную ошибку предсказания*:

$$S_p = S\sqrt{1+\frac{1}{n}+\frac{(x_p-\overline{x})^2}{n\,var(x)}}\,.$$

Доверительный интервал для действительного значения y_p определяется выражением

$$\hat{y}_p - t_{кр}S_p < y_p < \hat{y}_p + t_{кр}S_p\,,$$

где $t_{кр}$ – критическое значение t-статистики при заданном уровне значимости и числе $v = n - 2$ степени свободы.

На рис. 2.4 в общем виде показано соотношение между доверительным интервалом предсказания и значением объясняющей переменной.

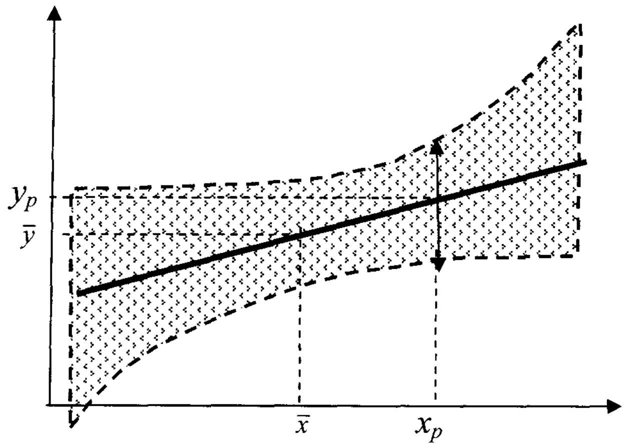

Рис. 2.4

Отрезок, помеченный на рис. 2.4 стрелками, определяет доверительный интервал предсказания в точке x_p.

Пример. 2.15. По данным о зависимости объема продаж y фирмы от затрат x на рекламу оценить объем продаж при затратах на рекламу, равных 5,5 усл. ед. Найти стандартную ошибку предсказания и 99%-ный доверительный интервал для полученной оценки. Построить график зависимости предсказания объема продаж с доверительным интервалом от затрат на рекламу.

Исходные данные (усл. ед.):

X	3	3	4	5	5	6	8	9	10	12
Y	68	62	65	72	70	78	76	80	90	82

▼Для решения задачи будем использовать функции Excel.

Объем продаж фирмы $\hat{y}_p$ =71,87 при затратах на рекламу x_p = 5,5 определяем функцией $\hat{y}_p$ = ПРЕДСКАЗ(x_p, массив_y, массив_x).

Для вычисления стандартной ошибки предсказания и его доверительного интервала предварительно получим значения

$$\bar{x} = 6{,}5,\ var(x) = 8{,}65,\ \ S = 4{,}24,\ t_{кр} = 3{,}36$$

с помощью функций:

$\bar{x}$ = СРЗНАЧ(массив_x); $var(x)$ = ДИСПР(массив_x);

S = СТОШУХ(массив_y, массив_x); $t_{кр}$= СТЬЮДРАСПОБР(α, v), где n = 10, $\nu = n - 2 = 8$, $\alpha = 1 - 0{,}99 = 0{,}01$.

Тогда

$$S_p = 4{,}24\sqrt{1 + \frac{1}{10} + \frac{(5{,}5 - 6{,}5)^2}{10 \cdot 8{,}65}} = 4{,}47\,;$$

$$71{,}87 - 3{,}36 \cdot 4{,}24 < y_p < 71{,}87 + 3{,}36 \cdot 4{,}24\text{, или } 56{,}64 < y_p < 86{,}09\,.$$

Для построения графика зависимости предсказания объема продаж с доверительным интервалом от затрат на рекламу строим расчетную таблицу:

x	y	$\hat{y}_p$	S_p	y_{min}	y_{max}
3	68	65,78	4,72	49,93	81,64
3	62	65,78	4,72	49,93	81,64

x	y	$\hat{y}_p$	S_p	y_{min}	y_{max}
4	65	68,22	4,59	52,81	83,62
5	72	70,65	4,50	55,55	85,75
5	70	70,65	4,50	55,55	85,75
6	78	73,08	4,45	58,14	88,03
8	76	77,95	4,50	62,85	93,05
9	80	80,38	4,59	64,98	95,79
10	90	82,82	4,72	66,96	98,67
12	82	87,68	5,11	70,55	104,82

где x, y — исходные данные, расположенные в порядке возрастания x;

$\hat{y}_p, S_p$ — предсказание и его стандартная ошибка;

$y_{min} = \hat{y}_p - t_{кр} S_p$, $y_{max} = \hat{y}_p + t_{кр} S_p$ — соответственно нижняя и верхняя границы доверительного интервала предсказания.

При построении совместных графиков $\hat{y}_p, y_{min}, y_{max}$ используем **Мастер диаграмм** Microsoft Excel. В результате получим график, показанный на рис. 2.5.

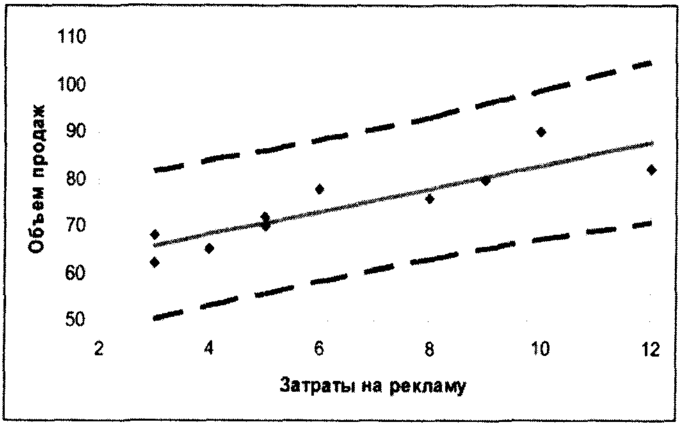

Рис. 2.5

3. МОДЕЛЬ МНОЖЕСТВЕННОЙ РЕГРЕССИИ

3.1. Классическая нормальная линейная регрессионная модель

Обобщением линейной регрессионной модели с одной объясняющей переменной является линейная регрессионная модель с k-объясняющими переменными (*модель множественной регрессии*):

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \varepsilon,$$

где $\beta_0, \beta_1, \ldots, \beta_k$ — параметры модели;

$x_1, x_2, \ldots, x_\kappa$ — объясняющие переменные;

ε — случайный член.

Случайный член ε удовлетворяет тем же предпосылкам (условия Гаусса–Маркова), что и в модели с парной регрессией, но на объясняющие переменные наложено условие: **случайные члены в любом наблюдении должны быть статистически независимы от объясняющих переменных.**

При выполнении условий Гаусса–Маркова модель называется *классической нормальной линейной регрессионной моделью*.

Кроме того, предполагается, что объясняющие переменные **некоррелированы** друг с другом.

На основе n наблюдений оценивается выборочное уравнение регрессии:

$$\hat{y} = b_0 + b_1 x_1 + \ldots + b_k x_k ,$$

где $b_0, b_1, \ldots, b_k$ оценки параметров $\beta_0, \beta_1, \ldots, \beta_k$.

Для оценки параметров регрессии используется метод наименьших квадратов, в соответствии с которым минимизируется сумма квадратов остатков:

$$Q = \sum e_i^2 = \sum (y_i - \hat{y}_i)^2 \rightarrow min .$$

Необходимым условием ее минимума является равенство нулю всех ее частных производных по $b_0, b_1, \ldots, b_k$.

В результате приходим к системе из $(k + 1)$ линейных уравнений с $(k + 1)$ неизвестными, называемой *системой нормальных уравнений*.

Ее решение в явном виде обычно записывается в матричной форме, иначе оно становится слишком громоздким.

Оценки параметров модели и их теоретические дисперсии в матричной форме определяются выражениями:

$$b = (X^T X)^{-1} X^T Y, \quad D(b_j) = (X^T X)_{jj}^{-1} \sigma^2,$$

где b — вектор с компонентами $b_0, b_1, \ldots, b_k$;

X — матрица значений объясняющих переменных;

Y — вектор значений зависимой переменной;

σ^2 — дисперсия случайного члена.

Несмещенной оценкой σ^2 является величина S^2 (остаточная дисперсия):

$$S^2 = \frac{\sum e_i^2}{n - k - 1}.$$

Величина S называется *стандартной ошибкой регрессии.*

Заменяя в теоретических дисперсиях неизвестную дисперсию σ^2 ее оценкой S^2 и извлекая квадратный корень, получим *стандартные ошибки коэффициентов регрессии*:

$$S_{bj} = S \sqrt{(X^T X)_{jj}^{-1}}.$$

Если предпосылки относительно случайного члена ε выполняются, оценки параметров множественной регрессии являются *несмещенными, состоятельными и эффективными.*

В дальнейшем определение коэффициентов регрессии и их стандартных ошибок производится без использования матричной алгебры.

3.2. Анализ вариации зависимой переменной

Пусть в уравнении регрессии содержится k объясняющих переменных и, допустим, можно разложить дисперсию зависимой переменной на объясненную и необъясненную составляющие:

$$var(y) = var(\hat{y}) + var(e).$$

Используя определение выборочной дисперсии, это уравнение можно представить в виде:

$$\sum(y_i - \bar{y})^2 = \sum(\hat{y}_i - \bar{y})^2 + \sum e_i^2.$$

Обозначим:

$TSS = \sum(y_i - \bar{y})^2$ — общий разброс зависимой переменной;

$ESS = \sum(\hat{y} - \bar{y})^2$ — разброс, объясненной регрессией;

$USS = \sum e_i^2$ — разброс, не объясненной регрессией.

Тогда

$$\underset{(n-1)}{TSS} = \underset{(k)}{ESS} + \underset{(n-k-1)}{USS}$$

(в скобках указано число степеней свободы, соответствующее каждому члену уравнения).

Замечание. Любая сумма квадратов связана с числом степеней свободы, т. е. с числом независимого варьирования переменной. Существует равенство между числом степеней свободы для этого уравнения. Отнесение каждой суммы квадратов этого уравнения на одну степень свободы приводит их к сравнимому виду.

Коэффициент детерминации есть доля объясненной части разброса зависимой переменной, т. е.

$$R^2 = ESS / TSS = 1 - USS / TSS.$$

Он является *мерой объясняющего качества* уравнения регрессии по сравнению с горизонтальной линией $\hat{y} = \bar{y}$.

Поскольку коэффициент R^2 измеряет долю дисперсии, совместно объясненной независимыми переменными, то, казалось бы, можно определить отдельный вклад каждой независимой переменной и таким образом получить меру ее относительной важности. Однако если объясняющие переменные, например x_1 и x_2, сильно коррелируют между собой, то они объясняют одну и ту же часть разброса переменной y, и в этом случае трудно оценить вклад каждой из переменных в объяснение поведения y. Геометрический смысл коэффициента детерминации при использовании одной и двух объясняющих переменных показан на рис. 3.1.

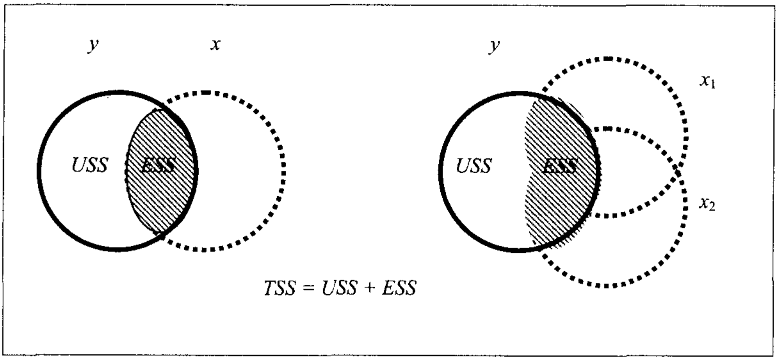

Рис. 3.1

С добавлением еще одной переменной R^2 обычно увеличивается, хотя это и не обязательно означает улучшение качества регрессионной модели.

Для компенсации такого увеличения R^2 вводится *скорректированный коэффициент детерминации* с поправкой на число степеней свободы:

$$\overline{R^2} = 1 - \frac{n-1}{n-k-1} \cdot \frac{USS}{TSS}.$$

В отличие от R^2, скорректированный коэффициент $\overline{R^2}$ может уменьшиться при введении в модель новой объясняющей переменной, следовательно, добавлять такую переменную нецелесообразно.

Проверка гипотезы H_0: $\beta_j = 0$

Статистическая значимость коэффициентов множественной линейной регрессии с k объясняющими переменными проверяется на основе t-статистики, имеющей распределение Стьюдента с $\nu = n - k - 1$ степенями свободы:

$$t_j = b_j / S_{b_j},$$

где t-тесты для коэффициентов множественной регрессии.

Определение доверительных интервалов выполняется так же, как и в парном регрессионном анализе с учетом числа степеней свободы.

Последовательный отсев несущественных факторов составляет основу многошагового регрессионного анализа.

Однако по коэффициентам регрессии нельзя сказать, какой из факторов оказывает наибольшее влияние на зависимую переменную, так как коэффициенты регрессии между собой несопоставимы (они измерены разными единицами).

Различия в единицах измерения факторов устраняют с помощью частных коэффициентов эластичности, рассчитываемых по формуле

$$Э_j = b_j \frac{\bar{x}_j}{\bar{y}},$$

где $\bar{x}_j$ — среднее значение изучаемого фактора.

Частные коэффициенты эластичности показывают, на сколько процентов в среднем изменяется зависимая переменная с изменением на 1% каждого фактора при фиксированном значении других факторов.

Проверка гипотезы H_0: $\beta_1 = \beta_2 = \ldots = \beta_k = 0$

Предположим, что в модель множественной регрессии включен свободный член, тогда

$$TSS = ESS + USS,$$

где ESS — объясняющая сумма квадратов отклонения с $\nu_1 = k$ числом степеней свободы;

USS — остаточная (необъясненная) сумма квадратов с $\nu_2 = n - k - 1$ степенями свободы.

Для определения того, действительно ли объясненный разброс ESS больше случайного USS, используется F-тест.

Построим F-статистику:

$$F = \frac{ESS / k}{USS / (n - k - 1)}$$

(для сопоставимости ESS и USS их значения привели на одну степень свободы).

После деления числителя и знаменателя этого выражения на TSS можно вычислить F-статистику на основе R^2:

$$F = \frac{R^2}{1 - R^2} \frac{n - k - 1}{k}.$$

Показатели F и R^2 равны или не равны нулю одновременно, поэтому принятие гипотезы H_0: $F = 0$ равнозначно статистической незначимости R^2.

Величина F имеет распределение Фишера с $\nu_1 = k$, $\nu_2 = n - k - 1$ степенями свободы.

F-тест для проверки значимости R^2 выполняется так же, как и в парном регрессионном анализе с поправкой на число степеней свободы.

Чаще всего F-тест используется для оценки того, значимо ли объяснение, даваемое уравнением в целом.

Проверка гипотезы H_0: $F = 0$ равнозначна проверке гипотезы H_0: $\beta_1 = \beta_2 = \ldots = \beta_k = 0$ об одновременном равенстве нулю всех коэффициентов линейной регрессии, за исключением свободного члена.

Если объясняющие способности независимых переменных перекрываются (сильная корреляция между ними), то t-тест для каждой переменной окажется незначим, в то время как F-тест для уравнения в целом вполне может быть значимым.

Проверка гипотезы H_0: $\beta' = \beta''$ (тест Чоу)

Пусть имеются две выборки объемов n_1 и n_2. Для каждой из этих выборок оценено уравнение регрессии с k-объясняющими переменными, т. е.

$$\hat{y}' = b_0' + b_1' x_1 + \ldots + b_k' x_k,$$

с необъясненной суммой квадратов USS_1, ($\nu = n_1 - k - 1$) и

$$\hat{y}'' = b_0'' + b_1'' x_1 + \ldots + b_k'' x_k$$

с необъясненной суммой квадратов USS_2, ($\nu = n_2 - k - 1$).

Проверяется нулевая гипотеза H_0: $\beta' = \beta''$, т. е. все соответствующие коэффициенты этих уравнений равны друг другу.

Пусть оценено уравнение регрессии того же вида сразу для всех (n_1 + n_2) наблюдений с необъясненной суммой квадратов USS_0, ($\nu = n_1 + n_2 - k - 1$).

Тогда рассматривается F-статистика:

$$F = \frac{USS_0 - (USS_1 + USS_2)}{USS_1 + USS_2} \frac{n_1 + n_2 - 2k - 2}{k + 1},$$

которая имеет распределение Фишера с $\nu_1 = k + 1$ и $\nu_2 = n_1 + n_2 - 2k - 2$ степенями свободы.

F-статистика будет близка к нулю, если USS_0 = USS_1 + USS_2, т. е. если уравнения регрессии для обеих выборок одинаковы.

Значимость F, соответствующая расчетному значению F сравнивается с заданным уровнем.

Если значимость F меньше заданного уровня α, то H_0 отклоняется, т. е. нельзя построить единое уравнение регрессии для обеих выборок.

Пакет анализа Excel (программа «Регрессия»)

При использовании Пакета анализа Excel (регрессия) коэффициенты множественной регрессии, их стандартные отклонения и другие показатели вычисляются автоматически и одновременно. При этом интерпретация получаемых показателей такая же, как в парной регрессии с учетом числа степеней свободы.

Пример 3.1. По данным бюджетного обследования семи случайно выбранных семей изучалась зависимость накопления y от дохода x_1 и стоимости имущества x_2.

Исходные данные (усл. ед.):

y	3	7	5	4	2	7	6
x_1	40	55	45	30	30	60	50
x_2	60	40	40	15	90	30	30

Используя Пакет анализа Excel (Регрессия), проанализировать зависимость накопления от дохода и имущества.

▼ Результаты расчетов с использованием инструмента **Регрессия** выводятся под общим названием **Вывод итогов** в виде таблиц.

Регрессионная статистика	
Множественный R	0,970
R-квадрат	**0,942**
Нормированный R-квадрат	0,913
Стандартная ошибка	0,577
Наблюдения	7

Дисперсионный анализ					
	Df	***SS***	***MS***	***F***	**Значимость *F***
Регрессия	2	21,527	10,763	32,355	**0,0034**
Остаток	4	1,331	0,333		
Итого	6	22,857			

	Коэффициенты	**Стандартная ошибка**	**t-статистика**	**Р-значение**	**Нижние 95%**	**Верхние 95%**
Y-пересечение	**0,493**	1,229	0,402	0,708	–2,917	3,904
X_1	**0,128**	**0,022**	5,866	**0,004**	0,067	0,188
X_2	**–0,030**	**0,010**	–2,860	**0,046**	–0,058	–0,001

Коэффициенты регрессии $a = 0{,}49$; $b_1 = 0{,}128$; $b_2 = -0{,}030$.

Запишем оцененное уравнение регрессии в виде

$$\hat{y} = 0{,}49 + 0{,}128x_1 - 0{,}030x_2\,,\ R^2 = 0{,}942\,.$$

Качество модели оценивается коэффициентом детерминации R^2.

Величина $R^2 = 0{,}942$ означает, что факторами дохода и имущества можно объяснить 94,2% вариации (разброса) накопления.

Установим статистическую значимость коэффициента R^2.

Поскольку Значимость $F = 0{,}0034 < \alpha = 0{,}05$, то $R^2 = 0{,}942$ значим при 5%-ном уровне, т. е. модель в целом адекватна исходным данным.

Коэффициент $b_1 = 0{,}128 > 0$ показывает, что при увеличении дохода на 1 усл. ед. накопление увеличивается на 0,128 усл. ед.

Коэффициент $b_2 = -0{,}030 < 0$ показывает, что при увеличении имущества на 1 усл. ед. накопление уменьшается на 0,030 усл. ед.

Оценим статистическую значимость коэффициентов b_1, b_2.

Поскольку P-значение (b_1) = $0{,}004 < \alpha = 0{,}05$, то $b_1 = 0{,}128$ значим на уровне 5%. Поскольку P-значение (b_2) = $0{,}046 < \alpha = 0{,}05$, то $b_2 = -0{,}030$ также значим.

Результаты оценивания регрессии совместимы не только с полученными значениями коэффициентов регрессии, но и с некоторым их множеством (доверительным интервалом). С вероятностью 95% доверительный интервал коэффициента b_1 есть (0,067 … 0,188), а для b_2 есть (–0,058 … –0,001).

3.3. Мультиколлинеарность

Мультиколлинеарность — это коррелированность двух или несколько объясняющих переменных в уравнении регрессии. При наличии мультиколлинеарности МНК-оценки формально существуют, но обладают рядом недостатков:

- небольшое изменение исходных данных приводит к существенному изменению оценок коэффициентов регрессии;
- оценки коэффициентов регрессии имеют большие стандартные ошибки, малую значимость, в то время как модель в целом является значимой (высокое значение R^2).

Если при оценке уравнения регрессии несколько факторов оказались незначимы, то нужно выяснить, нет ли среди них сильно коррелированных между собой.

Для отбора факторов в модель регрессии и оценки их мультиколлинеарности анализируют корреляционную матрицу.

Общий вид корреляционной матрицы, составленной из переменных y, x_1, x_2, x_3 приведен в следующей таблице.

	y	x_1	x_2	x_3
y	1			
x_1	$r(y,x_1)$	1		
x_2	$r(y,x_2)$	$r(x_1,x_2)$	1	
x_3	$r(y,x_3)$	$r(x_1,x_3)$	$r(x_2,x_3)$	1

При наличии корреляции один из пары связанных между собой факторов исключается. Если статистически незначим лишь один фактор, то он должен быть исключен.

В модель регрессии включаются те факторы, которые более сильно связаны с зависимой переменной, но слабо связаны с другими факторами.

Корреляционную матрицу можно получить, используя Пакет анализа Excel (Корреляция).

Пример 3.2. По данным бюджетного обследования семи случайно выбранных семей изучалась зависимость накопления y от дохода x_1, стоимости имущества x_2 и расходов на питание x_3. Исходные данные приведены в таблице (усл. ед.):

y	3	7	5	4	2	7	6
x_1	40	55	45	30	30	60	50
x_2	60	40	40	15	90	30	30
x_3	10	15	12	8	10	20	15

Используя Пакет анализа Excel (Корреляция), проанализируйте целесообразность включения в модель каждого фактора.

▼Выборочное уравнение регрессии:

$$\underset{}{\hat{y}} = 0{,}59 + \underset{(0{,}069)}{0{,}115x_1} - \underset{(0{,}012)}{0{,}030x_2} + \underset{(0{,}191)}{0{,}034x_3}, \quad R^2 = 0{,}941;\ \overline{R^2} = 0{,}885,$$

где коэффициенты при факторных переменных незначимы.

Корреляционная матрица из этих переменных:

	y	x_1	x_2	x_3
y	1			
x_1	0,907	1		
x_2	–0,664	–0,380	1	
x_3	0,830	0,935	–0,283	1

Из рассмотрения корреляционной матрицы заключаем, что факторы x_1, x_3 дублируют друг друга, $r(x_1, x_3) = 0{,}935$. В анализ целесообразно включить фактор x_1, а не x_3, так как его связь с результативным признаком у сильнее, чем у фактора x_3, $r(y, x_1) = 0{,}907 >$ $> r(y, x_3) = 0{,}830$.

Оценив регрессию y на x_1, x_2, получим:

$$\hat{y} = 0{,}49 + \underset{(0{,}021)}{0{,}127x_1} - \underset{(0{,}010)}{0{,}029x_2}, \quad R^2 = 0{,}941;\ \overline{R^2} = 0{,}912,$$

где коэффициенты при факторных переменных значимы.

Заметим, что первоначальное скорректированное значение $\overline{R^2} = 0{,}885$ при отбраковке дублирующего фактора x_3 увеличилось до $\overline{R^2} = 0{,}912$.▲

Частная корреляция

Если рассматриваемые случайные величины коррелируют друг с другом, то на значении коэффициента парной корреляции частично сказывается влияние других величин.

Коэффициенты частной корреляции характеризуют тесноту связи между результатом и соответствующим фактором при устранении влияния других факторов, включенных в уравнение регрессии.

Предположим, что имеется регрессионная модель

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

и необходимо оценить корреляцию между зависимой переменной y и объясняющей переменной x_1 после исключения влияния объясняющей переменной x_2.

С этой целью производим следующую процедуру:

- оцениваем регрессию $\hat{y} = \hat{\alpha}_1 + \hat{\alpha}_2 x_2$;
- оцениваем регрессию $\hat{x}_1 = \hat{\gamma}_1 + \hat{\gamma}_2 x_2$;
- удаляем влияние x_2, взяв остатки $e_y = y - \hat{y}$ и $e_{x1} = x_1 - \hat{x}_1$;
- определяем (выборочный) коэффициент частной корреляции между y и x_1 как (выборочный) коэффициент корреляции между e_y и e_{x1}:

$$r(y, x_1 | x_2) = r(e_y, e_{x1}).$$

Для коэффициента частной корреляции $r(y,x_1 | x_2)$ используется также обозначение $r_{yx1.x2}$.

Можно показать, что справедлива следующая формула, связывающая коэффициенты частной и обычной корреляции:

$$r_{yx1.x2} = \frac{r(y,x_1) - r(y,x_2)r(x_1,x_2)}{\sqrt{[1 - r^2(y,x_2)][1 - r^2(x_1,x_2)]}}.$$

Значения $r_{yx1.x2}$ лежат в интервале [-1, 1], а равенство нулю означает отсутствие прямого (линейного) влияния переменной x_1 на y.

Описанная выше процедура обобщается на случай, когда исключается влияние не одной, а нескольких переменных. Порядок частного коэффициента корреляции определяется количеством факторов, влияние которых исключается.

Пример 3.3. Исходя из корреляционной матрицы, полученной по данным примера 3.1:

	y	x_1	x_2
y	1		
x_1	0,907	1	
x_2	–0,664	–0,380	1

найти частные коэффициенты корреляции.

▼Приведем частные коэффициенты корреляции.

$$r_{yx1.x2} = \frac{r(y,x_1) - r(y,x_2)r(x_1,x_2)}{\sqrt{[1-r^2(y,x_2)][1-r^2(x_1,x_2)]}} =$$

$$= \frac{0{,}907 - (-0{,}664)\cdot(-0{,}380)}{\sqrt{(1-0{,}664^2)\cdot(1-0{,}380^2)}} = 0{,}947,$$

т. е. при закреплении фактора x_2 на постоянном уровне корреляция у и x_1 оказывается более высокой (0,947 против 0,907).

$$r_{yx2.x1} = \frac{r(y,x_2) - r(y,x_1)r(x_1,x_2)}{\sqrt{[1-r^2(y,x_1)][1-r^2(x_1,x_2)]}} =$$

$$= \frac{-0{,}664 - 0{,}907\cdot(-0{,}380)}{\sqrt{(1-0{,}907^2)\cdot(1-0{,}380^2)}} = -0{,}820,$$

т. е. при закреплении фактора x_1 на постоянном уровне корреляция y и x_2 по модулю оказывается более высокой (-0,820 против -0,664).▲

3.4. Спецификация модели

Построение экономической модели включает выбор объясняющих переменных. Свойства оценок коэффициентов регрессии в значительной мере зависят от правильной спецификации модели.

Влияние отсутствия в модели переменной, которая должна быть включена

Предположим, что переменная у зависит от двух переменных x_1 и x_2 в соответствии с соотношением

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \varepsilon.$$

Однако считается, что модель выглядит как

$$y = \alpha + \beta_1 x_1 + \varepsilon,$$

и оценивается регрессия

$$\hat{y} = a + b_1 x_1 .$$

В этом случае оценка b_1 является *смещенной.*

Смещенность оценки b_1 связана с тем, что если не учесть x_2 в регрессии, то переменная x_1 будет играть двойную роль: отражать свое прямое влияние и заменять переменную x_2 в описании ее влияния.

Коэффициент R^2 для данной регрессии отражает общую объясняющую способность переменной x_1 в обеих ролях и является завышенной оценкой.

Влияние включения в модель переменной, которая не должна быть включена

Допустим, что истинная модель есть

$$y = \alpha + \beta_1 x_1 + \varepsilon.$$

Однако считается, что ею является

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \varepsilon,$$

и оценивается регрессия

$$\hat{y} = a + b_1 x_1 + b_2 x_2 .$$

Оценки коэффициентов регрессии в этом случае являются *несмещенными*, но *неэффективными*. Практически обнаруживается, что коэффициент b_2 статистически незначим и переменная x_2 исключается из модели.

Замещающие переменные

Предположим, что истинной моделью является

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon,$$

и допустим, что не имеется данных по существенной переменной x_1.

Если не включить в модель эту переменную, то регрессия может пострадать от смещения оценок и статистическая проверка будет некорректной.

Если вместо отсутствующей переменной x_1 использовать ее заменитель z, линейно связанный с x_1, и построить регрессию

$$\hat{y} = a + b_2 x_2 + \ldots + b_k x_k + cz\,,$$

то оценки $b_2, \ldots, b_k$, их стандартные ошибки и коэффициент R^2 будут такими же, как с использованием x_1. Единственным недостатком является то, что отсутствует оценка коэффициента при самой величине x_1, а величина а не является оценкой α.

В качестве замещающей переменной, например, для показателя технического прогресса может использоваться время.

Лаговые переменные

При использовании данных временного ряда на текущие значения зависимой переменной могут влиять не только текущие значения объясняющих переменных, но также их значения с некоторым запаздыванием.

В общем случае, если какая-то переменная появляется в модели с запаздыванием на τ периодов, она записывается с нижним индексом $(t - \tau)$, например:

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t-\tau} + \varepsilon_t.$$

Сдвиг τ, характеризующий запаздывание в воздействии фактора на результат, называется *лагом*. Переменная, влияние которой характеризуется некоторым запаздыванием, называется *лаговой*.

Инструментальные переменные

При наличии *корреляции* между объясняющими переменными и случайным членом МНК-оценки являются *смещенными* и *несостоятельными*. Для получения состоятельных оценок используется **метод инструментальных переменных** (ИП).

Сущность метода ИП заключается в замене непригодной объясняющей переменной такой переменной, которая некоррелирована со случайным членом и коррелированна с исходной переменной.

Пусть в модели $y = \alpha + \beta x + \varepsilon$ переменная x коррелированна со случайным членом ε. Предположим, что можно найти другую переменную z, которая коррелированна с x, но не коррелированна с ε.

Основанная на использовании ИП оценка параметра β, определяемая как

$$b_{ИП} = \frac{cov(z,y)}{cov(z,x)},$$

является состоятельной.

В больших выборках ошибка исчезает при условии, что переменная z действительно распределена независимо от ε и $\text{cov}(z, x) \neq 0$. Следовательно, на больших выборках $b_{\text{ип}}$ будет стремиться к истинному значению β.

Таким образом, оценка $b_{\text{ип}}$ является *состоятельной*, но в общем случае *смещенной* и *неэффективной*.

Фиктивные переменные

При исследовании влияния качественных признаков в модель можно вводить фиктивные переменные, принимающие, как правило, два значения: единица, если данный признак присутствует в наблюдении, и нуль — при его отсутствии.

Если включаемый в рассмотрение качественный признак имеет не два, а несколько значений, то используют несколько фиктивных переменных, число которых должно быть на единицу меньше числа значений признака.

При назначении фиктивных переменных исследуемая совокупность по числу значений качественного признака разбивается на группы. Одну из групп выбирают как эталонную (группа 0) и определяют фиктивные переменные для остальных.

Например, если качественный признак имеет три значения, то две фиктивные переменные определяются следующим образом:

группа 0 : $z_1 = z_2 = 0$;
группа 1 : $z_1 = 1, z_2 = 0$ или
группа 2 : $z_1 = 0, z_2 = 1$,
или

$$z_1 = \begin{cases} 1, & (\text{группа } 1) \\ 0, & (\text{остальные}) \end{cases}, \quad z_2 = \begin{cases} 1, & (\text{группа } 2) \\ 0, & (\text{остальные}) \end{cases}.$$

Введение в регрессию фиктивных переменных существенно улучшает качество ее оценивания.

Пример 3.4. Построить линейную регрессионную модель зависимости заработной платы (y) 16 работников фирмы от возраста (x) с использованием фиктивной переменной (z) по фактору «пол», взяв женский пол в качестве эталонной группы.

▼Исходные данные представим в виде таблицы

N	*Пол*	*y*	*x*	*z*
1	Муж.	338	35	1
2	Жен.	329	55	0
3	Жен.	299	32	0
4	Муж.	334	44	1
5	Жен.	305	33	0
6	Жен.	304	58	0
7	Муж.	336	38	1
8	Жен.	310	50	0
9	Муж.	340	45	1
10	Жен.	300	47	0
11	Муж.	323	38	1
12	Муж.	321	35	1
13	Жен.	293	35	0
14	Жен.	313	48	0
15	Муж.	340	58	1
16	Муж.	340	54	1

Оценив регрессию между y и x, получим

$$\hat{y} = 299{,}19 + 0{,}48x, \; R^2 = 0{,}067.$$

Уравнение регрессии статистически незначимо.

Для учета качественного фактора «пол» введем в модель фиктивную переменную:

$$z = \begin{cases} 0, (\text{Жен.}) \\ 1, (\text{Муж.}) \end{cases}$$

Оценив регрессию между y и x, z, получим

$$\hat{y} = 279{,}64 + 0{,}60x + 28{,}2\,z\,,\ R^2 = 0{,}809.$$

Уравнение регрессии статистически значимо.

Из полученного уравнение регрессии следует, что при одном и том же возрасте заработная плата у мужчин на 28,2$ выше, чем у женщин.

Коэффициент 28,2 при фиктивной переменной z статистически значим, следовательно, имеем две различные модели:

$\hat{y} = 279{,}64 + 0{,}60x$ — для женщин, ($z = 0$);

$\hat{y} = 307{,}84 + 0{,}60x$ — для мужчин, ($z = 1$). ▲

Упражнение 3.1. Имеются данные о числе решенных задач (задание — 10 задач) на вступительных X и курсовых $У$ экзаменах по математике 16 студентами:

№	*X*	*Y*	*Пол*	*№*	*X*	*У*	*Пол*
1	10	9	Муж.	9	10	10	Муж.
2	8	8	Муж.	10	6	4	Жен.
3	4	2	Жен.	11	9	9	Муж.
4	10	10	Муж.	12	6	4	Жен.
5	6	4	Жен.	13	10	9	Муж.
6	6	4	Жен.	14	4	1	Жен.
7	10	9	Муж.	15	9	9	Муж.
8	4	2	Жен.	16	8	6	Жен.

Постройте линейную регрессионную модель $У$ по X с использованием фиктивной переменной по фактору «пол». Можно ли считать, что эта модель одна и та же для юношей и девушек?

Ответ. $\hat{y} = -3{,}30 + 1{,}22 \cdot x + 0{,}85 \cdot z,\ R^2 = 0{,}975$.

Имеем две различные модели для юношей и девушек:

$\hat{y} = -3{,}30 + 1{,}22x$ — для девушек;

$\hat{y} = \ -2{,}45 + 1{,}22x$ — для юношей.

3.5. Производственная функция Кобба–Дугласа

Макроэкономическая производственная функция — это статистически значимая связь между объемом выпуска Y, капитальными затратами K и затратами труда L.

Для моделирования и решения задач как на макро-, так и на микроэкономическом уровнях часто используют *производственную функцию Кобба–Дугласа (КД)*:

$$Y = AK^{\alpha}L^{\beta}\varepsilon,$$

где A, α, β — параметры модели, причем $A > 0$; $0 < \alpha < 1$; $0 < \beta < 1$.

***Свойства производственной функции Кобба–Дугласа*:**

1) *эластичность выпуска* продукции. Эластичность выпуска продукции по капиталу и труду равна соответственно α и β:

$$\frac{\partial Y}{Y} / \frac{\partial K}{K} = \alpha, \quad \frac{\partial Y}{Y} / \frac{\partial L}{L} = \beta$$

Это означает, что увеличение затрат капитала на 1% приведет к росту выпуска продукции на α %, а увеличение затрат труда на 1% приведет к росту выпуска на β %;

2) *эффект от масштаба* производства. При росте затрат каждого из факторов K, L в λ раз выпуск возрастает в $\lambda^{\alpha+\beta}$ раз. Это означает следующее:

• если $(\alpha + \beta) > 1$, то функция КД имеет возрастающую отдачу от масштаба производства;

• если $(\alpha + \beta) < 1$, то функция КД имеет убывающую отдачу от масштаба производства;

• если $(\alpha + \beta) = 1$, то функция КД имеет постоянную отдачу от масштаба производства;

3) *прогнозируемые доли производственных факторов*. В рыночной экономике оценки α и β интерпретируются как прогнозируемые доли дохода, полученные соответственно за счет капитала и труда.

Производственная функция КД нелинейна относительно параметров α, β. Для оценки параметров производственной функции КД с помощью МНК необходимо прологарифмировать уравнение:

$$\ln Y = \ln A + \alpha \ln K + \beta \ln L + \ln \varepsilon .$$

По рядам данных Y, K, L рассчитываются ряды их логарифмов lnY, lnK, lnL, и для них оценивается уравнение регрессии.

Пример 3.5. Известны данные (усл. ед.) об объеме производства Y, капитальных затрат K и затратах труда L некоторой страны за 12 лет. По выборочным данным произвести оценки параметров α, β производственной функции КД и сделать вывод.

▼Представим исходные и преобразованные данные в виде таблицы.

Y	***K***	***L***	***lnY***	***lnK***	***lnL***
100	100	100	4,605	4,605	4,605
112	114	110	4,718	4,736	4,700
124	131	123	4,820	4,875	4,812
143	149	125	4,963	5,004	4,828
151	176	138	5,017	5,170	4,927
155	198	140	5,043	5,288	4,942
153	216	145	5,030	5,375	4,977
184	236	154	5,215	5,464	5,037
189	266	154	5,242	5,583	5,037
227	335	196	5,425	5,814	5,278
218	397	193	5,384	5,984	5,263
179	417	147	5,187	6,033	4,990

Оцененное преобразованное уравнение регрессии:

$$ln\hat{Y} = -0{,}303 + 0{,}149 \cdot lnK + 0{,}922 \cdot lnL,\ R^2 = 0{,}969.$$

В целом полученное уравнение статистически значимо.

