Н.Ш. Кремер
Б.А. Путко

ЭКОНОМЕТРИКА

УЧЕБНИК

ТРЕТЬЕ ИЗДАНИЕ

ЮНИТИ
UNITY

Н.Ш. Кремер
Б.А. Путко

ЭКОНОМЕТРИКА

Под редакцией
профессора *Н.Ш. Кремера*

Третье издание,
переработанное и дополненное

Рекомендовано Министерством образования Российской Федерации в качестве ***учебника*** *для студентов высших учебных заведений*

Рекомендовано Учебно-методическим центром «Профессиональный учебник» в качестве ***учебника*** *для студентов высших учебных заведений, обучающихся по специальностям экономики и управления*

Москва • 2010

УДК 330.43(075.8)
ББК 65.в6.я73
К79

Рецензенты:
кафедра математической статистики и эконометрики Московского государственного университета экономики, статистики и информатики
(зав. кафедрой д-р экон. наук, проф. В.С. Мхитарян)
д-р физ.-мат. наук, проф. *Ю.С. Хохлов*

Главный редактор издательства *Н.Д. Эриашвили*,
кандидат юридических наук, доктор экономических наук, профессор,
лауреат премии Правительства РФ в области науки и техники

Кремер, Наум Шевелевич.

К79 Эконометрика: учебник для студентов вузов / Н.Ш. Кремер, Б.А. Путко; под ред. Н.Ш. Кремера. — 3-е изд., перераб. и доп. — М.: ЮНИТИ-ДАНА, 2010. — 328 с. — (Серия «Золотой фонд российских учебников»).

I. Путко, Борис Александрович.

ISBN 978-5-238-01720-4

В учебнике излагаются основы эконометрики. Большое внимание уделяется классической (парной и множественной) и обобщенной моделям линейной регрессии, классическому и обобщенному методам наименьших квадратов, анализу временных рядов и систем одновременных уравнений, моделям с панельными данными. Обсуждаются различные аспекты многомерной регрессии: мультиколлинеарность, фиктивные переменные, спецификация и линеаризация модели, частная корреляция. Учебный материал сопровождается достаточным числом решенных задач и задач для самостоятельной работы.

Для студентов, бакалавров и магистров экономических направлений и специальностей вузов, аспирантов, преподавателей и специалистов по прикладной экономике и финансам, лиц, обучающихся по программам МВА, второго высшего образования и проходящих профессиональную переподготовку или повышение квалификации.

ББК 65.в6.я73

ISBN 978-5-238-01720-4

Предисловие

«Эконометрика» как дисциплина федерального (регионального) компонента по циклу общих математических и естественно-научных дисциплин включена в основную образовательную программу подготовки экономистов, определяемую Государственными образовательными стандартами высшего образования. Однако в настоящее время ощущается нехватка доступных учебников и учебных пособий по эконометрике для студентов экономических специальностей вузов.

Авторы данного учебника попытались хотя бы в некоторой степени восполнить имеющийся пробел. Учебник написан в соответствии с требованиями Государственного образовательного стандарта по дисциплине «Эконометрика» для экономических специальностей вузов. При изложении учебного материала предполагается, что читатель владеет основами теории вероятностей, математической статистики и линейной алгебры в объеме курса математики экономического вуза (например, [4], или [5], и [17]).

Учебник состоит из введения, основного учебного материала (гл. 1—11) и приложения (гл. 12—13).

Во введении дано определение эконометрики, показано ее место в ряду математико-статистических и экономических дисциплин.

В *главе 1* изложены основные аспекты эконометрического моделирования, его предпосылки, типы выборочных данных, виды моделей, этапы проведения и возникающие при этом проблемы моделирования.

В связи с тем, что основой математического инструментария эконометрики является теория вероятностей и математическая статистика, в *главе 2* представлен краткий обзор ее основных понятий и результатов. Следует иметь в виду, что данный обзор не может заменить систематического изучения соответствующего вузовского курса.

В *главах 3,4* рассмотрены классические линейные регрессионные модели: в *главе 3* — парные регрессионные модели, на примере которых наиболее доступно и наглядно удается проследить базовые понятия регрессионного анализа, выяснить основные предпосылки классической модели, дать оценку ее параметров и геометрическую интерпретацию; в *главе 4* — обобщение

регрессии на случай нескольких объясняющих переменных. Применение в главе 4 аппарата матричной алгебры позволяет дать компактное описание и анализ множественной регрессии, доказательство ее основных положений.

В *главе 5* рассмотрен ряд проблем, связанных с использованием регрессионных моделей, таких, как мультиколлинеарность, фиктивные переменные, линеаризация модели, частная корреляция.

В *главе 6* даны общие понятия и проанализированы вопросы, связанные с временными (динамическими) рядами и использованием их моделей для прогнозирования.

В *главе 7* представлены обобщенная линейная модель множественной регрессии и обобщенный метод наименьших квадратов. Исследуется комплекс вопросов, связанных с нарушением предпосылок классической модели регрессии — гетероскедастичностью и автокоррелированностью остатков временного ряда, их тестированием и устранением, идентификацией временного ряда.

Глава 8 посвящена рассмотрению стохастических регрессоров и использованию специальных методов инструментальных переменных. Здесь же дано описание специальных моделей временных рядов (авторегрессионных, скользящей средней, с распределенными лагами и их модификаций), позволяющих наиболее эффективно решать задачи анализа и прогнозирования временных рядов.

В *главе 9* изучены эконометрические модели, выраженные системой одновременных уравнений. Рассмотрены проблемы идентифицируемости параметров модели, косвенный и трехшаговый метод наименьших квадратов.

В *главе 10* отражены проблемы спецификации эконометрических моделей.

В новой *главе 11*, включенной дополнительно в третье издание учебника, приводятся модели, основанные на панельных (пространственно-временных) данных, и модели с ограничениями на значения зависимой переменной и отбор наблюдений в выборку.

В главах (1—11) авторы ограничились рассмотрением в основном *линейных* эконометрических моделей как наиболее простых и обладающих меньшим риском получения значительных ошибок прогноза. По той же причине изучение временных рядов было ограничено рассмотрением в основном *стационарных* рядов.

Учитывая матричную форму изложения в учебнике вопросов множественной регрессии, в приложении (*главе 12*) приведены

основные сведения из линейной алгебры. Кроме того, в *главе 13* рассмотрено применение компьютерных пакетов для оценивания эконометрических моделей, а также проведение эксперимента по методу Монте-Карло, основанного на компьютерном моделировании случайных величин.

Изложение материала сопровождается иллюстрирующими его примерами и задачами. Решение этих задач проводится либо «вручную» — для отработки соответствующих методов их решения, либо с помощью компьютерного эконометрического пакета «*Econometric Views*». При подготовке задач были использованы различные пособия и методические материалы. Часть задач составлена авторами специально для учебника. Задачи с решениями приводятся в основном тексте данной главы, а задачи для самостоятельной работы — в конце главы в рубрике «Упражнения». (Нумерация задач по главе — единая.)

Необходимые для решения задач математико-статистические таблицы даны в приложении. В конце книги приведен развернутый предметный указатель основных понятий курса.

Авторы выражают глубокую благодарность проф. *В.С. Мхитаряну* и проф. *Ю.С. Хохлову* за рецензирование рукописи и сделанные ими замечания.

Авторы учебника:

профессор *Н.Ш. Кремер* — предисловие, гл. 1 (§ 1.5, 1.6), 2—6, 12 и введение, гл 7 (совместно Б.А. Путко);

доцент *Б.А. Путко* — гл. 1 (§ 1.1—1.4), 8—11, 13 и введение, гл. 7 (совместно с Н.Ш. Кремером).

Введение

Последние десятилетия эконометрика как научная дисциплина стремительно развивается. Растет число научных публикаций и исследований с применением эконометрических методов. Свидетельством всемирного признания эконометрики является присуждение за наиболее выдающиеся разработки в этой области Нобелевских премий по экономике Р. Фришу и Я. Тинбергу (1969), Л. Клейну (1980), Т. Хаавельмо (1989), Дж. Хекману и Д. Макфаддену (2000).

Язык экономики все больше становится языком математики, а экономику все чаще называют одной из наиболее математизированных наук.

Достижения современной экономической науки предъявляют новые требования к высшему профессиональному образованию экономистов. *Современное экономическое образование,* — утверждает директор ЦЭМИ РАН академик В. Л. Макаров, — *держится на трех китах: макроэкономике, микроэкономике и эконометрике.* Если в период централизованной плановой экономики упор делался на балансовых и оптимизационных методах исследования, на описании «системы функционирования социалистической экономики», построении оптимизационных моделей отраслей и предприятий, то *в период перехода к рыночной экономике возрастает роль эконометрических методов*. Без знания этих методов невозможно ни исследование и теоретическое обобщение эмпирических зависимостей экономических переменных, ни построение сколько-нибудь надежного прогноза в банковском деле, финансах или бизнесе.

Единое общепринятое определение эконометрики в настоящее время отсутствует. Сам термин «эконометрика»[1] был введен в 1926 г. норвежским ученым Р. Фришем и в дословном переводе означает «эконометрические измерения». Наряду с таким широким пониманием эконометрики, порождаемым переводом самого термина, встречается и весьма узкая трактовка эконометрики как набора математико-статистических методов, используемых в приложениях математики в экономике.

[1] В литературе используется и другой, менее употребительный, термин «эконометрия».

Приводимые ниже определения и высказывания известных ученых позволяют получить представление о различных толкованиях эконометрики.

Эконометрика — это раздел экономики, занимающийся разработкой и применением статистических методов для измерений взаимосвязей между экономическими переменными (*С. Фишер и др.*).

Основная задача эконометрики — наполнить эмпирическим содержанием априорные экономические рассуждения (*Л. Клейн*).

Цель эконометрики — эмпирический вывод экономических законов (*Э. Маленво*).

Эконометрика является не более чем набором инструментов, хотя и очень полезных... Эконометрика является одновременно нашим телескопом и нашим микроскопом для изучения окружающего экономического мира (*Ц. Грилихес*).

Р. Фриш указывает на то, что *эконометрика есть единство трех составляющих — статистики, экономической теории и математики.*

С.А. Айвазян полагает, что эконометрика объединяет совокупность методов и моделей, позволяющих на базе *экономической теории, экономической статистики и математико-статистического инструментария* придавать количественные выражения качественным зависимостям.

Основные результаты экономической теории носят качественный характер, а *эконометрика вносит в них эмпирическое содержание.* Математическая экономика выражает экономические законы в виде математических соотношений, а *эконометрика осуществляет опытную проверку этих законов.* Экономическая статистика дает информационное обеспечение исследуемого процесса в виде исходных (обработанных) статистических данных и экономических показателей, а *эконометрика,* используя традиционные математико-статистические и специально разработанные методы, *проводит анализ количественных взаимосвязей между этими показателями.*

Многие базовые понятия эконометрики имеют два определения — «экономическое» и «математическое». Подобная двойственность имеет место и в формулировках результатов. Характер научных работ по эконометрике варьируется от «классических» экономических работ, в которых почти не используется математический аппарат, до солидных математических трудов, использующих достаточно тонкий аппарат современной математики.

Экономическая составляющая эконометрики, безусловно, *является первичной*. Именно экономика определяет постановку задачи и исходные предпосылки, а результат, формируемый на математическом языке, представляет интерес лишь в том случае, если удается его экономическая интерпретация. В то же время многие эконометрические результаты носят характер математических утверждений (теорем).

Широкому *внедрению эконометрических методов способствовало появление* во второй половине XX в. *электронных вычислительных машин* и в частности *персональных компьютеров*. Компьютерные эконометрические пакеты сделали эти методы более доступными и наглядными, так как наиболее трудоемкую (рутинную) работу по расчету различных статистик, параметров, характеристик, построению таблиц и графиков в основном стал выполнять компьютер, а эконометристу осталась главным образом творческая работа: постановка задачи, выбор соответствующей модели и метода ее решения, интерпретация результатов.

Глава 1

Основные аспекты эконометрического моделирования

1.1. Введение в эконометрическое моделирование

Рассмотрим следующую ситуацию. Допустим, мы хотим продать автомобиль и решили дать объявление о продаже в газете «Из рук в руки». Естественно, перед нами встает вопрос: какую цену указать в объявлении?

Очевидно, мы будем руководствоваться информацией о цене, которую выставляют другие продавцы подобных автомобилей. Что значит «подобные автомобили»? — Очевидно, это автомобили, обладающие близкими значениями таких факторов, как год выпуска, пробег, мощность двигателя. Проглядев колонку объявлений, мы формируем свое мнение о рынке интересующего нас товара и, возможно, после некоторого размышления, назначаем цену.

На этом простейшем примере на самом деле можно проследить основные моменты эконометрического моделирования. Рассмотрим наши действия более формализованно.

Мы ставим задачу определить цену — величину, формируемую под воздействием некоторых факторов (года выпуска, пробега и т. д.). Такие зависимые величины обычно называются *зависимыми* (*объясняемыми*) переменными, а факторы, от которых они зависят, — *объясняющими*.

Формируя общее мнение о состоянии рынка, мы обращаемся к интересующему нас объекту и получаем *ожидаемое* значение зависимой переменной при заданных значениях объясняющих переменных.

Указанная конкретная цена — *наблюдаемое* значение зависимой переменной зависит также и от *случайных* явлений — таких, например, как характер продавца, его потребность в конкретной денежной сумме, возможные сроки продажи автомобиля и др.

Продавец-одиночка вряд ли будет строить какую-либо математическую модель, но менеджер крупного салона, специализирующегося на торговле автомобилями на вторичном рынке, скорее всего, захочет иметь более точное представление об ожидаемой цене и о возможном поведении случайной составляющей. Следующий шаг и есть эконометрическое моделирование.

Общим моментом для любой эконометрической модели является *разбиение зависимой переменной на две части — объясненную и случайную.* Сформулируем задачу моделирования самым общим, неформальным образом: *на основании экспериментальных данных определить объясненную часть и, рассматривая случайную составляющую как случайную величину, получить* (возможно, после некоторых предположений) *оценки параметров ее распределения.*

Таким образом, эконометрическая модель имеет следующий вид:

Наблюдаемое значение зависимой переменной	=	Объясненная часть, зависящая от значений объясняющих переменных	+	Случайная составляющая
Y	=	$f(X)$	+	ε. **(1.1)**

Остановимся теперь на целях моделирования. Предположим, получено следующее выражение для объясненной части переменной Y — цены автомобиля:

$$\hat{y} = 18000 - 1000x_1 - 0,5x_2,$$

где $\hat{y}$ — ожидаемая цена автомобиля (в усл. ден. ед., здесь и далее у.е.);

X_1 — срок эксплуатации автомобиля (в годах);

X_2 — пробег (в тыс. км)[1].

Каково практическое применение полученного результата? Очевидно, во-первых, он позволяет понять: как именно формируется рассматриваемая экономическая переменная — цена на автомобиль. Во-вторых, он дает возможность выявить влияние каждой из объясняющих переменных на цену автомобиля (так, в данном случае цена нового автомобиля (при x_1=0, x_2=0) 18000 у.е., при этом только за счет увеличения срока эксплуатации на 1 год цена автомобиля уменьшается в среднем на 1000 у.е., а только за счет увеличения пробега на 1 тыс. км — на 0,5 у.е.). В третьих, что, пожалуй, наиболее важно, этот результат позволяет п р о г н о з и р о в а т ь цену на автомобиль,

[1] Переменные обозначаем прописными буквами латинского алфавита, а их значения — строчными.

если известны его основные параметры. Теперь менеджеру не составит большого труда определить ожидаемую цену вновь поступившего для продажи автомобиля, даже если его год выпуска и пробег не встречались ранее в данном салоне.

1.2. Основные математические предпосылки эконометрического моделирования

Пусть имеется p объясняющих переменных $X_1,..., X_p$ и зависимая переменная Y. Переменная Y является *случайной величиной,* имеющей при заданных значениях факторов некоторое распределение. Если случайная величина Y непрерывна, то можно считать, что ее распределение при каждом допустимом наборе значений факторов (x_1, x_2,..., x_p) имеет условную плотность $f_{x_1,x_2,...,x_p}(y)$.

Обычно делается некоторое предположение относительно распределения Y. Чаще всего предполагается, что условные распределения Y при каждом допустимом значении факторов — *нормальные*. Подобное предположение позволяет получить значительно более «продвинутые» результаты. Впрочем, заметим здесь же, что порой предположение о нормальности условных распределений Y приходится отвергнуть.

Объясняющие переменные X_j ($j = 1,..., p$) могут считаться как *случайными,* так и *детерминированными,* т. е. принимающими определенные значения. Проиллюстрируем этот тезис на уже рассмотренном примере продажи автомобилей. Мы можем заранее определить для себя параметры автомобиля и искать объявления о продаже автомобиля с такими параметрами. В этом случае неуправляемой, случайной величиной остается только зависимая переменная — цена. Но мы можем также случайным образом выбирать объявления о продаже, в этом случае параметры автомобиля — объясняющие переменные — также оказываются случайными величинами.

Классическая эконометрическая модель рассматривает объясняющие переменные X_j как *детерминированные,* однако, как мы увидим в дальнейшем, основные результаты статистического исследования модели остаются в значительной степени теми же, что и в случае, если считать X_j с*лучайными* переменными.

Объясненная часть — обозначим ее Y_e — в любом случае представляет собой функцию от значений факторов — объясняющих переменных:

$$Y_e = f(X_1,..., X_p).$$

Таким образом, эконометрическая модель имеет вид

$$Y = f(X_1,..., X_p) + \varepsilon.$$

Наиболее естественным выбором объясненной части случайной величины Y является ее среднее значение — условное математическое ожидание $M_{x_1, x_2,..., x_p}(Y)$, полученное при данном наборе значений объясняющих переменных (x_1, x_2,..., x_p). (В дальнейшем математическое ожидание будем обозначать $M_x(Y)$.) В самом деле, по своему смыслу объясненная часть — это ожидаемое значение зависимой переменной при заданных значениях объясняющих.

Уравнение $M_x(Y) = f(x_1,..., x_p)$ называется *уравнением регрессии.*

При таком естественном выборе объясненной части эконометрическая модель имеет вид

$$Y = M_x(Y) + \varepsilon, \qquad \textbf{(1.2)}$$

где ε — случайная величина, называемая *возмущением* или *ошибкой.* В курсе математической статистики уравнение (**1.2**) называется *уравнением регрессионной модели.*

Сразу же отметим, что эконометрическая модель не обязательно является регрессионной, т.е. объясненная часть не всегда представляет собой условное математическое ожидание зависимой переменной.

Рассмотрим, например, следующую задачу: определить зависимость затрат Y на какой-либо товар от дохода X. Допустим, имеются данные опроса ста человек и сто пар чисел (x_1, y_1),..., (x_{100}, y_{100}). Анализируя эти данные, мы получаем (отложим пока вопрос — каким образом) зависимость $Y_e = f(X)$.

Однако может оказаться, что данные о доходе, полученные в результате опроса, на самом деле являются искаженными, — например, в среднем заниженными, т.е. объясняющие переменные измеряются с систематическими ошибками. В этом случае люди, действительно обладающие доходом X, будут на самом деле тратить на исследуемый товар в среднем величину, меньшую, чем $f(X)$, т.е. в рассмотренном примере объ-

ясненная часть не есть условное математическое ожидание зависимой переменной.

Систематические ошибки измерения объясняющих переменных — одна из возможных причин того, что эконометрическая модель не является регрессионной. В экономических исследованиях подобная ситуация встречается достаточно часто. Одним из возможных путей устранения этого, как правило, довольно неприятного обстоятельства, является выбор других объясняющих переменных (эти вопросы рассматриваются в гл. **8** настоящего учебника).

С математической точки зрения регрессионные модели оказываются существенно более простым объектом, чем эконометрическая модель общего типа. Отметим здесь некоторые свойства регрессионной модели.

Рассмотрим равенство $Y = M_x(Y) + \varepsilon$ и возьмем от обеих частей математическое ожидание при заданном наборе значений объясняющих переменных X. В этом случае $M_x(Y)$ есть числовая величина, равная своему математическому ожиданию, и мы получаем равенство

$$M_x(\varepsilon) = 0 \qquad \textbf{(1.3)}$$

(а значит, и $M(\varepsilon) = 0$), т.е. в регрессионной модели ожидаемое значение случайной ошибки равно нулю. Можно показать, что отсюда следует (если объясняющие переменные рассматриваются как случайные величины) *некоррелированность случайных ошибок и объясняющих переменных X*. Это обстоятельство оказывается наиболее существенным условием *состоятельности* получаемых количественных результатов анализа эконометрической модели.

1.3. Эконометрическая модель и экспериментальные данные

Чтобы получить достаточно достоверные и информативные данные о распределении какой-либо случайной величины, необходимо иметь *выборку* ее наблюдений достаточно большого объема. Выборка наблюдений зависимой переменной Y и объясняющих переменных X_j ($j = 1,..., p$) является отправной точкой любого эконометрического исследования.

Такие выборки представляют собой наборы значений ($x_{i1},..., x_{ip}; y_i$), где $i = 1,...,n$; p — количество объясняющих переменных, n — число наблюдений.

Как правило, число наблюдений n достаточно велико (десятки, сотни) и значительно превышает число p объясняющих переменных. Проблема, однако, заключается в том, что наблюдения y_i, рассматриваемые в разных выборках как случайные величины Y_i и получаемые при различных наборах значений объясняющих переменных X_j, имеют, вообще говоря, различное распределение. А это означает, что для каждой случайной величины Y_i мы имеем всего лишь одно наблюдение. Разумеется, на основании одного наблюдения никакого адекватного вывода о распределении случайной величины сделать нельзя, и нужны дополнительные предположения.

В классическом курсе эконометрики рассматривается три типа выборочных данных.

Пространственная выборка или пространственные данные (*cross-sectional data*). В экономике под *пространственной выборкой* понимают набор показателей экономических переменных, полученный в данный момент времени. Для эконометриста, однако, такое определение не очень удобно — из-за неоднозначности понятия «момент времени». Это может быть и день, и неделя, и год. Очевидно, о пространственной выборке имеет смысл говорить в том случае, если все наблюдения получены примерно в неизменных условиях, т. е. представляют собой набор независимых выборочных данных из некоторой генеральной совокупности.

Таким образом, мы будем называть *пространственной выборкой* серию из n независимых наблюдений $(p+1)$-мерной случайной величины $(X_1,...,X_p;\ Y)$. (При этом в дальнейшем можно не рассматривать X_j как случайные величины.) В этом случае различные случайные величины Y_i оказываются между собой независимыми, что влечет за собой некоррелированность их возмущений, т. е.

$$r(\varepsilon_i, \varepsilon_j) = 0 \text{ при } i = j, \qquad \textbf{(1.4)}$$

где $r(\varepsilon_i, \varepsilon_j)$ — коэффициент корреляции между возмущениями ε_i и ε_j.

Условие **(1.4)** существенно упрощает модель и ее статистический анализ.

Как определить, является ли выборка серией независимых наблюдений? — На этот вопрос нет однозначного ответа. Фор-

мальное определение независимости случайных величин, как правило, оказывается реально непроверяемым. Обычно за независимые принимаются величины, не связанные п р и ч и н н о. Однако на практике далеко не всегда вопрос о независимости оказывается бесспорным.

Вернемся к примеру о продаже машины (см. § **1.1**).

Пусть Y — цена машины, X — год выпуска, а $(x_1, y_1),\ldots, (x_n, y_n)$ — серия данных, полученная из газеты «Из рук в руки». Можно ли считать эти наблюдения независимыми?

Различные продавцы не знакомы между собой, они дают свои объявления независимо друг от друга, так что предположение о независимости наблюдений выглядит вполне разумно. С другой стороны, человек, назначающий цену за свой автомобиль, руководствуется ценами предыдущих объявлений, так что и возражение против независимости наблюдений также имеет право на существование.

Из этого можно сделать вывод, что решение о пространственном характере выборки в известной степени субъективно и связано с условиями используемой модели. Впрочем, то же самое можно сказать о многих предположениях, которые делаются в математической статистике и особенно ее приложениях.

Итак, эконометрическая модель, построенная на основе пространственной выборки экспериментальных данных (x_i, y_i), имеет вид:

$$y_i = f(x_i) + \varepsilon_i, \ \ i = 1,\ldots, n, \qquad \textbf{(1.5)}$$

где ошибки регрессии удовлетворяют условиям

$$M(\varepsilon_i) = 0\,, \qquad \textbf{(1.6)}$$

$$r(\varepsilon_i, \varepsilon_j) = 0, \qquad \textbf{(1.7)}$$

$$D(\varepsilon_i) = \sigma_i^2\,. \qquad \textbf{(1.8)}$$

Что касается условия **(1.8)**, то здесь возможны два случая:

а) $\sigma_i^2 = \sigma_j^2$ при всех i и j. Свойство постоянства дисперсий ошибок регрессии называется *гомоскедастичностью*. В этом случае распределения случайных величин Y_i отличаются только значением математического ожидания (объясненной части);

б) $\sigma_i^2 \neq \sigma_j^2$. В этом случае имеет место *гетероскедастичность модели*. Гетероскедастичность «портит» многие результаты статистического анализа и, как правило, требует устранения. (Подробнее об этом см. в гл. **7**.)

Как определить, является ли изучаемая модель гомо- или гетероскедастичной? — В некоторых случаях это достаточно очевидно. Например, цена автомобиля, которому пятнадцать лет, вряд ли может подняться выше 2000 у.е., так что стандартная ошибка цены в этом случае вряд ли может быть больше, чем 300—400 у.е. Между тем автомобиль, которому два года, может стоить и 7000, и 17 000 у.е., т.е. стандартная ошибка заведомо не меньше 1500—2000 у.е.

Однако во многих случаях гетероскедастичность модели далеко не столь очевидна, и требуется применение методов математической статистики для принятия решения о том, какой тип модели будет рассматриваться.

Временной (динамический) ряд (*time-series data*). Временным (динамическим) рядом называется выборка наблюдений, в которой важны не только сами наблюдаемые значения случайных величин, но и п о р я д о к их следования друг за другом. Чаще всего упорядоченность обусловлена тем, что экспериментальные данные представляют собой серию наблюдений одной и той же случайной величины в последовательные моменты времени. В этом случае динамический ряд называется *временным рядом*. При этом предполагается, что тип распределения наблюдаемой случайной величины остается одним и тем же (например, нормальным), но параметры его меняются в зависимости от времени.

Модели временных рядов, как правило, оказываются сложнее моделей пространственной выборки, так как наблюдения в случае временного ряда вообще говоря н е я в л я ю т с я н е з а в и с и м ы м и, а это значит, что ошибки регрессии могут коррелировать друг с другом, т. е. условие **(1.4)** вообще говоря не выполняется. В последующих главах мы увидим, что невыполнение условия **(1.4)** значительно усложняет статистический анализ модели.

Следует особенно отметить, что имея только ряд наблюдений без понимания их природы, невозможно определить, имеем мы дело с пространственной выборкой или временным рядом. Пусть, например, имеется 500 пар чисел $(x_1, y_1),\ldots, (x_{500}, y_{500})$, где Y — цена автомобиля, а X — год выпуска. Данные взяты из газеты «Из рук в руки». Возможны следующие варианты:

1) n газет было упорядочено по дате их выпуска, и из каждой газеты было выбрано (случайным образом) по одно-

му объявлению. — В этом случае мы, очевидно, можем считать, что имеем дело с временным рядом;

2) газеты были произвольным образом перемешаны, и невзирая на дату выпуска случайным образом было отобрано n объявлений. — В этом случае мы, скорее всего, можем считать, что наша выборка — пространственная.

При этом, вообще говоря, возможно, что в обоих случаях мы получим один и тот же набор числовых данных. Более того, теоретически возможно даже и то, что они окажутся в той же последовательности! Однако во втором случае мы должны постулировать некоррелированность ошибок регрессии (выполнение условия **(1.4)**), между тем как в первом случае подобная предпосылка может оказаться неправомерной.

Панельные (пространственно-временные) данные (рассмотрены в гл. 11).

1.4. Линейная регрессионная модель

Пусть определен характер экспериментальных данных и выделен определенный набор объясняющих переменных.

Для того, чтобы найти объясненную часть, т. е. величину $M_x(Y)$, требуется знание *условных распределений случайной величины Y*. На практике это почти никогда не имеет места, поэтому точное нахождение объясненной части невозможно.

В таких случаях применяется стандартная *процедура сглаживания* экспериментальных данных, подробно описанная, например, в [1]. Эта процедура состоит из двух этапов:

1) определяется параметрическое семейство, к которому принадлежит искомая функция $M_x(Y)$ (рассматриваемая как функция от значений объясняющих переменных X). Это может быть множество линейных функций, показательных функций и т.д.;
2) находятся оценки параметров этой функции с помощью одного из методов математической статистики.

Формально никаких способов выбора параметрического семейства не существует. Однако в подавляющем большинстве случаев эконометрические модели выбираются линейными.

Кроме вполне очевидного преимущества линейной модели — ее относительной *простоты*, — для такого выбора имеются, по крайней мере, две существенные причины.

Первая причина: если случайная величина (*X, Y*) имеет совместное *нормальное* распределение, то, как известно, у*равнения регрессии линейные* (см. § **2.5**). Предположение о нормальном распределении является вполне естественным и в ряде случаев может быть обосновано с помощью предельных теорем теории вероятностей (см. § **2.6**).

В других случаях сами величины *Y* или *X* могут не иметь нормального распределения, но некоторые функции от них распределены нормально. Например, известно, что логарифм доходов населения — нормально распределенная случайная величина. Вполне естественно считать нормально распределенной случайной величиной пробег автомобиля. Часто гипотеза о нормальном распределении принимается во многих случаях, когда нет явного ей противоречия, и, как показывает практика, подобная предпосылка оказывается вполне разумной.

Вторая причина, по которой линейная регрессионная модель оказывается предпочтительнее других, — это *меньший риск значительной ошибки прогноза.*

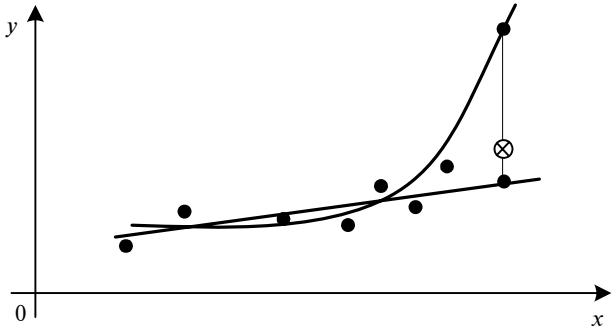


Рис. 1.1

Рис. 1.1 иллюстрирует два выбора функции регрессии — линейной и квадратичной. Как видно, имеющееся множество экспериментальных данных (точек) парабола сглаживает, пожалуй, даже лучше, чем прямая. Однако парабола быстро удаляется от корреляционного поля и для добавленного наблюдения (обозначенного крестиком) теоретическое значение может очень значительно отличаться от эмпирического.

Можно придать точный математический смысл этому утверждению: *ожидаемое значение ошибки прогноза,* т.е. математическое ожидание квадрата отклонения наблюдаемых значений от

сглаженных (или теоретических) $M\,(Y_{\text{набл}} - Y_{\text{теор}})^2$ *оказывается меньше в том случае, если уравнение регрессии выбрано линейным.*

В настоящем учебнике мы в основном будем рассматривать л и н е й н ы е регрессионные модели, и, по мнению авторов, это вполне соответствует той роли, которую играют линейные модели в эконометрике.

Наиболее хорошо изучены линейные регрессионные модели, удовлетворяющие условиям **(1.6)**, **(1.7)** и свойству постоянства дисперсии ошибок регрессии, — они называются *классическими моделями*.

Заметим, что условиям классической регрессионной модели удовлетворяют и гомоскедастичная модель пространственной выборки, и модель временного ряда, наблюдения которого не коррелируют, а дисперсии постоянны. С математической точки зрения они действительно неразличимы (хотя могут значительно различаться экономические интерпретации полученных математических результатов).

Подробному рассмотрению классической регрессионной модели посвящены гл. **3**, **4** настоящего учебника. Практически весь последующий материал посвящен моделям, которые так или иначе могут быть сведены к классической. Часто раздел эконометрики, изучающий классические регрессионные модели, называется «Эконометрикой-1», в то время как курс «Эконометрика-2» охватывает более сложные вопросы, связанные с временными рядами, а также более сложными, существенно нелинейными моделями.

1.5. Система одновременных уравнений

До сих пор мы рассматривали эконометрические модели, задаваемые уравнениями, выражающими зависимую (объясняемую) переменную через объясняющие переменные. Однако реальные экономические объекты, исследуемые с помощью эконометрических методов, приводят к расширению понятия эконометрической модели, описываемой *системой регрессионных уравнений* и тождеств[1].

Особенностью этих систем является то, что каждое из уравнений системы, кроме «своих» *объясняющих* переменных, может включать *объясняемые* переменные из других уравнений. Таким

[1] В отличие от регрессионных уравнений тождества не содержат подлежащих оценке параметров модели и не включают случайной составляющей.

образом, мы имеем не одну зависимую переменную, а набор зависимых (объясняемых) переменных, связанных уравнениями системы. Такую систему называют также *системой одновременных уравнений,* подчеркивая тот факт, что в системе одни и те же переменные *одновременно* рассматриваются как зависимые в одних уравнениях и независимые в других.

Системы одновременных уравнений наиболее полно описывают экономический объект, содержащий множество взаимосвязанных *эндогенных* (формирующихся внутри функционирования объекта) и *экзогенных* (задаваемых извне) переменных. При этом в качестве эндогенных и экзогенных могут выступать *лаговые* (взятые в предыдущий момент времени) переменные.

Классическим примером такой системы является модель спроса Q^d и предложения Q^s (см. § **9.1**), когда спрос на товар определятся его ценой P и доходом потребителя I, предложение товара — его ценой P и достигается равновесие между спросом и предложением:

$$Q^d = \beta_1 + \beta_2 P + \beta_3 I + \varepsilon_1;$$

$$Q^s = \beta_4 + \beta_5 P + \varepsilon_2;$$

$$Q^d = Q^s.$$

В этой системе экзогенной переменной выступает доход потребителя I, а эндогенными — спрос (предложение) товара $Q^d = Q^s = Q$ и цена товара (цена равновесия) P.

В другой модели спроса и предложения в качестве объясняющей предложение Q_t^s переменной может быть не только цена товара P в данный момент времени t, т.е. P_t, но и цена товара в предыдущий момент времени P_{t-1}, т.е. лаговая эндогенная переменная:

$$Q_t^s = \beta_4 + \beta_5 P_t + \beta_6 P_{t-1} + \varepsilon_2.$$

Обобщая изложенное, можно сказать, что *эконометрическая модель позволяет объяснить поведение эндогенных переменных в зависимости от значений экзогенных и лаговых эндогенных переменных* (иначе — в зависимости от *предопределенных,* т.е. заранее определенных, переменных).

Завершая рассмотрение понятия эконометрической модели, следует отметить следующее. Не всякая экономико-математическая модель, представляющая математико-статистическое описание исследуемого экономического объекта, может считаться эконометрической. *Она становится эконометрической только в том*

случае, если будет отражать этот объект на основе характеризующих именно его эмпирических (*статистических*) *данных.*

1.6. Основные этапы и проблемы эконометрического моделирования

Можно выделить шесть основных этапов эконометрического моделирования: постановочный, априорный, этап параметризации, информационный, этапы идентификации и верификации модели [2].

Остановимся подробнее на каждом из этих этапов и рассмотрим проблемы, связанные с их реализацией.

***1-й этап* (*постановочный*).** Формируется *цель* исследования, *набор* участвующих в модели *экономических* переменных.

В качестве цели эконометрического моделирования обычно рассматривают *анализ* исследуемого экономического объекта (процесса); *прогноз* его экономических показателей, *имитацию* развития объекта при различных значениях экзогенных переменных (отражая их случайный характер, изменение во времени), *выработку управленческих решений.*

При выборе экономических переменных необходимо теоретическое обоснование каждой переменной (при этом рекомендуется, чтобы число их было не очень большим и, как минимум, в несколько раз меньше числа наблюдений). Объясняющие переменные не должны быть связаны функциональной или тесной корреляционной зависимостью, так как это может привести к невозможности оценки параметров модели или к получению неустойчивых, не имеющим реального смысла оценок, т. е. к явлению *мультиколлинеарности* (см. об этом гл. **5**).

Забегая вперед, отметим, что для отбора переменных могут быть использованы различные методы, в частности процедуры *пошагового отбора переменных* (§ **5.2**). А для оценки влияния качественных признаков (например, пол, образование и т. п.) могут быть использованы *фиктивные* переменные (§ **5.3**). Но в любом случае определяющим при включении в модель тех или иных переменных является экономический (качественный) анализ исследуемого объекта.

***2-й этап* (*априорный*).** Проводится *анализ сущности* изучаемого *объекта,* формирование и формализация априорной (известной до начала моделирования) информации.

***3-й этап* (*параметризация*).** Осуществляется непосредственно *моделирование,* т.е. выбор общего *вида* модели, выявление входящих в нее связей.

Основная задача, решаемая на этом этапе, — выбор вида функции $f(X)$ в эконометрической модели (**1.1**), в частности, возможность использования линейной модели как наиболее простой и надежной (о некоторых вопросах линеаризации модели см. § **5.5**). Весьма важной проблемой на этом (и предыдущих) этапе эконометрического моделирования является проблема *спецификации* модели (см. гл. **10**), в частности: выражение в математической форме обнаруженных связей и соотношений; установление состава экзогенных и эндогенных переменных, в том числе лаговых; формулировка исходных предпосылок и ограничений модели. От того, насколько удачно решена проблема спецификации модели, в значительной степени зависит успех всего эконометрического моделирования.

***4-й этап* (*информационный*).** Осуществляется *сбор* необходимой статистической информации — наблюдаемых значений экономических переменных

$$(x_{i1}, x_{i2},...,x_{ip}; y_{i1},y_{i2},...,y_{iq}),\ i = 1,...,\ n.$$

Здесь могут быть наблюдения, полученные как с участием исследователя, так и без его участия (в условиях *активного* или *пассивного эксперимента*).

***5-й этап* (*идентификация модели*).** Осуществляется *статистический анализ модели* и *оценка ее параметров.* Реализации этого этапа посвящена основная часть учебника.

С проблемой идентификации модели не следует путать проблему ее *идентифицируемости* (гл. **9**), т. е. проблему возможности получения однозначно определенных параметров модели, заданной системой одновременных уравнений (точнее, параметров *структурной формы* модели, раскрывающей механизм формирования значений эндогенных переменных, по параметрам *приведенной формы* модели, в которой эндогенные переменные непосредственно выражаются через предопределенные переменные).

***6-й этап* (*верификация модели*).** Проводится *проверка истинности, адекватности модели.* Выясняется, насколько удачно решены проблемы спецификации, идентификации и идентифицируемости модели, какова точность расчетов по данной модели, в конечном счете, насколько соответствует построенная модель моделируемому реальному экономическому объекту или процес-

су. Обсуждение указанных вопросов проводится в большинстве глав настоящего учебника.

Следует заметить, что если имеются статистические данные, характеризующие моделируемый экономический объект в данный и предшествующие моменты времени, то для верификации модели, построенной для прогноза, достаточно сравнить реальные значения переменных в последующие моменты времени с соответствующими их значениями, полученными на основе рассматриваемой модели по данным предшествующих моментов.

Приведенное выше разделение эконометрического моделирования на отдельные этапы носит в известной степени условный характер, так как эти этапы могут пересекаться, взаимно дополнять друг друга и т. п.

Прежде чем изучать основные разделы эконометрики — классическую и обобщенную модели регрессии, временные ряды и системы одновременных уравнений, модели с различными типами выборочных данных (гл. **3—11**), рассмотрим в следующей главе (гл. **2**) основные понятия теории вероятностей и математической статистики, составляющие основу математического инструментария эконометрики. Подготовленный соответствующим образом читатель может сразу перейти к изучению гл. **3**.

Глава 2

Элементы теории вероятностей и математической статистики

В этой главе приводится краткий обзор основных понятий и результатов теории вероятностей и математической статистики, которые используются в курсе эконометрики. Цель этой главы — напомнить читателю некоторые сведения, но никак не заменить изучение курса теории вероятностей и математической статистики, например, в объеме учебника [17].

2.1. Случайные величины и их числовые характеристики

Вероятностью P(A) *события A называется численная мера степени объективной возможности появления этого события.*

Согласно классическому определению ***вероятность события*** *A равна отношению числа случаев m, благоприятствующих ему, к общему числу случаев n*, т.е. $P(A)= m/n$. При определенных условиях в качестве оценки вероятности события $P(A)$ может быть использована ***статистическая вероятность*** $P^*(A)$, т. е. *относительная частота* (*частость*) $W(A)$ *появления события A в n произведенных испытаниях.*

Одним из важнейших понятий теории вероятностей является понятие случайной величины.

Под случайной величиной понимается переменная, которая в результате испытания в зависимости от случая принимает одно из возможного множества своих значений (какое именно — заранее не известно).

Более строго ***случайная величина*** X определяется как *функция*, *заданная на множестве элементарных исходов* (*или в пространстве элементарных событий*), т.е.

$$X = f(\omega),$$

где ω — элементарный исход (или элементарное событие, принадлежащее пространству Ω, т. е. $\omega \in \Omega$).

Для *дискретной* случайной величины множество Ξ возможных значений случайной величины, т.е. функции $f(\omega)$, конечно или счетно[1], для *непрерывной* — бесконечно и несчетно.

П р и м е р ы случайных величин:

X — число родившихся детей в течение суток в г. Москве;

Y— число произведенных выстрелов до первого попадания;

Z — дальность полета артиллерийского снаряда.