Оценки $\alpha = 0{,}149$; $\beta = 0{,}922$. Это означает, что увеличение затрат капитала на 1% приводит к росту выпуска продукции на 0,15%, а увеличение затрат труда на 1% — к росту выпуска на 0,92%.▲

3.6. Линейные регрессионные модели финансового рынка

3.6.1. Формирование оптимального портфеля

Портфелем ценных бумаг инвестора называется совокупность ценных бумаг, принадлежащих данному инвестору.

При формировании портфеля следует различать рисковые и безрисковые активы.

Рисковые активы — это активы, доходность которых в будущем неопределенна.

Безрисковые активы — это активы, будущая доходность которых известна в момент погашения. Как правило, это краткосрочные правительственные облигации.

Портфель ценных бумаг, содержащий самые разнообразные типы ценных бумаг, называется *диверсифицированным портфелем.*

Структура портфеля акций или других ценных бумаг описывается показателями x_i, характеризующими долю стоимости акций данного вида в общей стоимости приобретаемого или имеющегося портфеля, причем выполняется соотношение

$$\sum_i x_i = 1 .$$

Доходность портфеля — это сумма доходностей всех составляющих его бумаг:

$$r_p = \sum_{i=1}^{n} x_i r_i ,$$

где x_i — доля i-го актива в портфеле;

r_i — доходность i-го актива;

n — число активов в портфеле.

Доходность i-го актива определяется выражением

$$r = \frac{d + (P_1 - P_0)}{P_0} \cdot 100 ,$$

где d — дивиденды, выплаченные в течение периода;

P_0 — курс акции на начало периода;

P_1 — курс акции на конец периода.

Поскольку доходность составляющих портфель ценных бумаг случайна, то и доходность r_p портфеля есть также случайная величина и имеет математическое ожидание и дисперсию.

Ожидаемая доходность портфеля акций в целом при заданной его структуре определяется как математическое ожидание значений его доходности, достигаемых в каждом возможном будущем состоянии экономики:

$$m_p = M(r_p) = \sum_i x_i M(r_i) = \sum_i x_i \bar{r}_i .$$

Дисперсия портфеля определяется выражением

$$\sigma_p^2 = \sigma^2(r_p) = \sum_{i,j} x_i x_j \sigma_{ij} = \sum_i x_i^2 \sigma_i^2 + \sum_{i \neq j} x_i x_j \sigma_{ij}$$

или

$$\sigma_p^2 = \sum_i x_i^2 \sigma_i^2 + 2\sum_{i<j} x_i x_j \sigma_{ij} ,$$

где по одинаковым нижним индексам производится суммирование.

Рискованность одного актива измеряется дисперсией или стандартным отклонением доходности по этому активу, а риск портфеля — дисперсией или стандартным отклонением доходности портфеля. Для оценки риска портфеля обычно используется стандартное отклонение $\sigma_p = \sqrt{\sigma_p^2}$.

Эффект диверсификации. Рассмотрим понижающий риск портфеля эффект диверсификации.

Из формулы для дисперсии портфеля σ_p^2 следует, что с увеличением ковариации (корреляции) доходностей активов, составляющих единый портфель, возрастает дисперсия, а следовательно, риск этого портфеля.

Эффект диверсификации проявляется при отрицательной корреляции доходностей активов. Инвестор может снизить риск портфеля, удерживая его ожидаемую доходность при помощи сочетания активов с низкой (лучше отрицательной) корреляцией.

Данная стратегия, приводящая к максимально возможному снижению риска при сохранении требуемого уровня доходности, состоит в выборе таких активов, доходности которых имеют возможно меньшую положительную корреляцию.

Инвесторы оценивают портфель по двум параметрам: ожидаемой доходности и риску. Каждый владелец портфеля ценных бумаг сталкивается с дилеммой: иметь доходность побольше, а риск поменьше. Сделать это одновременно нельзя, поэтому приходится выбирать между доходностью и риском; этот выбор в конечном счете определяется отношением инвестора к риску.

Эффективный портфель — это допустимый портфель с наибольшей ожидаемой доходностью для заданного уровня риска.

Совокупность всех эффективных портфелей называется *эффективным множеством портфелей*. Смысл эффективного множества состоит в том, что именно из него инвесторы выбирают свои оптимальные портфели.

Оптимизация портфеля обычно состоит из двух этапов:

1) выбора оптимальной комбинации рисковых активов;

2) объединения полученного оптимального набора рисковых активов с безрисковыми активами.

Пусть портфель составлен из двух видов рисковых активов:

$$r_p = x_1 r_1 + x_2 r_2 ,$$

где x_1 — доля рискового актива 1;

x_2 — доля рискового актива 2.

Ожидаемая доходность и дисперсия портфеля определяются следующим образом:

$$m_p = x_1 \bar{r}_1 + x_2 \bar{r}_2 ,\ \sigma_p^2 = x_1^2 \sigma_1^2 + x_2^2 \sigma_2^2 + 2 x_1 x_2 \sigma_{12} ,$$

где $\bar{r}_1, \bar{r}_2$ — ожидаемые доходности рисковых активов;

σ_1, σ_2 — стандартные отклонения активов;

σ_{12} — коэффициент ковариации между активами.

Пример 3.6. Пусть инвестор формирует портфель из рисковых активов двух видов: $\bar{r}_1 = 10\%$, $\sigma_1 = 2$; $\bar{r}_2 = 30\%$, $\sigma_2 = 3$ и $\sigma_{12} = -2$.

Ожидаемая доходность и дисперсия такого портфеля:

$m_p = 0{,}1x_1 + 0{,}3x_2$, $\sigma_p^2 = 4x_1^2 + 9x_2^2 - 4x_1x_2$, а риск равен σ_p.

Для определения эффективного множества портфелей, составленного из двух рисковых активов, построим по точкам зависимость «Риск — доходность». Все доступные комбинации риска и доходности показаны в таблице и на рис. 3.2.

Задано		**Вычислено**			**Портфель**
x_1	x_2	σ_p^2	σ_p	m_p	
1,00	0,00	4,00	2,00	0,100	*A*
0,80	0,20	2,28	1,51	0,140	
0,70	0,30	1,93	1,39	0,160	
0,60	0,40	1,92	1,39	0,180	
0,50	0,50	2,25	1,50	0,200	
0,40	0,60	2,92	1,71	0,220	
0,20	0,80	5,28	2,30	0,260	
0,00	1,00	9,00	3,00	0,300	*B*
Вычислено					
0,65	0,35	1,88	1,37	0,17	*D*

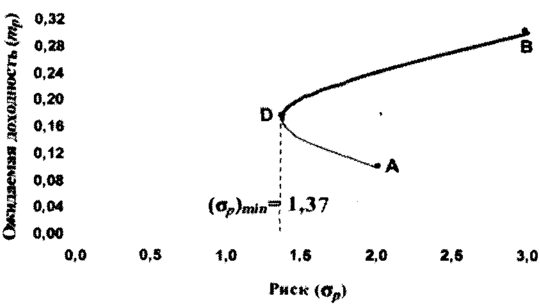

Рис. 3.2. Определение эффективного множества портфелей, составленного из двух рисковых активов

Дадим анализ кривой, приведенной на рис. 3.2. Точка A на графике соответствует портфелю, полностью составленному из первого актива, а точка B — полностью составленному из второго. При перемещении из точки A к точке B наблюдается не только повышение средней ставки доходности, но и снижение стандартного отклонения до значения, соответствующего точке D, после чего стандартное отклонение повышается.

Данный график показывает наличие структуры, обеспечивающей минимальный риск портфеля $(\sigma_p)_{\min}$, составленного из двух рисковых активов (самая левая точка D на графике). Найдем координаты точки D, соответствующей минимальному значению σ_p.

Для этого воспользуемся необходимым условием экстремума функции:

$$\sigma_p^2 = x_1^2\sigma_1^2 + (1-x_1)^2\sigma_2^2 + 2x_1(1-x_1)\sigma_{12}$$

от одной переменной x_1.

Приравняв к нулю ее производную по этой переменной, получим:

$$x_1 = \frac{\sigma_2^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}},\ x_2 = 1 - x_1.$$

Координаты точки D: $x_1 = 0{,}65$; $x_2 = 0{,}35$; при этом $(\sigma_p)_{\min} = 1{,}37$.

Портфели, расположенные от минимально рискованного портфеля и выше (кривой отрезок DB), принадлежат *эффективному множеству* в модели Марковица.

Теперь рассмотрим комбинации риск — доходность, которые можно получить посредством объединения безрискового актива с двумя рисковыми активами, параметры которых остаются прежними, а безрисковая ставка доходности $r_б = 0{,}08$.

На рис. 3.3 показано графическое представление всех возможных комбинаций риск – доходность.

Точка H на графике соответствует портфелю, составленному только из одной безрисковый ценной бумаги.

Кривая ADB показывает зависимость риск — доходность для портфелей, составленных из рисковых активов 1 и 2, взятых в разных соотношениях.

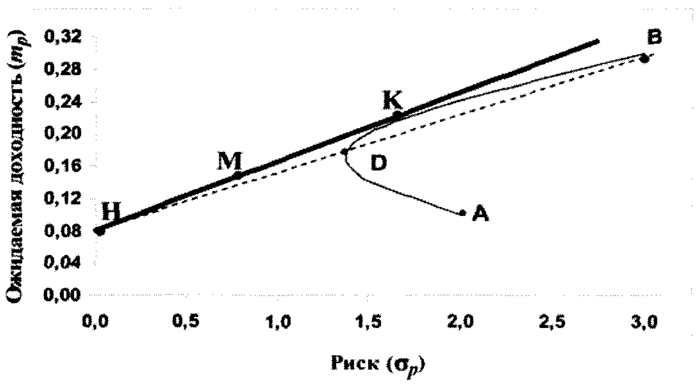

Рис. 3.3. Оптимальная комбинация рисковых активов

Прямая *HB* показывает ряд комбинаций риск — доходность, которые могут быть получены объединением безрискового актива с рисковым активом 2

Прямая линия, соединяющая точку *H* с любой точкой кривой *ADB*, представляет собой график, описывающий соотношение риск — доходность для всех комбинаций трех активов: объединения рисковых активов 1 и 2 с безрисковым активом. Наибольшее значение этого соотношения, какое только можно достичь, находится на линии *HK*, касательной к кривой *ADB*.

Рисковый портфель, соответствующий точке *K* на рис. 3.3, называется *касательным* (тангенциальным) *портфелем*. Его состав является оптимальной комбинацией рисковых активов и выражается формулой

$$x_1 = \frac{(\bar{r}_1 - r_б)\sigma_2^2 - (\bar{r}_2 - r_б)\sigma_{12}}{(\bar{r}_1 - r_б)\sigma_2^2 + (\bar{r}_2 - r_б)\sigma_1^2 - (\bar{r}_1 + \bar{r}_2 - 2r_б)\sigma_{12}}, \quad x_2 = 1 - x_1.$$

Состав касательного портфеля: $x_1 = 0{,}40$; $x_2 = 0{,}60$; при этом ожидаемая доходность и риск касательного портфеля: $m_K = 0{,}22$; $\sigma_K = 1{,}71$.

Это означает, что оптимальной комбинацией рисковых активов является $x_1 = 40\%$ рискованного актива 1 и $x_2 = 60\%$ рискованного актива 2.

Именно объединением этого портфеля рисковых активов с безрисковым активом достигается формирование *максимально эффективного портфеля.*

Уравнение прямой, проходящей через две заданные точки $H(0;r_б)$, $K(\sigma_K;m_K)$:

$$m_p = r_б + \frac{m_K - r_б}{\sigma_K}\sigma_p.$$

Это и будет *уравнение линии «Риск — доходность»* для комбинированного портфеля, составленного из безрискового актива и касательного портфеля.

Оптимальной для любого инвестора стратегией оказывается инвестирование части средств в касательный портфель, а части — в безрисковые активы.

Инвестор может выбрать позицию в любой точке на отрезке *НК* (см. рис. 3.3) в зависимости от своего отношения к риску. Пусть на этом отрезке выбрана точка *М.* Портфель, соответствующей этой точке, на 50% состоит из портфельных инвестиций в общей точке *К* (касательный портфель) и на 50% из инвестиций в безрисковый актив.

Доходность и риск портфеля в точке М:

$$m_M = r_б + 0{,}5\cdot(m_K - r_б) = 0{,}08 + 0{,}5\cdot(0{,}22 - 0{,}08) = 0{,}15;$$

$$\sigma_M = 0{,}5\cdot\sigma_K = 0{,}5\cdot1{,}71 = 0{,}855.$$

Учитывая, что касательный портфель состоит на 40% из рискового актива 1 и на 60% — из рискового актива 2 и на долю рисковых активов приходится 50% всего портфеля, определяем, что доля рискового актива 1 в портфеле *М* будет 0,5·40% = 20% .

Таким образом, состав портфеля *М* будет следующим: доля безрискового актива составляет 50%, доля рискового актива 1–20% и доля рискового актива 2–30%.

Следовательно, если инвестировать 10 000 у. е. в портфель *М*, то 5000 у. е. следует инвестировать в безрисковый актив, 2000 у. е. — в рисковый актив 1 и 3000 у. е. — в рисковый актив 2.

Таким образом, существует только один портфель с рискованными активами (касательный), который оптимальным образом можно объединить с безрисковым активом.

3.6.2. Рыночная модель доходности

Предположим, что случайный фактор F оказывает влияние на доходность r некоторой ценной бумаги.

Однофакторной моделью доходности называется уравнение регрессии

$$r = \alpha + \beta F + \varepsilon,$$

где α, β — константы (неизвестные параметры);

r, F — случайные величины;

ε — случайная ошибка.

Оценивая данное уравнение по выборочным наблюдениям методом наименьших квадратов, получают уравнение (линия регрессии):

$$m_r = \alpha + \beta F,$$

где α, β — оценки неизвестных параметров,

$$\beta = cov(r,F)/\sigma_F^2, \quad \alpha = \bar{r} - \beta \bar{F}.$$

Коэффициент β называется чувствительностью ценной бумаги к фактору F.

В роли ведущего фактора F наиболее удобно выбрать доходность рыночного индекса.

Рыночный индекс — это взвешенная сумма курсов наиболее важных для рынка ценных бумаг, а доходность рыночного индекса представляет собой их усредненную доходность.

Биржевые индексы различны по разным странам и разным биржам и определяются на основе различного числа компаний. Например, в США наиболее распространены *индекс Доу–Джонса* (*DJ*), рассчитываемый по 30 наиболее значимым корпорациям. В России наиболее известным является *индекс РТС*, определяемый по акциям наиболее крупных российских компаний.

Портфель ценных бумаг, определяемый по акциям компаний, учитываемых при формировании биржевых индексов, называется *индексным портфелем*.

Рыночный портфель — с теоретической точки зрения совокупность акций, обращающихся на фондовом рынке; с практической точки зрения в качестве рыночного рассматривается индексный портфель.

Предположим, что доходность любой ценной бумаги линейно зависит от доходности рыночного индекса. Рыночная модель доходности i-й ценной бумаги записывается в виде

$$r_i = \alpha_i + \beta_i r_F + \varepsilon_i,$$

где r_i, r_F — фактические доходности акции данного вида и рыночного портфеля в определенные моменты времени;

ε_i — случайная ошибка (невязка);

α, β — неизвестные параметры уравнения.

Рыночный индекс r_F в определенной степени отражает состояние экономики в целом, а рыночная модель показывает, насколько доходность ценной бумаги соответствует экономической динамике страны.

На основе выборочных наблюдений определяется *характеристическая линия* данной ценной бумаги:

$$m_i = \alpha_i + \beta_i m_F,$$

где m_i, m_F — ожидаемая доходность акции данного вида и рыночного портфеля.

Коэффициенты α_i и β_i рассчитываются по формулам

$$\beta_i = \sigma_{iF} / \sigma_F^2\,,\ \alpha = \bar{r}_i - \beta \bar{r}_F\,,$$

где σ_{iF} — ковариация между доходностью i-й ценной бумаги и доходностью индексного портфеля;

$\bar{r}_F$ — среднее значение, а σ_F — стандартное отклонение доходности индекса.

Коэффициент альфа характеризует ожидаемую доходность акции данного вида при условии нулевой доходности рыночного портфеля.

Коэффициент бета определяет влияние рынка на данные ценные бумаги.

Коэффициенты альфа и бета определяются для акций каждой компании, котировавшихся на фондовом рынке и приносивших доход в течение определенного периода.

Общий риск по акции данного вида состоит из рыночного и нерыночного (диверсифицируемого) риска.

Рыночный риск — часть риска по акции или иному рискованному активу или портфелю, которая определяется ковариацией доходности этого актива (портфеля) и доходности рыночного портфеля и не может быть снижена в процессе диверсификации портфеля.

Нерыночный риск по акциям данного вида определяется особенностями, относящимся только к данной компании, включая обновление производства, новые контракты, слияния и приобретения, ошибки и просчеты в управлении бизнесом и т.п. С нерыночным риском по акциям или портфелю инвестор должен бороться путем диверсификации своих вложений.

Дисперсия доходности i-й акции:

$$\sigma_i^2 = \beta_i^2 \sigma_F^2 + \sigma_{\varepsilon i}^2,$$

где $\sigma_{\varepsilon i}^2$ — остаточная дисперсия доходности данной акции.

Общий риск σ_i^2 по данной акции, выраженный дисперсией ее доходности, раскладывается на две части:

- рыночный риск $\sigma_{рын.i}^2 = \beta_i^2 \sigma_F^2$;
- нерыночный риск $\sigma_{\varepsilon i}^2$.

Соответственно, доли этих частей в общем риске по акции составляют:

- доля рыночного риска:

$$R_i^2 = \frac{\sigma_{рын.i}^2}{\sigma_i^2},$$

при этом $\sigma_{рын.i}^2 = R_i^2 \sigma_i^2$;

• доля нерыночного риска

$$1-R_i^2=\frac{\sigma_{\varepsilon i}^2}{\sigma_i^2},$$

при этом $\sigma_{\varepsilon i}^2=(1-R_i^2)\sigma_i^2$.

Общий риск σ_i по акции в форме стандартного отклонения *формально* можно разложить:

• на рыночный риск $\sigma_{рын.i}=\beta_i\sigma_F$;

• нерыночный риск $\sigma_{\varepsilon i}$.

Соответственно, доли этих частей в общем риске по акции составляют:

• доля рыночного риска:

$$R_i=\sqrt{R_i^2}=\frac{\sigma_{рын.i}}{\sigma_i},$$

при этом $\sigma_{рын.i}=R_i\sigma_i$;

• доля нерыночного риска:

$$1-R_i=\frac{\sigma_{\varepsilon i}}{\sigma_i},$$

но величина нерыночного риска $\sigma_{\varepsilon i}=\sqrt{1-R_i^2}\,\sigma_i$, поскольку в этом случае выполняется равенство $\sigma_i^2=\sigma_{рын.i}^2+\sigma_{\varepsilon i}^2$.

Значимым является не весь риск, связанный с владением акциями, а только неизбежный, или рыночный, риск. Чем больше рыночный риск ценной бумаги, тем большей доходности от нее будут ждать инвесторы. Связь между ожидаемой доходностью и рыночным риском, а также вытекающее из нее определение стоимости ценных бумаг составляет сущность данной модели.

Коэффициент бета определяет влияние рынка на данные ценные бумаги, причем, если:

• $\beta_i>0$, то доходность бумаг данного вида колеблется в такт с рынком;

• $\beta_i < 0$, то поведение бумаги прямо противоположно колебаниям доходности рынка в целом.

Коэффициент бета каждой акции характеризует соотношение рыночного риска по акции в форме ковариации и риска рыночного портфеля в форме дисперсии. Таким образом, *коэффициент бета* — показатель рыночного риска. Этот риск нельзя устранить путем диверсификации, инвестируя в большее число акций, поскольку он зависит от изменений в экономической и политической ситуации, влияющих на весь фондовый рынок.

Коэффициент бета определяется для акций каждой компании, котировавшихся на фондовом рынке и приносивших доход в течение определенного периода.

По определению коэффициент бета рыночного портфеля равен единице. Если для акции некоторой компании:

• $\beta_i = 1$, то доходность этой акции в среднем совпадает с доходностью рыночного портфеля, или, с другой стороны, акции данной компании имеют среднюю степень риска, сложившуюся на рынке в целом (такие акции называются среднерисковыми);

• $\beta_i > 1$, то доходность этой акции растет в среднем быстрее, чем по рыночному портфелю, или, с другой стороны, акции данной компании более рискованные, чем в среднем на рынке (такие акции называются агрессивными). Такие акции следует иметь в своем портфеле, когда ожидается рост доходности рыночного портфеля. Они могут обеспечить инвестору более высокий уровень доходности, чем в среднем по рынку;

• $\beta_i < 1$, то доходность этой акции растет в среднем медленнее, чем по рыночному портфелю, или, с другой стороны, акции данной компании менее рискованные, чем в среднем на рынке (такие акции называются оборонительными). Такие акции следует иметь в своем портфеле, когда ожидается падение доходности рыночного портфеля, поскольку сокращение доходности по ним будет меньше, чем по рынку.

Для определения коэффициентов α, β, R^2 и σ_ε необходимо иметь ряды наблюдений доходности отдельных акций и рыночного портфеля в целом, причем конкретные оценки указанных коэффициентов существенно зависит от продолжительности периода наблюдения и его разбиения на подпериоды.

Пример 3.7. Рассматриваются две акции: акция 1, акция 2. Пусть известны индекс РТС и курсы акций на конец месяца.

Индекс РТС и курсы акций на конец месяца, $

Период наблюдения	Индекс РТС	Курс акции 1	Курс акции 2
Январь	55,12	0,071	0,027
Февраль	70,03	0,099	0,044
Март	80,36	0,129	0,046
Апрель	91,83	0,153	0,05
Май	97,64	0,151	0,057
Июнь	125,65	0,16	0,087
Июль	116,49	0,183	0,083
Август	102,5	0,154	0,069
Сентябрь	83,12	0,128	0,052
Октябрь	97,8	0,167	0,058
Ноябрь	112,38	0,189	0,073
Декабрь	177,74	0,304	0,113

При определении доходности будем учитывать только изменения их курса (без учета дивидендов).

Преобразуя данные таблицы, определяем доходности индекса РТС и акций обоих видов в течение указанного периода.

Фактическая доходность акций и индекса РТС, %

Период наблюдения	Индекс РТС	Курс акции 1	Курс акции 2
Февраль	27,05	39,44	62,96
Март	14,75	30,30	4,55
Апрель	14,27	18,60	8,70
Май	6,33	–1,31	14,00
Июнь	28,69	5,96	52,63
Июль	–7,29	14,38	–4,60
Август	–12,01	–15,85	–16,87
Сентябрь	–18,91	–16,88	–24,64

Период наблюдения	Индекс РТС	Курс акции 1	Курс акции 2
октябрь	17,66	30,47	11,54
Ноябрь	14,91	13,17	25,86
Декабрь	58,16	60,85	54,79
Средняя доходность	13,06	16,28	17,18
Риск (σ)	21,48	23,42	29,15

Преобразование данных первой таблицы в данные второй таблицы удобно производить с использованием электронных таблиц Excel, а расчеты средней доходности (среднее значение доходности за рассматриваемый период) и риска σ_i (корень квадратный из несмещенной выборочной дисперсии) — с помощью функций «СРЗНАЧ» и «СТАНДОТКЛОН» Excel.

При использовании Пакета анализа «Excel» в режиме Анализ данных (Регрессия), получим характеристические линии соответствующих акций:

$$m_1 = 4{,}17 + 0{,}93 m_F ,$$

$$m_2 = 1{,}60 + 1{,}19 m_F ,$$

где m_i, m_F – ожидаемая доходность i-й акции и рыночного портфеля.

При этом:

- акция 1: $\alpha_1 = 4{,}17$; $\beta_1 = 0{,}93$; $R_1 = 0{,}85$; $R_1^2 = 0{,}72$; $\sigma_{\varepsilon 1} = 12{,}96$;
- акция 2: $\alpha_2 = 1{,}60$; $\beta_2 = 1{,}19$; $R_2 = 0{,}88$; $R_2^2 = 0{,}77$; $\sigma_{\varepsilon 2} = 14{,}65$.

На рис. 3.4 представлены графики характеристических линий обеих акций.

Полученные уравнения можно использовать для *прогнозирования* ожидаемой доходности акций в зависимости от прогноза ожидаемой доходности по фондовому рынку.

Анализ полученных результатов

- Расчетные значения коэффициентов альфа

$$\alpha_1 = 4{,}17 > 1{,}60 = \alpha_2$$

показывают, что при нулевой доходности фондового рынка большая доходность достигается по акциям 1.

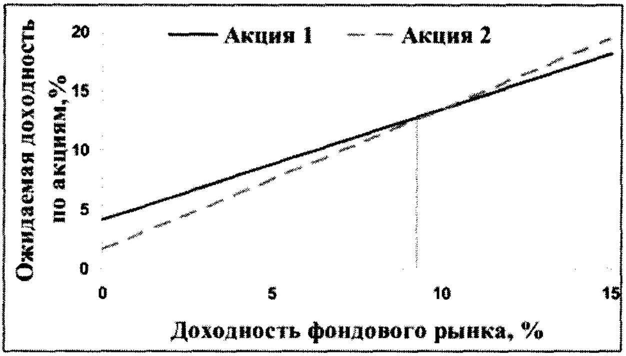

Рис. 3.4. Характеристические линии рассматриваемых акций

• Из сравнения значений коэффициентов бета для акций обоих видов

$$\beta_1 = 0{,}93 < 1, \beta_2 = 1{,}19 > 1$$

следует, что с ростом доходности фондового рынка доходность по акциям 2 будет возрастать быстрее, чем по акциям 1, а при падении доходности фондового рынка доходность по акциям 1 будет падать медленнее, чем по акциям 2.

• Коэффициенты $R_1 = 0{,}85$, и $R_2 = 0{,}88$ показывают долю рыночного риска в общем риске по соответствующим акциям в форме стандартного отклонения, а доля нерыночного риска: $1 - R_1 = 1 - 0{,}85 = 0{,}15$ и $1 - R_2 = 1 - 0{,}88 = 0{,}12$.

• Коэффициенты $\sigma_1 = 23{,}42$ и $\sigma_2 = 29{,}15$ определяют общий риск по соответствующим акциям в форме стандартного отклонения его доходности, которые можно разложить:

• на рыночный риск: $\sigma_{\text{рын.1}} = R_1\sigma_1 = 19{,}98$ и $\sigma_{\text{рын.2}} = R_2\sigma_2 = 25{,}56$;

• нерыночный риск: $\sigma_{\varepsilon1} = 12{,}96$; $\sigma_{\varepsilon2} = 14{,}65$, при этом соблюдается равенство $\sigma_i^2 = \sigma_{\text{рын.}i}^2 + \sigma_{\varepsilon1}^2$.

• Из графика (см. рис. 3.4) следует, что с ростом доходности фондового рынка ожидаемая доходность акций обоих видов возрастает. При относительно небольшой доходности фондового рынка большую ожидаемую доходность обеспечивают акции 1. В точке пересечения обеих прямых ожидаемые доходности по акциям обоих видов совпадают. При дальнейшем увеличении ожидаемой доходности фондового рынка большую доходность обеспечивают акции 2, для которых коэффициент бета больше единицы.

Анализ полученных данных показывает, что при ожидаемом уменьшении доходности рыночного портфеля (доходности на соответствующий биржевой индекс) целесообразнее иметь в портфеле акции с коэффициентом бета меньше единицы, а при прогнозируемом увеличении доходности рыночного портфеля — акции с коэффициентом бета больше единицы.

Информация о значениях коэффициентов α, β, R^2 и σ_ε для различных ценных бумаг, определенные подобным образом в зависимости от выбранного рыночного портфеля и установленного периода наблюдения, регулярно публикуются в специальных бюллетенях.

3.6.3. Модель оценки доходности финансовых активов (CAPM)[1]

Рыночный риск в рамках модели CAPM измеряется с помощью β-коэффициентов. Уравнение

$$m_i = r_б + \beta_i (m_F - r_б),$$

где $\beta_i = \sigma_{iF} / \sigma_F^2$, называется *рыночной линией* ценной бумаги. Оно показывает, что ожидаемая доходность по каждому рассматриваемому рисковому активу зависит от соотношения рыночного риска данного актива и риска рыночного портфеля в форме дисперсии.

Модель CAPM определяет доходности тех ценных бумаг, которые покупаются и продаются на идеальном рынке. Реальные ценные бумаги могут отклоняться от прямой, отвечающей модели идеального конкурентного рынка. Соответствующие этим отклонениям невязки $\varepsilon_i = r_i - m_i$ между фактическими значениями и модельными оценками могут отличаться от нуля.

Если $r_i = m_i$, то $\varepsilon_i = 0$. Такие ценные бумаги называются справедливо оцененными. Те же бумаги, у которых $\varepsilon_i > 0$, рынком недо-

[1] CAPM — Capital Asset Pricing Model.

оценены, а если $\varepsilon_i < 0$, то рынком переоценены. В частности, одна из задач финансового анализа состоит в нахождении недооцененных рынком бумаг и в рекомендации инвестору приобретать их.

Пример 3.8. Известны доходности (%) рискового актива r_i и рыночного индекса r_F на фондовой бирже за последние десять месяцев, а также безрисковая ставка доходности $r_б = 15\%$.

Месяц	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
r_i	17	18	19	17	16	20	23	22	17	16
r_F	16	17	17	18	18	19	21	20	18	16

Запишем уравнения характеристической линии и линии рынка ценных бумаг. Сделаем выводы.

▼ Исходные данные и расчетные показатели представим в виде расчетной таблицы.

n	r	r_F	$(r_F - 18)^2$	$(r - 18{,}5)(r_F - 18)$
1	17	16	4	3
2	18	17	1	0,5
3	19	17	1	–0,5
4	17	18	0	0
5	16	18	0	0
6	20	19	1	1,5
7	23	21	9	13,5
8	22	20	4	7
9	17	18	0	0
10	16	16	4	5
Итого	**185**	**180**	**24**	**30**
Среднее	18,5	18	2,4	3
	$\bar{r}_i$	m_F	σ_{iF}	σ_{iF}

Имеем

$\bar{r}_i = 18{,}5$ — средняя (фактическая) доходность акции;

$m_F = 18$ — средняя (ожидаемая) доходность рыночного индекса;

$\sigma_F^2 = 2{,}4$; $\sigma_{iF} = 3$; $\beta_i = \sigma_{iF}/\sigma_F^2 = 1{,}25$; $\alpha_i = \bar{r}_i - \beta_i m_F = 18{,}5 - 1{,}25 \cdot 18 = -4$.

Уравнение характеристической линией ценной бумаги:

$$m_i = \alpha_i + \beta_i m_F = -4 + 1{,}25 m_F.$$

Поскольку β_i=1,25 > 1, то бумага является *агрессивной*, ее доходность растет быстрее, чем доходность рыночного индекса.

Уравнение линии рынка ценных бумаг:

$$m_i = r_б + (m_F - r_б)\beta_j = 15 + (18 - 15)\cdot\beta_j = 15 + 3\beta_j.$$

Ожидаемое значение доходности данной ценной бумаги m_i = 15 + + 3·1,25 = 18,75. Поскольку $\varepsilon_i = \bar{r}_i - m_i$ = 18,5 – 18,75 = –0,25 < 0, то эта бумага переоценена рынком.

Линия рынка капитала. Важным свойством модели САРМ является ее линейность относительно степени риска.

Уравнение рыночной линии ценной бумаги для портфеля инвестора запишется как

$$m_p = r_б + \beta_p(m_F - r_б),$$

где $\beta_p = \Sigma\beta_i x_i$ — значение β-коэффициента в портфеле;

β_i — значение β-коэффициентов i-го актива в портфеле;

x_i — доля i-го актива в портфеле;

n — число различных финансовых активов в портфеле.

Для диверсифицируемого портфеля инвестора $\beta_p = \sigma_p/\sigma_F$. Подставляя β_p в уравнение, получим *уравнение линии капитала*:

$$m_p = r_б + \frac{(m_F - r_б)}{\sigma_F}\sigma_p,$$

где m_p, σ_p — ожидаемая доходность и риск эффективного портфеля;

m_F, σ_F — ожидаемая доходность и риск рыночного портфеля;

$r_б$ — доходность безрисковых ценных бумаг.

Уравнение линии капитала, которое связывает ожидаемую доходность и риск индивидуального портфеля инвестора, представляет лучшие из возможных для всех инвесторов комбинации «риск – доходность».

Исходное положение САРМ: при равновесии на рынке ценных бумаг рыночный портфель F как совокупность всех обращающихся на рынке рисковых активов совпадает с оптимальным эффективным для инвесторов касательным портфелем K.

Если для всех участников рынка имеется единая безрисковая ставка получения и предоставления займов $r_б$, то в состоянии равновесия оптимальным будет являться портфель, представляющий собой комбинацию рыночного портфеля и безрискового актива.

Это означает, что рыночный и касательные портфели имеют одинаковую структуру, в частности: $m_F = m_K$, $\sigma_M = \sigma_K$. В зависимости от своей меры неприятия риска инвесторы обладают различными наборами безрисковых и рисковых активов, однако процентное соотношение рисковых ценных бумаг в портфелях инвесторов оказывается для всех них одинаковым. При этом ожидаемая доходность и риск каждого оптимального индивидуального портфеля располагаются на одной прямой, которая содержит множество всех эффективных индивидуальных портфелей в рамках модели САРМ.

Очевидно, что рыночный портфель также принадлежит этому множеству портфелей, при этом $m_p = m_F$, $\sigma_p = \sigma_F$ и значение $\beta_p = 1$.

График линии капиталов представляет прямую линию (рис. 3.5).

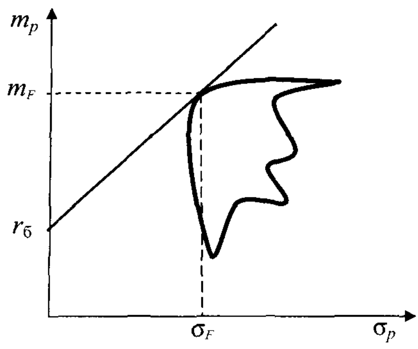

Рис. 3.5. Линия рынка капитала

Коэффициент при риске портфеля в уравнении линии капитала представляет собой тангенс угла наклона линии рынка капитала и называется *рыночной ценой риска*:

$$p_F = \frac{m_F - r_б}{\sigma_F}.$$

Рыночная цена риска показывает прирост ($m_F - r_б$) ожидаемой доходности рыночного портфеля над безрисковой ставкой процента (*премия за риск*) в среднем по рынку, приходящийся на один процент риска σ_F рыночного портфеля.

В этом случае уравнение линии рынка капитала можно записать так:

$$m_p = r_б + p_F \sigma_p .$$

Ни один инвестор, не склонный к риску, не будет осуществлять рискованные инвестиции, если его премия за риск отрицательна.

Линию рынка капитала можно использовать для сравнительного анализа портфельных инвестиций.

Пример 3.9. Ожидаемая доходность и риск рыночного портфеля: $m_F = 30\%$, $\sigma_F = 12\%$, а безрисковая ставка процента $r_б = 6\%$. Тогда рыночная цена риска $p_F = (30 - 6)/12 = 2$, и уравнение линии рынка капитала

$$m_p = 6 + 2\sigma_p.$$

Пусть в портфель включаются три рискованных актива с бета-коэффициентами $\beta_1 = 0,4$; $\beta_2 = 0,6$; $\beta_3 = 1,4$.

Риски и ожидаемые доходности каждого актива составляют:

$$\sigma_1 = \beta_1 \sigma_F = 0,4 \cdot 12 = 4,8; \; \sigma_2 = 7,2; \; \sigma_3 = 16,8;$$

$$m_1 = 6 + 2 \cdot 4,8 = 15,6; \; m_2 = 20,4; \; m_3 = 39,6.$$

Ожидаемые доходы и значения β различных активов зафиксированы с точки зрения портфельного менеджера, потому что они определяются рынком. Однако доходы и величина β портфеля могут формироваться посредством подбора доли каждого из активов в портфеле.

Пусть доли x_i каждого из активов портфеля: 0,10; 0,25; 0,65. Тогда инвестиционный портфель имеет $\beta_p = 0,10 \cdot 0,4 + 0,25 \cdot 0,6 + 0,65 \cdot 1,4 = 1,1$; риск $\sigma_p = 1,1 \cdot 12 = 13,2$; ожидаемую доходность $m_p = 32,4$.

4. ПОНЯТИЕ О ВРЕМЕННЫХ РЯДАХ

Временной ряд — это совокупность значений какого-либо показателя за несколько последовательных моментов времени.

Каждый уровень $y(t)$ временного ряда формируется под совместным влиянием длительных, кратковременных и случайных факторов.

Длительные, постоянно действующие, факторы оказывают на изучаемое явление определяющее влияние и формируют основную тенденцию ряда – тренд $T(t)$.

Кратковременные, периодические, факторы формируют сезонные колебания ряда $S(t)$.

Случайные факторы отражаются случайными изменениями уровней ряда — $\varepsilon(t)$.

Модель, в которой временной ряд представлен как сумма перечисленных компонентов, т. е. $y(t)=T(t)+S(t)+\varepsilon(t)$, называется *аддитивной*.

Модель, в которой временной ряд представлен как произведение перечисленных компонентов, т. е. $y(t) = T(t)S(t)\varepsilon(t)$, называется *мультипликативной*.

Выбор одной из двух моделей осуществляется на основе анализа структуры сезонных колебаний:

• если амплитуда сезонных колебаний приближенно постоянная, то используют аддитивную модель;

• если амплитуда возрастает или уменьшается, то используют мультипликативную модель.

Основная задача эконометрического исследования временного ряда — выявление каждой из перечисленных компонент ряда.

4.1. Моделирование основной тенденции развития

Основной тенденцией развития (трендом) называется плавное и устойчивое изменение уровня явления во времени, свободное от случайных колебаний.