Здесь X, Y — дискретные случайные величины, а Z — непрерывная случайная величина.

Наиболее полным, исчерпывающим описанием случайной величины является ее закон распределения.

Законом распределения *случайной величины называется всякое соотношение, устанавливающее связь между возможными значениями случайной величины и соответствующими им вероятностями.*

Для дискретной случайной величины закон распределения может быть задан в виде таблицы, аналитически (в виде формулы) и графически.

Например,

X:

x_1	x_2	…	x_i	…	x_n
p_1	p_2	…	p_i	…	p_n

,

или

$$X = \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_1 & p_2 & \dots & p_n \end{pmatrix}.$$

Такая таблица называется *рядом распределения* дискретной случайной величины.

Для любой дискретной случайной величины

$$\sum_{i=1}^{n} P(X = x_i) = \sum_{i=1}^{n} p_i = 1. \qquad \textbf{(2.1)}$$

Если по оси абсцисс откладывать значения случайной величины, по оси ординат — соответствующие их вероятности, то получаемая (соединением точек) ломаная называется *многоугольником* или *полигоном распределения вероятностей.*

---

[1] Напомним, что множество называется *счетным*, если его элементы можно перенумеровать натуральными числами.

▶ **Пример 2.1.** В лотерее разыгрывается: автомобиль стоимостью 5000 ден. ед., 4 телевизора стоимостью 250 ден. ед., 5 видеомагнитофонов стоимостью 200 ден. ед. Всего продается 1000 билетов по 7 ден. ед. Составить закон распределения чистого выигрыша, полученного участником лотереи, купившим один билет.

Р е ш е н и е. Возможные значения случайной величины X — чистого выигрыша на один билет — равны $0 - 7 = -7$ ден. ед. (если билет не выиграл), $200 - 7 = 193$, $250 - 7 = 243$, $5000 - 7 = 4993$ ден. ед. (если на билет выпал выигрыш — видеомагнитофон, телевизор или автомобиль соответственно). Учитывая, что из 1000 билетов число невыигравших составляет 990, а указанных выигрышей 5, 4 и 1 соответственно; используя классическое определение вероятности, получим:

$P(X=-7) = 990/1000 = 0{,}990$; $P(X=193) = 5/1000 = 0{,}005$;

$P(X=243) = 4/1000 = 0{,}004$; $P(X=4993) = 1/1000 = 0{,}001$,

т.е. ряд распределения

X:

x_i	-7	193	243	4993
p_i	0,990	0,005	0,004	0,001

. ▶

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

Закон (ряд) распределения дискретной случайной величины дает исчерпывающую информацию о ней, так как позволяет вычислить вероятности любых событий, связанных со случайной величиной. Однако такой закон (ряд) распределения бывает трудно обозримым, не всегда удобным (и даже необходимым) для анализа.

Поэтому для описания случайных величин часто используются их ***числовые характеристики*** *— числа, в сжатой форме выражающие наиболее существенные черты распределения случайной величины.* Наиболее важными из них являются математическое ожидание, дисперсия, среднее квадратическое отклонение и др. Обращаем внимание на то, что в силу определения, *числовые характеристики случайных величин являются числами н е с л у ч а й н ы м и*, определенными.

Математическим ожиданием, или ***средним значением***, $M(X)$ *дискретной случайной величины* X *называется сумма произведений всех ее значений на соответствующие им вероятности*:

$$M(X) = \sum_{i=1}^{n} x_i p_i. \qquad \textbf{(2.2)}$$

(Для математического ожидания используются также обозначения: $E(X), \overline{X}$.)

▶ **Пример 2.2**. Вычислить $M(X)$ для случайной величины X — чистого выигрыша по данным примера **2.1.**

Р е ш е н и е. По формуле (**2.2**)

$$M(X) = (-7)\cdot 0{,}990 + 193 \cdot 0{,}005 + 243 \cdot 0{,}004 + 4993 \cdot 0{,}001 = 0,$$

т.е. средний выигрыш равен нулю. Полученный результат означает, что вся выручка от продажи билета лотереи идет на выигрыши. ▶

При $n \to \infty$ математическое ожидание представляет сумму ряда $\sum_{i=1}^{\infty} x_i p_i$, если он абсолютно сходится.

С в о й с т в а математического ожидания:

1) $M(C) = C$, где C — постоянная величина;
2) $M(kX) = kM(X)$;
3) $M(X \pm Y) = M(X) \pm M(Y)$;
4) $M(XY) = M(X) \cdot M(Y)$, где X, Y — независимые случайные величины;
5) $M(X \pm C) = M(X) \pm C$;
6) $M(X - a) = 0$, где $a = M(X)$.

Дисперсией $D(X)$ *случайной величины* X *называется математическое ожидание квадрата ее отклонения от математического ожидания:*

$$D(X) = M[X - M(X)]^2, \qquad \textbf{(2.3)}$$

или $$D(X) = M(X - a)^2, \text{ где } a = M(X).$$

(Для дисперсии случайной величины X используется также обозначение $\mathrm{Var}(X)$.)

Дисперсия характеризует *отклонение* (*разброс*, *рассеяние*, *вариацию*) значений случайной величины относительно среднего значения.

Если случайная величина X — дискретная с конечным числом значений, то

$$D(X) = \sum_{i=1}^{n} (x_i - a)^2 p_i \,. \qquad \textbf{(2.4)}$$

Дисперсия $D(X)$ имеет размерность квадрата случайной величины, что не всегда удобно. Поэтому в качестве показателя рассеяния используют также величину $\sqrt{D(X)}$.

Средним квадратическим отклонением* (*стандартным отклонением или ***стандартом***) σ_x *случайной величины* X *называется арифметическое значение корня квадратного из ее дисперсии*:

$$\sigma_x = \sqrt{D(X)} \,. \qquad \textbf{(2.5)}$$

▶ **Пример 2.3.** Вычислить дисперсию и среднее квадратическое отклонение случайной величины X по данным примера **2.2.**

Р е ш е н и е. В примере **2.2** было вычислено $M(X) = 0$.
По формулам **(2.4)** и **(2.5)**

$$D(X) = (-7-0)^2 \cdot 0{,}990 + (193-0)^2 \cdot 0{,}005 +$$
$$+ (243-0)^2 \cdot 0{,}004 + (4993-0)^2 \cdot 0{,}001 = 25401,$$

$$\sigma_x = \sqrt{D(X)} = \sqrt{25401} = 159{,}38 \text{ (ден. ед.)} \,. \blacktriangleright$$

С в о й с т в а дисперсии случайной величины:
1) $D(C) = 0$, где C — постоянная величина;
2) $D(kX) = k^2 D(X)$;
3) $D(X) = M(X^2) - a^2$, где $a = M(X)$;
4) $D(X + Y) = D(X - Y) = D(X) + D(Y)$, где X и Y — независимые случайные величины.

▶ **Пример 2.4.** Найти математическое ожидание и дисперсию случайной величины $Z = 8X - 5Y + 7$, если даны $M(X) = 3$, $M(Y) = 2$, $D(X) = 1{,}5$ и $D(Y) = 1$ и известно, что X и Y — независимые случайные величины.

Р е ш е н и е. Используя свойства математического ожидания и дисперсии, вычислим:

$M(Z) = 8M(X) - 5M(Y) + 7 = 8 \cdot 3 - 5 \cdot 2 + 7 = 21;$
$D(Z) = 8^2 D(X) + 5^2 D(Y) + 0 = 8^2 \cdot 1{,}5 + 5^2 \cdot 1 = 121. \blacktriangleright$

2.2. Функция распределения случайной величины. Непрерывные случайные величины

Функцией распределения *случайной величины* X *называется функция* $F(x)$, *выражающая для каждого* x *вероятность того, что случайная величина* X *примет значение, меньшее* x:

$$F(x) = P(X < x). \qquad \textbf{(2.6)}$$

▶ **Пример 2.5.** Дан ряд распределения случайной величины

X:

x_i	1	4	5
p_i	0,4	0,1	0,5

Найти и изобразить графически ее функцию распределения.

Р е ш е н и е . В соответствии с определением

$$F(x) = 0 \text{ при } x \le 1;\ F(x) = 0{,}4 \text{ при } 1 < x < 4;$$

$$F(x) = 0{,}4 + 0{,}1 = 0{,}5 \text{ при } 4 < x \le 5;\ F(x) = 0{,}5 + 0{,}5 = 1 \text{ при } x > 5.$$

Итак (см. рис. 2.1):

$$F(x) = \begin{cases} 0 & \text{при } x \le 1; \\ 0{,}4 & \text{при } 1 < x \le 4; \\ 0{,}5 & \text{при } 4 < x \le 5; \\ 1 & \text{при } x > 5. \end{cases}$$

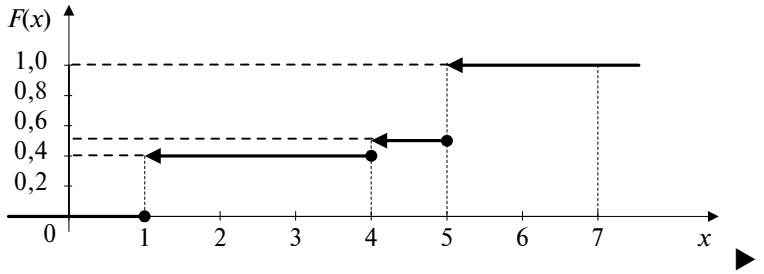


Рис. 2.1

С в о й с т в а функции распределения:

1. Функция распределения случайной величины есть неотрицательная функция, заключенная между нулем и единицей:

$$0 \le F(x) \le 1.$$

2. Функция распределения случайной величины есть неубывающая функция на всей числовой оси, т.е. при $x_2 > x_1$

$$F(x_2) \geq F(x_1).$$

3. На минус бесконечности функция распределения равна нулю, на плюс бесконечности — равна единице, т.е.

$$F(-\infty) = \lim_{x \to -\infty} F(x) = 0, \quad F(+\infty) = \lim_{x \to +\infty} F(x) = 1.$$

4. Вероятность попадания случайной величины X в интервал $[x_1, x_2)$ (включая x_1) равна приращению ее функции распределения на этом интервале, т.е.

$$P(x_1 \leq X < x_2) = F(x_2) - F(x_1). \qquad \textbf{(2.7)}$$

▶ **Пример 2.6**. Функция распределения случайной величины X имеет вид:

$$F(x) = \begin{cases} 0 & \text{при} \quad x \leq 0; \\ x/2 & \text{при} \quad 0 < x \leq 2; \\ 1 & \text{при} \quad x > 2. \end{cases}$$

Найти вероятность того, что случайная величина X примет значение в интервале [1;3).

Р е ш е н и е . По формуле (**2.7**)

$$P(1 \leq X < 3) = F(3) - F(1) = 1 - \frac{1}{2} = \frac{1}{2}. \quad ▶$$

Случайная величина X называется ***непрерывной****, если ее функция распределения непрерывна в любой точке и дифференцируема всюду, кроме, быть может, отдельных точек.*

Для непрерывной случайной величины X вероятность любого отдельно взятого значения равна нулю, т.е. $P(X = x_1) = 0$, а вероятность попадания X в интервал (x_1, x_2) не зависит от того, является ли этот интервал открытым или закрытым (т.е., например, $P(x_1 < X < x_2) = P(x_1 \leq X \leq x_2)$).

Плотностью вероятности (***плотностью распределения*** или просто ***плотностью***) $\varphi(x)$ *непрерывной случайной величины X называется производная ее функции распределения*

$$\varphi(x) = F'(x). \qquad \textbf{(2.8)}$$

Плотность вероятности $\varphi(x)$, как и функция распределения $F(x)$, является одной из форм закона распределения, но в отличие от функции распределения она существует только для н е п р е р ы в н ы х случайных величин.

График плотности вероятности называется *кривой распределения.*

▶ **Пример 2.7.** По данным примера **2.6** найти плотность вероятности случайной величины X.

Р е ш е н и е. Плотность вероятности $\varphi(x) = F'(x)$, т. е.

$$\varphi(x) = \begin{cases} 0 & \text{при} \quad x \le 0 \text{ и } x > 2, \\ 1/2 & \text{при} \quad 0 < x \le 2. \end{cases}$$ ▶

С в о й с т в а плотности вероятности непрерывной случайной величины:

1. Плотность вероятности — неотрицательная функция, т.е. $\varphi(x) \ge 0$.

2. Вероятность попадания непрерывной случайной величины в интервал $[a,b]$ равна определенному интегралу от ее плотности вероятности в пределах от a до b (см. рис. 2.2), т.е.

$$P(a \le X \le b) = \int_a^b \varphi(x)dx\,. \qquad \textbf{(2.9)}$$

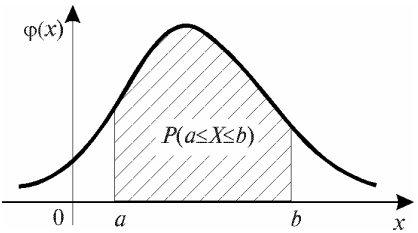


Рис. 2.2

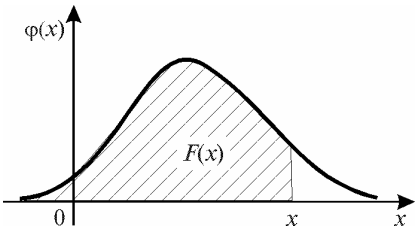


Рис. 2.3

3. Функция распределения непрерывной случайной величины (см. рис. 2.3) может быть выражена через плотность вероятности по формуле:

$$F(x) = \int_{-\infty}^{x} \varphi(x)\,dx\,. \qquad \textbf{(2.10)}$$

4. Несобственный интеграл в бесконечных пределах от плотности вероятности непрерывной случайной величины равен единице:

$$\int_{-\infty}^{+\infty} \varphi(x)\,dx = 1. \qquad \textbf{(2.11)}$$

Геометрически свойства *1* и *4* плотности вероятности означают, что ее график — *кривая распределения — лежит не ниже оси абсцисс, и полная площадь фигуры, ограниченной кривой распределения и осью абсцисс, равна единице*.

Для непрерывной случайной величины X математическое ожидание $M(X)$ и дисперсия $D(X)$ определяются по формулам:

$$a = M(X) = \int_{-\infty}^{+\infty} x\,\varphi(x)\,dx \qquad \textbf{(2.12)}$$

(если интеграл абсолютно сходится);

$$D(X) = \int_{-\infty}^{+\infty} (x-a)^2\,\varphi(x)\,dx\,, \qquad \textbf{(2.13)}$$

или $$D(X) = \int_{-\infty}^{+\infty} x^2\varphi(x)\,dx - a^2$$

(если приведенные интегралы сходятся).

Наряду с отмеченными выше числовыми характеристиками для описания случайной величины используется понятие квантилей и процентных точек.

Квантилем уровня q (или ***q-квантилем***) *называется такое значение x_q случайной величины, при котором функция ее распределения принимает значение, равное q*, т. е.

$$F(x_q) = P\,(X < x_q) = q. \qquad \textbf{(2.14)}$$

$100\boldsymbol{q}$***%-ой точкой*** *называется квантиль x_{1-q}.*

▶ **Пример 2.8.**

По данным примера **2.6** найти квантиль $x_{0,3}$ и 30%-ную точку случайной величины X.

Р е ш е н и е. По определению **(2.16)** $F(x_{0,3}) = 0{,}3$, т. е. $\dfrac{x_{0,3}}{2} = 0{,}3$, откуда квантиль $x_{0,3} = 0{,}6$. 30%-ная точка случайной величины X, или квантиль $x_{1-0,3} = x_{0,7}$, находится аналогично из уравнения $\dfrac{x_{0,7}}{2} = 0{,}7$, откуда $x_{0,7} = 1{,}4$. ▶

Среди числовых характеристик случайной величины выделяют ***начальные*** v_k и ***центральные*** μ_k ***моменты k-го порядка***, определяемые для дискретных и непрерывных случайных величин по формулам:

$$v_k = \sum_{i=1}^{n} x_i^k p_i ; \qquad v_k = \int_{-\infty}^{+\infty} x^k \varphi(x)\, dx ;$$

$$\mu_k = \sum_{i=1}^{n} (x_i - a)^k p_i ; \qquad \mu_k = \int_{-\infty}^{+\infty} (x-a)^k \varphi(x)\, dx .$$

2.3. Некоторые распределения случайных величин

Рассмотрим наиболее часто используемые в эконометрике распределения случайных величин.

1. *Дискретная случайная величина X имеет* ***биномиальный закон распределения****, если она принимает значения* 0, 1, 2,..., *m*,..., *n с вероятностями*

$$P(X = m) = C_n^m p^m q^{n-m}, \qquad \textbf{(2.15)}$$

где $0 < p < 1$, $q = 1 - p$, $m = 0, 1, \ldots, n$.

Биномиальный закон распределения представляет собой закон распределения числа $X = m$ наступлений события A в n независимых испытаниях, в каждом из которых оно может произойти с одной и той же вероятностью p.

Формула **(2.15)** называется *формулой Бернулли*.

Числовые характеристики: $M(X) = np$, $D(X) = npq$.

В частности, для *частости события* $W = \dfrac{m}{n}$ в n независимых испытаниях (где $X = m$ имеет биномиальный закон распределения с параметром p) числовые характеристики:

$$M(W) = p, \; D(W) = \frac{pq}{n}.$$

2. *Дискретная случайная величина X имеет* ***закон распределения Пуассона****, если она принимает значения* 0,1, 2,..., *m*,... (*бесконечное, но счетное множество значений*) *с вероятностями*

$$P(X = m) = \frac{\lambda^m e^{-\lambda}}{m!}, \qquad \textbf{(2.16)}$$

где $m = 0, 1, 2, \ldots$.

Числовые характеристики: $M(X) = \lambda$, $D(X) = \lambda$.

3. *Непрерывная случайная величина X имеет* ***равномерный закон распределения*** *на отрезке* [*a*, *b*], *если ее плотность вероятности постоянна на этом отрезке и равна нулю вне его,* т.е.

$$\varphi(x)=\begin{cases}\dfrac{1}{b-a} & \text{при} \quad a\le x\le b;\\ 0 & \text{при} \quad x<a,\ x>b.\end{cases} \qquad \textbf{(2.17)}$$

Числовые характеристики: $M(X)=\dfrac{a+b}{2}$, $D(X)=\dfrac{(b-a)^2}{12}$.

4. *Непрерывная случайная величина X имеет* ***показательный (экспоненциальный) закон распределения*** *с параметром* λ, *если ее плотность вероятности имеет вид:*

$$\varphi(x)=\begin{cases}\lambda e^{-\lambda x} & \text{при} \quad x\ge 0;\\ 0 & \text{при} \quad x<0.\end{cases} \qquad \textbf{(2.18)}$$

Числовые характеристики: $M(X)=\dfrac{1}{\lambda}$, $D(X)=\dfrac{1}{\lambda^2}$.

5. *Непрерывная случайная величина X имеет* ***нормальный закон распределения*** **(*закон Гаусса*)** *с параметрами a и* σ^2, *если ее плотность вероятности имеет вид* :

$$\varphi_N(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}. \qquad \textbf{(2.19)}$$

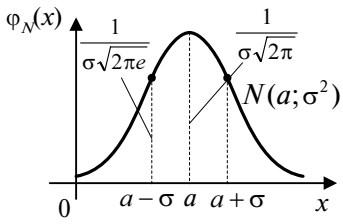


Рис. 2.4

Кривую нормального закона распределения называют *нормальной* или *гауссовой кривой* (рис. 2.4).

Числовые характеристики:

$$M(X)=a,\ D(X)=\sigma^2.$$

При изменении только параметра *a* нормальная кривая перемещается вдоль оси *Ox*, при изменении только параметра σ^2 меняется форма нормальной кривой.

Нормальный закон распределения с параметрами $a=0$, $\sigma^2=1$, т.е. $N(0;1)$, называется *стандартным* или *нормированным*.

Функция распределения случайной величины *X*, распределенной по нормальному закону, выражается через функцию Лапласа $\Phi(x)$ по формуле:

$$F_N(x) = \frac{1}{2} + \frac{1}{2}\Phi\left(\frac{x-a}{\sigma}\right), \qquad \textbf{(2.20)}$$

где $\Phi(x) = \frac{2}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt$ — *функция* (*интеграл вероятностей*) *Лапласа*, равная площади под стандартной нормальной кривой $N(0;1)$ на отрезке $[-x, x]$.

С в о й с т в а случайной величины, распределенной по нормальному закону:

1) Вероятность попадания случайной величины X, распределенной по нормальному закону, в интервал $[x_1, x_2]$, равна

$$P(x_1 \le X \le x_2) = \frac{1}{2}\left[\Phi(t_2) - \Phi(t_1)\right], \qquad \textbf{(2.21)}$$

где $t_1 = \frac{x_1 - a}{\sigma}$, $t_2 = \frac{x_2 - a}{\sigma}$.

2) Вероятность того, что отклонение случайной величины X, распределенной по нормальному закону, от математического ожидания a не превысит величину $\Delta > 0$ (по абсолютной величине), равна

$$P(|X - a| \le \Delta) = \Phi(t), \qquad \textbf{(2.22)}$$

где $t = \frac{\Delta}{\sigma}$.

Из второго свойства вытекает, в частности, **правило трех сигм**:

Если случайная величина X имеет нормальный закон распределения с параметрами a и σ^2, т.е. $N(a;\sigma^2)$, *то практически достоверно, что ее значения заключены в интервале* $(a - 3\sigma,\ a + 3\sigma)$.

6. *Непрерывная случайная величина X имеет* ***логарифмически нормальное*** (*сокращенно — **логнормальное распределение***), *если ее логарифм подчинен нормальному закону.*

7. ***Распределением*** χ^2 (***хи-квадрат***) с k *степенями свободы называется распределение суммы квадратов k независимых случайных величин, распределенных по стандартному нормальному закону*, т.е.

$$\chi^2 = \sum_{i=1}^{k} Z_i^2, \qquad \textbf{(2.23)}$$

где Z_i (i = 1, 2,..., k) имеет нормальное распределение $N(0;1)$.

При $k > 30$ распределение случайной величины $Z = \sqrt{2\chi^2} - \sqrt{2k-1}$ близко к стандартному нормальному закону, т.е. $N(0;1)$.

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

$$t = \frac{Z}{\sqrt{\frac{1}{k}\chi^2}}, \qquad \textbf{(2.24)}$$

где Z — случайная величина, распределенная по стандартному нормальному закону, т.е. $N(0;1)$; χ^2 — не зависимая от Z случайная величина, имеющая χ^2-распределение с k степенями свободы.

При $k \to +\infty$ t-распределение приближается к нормальному. Практически уже при $k > 30$ можно считать t-распределение приближенно нормальным.

9. ***Распределением Фишера—Снедекора*** (*или* ***F-распределением***) *называется распределение случайной величины*

$$F = \frac{\frac{1}{k_1}\chi^2(k_1)}{\frac{1}{k_2}\chi^2(k_2)}, \qquad \textbf{(2.25)}$$

где $\chi^2(k_1)$ и $\chi^2(k_2)$ — случайные величины, имеющие χ^2-распределение соответственно с k_1 и k_2 степенями свободы.

2.4. Многомерные случайные величины. Условные законы распределения

Упорядоченный набор $X=(X_1, X_2,..., X_n)$ случайных величин называется ***многомерной*** (***n-мерной***) ***случайной величиной*** (*или* ***системой случайных величин, n-мерным вектором***).

Например, погода в данном месте в определенное время суток может быть охарактеризована многомерной случайной величиной $X=(X_1, X_2,..., X_n)$, где X_1 — температура, X_2 — влажность, X_3 — давление, X_4 — скорость ветра и т.п.

Функцией распределения *n-мерной случайной величины* $(X_1, X_2,..., X_n)$ называется функция $F(x_1, x_2,...,x_n)$, выражающая вероятность совместного выполнения n неравенств $X_1<x_1$, $X_2<x_2$,..., $X_n<x_n$, т.е.

$$F(x_1, x_2,..., x_n) = P(X_1 < x_1, X_2 < x_2,..., X_n < x_n). \qquad \textbf{(2.26)}$$

В двумерном случае[1] для случайной величины (X, Y) функция распределения $F(x, y)$ определится равенством:

$$F(x, y) = P(X < x,\ Y < y). \qquad \textbf{(2.27)}$$

С в о й с т в а функции распределения $F(x, y)$, аналогичные свойствам одномерной случайной величины:

1) $0 \le F(x, y) \le 1$;

2) при $x_2 > x_1$ $F(x_2, y) \ge F(x_1, y)$; при $y_2 > y_1$ $F(x, y_2) \ge F(x, y_1)$;

3) $F(x, -\infty) = F(-\infty, y) = F(-\infty, -\infty) = 0$;

4) $F(x, +\infty) = F_1(x)$, $F(+\infty, y) = F_2(y)$, где $F_1(x)$ и $F_2(y)$ — функции распределения случайных величин X и Y;

5) $F(+\infty, +\infty) = 1$.

Плотностью вероятности (***плотностью распределения*** *или* ***совместной плотностью***) *непрерывной двумерной случайной величины* (X, Y) *называется вторая смешанная частная производная ее функции распределения*, т.е.

$$\varphi(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} = F''xy(x, y). \qquad \textbf{(2.28)}$$

С в о й с т в а плотности вероятности двумерной случайной величины $\varphi(x, y)$ аналогичны свойствам плотности вероятности одномерной случайной величины:

1) $\varphi(x, y) \ge 0$;

2) $P[(X, Y) \in D] = \iint\limits_D \varphi(x, y)\, dx\, dy$;

3) $F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} \varphi(x, y)\, dx\, dy$;

4) $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \varphi(x, y)\, dx\, dy = 1$.

Условным законом распределения *одной из одномерных составляющих двумерной случайной величины* (X, Y) *называется ее закон распределения, вычисленный при условии, что другая составляющая приняла определенное значение* (*или попала в какой-то интервал*).

Условные плотности вероятности $\varphi_y(x)$ и $\varphi_x(y)$ двумерной случайной величины (X, Y) определяются по формулам:

---

[1] В дальнейшем для простоты изложение ведем в основном для двумерной (n=2) случайной величины, при этом практически все понятия и утверждения могут быть перенесены на случай n>2.

$$\varphi_y(x) = \frac{\varphi(x,y)}{\varphi_2(y)}, \quad \varphi_x(y) = \frac{\varphi(x,y)}{\varphi_1(x)} \qquad \textbf{(2.29)}$$

или

$$\varphi_y(x) = \frac{\varphi(x,y)}{\int_{-\infty}^{+\infty}\varphi(x,y)\,dx}; \quad \varphi_x(y) = \frac{\varphi(x,y)}{\int_{-\infty}^{+\infty}\varphi(x,y)\,dy}.$$

Условные плотности $\varphi_y(x)$ и $\varphi_x(y)$ обладают всеми свойствами «безусловной» плотности, рассмотренной в § **2.2.**

Числовые характеристики условных распределений: условные математические ожидания $M_x(Y)$ и $M_y(X)$ и условные дисперсии $D_x(Y)$ и $D_y(X)$. Эти характеристики находятся по обычным формулам математического ожидания и дисперсии, в которых вместо вероятностей событий или плотностей вероятности используются условные вероятности или условные плотности вероятности.

Условное математическое ожидание случайной величины Y при X=x, т. е. $M_x(Y)$, *есть функция от x*, называемая ***функцией регрессии*** или просто ***регрессией*** Y по X; аналогично $M_y(X)$ называется *функцией регрессии* или просто регрессией X по Y. Графики этих функций называются соответственно *линиями регрессии* (или *кривыми регрессии*) Y по X и X по Y.

С в о й с т в а условного математического ожидания:

1. Если $Z = g(X)$, где g — некоторая неслучайная функция от X, то

$$M_Z(M_x(Y)) = M_Z(Y).$$

В частности,

$$M(M_x(Y) = M(Y).$$

2. Если $Z = g(X)$, то $M_x(ZY) = ZM_x(Y)$.
3. Если случайные величины X и Y независимы, то

$$M_x(Y) = M(Y).$$

Для независимых случайных величин

$$\varphi(x,y) = \varphi_1(x)\cdot\varphi_2(y) \text{ или } \varphi_y(x) = \varphi_1(x) \text{ и } \varphi_x(y) = \varphi_2(y),$$

где $\varphi_1(x), \varphi_2(y)$ — плотности соответствующих одномерных распределений случайных величин X и Y; $\varphi_y(x), \varphi_x(y)$ — плотности их условных распределений X по Y и Y по X.

Зависимость между двумя случайными величинами называется ***вероятностной*** (***стохастической*** или ***статистической***), *если каждому значению одной из них соответствует определенное* (*условное*) *распределение другой.*

Например, зависимость между урожайностью и количеством внесенных удобрений — вероятностная.

Ковариацией (*или* ***корреляционным моментом***) $\mathrm{Cov}(X, Y)$ *случайных величин* X *и* Y *называется математическое ожидание произведения отклонений этих величин от своих математических ожиданий*, т.е.

$$\mathrm{Cov}(X, Y) = M[(X - a_x)(Y - a_y)], \qquad \textbf{(2.30)}$$

где $a_x = M(X)$, $a_y = M(Y)$.

(Для ковариации случайных величин X и Y используются также обозначения K_{xy}, σ_{xy}.)

Ковариация двух случайных величин характеризует как *степень зависимости* случайных величин, так и их *рассеяние* вокруг точки (a_x, a_y). Ковариация — величина размерная, что затрудняет ее использование для оценки степени зависимости случайных величин. Этих недостатков лишен коэффициент корреляции.

Коэффициентом корреляции *двух случайных величин называется отношение их ковариации к произведению средних квадратических отклонений этих величин*:

$$\rho = \frac{\mathrm{Cov}(X, Y)}{\sigma_x \sigma_y}. \qquad \textbf{(2.31)}$$

Из определения следует, что коэффициент корреляции — величина *безразмерная* — характеризует *тесноту линейной зависимости* между случайными величинами.

С в о й с т в а ковариации двух случайных величин:

1) $\mathrm{Cov}(X, Y) = 0$, если X и Y независимы;

2) $\mathrm{Cov}(X, Y) = M(X, Y) - a_x a_y$;

3) $|\mathrm{Cov}(X, Y)| \le \sigma_x \sigma_y$.

С в о й с т в а коэффициента корреляции:

1) $-1 \le \rho \le 1$;

2) $\rho = 0$, если случайные величины X и Y независимы;

3) если $|\rho| = 1$, то между случайными величинами X и Y существует линейная функциональная зависимость.

Из *независимости двух случайных величин следует их некоррелированность*, т.е. равенство $\rho = 0$. Однако некоррелированность двух случайных величин еще не означает их независимость.

2.5. Двумерный (*n*-мерный) нормальный закон распределения

Случайная величина (*случайный вектор*) (*X, Y*) *называется распределенной по* ***двумерному нормальному закону****, если ее совместная плотность имеет вид*:

$$\varphi_N(x,y)=\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}e^{-L(x,y)}, \qquad \textbf{(2.32)}$$

где

$$L(x,y)=\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-a_x}{\sigma_x}\right)^2-2\rho\frac{x-a_x}{\sigma_x}\cdot\frac{y-a_y}{\sigma_y}+\left(\frac{y-a_y}{\sigma_y}\right)^2\right]. \qquad \textbf{(2.33)}$$

Числовые характеристики: $M(X) = a_x$, $M(Y) = a_y$, $D(X) = \sigma_x^2$, $D(Y) = \sigma_y^2$, $\rho_{xy} = \rho$.

При этом одномерные случайные величины X и Y распределены нормально с параметрами соответственно (a_x, σ_x^2), (a_y, σ_y^2).

Условные законы распределения X по Y и Y по X — также нормальные с числовыми характеристиками:

$$M_y(X)=a_x+\rho\frac{\sigma_x}{\sigma_y}(y-a_y), \qquad \textbf{(2.34)}$$

$$D_y(X)=\sigma_x^2(1-\rho^2); \qquad \textbf{(2.35)}$$

$$M_x(Y)=a_y+\rho\frac{\sigma_y}{\sigma_x}(x-a_x), \qquad \textbf{(2.36)}$$

$$D_x(Y)=\sigma_y^2(1-\rho^2). \qquad \textbf{(2.37)}$$

Из формул **(2.34)**, **(2.36)** следует, что *линии регрессии* $M_y(X)$ и $M_x(Y)$ *нормально распределенных случайных величин представляют собой прямые линии, т. е. нормальные регрессии Y по X и X по Y всегда линейны.*

Для *нормально распределенных случайных величин термины «некоррелированность» и «независимость» равносильны.*

Понятие двумерного ($n = 2$) нормального закона обобщается для любого натурального n.

Нормальный закон распределения n-мерной случайной величины (*n-мерного случайного вектора*) $X = (X_1, X_2, ..., X_n)$ характеризуется параметрами, задаваемыми вектором средних $a = (a_1, a_2, ..., a_n)'$ и ковариационной матрицей $\sum_x = (\sigma_{ij})_{n\times n}$, где $\sigma_{ij} = M[(X_i - a_i)(X_j - a_j)]$.

Ковариационная матрица и ее определитель, называемый *обобщенной дисперсией n-мерной случайной величины*, являются аналогами дисперсии одномерной случайной величины и характеризуют степень случайного разброса отдельно по каждой составляющей и в целом по *n*-мерной величине.

2.6. Закон больших чисел и предельные теоремы

Под ***законом больших чисел*** в широком смысле понимается *общий принцип, согласно которому*, по формулировке академика А.Н. Колмогорова, *совокупное действие большого числа случайных факторов приводит* (*при некоторых весьма общих условиях*) *к результату, почти не зависящему от случая*. Другими словами, при большом числе случайных величин их средний результат перестает быть случайным и может быть предсказан с большой степенью определенности.

Теорема Чебышева. *Если дисперсии n независимых случайных величин* $X_1, X_2, \ldots, X_n$ *ограничены одной и той же постоянной, то при неограниченном увеличении числа n средняя арифметическая случайных величин сходится по вероятности*[1] *к средней арифметической их математических ожиданий* a_1, a_2,...,a_n, т. е.

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - \frac{a_1 + a_2 + \ldots + a_n}{n}\right| \le \varepsilon\right) = 1, \quad \textbf{(2.38)}$$

или

$$\frac{\sum_{i=1}^{n} X_i}{n} \xrightarrow[n\to\infty]{\mathcal{P}} \frac{\sum_{i=1}^{n} a_i}{n}$$

Теорема Бернулли. *Частость события в n повторных независимых испытаниях, в каждом из которых оно может произойти с одной и той же вероятностью p, при неограниченном увеличении числа n сходится по вероятности к вероятности p этого события в отдельном испытании*, т.е.

$$\lim_{n\to\infty} P\left(\left|\frac{m}{n} - p\right| \le \varepsilon\right) = 1, \quad \textbf{(2.39)}$$

[1] Подробнее о понятии «сходимость по вероятности» см., например, [17].

или $$\frac{m}{n} \xrightarrow[n\to\infty]{\mathscr{P}} p\,.$$

Согласно **теореме Ляпунова**[1], *если независимые случайные величины* X_1, X_2,..., X_n *имеют конечные математические ожидания и дисперсии, по своему значению ни одна из этих случайных величин резко не выделяется среди остальных, то при* $n\to\infty$ *закон распределения их суммы* $\sum_{i=1}^{n} X_i$ *неограниченно приближается к нормальному.*

В частности, если X_1, X_2,..., X_n *одинаково распределены*, то закон распределения их суммы $\sum_{i=1}^{n} X_i$ при $n\to\infty$ неограниченно приближается к нормальному.

2.7. Точечные и интервальные оценки параметров

Оценкой $\widetilde{\theta}_n$ *параметра* θ *называют всякую функцию результатов наблюдений над случайной величиной* X (*иначе — статистику*), *с помощью которой судят о значениях параметра* θ.

В отличие от параметра, его оценка $\widetilde{\theta}_n$ — величина *случайная.* «*Наилучшая оценка*» $\widetilde{\theta}_n$ *должна обладать наименьшим рассеянием относительно оцениваемого параметра* θ, например, наименьшей величиной математического ожидания квадрата отклонения оценки от оцениваемого параметра $M(\widetilde{\theta}_n - \theta)^2$.

Оценка $\widetilde{\theta}_n$ *параметра* θ *называется* ***несмещенной***, *если ее математическое ожидание равно оцениваемому параметру*, т. е. $M(\widetilde{\theta}_n) = \theta$.

В противном случае оценка называется ***смещенной.***

Если это равенство не выполняется, то оценка $\widetilde{\theta}_n$, полученная по разным выборкам, будет в среднем либо завышать значение θ (если $M(\widetilde{\theta}_n) > \theta$, либо занижать его (если $M(\widetilde{\theta}_n) < 0$). Таким образом, требование несмещенности гарантирует отсутствие систематических ошибок при оценивании.

---

[1] Приведенная формулировка не является точной и дает лишь понятие о теореме Ляпунова.

Оценка $\widetilde{\theta}_n$ *параметра* θ *называется* ***состоятельной***, *если она удовлетворяет закону больших чисел, т.е. сходится по вероятности к оцениваемому параметру*:

$$\lim_{n\to\infty} P\left(\left|\widetilde{\theta}_n - \theta\right| \le \varepsilon\right) = 1, \qquad \textbf{(2.40)}$$

или

$$\widetilde{\theta}_n \xrightarrow[n\to\infty]{\mathscr{P}} \theta.$$

В случае использования состоятельных оценок оправдывается увеличение объема выборки, так как при этом становятся маловероятными значительные ошибки при оценивании. Поэтому практический смысл имеют только состоятельные оценки.

Несмещенная оценка $\widetilde{\theta}_n$ *параметра* θ *называется* ***эффективной***, *если она имеет наименьшую дисперсию среди всех возможных несмещенных оценок параметра* θ, *вычисленных по выборкам одного и того же объема n.*

Так как для несмещенной оценки $M(\widetilde{\theta}_n - \theta)^2$ есть ее дисперсия $\sigma^2_{\widetilde{\theta}_n}$, то эффективность является *решающим свойством*, определяющим качество оценки.

Для нахождения оценок параметров (характеристик) генеральной совокупности используется ряд методов.

Согласно **методу моментов** определенное количество выборочных моментов (начальных ν_k или центральных μ_k, или тех и других) приравнивается к соответствующим теоретическим моментам распределения ν_k или μ_k случайной величины *X*.

Основным методом получения оценок параметров генеральной совокупности по данным выборки является **метод максимального правдоподобия**.

Основу метода составляет ***функция правдоподобия,*** *выражающая плотность вероятности* (*вероятность*) *совместного появления результатов выборки* $x_1, x_2, \ldots, x_n$:

$$L(x_1, x_2, \ldots, x_n; \theta) = \varphi(x_1,\theta)\, \varphi(x_2,\theta) \ldots \varphi(x_i,\theta) \ldots \varphi(x_n,\theta),$$

или

$$L(x_1, x_2, \ldots, x_n; \theta) = \prod_{i=1}^{n} \varphi(x_i,\theta). \qquad \textbf{(2.41)}$$

Согласно методу максимального правдоподобия в качестве оценки неизвестного параметра θ принимается такое значение $\widetilde{\theta}_n$, которое *максимизирует* функцию L. Естественность подобного подхода к определению статистических оценок вытекает из смысла функции правдоподобия, которая при каждом фиксированном значении параметра является мерой правдоподобности получения наблюдений. И оценка $\widetilde{\theta}_n$ такова, что имеющиеся у нас наблюдения являются наиболее правдоподобными.

Достоинством метода максимального правдоподобия является то, что получаемые им *оценки состоятельны, асимптотически* (*при* $n\to\infty$) *эффективны и имеют асимптотически* (при $n\to\infty$) *нормальное распределение.*

Пусть для оценки параметров распределения случайной величины X отобрана случайная выборка, состоящая из значений $x_1, x_2,..., x_n$. Найдены выборочные характеристики: $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i n_i}{n}$ — выборочная средняя; $s^2 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2 n_i}{n}$ — выборочная дисперсия (где n_i — частоты значений x_i); $w = \dfrac{m}{n}$ — выборочная доля (где m элементов выборки из n обладают данным признаком).

Тогда $\bar{x}$ и w — *несмещенные, состоятельные и эффективные* (для нормально распределенной генеральной совокупности) оценки соответственно *математического ожидания* a и *вероятности* p, а s^2 — *смещенная, но состоятельная оценка дисперсии* σ^2.

Исправленная выборочная дисперсия $\hat{s}^2 = \dfrac{n}{n-1} s^2 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2 n_i}{n-1}$ — *несмещенная и состоятельная оценка дисперсии* σ^2.

Наряду с *точечными* оценками параметров (в виде одного числа) рассматривают интервальные оценки.

Интервальной оценкой *параметра* θ *называется числовой интервал* $(\widetilde{\theta}_n^{(1)}, \widetilde{\theta}_n^{(2)})$, *который с заданной вероятностью* γ *накрывает неизвестное значение параметра* θ.

Такой интервал ($\widetilde{\theta}_n^{(1)}$, $\widetilde{\theta}_n^{(2)}$) называется ***доверительным***, а вероятность γ — ***доверительной вероятностью*** или ***надежностью оценки.***

Величина доверительного интервала существенно зависит от объема выборки n (уменьшается с ростом n) и от значения доверительной вероятности γ (увеличивается с приближением γ к единице).

Приведем примеры построения доверительных интервалов (см. [17]).

Пусть: x_1, x_2,..., x_n — повторная выборка из *нормально распределенной* генеральной совокупности;

$\bar{x}$ — выборочная средняя;

$\bar{x}_0$ — генеральная средняя;

s — выборочное среднее квадратическое отклонение;

$\sigma_{\bar{x}} = \sqrt{\dfrac{s^2}{n-1}}$ — среднее квадратическое отклонение выборочной средней (средняя квадратическая ошибка выборки).