Методами выявления тренда являются:

- метод скользящей средней;
- аналитическое выравнивание.

Метод скользящей средней сводится к замене фактических уровней временного ряда расчетными уровнями, которые в меньшей степени подвержены колебаниям. Процедура метода заключается в том, что рассматривается средний уровень из определенного числа первых по счету уровней ряда, затем – из такого же числа уровней, но начиная со второго по счету и т. д. Средняя как бы «скользит» по ряду динамики, продвигаясь на один срок.

Пример 4.1. Рассчитаем скользящую среднюю (трех- и четырехлетние) по данным об урожайности зерновых культур (ц/га) за 10 лет.

▼Исходные данные и расчетные показатели представим в таблице.

Номер года	Фактический уровень	Скользящая средняя		Центрированная скользящая средняя
		Трехлетняя	Четырехлетняя	
1	15			
2	13	14,33	14,75	
3	15	14,67	15,50	15,125
4	16	16,33	16,50	16
5	18	17,00	16,75	16,625
6	17	17,00	17,50	17,125
7	16	17,33	17,25	17,375
8	19	17,33	18	17,625
9	17	18,67		
10	20			

Период скольжения может быть четным и нечетным.

Для нечетного периода (трехлетнего) первое значение скользящей средней есть (15 + 13 + 15)/3 = 14,33; второе — (13 + 15 + 16)/3 = 14,67 и т. д., причем полученные результаты скользящей средней отнесены к середине периода скольжения.

Для четного периода (четырехлетнего) первое значение скользящей средней есть (15 + 13 + 15 + 16)/4 = 14,75; второе — (13 + 15 + + 16 + 18)/4 = 15,50 и т. д. Однако рассчитанные усредненные значения нельзя сопоставить каким-либо определенным значениям t, поэтому применяют процедуру центрирования (вычисляют среднее значение из двух последовательных скользящих средних). Первое значение центрированной скользящей средней есть (14,75 + 15,50)/2 = = 15,125, и результат будет отнесена к третьему году. Второе — (15,50 + 16,50)/2 = 16 и т. д. ▲

Метод аналитического выравнивания заключается в том, что фактические уровни y_t ряда заменяются плавно изменяющимися теоретическими уровнями $\hat{y}_t$. Основная тенденция развития процесса (тренд) рассчитывается как функция времени $\hat{y}_t = f(t)$.

Для построения тренда чаще всего используют линейную функцию:

$$\hat{y}_t = a + bt .$$

Коэффициенты a, b определяют обычным МНК, используя в качестве независимой переменной время $t = 1,2,\ldots,n$, а в качестве зависимой переменной – фактические уровни временного ряда y_t.

В настоящее время компьютерные программы анализа временных рядов содержат широкий набор математических функций для построения уравнений тренда (линейный, логарифмический, полиномиальный, степенной, экспоненциальный, скользящее среднее).

Модели тенденции можно сравнивать по величине коэффициента детерминации R^2. Чем больше R^2, тем в большей мере уравнение тренда подходит для описания тенденции временного ряда.

Пример 4.2. Используя инструмент «Подбор линии тренда» из Мастера диаграмм Excel, по данным о розничном товарообороте y региона (усл. ед.) за 10 лет:

Номер года	1	2	3	4	5	6	7	8	9	10
Товарооборот	11	13	22	18,5	20	19	25	23	24,5	35

рассмотреть различные трендовые модели товарооборота.

▼Исходные данные (t, y) представляются в Excel.

Для построения линии тренда необходимо выделить временной ряд и выбрать в контекстном меню команду **Добавить линию трен-**

да. Будет вызвано диалоговое окно **Линия тренда**, содержащее вкладку **Тип** (рис. 4.1), на которой задается тип тренда:

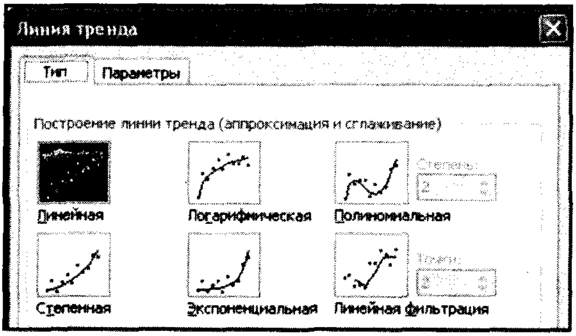

Рис. 4.1

Вкладка **Параметры** (рис. 4.2) предназначена для задания параметров тренда.

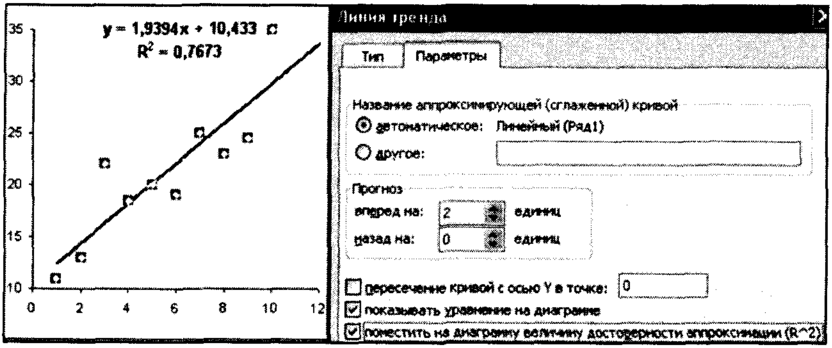

Рис. 4.2

Задавая различные типы тренда, получаем следующие трендовые модели товарооборота:

Вид уравнения	**Уравнение**	R^2
Линейное	$Y = 1{,}94 \cdot t + 10{,}43$	0,767
Полином 2-го порядка	$Y = 0{,}074 \cdot t^2 + 1{,}127 \cdot t + 12{,}06$	0,774
Полином 3-го порядка	$Y = 0{,}121 \cdot t^3 - 1{,}921 \cdot t^2 + 10{,}33 \cdot t + 1{,}68$	0,886

Вид уравнения	Уравнение	R^2
Логарифмическое	$Y = 7{,}70 \cdot Ln(t) + 9{,}46$	0,709
Степенное	$Y = 10{,}88 \cdot t^{0{,}407}$	0,815
Экспоненциальное	$Y = 11{,}84 \cdot e^{0{,}096 \cdot t}$	0,781

Уравнение тренда хорошо описывает тенденцию, если отсутствует автокорреляция в остатках $e_t = y_t - \hat{y}_t$, т. е. остатки текущего периода не коррелирует с остатками предыдущего периода.

Измерить автокорреляцию в остатках можно с помощью *коэффициента автокорреляции остатков*:

$$r = \frac{cov(e_t, e_{t-1})}{\sqrt{var(e_t)\,var(e_{t-1})}}.$$

Чтобы судить об отсутствии (наличии) автокорреляции остатков, используют критерий Дарбина-Уотсона. В дальнейшем (тема 5) будет подробно рассмотрен этот критерий. Если установлено наличие автокорреляции остатков, то уравнение тренда не является наилучшим, так как нарушена предпосылка МНК об отсутствии автокорреляции остатков.

4.2. Моделирование сезонных колебаний

Общий вид аддитивной модели: $Y = T + S + E$, а мультипликативной модели: $Y = T \cdot S \cdot E$, при этом для аддитивной модели характеристики сезонности и случайной составляющей измеряются в абсолютных величинах, а в мультипликативной — в относительных.

Построение моделей включает в себя следующие шаги:

- выравнивание исходного ряда методом скользящей средней;
- расчет значений сезонной компоненты S;
- устранение сезонной компоненты из исходных уровней ряда ($Y - S$) в аддитивной или (Y/S) в мультипликативной модели и получение выравненных данных ($T + E$) в аддитивной или ($T \cdot E$) в мультипликативной модели;
- аналитическое выравнивание уровней ($T + E$) или ($T \cdot E$) и расчет значений T с использованием полученного уравнения тренда;

• расчет полученных по модели значений $(T + S)$ или $(T \cdot E)$;

• расчет абсолютных или относительных ошибок, позволяющих оценить качество построенной модели.

Аддитивная модель временного ряда

Пример 4.3. Имеются поквартальные данные (усл. ед.) об объеме потребления электроэнергии **y** в некотором районе за четыре года.

Квартал	Год			
	1	2	3	4
I	60	72	80	90
II	44	48	56	66
III	50	60	64	70
IY	90	100	110	108

Построить аддитивную модель временного ряда. Дать прогнозное потребление электроэнергии в течение четырех кварталов ближайшего следующего года.

▼В качестве зависимой переменной при анализе временного ряда выступают фактические уровни ряда y_t, а в качестве независимой переменной — время (номер квартала) $t = 1, 2,\ldots,16$..

По графику ряда можно установить наличие приблизительно линейного тренда и сезонных колебаний (период равен 4) одинаковой амплитуды, поэтому используется аддитивная модель. Произведем декомпозицию ряда (разложение на компоненты).

Для исключения влияния сезонной компоненты произведем выравнивание исходного ряда методом скользящей средней за 4 квартала и процедуру центрирования. Результаты расчетов представлены в таблице.

Номер квартала	Потребление электроэнергии y_t	Скользящая средняя за 4 квартала	Центрированная скользящая средняя	Оценка сезонной компоненты
1	60			
2	44	61		
3	50	64	62,5	-12,5

Номер квартала	Потребление электроэнергии y_t	Скользящая средняя за 4 квартала	Центрированная скользящая средняя	Оценка сезонной компоненты
4	90	65	64,5	25,5
5	72	67,5	66,25	5,75
6	48	70	68,75	–20,75
7	60	72	71	–11
8	100	74	73	27
9	80	75	74,5	5,5
10	56	77,5	76,25	–20,25
11	64	80	78,75	–14,75
12	110	82,5	81,25	28,75
13	90	84	83,25	6,75
14	66	83,5	83,75	–17,75
15	70			
16	108			

Оценки сезонной компоненты определяются как разность между фактическими уровнями ряда y_t и центрированными скользящими средними.

Расчет сезонной компоненты произведем в следующей расчетной таблице, в которой оценки сезонной вариации записываются под соответствующим номером квартала в году.

Показатели	Год	Номер квартала в году				
		I	II	III	IY	
	1			–12,50	25,50	
	2	5,75	–20,75	–11,00	27,00	
	3	5,5	–20,25	–14,75	28,75	
	4	6,75	–17,75			
Итого		18,00	–58,75	–38,25	81,25	**Сумма**
Среднее		6,00	–19,58	–12,75	27,08	0,75
Скорректированное S_i		5,81	–19,77	–12,94	26,90	0

В строке **Среднее** рассчитаны средние сезонной вариации по годам за каждый квартал и их сумма, равная 0,75.

В аддитивной модели предполагается, что сумма всех сезонных компонент по всем кварталам должна быть равна нулю (условие взаимопогашаемости сезонных воздействий).

В строке **Скорректированное S_i** рассчитаны значения сезонных компонент S_i как разность между средней сезонной вариацией и корректирующим коэффициентом 0,75/4, при этом $\Sigma S_i = 0$.

Таким образом, получены следующие значения сезонной компоненты:

I квартал: $S_1 = 5,81$;

II квартал: $S_2 = -19,77$;

III квартал: $S_3 = -12,94$;

IV квартал: $S_4 = 26,90$.

Расчет трендовой компоненты и ошибок произведем в следующей таблице:

T	***Y***	***S***	***T + E =Y-S***	***T***	***Yp=T + S***	***E=Y- Yp***	***e =100·|E|/Y***
1	60	5,81	54,19	59,018	64,831	-4,831	8,05
2	44	-19,77	63,77	60,883	41,112	2,888	6,56
3	50	-12,94	62,94	62,747	49,809	0,191	0,38
4	90	26,90	63,10	64,611	91,507	-1,507	1,67
5	72	5,81	66,19	66,475	72,288	-0,288	0,40
6	48	-19,77	67,77	68,339	48,569	-0,569	1,18
7	60	-12,94	72,94	70,204	57,266	2,734	4,56
8	100	26,90	73,10	72,068	98,964	1,036	1,04
9	80	5,81	74,19	73,932	79,745	0,255	0,32
10	56	-19,77	75,77	75,796	56,025	-0,025	0,05
11	64	-12,94	76,94	77,661	64,723	-0,723	1,13
12	110	26,90	83,10	79,525	106,421	3,579	3,25
13	90	5,81	84,19	81,389	87,201	2,799	3,11
14	66	-19,77	85,77	83,253	63,482	2,518	3,81
15	70	-12,94	82,94	85,117	72,180	-2,180	3,11
16	108	26,90	81,10	86,982	113,877	-5,877	5,44
17		5,81		88,846	**94,658**		
18		-19,77		90,710	**70,939**	$\bar{e} =$	**2,75**
19		-12,94		92,574	**79,637**		
20		26,90		94,438	**121,334**		

В столбце «$T + E = Y - S$» исключается влияние сезонной компоненты: вычитая ее значение из каждого уровня исходного ряда, получим только тенденцию и случайную компоненту.

Проводя аналитическое выравнивание ряда ($T + E$) с помощью линейного тренда, получим следующее уравнение линии тренда:

$$T = 57{,}154 + 1{,}864 \cdot t.$$

Уровни ряда T для каждого $t = 1,2,\ldots,16$ приведены в таблице.

Расчетное значение (подгонка) уровня $Y_p = T + S$ временного ряда в аддитивной модели есть сумма трендовой и сезонной компонент.

Рассчитываются ошибка $E = Y - Y_p$, абсолютная относительная ошибки e,% и средняя абсолютная относительная ошибка $\bar{e} = 2{,}75\%$.

Прогнозные потребления электроэнергии в течение четырех кварталов ближайшего следующего года (17–20 кварталы) соответственно: 94,66; 70,94; 79,64; 121,34.

На рис. 4.3 показаны наблюдаемые и расчетные значения временного ряда.

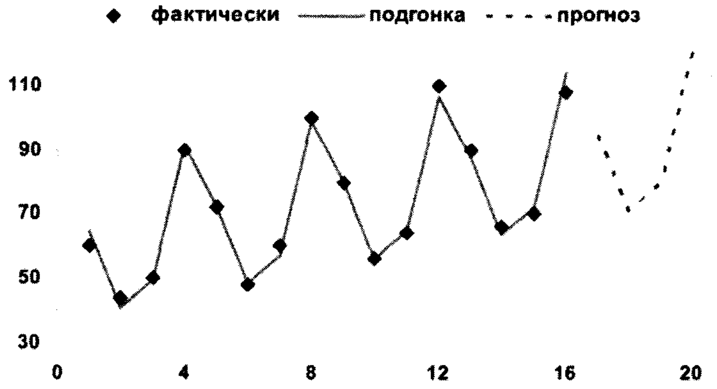

Рис. 4.3

Рассмотрим упрощенный метод разложения временного ряда (без сглаживания).

Пример 4.4. Решим пример 4.3 упрощенным метолом.

▼ Исходные данные и результаты расчета представлены в таблице.

Год	t	Y	T	S + E=Y-T	S	Y_p=T + S	$E=Y-Y_p$	Тест	e =100·|E|/Y
1	1	60	55,93	4,07	5,91	61,84	–1,84	60	3,07
	2	44	58,20	–14,20	–18,36	39,84	4,16	44	9,45
	3	50	60,48	–10,48	–13,14	47,34	2,66	50	5,32
	4	90	62,76	27,24	25,59	88,34	1,66	90	1,84
2	5	72	65,03	6,97	5,91	70,95	1,05	72	1,46
	6	48	67,31	–19,31	–18,36	48,95	–0,95	48	1,97
	7	60	69,59	–9,59	–13,14	56,45	3,55	60	5,92
	8	100	71,86	28,14	25,59	97,45	2,55	100	2,55
3	9	80	74,14	5,86	5,91	80,05	–0,05	80	0,07
	10	56	76,41	–20,41	–18,36	58,05	–2,05	56	3,67
	11	64	78,69	–14,69	–13,14	65,55	–1,55	64	2,43
	12	110	80,97	29,03	25,59	106,55	3,45	110	3,13
4	13	90	83,24	6,76	5,91	89,16	0,84	90	0,93
	14	66	85,52	–19,52	–18,36	67,16	–1,16	66	1,76
	15	70	87,80	–17,80	–13,14	74,66	–4,66	70	6,66
	16	108	90,07	17,93	25,59	115,66	–7,66	108	7,09
	17		92,35		5,91	**98,26**			
	18		94,63		–18,36	**76,26**		$\bar{e}$ =	**3,58**
	19		96,90		–13,14	**83,76**			
	20		99,18		25,59	**124,76**			

Производя аналитическое выравнивание ряда y_t с помощью линейной функции, получим следующее уравнение линии тренда: $T = 53,66 + 2,276 \cdot t$.

Уровни ряда T для каждого $t = 1,2,\ldots,16$ приведены в таблице.

Каждый столбец таблицы имеет обозначение и формулу расчета.

Столбец T — компонента тренда..

Столбец $S + E=Y - T$ — случайная величина сезонной компоненты.

Столбец S — сезонная компонента, рассчитанная как среднегодовая из значений $S + E$ каждого квартала.

Столбец $Y_p=T + S$ — расчетные значения временного ряда.

Столбец $E=Y- Y_p$ — случайная компонента.

Столбец Тест $=T + S + E$ — тест для проверки модели. В итоге значения Y и Test должны совпасть.

Столбец $e = 100\cdot|E|/Y$ — абсолютная величина относительной ошибки, %. Средняя абсолютная относительная ошибка подгонки $\bar{e} = 3,58\%$.

Прогнозное потребление электроэнергии в течение четырех кварталов ближайшего следующего года (17–20 кварталы) составляют соответственно значения 98,26; 76,26; 83,76; 124,76.

Сравним результаты расчета примеров 4.3 и 4.4. Установление значимости различий между ними произведем, используя программу «Парный двухвыборочный t-тест для средних» из Анализа данных Excel:

Парный двухвыборочный t-тест для средних		
	Метод	
	скользящее	**упрощенный**
Среднее	77,553	76,078
Дисперсия	535,488	486,098
Наблюдения	20	20
Df	19	
t-статистика	1,531	
$P(T \leq t)$ двухстороннее	0,142	
t критическое двухстороннее	2,093	

Поскольку $t = 1{,}53 < t_{кр} = 2{,}09 \Rightarrow$ нет разницы, или поскольку P-значение $= 0{,}142 > \alpha = 0{,}05 \Rightarrow$ нет разницы.

При рассмотрении помесячных данных упрощенный метод имеет преимущество во времени расчета.

Мультипликативная модель временного ряда

Пример 4.5. Имеются поквартальные данные (усл. ед.) о выплате доходов компании акционерам в форме дивидендов за последние четыре года.

Квартал	**Год**			
	1	**2**	**3**	**4**
I	40	60	50	30
II	50	80	70	50
III	60	100	80	60
IY	70	110	130	70

Построить мультипликативную модель временного ряда. Определить прогноз прибыли в течение ближайшего следующего года

▼По графику ряда можно установить наличие приблизительно линейного тренда и сезонных колебаний (период равен 4) возрастающей амплитуды, поэтому используется мультипликативная модель. Произведем декомпозицию ряда.

Для исключения влияния сезонной компоненты проведем выравнивание исходного ряда методом скользящей средней за 4 квартала и процедуру центрирования. Результаты расчетов представлены в таблице.

Номер квартала	**Размер дивидендов y_t**	**Скользящая средняя за 4 квартала**	**Центрированная скользящая средняя**	**Оценка сезонной компоненты**
1	40			
2	60	45		
3	50	47,5	46,25	1,081
4	30	52,5	50	0,600
5	50	57,5	55	0,909
6	80	62,5	60	1,333
7	70	65	63,75	1,098
8	50	70	67,5	0,741
9	60	72,5	71,25	0,842
10	100	75	73,75	1,356
11	80	77,5	76,25	1,049
12	60	80	78,75	0,762
13	70	92,5	86,25	0,811
14	110	95	93,75	1,173
15	130			
16	70			

Оценки сезонной компоненты для мультипликативной модели определяются как частное от деления фактических уровней ряда y_t на центрированные скользящие средние.

Расчет сезонной компоненты произведем в расчетной таблице, в которой оценки сезонной вариации записываются под соответствующим номером квартала в году.

Показатели	**Год**	**Номер квартала в году**				
		I	**II**	**III**	**IY**	
	1			1,081	0,600	
	2	0,909	1,333	1,098	0,741`	

Показатели	**Год**	**Номер квартала в году**				
		I	**II**	**III**	**IY**	
	3	0,842	1,356	1,049	0,762	
	4	0,811	1,173			
Итого		2,562	3,862	3,228	2,103	**Сумма**
Среднее		0,854	1,287	1,076	0,701	3,918
Скорректированное S_i		0,872	1,314	1,098	0,715	4

В строке **Среднее** рассчитаны средние сезонной вариации по годам за каждый квартал и их сумма, равная 3,918.

В мультипликативной модели предполагается, что сумма всех сезонных компонент по всем кварталам должна быть равна четырем — числу сезонов в году (условие взаимопогашаемости сезонных воздействий).

В строке **Скорректированное S_i** рассчитаны значения сезонных компонент S_i как произведение соответствующей средней сезонной вариации на корректирующий коэффициент 4 / 3,918 = 1,021, при этом $\Sigma S_i = 4$.

Получены следующие значения сезонной компоненты:

I квартал: S_1 =0,872;

II квартал: S_2 = 1,314;

III квартал: S_3 = 1,098;

IV квартал: S_4 = 0,715.

Расчет трендовой компоненты и ошибок произведем в следующей таблице:

t	Y	S	$T \cdot E = Y/S$	T	$Y_p = T \cdot S$	$E = Y - Y_p$	$e = 100 \cdot \lvert E \rvert / Y$
1	40	0,87	45,87	41,250	35,969	4,031	10,08
2	60	1,31	45,65	45,000	59,140	0,860	1,43
3	50	1,10	45,52	48,750	53,547	–3,547	7,09
4	30	0,72	41,93	52,500	37,559	-7,559	25,20
5	50	0,87	57,34	56,250	49,048	0,952	1,90
6	80	1,31	60,87	60,000	78,853	1,147	1,43
7	70	1,10	63,73	63,750	70,023	–0,023	0,03
8	50	0,72	69,89	67,500	48,290	1,710	3,42
9	60	0,87	68,81	71,250	62,128	–2,128	3,55
10	100	1,31	76,09	75,000	98,566	1,434	1,43

t	Y	S	$T \cdot E = Y/S$	T	$Y_p = T \cdot S$	$E = Y - Y_p$	$e = 100 \cdot \lvert E \rvert / Y$
11	80	1,10	72,83	78,750	86,499	–6,499	8,12
12	60	0,72	83,87	82,500	59,021	0,979	1,63
13	70	0,87	80,28	86,250	75,207	–5,207	7,44
14	110	1,31	83,70	90,000	118,280	–8,280	7,53
15	130	1,10	118,35	93,750	102,975	27,025	20,79
16	70	0,72	97,85	97,500	69,752	0,248	0,35
17		0,87		101,250	**88,287**		
18		1,31		105,000	**137,993**	$\bar{e} =$	**6,34**
19		1,10		108,750	**119,451**		
20		0,72		112,500	**80,483**		

Разделив каждый уровень исходного ряда на соответствующие значения сезонной компоненты (столбец $T \cdot E = Y/S$), исключим влияние сезонной компоненты и в результате получим только тенденцию и случайную компоненту.

Производя аналитическое выравнивание ряда ($T \cdot E$) с помощью линейного тренда, получим следующее уравнение линии тренда:

$$T = 37{,}5 + 3{,}75 \cdot t.$$

Уровни ряда T для каждого $t = 1,2,\ldots,16$ приведены в таблице.

Столбец $Y_p=T \cdot S$ — расчетные значения временного ряда.

Ошибка в мультипликативной модели определяется по формуле

$$E = Y/Y_p.$$

Но чтобы сравнивать мультипликативную модель с другими моделями временного ряда, расчет ошибки E произведем как $E = Y - Y_p$.

Средняя абсолютная относительная ошибка $\bar{e} = 6{,}34\%$.

Прогнозное значение уровня временного ряда в мультипликативной модели есть произведение трендовой и сезонной компонент.

Прогноз прибыли в течение ближайшего следующего года (17–20 кварталы) соответственно: 88,29; 137,99; 119,45; 80,48.

На рис. 4.4 показаны наблюдаемые и расчетные значения временного ряда.

Рис. 4.4

Пример 4.6. Решим пример 4.5 упрощенным метолом.

▼Исходные данные и результаты расчета представлены в таблице.

Год	*T*	*Y*	*T*	*S · E=Y/T*	*S*	$Y_p=T \cdot S$	$E=Y/Y_p$	*Тест*	$E=Y-Y_p$	*e =100·|E|/Y*
1	1	40	41,25	0,97	0,878	36,220	1,104	40,0	3,780	9,45
	2	60	45,00	1,33	1,305	58,750	1,021	60,0	1,250	2,08
	3	50	48,75	1,02	1,131	55,163	0,906	50,0	–5,163	10,33
	4	30	52,50	0,57	0,689	36,191	0,829	30,0	–6,191	20,64
2	5	50	56,25	0,88	0,878	49,392	1,012	50,0	0,608	1,22
	6	80	60,00	1,33	1,305	78,333	1,021	80,0	1,667	2,08
	7	70	63,75	1,09	1,131	72,137	0,970	70,0	–2,137	3,05
	8	50	67,50	0,74	0,689	46,531	1,075	50,0	3,469	6,94
3	9	60	71,25	0,84	0,878	62,563	0,959	60,0	–2,563	4,27
	10	100	75,00	1,33	1,305	97,917	1,021	100,0	2,083	2,08
	11	80	78,75	1,01	1,131	89,110	0,898	80,0	–9,110	11,39
	12	60	82,50	0,72	0,689	56,871	1,055	60,0	3,129	5,21
4	13	70	86,25	0,81	0,878	75,734	0,924	70,0	–5,734	8,19
	14	110	90,00	1,22	1,305	117,500	0,936	110,0	–7,500	6,82
	15	130	93,75	1,38	1,131	106,083	1,225	130,0	23,917	18,40
	16	70	97,50	0,71	0,689	67,211	1,041	70,0	2,789	3,98
	17		101,25		0,878	**88,905**				
	18		105,00		1,305	**137,083**			$\bar{\varepsilon} =$	**7,26**
	19		108,75		1,131	**123,057**				
	20		112,50		0,689	**77,552**				

Производя аналитическое выравнивание ряда y_t с помощью линейной функции, получим следующее уравнение линии тренда: $T = 37{,}5 + 3{,}75 \cdot t$.

Каждый столбец таблицы обозначается так же, как и в примере 4.5, но формулы расчета изменяются с учетом мультипликативности модели.

Чтобы сравнить мультипликативную модель с аддитивной, расчет ошибки E (предпоследний столбец таблицы) произведем по формуле: $E = Y - Y_p$.

Средняя абсолютная относительная ошибка подгонки $\bar{e}$ =7,26%.

Прогноз прибыли в течение ближайшего следующего года (17–20 кварталы) составляет соответственно значения 88,90; 137,08; 123,06; 77,55.

Сравнение результатов расчета примеров 4.5 и 4.6 показывает статистическую незначимость их различий.

Упражнение 4.1. По данным продаж продовольственных товаров в магазине (Y) за последние 11 кварталов разработать модель продаж и провести прогнозирование объема продаж на следующие два квартала.

Исходные данные (усл. ед.)

Квартал	1	2	3	4	5	6	7	8	9	10	11
Объем продаж	4	6	4	5	10	8	7	9	12	14	15

Указание. По графику ряда можно установить наличие приблизительно линейного тренда и сезонных колебаний (период равен 4) одинаковой амплитуды, поэтому используется *аддитивная* модель.

Ответ. Объем продаж в 12 и 13 кварталах $y(12) = 14{,}48$; $y(13) = 18{,}43$.

Упражнение 4.2. По данным продаж продовольственных товаров в магазине (Y) за последние 11 кварталов разработать модель продаж и провести прогнозирование объема продаж на следующие два квартала.

Исходные данные (усл. ед.)

Квартал	1	2	3	4	5	6	7	8	9	10	11
Объем продаж	63	74	79	120	67	79	88	140	60	82	90

Указание. По графику ряда можно установить наличие приблизительно линейного тренда и сезонных колебаний (период равен 4) возрастающей амплитуды, поэтому используется *мультипликативная* модель.

Ответ. Объем продаж в 12 и 13 кварталах $y(12) = 139{,}3$; $y(13) = 68{,}0$.

4.3. Адаптивное прогнозирование

Адаптивные модели прогнозирования — это модели дисконтирования данных, способные быстро приспосабливать свою структуру и параметры к изменению условий. Инструментом прогноза в адаптивных моделях является математическая модель с единственным фактором «время».

В основе адаптивных методов находится модель экспоненциального сглаживания.

Модель экспоненциального сглаживания описывается формулой:

$$S_t = \lambda y_t + (1 - \lambda)S_{t-1},$$

где S_t — значение экспоненциальной средней в момент t;

λ — параметр сглаживания, $0 < \lambda < 1$.

Когда эта формула применяется рекурсивно, то каждое новое сглаженное значение (которое является также прогнозом) вычисляется как взвешенное среднее текущего наблюдения и сглаженного ряда. Очевидно, результат сглаживания зависит от параметра λ. Если $\lambda = 1$, то предыдущие наблюдения полностью игнорируются. Если $\lambda = 0$, то игнорируются текущие наблюдения. Значения λ между 0 и 1 дают промежуточные результаты.

При использовании экспоненциальной средней для краткосрочного прогнозирования предполагается, что модель ряда (модель Брауна) имеет вид

$$y_t = a_t + \varepsilon_t,$$

где a_t – варьирующий во времени средний уровень ряда;

ε_t – случайные отклонения с нулевым математическим ожиданием и дисперсией σ^2.

Прогнозная модель определяется равенством

$$\hat{y}_\tau(t) = \hat{a}_t,$$

где $\hat{y}_\tau(t)$ — прогноз, сделанный в момент t на τ единиц времени вперед;

$\hat{a}_t$ — оценка a_t.

Единственный параметр модели $\hat{a}_t$ определяется экспоненциальной средней $\hat{a}_t = S_t$, $\hat{a}_0 = S_0$.

Начальное значение S_0 вычисляется как среднее всех наблюдений.

В качестве критерия оптимальности при выборе параметра сглаживания λ обычно принимают критерий минимума $\bar{\varepsilon}$ (средней абсолютной величины относительной ошибки).

Модель Брауна дает достаточно точный прогноз при отсутствии тренда и сезонности.

При прогнозировании временного ряда с линейной тенденцией рассмотренная модель будет давать смещенные прогнозы. Для таких временных рядов используются модели линейного роста, также применяющие процедуру экспоненциального сглаживания.

В этих моделях прогноз производится с помощью выражения

$$\hat{y}_\tau(t) = \hat{a}_t + \hat{b}_t\tau,$$

где $\hat{a}_t, \hat{b}_t$ — текущие оценки коэффициентов;

τ – время упреждения прогнозов.

Этот метод называется также двухпараметрическим методом Хольта.

Оценки коэффициентов в этой модели определяются с помощью выражений

$$\hat{a}_t = \lambda_1 y_t + (1-\lambda_1)(\hat{a}_{t-1} + \hat{b}_{t-1});$$
$$\hat{b}_t = \lambda_2(\hat{a}_t - \hat{a}_{t-1}) + (1-\lambda_2)\hat{b}_{t-1},$$

где $0 < \lambda_1, \lambda_2 < 1$.

В качестве начальных значений $\hat{a}_0, \hat{b}_0$ на практике берут оценки МНК коэффициентов линейного тренда, определенные по исходному временному ряду или его части.

Модель Хольта дает достаточно точный прогноз при наличии тренда и отсутствии периодических сезонных колебаний. Этот метод часто используется при прогнозировании экономических показателей торгового процесса с ярко выраженной тенденцией на повышение или понижение.

Многие экономические временные ряды содержат периодические сезонные колебания. В зависимости от характера этих колебаний их подразделяют на аддитивные и мультипликативные.

При *аддитивном* характере сезонности исходят из предположения о неизменности во времени, примерном постоянстве амплитуды периодических колебаний, а при *мультипликативном* — амплитуда колебаний изменяется во времени. При этом для аддитивной модели характеристики сезонности будут измеряться в абсолютных величинах и отражаться в модели в виде слагаемых, а для мультипликативной — в относительных и представляться в модели в виде сомножителей.

На практике наиболее часто используются две тренд-сезонные модели линейного роста: модель Тейла-Вейджа с аддитивной сезонностью и модель Хольта-Уинтерса с мультипликативной сезонностью.

Модель Тейла-Вейджа с аддитивной сезонностью (с - количество периодов в сезонном цикле)	
Оценка $\hat{a}_t, \hat{b}_t$	$\hat{a}_t = \lambda_1(y_t - \hat{w}_{t-c}) + (1-\lambda_1)(\hat{a}_{t-1} + \hat{b}_{t-1})$ $\hat{b}_t = \lambda_2(\hat{a}_t - \hat{a}_{t-1}) + (1-\lambda_2)\hat{b}_{t-1}$
Оценка сезонности	$\hat{w}_t = \lambda_3(y_t - \hat{a}_t) + (1-\lambda_3)\hat{w}_{t-c}$
Модель прогноза	$\hat{y}_\tau(t) = \hat{a}_t + \hat{b}_t\tau + \hat{w}_{t-c+\tau}$
Модель Хольта-Уинтерса с мультипликативной сезонностью	
Оценка $\hat{a}_t, \hat{b}_t$	$\hat{a}_t = \lambda_1 y_t / w_{t-c} + (1-\lambda_1)(\hat{a}_{t-1} + \hat{b}_{t-1})$ $\hat{b}_t = \lambda_2(\hat{a}_t - \hat{a}_{t-1}) + (1-\lambda_2)\hat{b}_{t-1}$
Оценка сезонности	$\hat{w}_t = \lambda_3 y_t / \hat{a}_t + (1-\lambda_3)\hat{w}_{t-c}$
Модель прогноза	$\hat{y}_\tau(t) = (\hat{a}_t + \hat{b}_t\tau)\hat{w}_{t-c+\tau}$

Пример 4.7. По данным примера 4.3 об объеме потребления электроэнергии построим модель Тейла-Вейджа с аддитивной сезонностью и сделаем прогноз на четыре квартала вперед.

▼Исходные данные (t, y_t) и расчетные показатели ($\hat{y}_t$, $\hat{a}_t$, $\hat{b}_t$, $\hat{w}_t$), полученные по приведенным выше рекуррентным формулам, представим в следующей таблице.

t	y_t	$\hat{y}_1(t)$	$\hat{a}_t$	$\hat{b}_t$	$\hat{w}_t$	e_t
0			53,65	2,28		
1	60	61,45	55,93	2,28	4,58	2,41
2	44	41,45	58,20	2,28	–15,09	5,80
3	50	50,45	60,48	2,28	–10,32	0,89
4	90	90,45	62,75	2,28	27,40	0,49
5	72	69,61	65,03	2,28	6,14	3,32
6	48	52,21	67,31	2,28	–17,84	8,78
7	60	59,26	69,58	2,28	–9,84	1,23
8	100	99,26	71,86	2,28	27,88	0,74
9	80	80,27	74,13	2,28	5,96	0,34
10	56	58,57	76,41	2,28	–19,51	4,59
11	64	68,85	78,69	2,28	–13,00	7,57
12	110	108,85	80,96	2,28	28,64	1,05
13	90	89,20	83,24	2,28	6,48	0,89
14	66	66,00	85,51	2,28	–19,51	0,00
15	70	74,79	87,79	2,28	–16,12	6,85
16	108	118,70	90,07	2,28	21,66	9,91
17		**98,82**	λ_1	λ_2	λ_3	$\bar{e}$
18		**75,10**	0	0,238	0,652	**3,430**
19		**80,77**				
20		**120,83**	Модель Тейла-Вейджа			

Начальные значения $\hat{a}_0 = 53,65$ и $\hat{b}_0 = 2,28$ получены из оценок МНК коэффициентов линейного тренда, определяемых по исходным данным.

Ошибка подгонки $\bar{e}$=3,43%. Результаты расчета по модели Тейла-Вейджа представлены на рис. 4.5, из которого видно, что значения подгонки хорошо воспроизводят сезонные колебания.

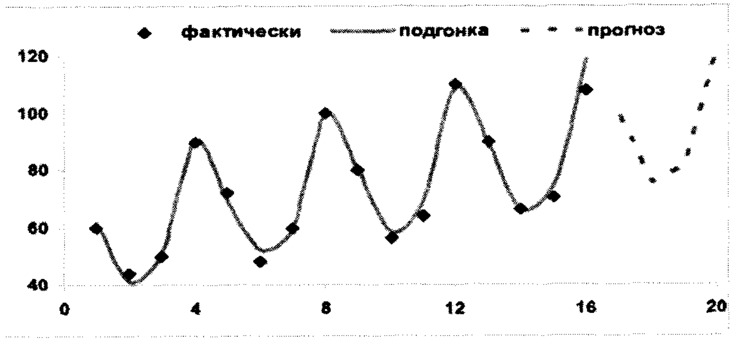

Рис. 4.5

Сравните полученные результаты прогноза по модели Тейла-Вейджа можно с результатами примера 4.3.

Пример 4.8. По данным пассажирских авиаперевозок y_t одной из авиакомпаний (тыс. чел.) за 2007–2012 гг. сделаем прогноз на следующий 2013 год.

Исходные данные

Месяц	Год					
	2007	**2008**	**2009**	**2010**	**2011**	**2012**
Январь	272	238	285	325	367	407
Февраль	224	190	272	300	329	357
Март	284	252	300	316	378	416
Апрель	272	252	319	349	410	452
Май	292	290	363	371	441	485
Июнь	330	347	399	433	520	520
Июль	385	409	487	541	590	590
Август	441	476	521	563	600	617
Сентябрь	352	390	402	465	510	521
Октябрь	289	332	351	402	442	460
Ноябрь	237	308	303	350	390	420
Декабрь	251	300	329	378	400	433

▼В качестве зависимой переменной при анализе временного ряда выступают фактические уровни ряда y_t, а в качестве независимой переменной — время (номер месяца) $t = 1, 2,\ldots,72$.