Учитывая, что статистика $\dfrac{\bar{x}-\bar{x}_0}{\sigma_{\bar{x}}} = \dfrac{\bar{x}-\bar{x}_0}{s}\sqrt{n-1}$ имеет t-распределение Стьюдента с $n-1$ степенями свободы, *доверительный интервал для генеральной средней* $\bar{x}_0$ на уровне значимости α определяется по формуле:

$$\left(\bar{x} - t_{1-\alpha,n-1}\frac{s}{\sqrt{n-1}}, \bar{x} + t_{1-\alpha,n-1}\frac{s}{\sqrt{n-1}}\right). \qquad \textbf{(2.42)}$$

Учитывая, что статистика $\dfrac{ns^2}{\sigma^2}$ имеет χ^2-распределение с $n-1$ степенями свободы, *доверительный интервал для генеральной дисперсии* σ^2 на уровне значимости α определяется по формуле:

$$\left(\frac{ns^2}{\chi^2_{\alpha/2,\,n-1}}, \frac{ns^2}{\chi^2_{1-\alpha/2,\,n-1}}\right). \qquad \textbf{(2.43)}$$

2.8. Проверка (тестирование) статистических гипотез

Статистической гипотезой *называется любое предположение о виде или параметре неизвестного закона распределения.*

Проверяемую гипотезу обычно называют *нулевой* и обозначают H_0. Наряду с нулевой гипотезой H_0 рассматривают *альтернативную*, или *конкурирующую*, гипотезу H_1, являющуюся логическим отрицанием H_0. Нулевая и альтернативная гипотезы представляют собой две возможности выбора, осуществляемого в задачах проверки статистических гипотез.

Суть п р о в е р к и (т е с т и р о в а н и я) статистической гипотезы заключается в том, что используется специально составленная выборочная характеристика (статистика) $\widetilde{\theta}_n$ (x_1, x_2,..., x_n), полученная по выборке X_1, X_2,..., X_n, точное или приближенное распределение которой известно. Затем по этому выборочному распределению определяется критическое значение $\theta_{кр}$ — такое, что если гипотеза H_0 верна, то вероятность $P(\widetilde{\theta}_n > \theta_{кр})$ мала; так что в соответствии с принципом практической уверенности в условиях данного исследования событие $\widetilde{\theta}_n > \theta_{кр}$ можно (с некоторым риском) считать практически невозможным. Поэтому, если в данном конкретном случае обнаруживается отклонение $\widetilde{\theta}_n > \theta_{кр}$, то гипотеза H_0 отвергается, в то время как появление значения $\widetilde{\theta}_n \le \theta_{кр}$ считается совместимым с гипотезой H_0, которая тогда принимается (точнее, не отвергается). *Правило, по которому гипотеза H_0 отвергается или принимается, называется* ***статистическим критерием*** *или* ***статистическим тестом***.

Таким образом, множество возможных значений статистики критерия (критической статистики) $\widetilde{\theta}_n$ разбивается на два непересекающихся подмножества: *критическую область* (*область отклонения гипотезы*) W и *область допустимых значений* (*область принятия гипотезы*) $\overline{W}$. Если фактически наблюдаемое значение статистики критерия $\widetilde{\theta}_n$ попадает в критическую область W, то гипотезу H_0 отвергают. При этом возможны четыре случая:

Гипотеза H_0	Принимается	Отвергается
Верна	Правильное решение	Ошибка 1-го рода
Неверна	Ошибка 2-го рода	Правильное решение

Вероятность α допустить ошибку 1-го рода, т. е. отвергнуть гипотезу H_0, когда она верна, называется ***уровнем значимости критерия***.

Вероятность допустить ошибку 2-го рода, т. е. принять гипотезу H_0, когда она неверна, обычно обозначают β.

Вероятность $(1-\beta)$ *не допустить ошибку 2-го рода*, т. е. *отвергнуть гипотезу* H_0, *когда она неверна, называется* ***мощностью*** (*или* ***функцией мощности***) ***критерия***.

Вероятности ошибок 1-го и 2-го рода (α и β) однозначно определяются выбором критической области. Очевидно, желательно сделать как угодно малыми α и β. Однако это противоречивые требования: при фиксированном объеме выборки можно сделать как угодно малой лишь одну из величин — α или β, что сопряжено с неизбежным увеличением другой. *Лишь при увеличении объема выборки возможно одновременное уменьшение вероятностей* α и β.

Критическую область W следует выбирать так, чтобы вероятность попадания в нее статистики критерия $\widetilde{\theta}_n$ *была минимальной и равной* α, *если верна нулевая гипотеза* H_0, и *максимальной в противоположном случае*:

$$\begin{aligned} P\left(\widetilde{\theta}_n \in W / H_0\right) &= \alpha, \\ P\left(\widetilde{\theta}_n \in W / H_1\right) &= \max. \end{aligned} \qquad \textbf{(2.44)}$$

Другими словами, *критическая область должна быть такой, чтобы при заданном уровне значимости мощность критерия* $1-\beta$ *была максимальной.* Задача построения такой критической области (или, как говорят, построения наиболее мощного критерия) для простых гипотез решается с помощью теоремы Неймана—Пирсона, излагаемой в более полных курсах математической статистики.

В зависимости от вида конкурирующей гипотезы H_1 выбирают *правостороннюю*, *левостороннюю* или *двустороннюю критическую область.* Границы критической области при заданном уровне значимости α определяются соответственно из соотношений:

- для *правосторонней* критической области

$$P(\widetilde{\theta}_n > \theta_{\text{кр}}) = \alpha; \qquad \textbf{(2.45)}$$

- для *левосторонней* критической области

$$P(\widetilde{\theta}_n < \theta_{\text{кр}}) = \alpha; \qquad \textbf{(2.46)}$$

- для *двусторонней* критической области

$$P(\widetilde{\theta}_n < \theta_{\text{кр.1}}) = P(\widetilde{\theta}_n > \theta_{\text{кр.2}}) = \frac{\alpha}{2}. \qquad \textbf{(2.47)}$$

Следует отметить, что в компьютерных эконометрических пакетах обычно не находятся границы критической области $\theta_{кр}$, необходимые для сравнения их с фактически наблюдаемыми значениями выборочных характеристик $\widetilde{\theta}_{набл}$ и принятия решения о справедливости гипотезы H_0. А рассчитываются точные значения уровня значимости (*p-valio*), исходя из соотношения $P(\widetilde{\theta}_n > \widetilde{\theta}_{набл}) = p$. Если вероятность p очень мала, то гипотезу H_0 отвергают, в противном случае H_0 принимают.

Принцип проверки (*тестирования*) *статистической гипотезы не дает логического доказательства ее верности или неверности.* Принятие гипотезы H_0 следует расценивать не как раз и навсегда установленный, абсолютно верный содержащийся в ней факт, а лишь как достаточно правдоподобное, не противоречащее опыту утверждение.

Упражнения

2.9. Дан ряд распределения случайной величины X:

x_i	0	1	2	3
p_i	0,06	0,29	0,44	0,21

Необходимо: *а*) найти математическое ожидание $M(X)$, дисперсию $D(X)$ и среднее квадратическое (стандартное) отклонение σ случайной величины X; *б*) определить функцию распределения $F(x)$ и построить ее график.

2.10. Дана функция распределения случайной величины X:

$$F(x) = \begin{cases} 0 \quad \text{при} \quad x < 1; \\ 0{,}3 \text{ при } 1 < x \le 2; \\ 0{,}7 \text{ при } 2 < x \le 3; \\ 1 \quad \text{при} \quad x > 3. \end{cases}$$

Найти: *а*) ряд распределения; *б*) математическое ожидание $M(X)$ и дисперсию $D(X)$; *в*) построить многоугольник распределения и график $F(x)$.

2.11. Случайная величина X, сосредоточенная на интервале $[-1;3]$, задана функцией распределения $F(x) = \frac{1}{4}x + \frac{1}{4}$.

Найти вероятность попадания случайной величины X в интервал [0;2].

2.12. Случайная величина X задана функцией распределения

$$F(x)=\begin{cases}0 & \text{при} \quad x\le 0;\\ x^2 & \text{при} \quad 0<x\le 1;\\ 1 & \text{при} \quad x>1.\end{cases}$$

Найти: *а*) плотность вероятности $\varphi(x)$; *б*) математическое ожидание $M(X)$ и дисперсию $D(X)$; *в*) вероятности $P(X=0{,}5)$, $P(X<0{,}5)$, $P(0{,}5<X<1)$; *г*) построить графики $\varphi(x)$ и $F(x)$ и показать на них математическое ожидание $M(X)$ и вероятности, найденные в п. *в*); *д*) квантиль $x_{0,3}$ и 20%-ную точку распределения X.

2.13. Дана функция

$$\varphi(x)=\begin{cases}0 & \text{при} \quad x<0;\\ Cxe^{-x} & \text{при} \quad x\ge 0.\end{cases}$$

При каком значении параметра C эта функция является плотностью распределения некоторой случайной величины? Найти математическое ожидание и дисперсию случайной величины X.

2.14. Даны две случайные величины X и Y. Величина X распределена по биномиальному закону с параметрами $n = 19$, $p = 0{,}1$; величина Y распределена по закону Пуассона с параметром $\lambda=2$.

Построить ряды распределения случайных величин X и Y. Найти $M(X)$, $D(X)$; $M(Y)$, $D(Y)$; $P(X<2)$, $P(Y>1)$.

2.15. Даны две случайные величины X и Y; величина X распределена по равномерному закону на отрезке [0;1]; величина Y распределена по показательному закону с параметром $\lambda=\dfrac{1}{80}$.

Определить плотности вероятности и функции распределения случайных величин X и Y. Найти $P(X\ge 0{,}05)$, $P(Y\le 100)$.

2.16. Случайная величина X распределена по нормальному закону с параметрами $a = 15$, $\sigma^2 = 0{,}04$.

Написать выражения плотности и функции распределения случайной величины X. Найти вероятности $P(X\le 15{,}3)$, $P(X\ge 15{,}4)$, $P(14{,}9\le X\le 15{,}3)$, $P(X-15)\le 0{,}3$; квантиль $x_{0,6}$, 30%-ную точку распределения X. С помощью правила трех сигм определить границы для значения случайной величины X.

Глава 3
Парный регрессионный анализ

В практике экономических исследований имеющиеся данные не всегда можно считать выборкой из многомерной нормальной совокупности, когда одна из рассматриваемых переменных не является случайной или когда линия регрессии явно не прямая и т.п. В этих случаях пытаются определить кривую (поверхность), которая дает наилучшее (в смысле метода наименьших квадратов) приближение к исходным данным. Соответствующие методы приближения получили название *регрессионного анализа*.

Методы и модели регрессионного анализа занимают центральное место в математическом аппарате эконометрики.

Задачами регрессионного анализа *являются установление формы зависимости между переменными, оценка функции регрессии, оценка неизвестных значений* (*прогноз значений*) *зависимой переменной.*

3.1. Функциональная, статистическая и корреляционная зависимости

В естественных науках часто речь идет о ***функциональной зависимости*** (связи), когда каждому значению одной переменной соответствует вполне *определенное значение другой* (например, скорость свободного падения в вакууме в зависимости от времени и т.д.).

В экономике в большинстве случаев между переменными величинами существуют зависимости, когда каждому значению одной переменной соответствует не какое-то определенное, а м н о ж е с т в о возможных значений другой переменной. Иначе говоря, каждому значению одной переменной соответствует *определенное* (*условное*) *распределение другой переменной*. Такая зависимость получила название ***статистической*** (или ***стохастической***, ***вероятностной***) (о ней упоминалось в § **2.4**).

Возникновение понятия статистической связи обуславливается тем, что зависимая переменная подвержена влиянию ряда неконтролируемых или неучтенных факторов, а также тем, что измерение значений переменных неизбежно сопровождается некоторыми случайными ошибками. Примером статистической связи является зависимость урожайности от количества внесенных удобрений, производительности труда на предприятии от его энерговооруженности и т.п.

В силу неоднозначности статистической зависимости между Y и X для исследователя, в частности, представляет интерес усредненная по X схема зависимости, т. е. закономерность в изменении условного математического ожидания $M_x(Y)$ или $M(Y/X = x)$ (математического ожидания случайной переменной Y, вычисленного в предположении, что переменная X приняла значение x) в зависимости от x.

Если зависимость между двумя переменными такова, что *каждому значению одной переменной соответствует определенное условное математическое ожидание* (*среднее значение*) *другой, то такая статистическая зависимость называется* ***корреляционной.***

Иначе, ***корреляционной зависимостью*** *между двумя переменными называется функциональная зависимость между значениями одной из них и условным математическим ожиданием другой.*

Корреляционная зависимость может быть представлена в виде

$$M_x(Y) = \varphi(x) \qquad \textbf{(3.1)}$$

или

$$M_y(X) = \psi(y),$$

где $\varphi(x) \neq \text{const}$, $\psi(y) \neq \text{const}$.

В регрессионном анализе рассматриваются *односторонняя зависимость случайной переменной Y от одной* (*или нескольких*) *неслучайной независимой переменной X.* Такая зависимость может возникнуть, например, в случае, когда при каждом фиксированном значении X соответствующие значения Y подвержены случайному разбросу за счет действия ряда неконтролируемых факторов. Такая зависимость Y от X (иногда ее называют ***регрессионной***) может быть также представлена в виде модельного уравнения регрессии Y по X (**3.1**). При этом *зависимую* переменную Y называют также *функцией отклика*, *объясняемой*, *выходной*, *результирующей*, *эндогенной переменной*, *результативным признаком*, а независимую переменную X — *объясняющей, входной,*

предсказывающей, предикторной, экзогенной переменной, фактором, регрессором, факторным признаком.

Уравнение (**3.1**) называется *модельным уравнением регрессии* (или просто *уравнением регрессии*), а функция $\varphi(x)$ — *модельной функцией* регрессии (или просто *функцией регрессии*), а ее график — *модельной линией регрессии* (или просто *линией регрессии*).

Для точного описания уравнения регрессии необходимо знать условный закон распределения зависимой переменной Y при условии, что переменная X примет значение x, т. е. $X=x$. В статистической практике такую информацию получить, как правило, не удается, так как обычно исследователь располагает лишь выборкой пар значений (x_i, y_i) ограниченного объема n. В этом случае речь может идти об *оценке* (*приближенном выражении*, *аппроксимации*) по выборке функции регрессии. Такой оценкой[1] является *выборочная линия* (*кривая*) *регрессии*:

$$\hat{y} = \hat{\varphi}\left(x, b_0, b_1, ..., b_p\right), \qquad \textbf{(3.2)}$$

где $\hat{y}$ — *условная* (*групповая*) *средняя* переменной Y при фиксированном значении переменной $X = x$, b_0, b_1,..., b_p — параметры кривой.

Уравнение (**3.2**) называется *выборочным уравнением регрессии*.

При правильно определенной аппроксимирующей функции $\hat{\varphi}\left(x, b_0, b_1, ..., b_p\right)$ с увеличением объема выборки ($n \to \infty$) она будет сходиться по вероятности к функции регрессии $\varphi(x)$.

3.2. Линейная парная регрессия

Рассмотрим в качестве примера зависимость между сменной добычей угля на одного рабочего Y(т) и мощностью пласта X(м) по следующим (условным) данным, характеризующим процесс добычи угля в $n = 10$ шахтах.

Таблица 3.1

i	1	2	3	4	5	6	7	8	9	10
x_i	8	11	12	9	8	8	9	9	8	12
y_i	5	10	10	7	5	6	6	5	6	8

[1] Как мы увидим дальше, наилучшей оценкой является линия *среднеквадратической* регрессии.

Изобразим полученную зависимость графически точками координатной плоскости (рис. 3.1). Такое изображение статистической зависимости называется *полем корреляции*.

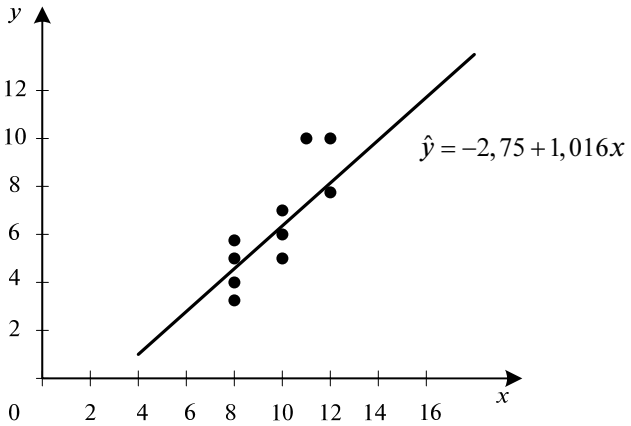


Рис. 3.1

По расположению эмпирических точек можно предполагать наличие л и н е й н о й корреляционной (регрессионной) зависимости между переменными X и Y.

Поэтому уравнение регрессии (**3.2**) будем искать в виде линейного уравнения

$$\hat{y} = b_0 + b_1 x . \qquad \textbf{(3.3)}$$

Отвлечемся на время от рассматриваемого примера и найдем формулы расчета неизвестных параметров уравнения линейной регрессии.

Согласно **методу наименьших квадратов** *неизвестные параметры* b_0 и b_1 *выбираются таким образом, чтобы сумма квадратов отклонений эмпирических значений* y_i от *значений* $\hat{y}_i$, *найденных по уравнению регрессии* (**3.3**), *была минимальной*:

$$S = \sum_{i=1}^{n} \left(\hat{y}_i - y_i \right)^2 = \sum_{i=1}^{n} \left(b_0 + b_1 x_i - y_i \right)^2 \to \min . \qquad \textbf{(3.4)}$$

Следует отметить, что для оценки параметров b_0 и b_1 возможны и другие подходы. Так, например, согласно *методу наименьших модулей* следует минимизировать сумму абсолютных величин отклонений $\sum_{i=1}^{n} \left| \hat{y}_i - y_i \right|$. Однако метод наименьших квадратов существенно проще при проведении вычислительной процедуры и дает, как мы увидим далее, хорошие по статистиче-

ским свойствам оценки. Этим и объясняется его широкое применение в статистическом анализе.

На основании необходимого условия экстремума функции двух переменных $S=S(b_0, b_1)$ (**3.4**) приравниваем к нулю ее частные производные, т. е.

$$\begin{cases} \dfrac{\partial S}{\partial b_0} = 2\sum\limits_{i=1}^{n}\left(b_0 + b_1 x_i - y_i\right) = 0; \\ \dfrac{\partial S}{\partial b_1} = 2\sum\limits_{i=1}^{n}\left(b_0 + b_1 x_i - y_i\right)x_i = 0, \end{cases}$$

откуда после преобразований получим *систему нормальных уравнений* для определения параметров линейной регрессии:

$$\begin{cases} b_0 n + b_1 \sum\limits_{i=1}^{n} x_i = \sum\limits_{i=1}^{n} y_i; \\ b_0 \sum\limits_{i=1}^{n} x_i + b_1 \sum\limits_{i=1}^{n} x_i^2 = \sum\limits_{i=1}^{n} x_i y_i. \end{cases} \qquad \textbf{(3.5)}$$

Теперь, разделив обе части уравнений (**3.5**) на n, получим систему нормальных уравнений в виде:

$$\begin{cases} b_0 + b_1 \overline{x} = \overline{y}; \\ b_0 \overline{x} + b_1 \overline{x}^2 = \overline{xy}, \end{cases} \qquad \textbf{(3.6)}$$

где соответствующие средние определяются по формулам:

$$\overline{x} = \frac{\sum\limits_{i=1}^{n} x_i}{n}; \qquad \textbf{(3.7)}$$

$$\overline{y} = \frac{\sum\limits_{i=1}^{n} y_i}{n}; \qquad \textbf{(3.8)}$$

$$\overline{xy} = \frac{\sum\limits_{i=1}^{n} x_i y_i}{n}; \qquad \textbf{(3.9)}$$

$$\overline{x^2} = \frac{\sum\limits_{i=1}^{n} x_i^2}{n}. \qquad \textbf{(3.10)}$$

Подставляя значение

$$b_0 = \overline{y} - b_1 \overline{x} \qquad \textbf{(3.11)}$$

из первого уравнения системы (**3.6**) в уравнение регрессии (**3.3**), получим

$$\hat{y} = \overline{y} - b_1 \overline{x} + b_1 x,$$

или

$$\hat{y} - \overline{y} = b_1 \left(x - \overline{x}\right). \qquad \textbf{(3.12)}$$

Коэффициент b_1 называется *выборочным коэффициентом регрессии* (или просто *коэффициентом регрессии*) Y по X.

Коэффициент регрессии Y по X показывает, на сколько единиц в среднем изменяется переменная Y при увеличении переменной X на одну единицу.

Решая систему (**3.6**), найдем

$$b_1 = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - \overline{x}^2} = \frac{\hat{\mathrm{Cov}}(X,Y)}{s_x^2}, \qquad \textbf{(3.13)}$$

где s_x^2 — *выборочная дисперсия переменной X*:

$$s_x^2 = \overline{x^2} - \overline{x}^2 = \frac{\sum\limits_{i=1}^{n} x_i^2}{n} - (\overline{x})^2, \qquad \textbf{(3.14)}$$

$\hat{\mathrm{Cov}}(X,Y)$ — *выборочный корреляционный момент* или *выборочная ковариация*:

$$\hat{\mathrm{Cov}}(X,Y) = \overline{xy} - \overline{x}\,\overline{y} = \frac{\sum\limits_{i=1}^{n} x_i y_i}{n} - \overline{x}\,\overline{y}. \qquad \textbf{(3.15)}$$

Отметим, что из полученного уравнения регрессии (**3.12**) следует, что линия регрессии проходит через точку $(\overline{x}, \overline{y})$, т. е. $\overline{y} = b_0 + b_1\overline{x}$.

▶ **Пример 3.1.** По данным табл. 3.1 найти уравнение регрессии Y по X.

Р е ш е н и е. Вычислим все необходимые суммы:

$$\sum_{i=1}^{10} x_i = 8+11+12+9+8+8+9+9+8+12=94;.$$

$$\sum_{i=1}^{10} x_i^2 = 8^2+11^2+12^2+9^2+8^2+8^2+9^2+9^2+8^2+12^2=908;$$

$$\sum_{i=1}^{10} y_i = 5+10+10+7+5+6+6+5+6+8=68;$$

$$\sum_{i=1}^{10} x_i y_i = 8\cdot 5+11\cdot 10+12\cdot 10+9\cdot 7+8\cdot 5+8\cdot 6+9\cdot 6+$$

$$+9\cdot 5+8\cdot 6+12\cdot 8 = 664.$$

Затем по формулам (**3.7**)—(**3.15**) находим выборочные характеристики и параметры уравнений регрессии:

$\bar{x} = 94/10 = 9{,}4$ (м); $\bar{y} = 68/10 = 6{,}8$ (т);

$s_x^2 = 908/10 - 9{,}4^2 = 2{,}44$;

$\hat{\mathrm{C}}\mathrm{ov}(X,Y) = 664/10 - 9{,}4 \cdot 6{,}8 = 2{,}48$; $b_1 = 2{,}48/2{,}44 = 1{,}016$.

Итак, уравнение регрессии Y по X:

$$\hat{y} - 6{,}8 = 1{,}016(x - 9{,}4) \quad \text{или} \quad \hat{y} = 2{,}75 + 1{,}016x.$$

Из полученного уравнения регрессии (см. рис. 3.1) следует, что при увеличении мощности пласта X на 1 м добыча угля на одного рабочего Y увеличивается в среднем на 1,016 т (в усл. ед.) (отметим, что свободный член в данном уравнении регрессии не имеет реального смысла). ▶

З а м е ч а н и е. Значения переменных x_i и y_i могут быть измерены в о т к л о н е н и я х от средних значений, т. е. как $x'_i = x_i - \bar{x}$, $y'_i = y_i - \bar{y}$. Начало координат при этом переместится в точку $(\bar{x}, \bar{y})$, а линией регрессии будет та же прямая на плоскости, что и для исходных данных x_i, y_i. Следовательно, $\overline{x'} = 0$, $\overline{y'} = 0$, и *уравнение регрессии* (**3.12**) *в отклонениях* примет вид

$$\hat{y}' = b_1 x',$$

а формула (**3.13**) того же коэффициента регрессии b_1 упростится:

$$b_1 = \frac{\overline{x'y'}}{\overline{x'^2}} = \frac{\sum_{i=1}^{n} x'y'}{\sum_{i=1}^{n} x'^2}, \qquad \textbf{(3.16)}$$

ибо $\overline{x'y'} = \left(\sum_{i=1}^{n} x'y'_i\right)/n, \quad \overline{x'^2} = \left(\sum_{i=1}^{n} x'^2\right)/n.$

3.3. Коэффициент корреляции

Перейдем к о ц е н к е т е с н о т ы корреляционной зависимости. Рассмотрим наиболее важный для практики и теории случай *линейной зависимости* вида (**3.12**).

На первый взгляд, подходящим измерителем тесноты связи Y от X является коэффициент регрессии b_1, ибо, как уже было отмечено, он показывает, на сколько единиц в среднем изменя-

ется Y, когда X увеличивается на одну единицу. Однако b_1 зависит от единиц измерения переменных. Например, в полученной ранее зависимости он увеличится в 100 раз, если мощность пласта X выразить не в метрах, а в сантиметрах.

Очевидно, что для «исправления» b_1 как показателя тесноты связи нужная такая стандартная система единиц измерения, в которой данные по различным характеристикам оказались бы сравнимы между собой. Статистика знает такую систему единиц. Эта система использует в качестве единицы измерения переменной ее *среднее квадратическое отклонение s.*

Представим уравнение **(3.12)** в эквивалентном виде:

$$\frac{\hat{y}-\overline{y}}{s_y} = b_1 \frac{s_x}{s_y} \cdot \frac{x-\overline{x}}{s_x}.$$

В этой системе величина

$$r = b_1 \frac{s_x}{s_y} \qquad \textbf{(3.17)}$$

показывает, *на сколько величин s_y изменится в среднем Y, когда X увеличится на одно s_x.*

Величина r является показателем тесноты линейной связи и называется ***выборочным коэффициентом корреляции*** (или просто ***коэффициентом корреляции***).

Две корреляционные зависимости переменной Y от X приведены на рис. 3.2. Очевидно, что в случае *а* зависимость между переменными менее тесная и коэффициент корреляции должен быть меньше, чем в случае *б*, так как точки корреляционного поля *а* дальше отстоят от линии регрессии, чем точки поля *б*.

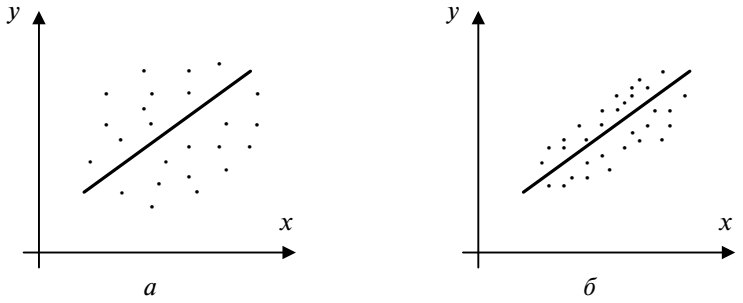


Рис. 3.2

Если $r > 0$ ($b_1 > 0$), то корреляционная связь между переменными называется *прямой*, если $r < 0$ ($b_1 < 0$), — *обратной*. При прямой (обратной) связи увеличение одной из переменных ведет к увеличению (уменьшению) условной (групповой) средней другой.

Учитывая (**3.13**), формулу для r представим в виде:

$$r = \frac{\overline{xy} - \bar{x}\,\bar{y}}{s_x s_y} = \frac{\hat{\mathrm{Cov}}(X,Y)}{s_x s_y}. \qquad \textbf{(3.18)}$$

Отметим другие модификации формулы r, полученные из формулы (**3.18**) с помощью формул (**3.7**)—(**3.10**), (**3.13**)—(**3.15**):

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n s_x s_y}; \qquad \textbf{(3.19)}$$

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\cdot\sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}}. \qquad \textbf{(3.20)}$$

Для практических расчетов наиболее удобна формула (**3.20**), так как по ней r находится непосредственно из данных наблюдений и на значении r не скажутся округления данных, связанные с расчетом средних и отклонений от них.

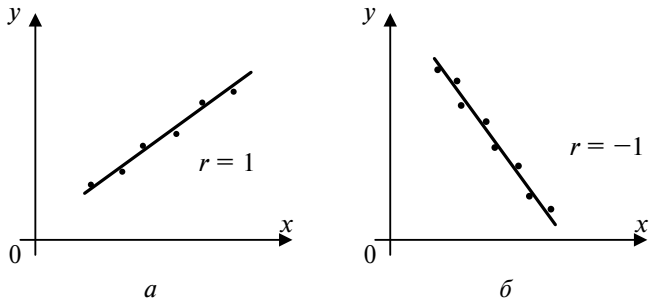
Выборочный коэффициент корреляции r (при достаточно большом объеме выборки n) так же, как и коэффициент корреляции двух случайных величин (§ **2.5**), обладает следующими с в о й с т в а м и.

1. *Коэффициент корреляции принимает значения на отрезке* [—1;1], т. е. $-1 \le r \le 1$.

Чем ближе $|r|$ к единице, тем теснее связь.

2. *При* $r = \pm 1$ *корреляционная связь представляет линейную функциональную зависимость. При этом все наблюдаемые значения располагаются на прямой линии* (рис. 3.3).

3. *При* $r = 0$ *линейная корреляционная связь отсутствует. При этом линия регрессии параллельна оси Ox* (рис. 3.4).



а *б*

Рис. 3.3

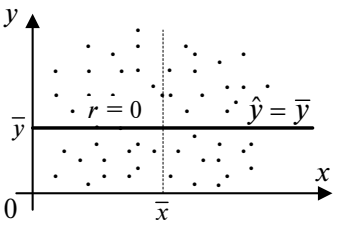


Рис. 3.4

Следует отметить, что мы ввели выборочный коэффициент корреляции r исходя из оценки близости точек корреляционного поля к прямой регрессии Y по X. Однако r является непосредственно оценкой генерального коэффициента корреляции ρ между X и Y лишь в случае двумерного нормального закона распределения случайных величин X и Y. В других случаях (когда распределения X и Y отклоняются от нормального, одна из исследуемых величин, например X, не является случайной и т.п.) выборочный коэффициент корреляции не следует рассматривать как строгую меру взаимосвязи переменных.

▶ **Пример 3.2.**

По данным табл. 3.1 вычислить коэффициент корреляции между переменными X и Y.

Р е ш е н и е. В примере **3.1** были вычислены $\sum_{i=1}^{10} x_i = 94$;

$\sum_{i=1}^{10} x_i^2 = 908$; $\sum_{i=1}^{10} y_i = 68$, $\sum_{i=1}^{10} x_i y_i = 664$. Вычислим сумму

$$\sum_{i=1}^{10} y_i^2 = 5^2 + 10^2 + 10^2 + 7^2 + 5^2 + 6^2 + 6^2 + 5^2 + 6^2 + 8^2 = 496.$$

По формуле (**3.20**)

$$r = \frac{10 \cdot 664 - 94 \cdot 68}{\sqrt{10 \cdot 908 - 94^2}\sqrt{10 \cdot 496 - 68^2}} = 0{,}866,$$

т. е. связь между переменными достаточно тесная. ▶

3.4. Основные положения регрессионного анализа. Оценка параметров парной регрессионной модели. Теорема Гаусса—Маркова

Как отмечено в §**3.2**, рассматриваемая в регрессионном анализе зависимость Y от X может быть представлена в виде модельного уравнения регрессии (**3.1**).

В силу воздействия неучтенных случайных факторов и причин отдельные наблюдения переменной Y будут в бóльшей или меньшей мере отклоняться от функции регрессии $\varphi(X)$. В этом случае *уравнение взаимосвязи двух переменных* (*парная регрессионная модель*) может быть представлено в виде:

$$Y = \varphi(X) + \varepsilon,$$

где ε — *случайная переменная* (*случайный член*), характеризующая отклонение от функции регрессии. Эту переменную будем называть *возмущающей* или просто *возмущением* (либо *ошибкой*)[1]. *Таким образом, в регрессионной модели зависимая переменная Y есть некоторая функция* $\varphi(X)$ *с точностью до случайного возмущения* ε.

Рассмотрим *линейный регрессионный анализ,* для которого функции $\varphi(X)$ л и н е й н а относительно оцениваемых параметров:

$$M_x(Y) = \beta_0 + \beta_1 x. \qquad \textbf{(3.21)}$$

Предположим, что для оценки параметров линейной функции регрессии (**3.21**) взята выборка, содержащая n пар значений переменных (x_i, y_i), где i=1,2,..., n. В этом случае *линейная парная регрессионная модель* имеет вид:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i. \qquad \textbf{(3.22)}$$

[1] В литературе переменную ε называют также *остаточной* или *остатком.*

Отметим о с н о в н ы е п р е д п о с ы л к и регрессионного анализа.

1. *В модели* (**3.22**) *возмущение*[1] ε_i (или зависимая переменная y_i) *есть величина случайная*, а *объясняющая переменная* x_i *— величина неслучайная*[2].

2. *Математическое ожидание возмущения* ε_i *равно нулю*:

$$M(\varepsilon_i) = 0 \qquad \textbf{(3.23)}$$

(или математическое ожидание зависимой переменной y_i равно линейной функции регрессии: $M(y_i) = \beta_0 + \beta_1 x_i$).

3. *Дисперсия возмущения* ε_i (или зависимой переменной y_i) *постоянна для любого i*:

$$D(\varepsilon_i) = \sigma^2 \qquad \textbf{(3.24)}$$

(или $D(y_i) = \sigma^2$) — условие *гомоскедастичности* или *равноизменчивости* возмущения (зависимой переменной)).

4. *Возмущения* ε_i и ε_j (или переменные y_i и y_j) *не коррелированы*[3]:

$$M(\varepsilon_i \varepsilon_j) = 0 \ (i \neq j). \qquad \textbf{(3.25)}$$

5. *Возмущение* ε_i (или зависимая переменная y_i) *есть нормально распределенная случайная величина.*

В этом случае модель (**3.22**) называется *классической нормальной линейной регрессионной моделью* (*Classical Normal Linear Regression model*).

Для получения уравнения регрессии достаточно предпосылок *1—4*. Требование выполнения предпосылки *5* (т. е. рассмотрение «*нормальной регрессии*») необходимо для оценки т о ч н о с т и уравнения регрессии и его параметров.

Оценкой модели (**3.22**) по выборке является уравнение регрессии $\hat{y} = b_0 + b_1 x$ (**3.3**). Параметры этого уравнения b_0 и b_1 определяются на основе метода наименьших квадратов. Об их нахождении подробно см. § **3.2.**

---

[1] Во всех предпосылках i=1,2,..., n.

[2] При этом предполагается, что среди значений x_i (i=1,2,...,n) не все одинаковые, так что имеет смысл формула (**3.13**) для коэффициента регрессии.

[3] Требование некоррелированности $\mathrm{Cov}(\varepsilon_i, \varepsilon_j)$=0 с учетом (**2.30**) и (**3.23**) приводит к условию (**3.25**): $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = M[(\varepsilon_i - 0)(\varepsilon_j - 0)] = M(\varepsilon_i \varepsilon_j) = 0$. При выполнении предпосылки *5* это требование равносильно независимости переменных ε_i и ε_j (y_i и y_j).

Воздействие неучтенных случайных факторов и ошибок наблюдений в модели (**3.22**) определяется с помощью *дисперсии возмущений* (*ошибок*) или *остаточной дисперсии* σ^2. Несмещенной оценкой этой дисперсии является *выборочная остаточная дисперсия*[1].

$$s^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n-2} = \frac{\sum_{i=1}^{n} e_i^2}{n-2}, \qquad \textbf{(3.26)}$$

где $\hat{y}_i$ — *групповая средняя*, найденная по уравнению регрессии; $e_i = \hat{y}_i - y_i$ — *выборочная оценка возмущения*[2] ε_i или *остаток регрессии*.

Напомним, что в математической статистике для получения несмещенной оценки дисперсии случайной величины соответствующую сумму квадратов отклонений от средней делят не на число наблюдений n, а на *число степеней свободы* (degress of freedom) $n - m$, равное разности между числом независимых наблюдений случайной величины n и числом связей, ограничивающих свободу их изменения, т. е. число m уравнений, связывающих эти наблюдения. Поэтому в знаменателе выражения (**3.26**) стоит число степеней свободы $n - 2$, так как две степени свободы теряются при определении двух параметров прямой из системы нормальных уравнений (**3.5**).

Возникает вопрос, являются ли оценки b_0, b_1, s^2 параметров β_0, β_1 σ^2 «наилучшими»? Ответ на этот вопрос дает следующая теорема.

Теорема Гаусса—Маркова. *Если регрессионная модель* (**3.22**) *удовлетворяет предпосылкам 1—4* (с. 61), *то оценки* b_0 (**3.11**), b_1 (**3.13**) *имеют наименьшую дисперсию в классе всех линейных несмещенных оценок* (*Best Linear Unbiased Estimator,* или *BLUE*).[3]

Таким образом, оценки b_0 и b_1 в определенном смысле являются наиболее *эффективными* линейными оценками параметров β_0 и β_1.

До сих пор мы использовали оценки параметров, полученные методом наименьших квадратов. Рассмотрим еще один

---

[1] Формула (**3.26**) при $p = 1$ является частным случаем формулы (**4.21**), доказанной ниже в § **4.4**.

[2] e_i называют также *невязкой*.

[3] Доказательство теоремы Гаусса—Маркова в общем виде приведено в § **4.4.**

важный метод получения оценок, широко используемый в эконометрике, — метод максимального правдоподобия.

Метод максимального правдоподобия. Для его применения должен быть известен в и д закона распределения вероятностей имеющихся выборочных данных.

Полагая выполнение предпосылки *5* (с. 61) регрессионного анализа, т. е. н о р м а л ь н у ю классическую регрессионную модель (**3.22**), будем рассматривать значения y_i как независимые нормально распределенные случайные величины с математическим ожиданием $M(y_i) = \beta_0 + \beta_1 x_i$, являющимся функцией от x_i, и постоянной дисперсией σ^2.

Следовательно, плотность нормально распределенной случайной величины y_i

$$\varphi_N(y_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}}.$$

Функция правдоподобия, выражающая *плотность вероятности совместного появления результатов выборки*, имеет вид

$$L(y_i x_i; ...; y_n, x_n; \beta_0, \beta_1; \sigma^2) = \prod_{i=1}^{n} \frac{1}{\left(\sigma\sqrt{2\pi}\right)^n} e^{-\frac{\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}}.$$

Согласно методу максимального правдоподобия в качестве оценок параметров β_0, β_1 и σ^2 *принимаются такие значения* $\hat{\beta}_0$, $\hat{\beta}_1$ и $\hat{\sigma}^2$, *которые максимизируют функцию правдоподобия L.*

Очевидно, что при заданных значениях x_1, x_2,..., x_n объясняющей переменной X и постоянной дисперсии σ^2 функция правдоподобия L достигает максимума, когда показатель степени при e будет минимальным по абсолютной величине, т. е. при условии минимума функции

$$\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2,$$

что совпадает с условием (**3.4**) нахождения оценок b_0 и b_1 методом наименьших квадратов. Следовательно, *оценки* b_0 (**3.11**) и b_1

(**3.13**) *параметров* β_0, β_1 *совпадают с оценками метода максимального правдоподобия* $\hat{\beta}_0$ и $\hat{\beta}_1$.

Для нахождения оценки $\hat{\sigma}^2$ максимального правдоподобия параметра σ^2, максимизирующей функцию L, качественных соображений уже недостаточно, и необходимо прибегнуть к методам дифференциального исчисления. Приравняв частную производную $\frac{\partial L}{\partial \sigma^2} = 0$ (соответствующие выкладки предлагаем провести читателю самостоятельно), получим

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - b_0 - b_1 x_i\right)^2 = \frac{\sum_{i=1}^{n} e_i^2}{n}, \qquad \textbf{(3.27)}$$

где параметры β_0 и β_1 заменены их оценками b_0 и b_1. Сравнивая с полученной ранее несмещенной оценкой s^2 (**3.26**), видим, что оценка $\hat{\sigma}^2$ (**3.27**) метода максимального правдоподобия параметра σ^2 является *смещенной*.

В соответствии со свойствами оценок максимального правдоподобия *оценки* (b_0, b_1) и $\hat{\sigma}^2$ (а значит, и s^2) являются *состоятельными оценками*. Можно показать, что *при выполнении предпосылки 5 о нормальном законе распределения возмущения* ε_i (i=1,..., n) *эти оценки являются независимыми*.

3.5. Интервальная оценка функции регрессии и ее параметров

Доверительный интервал для функции регрессии. Построим *доверительный интервал для функции регрессии,* т.е. для условного математического ожидания $M_x(Y)$, который с заданной надежностью (доверительной вероятностью) $\gamma = 1-\alpha$ накрывает неизвестное значение $M_x(Y)$.

Найдем дисперсию групповой средней $\hat{y}$, представляющей выборочную оценку $M_x(Y)$. С этой целью уравнение регрессии **(3.12)** представим в виде:

$$\hat{y} = \overline{y} + b_1(x - \overline{x}). \qquad \textbf{(3.28)}$$

На рис. 3.5 линия регрессии **(3.28)** изображена графически. Для произвольного наблюдаемого значения y_i выделены его

составляющие: средняя $\bar{y}$, приращение $b_1(x_i - \bar{x})$, образующие расчетное значение $\hat{y}_i$, и остаток e_i.