График ряда имеет вид (рис. 4.6)

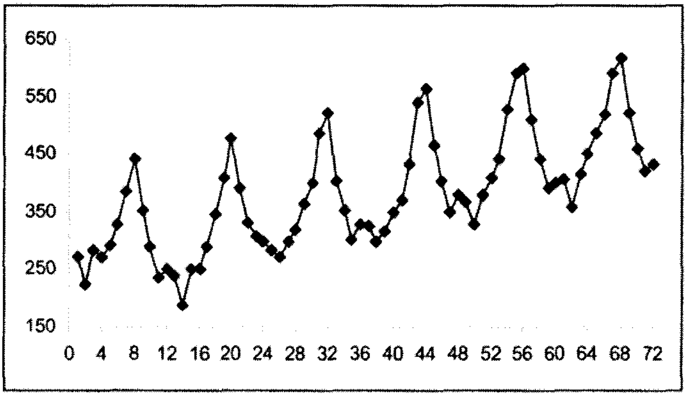

Рис. 4.6

По графику ряда можно установить наличие приблизительно линейного тренда и сезонных колебаний (период равен 12) одинаковой амплитуды, поэтому используется аддитивная модель (Тейла-Вейджа).

Исходные данные и расчетные показатели представим в следующей таблице.

t	y_t	$\hat{y}_t$	$\hat{a}_t$	$\hat{b}_t$	$\hat{w}_t$	e_t
0			266,01	3,21		
1	272	235,71	291,60	3,21	–19,60	13,34
2	224	210,09	303,39	3,21	–79,39	6,21
3	284	279,67	309,28	3,21	–25,28	1,53
4	272	276,34	309,81	3,21	–37,81	1,60
5	292	302,66	306,45	3,21	–14,45	3,65
6	330	343,59	301,29	3,21	28,71	4,12
7	385	393,71	299,13	3,21	85,87	2,26
8	441	449,84	296,90	3,21	144,10	2,00
9	352	356,89	297,10	3,21	54,90	1,39

t	y_t	$\hat{y}_t$	$\hat{a}_t$	$\hat{b}_t$	$\hat{w}_t$	e_t
10	289	293,37	297,61	3,21	–8,61	1,51
11	237	252,68	291,16	3,21	–54,16	6,61
12	251	246,01	297,45	3,21	–46,45	1,99
13	238	281,07	274,11	3,21	–36,11	18,10
14	190	197,93	272,43	3,21	–82,43	4,18
15	252	250,37	276,65	3,21	–24,65	0,65
16	252	242,05	286,00	3,21	–34,00	3,95
17	290	274,76	298,61	3,21	–8,61	5,25
18	347	330,54	311,98	3,21	35,02	4,74
19	409	401,06	320,09	3,21	88,91	1,94
20	476	467,41	328,60	3,21	147,40	1,81
21	390	386,72	333,84	3,21	56,16	0,84
22	332	328,44	339,25	3,21	–7,25	1,07
23	308	288,30	354,61	3,21	–46,61	6,40
24	300	311,37	350,81	3,21	–50,81	3,79
25	285	317,92	333,73	3,21	–48,73	11,55
26	272	254,51	347,73	3,21	–75,73	6,43
27	300	326,29	334,73	3,21	–34,73	8,76
28	319	303,94	347,23	3,21	–28,23	4,72
29	363	341,83	363,50	3,21	–0,50	5,83
30	399	401,73	365,03	3,21	33,97	0,69
31	487	457,15	386,65	3,21	100,35	6,13
32	521	537,26	379,84	3,21	141,16	3,12
33	402	439,21	360,10	3,21	41,90	9,26
34	351	356,07	360,19	3,21	–9,19	1,44
35	303	316,80	354,90	3,21	–51,90	4,55
36	329	307,30	371,49	3,21	–42,49	6,60
37	325	325,98	374,10	3,21	–49,10	0,30
38	300	301,59	376,34	3,21	–76,34	0,53
39	316	344,82	361,78	3,21	–45,78	9,12
40	349	336,76	372,54	3,21	–23,54	3,51
41	371	375,26	373,13	3,21	–2,13	1,15
42	433	410,32	390,33	3,21	42,67	5,24

t	y_t	$\hat{y}_t$	$\hat{a}_t$	$\hat{b}_t$	$\hat{w}_t$	e_t
43	541	493,90	422,59	3,21	118,41	8,71
44	563	566,97	423,36	3,21	139,64	0,70
45	465	468,47	424,43	3,21	40,57	0,75
46	402	418,45	417,50	3,21	–15,50	4,09
47	350	368,82	409,11	3,21	–59,11	5,38
48	378	369,83	417,36	3,21	–39,36	2,16
49	367	371,47	417,82	3,21	–50,82	1,22
50	329	344,70	411,35	3,21	–82,35	4,77
51	378	368,79	420,25	3,21	–42,25	2,44
52	410	399,92	429,68	3,21	–19,68	2,46
53	441	430,76	439,20	3,21	1,80	2,32
54	528	485,09	468,88	3,21	59,12	8,13
55	590	590,50	471,78	3,21	118,22	0,09
56	600	614,64	465,97	3,21	134,03	2,44
57	510	509,75	469,34	3,21	40,66	0,05
58	442	457,05	463,27	3,21	–21,27	3,41
59	390	407,37	455,77	3,21	–65,77	4,45
60	400	419,62	446,89	3,21	–46,89	4,91
61	407	399,28	454,86	3,21	–47,86	1,90
62	357	375,72	446,53	3,21	–89,53	5,24
63	416	407,50	454,99	3,21	–38,99	2,04
64	452	438,53	466,51	3,21	–14,51	2,98
65	485	471,52	478,04	3,21	6,96	2,78
66	520	540,37	468,69	3,21	51,31	3,92
67	590	590,12	471,83	3,21	118,17	0,02
68	617	609,07	479,93	3,21	137,07	1,28
69	521	523,81	481,41	3,21	39,59	0,54
70	460	463,36	482,56	3,21	–22,56	0,73
71	420	420,00	485,77	3,21	–65,77	0,00
73	433	442,10	483,37	3,21	–50,37	2,10
73		**438,73**	λ_1	λ_2	λ_3	$\bar{e}$
74		**400,27**	0,61665087	**0**	1,000	**3,748**

t	y_t	$\hat{y}_t$	$\hat{a}_t$	$\hat{b}_t$	$\hat{w}_t$	e_t
75		**454,03**				
76		**481,72**	Модель Тейла-Вейджа			
77		**506,41**				
78		**553,97**				
79		**624,05**				
80		**646,16**				
81		**551,89**				
82		**492,96**				
83		**452,96**				
84		**471,57**				

Начальные значения $\hat{a}_0$=266,01 и $\hat{b}_0$=3,21 получены из оценок МНК коэффициентов линейного тренда, определяемых по всем наблюдениям исходных данных. Оптимальные значения параметров сглаживания λ_1= 0,617; λ_2= 0,0; λ_3= 1,0 определяются из минимума средней абсолютной относительной ошибки $\bar{e}$ с помощью программы «Поиск решения», при этом $\bar{e}$ = 3,75, что свидетельствует о хорошем качестве прогнозной модели.

Результаты расчета по модели представлены на рис. 4.7, из которого видно, что прогнозные значения хорошо воспроизводят сезонные колебания. Данную модель можно считать пригодной для прогнозирования.

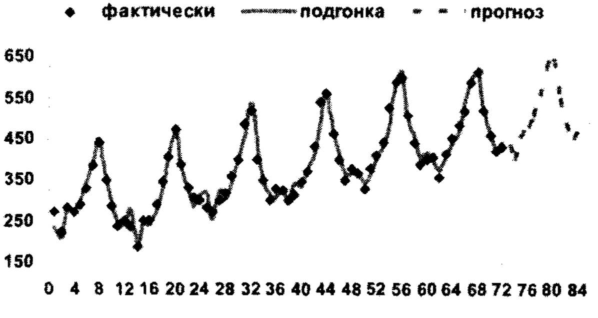

Рис. 4.7

Пример 4.9. По данным примера 4.5 о выплате дивидендов построим модель Хольта-Уинтерса с мультипликативной сезонностью и сделаем прогноз на четыре квартала вперед.

▼Исходные данные и расчетные показатели представим в следующей таблице.

t	y_t	$\hat{y}_1(t)$	$\hat{a}_t$	$\hat{b}_t$	$\hat{w}_t$	e_t
0			37,50	3,75		
1	40	38,33	41,25	3,75	0,96	4,17
2	60	60,00	45,00	3,75	1,33	0,00
3	50	51,76	48,75	3,75	1,04	3,53
4	30	34,44	52,50	3,75	0,60	14,81
5	50	53,85	56,25	3,75	0,91	7,70
6	80	80,00	60,00	3,75	1,33	0,00
7	70	66,09	63,75	3,75	1,08	5,58
8	50	40,32	67,50	3,75	0,70	19,36
9	60	64,83	71,25	3,75	0,86	8,04
10	100	100,00	75,00	3,75	1,33	0,00
11	80	84,99	78,75	3,75	1,04	6,24
12	60	57,49	82,50	3,75	0,72	4,19
13	70	74,42	86,25	3,75	0,83	6,32
14	110	120,00	90,00	3,75	1,26	9,09
15	130	97,06	93,75	3,75	1,28	25,34
16	70	70,00	97,50	3,75	0,72	0,00
17		**83,76**	λ_1	λ_2	λ_3	$\bar{e}$
18		**131,91**	**0**	**0**	**0,694**	**7,148**
19		**139,10**				
20		**80,77**	**Модель Хольта-Уинтерса**			

Начальные значения $\hat{a}_0 = 37{,}50$ и $\hat{b}_0 = 3{,}75$ получены из оценок МНК коэффициентов линейного тренда, определяемых по исходным данным.

Ошибка подгонки $\bar{e} = 7{,}15\%$. Результаты расчета по модели Хольта-Уинтерса представлены на рис. 4.8, из которого видно, что значения подгонки хорошо воспроизводят сезонные колебания.

Рис. 4.8

Полученные результаты прогноза по модели Хольта-Уинтерса сравните с результатами примера 4.5.

Экспоненциальное сглаживание — наиболее простой способ построения прогнозов, часто дающий быстрые эффективные результаты. Однако этот метод не позволяет строить доверительные интервалы и, следовательно, рассчитать риски при использовании прогнозов

4.4. Автокорреляция уровней временного ряда

При наличии во временном ряде тенденции и сезонных колебаний значения каждого последующего уровня ряда зависит от предыдущих.

Корреляционная зависимость между последовательными уровнями временного ряда называется *автокорреляцией* уровней ряда.

Для измерения автокорреляции уровней ряда используется коэффициент автокорреляции уровней.

Коэффициент автокорреляции порядка τ определяется как коэффициент корреляции между рядами y_t, $y_{t-\tau}$:

$$r_\tau = \frac{cov(y_t, y_{t-\tau})}{\sqrt{var(y_t)var(y_{t-\tau})}}.$$

Число периодов τ, по которым рассчитывается коэффициент автокорреляции, называется *лагом*.

Коэффициенты автокорреляции разных порядков принято обозначать как r_1, r_2,..., r_k, где 1, 2,..., k указывают на номер порядка коэффициента автокорреляции.

Пример 4.10. По данным примера 4.3 об объеме потребления электроэнергии (y) в некотором районе за четыре года определите коэффициенты автокорреляции до 4-го порядка.

▼Создадим следующую таблицу уровней временного ряда:

y_t	y_{t-1}	y_{t-2}	y_{t-3}	y_{t-4}
60				
44	60			
50	44	60		
90	50	44	60	
72	90	50	44	60
48	72	90	50	44
60	48	72	90	50
100	60	48	72	90
80	100	60	48	72
56	80	100	60	48
64	56	80	100	60
110	64	56	80	100
90	110	64	56	80
66	90	110	64	56
70	66	90	110	64
108	70	66	90	110

Последовательно вводя данные (y_t,y_{t-1}),(y_t,y_{t-2}),(y_t,y_{t-3}),(y_t,y_{t-4}), с помощью функции Excel: «Коррел» получим следующие коэффициенты автокорреляции: $r_1 = 0{,}165$; $r_2 = -0{,}566$; $r_3 = 0{,}113$; $r_4 = 0{,}983$. ▲

Последовательность коэффициентов автокорреляции первого, второго и более высоких порядков называется *автокорреляционной функцией (АКФ)* временного ряда. График зависимости ее значений от величины лага (порядка коэффициента автокорреляции) называется *коррелограммой.*

Автокорреляционную функцию обычно используют для выявления во временном ряде наличия или отсутствия трендовой и сезонной компонент:

• если наиболее высоким оказался коэффициент автокорреляции первого порядка, то исследуемый ряд содержит только тенденцию;

• если наиболее высоким оказался коэффициент автокорреляции порядка $\tau > 1$, то ряд содержит сезонные колебания с периодом τ;

• если ни один из коэффициентов автокорреляции не является значимым, то либо ряд не содержит тенденции и сезонных колебаний, либо ряд содержит сильную нелинейную тенденцию, для выявления которой нужно провести дополнительный анализ.

4.5. Учет тенденции при построении модели регрессии

При изучении развития явления во времени иногда возникает необходимость оценить степень взаимосвязи в изменениях уровней двух и более временных рядов различного содержания, но связанных один с другим.

При установлении связи между уровнями временных рядов может иметь место ложная корреляция — установление связи там, где ее на самом деле нет.

Наличие ложной корреляции связано с тенденцией каждого из рядов динамики, с автокорреляцией их уровней. Если ряды динамики характеризуются наличием тренда, то при построении модели регрессии надо исключить тренд, в противном случае корреляция уровней рядов динамики будет преувеличена.

При построении модели регрессии по временным рядам при наличии тенденции используются следующие методы:

• метод последовательных разностей;

• метод отклонений от тренда;

• метод включения в уравнение регрессии фактора времени.

Метод отклонений от тренда. Предположим, имеются два ряда динамики y_t, x_t, содержащие трендовую и случайную компоненты.

Проведение аналитического выравнивания по каждому из этих рядов позволяет найти параметры соответствующих уравнений трендов и определить их расчетные уровни $\hat{y}_t$, $\hat{x}_t$. Эти расчетные значения можно принять за оценку трендовой компоненты каждого ряда. Поэтому влияние тенденции можно устранить путем вычитания этих расчетных значений из фактических уровней ряда.

Дальнейший анализ взаимосвязи рядов проводят с использованием не исходных уровней, а рассчитанных отклонений от тренда $dy_t = y_t - \hat{y}_t$, $dx_t = x_t - \hat{x}_t$, при условии что последние не содержат тенденцию.

Метод последовательных разностей. Исключить влияние автокорреляции можно путем перехода от исходных уровней ряда y_t, x_t к первым разностям $\Delta y_t = y_t - y_{t-1}$, $\Delta x_t = x_t - x_{t-1}$ в случае линейного тренда, и ко вторым разностям — в случае параболического тренда. В случае экспоненциального и степенного трендов метод последовательных разностей применяется к логарифмам исходных данных.

Метод включения в уравнение регрессии фактора времени. Модель регрессии по временным рядам может быть построена с включением в нее как отдельной независимой переменной фактора времени t. Считается, что введение фактора времени исключает тенденцию развития всех явлений представленными исследуемыми рядами динамики.

Доказано, что если временные ряды характеризуются линейной тенденцией, то включение в модель фактора времени t равносильно построению модели регрессии по отклонениям от трендов с последующим переходом от нее к исходным уровням временного ряда зависимой переменной y.

Пример 4.11. Имеются данные о расходах на конечное потребление y_t и совокупном доходе x_t за 10 лет (усл. ед.). Построим модель регрессии и сделаем прогноз на 11-й год, используя методы исключения тенденции и включение фактора времени.

▼Исходные данные и расчетные показатели представим в виде расчетной таблицы.

t	y_t	x_t	$\hat{y}_t$	$\hat{x}_t$	dy_t	dx_t	Δy_t	Δx_t
1	6	24	6,073	23,00	–0,073	1,00		
2	10	29	8,945	27,40	1,055	1,60	4	5
3	12	33	11,818	31,80	0,182	1,20	2	4
4	14	35	14,691	36,20	–0,691	–1,20	2	2
5	17	39	17,564	40,60	–0,564	–1,60	3	4

t	y_t	x_t	$\hat{y}_t$	$\hat{x}_t$	dy_t	dx_t	Δy_t	Δx_t
6	21	45	20,436	45,00	0,564	0,00	4	6
7	23	48	23,309	49,40	–0,309	–1,40	2	3
8	25	50	26,182	53,80	–1,182	–3,80	2	2
9	28	55	29,055	58,20	–1,055	–3,20	3	5
10	34	70	31,927	62,60	2,073	7,40	6	15

Если к исходным данным y_t, x_t применить МНК, то получим уравнение регрессии:

$$\hat{y}_t = -8{,}14 + 0{,}634 x_t,\ R^2 = 0{,}983.$$

Полученные результаты содержат ложную корреляцию ввиду наличия в каждом из рядов y_t, x_t линейной тенденции.

Применим различные методы устранения тенденции.

1. **Метод отклонения от тренда**. По трендам

$$\hat{y}_t = 3{,}2 + 2{,}87t,\ \hat{x}_t = 18{,}6 + 4{,}4t$$

определим расчетные значения $\hat{y}_t$, $\hat{x}_t$ и отклонения от трендов dy_t, dx_t.

Применяя к рядам dy и dx МНК, получим уравнение линейной регрессии:

$$d\hat{y}_t = 0{,}30 \cdot dx_t,\ R^2 = 0{,}893.$$

Для прогноза значений y перейдем к уравнению, связывающему между собой уровни временных рядов:

$$y_p - \hat{y}_{t=p} = 0{,}30 \cdot (x_p - \hat{x}_{t=p}),$$

где y_p, x_p — прогнозные значения y, x;

$\hat{y}_{t=p}$, $\hat{x}_{t=p}$ — прогнозы y, x по тренду при $t = p$.

Для прогноза на 11-й год имеем:

$\hat{y}_{t=p} = 3{,}2 + 2{,}87 \cdot 11 = 34{,}77$; $\hat{x}_{t=p} = 18{,}6 + 4{,}4 \cdot 11 = 67$; $x_p = 71$ (принято представлять как $x_n = 70$ и плюс $\Delta n = 1$), тогда

$$y_p = 34{,}77 + 0{,}30 \cdot (71 - 67) = 35{,}95.$$

2. **Метод последовательных разностей.** Найдем первые разности Δy_t, Δx_t. Используя МНК, получим уравнение регрессии

$$\Delta y_t = 1{,}48 + 0{,}32 \cdot \Delta x_t,\ R^2 = 0{,}856.$$

Для прогноза значений y перейдем к уравнению, связывающему между собой уровни временных рядов:

Прогноз на 11-й год выполним по уравнению

$$y_p - y_n = 1{,}48 + 0{,}32 \cdot (x_p - x_n),$$

где $y_n = 34$, $x_n = 70$ — конечные уровни рядов.

По уравнению линейного тренда найдем $x_p = 18{,}6 + 4{,}4 \cdot 11 = 67$, тогда

$$y_p = 34 + 1{,}48 + 0{,}32 \cdot (67 - 70) = 34{,}52.$$

3. **Метод включения фактора времени.** Модель регрессии с включением в нее фактора времени имеет вид

$$\hat{y}_t = -2{,}4 + 0{,}30 \cdot x_t + 1{,}55 \cdot t,\ R^2 = 0{,}998.$$

Значение прогноза y на 11-й год составляет:

$$y_p = -2{,}4 + 0{,}30 \cdot 71 + 1{,}55 \cdot 11 = 35{,}95.$$

Результат прогноза совпал с прогнозом при использовании метода отклонения от тренда.

Полученные прогнозные уравнения хорошо учитывают тенденцию, если отсутствует автокорреляция в остатках $e_t = y_t - \hat{y}_t$, т. е. остатки текущего периода не коррелирует с остатками предыдущего периода. Чтобы судить об отсутствии (наличии) автокорреляции остатков, используют критерий Дарбина-Уотсона. В дальнейшем (тема 5) будет подробно рассмотрен этот критерий. В данных уравнениях отсутствует автокорреляция в остатках.

4.6. Учет сезонности при построении модели регрессии

Если в рядах динамики наблюдается тенденция и сезонная компонента одновременно, то в модель можно включить фактор времени и фиктивные переменные.

Пример 4.12. По данным примера 4.3 об объеме потребления электроэнергии рассмотрим построение модели временного ряда с использованием фактора времени и фиктивных переменных.

Возьмем I квартал каждого года в качестве эталонного, и будем использовать фиктивные переменные для оценки разности между ним и другими кварталами.

Запишем модель как

$$y = \alpha + \beta\, t + \delta_1 z_1 + \delta_2 z_2 + \delta_3 z_3 + \varepsilon,$$

где z_1, z_2, z_3 — фиктивные переменные, определяемые следующим образом:

$$z_1 = \begin{cases} 1 & (\text{II квартал}) \\ 0 & (\text{остальные}) \end{cases}, \quad z_2 = \begin{cases} 1 & (\text{III квартал}) \\ 0 & (\text{остальные}) \end{cases}, \quad z_3 = \begin{cases} 1 & (\text{IY квартал}) \\ 0 & (\text{остальные}) \end{cases}.$$

Коэффициенты δ_1, δ_2, δ_3 дают численную величину эффекта, вызываемого сменой года.

Исходные данные с фиктивными переменными представим таблицей

y	t	z_1	z_2	z_3
60	1	0	0	0
44	2	1	0	0
50	3	0	1	0
90	4	0	0	1
72	5	0	0	0
48	6	1	0	0
60	7	0	1	0
100	8	0	0	1
80	9	0	0	0
56	10	1	0	0
64	11	0	1	0
110	12	0	0	1
90	13	0	0	0
66	14	1	0	0
70	15	0	1	0
108	16	0	0	1

Оцененное уравнение регрессии:

$$\hat{y} = 62{,}37 + 1{,}87t - 23{,}87z_1 - 18{,}25z_2 + 20{,}87z_3,\ R^2 = 0{,}985,$$

где все коэффициенты при переменных статистически значимы.

Отдельные уравнения для каждого квартала таковы:

$\hat{y} = 62{,}37 + 1{,}87t$ (I квартал);

$\hat{y} = 38{,}50 + 1{,}87t$ (II квартал);

$\hat{y} = 44{,}12 + 1{,}87t$ (III квартал);

$\hat{y} = 83{,}24 + 1{,}87t$ (IY квартал).

Усредняя эти уравнения, получим линейный тренд:

$$T = \frac{62{,}37 + 38{,}50 + 44.12 + 83{,}24}{4} + 1{,}87t = 57{,}06 + 1{,}87t.$$

Расстояние между линией регрессии каждого квартала и трендом дает оценку сезонной компоненты в данном квартале:

$S_1 = 62{,}37 - 57{,}06 = 5{,}31$ (I квартал);

$S_2 = 38{,}50 - 57{,}06 = -18{,}56$ (II квартал);

$S_3 = 44{,}12 - 57{,}06 = -12{,}94$ (III квартал);

$S_4 = 83{,}24 - 57{,}06 = 26{,}18$ (IY квартал),

причем сумма сезонных отклонений должна равняться нулю, т. е.

$$S_1 + S_2 + S_3 + S_4 = 0.$$

Прогнозное значение уровня временного ряда в аддитивной модели есть сумма трендовой и сезонной компонент.

Прогнозное потребление электроэнергии в 17–20 кварталах:

$y(17) = T(17) + S_1 = 57{,}06 + 1{,}87 \cdot 17 + 5{,}31 = 94{,}16$;

$y(18) = T(18) + S_2 = 57{,}06 + 1{,}87 \cdot 18 - 18{,}56 = 72{,}16$;

$y(19) = T(19) + S_3 = 57{,}06 + 1{,}87 \cdot 19 - 12{,}94 = 79{,}65$;

$y(20) = T(20) + S_4 = 57{,}06 + 1{,}87 \cdot 20 + 26{,}18 = 120{,}64$.

Средняя абсолютная относительная ошибка $\bar{e} = 2{,}8\%$.

5. ГЕТЕРОСКЕДАСТИЧНОСТЬ И АВТОКОРРЕЛИРОВАННОСТЬ СЛУЧАЙНОГО ЧЛЕНА

5.1. Обнаружение гетероскедастичности

Одной из предпосылок регрессионного анализа является предположение о *постоянстве* дисперсии случайного члена для всех наблюдений (*гомоскедастичность*). Это значит, что для каждого значения объясняющей переменной случайные члены имеют одинаковые дисперсии. Если это условие не соблюдается, то имеет место *гетероскедастичность.*

При *отсутствии гетероскедастичности* коэффициенты регрессии имеют *наименьшую* дисперсию среди всех несмещенных оценок, являющихся линейными функциями от наблюдений y.

Если наблюдается *гетероскедастичность*, то МНК-оценки будут *неэффективными* (они не будут иметь наименьшую дисперсию по сравнению с другими оценками данного параметра).

Оценки стандартных ошибок коэффициентов регрессии вычисляются в предположении, что распределение случайного члена гомоскедастично; если это не так, то они неверны (занижены), а, следовательно, t-статистика завышена. Это может привести к статистически значимым коэффициентам регрессии, тогда как в действительности это неверно.

Проблема гетероскедастичности характерна для пространственных данных, полученных от неоднородных объектов. Например, если исследуется зависимость прибыли предприятий от размера основного фонда, то можно ожидать, что для больших предприятий колебание прибыли будет выше, чем для малых.

Рассмотрим ряд тестов для обнаружения гетероскедастичности.

I. Тест ранговой корреляции Спирмена. Выдвигается нулевая гипотеза об отсутствии гетероскедастичности случайного члена.

Предполагается, что дисперсия случайного члена будет либо увеличиваться, либо уменьшаться по мере увеличения x, и поэтому в регрессии, оцениваемой с помощью МНК, абсолютные величины остатков $|e|$ и значения x будут коррелированны.

Данные по x и остатки $|e|$ ранжируются по переменной x, и определяются их ранги.

Ранг — это порядковый номер значений переменной в ранжированном ряду.

Определяется *коэффициент ранговой корреляции Спирмена* по формуле

$$r = 1 - \frac{6\sum D_i^2}{n(n^2-1)},$$

где D_i – разность между рангами x и $|e|$.

Значимость r проверяется по t-тесту. В случае установления значимости коэффициента ранговой корреляции Спирмена делается вывод о наличии гетероскедастичности.

Пример 5.1. Имеются данные (усл. ед.) зависимости выпуска продукции обрабатывающей промышленности на душу населения y от валового внутреннего продукта x на душу населения в том же году для 17 стран:

№	y	x		№	y	x
1	18	3		10	100	24
2	27	6		11	63	25
3	18	7		12	130	26
4	45	9		13	135	27
5	55	13		14	60	28
6	68	15		15	70	35
7	51	18		16	80	37
8	84	21		17	180	44
9	85	22				

Построить регрессионную модель выпуска продукции от валового внутреннего продукта. Проверить наличие гетероскедастичности, используя тест Спирмена.

▼Исходные данные и расчетные показатели представим в виде расчетной таблицы.

y	x	Ранг	e	$\lvert e\rvert$	Ранг	D	D^2
18	3	1	–3,60	3,60	2	–1	1
27	6	2	–3,35	3,35	1	1	1
18	7	3	–15,27	15,27	9	–6	36
45	9	4	5,89	5,89	4	0	0
55	13	5	4,22	4,22	3	2	4
68	15	6	11,38	11,38	7	–1	1
51	18	7	–14,38	14,38	8	–1	1
84	21	8	9,87	9,87	6	2	4
85	22	9	7,95	7,95	5	4	16
100	24	10	17,11	17,11	10	0	0
63	25	11	–22,81	22,81	11	0	0
130	26	12	41,28	41,28	15	–3	9
135	27	13	43,36	43,36	16	–3	9
60	28	14	–34,56	34,56	12	2	4
70	35	15	–44,99	44,99	17	–2	4
80	37	16	–40,83	40,83	14	2	4
180	44	17	*38,74*	38,74	13	4	16
						Итого	**110**

Пусть модель описывается выражением $y = \alpha + \beta x + \varepsilon$.

Оцененное уравнение регрессии обычным МНК:

$$\hat{y} = 12{,}84 + 2{,}92\,x\,,\ \alpha = 12{,}84;\ \beta = 2{,}92.$$

Из графика остатков (рис. 5.1), видно, что с увеличением переменной x размах колебаний остатков e тоже возрастает, поэтому есть предположение о зависимости ошибки регрессии от независимой переменной (гетероскедастичность).

Коэффициент ранговой корреляции:

$$r = 1 - \frac{6\sum D_i^2}{n(n^2-1)} = 1 - \frac{6\cdot 110}{17\cdot 288} = 0{,}865\,.$$

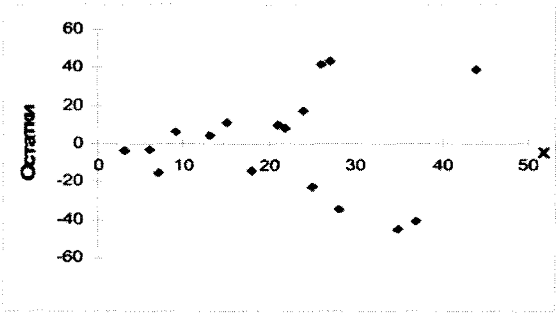

Рис. 5.1

Для проверки нулевой гипотезы используем t-тест.

Расчетное значение критерия:

$$t_p = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0{,}865\sqrt{17-2}}{\sqrt{1-0{,}865^2}} = 6{,}68.$$

Установим $\alpha = 0{,}05$. Число степени свободы $v = n - 2 = 15$.

Критическую точку находим с помощью функции Excel:

$$t_{кр} = \text{СТЬЮДРАСПОБР}(0{,}05;\ 15) = 2{,}13.$$

Поскольку $|t_p| = 6{,}68 > t_{кр} = 2{,}13 \Rightarrow$ гипотеза H_0 отклоняется, следовательно, имеет место гетероскедастичность. ▼

II. Тест Глейзера. Тест основывается на более общих представлениях о зависимости стандартной ошибки случайного члена от значений объясняющей переменной. Например, зависимость может быть представлена в виде $\sigma_i = \alpha + \beta x_i^{\gamma} + \varepsilon_i$.

Используя абсолютные значения остатков в качестве оценки σ_i, определяют данную регрессионную зависимость при различных значениях γ и выбирают наилучшую из них.

Таким образом, гетероскедастичность аппроксимируется уравнением

$$s_i = a + bx_i^{\gamma},$$

где $s_i = |e_i|$ — оценка σ_i.

Нулевая гипотеза об отсутствии гетероскедастичности отклоняется, если оценка b значимо отличается от нуля.

Пример 5.2. По данным примера 5.1 установить наличие гетероскедастичности, используя тест Глейзера.

▼ Оценив регрессию между $|e|$ и x, получим оцененное уравнение:

$$s = -3{,}13 + 1{,}15x,$$

где коэффициент b = 1,15 значим, следовательно, имеет место гетероскедастичность. ▼

III. Тест Голдфельда-Квандта. При проведении проверки по этому тесту предполагается, что стандартное отклонение σ случайного члена пропорционально значению независимой переменной x.

Тест включает следующие шаги:

- все n наблюдений в выборке упорядочиваются по возрастанию переменной x;
- оцениваются отдельные регрессии для первых n_0 и для последних n_0 наблюдений, а средние $(n - 2n_0)$ наблюдений отбрасываются;
- составляется статистика: $F = RSS_2/RSS_1$, где RSS_1, RSS_2 — суммы квадратов остатков для первых и последних n_0 наблюдений соответственно.

Если верна гипотеза H_0 об отсутствии гетероскедастичности, то F имеет распределение Фишера с $v_1 = n_0 - k - 1$, $v_2 = n_0 - k - 1$ степенями свободы, где k — число объясняющих переменных.

Определяется критическое значение критерия $F_{кр}$.

Если $F > F_{кр}$, то нулевая гипотеза об отсутствии гетероскедастичности отклоняется.

Замечание. Тест Голфелда–Квандта можно также использовать для проверки на гетероскедастичность при предположении, что σ_i обратно пропорционально x_i. В этом случае тестовой статистикой является величина $F = RSS_1/RSS_2$.

Пример 5.3. По данным примера 5.1 установить наличие гетероскедастичности, используя тест Голдфельда–Квандта.

▼ С помощью обычного МНК оценим регрессии для шести стран с наименьшими значениями показателя x и для шести стран с наибольшими значениями этого показателя.

Получим соответственно уравнения:

$$\hat{y}_1 = -0{,}18 + 4{,}38x;$$
$$\hat{y}_2 = 39{,}9 + 2{,}11x.$$

Суммы квадратов отклонений составляют RSS_1 = 224,4; RSS_2 = = 9804, при этом F = 9804/224,4 = 43,7. Число степеней свободы $\nu_1 = \nu_2 = 4$. Критическое значение $F_{кр}$ = 6,39 при 5%-ном уровне значимости.

Поскольку $F = 43{,}7 > F_{кр} = 6{,}39$, то нулевая гипотеза об отсутствии гетероскедастичности отклоняется. ▲

5.2. Обобщенный метод наименьших квадратов

При наличии гетероскедастичности используют *обобщенный (взвешенный) МНК.*

Суть метода заключается в уменьшении вклада данных наблюдений, имеющих большую дисперсию в результат расчета.

Пусть в исходной модели регрессии

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

случайные члены гетероскедастичны, т. е.

$$M(\varepsilon_i) = 0,\ D(\varepsilon_i) = \sigma_i^2,\ (i=1,n).$$

Допустим, что дисперсии σ_i^2 в каждом наблюдении известны.

Разделив каждое наблюдение на соответствующее ему значение σ_i, получим преобразованную модель:

$$\frac{y_i}{\sigma_i} = \alpha \cdot \frac{1}{\sigma_i} + \beta \cdot \frac{x_i}{\sigma_i} + \frac{\varepsilon_i}{\sigma_i},$$

для которой выполнены условия гомоскедастичности, так как

$$M\left(\frac{\varepsilon_i}{\sigma_i}\right) = \frac{M(\varepsilon_i)}{\sigma_i} = 0,\ D\left(\frac{\varepsilon_i}{\sigma_i}\right) = \frac{D(\varepsilon_i)}{\sigma_i^2} = \frac{\sigma_i^2}{\sigma_i^2} = 1,\ (i = 1,n).$$

Оценки параметров преобразованной модели обычным МНК дают непосредственно оценки параметров исходной модели.

На практике дисперсии σ_i^2 неизвестны, поэтому заменяют их оценками s_i^2, т. е. оценивается обычным МНК преобразованная модель

$$\frac{y}{s} = \alpha \cdot \frac{1}{s} + \beta \cdot \frac{x}{s} + \frac{\varepsilon}{s},$$

в которой отсутствует постоянный член.

Для экономических данных σ_i часто пропорциональны значениям объясняющей переменной x. При этом оценивается обычным МНК преобразованная модель:

$$\frac{y}{x} = \alpha \cdot \frac{1}{x} + \beta + \frac{\varepsilon}{x}.$$

Коэффициент при $1/x$ будет *эффективной оценкой* α, а постоянный член — *эффективной оценкой* β.

При применении взвешенного МНК оценки параметров будут несмещенными, кроме того, они имеют меньшую дисперсию, чем невзвешенные оценки.

Пример 5.4. В примере 5.1 обычным МНК получена регрессионная зависимость

$$\hat{y} = 12{,}84 + 2{,}92\,x,\ \alpha = 12{,}84;\ \beta = 2{,}92.$$

По этим же данным оценим регрессионную зависимость взвешенным МНК в двух вариантах:

а) с делением на s, где $s = -\,3{,}13 + 1{,}15x$;

б) с делением на x.

▼Исходные и преобразованные данные представлены в таблице.

Y	*X*	*s*	*y/s*	*y/x*	1/*s*	*x/s*	1/*x*
18	3	0,32	56,25	6,00	3,125	9,375	0,333
27	6	3,73	7,25	4,50	0,268	1,610	0,167
18	7	4,87	3,69	2,57	0,205	1,437	0,143
45	9	7,16	6,28	5,00	0,140	1,256	0,111
55	13	11,75	4,68	4,23	0,085	1,107	0,077
68	15	14,04	4,84	4,53	0,071	1,068	0,067
51	18	17,48	2,92	2,83	0,057	1,030	0,056
84	21	20,92	4,02	4,00	0,048	1,004	0,048

85	22	22,06	3,85	3,86	0,045	0,997	0,045
100	24	24,35	4,11	4,17	0,041	0,985	0,042
63	25	25,50	2,47	2,52	0,039	0,980	0,040
130	26	26,65	4,88	5,00	0,038	0,976	0,038
135	27	27,79	4,86	5,00	0,036	0,972	0,037
60	28	28,94	2,07	2,14	0,035	0,968	0,036
70	35	36,96	1,89	2,00	0,027	0,947	0,029
80	37	39,25	2,04	2,16	0,025	0,943	0,027
180	44	47,27	3,81	4,09	0,021	0,931	0,023

Регрессионные зависимости, полученные взвешенным МНК в двух вариантах, следующие:

а) $\dfrac{\hat{y}}{s} = 8{,}64 \cdot \dfrac{1}{s} + 3{,}11 \cdot \dfrac{x}{s}$; $\alpha = 8{,}64$; $\beta = 3{,}11$;

б) $\dfrac{\hat{y}}{x} = 7{,}78 \cdot \dfrac{1}{x} + 3{,}20$; $\alpha = 7{,}78$; $\beta = 3{,}20$,

где вариант «а» — с делением на s, а вариант «б» — с делением на x.

Полученные регрессионные зависимости не обнаруживают гетероскедастичности и дают более эффективные оценки коэффициентов регрессии, чем с использованием обычного МНК ▲.