Дисперсия групповой средней равна сумме дисперсий двух н е з а в и с и м ы х [1] слагаемых выражения **(3.28)**:

$$\sigma^2_{\hat{y}} = \sigma^2_{\bar{y}} + \sigma^2_{b_1}(x - \bar{x})^2. \qquad \textbf{(3.29)}$$

(Здесь учтено, что $(x - \bar{x})$ — неслучайная величина, при вынесении которой за знак дисперсии ее необходимо возвести в квадрат.)

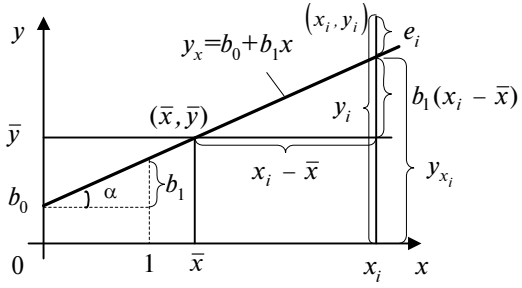


Рис. 3.5

Дисперсия выборочной средней $\bar{y}$

$$\sigma^2_{\bar{y}} = \sigma^2\left(\frac{\sum_{i=1}^{n} y_i}{n}\right) = \frac{\sum_{i=1}^{n}\sigma^2_{y_i}}{n^2} = \frac{\sum_{i=1}^{n}\sigma^2}{n^2} = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}. \qquad \textbf{(3.30)}$$

Для нахождения дисперсии $\sigma^2_{b_1}$ представим коэффициент регрессии (**3.16**) в виде

$$b_1 = \frac{\sum_{i=1}^{n} x'_i y'_i}{\sum_{i=1}^{n} x'^2_i} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}. \qquad \textbf{(3.31)}$$

[1] Доказательство этого факта здесь не приводится.

Тогда

$$\sigma_{b_1}^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sigma^2}{\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^2} = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}. \qquad \textbf{(3.32)}$$

Найдем оценку дисперсии групповых средних **(3.29)**, учитывая **(3.30)** и **(3.32)** и заменяя σ^2 ее оценкой s^2:

$$s_{\hat{y}}^2 = s^2 \left(\frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \right). \qquad \textbf{(3.33)}$$

Основываясь на предпосылках *1—5* регрессионного анализа (§ **3.4**), можно показать, что *статистика* $t = \dfrac{\hat{y} - M_x(Y)}{s_{\hat{y}}}$ имеет t-распределение Стьюдента с $k = n - 2$ *степенями свободы* и построить доверительный интервал для условного математического ожидания $M_x(Y)$:

$$\hat{y} - t_{1-\alpha;k} \cdot s_{\hat{y}} \le M_x(Y) \le \hat{y} + t_{1-\alpha;k} \cdot s_{\hat{y}}, \qquad \textbf{(3.34)}$$

где $s_{\hat{y}} = \sqrt{s_{\hat{y}}^2}$ — стандартная ошибка групповой средней $\hat{y}$.

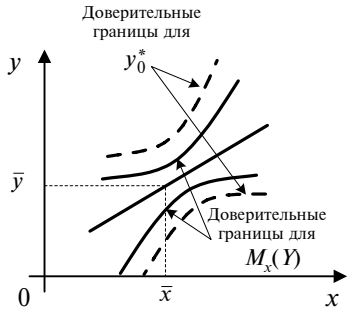


Рис. 3.6

Из формул **(3.33)** и **(3.34)** видно, что величина (длина) доверительного интервала зависит от значения объясняющей переменной x: при $x = \bar{x}$ она минимальна, а по мере удаления x от $\bar{x}$ величина доверительного интервала увеличивается (рис. 3.6). Таким образом, *прогноз* значений (определение неизвестных значений) зависимой переменной Y по уравнению регрессии оправдан, если значение x объясняющей переменной X не выходит за диапазон ее значений по выборке (причем тем более точный, чем ближе x

к $\bar{x}$). Другими словами, *экстраполяция кривой регрессии, т.е. ее использование вне пределов обследованного диапазона значений объясняющей переменной* (даже если она оправдана для рассматриваемой переменной исходя из смысла решаемой задачи) *может привести к значительным погрешностям.*

Доверительный интервал для индивидуальных значений зависимой переменной. Построенная доверительная область для $M_x(Y)$ (см. рис. 3.6) определяет местоположение модельной линии регрессии (т.е. условного математического ожидания), но не отдельных возможных значений зависимой переменной, которые отклоняются от средней. Поэтому при определении *доверительного интервала для индивидуальных значений* y_0^* *зависимой переменной* необходимо учитывать еще один источник вариации — *рассеяние вокруг линии регрессии*, т.е. в оценку суммарной дисперсии $s_{\hat{y}}^2$ следует включить величину s^2. В результате оценка дисперсии индивидуальных значений y_0 при $x = x_0$ равна

$$s_{\hat{y}_0}^2 = s^2 \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right), \qquad \textbf{(3.35)}$$

а соответствующий доверительный интервал для прогнозов индивидуальных значений y_0^* будет определяться по формуле:

$$\hat{y}_0 - t_{1-\alpha;\, n-2} s_{\hat{y}_0} \le y_0^* \le \hat{y}_0 + t_{1-\alpha;\, n-2} s_{\hat{y}_0}. \qquad \textbf{(3.36)}$$

Доверительный интервал для параметров регрессионной модели. Наряду с интервальным оцениванием функции регрессии иногда представляет интерес построение доверительных интервалов для параметров регрессионной модели, в частности для параметров регрессионной модели, в частности для β_1 и σ^2.

Можно показать, что при выполнении предпосылки *5* (с. 61) регрессионного анализа статистика $t = \dfrac{b_1 - \beta_1}{\sigma_{b_1}}$ имеет стандартный нормальный закон распределения, а если в выражении (**3.32**) для $\sigma_{b_1}^2$ заменить σ^2 ее оценкой s^2, то статистика

$$t = \frac{b_1 - \beta_1}{s} \sqrt{\sum_{i=1}^{n} \left(x_i - \bar{x} \right)^2} \qquad \textbf{(3.37)}$$

имеет t-распределение Стьюдента с $k = n - 2$ степенями свободы. Поэтому (см. § **2.7**) *интервальная оценка параметра* β_1 на уровне значимости α имеет вид:

$$b_1 - t_{1-\alpha;n-2}\frac{s}{\sqrt{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}} \le \beta_1 \le b_1 + t_{1-\alpha;n-2}\frac{s}{\sqrt{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}}. \quad \textbf{(3.38)}$$

При построении доверительного интервала для параметра σ^2 исходят из того, что статистика $\frac{ns^2}{\sigma^2}$ имеет χ^2-распределение с $k = n - 2$ *степенями свободы*. Поэтому *интервальная оценка для* σ^2 на уровне значимости α имеет вид (см. (**2.43**)):

$$\frac{ns^2}{\chi^2_{\alpha/2;n-2}} \le \sigma^2 \le \frac{ns^2}{\chi^2_{1-\alpha/2;n-2}} \quad \textbf{(3.39)}$$

(доверительный интервал выбирается таким образом, чтобы

$$P\left(\chi^2 < \chi^2_{1-\alpha/2;n-2}\right) = P\left(\chi^2 > \chi^2_{\alpha/2;n-2}\right) = \frac{\alpha}{2}).$$

▶ **Пример 3.3.**

По данным табл. 3.1: 1) оценить сменную среднюю добычу угля на одного рабочего для шахт с мощностью пласта 8 м;

2) найти 95%-ные доверительные интервалы для индивидуального и среднего значений сменной добычи угля на 1 рабочего для таких же шахт;

3) найти с надежностью 0,95 интервальные оценки коэффициента регрессии β_1 и дисперсии σ^2.

Р е ш е н и е. Уравнение регрессии Y по X было получено в примере **3.1**: $\hat{y} = -2,75 + 1,016x$, т.е. при увеличении мощности пласта X на 1 м добыча угля на одного рабочего Y увеличивается *в среднем* на 1,016 т (в усл. ед.).

1. Оценим условное математическое ожидание $M_{x=8}(Y)$. Выборочной оценкой $M_{x=8}(Y)$ является групповая средняя $\hat{y}_{x=8}$, которую найдем по уравнению регрессии:

$$\hat{y}_{x=8} = -2,75 + 1,016 \cdot 8 = 5,38 \text{ (т)}.$$

Для построения доверительного интервала для $M_{x=8}(Y)$ необходимо знать дисперсию его оценки, т.е. $s^2_{\hat{y}_{x=8}}$. Составим

вспомогательную таблицу (табл. 3.2) с учетом того, что $\bar{x}$=9,4 (м), а значения определяются по полученному уравнению регрессии.

Таблица 3.2

x_i	8	11	12	9	8	8	9	9	8	12	$\sum$
$(x_i - \bar{x})^2$	1,96	2,56	6,76	0,16	1,96	1,96	0,16	0,16	1,96	6,76	24,40
$\hat{y}_i = -2,75 + +1,016x_i$	5,38	8,43	9,44	6,39	5,38	5,38	6,39	6,39	5,38	9,44	—
$e_i^2 = (\hat{y}_i - y_i)^2$	0,14	2,48	0,31	0,37	0,14	0,39	0,15	1,94	0,39	2,08	8,39

Теперь имеем по **(3.26)**: $s^2=\frac{8,39}{10-2}$=1,049, по **(3.32)**:

$$s^2_{\hat{y}_{x=8}} = 1,049\left[\frac{1}{10} + \frac{(8-9,4)^2}{24,4}\right] = 0,189$$

и $s_{\hat{y}_{x=8}} = \sqrt{0,189} = 0,435$ (т).

По табл. II приложений $t_{0,95;8}$=2,31. Теперь по **(3.34)** искомый доверительный интервал

$$5,38-2,31 \cdot 0,435 \le M_{x=8}(Y) \le 5,38+2,31 \cdot 0,435,$$

или $$4,38 \le M_{x=8}(Y) \le 6,38 \text{ (т).}$$

Итак, средняя сменная добыча угля на одного рабочего для шахт с мощностью пласта 8 м с надежностью 0,95 находится в пределах от 4,38 до 6,38 т.

2. Чтобы построить доверительный интервал для индивидуального значения $y^*_{x_0=8}$, найдем дисперсию его оценки по **(3.35)**:

$$s^2_{y_{x_0=8}} = 1,049\left(1 + \frac{1}{10} + \frac{(8-9,4)^2}{24,4}\right) = 1,238$$

и $s_{y_{x_0=8}} = \sqrt{1,238} = 1,113$ (т).

Далее искомый доверительный интервал получим по **(3.36)**:

$$5,38-2,31 \cdot 1,113 \le y^*_{x_0=8} \le 5,38+2,31 \cdot 1,113,$$

или $$2,81 \le y^*_{x_0=8} \le 7,95.$$

Таким образом, индивидуальная сменная добыча угля на одного рабочего для шахт с мощностью пласта 8 м с надежностью 0,95 находится в пределах от 2,81 до 7,95 т.

3. Найдем 95%-ный доверительный интервал для параметра β_1.

По формуле (**3.38**)

$$1{,}016 - 2{,}31\frac{\sqrt{1{,}049}}{\sqrt{24{,}4}} \le \beta_1 \le 1{,}016 + 2{,}31\frac{\sqrt{1{,}049}}{\sqrt{24{,}4}},$$

или $0{,}537 \le \beta_1 \le 1{,}495$, т. е. с надежностью 0,95 при изменении мощности пласта X на 1 м суточная выработка Y будет изменяться на величину, заключенную в интервале от 0,537 до 1,495 (т).

Найдем 95%-ный доверительный интервал для параметра σ^2.

Учитывая, что $\alpha=1-0{,}95=0{,}05$, найдем по таблице III приложений

$$\chi^2_{\alpha/2;n-2} = \chi^2_{0{,}025;8} = 17{,}53;$$
$$\chi^2_{1-\alpha/2;n-2} = \chi^2_{0{,}975;8} = 2{,}18.$$

По формуле (**3.39**)

$$\frac{10 \cdot 1{,}049}{17{,}53} \le \sigma^2 \le \frac{10 \cdot 1{,}049}{2{,}18},$$

или $0{,}598 \le \sigma^2 \le 4{,}81$, и $0{,}773 \le \sigma \le 2{,}19$.

Таким образом, с надежностью 0,95 дисперсия возмущений заключена в пределах от 0,598 до 4,81, а их стандартное отклонение — от 0,773 до 2,19 (т). ▶

3.6. Оценка значимости уравнения регрессии. Коэффициент детерминации

Проверить **значимость уравнения регрессии** — значит *установить, соответствует ли математическая модель, выражающая зависимость между переменными, экспериментальным данным и достаточно ли включенных в уравнение объясняющих переменных* (*одной или нескольких*) *для описания зависимой переменной.*

Проверка значимости уравнения регрессии производится на основе дисперсионного анализа.

В математической статистике дисперсионный анализ рассмотрен как самостоятельный инструмент (метод) статистического анализа.

Здесь же он применяется как вспомогательное средство для изучения качества регрессионной модели.

Согласно основной идее дисперсионного анализа (см., например, [17])

$$\sum_{i=1}^{n}(y_i-\overline{y})^2=\sum_{i=1}^{n}\left[(\hat{y}_i-\overline{y})+(y_i-\hat{y}_i)\right]^2= \\ =\sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2+\sum_{i=1}^{n}(y_i-\hat{y}_i)^2+2\sum_{i=1}^{n}(\hat{y}_i-\overline{y})(y_i-\hat{y}_i), \tag{3.40}$$

или

$$Q = Q_R+Q_e, \tag{3.41}$$

где Q — общая сумма квадратов отклонений зависимой переменной от средней, а Q_R и Q_e — соответственно сумма квадратов, обусловленная регрессией, и остаточная сумма квадратов, характеризующая влияние неучтенных факторов[1].

Убедимся в том, что пропущенное в (**3.41**) третье слагаемое $Q_3=2\sum_{i=1}^{n}(\hat{y}_i-\overline{y})(y_i-\hat{y}_i)$ равно 0. Учитывая (**3.28**), (**3.11**), имеем:

$$\hat{y}_i-\overline{y}=b_1(x_i-\overline{x});$$

$$y_i-\hat{y}_i=y_i-b_0-b_1x_i=y_i-(\overline{y}-b_1\overline{x})-b_1x_i=(y_i-\overline{y})-b_1(x_i-\overline{x}).$$

Теперь[2]

$$Q_3=2\sum_{i=1}^{n}(\hat{y}_i-\overline{y})(y_i-\hat{y}_i)=2b_1\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})-2b_1^2\sum_{i=1}^{n}(x_i-\overline{x})^2=0$$

(с учетом соотношения (**3.31**)).

Схема дисперсионного анализа имеет вид, представленный в табл. 3.3.

---

[1] В переводной литературе Q, Q_R, Q_e обозначаются соответственно *TSS* (*total sum of squares*), *RSS* (*regression sum of squares*) и *ESS* (*error sum of squares*).

[2] Из полученного соотношения видно, что $\sum_{i=1}^{n}(y_i-\hat{y}_i)=0$. Вообще говоря, это равенство, а с ним в конечном счете и разложение (**3.41**), выполняется только при *наличии свободного члена* в регрессионной модели.

Таблица 3.3

Компоненты дисперсии	*Сумма квадратов*	*Число степеней свободы*	*Средние квадраты*
Регрессия	$Q_R = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$	$m-1$	$s_R^2 = \dfrac{Q_R}{m-1}$
Остаточная	$Q_e = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$	$n-m$	$s^2 = \dfrac{Q_e}{n-m}$
Общая	$Q = \sum_{i=1}^{n} (y_i - \bar{y})^2$	$n-1$	

Средние квадраты s_R^2 и s^2 (табл. 3.3) представляют собой несмещенные оценки дисперсий зависимой переменной, обусловленных соответственно регрессий или объясняющей переменной X и воздействием неучтенных случайных факторов и ошибок; m — число оцениваемых параметров уравнения регрессии; n — число наблюдений.

Замечание. При расчете общей суммы квадратов Q полезно иметь в виду, что

$$Q = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}. \qquad \textbf{(3.42)}$$

(Формула (**3.42**) следует из разложения

$Q = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - 2\bar{y}\sum_{i=1}^{n} y_i + n\bar{y}^2$ с учетом (**3.8**).)

При отсутствии линейной зависимости между зависимой и объясняющими(ей) переменными случайные величины $s_R^2 = Q_R/(m-1)$ и $s^2 = Q_e/(n-m)$ имеют χ^2-распределение соответственно с $m-1$ и $n-m$ степенями свободы, а их отношение — F-распределение с теми же степенями свободы (см. § **2.3**). Поэтому уравнение регрессии значимо на уровне α, если фактически наблюдаемое значение статистики

$$F = \frac{Q_R(n-m)}{Q_e(m-1)} = \frac{s_R^2}{s^2} > F_{\alpha;k_1;k_2}, \qquad \textbf{(3.43)}$$

где $F_{\alpha;k_1;k_2}$ — табличное значение F-критерия Фишера—Снедекора, определенное на уровне значимости α при $k_1 = m-1$ и $k_2 = n-m$ степенях свободы.

Учитывая смысл величин s_R^2 и s^2, можно сказать, что *значение F показывает, в какой мере регрессия лучше оценивает значение зависимой переменной по сравнению с ее средней.*

В случае линейной парной регрессии $m = 2$, и уравнение регрессии значимо на уровне α, если

$$F = \frac{Q_R(n-2)}{Q_e} > F_{\alpha;1;n-2}. \qquad \textbf{(3.44)}$$

Следует отметить, что значимость уравнения парной линейной регрессии может быть проведена и другим способом, если оценить *значимость коэффициента регрессии* b_1, который, как отмечено в § **3.4**, имеет *t*-распределение Стьюдента с $k=n-2$ степенями свободы.

Уравнение парной линейной регрессии или коэффициент регрессии b_1 значимы на уровне α (иначе — гипотеза H_0 о равенстве параметра β_1 нулю, т. е. H_0: $\beta_1=0$, отвергается), если фактически наблюдаемое значение статистики (**3.37**)

$$t = \frac{b_1 - 0}{s}\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad \textbf{(3.45)}$$

больше критического (по абсолютной величине), т. е. $|t| > t_{1-\alpha;\,n-2}$.

Можно показать, что для *парной* линейной модели оба способа проверки значимости с использованием *F*- и *t*-критериев равносильны, ибо эти критерии связаны соотношением $F = t^2$.

В ряде прикладных задач требуется оценить *значимость коэффициента корреляции r* (§ **3.3**). При этом исходят из того, что при отсутствии корреляционной связи статистика $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ имеет *t*-распределение Стьюдента с $n - 2$ степенями свободы.

Коэффициент корреляции *r* значим на уровне α (иначе — гипотеза H_0 о равенстве генерального коэффициента корреляции ρ нулю, т. е. H_0: $\rho=0$, отвергается), если

$$|t| = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}} > t_{1-\alpha;n-2}, \qquad \textbf{(3.46)}$$

где $t_{1-\alpha;n-2}$ — табличное значение *t*-критерия Стьюдента, определенное на уровне значимости α при числе степеней свободы $n-2$.

Легко показать, что получаемые значения t-критерия для проверки гипотез $\beta=0$ по (**3.45**) и $\rho=0$ по (**3.46**) одинаковы.

▶ **Пример 3.4**.

По данным табл. 3.1 оценить на уровне $\alpha=0{,}05$ значимость уравнения регрессии Y по X.

Р е ш е н и е. *1-й способ.* Выше, в примерах **3.1, 3.2** были найдены: $\sum_{i=1}^{10} y_i = 68$, $\sum_{i=1}^{10} y_i^2 = 496$.

Вычислим необходимые суммы квадратов по формулам (**3.40**), (**3.42**):

$$Q = \sum_{i=1}^{10}\left(y_i - \overline{y}\right)^2 = \sum_{i=1}^{10} y_i^2 - \frac{\left(\sum_{i=1}^{10} y_i\right)^2}{10} = 496 - \frac{68^2}{10} = 33{,}6;$$

$$Q_e = \sum_{i=1}^{10}\left(\hat{y}_i - \overline{y}\right)^2 = \sum_{i=1}^{10} e_i^2 = 8{,}39 \text{ (см. табл. 3.2)};$$

$$Q_R = Q - Q_e = 33{,}6 - 8{,}39 = 25{,}21.$$

По формуле (**3.43**)

$$F = \frac{25{,}21(10-2)}{8{,}39} = 24{,}04.$$

По таблице F-распределения (табл. IV приложений) $F_{0{,}05;1;8}=4{,}20$. Так как $F>F_{0{,}05;1;8}$, то уравнение регрессии значимо.