Упражнение 5.1. Оценить регрессионную зависимость поступления доходов y в бюджет некоторого региона от численности x работающих на крупных и средних предприятиях экономики районов. Исходные данные (усл. ед.):

№	*y*	*x*	№	*y*	*x*
1	2,3	3	11	11,4	79
2	5	6	12	263,4	95
3	16	8	13	193,6	106
4	42,3	18	14	127	112
5	37	20	15	322,8	115
6	31,4	23	16	281,5	125
7	3,2	39	17	85,5	132
8	104,3	49	18	172	149
9	149,7	60	19	469,9	157
10	117,6	74	20	15,2	180

Ответ. $\dfrac{\hat{y}}{x} = 1{,}74 - 2{,}75 \cdot \dfrac{1}{x}$, $\alpha = -2{,}75$, $\beta = 1{,}74$.

5.3. Обнаружение автокорреляции

Одной из предпосылок регрессионного анализа является *независимость* случайного члена в любом наблюдении от его значений во всех других наблюдениях, т. е. $M(\varepsilon_i \varepsilon_j) = 0$, $(i \neq j)$.

Если данное условие не выполняется, то говорят, что случайный член подвержен автокорреляции. В этом случае коэффициенты регрессии, получаемые по МНК, оказываются *неэффективными*, хотя и несмещенными, а их стандартные ошибки рассчитываются некорректно (занижаются).

Причиной автокорреляции может быть либо неверная спецификация модели, либо наличие неучтенных факторов. Устранение этих причин не всегда дает нужные результаты. Автокорреляция имеет собственные внутренние причины, связанные с автокорреляционной зависимостью.

Автокорреляция обычно встречается в регрессионном анализе при использовании данных временных рядов. В силу этого в дальнейшем вместо символа i (порядковый номер наблюдения) будем использовать символ t (момента наблюдения).

Рассмотрим модель

$$y_t = \alpha + \beta x_t + \varepsilon_t,$$

где ε — случайный член.

Необходимым условием независимости случайных членов является их некоррелированность для каждых двух соседних значений.

Пусть ρ — коэффициент корреляции между двумя случайными соседними членами ε_t и ε_{t-1}, причем:

- если $\rho > 0$, то автокорреляция *положительная*;
- если $\rho < 0$, то автокорреляция *отрицательная*;
- если $\rho = 0$, то автокорреляция *отсутствует*.

Поскольку значения случайных членов ε_t, ε_{t-1} неизвестны, то проверяется статистическая некоррелированность остатков e_t, e_{t-1} с использование обычного МНК.

Соответствующей оценкой коэффициента корреляции ρ является **коэффициент автокорреляции остатков первого порядка**, который при достаточно большом числе наблюдений имеет вид

$$r = \frac{cov(e_t, e_{t-1})}{\sqrt{var(e_t) var(e_{t-1})}} \approx \frac{\sum e_t \, e_{t-1}}{\sum e_t^2}.$$

Выдвигается нулевая гипотеза об отсутствии корреляции первого порядка $H_0 : \rho = 0$. В качестве альтернативной гипотезы может выступать либо $H_1 : \rho > 0$, либо $H_2 : \rho < 0$.

Для проверки нулевой гипотезы используют **статистику Дарбина–Уотсона**, рассчитываемую по формуле

$$DW = \frac{\sum (e_t - e_{t-1})^2}{\sum e_t^2} \simeq 2(1 - r), \quad 0 \le DW \le 4.$$

По таблице определяются критические значения критерия Дарбина–Уотсона d_1 и d_2 для заданного числа наблюдений, числа объясняющих переменных и уровня значимости. По этим значениям отрезок [0; 4] разбивается на пять зон (рис. 5.2).

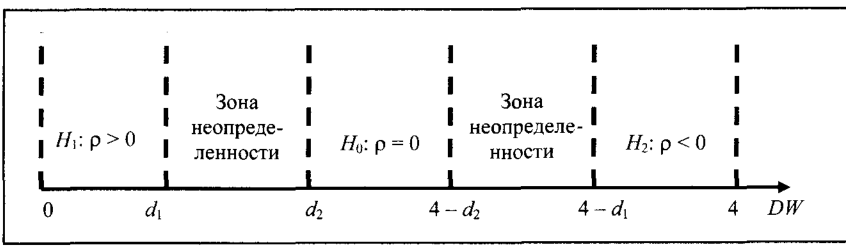

Рис. 5.2

В зависимости от того, в какую зону попадает расчетное значение критерия DW, принимают или отвергают соответствующую гипотезу.

Замечание. Тест Дарбина–Уотсона построен в предположении, что объясняющие переменные некоррелированы со случайным членом. Поэтому этот тест не применим к моделям, включающим в качестве объясняющих переменных лаговые значения зависимой переменной y.

Пример 5.5. Имеются данные об объеме предложения y товара, его цены x_1 и зарплаты сотрудников x_2 за 10 месяцев. Выявить на уровне значимости 0,05 наличие автокорреляции остатков в модели регрессии

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \varepsilon.$$

▼Исходные данные (усл. ед.) и результаты промежуточных расчетов представлены в таблице.

t	y	x_1	x_2	e_t	e_{t-1}
1	20	10	12	8,30	
2	35	15	10	4,26	8,30
3	30	20	9	–12,46	4,26
4	45	25	9	–1,86	–12,46
5	60	40	8	–7,38	–1,86
6	70	37	8	5,26	–7,38
7	75	43	6	–9,66	5,26
8	90	35	4	–2,26	–9,66
9	105	40	4	8,34	–2,26
10	110	55	5	7,46	8,34

Выборочная регрессия для этой модели:

$$\hat{y} = 90{,}72 + 0{,}88x_1 - 7{,}32x_2,$$

где все коэффициенты уравнения статистически значимы.

Оценивая корреляцию между остатками e_t, e_{t-1} (функция «Коррел»), получим коэффициент автокорреляции остатков первого порядка $r = -0{,}0252$. Значение критерия $DW = 2(1 - r) = 2{,}05$.

По таблице распределения Дарбина–Уотсона находим $d_1 = 0{,}70$ и $d_2 = 1{,}64$.

Поскольку $d_2 < DW < 4 - d_2$, то нет оснований отклонять гипотезу H_0 об отсутствии автокорреляции в остатках. ▲

5.4. Авторегрессионное преобразование

Пусть исходное уравнение регрессии

$$y_t = \alpha + \beta x_t + \varepsilon_t$$

содержит автокорреляцию случайных членов.

Допустим, что автокорреляция подчиняется авторегрессионной схеме первого порядка:

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t ,$$

где ρ — коэффициент авторегрессии;

u_t — случайный член, удовлетворяющий предпосылкам МНК.

Коэффициент авторегрессии ρ есть в точности коэффициент корреляции между двумя соседними ошибками ε_t, ε_{t-1}.

Применение обычного МНК в этом случае дает несмещенные, состоятельные, но неэффективные оценки.

Пусть ρ известен. Обратимся к исходной модели.

Для момента времени $t-1$ эта модель примет вид:

$$y_{t-1} = \alpha + \beta x_{t-1} + \varepsilon_{t-1} .$$

Вычтем из обеих частей исходного уравнения последнее уравнение, умноженное на ρ, получим:

$$y_t - \rho y_{t-1} = \alpha(1-\rho) + \beta(x_t - \rho x_{t-1}) + (\varepsilon_t - \rho\varepsilon_{t-1}) .$$

Обозначим:

$$\begin{cases} y'_t = y_t - \rho y_{t-1} \\ x'_t = x_t - \rho x_{t-1} \\ \alpha' = \alpha(1-\rho) \end{cases} .$$

Это преобразование переменных называется *авторегрессионным преобразованием Бокса–Дженкинса.*

Тогда получим преобразованное уравнение:

$$y'_t = \alpha' + \beta x'_t + u_t ,$$

где $t \geq 2$, не содержит автокорреляции, и для оценки ее параметров α', β используется обычный МНК.

Оценка коэффициента β из этой зависимости непосредственно используется и для исходного уравнения, а коэффициент α рассчитывается по формуле $\alpha = \alpha'/(1-\rho)$.

Первое наблюдение в преобразованном уравнении можно сохранить, если положить:

$$y_1' = \sqrt{1-\rho^2}\, y_1;$$

$$x_1' = \sqrt{1-\rho^2}\, x_1.$$

Обычно первое наблюдение опускают, поэтому преобразование исходной модели становится единообразным.

На практике величина ρ неизвестна, ее оценка получается одновременно с оценками α, β в результате следующей итеративной процедуры:

• применяя МНК к исходному уравнению регрессии, получают первоначальные оценки параметров α, β;

• вычисляют остатки **e** и в качестве оценки ρ используют коэффициент автокорреляции остатков первого порядка, т. е. полагают $\rho = r$;

• применяя МНК к преобразованному уравнению, получают новые оценки параметров α, β.

Процесс обычно заканчивается, когда очередное приближение ρ мало отличается от предыдущего.

Указанная процедура является *обобщенным МНК* при наличии автокорреляции.

Пример 5.6. Оценить зависимость расходов на конечное потребление y от дохода x по данным некоторой страны за 15 лет. Исходные данные (усл. ед.):

t	y_t	x_t	t	y_t	x_t
1	70	73	9	103	113
2	73	76	10	109	119
3	78	83	11	112	121
4	83	89	12	114	122
5	86	95	13	115	131
6	89	100	14	118	135
7	96	107	15	122	139
8	96	108			

▼Исходные данные (y_t, x_t) и расчетные показатели приведены в таблице.

T	y_t	x_t	e_t	e_{t-1}	y_t'	x_t'
1	70	73	0,273			
2	73	76	0,842	0,273	38,327	39,841
3	78	83	0,170	0,842	41,841	45,355
4	83	89	0,309	0,170	44,364	47,888
5	86	95	–1,553	0,309	44,888	50,916
6	89	100	–2,604	–1,553	46,402	52,944
7	96	107	–1,276	–2,604	51,916	57,467
8	96	108	–2,086	–1,276	48,448	55,000
9	103	113	0,863	–2,086	55,448	59,505
10	109	119	2,001	0,863	57,981	63,028
11	112	121	3,381	2,001	58,009	62,056
12	114	122	4,570	3,381	58,523	62,065
13	115	131	–1,722	4,570	58,533	70,570
14	118	135	–1,963	–1,722	61,037	70,112
15	122	139	–1,204	–1,963	63,551	72,131

Пусть исходная модель имеет вид $y_t = \alpha + \beta x_t + \varepsilon_t$.

Оцененное уравнение регрессии обычным МНК:

$$\widehat{y}_t = 10{,}58 + 0{,}81 x_t,\ R^2 = 0{,}985$$

и статистически значимо.

Коэффициент автокорреляции остатков первого порядка составляет $r = 0{,}495$, следовательно, $DW \approx 2(1 - r) = 1{,}01$.

По таблице распределения Дарбина–Уотсона находим значение $d_1 = 1{,}08$ и $d_2 = 1{,}36$. Поскольку $0 < DW < d_1$, то имеется положительная автокорреляция остатков.

Для устранения автокорреляции воспользуемся обобщенным МНК.

Применяя обычный МНК к преобразованным данным

$$y_t' = y_t - 0{,}495 \cdot y_{t-1};$$
$$x_t' = x_t - 0{,}495 \cdot x_{t-1},\ (t \geq 2),$$

получим оценку преобразованного уравнения:

$$\widehat{y}_t' = 5{,}63 + 0{,}80 x_t',\ R^2 = 0{,}927,$$

которое статистически значимо.

Коэффициент автокорреляции остатков первого порядка $r = 0{,}148$, следовательно, $DW = 1{,}7$ (отсутствует автокорреляция).

Пересчитывая оценку $\alpha = \alpha'/(1 - r) = 5{,}63/(1 - 0{,}148) = 6{,}61$, получим следующую оценку исходной модели:

$$\hat{y}_t = 6{,}61 + 0{,}80 \cdot x_t. \blacktriangle$$

Упражнение 5.2. Постройте выборочное уравнение линейной регрессии по следующим данным:

№	x	y		№	x	y		№	x	y
1	2,31	3,12		11	2,88	2,11		21	3,06	2,41
2	3,21	1,64		12	2,46	3,24		22	3,17	2,11
3	2,37	3,41		13	3,22	2,09		23	2,98	2,25
4	2,87	2,79		14	2,75	2,77		24	2,28	4
5	2,53	2,87		15	2,29	3,04		25	2,29	3,67
6	3,14	1,89		16	2,99	2,54				
7	3,26	2,1		17	3,27	1,7				
8	2,31	3,63		18	2,29	3,43				
9	2,46	1,93		19	3,28	1,93				
10	2,28	2,93		20	2,84	2,58				

Ответ. Если сохранить первое наблюдение в преобразованном уравнении, то оценка модели есть $\hat{y}_t = 7{,}14 - 1{,}62x_t$, в которой автокорреляция отсутствует.

5.5. Автокорреляция с лаговой зависимой переменной

Пусть имеется модель, в которой зависимая переменная, взятая с некоторым лагом, используется в качестве одной из объясняющих переменных.

Например, модель имеет вид

$$y_t = \alpha + \beta y_{t-1} + \varepsilon_t,$$

и допусти, что случайный член ε_t подвержен воздействию автокорреляции первого порядка.

В этом случае влияние автокорреляции сделает оценки по обычному МНК несостоятельными.

Для обнаружения автокорреляции в случае, когда уравнение регрессии включает лаговую зависимую переменную, используется *h-статистика Дарбина*, которая также вычисляется на основе остатков:

$$h \approx r\sqrt{\frac{n}{1 - n\,var(b)}},$$

где r — коэффициент автокорреляции остатков первого порядка;

n — число наблюдений в выборке;

$var(b)$ — оцененная дисперсия коэффициента β при лаговой зависимой переменной.

Значение h можно вычислить на основе обычных результатов оценивания регрессии. Этот тест предназначен только для проверки на наличие автокорреляции первого порядка.

При больших выборках h распределена как $N(0;\ 1)$ по нулевой гипотезе об отсутствия автокорреляции, следовательно, при применении двустороннего критерия и большой выборке гипотеза об отсутствии автокорреляции может быть отклонена:

- если $|h| > 1{,}96$ при уровне значимости 5%;
- если $|h| > 2{,}58$ при уровне значимости 1%.

Тест Дарбина неприменим, если $n\,var(b) \geq 1$.

Пример 5.7. Исходные данные из 31 наблюдений для модели

$$y_t = \alpha + \beta y_{t-1} + \varepsilon_t,$$

где истинные значения $\alpha = 13$, $\beta = 0{,}8$, получены по методу Монте-Карло.Оценить модель с использованием МНК и авторегрессионного преобразования.

▼Исходные данные и расчетные показатели представлены в таблице.

t	y_t	y_{t-1}	e_t	y'_t	y'_{t-1}
1	13,31			12,0293	
2	20,53	13,31	–1,863	14,8333	12,0293

t	y_t	y_{t-1}	e_t	y'_t	y'_{t-1}
3	25,94	20,53	–2,587	17,1532	14,8333
4	41,45	25,94	8,326	30,3477	17,1532
5	48,49	41,45	2,185	30,7494	30,3477
6	53,17	48,49	0,888	32,4163	30,7494
7	58,38	53,17	2,122	35,6232	32,4163
8	61,09	58,38	0,406	36,1034	35,6232
9	62,11	61,09	–0,877	35,9635	36,1034
10	67,06	62,11	3,207	40,4769	35,9635
11	71,14	67,06	3,081	42,4383	40,4769
12	81,61	71,14	10,08	51,1621	42,4383
13	83,87	81,61	3,45	48,9409	51,1621
14	91,47	83,87	9,13	55,5736	48,9409
15	94,96	91,47	6,163	55,8108	55,5736
16	93,93	94,96	2,168	53,2871	55,8108
17	85,31	93,93	–5,577	45,1080	53,2871
18	78,75	85,31	–4,814	42,2373	45,1080
19	79,15	78,75	1,16	45,4450	42,2373
20	81,87	79,15	3,54	47,9938	45,4450
21	74,2	81,87	–6,441	39,1596	47,9938
22	72,89	74,2	–1,235	41,1324	39,1596
23	69,13	72,89	–3,882	37,9331	41,1324
24	66,71	69,13	–3,107	37,1224	37,9331
25	61,25	66,71	–6,511	32,6981	37,1224
26	51,14	61,25	11,98	24,9250	32,6981
27	51,2	51,14	–3,333	29,3121	24,9250
28	60,72	51,2	6,136	38,8064	29,3121
29	60,31	60,72	–2,362	34,3218	38,8064
30	59,49	60,31	–2,834	33,6773	34,3218
31	56,98	59,49	–4,647	31,5183	33,6773

По исходным данным оцененное уравнение регрессии:

$$\underset{}{\hat{y}_t = 11{,}08 + \underset{(0{,}047)}{0{,}849\, y_{t-1}}},$$

где оценки $\alpha = 11{,}08$; $\beta = 0{,}849$; $var(b) = 0{,}047^2$.

Коэффициент автокорреляции остатков первого порядка составляет $r = 0{,}428$. Значение h-критерия Дарбина:

$$h = 0{,}428\sqrt{\frac{30}{1-30\cdot 0{,}047^2}} = 2{,}42 > 1{,}96,$$

что указывает на автокорреляцию.

Используя авторегрессионное преобразование (первое наблюдение в преобразованном уравнении сохранено), получим следующую оценку уравнения:

$$\hat{y}_t = \underset{}{7{,}925} + \underset{(0{,}078)}{0{,}808\, y_{t-1}},$$

где оценки $\alpha' = 7{,}925$; $\beta = 0{,}808$; $var(b) = 0{,}078^2$.

Пересчитывая оценку $\alpha = \alpha'/(1 - r) = 7{,}925/(1 - 0{,}428) = 13{,}85$, получим следующую оценку исходной модели:

$$\hat{y}_t = 13{,}85 + 0{,}808x_t;\ \alpha = 13{,}85;\ \beta = 0{,}808.$$

Коэффициент автокорреляции остатков первого порядка составляет $r = 0{,}0358$. Значение h-критерия Дарбина:

$$h = 0{,}0358\sqrt{\frac{30}{1-30\cdot 0{,}078^2}} = 0{,}22 < 1{,}96,$$

что указывает на отсутствие автокорреляции.

Ближе к истинным являются оценки $\alpha = 13{,}85$ и $\beta = 0{,}808$, полученные с учетом авторегрессии, чем без них. ▲

6. ДИНАМИЧЕСКИЕ ЭКОНОМЕТРИЧЕСКИЕ МОДЕЛИ

6.1. Типы моделей

Во многих экономических задачах встречаются лагированные (взятые в предыдущий момент времени) переменные. Например, y_t — выпуск предприятия за год t — может зависеть не только от инвестиции I_t в этот год, но и от инвестиций в предыдущие годы.

Эконометрическая модель, содержащая в качестве факторов не только текущие переменные, но и лаговые их значения, называется *динамической*.

Выделим два типа динамических эконометрических моделей:

•модели с распределенным лагом;

•модели авторегрессии.

Моделями с распределенным лагом называются модели, содержащие в качестве факторов лаговые значения *факторных* переменных например, модель вида

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t .$$

Моделями авторегрессии называются модели, содержащие в качестве факторов лаговые значения *зависимой* переменной, например, модель вида

$$y_t = \alpha + \beta_0 x_t + \beta_1 y_{t-1} + \varepsilon_t .$$

Обе модели включают в себя лаговые значения переменных, но существенно различаются с точки зрения статистического оценивания параметров.

6.2. Модели с распределенным лагом

Модель с распределенным лагом в предположении, что максимальная величина лага конечна, имеет вид

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \ldots + \beta_p x_{t-p} + \varepsilon_t .$$

В этой модели влияние x на y сохраняется в течение времени p.

В краткосрочном (текущем) периоде влияние x на y отражается величиной β_0, называемой *краткосрочным мультипликатором*. Он характеризует среднее абсолютное изменение y_t при изменении x_t на единицу в некоторый фиксированный момент t без учета воздействия лаговых значений фактора x.

В долгосрочном периоде (через p моментов времени) суммарное влияние x на y отражается величиной $\beta = \beta_0 + \beta_1 + \ldots + \beta_p$, называемой *долгосрочным мультипликатором*. Он характеризует общее изменение результата y в долгосрочном периоде $(t + p)$ под влиянием изменения на единицу фактора x.

В моделях с распределенным лагом объясняющие переменные *некоррелированы* со случайным членом, поэтому модель можно оценивать с помощью обычного МНК. Однако на практике оценка параметров модели затруднительна из-за высокой мультиколлинеарности факторов.

Для уменьшения числа объясняющих переменных и уменьшения эффекта мультиколлинеарности разработан ряд подходов, например модель геометрических лагов и модель полиномиальных лагов.

Модель геометрических лагов (модель Койка)

Предположим, что в исходной модели с бесконечным лагом

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \ldots + \varepsilon_t$$

коэффициенты при лаговых значениях объясняющих переменных убывают в геометрической прогрессии.

Модель имеет вид

$$y_t = \alpha + \beta_0 x_t + \beta_0 \delta x_{t-1} + \beta_0 \delta^2 x_{t-2} + \beta_0 \delta^3 x_{t-3} + \ldots + \varepsilon_t, \; 0 < \delta < 1.$$

В этой модели влияние x на y продолжается бесконечно.

В краткосрочном (текущем) периоде влияние x на y отражается коэффициентом β_0.

В долгосрочном периоде суммарное влияние x на y:

$$\beta = \sum_{k=0}^{\infty} \beta_0 \delta^k = \frac{\beta_0}{1-\delta}.$$

Модель содержит только три параметра (α, β_0, δ). Оценив эти три параметра, можно получить параметры исходной модели с бесконечным лагом из соотношений: $\beta_j = \beta_0\delta^j$, $j = 0, 1, 2,\ldots$

Однако полученная модель является нелинейной, и наличие в модели бесконечного числа лаговых переменных затрудняет практическую ее реализацию.

Процедура оценивания нелинейной модели:

• перебирается с некоторым шагом значение δ из интервала (0;1);

• для каждого δ рассчитывается $z_t = x_t + \delta x_{t-1} + \delta^2 x_{t-2} + \delta^3 x_{t-3} + \ldots + \delta^p x_{t-p}$

с таким значением p, при котором дальнейшие лаговые значения x не оказывают существенного воздействия на z;

• оценивается уравнение регрессии $y_t = \alpha + \beta_0 z_t + \varepsilon_t$;

• выбирается такое значение δ, которое обеспечивает наибольший коэффициент детерминации R^2 при оценке уравнения. Выбранному δ соответствуют вычисленные значения α, β_0 этого уравнения.

Использование этого метода при оценке параметров позволяет избежать проблемы мультиколлинеарности объясняющих переменных.

Другой метод использует так называемое преобразование Койка.

Если исходное уравнение

$$y_t = \alpha + \beta_0 x_t + \beta_0\delta x_{t-1} + \beta_0\delta^2 x_{t-2} + \beta_0\delta^3 x_{t-3} + \ldots + \varepsilon_t$$

выполняется для периода t, то оно также выполняется для периода $t - 1$:

$$y_{t-1} = \alpha + \beta_0 x_{t-1} + \beta_0\delta x_{t-2} + \beta_0\delta^2 x_{t-3} + \beta_0\delta^3 x_{t-4} + \ldots + \varepsilon_{t-1}.$$

Умножив обе части этого уравнения на δ и вычтя их из исходного уравнения, получим преобразованное уравнение:

$$y_t = \alpha(1-\delta) + \beta_0 x_t + \delta y_{t-1} + u_t,$$

где $u_t = \varepsilon_t - \delta\varepsilon_{t-1}$.

Преобразование Койка приводит модель с распределенными лагами к модели авторегрессии, в которой требуется оценить всего три параметра: α, β_0, δ.

Одна из объясняющих переменных y_{t-1} преобразованного уравнения коррелирует с одной из составляющих случайного члена $\delta\varepsilon_{t-1}$, поэтому оценки α, β_0, δ при наличии автокорреляции оказываются смещенными и несостоятельными.

Пример 6.1. Имеются данные об объеме валового внутреннего продукта y некоторой страны в зависимости от инвестиций x в ее экономику за 25 лет.

Построить модель геометрических лагов, используя: а) метод нелинейного оценивания параметров; б) преобразование Койка.

▼Исходные данные и необходимые расчетные показатели представлены в таблице.

t	y_t	x_t	y_{t-1}	z_t
1	193	30		
2	197	29	193	
3	202	29	197	
4	213	32	202	
5	222	34	213	
6	234	37	222	
7	247	41	234	
8	262	44	247	
9	269	42	262	
10	280	44	269	
11	287	46	280	
12	287	43	287	
13	296	48	287	194,67
14	310	53	296	204
15	326	59	310	219,3
16	325	54	326	225,63
17	322	44	325	219,41

18	338	52	322	225,19
19	353	60	338	236,07
20	370	66	353	250,87
21	380	67	370	260,32
22	377	54	380	255,74
23	384	63	377	266,3
24	376	54	384	262,4
25	390	60	376	265,17

а) Ограничиваясь значением $p = 12$, рассчитываем значения z:

$$z_t = x_t + \delta x_{t-1} + \delta^2 x_{t-2} + \ldots + \delta^{12} x_{t-12},$$

при этом $\delta = 0{,}8$ обеспечивала max $R^2 = 0{,}979$ при оценке уравнения

$$y_t = \alpha + \beta_0 z_t + \varepsilon_t.$$

Оцененное уравнение при таком δ имеет вид

$$\hat{y}_t = 46{,}32 + 1{,}232 \cdot z_t,$$

где $\alpha = 46{,}32$; $\beta_0 = 1{,}232$.

Соответственно, оцененное уравнение геометрических лагов есть

$$\hat{y}_t = 46{,}32 + 1{,}232x_t + 0{,}986x_{t-1} + 0{,}788x_{t-2} + \ldots$$

В краткосрочном (текущем) периоде влияние x на y отражается коэффициентом $\beta_0 = 1{,}232$.

В долгосрочном периоде суммарное влияние x на y равно:

$$\beta = \frac{\beta_0}{1-\delta} = \frac{1{,}232}{1-0{,}8} = 6{,}2.$$

б) Оценивая преобразованное уравнение

$$y_t = \alpha(1 - \delta) + \beta_0 x_t + \delta y_{t-1} + u_t,$$

получим

$$\hat{y}_t = 13{,}73 + 1{,}197 \cdot x_t + \underset{(0{,}032)}{0{,}785} \cdot y_{t-1}, \quad R^2 = 0{,}996,$$

где $\alpha(1 - \delta) = 13{,}73$; $\beta_0 = 1{,}197$; $\delta = 0{,}785$.

Проверка на автокоррелированность остатков при наличии лаговой зависимой переменной показывает, что критерий Дарбина $h = 0{,}115 < 1{,}96$, т. е. автокорреляция отсутствует.

Оценки $\alpha = \dfrac{13{,}73}{1-0{,}785} = 63{,}86$; $\beta = \dfrac{\beta_0}{1-\delta} = \dfrac{1{,}197}{1-0{,}785} = 5{,}6$.

Соответственно, оцененное уравнение геометрических лагов:

$$\hat{y}_t = 63{,}86 + 1{,}197x_t + 0{,}939x_{t-1} + 0{,}737x_{t-2} + \dots$$

В краткосрочном (текущем) периоде влияние x на y отражается коэффициентом $\beta_0 = 1{,}197$, а в долгосрочном периоде суммарное влияние x на y отражается коэффициентом $\beta = 5{,}6$. ▲

Модель полиномиальных лагов (метод Алмон)

В модели полиномиальных лагов предполагается, что зависимость коэффициентов при лаговых значениях объясняющей переменной от величины лага описывается полиномом m-й степени. Модель имеет вид

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \dots + \beta_p x_{t-p} + \varepsilon_t,$$

где

$$\beta_s = \gamma_0 + \gamma_1 S + \gamma_2 S^2 + \dots + \gamma_m S^m, \quad m \le p\,.$$

Предположим, что величина лага p известна. Кроме того, необходимо установить степень полинома m. Обычно на практике ограничиваются рассмотрением полиномов 2-й и 3-й степеней.

Рассмотрим следующие варианты.

I. Пусть $p = 3$, $m = 2$, тогда исходная модель:

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \beta_3 x_{t-3} + \varepsilon_t,$$

где $\beta_0 = \gamma_0$;

$\beta_1 = \gamma_0 + \gamma_1 + \gamma_2$;

$\beta_2 = \gamma_0 + 2\gamma_1 + 4\gamma_2$;

$\beta_3 = \gamma_0 + 3\gamma_1 + 9\gamma_2$.

Преобразованная модель имеет вид

$$y_t = \alpha + \gamma_0 z_0 + \gamma_1 z_1 + \gamma_2 z_2 + \varepsilon_t;$$

где $z_0 = x_t + x_{t-1} + x_{t-2} + x_{t-3}$;

$z_1 = x_{t-1} + 2x_{t-2} + 3x_{t-3}$;

$z_2 = x_{t-1} + 4x_{t-2} + 9x_{t-3}$.

II. Пусть $p = 4$, $m = 2$, тогда исходная модель:

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \beta_3 x_{t-3} + \beta_4 x_{t-4} + \varepsilon_t,$$

где $\beta_0 = \gamma_0$;

$\beta_1 = \gamma_0 + \gamma_1 + \gamma_2$;

$\beta_2 = \gamma_0 + 2\gamma_1 + 4\gamma_2$;

$\beta_3 = \gamma_0 + 3\gamma_1 + 9\gamma_2$;

$\beta_4 = \gamma_0 + 4\gamma_1 + 16\gamma_2$.

Преобразованная модель имеет вид

$$y_t = \alpha + \gamma_0 z_0 + \gamma_1 z_1 + \gamma_2 z_2 + \varepsilon_t,$$

где $z_0 = x_t + x_{t-1} + x_{t-2} + x_{t-3} + x_{t-4}$;

$z_1 = x_{t-1} + 2x_{t-2} + 3x_{t-3} + 4x_{t-4}$;

$z_2 = x_{t-1} + 4x_{t-2} + 9x_{t-3} + 16x_{t-4}$.

Используя МНК, оцениваем параметры преобразованной модели и затем рассчитываем параметры исходной модели с распределенным лагом.

Пример 6.2. По данным примера 6.1 об объеме валового внутреннего продукта y от инвестиций x построить модель с распределенным лагом для $p = 3$, $m = 2$.

▼Общий вид исходной модели:

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \beta_3 x_{t-3} + \varepsilon_t .$$

Исходные данные (y_t, x_t) и преобразованные (z_0, z_1, z_2) представлены в таблице:

t	y_t	x_t	x_{t-1}	x_{t-2}	x_{t-3}	z_0	z_1	z_2
1	193	30						
2	197	29	30					
3	202	29	29	30				
4	213	32	29	29	30	120	177	415

t	y_t	x_t	x_{t-1}	x_{t-2}	x_{t-3}	z_0	z_1	z_2
5	222	34	32	29	29	124	177	409
6	234	37	34	32	29	132	185	423
7	247	41	37	34	32	144	201	461
8	262	44	41	37	34	156	217	495
9	269	42	44	41	37	164	237	541
10	280	44	42	44	41	171	253	587
11	287	46	44	42	44	176	260	608
12	287	43	46	44	42	175	260	600
13	296	48	43	46	44	181	267	623
14	310	53	48	43	46	190	272	634
15	326	59	53	48	43	203	278	632
16	325	54	59	53	48	214	309	703
17	322	44	54	59	53	210	331	767
18	338	52	44	54	59	209	329	791
19	353	60	52	44	54	210	302	714
20	370	66	60	52	44	222	296	664
21	380	67	66	60	52	245	342	774
22	377	54	67	66	60	247	379	871
23	384	63	54	67	66	250	386	916
24	376	54	63	54	67	238	372	882
25	390	60	54	63	54	231	342	792

Оцененная исходная модель имеет вид

$$\hat{y}_t = \underset{(9{,}4)}{41{,}73} + \underset{(0{,}33)}{2{,}42}x_t + \underset{(0{,}39)}{0{,}71}x_{t-1} + \underset{(0{,}41)}{0{,}99}x_{t-2} + \underset{(0{,}34)}{1{,}47}x_{t-3}, \quad R^2 = 0{,}981.$$

в которой коэффициент 0,71 при переменной x_{t-1} незначим.

Оцененная преобразованная модель имеет вид

$$\hat{y}_t = \underset{(9{,}27)}{41{,}6} + \underset{(0{,}30)}{2{,}32}\, z_0 - \underset{(0{,}66)}{1{,}92}\, z_1 + \underset{(0{,}21)}{0{,}56}\, z_2, \quad R^2 = 0{,}981,$$

и все коэффициенты при переменных значимы.

Получили следующие оценки параметров преобразованной модели:

$$\gamma_0 = 2{,}32,\ \gamma_1 = -1{,}92,\ \gamma_2 = 0{,}56.$$

Коэффициенты регрессии исходной модели:

$$\beta_0 = 2{,}32;$$

$$\beta_1 = 2{,}32 - 1{,}92 + 0{,}56 = 0{,}96;$$

$$\beta_2 = 2{,}32 - 2 \cdot 1{,}92 + 4 \cdot 0{,}56 = 0{,}72;$$

$$\beta_3 = 2{,}32 - 3 \cdot 1{,}92 + 9 \cdot 0{,}56 = 1{,}6.$$

Таким образом, модель с распределенным лагом имеет вид

$$\hat{y}_t = 41{,}6 + 2{,}32x_t + 0{,}96x_{t-1} + 0{,}72x_{t-2} + 1{,}6x_{t-3}.$$

Краткосрочный мультипликатор равен 2,32, а долгосрочный мультипликатор равен 2,32 + 0,96 + 0,72 + 1,6=5,6. Это означает, что увеличение инвестиций в экономику страны на 1 усл. ед. приведет к росту валового внутреннего продукта в среднем на 2,32 усл. ед. в текущем периоде, и на 5,6 усл. ед. через 3 года. ▲

Упражнение 6.1. По данным примера 6.1 об объеме валового внутреннего продукта y от инвестиций x построить модель с распределенным лагом для p =4, m = 2.

Ответ. Модель с распределенным лагом имеет вид

$$\hat{y}_t = 46{,}2 + 1{,}87x_t + 1{,}08x_{t-1} + 0{,}70x_{t-2} + 0{,}73x_{t-3} + 1{,}17x_{t-4}$$

6.3. Модели авторегрессии

Пусть имеется модель авторегрессии вида

$$y_t = \alpha + \beta_0 x_t + \beta_1 y_{t-1} + \varepsilon_t.$$

Для интерпретации коэффициентов модели авторегрессии сделаем предположение о наличии бесконечного лага в воздействии текущего значения зависимой переменной на ее последующие значения и о выполнении неравенства $|\beta_1| < 1$ (*условие устойчивости*).

В краткосрочном (текущем) периоде влияние x на y отражается величиной β_0 (*краткосрочный мультипликатор*). Он характеризует краткосрочное изменение y под влиянием изменения x на единицу своего измерения.

В долгосрочной перспективе суммарное влияние x на y отражается величиной β (*долгосрочный мультипликатор*):

$$\beta = \beta_0 + \beta_0\beta_1 + \beta_0\beta_1^2 + \beta_0\beta_1^3 + ... = \frac{\beta_0}{1-\beta_1}.$$

При наличии в правой части модели авторегрессии лаговой зависимой переменной y_{t-1} может иметь место автокорреляция остатков. Кроме того, может иметь место зависимость между объясняющей переменной y_{t-1} и случайным членом ε_t. Применение обычного МНК при оценке параметров модели авторегрессии дает неудовлетворительные результаты.

Одним из возможных методов оценки параметров уравнения авторегрессии является **метод инструментальных переменных**.

В качестве инструментальной переменной можно взять переменную x_{t-1}, которая коррелированна с y_{t-1} и некоррелированы с ε_t. Практически, в качестве инструментальной переменной можно взять оценку

$$\hat{y}_{t-1} = \gamma_0 + \gamma_1 x_{t-1},$$

полученную из предполагаемой линейной зависимости y_{t-1} от x_{t-1}.

Тогда оценку параметров модели авторегрессии можно найти обычным МНК из соотношения

$$y_t = \alpha + \beta_0 x_t + \beta_1 \hat{y}_{t-1} + \varepsilon_t,$$

где x_t y_t — исходные, а $\hat{y}_{t-1}$ — расчетные данные.

Практическая реализация метода инструментальных переменных осложняется появлением мультиколлинеарности факторов x_t и $\hat{y}_{t-1}$ в модели.

Однако если коллинеарность факторов не привела к большим стандартным ошибкам оценок, то применение инструментальной переменной можно считать возможным.

Пример 6.3. По месячным данным объема продаж продукции фирмы (y) и инвестициях (x) за два года построить модель авторегрессии.

▼Исходные данные (усл. ед.) и расчетные показатели представим в таблице.

t	y_t	x_t	y_{t-1}	$\widehat{y}_{t-1}$	e_t
1	164	78			
2	162	81	164	170,30	–7,36
3	165	89	162	172,91	–8,54
4	168	76	165	179,88	–7,65
5	172	105	168	168,56	–2,30
6	177	101	172	193,80	–16,30
7	182	93	177	190,32	–6,43
8	186	94	182	183,36	2,84
9	187	107	186	184,23	–0,28
10	191	103	187	195,55	–4,21
11	196	116	191	192,06	0,13
12	201	170	196	203,38	–18,10
13	213	101	201	250,39	–25,26
14	211	110	213	190,32	18,09
15	219	138	211	198,16	12,49
16	228	145	219	222,53	0,27
17	232	180	228	228,62	–9,79
18	239	165	232	259,09	–23,05
19	244	144	239	246,03	–2,14
20	249	130	244	227,75	21,08
21	255	155	249	215,57	30,17
22	264	142	255	237,33	25,30
23	265	151	264	226,01	32,92
24	267	305	265	233,85	–11,89

Применение обычного МНК к модели

$$y_t = \alpha + \beta_0 x_t + \beta_1 y_{t-1} + \varepsilon_t$$

приводит к результатам:

$$\widehat{y}_t = -0{,}495 - \underset{(0,022)}{0{,}012x_t} + \underset{(0,034)}{1{,}031y_{t-1}}, \quad R^2 = 0{,}991,$$

где коэффициент 0,012 при переменной x_t статистически незначим.

Для оценки параметров уравнения авторегрессии используем метод инструментальных переменных.

Оценка уравнения регрессии $y_{t-1} = \gamma_0 + \gamma_1 x_{t-1} + \varepsilon_{t-1}$ обычным МНК дает

$$\hat{y}_{t-1} = \underset{(17,74)}{102,40} + \underset{(0,142)}{0,870} \cdot x_{t-1},$$

где все коэффициенты статистически значимы.

Тогда оценка модели

$$y_t = \alpha + \beta_0 x_t + \beta_1 \hat{y}_{t-1} + \varepsilon_t$$

обычным МНК дает следующие результаты:

$$\hat{y}_t = 12,68 + \underset{(0,092)}{0,263} \cdot x_t + \underset{(0,163)}{0,795} \cdot y_{t-1}, \quad R^2 = 0,781.$$

Уравнение авторегрессии в целом значимо, значимы также коэффициенты при переменных, однако автокорреляция остатков не устранена.

Действительно, коэффициент автокорреляции остатков первого порядка $r = 0,447$. Значение h-критерия Дарбина:

$$h = 0,447\sqrt{\frac{23}{1 - 23 \cdot 0,163^2}} = 3,44 > 1,96,$$

что указывает на автокорреляцию.

Автокорреляция в остатках по авторегрессионным моделям можно устранить с помощью авторегрессионных преобразований, используя модели ARMA и ARIMA (см. разд. 7).

6.4. Примеры моделей с лагированными переменными

Модель частичной корректировки

В модели частичной корректировки предполагается, что поведенческое уравнение определяет не фактическое значение зависимой переменной y_t, а ее *желаемый* (целевой) уровень:

$$y_t^* = \alpha + \beta x_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0; \sigma^2).$$

Предполагается также, что фактическое значение зависимой переменной не выходит мгновенно на желаемый уровень, а изменяется только на долю λ в нужном направлении:

$$y_t - y_{t-1} = \lambda\,(y_t^* - y_{t-1}), \quad (0 \le \lambda \le 1).$$

Это выражение можно переписать следующим образом:

$$y_t = \lambda\, y_t^* + (1 - \lambda)\, y_{t-1},$$

откуда видно, что y_t получается как взвешенное среднее желаемого уровня и фактического значения этой переменной в предыдущем периоде.

Параметр λ называется *корректирующим коэффициентом*. Чем больше λ, тем быстрее происходит процесс корректировки, причем:

- если $\lambda = 1$, то $y_t = y_t^*$ и полная корректировка происходит за один период;
- если $\lambda = 0$, то корректировка y_t не происходит совсем.

Подставляя y_t^* в выражение для y_t, получим

$$y_t = \alpha\,\lambda + \beta\,\lambda\, x_t + (1 - \lambda)\, y_{t-1} + \lambda\,\varepsilon_t.$$

Полученное уравнение включает только фактические значения переменных.

Поскольку случайные члены некоррелированы, состоятельные оценки параметров можно получить, применяя МНК к оцениванию составных параметров $\alpha\lambda$, $\beta\lambda$ и $(1 - \lambda)$ в полученном уравнении.

Пример 6.4. Производственные компании распределяют прибыль *П*, оставшуюся после уплаты налогов: одну часть на выплату доходов акционерам в форме дивидендов (*D*), другую — на финансирование инвестиций.

Когда прибыль растет, дивиденды тоже увеличиваются, но, как правило, не в той пропорции. Причиной этого в основном является осторожность руководства компании.

Предположим, что у фирмы имеется целевая долгосрочная доля выплат γ и что желаемый объем дивидендов D_t^* соотносится с текущей прибылью $П_t$ как $D_t^* = \gamma\, П_t + \varepsilon_t$. Однако реальный объем дивидендов подвержен процессу частичной корректировки:

$$D_t - D_{t-1} = \lambda\,(D_t^* - D_{t-1}),$$

или

$$D_t = \gamma\,\lambda\,П_t + (1-\lambda)\,D_{t-1} + \lambda\,\varepsilon_t,\ \ (0 \le \lambda \le 1).$$

Используя данные о деятельности производственных компаний за ряд предыдущих лет:

t	*D*	*П*
1	100	400
2	300	600
3	450	700
4	550	800
5	700	100
6	800	1100
7	900	1300
8	1000	1400
9	1100	1500
10	1200	1700

можно построить уравнение регрессии

$$\underset{(28{,}9)}{D_t = 68} + \underset{(0{,}071)}{0{,}293П_t} + \underset{(0{,}081)}{0{,}582D_{t-1}},\ R^2 = 0{,}999,$$

где все коэффициенты значимы.

Из соотношения $1 - \lambda = 0{,}582$ определяется корректирующий коэффициент $\lambda = 0{,}418$, а из соотношения $\gamma\lambda = 0{,}293$ — оценка доли выплат $\gamma = 0{,}7$.

Модель адаптивных ожиданий

Предположим, что зависимая переменная y_t связана с ожидаемым значением объясняющей переменной x в $(t + 1)$-м периоде соотношением

$$y_t = \alpha + \beta\,x^*_{t+1} + \varepsilon_t,$$

где x^*_{t+1} — ожидаемое значение объясняющей переменной, которую необходимо заменить наблюдаемыми переменными.

Такая модель возникает, например, в случае, когда фирма принимает решение об объеме производимой в период t продукции y_t до того, как известна цена x_{t+1}, по которой эта продукция может быть продана в следующем периоде. Поскольку цена x_{t+1} неизвестна в период t, то решение принимается на основе ожидаемого значения x^*_{t+1}.

Процесс формирования ожиданий следующий:

$$x^*_{t+1} - x_t^* = \lambda\ (x_t - x_t^*),$$

или

$$x^*_{t+1} = \lambda\, x_t + (1 - \lambda)\, x_t^*,\ \ (0 \le \lambda \le 1),$$

т. е. ожидаемое значение переменной x_t^* в следующем периоде является взвешенным средним ее фактического и ожидаемого значений в текущем периоде.

Параметр λ называется *коэффициентом ожидания*. Чем больше λ, тем быстрее ожидаемое значение адаптируется предыдущим реальным значениям. Чем меньше λ, тем ожидаемое значение в будущем ближе к ожидаемому значению предыдущего периода x^*_t, т. е. тенденции в ожиданиях сохраняются.

Величину x_t^* можно выразить через x^*_{t-1} и т. д. Повторяя эту процедуру бесконечное число раз, получим

$$x^*_{t+1} = \lambda\ [\ x_t + (1 - \lambda)\, x_{t-1} + (1 - \lambda\ )^2 x_{t-2} + \ldots].$$

В итоге получаем модель адаптивных ожиданий, в которой ожидаемое значение переменной является взвешенным средним ее прошлых значений с геометрически убывающими весами.

Подставим выражение для x^*_{t+1} в исходную модель и заменим $(1 - \lambda\ )$ на δ:

$$y_t = \alpha + \beta\ \lambda\ (x_t + \delta\, x_{t-1} + \delta^2 x_{t-2} + \ldots\ ) + \varepsilon_t,$$

т. е. пришли к модели *геометрических лагов*. Параметры уравнения можно оценить с помощью **метода нелинейного оценивания**.

Полученное уравнение может быть преобразовано к виду:

$$y_t = \alpha\lambda\ +\ \beta\lambda x_t + (1 - \lambda)y_{t-1} + u_t,$$

где $u_t = \varepsilon_t - (1 - \lambda)\ \varepsilon_{t-1}$.

Явный вид случайного члена показывает, что имеет место корреляция между лаговой переменной y_{t-1} и ошибкой регрессии u_t, т. е. оценки МНК будут несостоятельными. Для получения состоятельных оценок следует оценивать модель геометрических лагов.

Пример 6.5. По данным цены x_t на сырье в момент времени t и цены y_t форвардной сделки на товар в момент времени $t + 1$ оценить параметры адаптивной модели ожиданий. Исходные данные (усл. ед.) представлены в таблице.

t	y_t	x_t	t	y_t	x_t
1	210,3	33,1	14	334,0	42,2
2	237,0	34,2	15	334,0	42,5
3	276,5	35,4	16	325,2	42,7
4	280,9	36,5	17	338,2	42,9
5	289,6	37,1	18	345,7	43,3
6	297,6	38,4	19	339,5	43,5
7	320,6	39,6	20	344,4	43,8
8	319,5	40,1	21	371,8	44,3
9	326,3	41,2	22	346,8	44,6
10	319,1	41,4	23	349,7	45,1
11	341,6	41,6	24	348,9	46,2
12	333,7	41,8	25	362,8	47,3
13	334,0	42,2	26	383,4	48,2

▼Величина y_t зависит в большей степени от ожидаемого значения x^*_{t+1} на сырье и может быть описана моделью адаптивных ожиданий

$$y_t = \alpha + \beta x^*_{t+1} + \varepsilon_t,$$

или (в преобразованном виде) уравнением

$$y_t = \alpha\lambda + \beta\lambda x_t + (1 - \lambda)y_{t-1} + u_t.$$

Оценка модели адаптивных ожиданий в преобразованном виде есть:

$$\hat{y}_t = \underset{}{-5{,}23} + \underset{(1{,}42)}{5{,}69}\cdot x_t + \underset{(0{,}13)}{0{,}30}\cdot y_{t-1}, \quad R^2 = 0{,}922.$$

Оценки параметров α, β, λ находятся из решения системы:

$$\begin{cases} \alpha\lambda = -5{,}23 \\ \beta\lambda = 5{,}69 \\ 1-\lambda = 0{,}30 \end{cases} \Rightarrow \alpha = -7{,}47;\ \beta = 8{,}13;\ \lambda = 0{,}7.$$

Но эти оценки несостоятельные, поскольку значение h-критерия Дарбина $h = 2{,}05 > 1{,}96$, что указывает на автокорреляцию остатков.

Для получения состоятельных оценок применим к уравнению

$$y_t = \alpha\lambda + \beta\lambda x_t + (1-\lambda)y_{t-1} + u_t$$

обратное преобразование Койка.

Получим модель геометрических лагов, которое оцениваем нелинейным методом, ограничиваясь первыми пятью членами, т. е.

$$y_t = \alpha_0 + \beta_0 z_t + \varepsilon_t,$$

где $z_t = x_t + \delta x_{t-1} + \delta^2 x_{t-2} + \delta^3 x_{t-3} + \delta^4 x_{t-4} + \delta^5 x_{t-5}$.

При $\delta = 0{,}3$ обеспечивался $max\ R^2 = 0{,}814$ оценки уравнения.

Исходные данные для такой оценки приведены в таблице.

t	y_t	x_t	z_t
1	210,3	33,1	
2	237,0	34,2	
3	276,5	35,4	
4	280,9	36,5	
5	289,6	37,1	
6	297,6	38,4	54,1
7	320,6	39,6	55,8
8	319,5	40,1	56,8
9	326,3	41,2	58,2
10	319,1	41,4	58,8
11	341,6	41,6	59,2
12	333,7	41,8	59,5
13	334,0	42,2	60,0

t	y_t	x_t	z_t
14	334,0	42,5	60,5
15	325,2	42,7	60,8
16	338,2	42,9	61,1
17	345,7	43,3	61,6
18	339,5	43,5	62,0
19	344,4	43,8	62,4
20	371,8	44,3	63,0
21	346,8	44,6	63,5
22	349,7	45,1	64,1
23	348,9	46,2	65,4
24	362,8	47,3	66,9
25	383,4	48,2	68,2

Оцененное уравнение при таком δ имеет вид

$$\hat{y}_t = 36,35 + 4,96 \cdot z_t,$$

где $\alpha_0 = 36,35$; $\beta_0 = 4,96$.

Оценки параметров α, β, λ находятся из решения системы:

$$\begin{cases} \alpha\lambda = 36,35 \\ \beta\lambda = 4,96 \\ 1 - \lambda = 0,30 \end{cases} \Rightarrow \alpha = 51,9;\ \beta = 7,08;\ \lambda = 0,7$$

и эти оценки состоятельные. ▲

7. АВТОРЕГРЕССИОННЫЕ ПРОЦЕССЫ И ИХ МОДЕЛИРОВАНИЕ

7.1. Понятие стационарности

Часто экономические показатели, представленные временными рядами, имеют настолько сложную структуру, что моделирование таких рядов путем построения моделей тренда, сезонности и применение других традиционных подходов не приводит к удовлетворительным результатам. Ряд остатков часто имеет статистические закономерности, которые можно моделировать. Как правило, ряд остатков — это стационарный ряд.

Будем рассматривать класс стационарных временных рядов. Задача состоит в построении модели для случайных остатков временного ряда и прогнозирования его значений.

Ряд y_t называется *строго стационарным* (или стационарным в узком смысле), если совместное распределение вероятностей m наблюдений y_{t_1}, $y_{t_2}, \ldots, y_{t_m}$ такое же, как и для m наблюдений $y_{t_1+\tau}$, $y_{t_2+\tau}, \ldots$, $y_{t_m+\tau}$ при любых m, t_1, $t_2, \ldots$, t_m, τ.

Таким образом, свойства строго стационарного временного ряда не зависят от начала отсчета времени.

Обычно интересует не все распределение, а средние значения и ковариации. Поэтому часто используется понятие *слабой стационарности* (или стационарности в широком смысле), которое состоит в том, что среднее, дисперсия и ковариации y_t не зависят от момента времени t:

$$M(y_t) = \mu,\ D(y_t) = \gamma_0,\ cov(y_t, y_{t-\tau}) = \gamma_\tau.$$

В дальнейшем под стационарностью будем понимать слабую стационарность. Временной ряд, не удовлетворяющий перечисленным выше свойствам, называется *нестационарным* временным рядом.

Введем понятие автокорреляционной функции (АКФ):

$$\rho(\tau) = cov(y_t, y_{t-\tau})/\ D(y_t) = \gamma_\tau / \gamma_0.$$

Значения АКФ характеризуют тесноту (степень) статистической связи между уровнями временного ряда, разделенными τ временными тактами.

Проверка на стационарность. *Первое*, что следует сделать, — построить график временного ряда.

График может содержать очевидный на глаз тренд или сезонную компоненту. Также возможно, что разброс наблюдений возрастает или убывает со временем. Это может служить указанием на зависимость среднего или дисперсии от времени, т. е. ряд будет, скорее всего, нестационарным.

Второе — построить график выборочной АКФ или коррелограммы $r(\tau)$, являющейся статистической оценкой $\rho(\tau)$.

Коррелограмма стационарного временного ряда быстро убывает с ростом τ после нескольких первых значений. Если график убывает достаточно медленно, есть основание предположить нестационарность ряда.

Наряду с автокорреляционной функцией при исследовании стационарных временных рядов рассматривается частная автокорреляционная функция $\rho_ч(\tau)$ (ЧАКФ). С помощью ЧАКФ измеряется корреляция между уровнями ряда y_t и $y_{t-\tau}$, разделенными τ временными тактами, при исключении влияния на эту взаимосвязь всех промежуточных уровней ряда $y_{t-1}, y_{t-2},\ldots, y_{t-\tau+1}$.

Кроме выборочной АКФ можно также построить график выборочной ЧАКФ $r_ч(\tau)$, которая также должна быстро убывать для стационарного процесса.

В случае стационарного ряда y_t значение выборочной ЧАКФ вычисляются как МНК-оценка последнего коэффициента β_τ в регрессионном уравнении: $y_t = \beta_0 + \beta_1 y_{t-1} + \ldots + \beta_\tau y_{t-\tau} + \varepsilon_t$.

Третье — можно использовать тесты на наличие единичного корня, часть из которых будет в дальнейшем рассмотрена.

Примером стационарного по определению временного ряда является ряд с независимыми одинаково распределенными наблюдениями:

$$y_t = \varepsilon_t,\ \varepsilon_t \sim iid(0, \sigma^2),\ t = 1,\ldots,n.$$

Этот процесс называется «белым шумом», у него

$$\mu = 0,\ \gamma_0 = \sigma^2,\ \gamma_\tau = 0,\ \tau > 0.$$

Большое значение в анализе временных рядов имеют стационарные временные ряды, вероятностные свойства которых не изменяются во времени. Это объясняется тем, что многие временные ряды могут быть приведены к стационарному ряду после выделения и удаления из них тренда, сезонной компоненты или взятия разностей. Как правило, ряд ошибок является стационарным рядом.

Наиболее распространенными моделями стационарных рядов являются *модели авторегрессии* и *модели скользящего среднего.*

7.2. Модель авторегрессии $AR(p)$

Рассмотрим класс авторегрессионных моделей, обозначаемых $AR(p)$-модели (число в скобках указывает порядок авторегрессии).

В авторегрессионной модели каждое значение ряда y_t зависит только от конечного числа p предыдущих своих значений.

Авторегрессионной процесс $AR(p)$ может быть представлен в виде

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p} + \varepsilon_t$$

или в более короткой форме

$$(1 - \alpha_1 L - \alpha_2 L^2 - \ldots - \alpha_p L^p) y_t \equiv \Phi(L) y_t = \varepsilon_t,$$

где L — оператор сдвига, т. е. преобразование ряда, смещающее его на один временной такт;

$\Phi(L)$ – полином от оператора сдвига.

Удобным и полезным инструментом для изучения процессов авторегрессии является *характеристический многочлен* (характеристический полином)

$$\Phi(z) = 1 - \alpha_1 z - \alpha_2 z^2 - \ldots - \alpha_p z^p$$

и связанное с ним *характеристическое уравнение*

$$1 - \alpha_1 z - \alpha_2 z^2 - \ldots - \alpha_p z^p = 0.$$

Для выполнения условия стационарности все корни многочлена $\Phi(z)$ должны лежать вне единичного круга, т. е. все корни соответствующего характеристического уравнения должны быть по модулю больше 1 и различны.

Например, рассмотрим процесс авторегрессии $AR(2)$:

$$y_t = -0{,}9y_{t-1} - 0{,}2y_{t-2} + \varepsilon_t.$$

Характеристическое уравнение принимает вид: $1 + 0{,}9z + 0{,}2z^2 = 0$, или $z^2 + 4{,}5z + 5 = 0$. При этом его корни $z_1 = -2{,}5$ и $z_2 = -2$ по абсолютной величине больше единицы, процесс — стационарный.

Простейшим примером является модель $AR(1)$, или *марковский процесс*:

$$y_t = \alpha y_{t-1} + \varepsilon_t,\ \varepsilon_t \sim iid(0, \sigma^2),\ t = 1,\ldots,n.$$

Условие стационарности ряда для $AR(1)$ определяется условием $|\alpha| < 1$, или, что то же самое, корень характеристического уравнения $1 - \alpha z = 0$ должен быть по абсолютной величине больше 1.

Используя оператор сдвига, запишем уравнение в виде

$$(1 - \alpha L)y_t = \varepsilon_t.$$

Разлагая оператор $(1 - \alpha L)^{-1}$ в ряд по αL, получим

$$y_t = (1 - \alpha L)^{-1}\,\varepsilon_t = (1 + \alpha L + \alpha^2 L^2 + \ldots)\,\varepsilon_t = \varepsilon_t + \alpha\varepsilon_{t-1} + \alpha^2\varepsilon_{t-2} + \ldots,$$

откуда $M(y_t) = 0$, $D(y_t) = \sigma^2/(1 - \alpha^2)$, $cov(y_t, y_{t-k}) = \alpha^k\sigma^2/(1 - \alpha^2)$.

Если в уравнение добавить константу: $y_t = \delta + \alpha y_{t-1} + \varepsilon_t$, то $M(y_t) = \delta/(1 - \alpha)$.

Поскольку среднее, дисперсия и ковариация не зависят от времени t, процесс $AR(1)$ является стационарным. Неравенство $|\alpha| < 1$ является необходимым условием стационарности процесса y_t.

Автокорреляционная функция равна $\rho(\kappa) = \alpha^\kappa$, поэтому степень тесноты корреляционной связи между членами последовательности экспоненциально убывает по мере их взаимного удаления друг от друга по времени.

Значения частной автокорреляционной функции $\rho_ч(\kappa)$ равны нулю для всех лагов $\kappa > 1$.

Важным примером является процесс, называемый случайным блужданием:

$$y_t = y_{t-1} + \varepsilon_t,\ ,\ \varepsilon_t \sim iid(0, \sigma^2),\ t = 1,\ldots,n.$$

Этот процесс по виду похож на $AR(1)$ с $\alpha = 1$, однако существенно отличается по своим свойствам от стационарного процесса $AR(1)$ с $|\alpha| < 1$.

Учитывая, что ошибка ε_t не коррелирована с y_{t-1}, можно получить

$$M(y_t) = M(y_{t-1}); \; D(y_t) = D(y_{t-1}) + \sigma^2,$$

т. е. случайное блуждание нестационарное, так как $D(y_t) \neq D(y_{t-1})$. При этом дисперсия может неограниченно возрастать со временем.

Процесс случайного блуждания отличается от стационарного процесса $AR(1)$ тем, что влияние возмущения ε_t не ослабевает: $y_t = \varepsilon_t + \varepsilon_{t-1} + \ldots$, в то время как в $AR(1)$ их влияние падает: $y_t = \varepsilon_t + \alpha\varepsilon_{t-1} + \alpha^2\varepsilon_{t-2} + \ldots$

В общем случае для процесса $AR(p)$ вытекают следующие *практические рекомендации по их идентификации*:

• значения коэффициентов АКФ экспоненциально затухают (либо монотонно, либо попеременно меняя знак);

• значения коэффициентов ЧАКФ имеют выбросы (пики) на первых p-лагах, а значения коэффициентов для лагов, больших порядка авторегрессии, статистически незначимы.

Тестирование на единичные корни

Рассмотрим $AR(1)$ с нулевым средним:

$$y_t = \alpha y_{t-1} + \varepsilon_t, \; \varepsilon_t \sim iid(0, \sigma^2).$$

Ряд y_t является стационарным, если $|\alpha| < 1$. Если $\alpha = 1$, то y_t — нестационарный временной ряд случайного блуждания. В этом случае временной ряд y_t имеет единичный корень. Как определить по имеющимся наблюдениям, верно ли, что в $AR(1)$ процессе $\alpha = 1$?

Вычтем y_{t-1} из обеих частей уравнения для y_t (взятие первой разности):

$$\Delta y_t = \beta y_{t-1} + \varepsilon_t, \text{ где } \Delta y_t = y_t - y_{t-1}, \; \beta = \alpha - 1.$$

Дики и Фуллер рассмотрели три регрессии:

$$\Delta y_t = \beta y_{t-1} + \varepsilon_t; \qquad (7.1)$$

$$\Delta y_t = \alpha_0 + \beta y_{t-1} + \varepsilon_t; \qquad (7.2)$$

$$\Delta y_t = \alpha_0 + \beta y_{t-1} + \alpha_2 t + \varepsilon_t. \qquad (7.3)$$

Во всех трех регрессиях интересующий параметр β.

Нулевая гипотеза H_0: $\beta = 0$ (имеется единичный корень, временной ряд нестационарен) против альтернативной H_1: $\beta < 0$.

Оценивается методом наименьших квадратов указанное уравнение (7.1) (или 7.2, или 7.3). Получают оценку $\hat{y}$, стандартную ошибку и соответствующее значение t-статистики. Нулевая гипотеза будет отвергнута, если наблюдаемое значение t-статистики меньше критического значения $t_{кр}$, взятого из таблиц Дики и Фуллера. Значение $t_{кр}$ зависит от фиксированного уровня значимости, объема выборки и вида регрессии.

Если нулевая гипотеза отвергается, то в этом случае ряд может быть отнесен к стационарным. Если нулевая гипотеза не отвергается, то имеется единичный корень и исследуемый ряд соответствует модели случайного блуждания.

Критические значения не изменятся, если в правые части регрессий (7.1), (7.2), (7.3) добавить слагаемые вида Δy_{t-1}, Δy_{t-2},... Это позволяет тестировать наличие единичного корня в AR-моделях произвольного порядка. Тест, соответствующий уравнению с лагированными значениями приращений Δy_{t-1}, Δy_{t-2},... в правой части, называется *расширенным тестом Дики–Фуллера*.

Если порядок процесса $AR(p)$ заранее неизвестен, то рекомендуется включать возможно большее количество лагов, чтобы устранить возможную автокорреляцию ошибок. Однако включение чрезмерного количества лагов снижает мощность теста.

Тест Дики–Фуллера включен во все современные эконометрические пакеты.

7.3. Модель скользящего среднего $MA(q)$

При моделировании стационарных временных рядов широкое распространение получили модели скользящего среднего, когда y_t линейно зависит от конечного числа q предыдущих значений ε_t:

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q},\ \varepsilon_t \sim iid(0, \sigma^2).$$

Модель скользящего среднего порядка q обозначается $MA(q)$. Согласно определению процесс $MA(q)$ стационарен при любом q и любых θ_i.

Сформируем условие обратимости процесса, т. е. возможности его представления в виде AR-процесса.

Рассмотрим модель $MA(1)$:

$$y_t = \Theta(L)\varepsilon_t = \varepsilon_t - \theta\varepsilon_{t-1}.$$

Представим $MA(1)$ процесс в виде авторегрессионного процесса:

$$\Theta(L)^{-1} y_t = \varepsilon_t \text{, или } y_t = \theta y_{t-1} - \theta^2 y_{t-2} - \ldots + \varepsilon_t .$$

Такое $AR(\infty)$ представление процесса $MA(1)$ возможно только в случае обратимости оператора $\Theta(L) = 1 - \theta L$, т. е. когда выполняется условие обратимости $|\theta| < 1$.

Легко определить среднее и дисперсию процесса $MA(1)$:

$$M(y_t) = 0,\ D(y_t) = \sigma^2(1 + \theta^2).$$

Если в уравнение добавить константу: $y_t = \delta + \varepsilon_t - \theta\varepsilon_{t-1}$, то $M(y_t) = \delta$.

АКФ и ЧАКФ модели $MA(1)$ задаются выражениями

$$\rho(\tau) = \begin{cases} -\theta/(1+\theta^2),\ \tau = 1 \\ 0,\ \tau > 1 \end{cases};\ \rho_ч(\tau) = -\theta^\tau \frac{1-\theta^2}{1-\theta^{2(\tau+1)}}.$$

Поведение ЧАКФ определяется затухающей экспонентой. Если значение $\rho(1) > 0$, то параметр $\theta < 0$, следовательно, $\rho_ч(\tau)$ осциллирует с переменным знаком. Если значение $\rho(1) < 0$, то параметр $\theta > 0$, следовательно, все значения $\rho_ч(\tau)$ отрицательны.

В общем случае для процесса $MA(q)$ вытекают следующие *практические рекомендации* по их идентификации:

• автокорреляционная функция имеет выбросы (пики) на первых q-лагах, а остальные значения статистически незначимы;

• частная автокорреляционная функция экспоненциально затухает (либо монотонно, либо осциллируя, т. е. меняя знак).

7.4. Модель авторегрессии и скользящего среднего *ARMA*

Для описания стационарных процессов также может использоваться модель авторегрессии и скользящего среднего порядка (p, q), или модель $ARMA(p, q)$, включающая как члены, описывающие авторегрессионные составляющие, так и члены, моделирующие остаток в виде процесса скользящих средних.

Модель $ARMA(p, q)$ имеет вид

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q},$$

где $\varepsilon_t \sim iid(0, \sigma^2)$, или в более короткой форме

$$\Phi(L) y_t = \Theta(L) \varepsilon_t,$$

где $\Phi(L), \Theta(L)$ — полиномы от оператора сдвига.

По аналогии с $AR(p)$ условия стационарности $ARMA(p, q)$ определяются корнями характеристического уравнения $\Phi(z) = 0$: если эти корни лежат вне единичного круга, то процесс стационарен. Для обратимости процесса необходимо и достаточно, чтобы все корни характеристического уравнения $\Theta(z) = 0$ лежали вне единичного круга.

Рассмотрим процесс $ARMA(1, 1)$:

$$y_t = \alpha y_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}$$

или через оператор сдвига:

$$(1 - \alpha L) y_t = (1 - \theta L) \varepsilon_t.$$

Процесс стационарен, если $|\alpha| < 1$, и обратим, если $|\theta| < 1$.

Применяя те же методы, что и ранее, получим следующие выражения для среднего и дисперсии $ARMA(1, 1)$ процесса: $M(y_t) = \delta/(1 - \alpha)$, если константа δ включена в модель, и $D(y_t) = \sigma^2(1 + \theta^2 - 2\alpha\theta)/(1 - \alpha^2)$.

Значения автокорреляционной функции для $ARMA(1,1)$ имеет вид

$$\rho(\tau) = \begin{cases} \dfrac{(1-\alpha\theta)(\alpha-\theta)}{1+\theta^2-2\alpha\theta}, & \tau = 1 \\ \alpha^{\tau-1}\rho(1), & \tau > 1 \end{cases},$$

т. е. значения АКФ будут экспоненциально убывать от значения $\rho(1)$, причем если α положительно — то монотонно, если отрицательно — то знакопеременно.

Поведение ЧАКФ определяется начальным значением $\rho_ч(1)$, после которого функция экспоненциально убывает. Если $\theta > 0$, то функция убывает монотонно, если $\theta < 0$ — то знакопеременно.

Обычно число параметров p или q не бывает больше 2.

Для процесса $ARMA(p,q)$ вытекают следующие практические рекомендации по их идентификации:

• ***ARMA*(1,0):** АКФ экспоненциально затухает, ЧАКФ имеет выброс на лаге 1;

• ***ARMA*(2,0):** АКФ имеет форму затухающей синусоидальной волны или экспоненциально затухает, ЧАКФ имеет выбросы на лагах 1 и 2;

• ***ARMA*(0,1):** АКФ имеет выброс на лаге 1, ЧАКФ экспоненциально затухает;

• ***ARMA*(0,2):** АКФ имеет выбросы для лагов 1 и 2, ЧАКФ имеет форму затухающей синусоидальной волны или экспоненциально затухает;

• ***ARMA*(1,1):** АКФ экспоненциально затухает от значения $\rho(1)$,ЧАКФ экспоненциально убывает от значения $\rho_ч(1)$.

7.5. *ARIMA*-модели

Некоторые нестационарные временные ряды могут быть сведены к стационарным с помощью оператора последовательной разности.

Пусть временной ряд y_t после применения к нему d раз оператора последовательной разности стал стационарным рядом $\Delta^d y_t$, удовлетворяющим $ARMA(p,\ q)$-модели. В этом случае процесс y_t принято называть *интегрированным процессом авторегрессии и скользящего среднего*, или $ARIMA(p,\ d,\ q)$. В специальной литературе — эта **модель Бокса–Дженкинса**.

Модель *ARIMA* обладает тремя параметрами: p — порядок авторегрессии *AR*; d — порядок последовательных разностей уровней временных рядов, обеспечивающий стационарность ряда; q — порядок скользящей средней *MA*. Из модели для ряда Δy_t можно получить модель для исходного ряда y_t, используя соотношение $y_t = y_{t-1} + \Delta y_t$.

Например, модель *ARMA* (2, 1): $y_t = 1{,}2y_{t-1} - 0{,}2y_{t-2} + \varepsilon_t - 0{,}5\varepsilon_{t-1}$ имеет для *AR*-части характеристическое уравнение

$$1 - 1{,}2z + 0{,}2z^2 = (1 - 0{,}2z)(1 - z) = 0,$$

корни которого $z_1 = 5$, $z_2 = 1$.

Поскольку один из корней равен 1, то процесс *ARMA*(2, 1) является нестационарным. Оператор авторегрессии этого процесса можно представить в следующем виде: $(1 - 0{,}2L)(1 - L) = (1 - 0{,}2L)\Delta$.

Введем обозначение $\Delta y_t = y_t - y_{t-1}$. Полученный процесс Δy_t является стационарным процессом *ARMA* (1, 1), задаваемым уравнением:

$$\Delta y_t = 0{,}2\Delta y_{t-1} + \varepsilon_t - 0{,}5\varepsilon_{t-1},$$

при этом исходный ряд y_t, является рядом *ARIMA*(1, 1, 1), или проинтегрированным первого порядка *ARMA*(1, 1)-рядом.

В общем случае, если характеристическое уравнение процесса *ARMA*($p + d$, q) содержит d единичных корней, а все остальные корни по модулю больше единицы, то d-я разность этого временного ряда

$$\Delta^d y_t = \Phi(L)(1 - L)^d y_t$$

может быть представлена как стационарный процесс *ARMA*(p, q):

$$\Phi(L)\Delta^d y_t = \Theta(L)\varepsilon_t,$$

при этом исходный ряд y_t, является рядом типа *ARIMA*(p, d, q), или d раз проинтегрированным ARMA(p, q)-рядом.

Методология Бокса–Дженкинса подбора *ARIMA*-модели для описания и прогнозирования временного ряда включает следующие этапы:

- идентификация модели;
- оценивание модели и проверка ее адекватности;
- прогнозирование.

Пример 7.1. Проведем подбор *ARIMA*-модели по данным золотовалютных резервов y_t России с 31.12.05 по 12.10.07 и сделаем прогноз на 10 недель вперед.

Исходные данные и расчетные показатели приведены в следующей таблице.

Дата	t	y_t	$z_t = \Delta y_t$	z_{t-1}	e_t
31.12.05	1	182,2			
06.01.06	2	182,3	0,1		
13.01.06	3	184,6	2,3	0,1	0,2693
20.01.06	4	185,2	0,6	2,3	–2,0422
27.01.06	5	188,2	3,0	0,6	0,8303
03.02.06	6	188,5	0,3	3,0	–2,5368
10.02.06	7	194,2	5,7	0,3	3,6137
17.02.06	8	195,4	1,2	5,7	–2,3872
24.02.06	9	195,6	0,2	1,2	–2,1365
03.03.06	10	197,9	2,3	0,2	0,2415
10.03.06	11	201,7	3,8	2,3	1,1578
17.03.06	12	204,1	2,4	3,8	–0,6591
24.03.06	13	204,4	0,3	2,4	–2,3700
31.03.06	14	205,9	1,5	0,3	–0,5863
07.04.06	15	208,1	2,2	1,5	–0,2199
14.04.06	16	212,0	3,9	2,2	1,2856
21.04.06	17	217,1	5,1	3,9	2,0131
28.04.06	18	225,7	8,6	5,1	5,1795
05.05.06	19	231,1	5,4	8,6	1,0067
12.05.06	20	236,1	5,0	5,4	1,4961
19.05.06	21	236,7	0,6	5,0	–2,7927
26.05.06	22	243,3	6,6	0,6	4,4303
02.06.06	23	247,0	3,7	6,6	–0,1374
09.06.06	24	247,9	0,9	3,7	–2,1313
16.06.06	25	246,0	–1,9	0,9	–4,1531
23.06.06	26	247,2	1,2	–1,9	–0,2748
30.06.06	27	250,6	3,4	1,2	1,0635
07.07.06	28	253,2	2,6	3,4	–0,3480

Дата	t	y_t	$z_t = \Delta y_t$	z_{t-1}	e_t
15.07.06	29	255,7	2,5	2,6	–0,2256
21.07.06	30	262,9	7,2	2,5	4,5022
28.07.06	31	265,6	2,7	7,2	–1,3042
04.08.06	32	266,9	1,3	2,7	–1,4534
11.08.06	33	277,0	10,1	1,3	7,7357
18.08.06	34	275,0	–2,0	10,1	–6,8102
25.08.06	35	258,5	–16,5	-2,0	–17,9470
01.09.06	36	260,4	1,9	–16,5	4,4832
08.09.06	37	260,7	0,3	1,9	–2,2310
15.09.06	38	259,0	–1,7	0,3	–3,7863
22.09.06	39	261,0	2,0	–1,7	0,4696
29.09.06	40	266,6	5,6	2,0	3,0412
06.10.06	41	267,9	1,3	5,6	–2,2594
13.10.06	42	266,5	–1,4	1,3	–3,7643
20.10.06	43	267,3	0,8	–1,4	–0,8138
27.10.06	44	269,1	1,8	0,8	–0,4253
03.11.06	45	274,2	5,1	1,8	2,5968
10.11.06	46	277,0	2,8	5,1	–0,6205
17.11.06	47	278,9	1,9	2,8	–0,8812
24.11.06	48	283,4	4,5	1,9	1,9690
01.12.06	49	290,1	6,7	4,5	3,4463
08.12.06	50	293,8	3,7	6,7	–0,1652
15.12.06	51	295,8	2,0	3,7	–1,0313
22.12.06	52	299,2	3,4	2,0	0,8412
29.12.06	53	303,0	3,8	3,4	0,8520
05.01.07	54	303,9	0,9	3,8	–2,1591
12.01.07	55	301,7	–2,2	0,9	–4,4531
19.01.07	56	302,7	1,0	–2,2	–0,3914
26.01.07	57	303,8	1,1	1,0	–1,1809
02.02.07	58	304,6	0,8	1,1	–1,5087
09.02.07	59	309,5	4,9	0,8	2,6747
16.02.07	60	311,2	1,7	4,9	–1,6649
23.02.07	61	311,1	–0,1	1,7	–2,5754

Дата	t	y_t	$z_t = \Delta y_t$	z_{t-1}	e_t
02.03.07	62	315,3	4,2	–0,1	2,2249
09.03.07	63	317,3	2,0	4,2	–1,1703
16.03.07	64	321,7	4,4	2,0	1,8412
23.03.07	65	332,6	10,9	4,4	7,6741
30.03.07	66	338,7	6,1	10,9	1,0674
06.04.07	67	346,3	7,6	6,1	3,9016
13.04.07	68	356,6	10,3	7,6	6,1847
20.04.07	69	361,2	4,6	10,3	–0,2658
27.04.07	70	369,0	7,8	4,6	4,5185
04.05.07	71	372,1	3,1	7,8	–1,0709
11.05.07	72	386,3	14,2	3,1	11,3354
18.05.07	73	394,3	8,0	14,2	2,0502
25.05.07	74	402,2	7,9	8,0	3,6735
01.06.07	75	403,6	1,4	7,9	–2,7987
08.06.07	76	406,5	2,9	1,4	0,5079
15.06.07	77	405,0	–1,5	2,9	–4,3090
22.06.07	78	406,6	1,6	–1,5	0,0140
29.06.07	79	406,0	–0,6	1,6	–3,0476
06.07.07	80	408,4	2,4	–0,6	0,5638
13.07.07	81	411,2	2,8	2,4	0,1300
20.07.07	82	413,1	1,9	2,8	–0,8812
27.07.07	83	417,3	4,2	1,9	1,6690
03.08.07	84	416,8	–0,5	4,2	–3,6703
10.08.07	85	420,2	3,4	–0,5	1,5360
17.08.07	86	414,7	–5,5	3,4	–8,4480
24.08.07	87	413,8	–0,9	-5,5	–1,3742
31.08.07	88	416,0	2,2	–0,9	0,4472
07.09.07	89	417,1	1,1	2,2	–1,5144
14.09.07	90	420,9	3,8	1,1	1,4913
21.09.07	91	422,5	1,6	3,8	–1,4591
28.09.07	92	425,1	2,6	1,6	0,1524
05.10.07	93	424,8	–0,3	2,6	–3,0256
12.10.07	94	434,0	9,2	–0,3	7,2805

Идентификация модели. Первый шаг идентификации — получение стационарного ряда. Исходный ряд y_t не является стационарным, поскольку имеет возрастающий тренд (рис. 7.1).

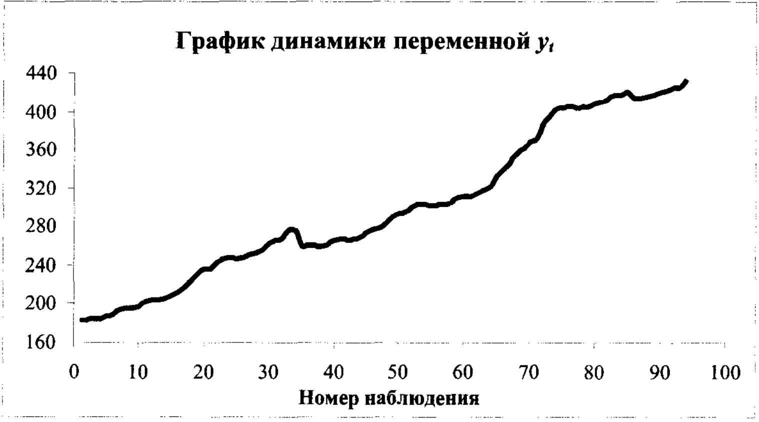

Рис. 7.1

Чтобы ряд стал стационарным, необходимо брать последовательные разности до тех пор, пока он не станет стационарным.

Для определения порядка разности нужно исследовать автокоррелограмму. Если наблюдается медленное убывание выборочных коэффициентов автокорреляции в зависимости от лага, обычно берут разность первого порядка.

На рис. 7.2 показана АКФ переменной y_t, где коэффициенты выборочной АКФ вычисляются по формуле

$$r_k = \sum_{t=k+1}^{n} (y_t - \bar{y})(y_{t-k} - \bar{y}) / \sum_{t=1}^{n} (y_t - \bar{y})^2, \; k = 1,2,\ldots$$

Из графика видно, что автокорреляции в зависимости от лага убывают медленно, что говорит о взятии разности первого порядка ($d = 1$) для идентификации модели $ARIMA(p,d,q)$.

Найдем первую разность $z_t = \Delta y_t$, где $\Delta y_t = y_t - y_{t-1}$ и построим ее график в зависимости от номера наблюдений (рис. 7.3), из которого видно, что ряд стал стационарным, так как тренд отсутствует.

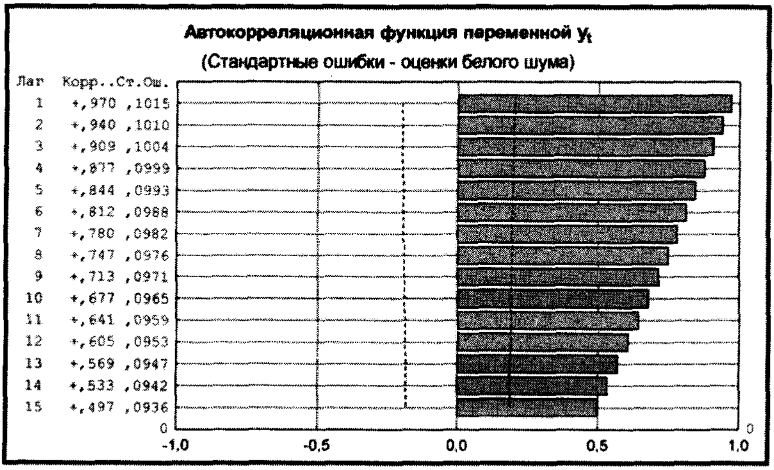

Рис. 7.2

Рис. 7.3

Для стационарного ряда z_t исследуется характер поведения выборочных АКФ и ЧАКФ, которые позволяют сформулировать несколько гипотез о возможных порядках авторегрессии (p) и скользящего среднего (q).

Для стационарного ряда z_t значение выборочной ЧАКФ вычисляются как МНК-оценка последнего коэффициента β_κ в регрессионном уравнении: $z_t = \beta_0 + \beta_1 z_{t-1} + \ldots + \beta_\tau z_{t-\kappa} + \varepsilon_t$.

На рис. 7.4 приведены автокорреляционная и частная автокорреляционная функции переменной z_t.

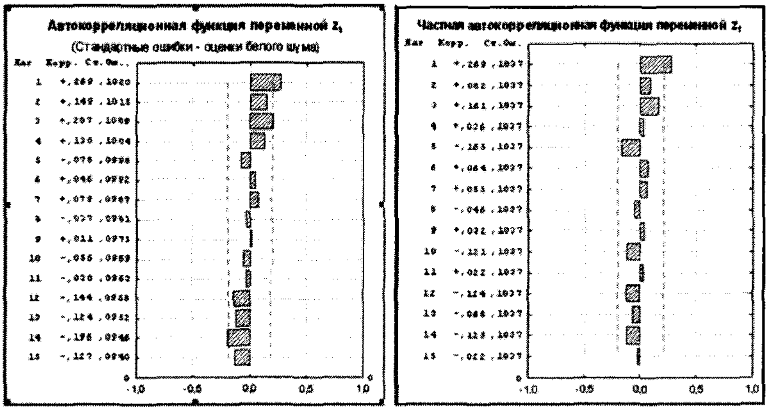

Рис. 7.4

На рис. 7.4 в *ЧАКФ* значимо отличается от нуля только значение корреляции для первого лага, АКФ имеет небольшой выброс на первом лаге и заметную тенденцию к затуханию.

В соответствии с указанными ранее практическими рекомендациями по идентификации модели *ARMA*, выбираем модель *ARIMA*(1,1,0), но можно также использовать и модель *ARIMA*(0,1,1).

Оценивание ARMA-моделей производится различными методами (линейный и нелинейный МНК, полный или условный метод максимального правдоподобия).

Рассмотрим модель *ARIMA*(1,1,0). Оценим модель авторегрессии первого порядка со свободным членом $z_t = \delta + \alpha \cdot z_{t-1} + \varepsilon_t$ методом МНК.

В таблице исходных данных приведены расчетные показатели, необходимые для оценивания параметров уравнения в Excel.

Оцененная статистически значимая модель есть

$$\hat{z}_t = 1{,}966 + 0{,}277 \cdot z_{t-1},$$

где $\delta = 1{,}966$; $\alpha = 0{,}277$, а остаточная дисперсия (остаток) = 13,3.

Коэффициенты модели статистически значимые.

Учитывая, что $\mu = \delta/(1 - \alpha) = 2{,}719$, запишем преобразованную модель в виде

$$z_t = 2{,}719 + 0{,}277 \cdot z_{t-1} + \varepsilon_t,$$

где $\mu = 2{,}719$; $\alpha = 0{,}277$.

Такие же оценки параметров этой модели получаются при использовании пакета STATISTICA.

Если имеется несколько моделей, успешно прошедших проверку условия адекватности, то выбираем ту модель, у которой дисперсия остатков минимальна.

Существуют различные критерии для *проверки адекватности ARMA-моделей*:

1. Оценки коэффициентов модели должны статистически значимо отличаться от нуля.

2. Остатки модели e_t должны быть похожи на белый шум, т. е. иметь нулевую автокорреляцию.

Проверим адекватность модели *ARIMA*(1,1,0).

Коэффициенты $\mu = 2{,}719$ и $\alpha = 0{,}277$ статистически значимы (первое условие по проверке адекватности модели выполнено).

При проверке значимости коэффициентов АКФ остатков используются два подхода:

• проверка значимости каждого коэффициента автокорреляции отдельно;

• проверка значимости группы коэффициентов автокорреляции с помощью теста Бокса–Льюнга.

Для проверки выполнения второго условия рассмотрим следующую таблицу результатов, которую можно получить расчетным способом на основании остатков e_t модели *ARIMA*(1,1,0) данной задачи или сразу с помощью пакета STATISTICA.

Таблица результатов

	Автокорреляционная функция остатков модели *ARIMA*(1,1,0) Стандартные ошибки — ошибки белого шума			
Лаг	**Авто-корр.**	**Ст. ошиб.**	**Бокса-Льюнга *Q***	***p***
1	–0,021	0,1020	0,04	0,8377
2	0,026	0,1015	0,11	0,9487

	Автокорреляционная функция остатков модели *ARIMA*(1,1,0) Стандартные ошибки — ошибки белого шума			
Лаг	**Авто-корр.**	**Ст. ошиб.**	**Бокса-Льюнга *Q***	***p***
3	0,158	0,1009	2,56	0,4652
4	0,111	0,1004	3,77	0,4375
5	–0,141	0,0998	5,78	0,3286
6	0,051	0,0992	6,04	0,4185
7	0,089	0,0987	6,86	0,4432
8	–0,072	0,0981	7,40	0,4940
9	0,029	0,0975	7,49	0,5865
10	–0,058	0,0965	7,84	0,6443
11	0,019	0,0963	7,88	0,7236
12	–0,119	0,0958	9,42	0,6668
13	–0,044	0,0952	9,63	0,7237
14	–0,153	0,0946	12,25	0,5864
15	–0,064	0,0940	12,71	0,6249

В таблице результатов приведены автокорреляции, их стандартные ошибки, статистика Бокса-Льюнга Q и уровень значимости для этой статистики.

Автокорреляция — это корреляция исходного ряда с самим собой, сдвинутым на определенный лаг k. Коэффициенты выборочной автокорреляционной функции остатков определяются по формуле

$$r_k = \sum_{t=k+1}^{n} e_t e_{t-k} / \sum_{t=1}^{n} e_t^2 \text{ , } k = 1, 2, \ldots$$

В предположении, что процесс — белый шум (в этом процессе все автокорреляции равны нулю), стандартные ошибки r_k определяются как Ст.ош.$(r_k) = \sqrt{(1/n)\cdot(n-k)/(n+2)}$, где n — число наблюдений ряда.

Из сопоставления полученных значений первых двух столбцов таблицы результатов следует, что коэффициенты автокорреляции незначимы на всех 15 лагах.

Для проверки равенства нулю k первых значений автокорреляционной функции остатков используется Q-статистика Бокса–Льюнга.

На данном лаге k-статистика Бокса-Льюнга Q определяется как $Q_k = n(n+2)\sum_{i=1}^{k} r_i^2 /(n-i)$. При выполнении нулевой гипотезы отсутствия автокорреляции Q-статистика имеет распределение $\chi^2(\tau - p - q)$.

Уровни значимости $p_к$, соответствующие статистике Q_k, можно определить с помощью функции Excel: = ХИ2РАСП(Q_k, k). Если $p_к$ больше заданного уровня значимости, то k первых значений автокорреляционной функции остатков статистически незначимы.

В следующей таблице показан пример расчета значений Q_k, p_k для лагов $к$ = 1,2,3 по приведенным формулам, n = 92.

k	r_k	Q_k	p_k
1	–0,021	$Q_1 = 92 \cdot 94 \cdot 0{,}021^2 / 91 = 0{,}04$	$ХИ2РАСП(0{,}04;1) = 0{,}838$
2	0,026	$Q_2 = Q_1 + 92 \cdot 94 \cdot 0{,}026^2 / 90 = 0{,}11$	$ХИ2РАСП(0{,}11;2) = 0{,}949$
3	0,158	$Q_3 = Q_2 + 92 \cdot 94 \cdot 0{,}158^2 / 89 = 2{,}56$	$ХИ2РАСП(2{,}56;3) = 0{,}465$

Из рассмотрения полученных значений последнего столбца таблицы результатов следует, что все k первых значений автокорреляционной функции остатков статистически незначимы.

Таким образом, второе условие по проверке адекватности модели выполнено.

Прогнозирование в модели *ARIMA*(1,1,0)

Рассмотрим нестационарный временной ряд y_t, первые разности которого z_t являются $AR(1)$ процессом: $z_t = y_t - y_{t-1}$, $z_t - \mu = \alpha(z_{t-1} - \mu) + \varepsilon_t$. Многократное применение этих выражений дает: $y_{n+s} = y_n + z_{n+1} + z_{n+2} + \ldots + z_{n+s} = (y_n + s\mu) + (z_{n+1} - \mu) + \ldots + (z_{n+s} - \mu)$, или

$$y_{n+s} = y_n + s\mu + \frac{\alpha(1-\alpha^s)}{1-\alpha}(y_n - y_{n-1} - \mu) + e_{n+s},$$

где $e_{n+s} = \varepsilon_{n+s} + (1+\alpha)\varepsilon_{n+c-1} + \ldots + (1+\alpha+\alpha^2+\ldots+\alpha^{s-1})\varepsilon_{n+1}$.

Прогноз $\hat{y}_{n+s}$, минимизирующий среднеквадратическое отклонение, равен сумме первых трех слагаемых в y_{n+s}, т. е.

$$\hat{y}_{n+s} = y_n + s\mu + \frac{\alpha(1-\alpha^s)}{1-\alpha}(y_n - y_{n-1} - \mu).$$

Прогнозируемое значение $\hat{y}_{n+s}$ и ошибка прогноза e_{n+s} увеличиваются по мере роста s.

Произведем прогноз на 1 и 2 шага. Для двух последних наблюдений имеем $y_{93} = 424{,}8$ и $y_{94} = 434$.

Прогноз на 1 шаг: $\hat{y}_{95} = y_{94} + \mu + \alpha\cdot(y_{94} - y_{93} - \mu) = 434 + 2{,}7193 +$ $+ 0{,}2777\cdot(434 - 424{,}8 - 2{,}7193) = 438{,}519$.

Прогноз на 2 шага: $\hat{y}_{96} = y_{94} + 2\cdot\mu + \alpha\cdot(1 - \alpha^2)/(1 - \alpha)\cdot(y_{94} - y_{93} -$ $- \mu) = 434 + 2\cdot 2{,}7193 + 0{,}2777\cdot(1 - 0{,}2777^2)/(1 - 0{,}2777)\cdot(434 - 424{,}8 -$ $- 2{,}7193) = 441{,}738$.

При использовании пакета STATISTICA сразу получаем таблицу прогноза на требуемое число шагов и доверительные границы прогнозных значений.

	Прогнозы; Модель:(1,1,0) Сезонный лаг: 12 (Резервы Исход.:РЕЗЕРВЫ Начало исходных: 1 Конец исходн.: 94			
Набл. N	Прогноз	Нижний 90,0000%	Верхний 90,0000%	Ст.ошиб.
95	438,5191	432,4410	444,5972	3,65758
96	441,7383	431,8765	451,6000	5,93448
97	444,5964	431,7486	457,4442	7,73139
98	447,3543	432,0233	462,6853	9,22569
99	450,0843	432,6024	467,5662	10,52007
100	452,8066	433,4064	472,2068	11,67439
101	455,5268	434,3806	476,6729	12,72509
102	458,2463	435,4874	481,0053	13,69559
103	460,9657	436,7009	485,2305	14,60178
104	463,6850	438,0025	489,3676	15,45493

На рис. 7.5 приведен график динамики переменной y_t (Резервы) и прогноз с доверительным интервалом на 10 шагов вперед. Доверительный интервал показывает, что с вероятностью 90% прогнозируемое значение попадет в заданный интервал.

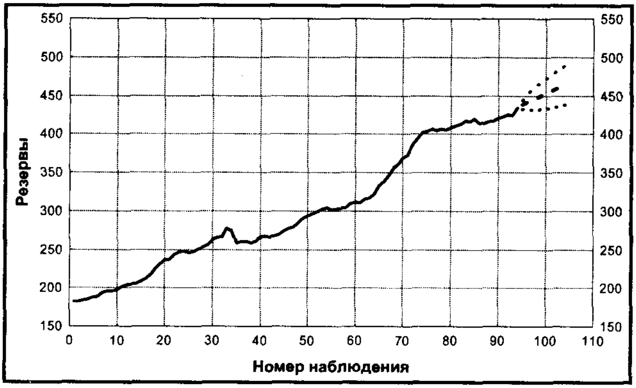

Рис. 7.5

7.6. Сезонные модели *ARIMA*

Сезонная модель представляется в виде $ARIMA(p, d, q)(P, D, Q)_s$, где к параметрам модели p, d, q добавлены сезонные параметры: P, D, Q и s-сезонная авторегрессия, сезонная разность, сезонное скользящее среднее и сезонный период соответственно.

Идентификация сезонной модели производится тем же способом, что и идентификация несезонной модели. Поведение автокорреляционной и частной автокорреляционной функций на начальных лагах позволяет идентифицировать стандартным образом несезонную компоненту, а на лагах, кратных сезонному лагу, — сезонную составляющую.

При наличие ярковыраженной сезонной компоненты целесообразно включение в модель сезонного дифференцирования, но при этом желательно, чтобы $d + D \leq 2$.

Существенно облегчить решение задач анализа и прогнозирования финансово-экономических показателей поможет использование современных компьютерных статистических пакетов. В некоторых компьютерных пакетах, например *SPSS*, реализованы процедуры автоматического подбора структуры модели Бокса–Дженкинса. В работе [3] подробно изложены прикладные процедуры обработки данных в пакете STATISTICA, в том числе и подбор *ARIMA*-модели.

Пример 7.2. Проведем подбор *ARIMA*-модели по данным примера 4.8 пассажирских авиаперевозок y_t одной из авиакомпаний (тыс. чел.) за 2007–2012 гг. и сделаем прогноз на следующий 2013 год.

▼Исходный ряд y_t не является стационарным, поскольку имеет возрастающий тренд (см. рис. 4.6). Динамика имеет явный сезонный характер с периодом в один год.

Используя пакет STATISTICA, легко перебрать множество вариантов моделей $ARIMA(p, d, q)(P, D, Q)$ при не очень больших значениях параметров $p \le 2, d \le 2, q \le 2, P \le 2, D \le 2, Q \le 2$.

Основными критериями отбора являются:

• остаток (остаточная дисперсия) должен быть по возможности минимизирован;

• все оцениваемые параметры (включая свободный параметр модели) должны быть высокозначимыми;

• прогноз должен учитывать сезонный характер данных.

Применяя указанные критерии, найдем следующую идентификацию искомой модели: $ARIMA(0,1,0)(1,0,0)$ (без свободного параметра). Данная модель содержит первую разность для учета линейного тренда в динамике авиаперевозок y_t и коэффициент сезонной авторегрессии.

В таблице приведены результаты оценивания параметров модели $ARIMA(0,1,0)(1,0,0)$.

	Исход.:Авиаперевозки Преобразования: D(1) Модель(0,1,0)(1,0,0) Сезонный лаг: 12 Остаток= 714,24					
Параметр	Парам.	Асимпт. ст.ошиб.	Асимпт. t(70)	p	Нижняя 95% дов.	Верхняя 95% дов.
Ps(1)	0,935968	0,066979	13,97397	0,000000	0,802382	1,069555

Приведенные в таблице параметры модели высокозначимы.

На рис. 7.6 приведены автокорреляционная и частная автокорреляционная функции остатков модели.

Из графиков видно, что эти корреляции незначимо отличаются от нуля, т. е. выбранная модель ARIMA(0,1,0)(1,0,0) является адекватной.

Рис. 7.6

В следующей таблице приведены результаты прогноза Авиаперевозок на 12 месяцев вперед и доверительные границы прогнозных значений.

	Прогнозы; Модель:(0,1,0)(1,0,0) Исход.:Авиаперевозки Начало исходных: 1 Конец исходн.:72			
Набл. №	Прогноз	Нижний 90%	Верхний 90%	Ст. ошиб.
73	439,5	392,6	486,5	28,2
74	392,8	326,4	459,2	39,9
75	448,0	366,6	529,3	48,8
76	453,6	359,6	547,5	56,4
77	484,4	379,4	589,5	63,0
78	534,9	419,8	650,0	69,0
79	601,3	477,0	725,6	74,6
80	635,9	503,0	768,8	79,7
81	546,1	405,2	687,1	84,6
82	489,1	340,5	637,7	89,1
83	438,6	282,8	594,4	93,5
84	463,9	301,1	626,6	97,6

Результаты прогноза близки к прогнозам, полученными при использовании модели Тейла-Вейджа (пример 4.8), но в данном случае имеем также и доверительные границы прогнозных значений (риски).

На рис. 7.7 приведен график динамики переменной y_t (Авиаперевозки) и прогноз с доверительным интервалом на 12 месяцев вперед. Доверительный интервал показывает, что с вероятностью 90% прогнозируемое значение попадет в заданный интервал.

Рис. 7.7

В настоящее время прогнозирование на основе авторегрессионных моделей стало важным инструментом в деятельности плановых, аналитических, маркетинговых отделов производственных предприятий, банков, торговых и страховых компаний.

8. СИСТЕМЫ ОДНОВРЕМЕННЫХ УРАВНЕНИЙ

8.1. Структурная и приведенная формы уравнений

Системный подход в экономическом анализе предполагает сложную структуру взаимосвязей между признаками, когда эффективность деятельности экономического объекта характеризуется несколькими показателями. В этом случае одного регрессионного уравнения может быть недостаточно, и для описания явления или процесса может потребоваться система уравнений и тождеств.

Система одновременных уравнений получила название также структурной формы модели.

Структурной формой модели (системой одновременных уравнений) называется система уравнений, в каждом из которых аргументы помимо объясняющих переменных могут включать в себя также объясняемые переменные из других уравнений системы.

Уравнения, составляющие исходную модель, называются *структурными уравнениями модели*.

Простейшая структурная форма модели имеет вид

$$\begin{cases} y_1 = \alpha_1 + \beta_{12} y_2 + \alpha_{11} x_1 + \alpha_{12} x_2 + \varepsilon_1; \\ y_2 = \alpha_2 + \beta_{21} y_1 + \alpha_{21} x_1 + \alpha_{22} x_2 + \varepsilon_2, \end{cases}$$

где y и x — зависимая и независимая переменные;

ε_1 и ε_2 — случайные члены;

α, β — параметры модели.

Параметры структурной формы модели называются *структурными коэффициентами.*

Структурная форма модели обычно включает в систему не только уравнения, отражающие взаимосвязи между отдельными переменными, но и уравнения, отражающие тенденцию развития явления, а также разного рода уравнения-тождества. Тождества не содержат каких-либо подлежащих оценке параметров, а также не включают случайного члена.

В процессе оценивания параметров одновременных уравнений следует различать эндогенные и экзогенные переменные. Приставки «эндо» и «экзо» означают соответственно внутреннее и внешнее.

Эндогенными считаются переменные, значения которых определяются *внутри* модели и являются зависимыми переменными.

Экзогенными считаются переменные, значения которых определяются *вне* модели и являются независимыми переменными.

В качестве экзогенных переменных могут рассматриваться значения эндогенных переменных за предшествующий период времени (лаговые переменные).

Предполагается, что в каждом уравнении экзогенные переменные некоррелированы со случайным членом.

В общем случае эндогенные переменные коррелированы со случайным членом, поэтому применение МНК к структурной форме модели приводит к *смещенным* и *несостоятельным* оценкам структурных коэффициентов.

Для определения структурных коэффициентов структурная форма модели преобразуется в приведенную форму.

Приведенной формой модели называется система уравнений, в каждом из которых эндогенные переменные выражены только через экзогенные переменные и случайные составляющие.

Например, приведенная форма исходной модели имеет вид

$$\begin{cases} y_1 = \alpha'_1 + \alpha'_{11}x_1 + \alpha'_{12}x_2 + \nu_1; \\ y_2 = \alpha'_2 + \alpha'_{21}x_1 + \alpha'_{22}x_2 + \nu_2, \end{cases}$$

где α' — параметры приведенной формы;

ν_1 и ν_2 — случайные члены.

Параметры приведенной формы модели называются *коэффициентами приведенной формы* (приведенными коэффициентами).

Коэффициенты приведенной формы оцениваются обычным МНК, поскольку экзогенные переменные некоррелированы со случайным членом.

Оцененные коэффициенты приведенной формы могут быть использованы для оценивания структурных коэффициентов. Такой способ оценивания структурных коэффициентов называется *косвенным* МНК.

Приведенная форма модели аналитически уступает структурной форме модели, так как в ней отсутствуют оценки взаимосвязи между эндогенными переменными.

При переходе от приведенной формы модели у структурной возникает проблема идентификации.

Идентификация — это единственность соответствия между приведенной и структурной формами модели.

Тот или иной структурный коэффициент может либо однозначно выражаться через приведенные коэффициенты, либо иметь несколько разных оценок, либо совсем не выражаться через них.

Структурный коэффициент называется *идентифицируемым*, если его можно вычислить на основе приведенных коэффициентов, причем *точно идентифицируемым*, если он единственен, и *сверхидентифицируемым*, если имеет несколько разных оценок; в противном случае он называется *неидентифицируемым*.

Какое-либо структурное уравнение называется идентифицируемым, если идентифицируемы все его коэффициенты. Если хотя бы один структурный коэффициент неидентифицируем, то и все уравнение является неидентифицируемым.

Модель считается идентифицируемой, если каждое ее уравнение идентифицируемо. Если хотя бы одно из уравнений системы неидентифицируемо, то вся модель неидентифицируема.

8.2. Методы оценивания структурных уравнений

Рассмотрим различные виды структурных уравнений.

I. Точная идентифицируемость.

Допустим, требуется оценить параметры уравнения функции потребления в простой модели Кейнса формирования доходов:

$$\begin{cases} C_t = \alpha + \beta Y_t + \varepsilon_t & \text{(функция потребления)} \\ Y_t = C_t + I_t & \text{(тождество дохода)} \end{cases},$$

где C_t, Y_t, I_t — объем потребления, совокупный доход, и инвестиции соответственно;

ε_t — случайный член.

Структурный коэффициент β характеризует предельную склонность к потреблению. т. е. из каждой единицы валового внутреннего продукта расходуется β единиц на конечное потребление.

В данной модели C_t, Y_t — эндогенные переменные, а I_t — экзогенная.

Непосредственное оценивание параметров α, β в структурном уравнении функции потребления дает *смещенные* и *несостоятельные* оценки, так как объясняющая переменная Y_t является эндогенной.

Разрешая структурную систему относительно эндогенных переменных, получим приведенную систему:

$$\begin{cases} C_t = \dfrac{\alpha}{1-\beta} + \dfrac{\beta}{1-\beta} I_t + \dfrac{\varepsilon_t}{1-\beta} \\ Y_t = \dfrac{\alpha}{1-\beta} + \dfrac{1}{1-\beta} I_t + \dfrac{\varepsilon_t}{1-\beta}. \end{cases}$$

В приведенной системе коэффициенты при переменной I_t, равные $M_C = \beta/(1-\beta)$ и $M_Y = 1/(1-\beta)$ — *инвестиционные мультипликаторы потребления* и *дохода* соответственно. Это значит, что если объем инвестиций возрастает на единицу, то объем потребления увеличится на $\beta/(1-\beta)$ единиц, а совокупный доход — на $1/(1-\beta)$ единиц.

Косвенный метод наименьших квадратов. Уравнение для C_t в приведенной форме можно также представить в виде

$$C_t = \alpha' + \beta' I_t + \varepsilon'_t,$$

где $\alpha' = \dfrac{\alpha}{1-\beta}$;

$\beta' = \dfrac{\beta}{1-\beta}$;

$\varepsilon'_t = \dfrac{\varepsilon_t}{1-\beta}$.

В этом уравнении экзогенная переменная I_t некоррелирована со случайным членом ε'_t, поэтому для оценки параметров (α', β') можно использовать обычный МНК.

Оцененное уравнение $\widehat{C}_t = \alpha' + \beta' I_t$, полученное по выборочным данным с помощью МНК, дает *несмещенные* и *состоятельные* оценки параметров.

Поскольку оценки (α, β) структурных коэффициентов

$$\alpha = \frac{\alpha'}{1+\beta'}, \quad \beta = \frac{\beta'}{1+\beta'}$$

однозначно выражаются через оценки (α', β') приведенных коэффициентов, то структурное уравнение функции потребления является *точно идентифицируемым.*

Таким образом, для решения точно идентифицируемого уравнения применяется косвенный метод наименьших квадратов (КМНК).

Процедура КМНК производится в несколько этапов.

1. Структурная модель преобразуется в приведенную форму.
2. Для каждого приведенного уравнения обычным МНК оцениваются приведенные коэффициенты.
3. Оценки приведенных коэффициентов преобразуются в оценки параметров структурных уравнений.

Метод инструментальных переменных. Проблема коррелированности объясняющей переменной Y_t со случайным членом ε_t в структурном уравнении для C_t может быть разрешена с помощью метода ИП.

Для применения метода ИП необходимо найти такую инструментальную переменную, которая обладает следующими свойствами:

1) коррелирует с неудачно объясняющей переменной Y_t;

2) не коррелирует со случайным членом ε_t.

В данном случае модель сама предоставляет такую переменную. Величина I_t коррелирует с Y_t, поскольку Y_t зависит от I_t в приведенном уравнении, и I_t не коррелирует с ε_t, поскольку является экзогенной переменной.В общем случае, когда оценка, полученная косвенным методом, единственна, она совпадает с оценкой, полученной методом ИП. Поэтому КМНК можно рассматривать как частный случай метода ИП.

Структурное уравнение модели, в которой число экзогенных переменных, которые могут использоваться как инструментальные, равно числу объясняющих эндогенных переменных, является *точно идентифицируемым.*

Пример 8.1. Для некоторой страны имеются данные (усл. ед.) о совокупном доходе Y, объеме потребления C и инвестициях I, полученные за 10 лет.

C_t	190	198	200	180	200	210	220	210	205	210
I_t	10	20	30	20	10	20	30	20	15	30
Y_t	200	218	230	200	210	230	250	230	220	240

Построить функцию потребления, используя модель Кейнса формирования доходов:

$$\begin{cases} C_t = \alpha + \beta Y_t + \varepsilon_t & (\text{функция потребления}) \\ Y_t = C_t + I_t & (\text{тождество дохода}) \end{cases}.$$

▼ В данной модели C_t, Y_t — эндогенные переменные, а I_t — экзогенная.

Первое уравнение является идентифицируемым, поскольку переменная I_t не включена в него и может выступать как инструментальная для Y_t.

Второе уравнение представляет собой тождество, параметры которого известны, поэтому необходимости в его идентификации нет.

Непосредственное оценивание структурного уравнения функции потребления обычным МНК приводит к следующему:

$$\hat{C} = 60{,}9 + 0{,}635Y,$$

где оценки $\alpha = 60{,}9$; $\beta = 0{,}635$ — смещенные и несостоятельные.

Оцененная система приведенных уравнений:

$$\hat{C} = 188 + 0{,}695\,I;$$

$$\hat{Y} = 188 + 1{,}695\,I.$$

Выразив I из второго уравнения системы в виде $I = (Y - 188)$: : 1,695 и подставив его в первое, получим

$$\hat{C} = 188 + \frac{0,695}{1,695}(Y - 188) = 110,9 + 0,41Y,$$

где оценки $\alpha = 110,9$; $\beta = 0,41$ — несмещенные и состоятельные.

Коэффициент $\beta = 0,41$ определяет склонность к потреблению, а $M_C = 0,695$ и $M_Y = 1,695$ — инвестиционные мультипликаторы потребления и дохода соответственно. Это значит, что если объем инвестиций возрастает на единицу, то объем потребления увеличится на 0,695 единиц, а совокупный доход — на 1,695 единиц. ▲

Пример 8.2. По данным 15 торговых предприятий получены сведения о показателях, характеризующих объем продаж, интенсивность рекламы и динамику цен: y_1 — объем продаж, млн. руб.; y_2 — число рекламных сообщений; x_1 — индекс цен на продукцию, %; x_2 — индекс цен на рекламу.

Пусть исходная модель имеет вид:

$$\begin{cases} y_{1t} = \alpha_{10} + b_{11}y_{2t} + \alpha_{11}x_{1t} + \varepsilon_{1t} \\ y_{2t} = \alpha_{20} + b_{21}y_{1t} + \alpha_{21}x_{2t} + \varepsilon_{2t} \end{cases}.$$

На основании данных (усл. ед.) оцените структурную модель

№	y_{1t}	y_{2t}	x_{1t}	x_{2t}	№	y_{1t}	y_{2t}	x_{1t}	x_{2t}
1	56,7	270	104,3	97,8	9	69,7	409	100,3	102
2	64,5	172	94,2	105,7	10	46,4	191	105	101,9
3	53,3	324	102,8	103,3	11	53,5	231	105,6	106,5
4	82,6	428	98,7	95,1	12	41,2	131	106,2	107,7
5	62	420	99,8	100,5	13	45,6	115	110,3	109,1
6	61,3	473	100,5	101,4	14	48,3	202	105,9	102,9
7	25,7	82	112,8	110,1	15	56,1	223	105,8	100,2
8	36,1	276	106,7	100,7					

▼ В исходной модели y_1, y_2 — эндогенные переменные, а x_1, x_2 — экзогенные. Первое уравнение является идентифицируемым, поскольку переменная x_2 не включена в него и может выступать как инструментальная для y_2. Аналогично второе уравнение также является идентифицируемым.

Для оценки структурных коэффициентов используем КМНК. Оценивая систему приведенных уравнений, получим

$$\begin{cases} \hat{y}_{1t} = 369{,}634 - 2{,}013 \cdot x_{1t} - 1{,}038 \cdot x_{2t} \\ y_{2t} = 2887{,}483 - 9{,}259 \cdot x_{1t} - 16{,}138 \cdot x_{2t} \end{cases},$$

где коэффициенты уравнений статистически значимы.

Выражая переменную x_2 из второго уравнения и подставляя ее в первое, а переменную x_1 из первого уравнения и подставляя ее во второе уравнение, получим:

$$\begin{cases} \hat{y}_{1t} = 183{,}933 + 0{,}064 \cdot y_{2t} - 1{,}418 \cdot x_{1t} \\ \hat{y}_{2t} = 1187{,}368 + 4{,}599 \cdot y_{1t} - 11{,}364 \cdot x_{2t} \end{cases}$$ ▲

Пример 8.3. Рассматривается модель

$$\begin{cases} y_1 = \beta_{12} y_2 + \alpha_{11} x_1 + \varepsilon_1 \\ y_2 = \beta_{21} y_1 + \alpha_{22} x_2 + \varepsilon_2 \end{cases},$$

где y_1 – годовое потребление свинины на душу населения;

y_2 – оптовая цена;

x_1 – доход на душу населения;

x_2 – расходы по обработке мяса.

На основании данных таблицы (усл. ед.) оцените структурную модель.

t	y_1	y_2	x_1	x_2
1	60	5	1300	60
2	62	4	1300	56
3	65	4,2	1500	56
4	62	5	1600	63
5	66	3,8	1800	50
Среднее	63	4,4	1500	57

▼ В исходной модели y_1, y_2 — эндогенные переменные, а x_1, x_2 — экзогенные. Первое уравнение является идентифицируемым, поскольку переменная x_2 не включена в него и может выступать как инструментальная для y_2. Аналогично второе уравнение также является идентифицируемым.

Для оценки структурных коэффициентов используем КМНК.

Структурная форма модели преобразуется в приведенную форму:

$$\begin{cases} y_1 = \delta_{11}x_1 + \delta_{12}x_2 + u_1 \\ y_2 = \delta_{21}x_1 + \delta_{22}x_2 + u_2 \end{cases}.$$

При расчете параметров приведенной формы модели преобразуем исходные данные в центрированные, т. е.

t	y_1	y_2	x_1	x_2
1	–3	0,6	–200	3
2	–1	–0,4	–200	–1
3	2	–0,2	0	–1
4	–1	0,6	100	6
5	3	–0,6	300	–7
Итого	0	0,0	0	0

Оцененная приведенная форма модели имеет вид

$$\begin{cases} \hat{y}_1 = 0{,}00609 \cdot x_1 - 0{,}26481 \cdot x_2 \\ \hat{y}_2 = 0{,}00029 \cdot x_1 + 0{,}11207 \cdot x_2 \end{cases}.$$

Выражая переменную x_2 из второго уравнения и подставляя ее в первое, а переменную x_1 из первого уравнения и подставляя ее во второе, получим следующую структурную форму модели:

$$\begin{cases} \hat{y}_1 = -2{,}362 \cdot y_2 + 0{,}0068 \cdot x_1 \\ \hat{y}_2 = 0{,}048 \cdot y_1 + 0{,}1247 \cdot x_2 \end{cases}.$$

Если в структурную модель включить свободные члены, т. е.

$$\begin{cases} y_1 = \alpha_{01} + \beta_{12}y_2 + \alpha_{11}x_1 + \varepsilon_1 \\ y_2 = \alpha_{02} + \beta_{21}y_1 + \alpha_{22}x_2 + \varepsilon_2 \end{cases},$$

то свободные члены уравнений определим по формулам:

$$a_{01} = \bar{y}_1 - b_{12}\bar{y}_2 - a_{11}\bar{x}_1 = 63{,}19$$
$$a_{02} = \bar{y}_2 - b_{21}\bar{y}_1 - a_{22}\bar{x}_2 = -5{,}73.$$

Тогда структурная модель имеет вид:

$$\begin{cases} \hat{y}_1 = 63{,}19 - 2{,}362 \cdot y_2 + 0{,}0068 \cdot x_1 \\ \hat{y}_2 = -5{,}73 + 0{,}048 \cdot y_1 + 0{,}1247 \cdot x_2 \end{cases} \quad \blacktriangle$$

II. Сверхидентифицированность

Рассмотрим следующую простую модель Кейнса формирования доходов:

$$\begin{cases} C_t = \alpha + \beta Y_t + \varepsilon_t & (\text{функция потребления}) \\ Y_t = C_t + I_t + G_t & (\text{тождество дохода}) \end{cases},$$

где G_t — объем государственных расходов.

В исходной модели C_t, Y_t — эндогенные переменные, а I_t, G_t — экзогенные. Обе экзогенные переменные I_t, G_t не присутствуют в структурном уравнении функции потребления и могут использоваться как инструментальные для эндогенной переменной Y.

Структурное уравнение модели, в которой число экзогенных переменных, которые могут использоваться как инструментальные, больше, чем необходимо, является *сверхидентифицируемым.*

Наилучшим решением в данном случае является применение двухшагового метода наименьших квадратов (ДМНК) и построение инструментальной переменной, которая является комбинацией I_t, G_t.

Процедура ДМНК производится в несколько этапов.

1. На основе приведенной формы модели получают для сверхидентифицированного уравнения теоретические (расчетные) значения эндогенных переменных, содержащихся в правой части уравнения.

2. Подставляя теоретические значения эндогенных переменных вместо их фактических значений в сверхидентифицируемое уравнение и применяя обычный МНК, определяют его структурные коэффициенты.

ДМНК можно рассматривать как способ конструирования наилучшей из возможных комбинаций инструментальных переменных в случае, когда в уравнении имеется избыток экзогенных переменных, которые можно использовать как инструментальные для объясняющей эндогенной переменной.

Пример 8.4. Для некоторой страны имеются данные (усл. ед.) о совокупном доходе Y, объеме потребления C, инвестициях I и государственных расходах G, полученные за 10 лет:

C_t	195	203	210	200	215	215	210	215	225	220
I_t	10	20	30	20	10	20	30	20	15	30
G_t	20	10	20	40	30	10	20	10	40	20
Y_t	225	233	260	260	255	245	260	245	280	270

Построить функцию потребления, используя модель Кейнса формирования доходов:

$$\begin{cases} C_t = \alpha + \beta Y_t + \varepsilon_t & \text{(функция потребления)} \\ Y_t = C_t + I_t + G_t & \text{(тождество дохода)} \end{cases}.$$

▼Исходные данные и расчетные показатели представим в таблице.

C_t	I_t	G_t	Y_t	$\hat{Y}_t$
195	10	20	225	237,486
203	20	10	233	238,953
210	30	20	260	263,258
200	20	40	260	273,210
215	10	30	255	248,905
215	20	10	245	238,953
210	30	20	260	263,258
215	20	10	245	238,953
225	15	40	280	266,767
220	30	20	270	263,258

В данной модели C_t, Y_t — эндогенные переменные, а I_t , G_t — экзогенные.

Непосредственное оценивание структурного уравнения функции потребления обычным МНК приводит к следующим результатам:

$$\hat{C} = 109{,}8 + 0{,}4Y,$$

где оценки $\alpha = 109{,}8$; $\beta = 0{,}4$ — смещенные и несостоятельные.

Структурное уравнение для функции потребления является сверхидентифицируемым, поскольку переменные I_t, G_t не включены в него и могут выступать как инструментальные для Y_t. Для оценки параметров сверхидентифицируемого уравнения используем ДМНК.

Оценивая приведенное уравнение для совокупного дохода

$$Y_t = \gamma_0 + \gamma_1 I_t + \gamma_2 G_t + \varepsilon_t$$

обычным МНК, получим

$$\hat{Y} = 201{,}7 + 1{,}29 \cdot I + 1{,}14 \cdot G .$$

Подставив расчетные значения $\hat{Y}$ вместо фактических значений в структурное уравнение функции потребления, получим уравнение

$$C_t = \alpha + \beta \hat{Y}_t + \varepsilon_t,$$

и оценка полученного уравнение обычным МНК дает

$$\hat{C} = 171{,}3 + 0{,}156\, Y ,$$

где оценки $\alpha = 171{,}3$; $\beta = 0{,}156$ — состоятельны. ▲

Пример 8.5. Имеются поквартальные данные по РФ (1995–2005 гг., трлн. руб.) об объемах валового внутреннего продукта (ВВП), расходов на конечное потребление (КП), валового накопления (ВН) и чистого экспорта (Э) в среднегодовых ценах 1995 г. Оцените следующую структурную модель:

$$\begin{cases} КП_t = \alpha_1 + \beta_{11} ВВП_t + \varepsilon_{1t} \\ ВН_t = \alpha_2 + \beta_{21} ВВП_{t-4} + \varepsilon_{2t} \\ ВВП_t + КП_t + ВН_t + Э_t \end{cases},$$

где $ВВП_{t-4}$ – объем $ВВП$ за аналогичный квартал предыдущего года.

▼Исходные данные и расчетные показатели представим в виде таблицы.

Квартал	*t*	$ВВП_t$	$КП_t$	$ВН_t$	$Э_t$	$ВВП_{t-4}$	$\widehat{ВВП}_t$
1	1	329	260	43	26		
2	2	341	248	73	20		
3	3	395	266	124	5		
4	4	360	251	105	4		

Квартал	t	$ВВП_t$	$КП_t$	$ВН_t$	$Э_t$	$ВВП_{t-4}$	$\widehat{ВВП}_t$
1	5	322	257	55	10	329	322
2	6	329	254	64	11	341	335
3	7	373	244	116	13	395	393
4	8	349	241	86	22	360	361
1	9	320	255	52	13	322	316
2	10	327	258	63	6	329	320
3	11	385	267	113	5	373	366
4	12	362	278	78	6	349	341
1	13	316	247	73	–4	320	306
2	14	323	237	87	–1	327	314
3	15	350	270	62	18	385	385
4	16	329	275	–7	61	362	383
1	17	311	244	24	43	316	325
2	18	333	232	51	50	323	336
3	19	390	242	89	59	350	369
4	20	369	245	41	83	329	359
1	21	346	222	40	84	311	340
2	22	367	222	61	84	333	364
3	23	432	253	100	79	390	421
4	24	399	250	85	64	369	392
1	25	346	235	46	65	346	368
2	26	364	242	71	51	367	383
3	27	424	261	116	47	432	450
4	28	400	265	101	34	399	408
1	29	363	272	53	38	346	354
2	30	382	269	69	44	364	376
3	31	449	289	111	49	424	442
4	32	418	287	88	43	400	414
1	33	377	272	50	55	363	381
2	34	399	280	74	45	382	396
3	35	469	296	127	46	449	467
4	36	445	300	100	45	418	434
1	37	405	298	58	49	377	392

Квартал	t	$ВВП_t$	$КП_t$	$ВН_t$	$Э_t$	$ВВП_{t-4}$	$\widehat{ВВП}_t$
2	38	430	296	80	54	399	418
3	39	498	316	122	60	469	495
4	40	478	306	113	59	445	469
1	41	435	314	61	60	405	428
2	42	463	309	84	70	430	459
3	43	534	336	127	71	498	531
4	44	509	318	128	63	478	506

В структурной модели три эндогенные переменные ($КП_t$, $ВН_t$, $ВВП_t$) и две экзогенные переменные ($Э_t$, $ВВП_{t\text{-}4}$).

Первое структурное уравнение является сверхидентифицируемым, поскольку переменные $Э_t$, $ВВП_{t\text{-}4}$ не включены в него и могут выступать как инструментальные для $ВВП_t$. Для оценки параметров сверхидентифицируемого уравнения используем *ДМНК*.

Оценивая приведенное уравнение

$$ВВП_t = \gamma_0 + \gamma_1 Э_t + GВВП_{t\text{-}3} + \varepsilon_t$$

обычным МНК, получим

$$\widehat{ВВП}_t = -30{,}66 + 0{,}50 \cdot Э_t + 1{,}06 \cdot ВВП_{t-4}\,.$$

Подставив расчетные значения $\widehat{ВВП}_t$ вместо фактических значений в первое структурное уравнение функции $КП_t$, получим уравнение

$$КП_t = \alpha_1 + \beta_{11}\widehat{ВВП}_t + \varepsilon_{1t}\,.$$

и оценка полученного уравнение обычным МНК дает

$$\widehat{КП}_t = 110{,}72 + 0{,}40 \cdot ВВП_t\,.$$

Поскольку второе структурное уравнение не содержит эндогенных переменных в качестве факторов, то для оценки его параметров используем обычный МНК и в результате получим

$$\widehat{ВН}_t = -103{,}66 + 0{,}48 \cdot ВВП_{t-4}\,.$$

Третье структурное уравнение-тождество не содержит неизвестных параметров.

Таким образом, получены следующие параметры структурной модели:

$$\begin{cases} \widehat{КП}_t = 110{,}72 + 0{,}40 \cdot ВВП_t \\ \widehat{ВН}_t = -103{,}66 + 0{,}48 \cdot ВВП_{t-4}, \\ ВВП_t = КП_t + ВН_t + Э_t \end{cases}$$

где все коэффициенты модели статистически значимы. ▲

III. Неидентифицируемость

Рассмотрим следующую модель спроса и предложения:

$$\begin{cases} y^D = \alpha + \beta P + u^D & \text{(спрос)} \\ y^S = \delta + \varepsilon P + u^S & \text{(предложение)}, \\ y^D = y^S = y & \text{(равновесие)} \end{cases}$$

где P — цена товара;

u^D, u^S — случайные члены.

Переменные Y, P являются эндогенными, и их значения определяются в процессе установления равновесия.

Структурное уравнение модели, в которой число экзогенных переменных, которые могут использоваться как инструментальные меньше числа объясняющих эндогенных переменных, является *неидентифицируемым.*

В рассматриваемой модели нет экзогенных переменных, поэтому ни одно из этих уравнений *не является идентифицируемым.*

Чтобы модель имела статистическое решение, в нее вводятся экзогенные переменные.

Предположим, что продавцы товара облагаются специальным налогом T, который они должны платить с выручки. При этом уравнение спроса останется неизменным, если переменная P означает рыночную цену, а уравнение предложения изменится:

$$\begin{cases} y^D = \alpha + \beta P + u^D & \text{(спрос)} \\ y^S = \delta + \varepsilon P + \sigma T + u^S & \text{(предложение)}, \\ y^D = y^S = y & \text{(равновесие)} \end{cases}$$

где T — экзогенная переменная.

Уравнение спроса будет идентифицируемым, поскольку переменная T не включена в него и может выступать как инструментальная для P, а уравнение предложения — неидентифицируемым.

Включим в уравнение спроса экзогенную переменную x — доход на душу населения:

$$\begin{cases} y^D = \alpha + \beta P + \gamma x + u^D (\text{спрос}) \\ y^S = \delta + \varepsilon P + \sigma T + u^S \ (\text{предложение}). \\ y^D = y^S = y \qquad (\text{равновесие}) \end{cases}$$

Экзогенную переменную x можно использовать как инструментальную вместо P для уравнения предложения.

В итоге получили в целом точно идентифицируемую модель спроса и предложения.

Пусть структурное уравнение спроса имеет временной тренд (скажем, потому что привычки медленно меняются со временем):

$$\begin{cases} y^D = \alpha + \beta P + \gamma x + \rho t + u^D \ (\text{спрос}) \\ y^S = \delta + \varepsilon P + \sigma T + u^S \qquad (\text{предложение}), \\ y^D = y^S = y \qquad (\text{равновесие}) \end{cases}$$

где t — переменная времени;

ρ — коэффициент при ней.

В модели спроса имеются две экзогенные переменные x, t, которые можно использовать в качестве инструментальных для P в уравнении предложения.

В итоге получили сверхидентифицируемое уравнение предложения и точно идентифицируемое уравнение спроса.

Пример 8.6. Имеется следующая модель «спрос – предложение»:

$$\begin{cases} y = \alpha + \beta P + \gamma x + u_1 \ (\text{спрос}) \\ y = \delta + \varepsilon P + u_2 \qquad (\text{предложение}), \end{cases}$$

где y, P —эндогенные переменные: количество товара и цена;

x – экзогенная переменная: доход потребителей.

На основании данных (усл. ед.) оцените структурные уравнения.

Y	5	4	8	5	9	7	11
P	2	3	4	3	7	5	6
X	1	2	3	4	5	6	7

▼ В исходной модели y, P — эндогенные переменные, а x — экзогенная.

Уравнение спроса является неидентифицируемым, а уравнение предложения — идентифицируемым, поскольку переменная x не включена в него и может выступать как инструментальная для P.

Оценивая систему приведенных уравнений, получим

$$\widehat{P} = 1{,}571 + 0{,}678\,x;$$

$$\widehat{Y} = 3{,}428 + 0{,}892\,x.$$

Выразив x из первого уравнения системы в виде $x = \dfrac{P - 1{,}571}{0{,}678}$ и подставив его во второе, получим структурное уравнение предложения

$$\hat{Y} = 3{,}428 + 0{,}892 \cdot \frac{P - 1{,}571}{0{,}678} = 1{,}362 + 1{,}315P,$$

где оценки $\delta = 1{,}362$; $\varepsilon = 1{,}315$.

Функция спроса неидентифицируема. ▲

Упражнение 8.1. Имеется следующая модель «спрос – предложение»:

$$\begin{cases} y = \alpha + \beta P + \gamma x + u_1 & (\text{спрос}) \\ y = \delta + \varepsilon P + \sigma I + u_2 & (\text{предложение}), \end{cases}$$

где y, P – эндогенные переменные: количество некоторого товара и цена;

x, I – экзогенные переменные: доход на душу населения и инвестиции в производство.

На основании данных (усл. ед.) оцените структурные уравнения.

Y	P	X	I
20	3	34	5
33	3	43	6
28	5	51	6
41	4	49	7
40	5	55	7
36	6	62	6
42	6	70	8
38	7	68	8
51	7	78	12

Ответ. $\begin{cases} \hat{y} = -2{,}743 - 12{,}667P + 1{,}842\,x \ \text{(спрос)} \\ \hat{y} = 7{,}297 + 1{,}843P + 2{,}742I \ \text{(предложение)} \end{cases}$ ▲

Упражнение 8.2. Имеется следующая модель «спрос – предложение»:

$$\begin{cases} y = \alpha + \beta P + u_1 \quad \text{(спрос)} \\ y = \delta + \varepsilon P + \sigma W + u_2 \ \text{(предложение)}, \end{cases}$$

где y, P — количество товара и цена;

W — зарплата потребителей.

На основании данных (усл. ед.) оцените структурные уравнения.

Y	10	12	10	8	9	7	6
P	7	6	9	7	8	11	10
W	1	2	3	4	5	6	7

Ответ. Структурное уравнение спроса

$$\hat{Y} = 19{,}45 - 1{,}28P,$$

а функция предложения неидентифицируема. ▲

8.3. Ненулевое ограничение

Добавление экзогенной переменной не единственный способ, который может привести к идентифицируемости уравнения. В некоторых случаях неидентифицируемая модель может быть идентифицируема путем задания соотношения между структурными коэффициентами.

Рассмотрим неидентифицируемую модель спроса и предложения

$$\begin{cases} y^D = \alpha + \beta P + u^D & (\text{спрос}) \\ y^S = \delta + \varepsilon P + \sigma T + u^S & (\text{предложение}) \\ y^D = y^S = y & (\text{равновесие}) \end{cases}.$$

Улучшим спецификацию модели, введя ограничения $\sigma = -\varepsilon$:

$$\begin{cases} y^D = \alpha + \beta P + u^D & (\text{спрос}) \\ y^S = \delta + \varepsilon(P - T) + u^S & (\text{предложение}) \\ y^D = y^S = y & (\text{равновесие}) \end{cases}. \qquad (8.1)$$

Благодаря введению ограничения на коэффициенты $\sigma = -\varepsilon$ уравнение предложения также стало идентифицируемым.

Действительно, при использовании ИП можно рассмотреть новую версию модели как систему из четырех уравнений:

$$\begin{cases} y^D = \alpha + \beta P + u^D \\ y^S = \delta + \varepsilon P_1 + u^S \\ P_1 = P - T \\ y^D = y^S \end{cases}, \qquad (8.2)$$

где P_1 — цена товара для продавца (сумма, остающаяся у него после уплаты налога).

Последние два уравнения системы являются уравнениями-тождествами и не требуют проверки на идентификацию.

Переменная T не включена в уравнение спроса, поэтому она может использоваться как инструментальная для P. Точно так же эта переменная не включена в уравнение предложения, поэтому она может использоваться как инструментальная для P_1.

В итоге модель в целом является *точно идентифицируемой.*

Вывод. Ограничения на коэффициенты позволяют исключить одну объясняющую переменную из уравнения. Если эта переменная *эндогенная*, для нее не нужно искать инструментальную переменную, если *экзогенная*, то она освобождается на роль инструментальной для одной из эндогенных переменных, оставшихся в уравнении.

Пример 8.7. Оценить структурную модель «спрос – предложение» (7.1) по исходным данным.

T	0	2	5	8	10	12	14
P	40	42	43	44	45	48	49
Y	70	68	63	61	60	56	52

▼Рассмотрим как исходную (8.1) так и новую (8.2) версии модели.

Было показано, что исходная модель точно идентифицируема, и поэтому для оценки ее структурных коэффициентов используем КМНК.

Оцененные уравнения приведенной системы, полученные по выборочным данным обычным МНК:

$$\begin{cases} \widehat{P} = 40 + 0{,}6T \\ \widehat{Y} = 70 - 1{,}2T \end{cases}, \text{ или } \begin{cases} \widehat{P}_1 = P - T = 40 - 0{,}4T \\ \widehat{Y} = 70 - 1{,}2T \end{cases}.$$

Перейти от приведенной формы модели к структурной можно следующим образом. Выразив T из первого уравнения приведенной формы в виде $T = \dfrac{P - 40}{0{,}6}$ и подставив его во второе, получим

$$\widehat{Y} = 70 - 1{,}2 \cdot \frac{P - 40}{0{,}6} = 150 - 2P,$$

т. е. $\alpha = 150$, $\beta = -2$.

Выразив T из первого уравнения приведенной формы в виде $T = \dfrac{40 - P_1}{0{,}4}$ и подставив его во второе, получим

$$\widehat{Y} = 70 - 1{,}2 \cdot \frac{40 - P_1}{0{,}4} = -50 + 3P_1,$$

т. е. $\delta = -50$, $\varepsilon = 3$.

Таким образом, получили следующую оцененную модель «спрос – предложение»:

$$\begin{cases} \widehat{Y} = 150 - 2P & (\text{спрос}) \\ \widehat{Y} = -50 + 3(P - T) & (\text{предложение}) \end{cases} \blacktriangle$$

Пример 8.8. Оценить структурную модель вида

$$\begin{cases} y = \alpha_1 + \beta_1(C + D) + \varepsilon_1 \\ C = \alpha_2 + \beta_2 y + \beta_3 y_{-1} + \varepsilon_2 \end{cases},$$

где y — валовой доход;

y_{-1} — валовой доход предшествующего года;

C — личное потребление;

D — конечный спрос (помимо личного потребления);

$\varepsilon_1, \varepsilon_2$ — случайные члены.

Исходные данные (усл. ед.) следующие:

t	*y*	*C*	*D*	y_{-1}
1	3	7	5	47
2	23	30	23	3
3	8	1	2	23
4	21	9	12	8
5	18	25	6	21
6	37	8	45	18
7	36	30	24	37
8	47	32	50	36
9	56	40	32	47
10	32	26	20	56

▼В данной модели две эндогенные переменные (y, C) и две экзогенные переменные (D, y_{-1}).

Учитывая ограничение на коэффициенты в первом уравнении, а также, что переменная y_{-1} на включена в это уравнение, переменные D, y_{-1} могут выступать как инструментальные для C_t, следовательно, первое уравнение сверхидентифицировано. Переменная D не включена во второе уравнение, поэтому она может использоваться как инструментальная для y, т. е. второе уравнение точно идентифицировано.

Оценивая систему приведенных уравнений, получим

$$\begin{cases} \hat{y} = 4{,}146 + 0{,}8249 \cdot D + 0{,}1989 \cdot y_{-1} \\ \hat{C} = 7{,}328 + 0{,}3521 \cdot D + 0{,}1946 \cdot y_{-1} \end{cases}.$$

Для определения параметров сверхидентифицированного уравнения используем двухшаговый МНК.

По сверхидентифицированному уравнению структурной формы модели заменяем фактические значения C теоретическим $\hat{C}$ и рассчитываем новую переменную $\hat{C} + D$ (см. таблицу).

Y	$\hat{C}$	D	$\hat{C} + D$
3	18,235	5	23,235
23	16,010	23	39,010
8	12,508	2	14,508
21	13,110	12	25,110
18	13,527	6	19,527
37	26,675	45	71,675
36	22,979	24	46,979
47	31,939	50	81,939
56	27,741	32	59,741
32	25,268	20	45,268

Оценивая полученное структурное уравнение обычным МНК, получим

$$\hat{y} = 1{,}443 + 0{,}624 \cdot (C + D).$$

Для определения параметров точно идентифицированного второго уравнения используем косвенный МНК.

Из первого уравнения приведенной системы выразим D в виде $D = \dfrac{y - 4{,}146 - 0{,}1989 \cdot y_{-1}}{0{,}8249}$ и, подставив его во второе, получим

$$\hat{C} = 5{,}559 + 0{,}427 \cdot y + 0{,}110 \cdot y_{-1}.$$

Таким образом, получили следующую оцененную структурную модель:

$$\begin{cases} \hat{y} = 1{,}443 + 0{,}624 \cdot (C + D) \\ \hat{C} = 5{,}559 + 0{,}427 \cdot y + 0{,}110 \cdot y_{-1} \end{cases} \blacktriangle$$

Пример 8.9. Оценить следующую идентифицируемую эконометрическую модель с двумя эндогенными (y_1, y_2) и двумя экзогенными (x_1, x_2) переменными:

$$\begin{cases} y_1 = \alpha_1 + \beta_{12} y_2 + \alpha_{11} x_1 + \varepsilon_1 \\ y_2 = \alpha_2 + \beta_{21} y_1 + \alpha_{22} x_2 + \varepsilon_2 \end{cases}$$

по следующим выборочным данным (усл. ед.):

y_1	y_2	x_1	x_2
2	5	1	3
3	6	2	1
4	7	3	2
5	8	2	5
6	5	4	6

▼Для точно идентифицируемой структурной модели применим КМНК.

Приведенная форма модели:

$$\begin{cases} y_1 = \delta_1 + \delta_{11} x_1 + \delta_{12} x_2 + \nu_1 \\ y_2 = \delta_2 + \delta_{21} x_1 + \delta_{22} x_2 + \nu_2 \end{cases}.$$

Оцененные уравнения приведенной системы, полученные по выборочным данным с использованием МНК:

$$\begin{cases} \hat{y}_1 = 0{,}685 + 0{,}8524 x_1 + 0{,}3733 x_2 \\ \hat{y}_2 = 6{,}394 - 0{,}0724 x_1 - 0{,}0056 x_2 \end{cases}.$$

Перейдем от приведенной формы модели к структурной.

Для этой цели из первого уравнения приведенной формы модели надо исключить x_2, выразив его из второго уравнения:

$x_2 = \dfrac{6{,}394 - 0{,}0724x_1 - y_2}{0{,}0056}$, и подставить в первое, а из второго уравнения приведенной формы модели следует исключить x_1, выразив его из первого уравнения приведенной формы:

$x_1 = \dfrac{y_1 - 0{,}685 - 0{,}3733x_2}{0{,}8524}$, и подставить во второе.

В результате получим следующую структурную форму модели:

$$\begin{cases} y_1 = 426{,}91 - 66{,}96\,y_2 - 3{,}97\,x_1 + \varepsilon_1 \\ y_2 = 6{,}45 - 0{,}085\,y_1 + 0{,}026\,x_2 + \varepsilon_2 \end{cases}.$$

Покажем, что для точно идентифицируемых уравнений ДМНК дает тот же результат, что и КМНК.

Из приведенной системы уравнений можно найти расчетные значения эндогенных переменных $\hat{y}_1, \hat{y}_2$. Подставляя их вместо фактических y_1, y_2 в правую часть структурной модели и применяя обычный МНК к каждому уравнению модели, получим тот же результат, что и при КМНК.

Расчетные данные для использования ДМНК приведены в таблице:

y_1	$\hat{y}_2$	x_1		y_2	$\hat{y}_1$	x_2
2	6,303	1		5	2,657	3
3	6,242	2		6	2,763	1
4	6,164	3		7	3,989	2
5	6,220	2		8	4,256	5
6	6,070	4		5	6,334	6

Пусть в исходной идентифицируемой модели наложены ограничения на ее параметры $\beta_{12} = \alpha_{11}$, тогда придем к новой модели:

$$\begin{cases} y_1 = \alpha_1 + \beta_{12}(y_2 + x_1) + \varepsilon_1 \\ y_2 = \alpha_2 + \beta_{21}y_1 + \alpha_{22}x_2 + \varepsilon_2 \end{cases}.$$

В результате второе (точно идентифицируемое) уравнение не изменилось, следовательно, структурные коэффициенты не изменятся, а первое уравнение стало сверхидентифицируемым.

Для определения структурных коэффициентов первого, сверхидентифицируемого уравнения новой системы используем ДМНК.

Для этого на основе второго уравнения приведенной системы находим расчетные значения $\widehat{y}_2$ эндогенной переменной. Подставляя их вместо фактических y_2 в первое уравнение полученной системы и применяя обычный МНК, получим решение поставленной задачи.

Исходные данные при использовании ДМНК следующие:

y_1	x_1	$\widehat{y}_2$	$\widehat{y}_2 + x_1$
2	1	6,303	7,303
3	2	6,242	8,242
4	3	6,146	9,164
5	2	6,220	8,220
6	4	6,070	10,070

Окончательно рассматриваемая структурная система уравнений:

$$\begin{cases} y_1 = -6{,}693 + 1{,}244(y_2 + x_1) \\ y_2 = 6{,}45 - 0{,}085\,y_1 + 0{,}026\,x_2 \end{cases} \blacktriangle$$

8.4. Условия для идентификации

Приступать к оцениванию того или иного структурного уравнения системы имеет смысл после того, как установлено его идентифицируемость.

Для установления идентифицируемости можно использовать инструментальные переменные (ИП).

В полностью определенной модели будет столько уравнений, сколько имеется эндогенных переменных.

Пусть D — число не включенных в уравнение, но присутствующих в системе экзогенных переменных, а H — число включенных в уравнение эндогенных переменных.

Необходимое условие идентификации. Уравнение в структурной модели может быть идентифицировано, если число не включенных в него экзогенных переменных не меньше числа включенных в него объясняющих эндогенных переменных, т. е.

$$\mathrm{D} \geq \mathrm{H} - 1 \quad (\textit{порядковое условие}).$$

Данное условие является *необходимым*, но не *достаточным* для идентификации.

В частности:

- если $D = H - 1$, то уравнение *точно идентифицируемо*;
- если $D > H - 1$, то уравнение *сверхидентифицируемо*;
- если $D < H - 1$, то уравнение *неидентифицируемо*.

Достаточное условие идентификации. Уравнение идентифицируемо, если ранг матрицы, составленной из коэффициентов при переменных (эндогенных и экзогенных), отсутствующих в исследуемом уравнении, не меньше $N - 1$, где N — число эндогенных переменных системы.

Пример 8.10. Проверим на идентификацию каждое уравнение модели:

$$\begin{cases} y_1 = \alpha_{01} + \beta_{13}y_3 + \beta_{14}y_4 + \varepsilon_1 \\ y_2 = \alpha_{02} + \beta_{23}y_3 + \alpha_{21}x_1 + \varepsilon_2 \\ y_3 = \alpha_{03} + \beta_{34}y_4 + \alpha_{31}x_1 + \varepsilon_3 \\ y_4 = y_1 + y_2 + x_2 \end{cases},$$

где y_1 — расходы на конечное потребление текущего года;

y_2 — валовые инвестиции в текущем году;

y_3 — расходы на заработную плату в текущем году;

y_4 — валовой доход за текущий год;

x_1 — валовой доход предыдущего года;

x_2 — государственные расходы текущего года;

$\varepsilon_1, \varepsilon_2, \varepsilon_3$ — случайные ошибки.

В данной модели четыре эндогенные переменные (y_1, y_2, y_3, y_4), т. е. $N = 4$, и две экзогенные (x_1, x_2).

Первое уравнение: $H = 3$ (y_1, y_3, y_4 присутствуют), $D = 2$ (x_1, x_2 отсутствуют) и $D = H - 1$, поэтому уравнение точно идентифицируемо (необходимое условие).

Для проверки на достаточное условие идентификации выпишем матрицу A коэффициентов при переменных, не входящих в первое уравнение:

Уравнение	y_2	x_1	x_2
2	-1	α_{21}	0
3	0	α_{31}	0
4	1	0	1

Определитель матрицы $\det A = -\alpha_{31} \neq 0$, следовательно, ранг матрицы равен $3 \geq N - 1$, т. е. достаточное условие идентификации выполняется, и первое уравнение точно идентифицируемо.

Второе уравнение: $H = 2$ (y_2, y_3 присутствуют), $D = 1$ (x_2 отсутствует) и $D = H - 1$, поэтому уравнение точно идентифицируемо (необходимое условие).

Выпишем матрицу A коэффициентов при переменных, не входящих во второе уравнение:

Уравнение	y_1	y_4	x_2
1	–1	β_{14}	0
3	0	β_{34}	0
4	1	–1	1

Выполняется также достаточное условие идентификации: $\det A = -\beta_{34} \neq 0$, ранг матрицы равен $3 \geq N - 1$.

Третье уравнение: $H = 2$ (y_3, y_4 присутствуют), $D = 1$ (x_2 отсутствует) и $D = H - 1$, поэтому уравнение точно идентифицируемо (необходимое условие).

Выпишем матрицу A коэффициентов при переменных, не входящих в третье уравнение:

Уравнение	y_1	y_2	x_2
1	–1	0	0
2	0	–1	0
4	1	1	1

Здесь также выполняется достаточное условие идентификации: $\det A = 1$, ранг матрицы равен $3 \geq N - 1$.

Четвертое уравнение представляет собой тождество, параметры которого известны, поэтому необходимости в его идентификации нет.

Таким образом, все уравнения модели точно идентифицированы. ▲

Пример 8.11. Выполним идентификацию следующей модели:

$$\begin{cases} C_t = \alpha_{01} + \beta_{11} Y_t + \alpha_{12} C_{t-1} + \varepsilon_1 & \text{(функция потребления)} \\ I_t = \alpha_{02} + \beta_{21} r_t + \alpha_{22} I_{t-1} + \varepsilon_2 & \text{(функция инвестиций)} \\ r_t = \alpha_{03} + \beta_{31} Y_t + \alpha_{32} M_t + \varepsilon_3 & \text{(функция денежного рынка)}, \\ Y_t = C_t + I_t + G_t & \text{(тождество дохода)} \end{cases}$$

где C — расходы на потребление;

Y — совокупный доход;

I — инвестиции;

r — процентная ставка;

M — денежная масса;

G — государственные расходы;

t — текущий период;

t-1 — предыдущий период.

В данной модели четыре эндогенные переменные (C_t, I_t, Y_t, r_t), $N = 4$, и четыре экзогенных (M_t, G_t, C_{t-1}, I_{t-1}).

Для ***первого уравнения***: $H = 2$ (C_t, Y_t присутствуют), $D = 3$ (M_t, G_t, I_{t-1} отсутствуют) и $D > H - 1$, поэтому уравнение сверхидентифицируемо (необходимое условие).

Для проверки на достаточное условие идентификации выпишем матрицу коэффициентов при переменных, не входящих в первое уравнение.

Уравнение	I_t	r_t	I_{t-1}	M_t	G_t
2	–1	β_{21}	α_{22}	0	0
3	0	–1	0	α_{32}	0
4	1	0	0	0	1

Минор 3-го порядка данной матрицы $\begin{vmatrix} -1 & \beta_{21} & 0 \\ 0 & -1 & 0 \\ 1 & 0 & 1 \end{vmatrix} \neq 0$, следовательно, ранг матрицы равен $3 \geq N - 1$, т. е. достаточное условие идентификации выполняется.

Для ***второго уравнения***: $H = 2$ (I_t, r_t присутствуют), $D = 3$ (M_t, G_t, C_{t-1} отсутствуют) и $D > H - 1$, поэтому уравнение сверхидентифицируемо.

Выпишем матрицу коэффициентов при переменных, не входящих во второе уравнение:

Уравнение	C_t	Y_t	C_{t-1}	M_t	G_t
1	–1	β_{11}	α_{12}	0	0
3	0	β_{31}	0	α_{32}	0
4	1	–1	0	0	1

Минор 3-го порядка данной матрицы $\begin{vmatrix} -1 & 0 & 0 \\ 0 & \alpha_{32} & 0 \\ 1 & 0 & 1 \end{vmatrix} \neq 0$, следовательно, ранг матрицы равен $3 \geq N - 1$, т. е. достаточное условие идентификации выполняется.

Для ***третьего уравнения***: $H = 2$ (Y_t, r_t присутствуют), $D = 3$ (G_t, C_{t-1}, I_{t-1} отсутствуют) и $D > H - 1$, поэтому уравнение сверхидентифицируемо.

Выпишем матрицу коэффициентов при переменных, не входящих в третье уравнение:

Уравнение	C_t	C_{t-1}	I_t	I_{t-1}	G_t
1	–1	α_{12}	0	0	0
2	0	0	–1	α_{22}	0
4	1	0	1	0	1

Минор 3-го порядка данной матрицы $\begin{vmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 1 & 1 & 1 \end{vmatrix} \neq 0$, следовательно, ранг матрицы равен $3 \geq N - 1$, т. е. достаточное условие идентификации выполняется.

Четвертое уравнение представляет собой тождество, параметры которого известны и необходимости в его идентификации нет.

Таким образом, все уравнения модели сверхидентифицированы. ▲

Выводы

Для установления идентифицируемости можно использовать ИП или условия для идентификации.

Для решения *точно идентифицируемого уравнения* применяетс КМНК, а для решения *сверхидентифицированного уравнения* — ДМНК.

Для точно идентифицированной системы применение к не КМНК или ДМНК дает одинаковые результаты.

Сверхидентифицируемая структурная модель может быть дву типов:

- все уравнения системы сверхидентифицируемы;
- система наряду со сверхидентифицируемыми содержит точн идентифицируемые уравнения.

Если все уравнения системы сверидентифицируемы, то дл оценки структурных коэффициентов каждого уравнения использу ется ДМНК.

Если в системе есть точно идентифицируемые уравнения, т структурные коэффициенты по ним находятся из системы приве денных уравнений.

Для точно идентифицируемых уравнений ДМНК дает тот ж результат, что и КМНК.

ЛИТЕРАТУРА

1. Эконометрика / Под ред. И. И. Елисеевой. — М.: Проспект 2009.

2. Эконометрика / Под ред. В. С. Мхитаряна. — М.: Проспект 2009.

3. *Плохотников К. Э.* Основы эконометрики в пакете STATISTICA: Учеб. пособие. — М.: Вузовский учебник, 2010.

4. *Замков О. О.* Эконометрические методы в макроэкономическом анализе: Курс лекций. — М.: ГУ ВШЭ, 2001.

5. *Просветов Г. И.* Эконометрика: Учебно-методическое пособие. — М.: РДЛ, 2005.

6. *Доугерти К.* Введение в эконометрику: Пер. с англ. — М. ИНФРА-М, 1999.

7. *Магнус Я. Р., Катышев П. К., Пересецкий А. А.* Эконометрика: Начальный курс. — М.: Дело, 2005.

8. *Винстон У.* Excel 2007. Анализ данных и бизнес моделирование: Пер. с англ. — М.: БХВ-Петербург, 2008.

ПРИЛОЖЕНИЕ

Значения d_1 и d_2 критерия Дарбина–Уотсона при уровне значимости 0,05

(n — число наблюдений, m — число объясняющих переменных)

N	$m = 1$		$m = 2$		$m = 3$		$m = 4$		$M = 5$	
	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2
6	0,61	1,40								
7	0,70	1,36	0,47	1,90						
8	0,76	1,33	0,56	1,78	0,37	2,29				
9	0,82	1,32	0,63	1,70	0,46	2,13				
10	0,88	1,32	0,70	1,64	0,53	2,02				
11	0,93	1,32	0,66	1,60	0,60	1,93				
12	0,97	1,33	0,81	1,58	0,66	1,86				
13	1,01	1,34	0,86	1,56	0,72	1,82				
14	1,05	1,35	0,91	1,55	0,77	1,78				
15	1,08	1,36	0,95	1,54	0,82	1,75	0,69	1,97	0,56	2,21
16	1,10	1,37	0,98	1,54	0,86	1,73	0,74	1,93	0,62	2,15
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,67	2,10
18	1,16	1,39	1,05	1,53	0,93	1,69	0,82	1,87	0,71	2,06
19	1,18	1,40	1,08	1,53	0,97	1,68	0,86	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,90	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,90
25	1,29	1,45	1,21	1,55	1,12	1,66	1,04	1,77	0,95	1,89
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,88
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,76	1,01	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83
50	1,50	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,77
100	1,65	1,69	1,63	1,72	161	1,74	1,59	1,76	1,57	1,78

Редактор — *В. Н. Рогожкин*
Художник — *В. А. Антипов*
Верстка — *Н. А. Кирьянова*

Ответственный за выпуск — *М. Д. Писарева*

Учебное издание

Новиков Анатолий Иванович

Эконометрика

Учебное пособие

Сертификат соответствия № РОСС RU.AB51.HO5316

Подписано в печать 10.09.2020. Формат 60×84 1/16.
Печать офсетная. Бумага газетная. Печ. л. 14.
Тираж 100 экз.

Издательско-торговая корпорация «Дашков и К°»
129347, Москва, Ярославское шоссе, д. 142, к. 732
Тел.: 8 (495) 668-12-30, 8 (499) 183-93-23
Email: sales@dashkov.ru — отдел продаж;
office@dashkov.ru — офис; http://www.dashkov.ru

9785394040511

535,000 [redacted]