2-й способ. Учитывая, что $b_1=1{,}016$, $\sum_{i=1}^{10}\left(x_i - \overline{x}\right)^2 = 24{,}40$, $s^2=1{,}049$ (см. пример **3.3,** табл. 3.2), по формуле (**3.45**)

$$t = \frac{1{,}016}{\sqrt{1{,}049}}\sqrt{24{,}40} = 4{,}90.^{1}$$

По таблицам t-распределения (табл. II приложений) $t_{0{,}95;8}=2{,}31$. Так как $t > t_{0{,}95;8}$, то коэффициент регрессии b_1, а значит, и уравнение парной линейной регрессии Y по X значимы. ▶

Одной из *наиболее эффективных оценок адекватности регрессионной модели, мерой качества уравнения регрессии*, (или, как

---

[1] Тот же результат может быть получен по формуле (**3.46**), учитывая, что $r = 0{,}866$ (см. пример **3.2**): $t = \frac{0{,}866\sqrt{10-2}}{\sqrt{1-0{,}866^2}} = 4{,}90.$

говорят, мерой качества подгонки регрессионной модели к наблюденным значениям y_i), *характеристикой прогностической силы анализируемой регрессионной модели является* **коэффициент детерминации**, определяемый по формуле

$$R^2 = \frac{Q_R}{Q} = 1 - \frac{Q_e}{Q}. \qquad \textbf{(3.47)}$$

Величина R^2 показывает, какая часть (доля) вариации зависимой переменной обусловлена вариацией объясняющей переменной.

Так как $0 \le Q_R \le Q$, то $0 \le R^2 \le 1$.

Чем ближе R^2 к единице, тем лучше регрессия аппроксимирует эмпирические данные, тем теснее наблюдения примыкают к линии регрессии. *Если $R^2=1$, то эмпирические точки (x_i, y_i) лежат на линии регрессии* (см. рис. 3.3) *и между переменными Y и X существует линейная функциональная зависимость. Если $R^2 = 0$, то вариация зависимой переменной полностью обусловлена воздействием неучтенных в модели переменных, и линия регрессии параллельна оси абсцисс* (см. рис. 3.4).

Заметим, что коэффициент R^2 имеет смысл рассматривать только при *наличии свободного члена* в уравнении регрессии, так как лишь в этом случае, как уже отмечалось, верно равенство (**3.41**), а следовательно, и (**3.47**).

Если известен коэффициент детерминации R^2, то критерий значимости (**3.43**) уравнения регрессии или самого коэффициента детерминации может быть записан в виде

$$F = \frac{R^2 (n-m)}{(1-R^2)(m-1)} > F_{\alpha;k_1;k_2}. \qquad \textbf{(3.48)}$$

В случае парной линейной регрессионной модели коэффициент детерминации равен квадрату коэффициента корреляции, т. е. $R^2 = r^2$.

Действительно, учитывая (**3.12**), (**3.17**),

$$R^2 = \frac{Q_R}{Q} = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = \frac{\sum_{i=1}^{n} b_1^2 (x_i - \bar{x})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} = \frac{b_1^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 / n}{\sum_{i=1}^{n} (y_i - \bar{y})^2 / n} =$$

$$= \frac{b_1^2 s_x^2}{s_y^2} = \left(\frac{b_1 s_x}{s_y} \right)^2 = r^2.$$

▶ **Пример 3.5.**

По данным табл. 3.1 найти коэффициент детерминации и пояснить его смысл.

Р е ш е н и е. В примере **3.4** было получено Q_R =25,21, Q = 33,6.

По формуле (**3.47**) $R^2 = \frac{Q_R}{Q} = \frac{25{,}21}{33{,}6} = 0{,}750$. (Коэффициент детерминации можно было вычислить и иначе, если учесть, что в примере **3.2** был вычислен коэффициент корреляции r=0,866. Тогда R^2=r^2=0,866^2=0,750.)

Это означает, что вариация зависимой переменной Y — сменной добычи угля на одного рабочего — на 75,0% объясняется изменчивостью объясняющей переменной X — мощностью пласта. ▶

3.7. Геометрическая интерпретация регрессии и коэффициента детерминации

Рассмотрим геометрическую интерпретацию регрессии. Предположим, что мы имеем n=3 наблюдения: y_1, y_2, y_3 — зависимой переменной Y и x_1, x_2, x_3 [1] — объясняющей переменной X. Рассматривая трехмерное пространство с осями координат *1*, *2*, *3,* можно построить векторы Y=(y_1, y_2, y_3), X=(x_1, x_2, x_3), а также вектор S=(1;1;1) (рис. 3.7). Тогда значения $\hat{y}_1, \hat{y}_2, \hat{y}_3$, получаемые по уравнению регрессии $\hat{y} = b_0 + b_1 x$, можно рассматривать как компоненты вектора $\hat{Y}$, представляющего собой линейную комбинацию векторов S и X, т. е. $\hat{Y} = b_0 S + b_1 X$.

Необходимо найти такие значения оценок b_0 и b_1, при которых вектор $\hat{Y}$ наилучшим образом аппроксимирует (заменяет) вектор Y, т.е. вектор остатков e = (e_1, e_2, e_3), где $e_i = \hat{y}_i - y_i$ (i = 1, 2, 3) будет иметь *минимальную* длину. Очевидно, решением задачи будет такой вектор Y, для которого вектор e перпендикулярен плоскости π, образуемой векторами X и S (рис. 3.7), а значит, перпендикулярен и самим векторам X и S, т.е $e \perp X$ и $e \perp S$.

[1] Полагаем, что равенство x_1=x_2=x_3 не выполняется.

Условием перпендикулярности пары векторов является равенство нулю их скалярного произведения (**11.27**):

$$(e,\ S)=0 \text{ или } \sum_{i=1}^{n} e_i \cdot 1 = \sum_{i=1}^{n} \left(b_0 + b_1 x_i - y_i\right) = 0;$$

$$(e,\ X)=0 \text{ или } \sum_{i=1}^{n} e_i x_i = \sum_{i=1}^{n} \left(b_0 + b_1 x_i - y_i\right) x_i = 0$$

(где n=3), т. е. мы получим те же условия, из которых находятся «наилучшие» оценки b_0, b_1 (см. (**3.13**), (**3.11**)) метода наименьших квадратов.

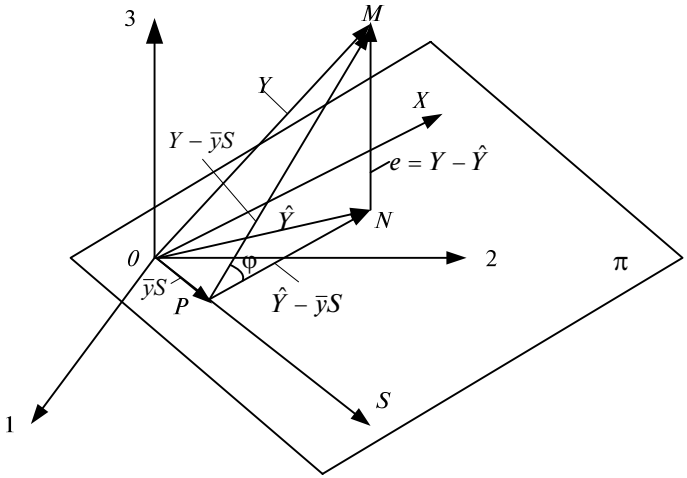


Рис. 3.7

Вектор OP есть ортогональная проекция вектора Y на вектор S. Из векторной алгебры известно, что длина такого вектора равна отношению скалярного произведения векторов Y и S к длине вектора S, т. е.

$$|OP| = \frac{(Y,S)}{|S|} = \frac{y_1 \cdot 1 + y_2 \cdot 1 + y_3 \cdot 1}{\sqrt{1^2 + 1^2 + 1^2}} = \frac{\sum_{i=1}^{3} y_i}{3} \cdot \sqrt{3} = \bar{y}|S|,$$

где $\bar{y}$ — среднее значение переменной Y.

Следовательно, вектор $OP = \bar{y}S$.

Вектор $\hat{Y}$ есть ортогональная проекция вектора Y на плоскость π. По известной в стереометрии теореме о трех перпендикулярах проекция вектора $\hat{Y}$ на вектор S совпадает с OP. Следовательно, прямоугольный треугольник PMN образуют векторы $PM = Y - \bar{y}S$, $PN = \hat{Y} - \bar{y}S$, $NM = Y - \hat{Y} = e$ (см. рис. 3.7). По теореме Пифагора $|PM|^2 = |PN|^2 + |NM|^2$, где вертикальными чертами отмечены длины векторов. Это равенство соответствует разложению (**3.41**) общей суммы Q квадратов отклонений зависимой переменной Y от средней $\bar{y}$ на сумму квадратов Q_R, обусловленную регрессией, и остаточную сумму квадратов Q_e, т. е. $Q = Q_R + Q_e$. Поэтому коэффициент детерминации R^2, определяемый по (**3.47**), примет вид:

$$R^2 = |PN|^2 / |PM|^2 = \cos^2 \varphi,$$

где φ — угол между векторами PN и PM.

Геометрическая интерпретация регрессии, проведенная нами при $n=3$ наблюдениях, в принципе сохраняется и при $n>3$, однако при этом она теряет свою наглядность.

3.8. Коэффициент ранговой корреляции Спирмена

До сих пор мы анализировали зависимость между двумя *количественными* переменными. Вместе с тем в практике эконометриста иногда встречаются случаи, когда необходимо установить тесноту связи между *ординальными* (*порядковыми*) переменными (например, качество жилищных условий, тестовые баллы, экзаменационные оценки и т. п.). В этом случае объекты анализа упорядочивают или *ранжируют* по степени выраженности измеряемых переменных. При этом каждому объекту присваивается определенный номер, называемый *рангом*. Например, объекту с наименьшим проявлением (значением) признака присваивается ранг 1, следующему за ним — ранг 2 и т. д. Если объекты ранжированы по двум признакам, то имеется возможность оценить тесноту связи между переменными, основываясь на рангах, т. е. тесноту *ранговой корреляции*.

Коэффициент ранговой корреляции Спирмена находится по формуле

$$\rho = 1 - \frac{6\sum_{i=1}^{n}(r_i - s_i)^2}{n^3 - n}, \qquad \textbf{(3.49)}$$

где r_i и s_i ранги i-го объекта по переменным X и Y; n — число пар наблюдений.

Если ранги всех объектов равны ($r_i = s_i$, i=1,2,..., n), то ρ=1, т. е. при полной прямой связи ρ=1. При полной обратной связи, когда ранги объектов по двум переменным расположены в обратном порядке, можно показать, что ρ=−1. Во всех остальных случаях $|\rho| < 1$.

При ранжировании иногда сталкиваются со случаями, когда невозможно найти существенные различия между объектами по величине проявления рассматриваемого признака: объекты, как говорят, оказываются *связанными*. Связанным объектам приписывают одинаковые средние ранги, такие, чтобы сумма всех рангов оставалась такой же, как и при отсутствии связанных рангов. Например, если четыре объекта оказались равнозначными в отношении рассматриваемого признака и невозможно определить, какие из четырех рангов (4,5,6,7) приписать этим объектам, то каждому объекту приписывается средний ранг, равный (4+5+6+7)/4=5,5. В модификациях формулы (**3.49**) на связанные ранги вводятся поправки.

При проверке значимости ρ исходят из того, что в случае справедливости нулевой гипотезы об отсутствии корреляционной связи между переменными при n>10 статистика

$$t = \frac{\rho\sqrt{n-2}}{\sqrt{1-\rho^2}} \qquad \textbf{(3.50)}$$

имеет t-распределение Стьюдента с (n−2) степенями свободы. Поэтому ρ значим на уровне α, если $|t| > t_{1-\alpha;\, n-2}$, где $t_{1-\alpha;\, n-2}$ — табличное значение t-критерия Стьюдента, определенное на уровне значимости α при числе степеней свободы (n−2).

▶ **Пример 3.6.** По результатам тестирования 10 студентов по двум дисциплинам *A* и *B* на основе набранных баллов получены следующие ранги (табл. 3.4). Вычислить коэффициент ранговой корреляции Спирмена и проверить его значимость на уровне α=0,05.

Р е ш е н и е. Разности рангов и их квадраты поместим в последних двух строках табл. 3.4.

Т а б л и ц а 3.4

Ранги по дисциплинам		*Результаты тестирования студентов*										*Всего*
		1-й	*2-й*	*3-й*	*4- й*	*5-й*	*6-й*	*7-й*	*8-й*	*9-й*	*10-й*	
A	r_i	2	4	5	1	7,5	7,5	7,5	7,5	3	10	55
B	s_i	2,5	6	4	1	2,5	7	8	9,5	5	9,5	55
$r_i - s_i$		−0,5	−2	1	0	5	0,5	−0,5	−2	−2	0,5	—
$(r_i - s_i)^2$		0,25	4	1	0	25	0,25	0,25	4	4	0,25	39

По формуле (**3.49**) $\rho = 1 - \dfrac{6 \cdot 39}{10^3 - 10} = 0{,}763$.

Для проверки значимости ρ по формуле (**3.50**) вычислим $t = 0{,}763 \dfrac{\sqrt{10-2}}{\sqrt{1-0{,}763^2}} = 3{,}34$ и найдем по табл. II приложений $t_{0,95;8} = 2{,}31$. Так как $t > t_{0,95;8}$, то коэффициент ранговой корреляции ρ значим на 5%-ном уровне. Связь между оценками дисциплин достаточно тесная. ▶

Ранговый коэффициент корреляции ρ может быть использован и для оценки тесноты связи между обычными количественными переменными. Достоинство ρ здесь заключается в том, что нахождение этого коэффициента не требует нормального распределения переменных, линейной связи между ними. Однако необходимо учитывать, что при переходе от первоначальных значений переменных к их рангам происходит определенная потеря информации. Чем теснее связь, чем меньше корреляционная зависимость между переменными отличается от линейной, тем ближе коэффициент корреляции Спирмена ρ к коэффициенту парной корреляции *r*.

Упражнения

3.7. Имеются следующие данные об уровне механизации работ *X*(%) и производительности труда *Y* (т/ч) для 14 однотипных предприятий:

x_i	32	30	36	40	41	47	56	54	60	55	61	67	69	76
y_i	20	24	28	30	31	33	34	37	38	40	41	43	45	48

Необходимо: *а*) оценить тесноту и направление связи между переменными с помощью коэффициента корреляции; *б*) найти уравнение регрессии Y по X.

3.8. При исследовании корреляционной зависимости между ценой на нефть X и индексом нефтяных компаний Y получены следующие данные:

$\bar{x}$ =16,2 (ден. ед.), $\bar{y} = 4000$ (усл. ед.), $s_x^2 = 4$, $s_y^2 = 500$, $\hat{\mathrm{Cov}}(X, Y) = 40$.

Необходимо: *а*) составить уравнение регрессии Y по X; *б*) используя уравнение регрессии, найти среднее значение индекса при цене на нефть 16,5 ден. ед.

3.9. По данным примера **3.7**: *а*) найти уравнение регрессии Y по X; *б*) найти коэффициент детерминации R^2 и пояснить его смысл; *в*) проверить значимость уравнения регрессии на 5%-ном уровне по F-критерию; *г*) оценить среднюю производительность труда на предприятиях с уровнем механизации работ 60% и построить для нее 95%-ный доверительный интервал; аналогичный доверительный интервал найти для индивидуальных значений производительности труда на тех же предприятиях.

3.10. По данным 30 нефтяных компаний получено следующее уравнение регрессии между оценкой Y (ден. ед.) и фактической стоимостью X (ден. ед.) этих компаний: $y_x = 0{,}8750x + 295$. Найти: 95%-ные доверительные интервалы для среднего и индивидуального значений оценки предприятий, фактическая стоимость которых составила 1300 ден. ед., если коэффициент корреляции между переменными равен 0,76, а среднее квадратическое отклонение переменной X равно 270 ден. ед.

3.11. При приеме на работу семи кандидатам было предложено два теста. Результаты тестирования приведены в таблице:

Тест	*Результаты тестирования кандидатов* (*в баллах*)						
	1-й	*2-й*	*3-й*	*4-й*	*5-й*	*6-й*	*7-й*
1	31	82	25	26	53	30	29
2	21	55	8	27	32	42	26

Вычислить коэффициент ранговой корреляции Спирмена между результатам тестирования по двум тестам и на уровне α=0,05 оценить его значимость.

Глава 4
Множественный регрессионный анализ

4.1. Классическая нормальная линейная модель множественной регрессии

Экономические явления, как правило, определяются большим числом одновременно и совокупно действующих факторов. В связи с этим часто возникает задача исследования зависимости одной зависимой переменной Y от нескольких объясняющих переменных X_1, X_2,..., X_n. Эта задача решается с *помощью множественного регрессионного анализа*.

Обозначим i-е наблюдение зависимой переменной y_i, а объясняющих переменных — x_{i1}, x_{i2},..., x_{ip}. Тогда модель множественной линейной регрессии можно представить в виде:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_p x_{ip} + \varepsilon_i, \qquad \textbf{(4.1)}$$

где $i = 1,2,..., n$; ε_i удовлетворяет приведенным выше предпосылкам **(3.23)—(3.25)**.

Модель (**4.1**), в которой зависимая переменная y_i, возмущения ε_i и объясняющие переменные x_{i1}, x_{i2},..., x_{ip} удовлетворяют приведенным выше (**§ 3.4**) предпосылкам *1—5* регрессионного анализа и, кроме того, предпосылке *6* о невырожденности матрицы (независимости столбцов) значений объясняющих переменных[1], называется *классической нормальной линейной моделью множественной регрессии* (*Classic Normal Linear Multiple Regression model*).

Включение в регрессионную модель новых объясняющих переменных усложняет получаемые формулы и вычисления. Это приводит к целесообразности использования матричных обозначений. Матричное описание регрессии облегчает как теоретические концепции анализа, так и необходимые расчетные процедуры.

[1] См. дальше, с. 86.

Введем обозначения: $Y=(y_1\ y_2 \ldots\ y_n)'$ — *матрица-столбец, или вектор, значений зависимой переменной размера* n[1];

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1p} \\ 1 & x_{21} & x_{22} & \ldots & x_{2p} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{np} \end{pmatrix}$$

— матрица значений объясняющих переменных, или матрица плана размера $n\times(p+1)$ (обращаем внимание на то, что в матрицу X дополнительно введен столбец, все элементы которого равны 1, т.е. условно полагается, что в модели **(4.1)** свободный член β_0 умножается на фиктивную переменную x_{i0}, принимающую значение 1 для всех i: $x_{i0} = 1$ ($i = 1,2,\ldots, n$);

$\beta = (\beta_0\ \beta_1 \ldots\ \beta_p)'$ — *матрица-столбец, или вектор, параметров* размера $(p+1)$; $\varepsilon = (\varepsilon_1\ \varepsilon_2 \ldots\ \varepsilon_n)'$ — *матрица-столбец, или вектор, возмущений* (*случайных ошибок*, *остатков*) *размера* n.

Тогда в матричной форме модель **(4.1)** примет вид:

$$Y = X\beta+\varepsilon. \qquad \textbf{(4.2)}$$

Оценкой этой модели по выборке является уравнение

$$Y = Xb+e, \qquad \textbf{(4.2′)}$$

где $b = (b_0\ b_1 \ldots\ b_p)'$, $e = (e_1\ e_2 \ldots\ e_n)'$.

4.2. Оценка параметров классической регрессионной модели методом наименьших квадратов

Для оценки вектора неизвестных параметров β применим *метод наименьших квадратов*. Так как произведение транспонированной матрицы e' на саму матрицу e

$$e'e = (e_1\ e_2 \ldots e_n)\begin{pmatrix} e_1 \\ e_2 \\ \ldots \\ e_n \end{pmatrix} = e_1^2 + e_2^2 + \ldots + e_n^2 = \sum_{i=1}^{n} e_i^2,$$

[1] Знаком «′» обозначается операция транспонирования матриц.

то условие минимизации остаточной суммы квадратов запишется в виде:

$$S = \sum_{i=1}^{n} (\hat{y}_{x_i} - y_i)^2 = \sum_{i=1}^{n} e_i^2 = e'e = (Y - Xb)'(Y - Xb) \to \min. \textbf{(4.3)}$$

Учитывая, что при транспонировании произведения матриц получается произведение транспонированных матриц, взятых в обратном порядке, т.е. $(Xb)'=b'X$; после раскрытия скобок получим:

$$S = Y'Y - b'X'Y - Y'Xb + b'X'Xb.$$

Произведение $Y'Xb$ есть матрица размера $(1\times n)[n\times(p+1)]\times[(p+1)\times 1]=(1\times 1)$, т.е. величина скалярная, следовательно, оно не меняется при транспонировании, т.е. $Y'Xb = (Y'Xb)' = b'X'Y$. Поэтому условие минимизации **(4.3)** примет вид:

$$S = Y'Y - 2b'X'Y + b'X'Xb \to \min. \qquad \textbf{(4.4)}$$

На основании необходимого условия экстремума функции нескольких переменных $S(b_0, b_1,..., b_p)$, представляющей **(4.3)**, необходимо приравнять нулю частные производные по этим переменным или в матричной форме — вектор частных производных

$$\frac{\partial S}{\partial b} = \left(\frac{\partial S}{\partial b_0} \; \frac{\partial S}{\partial b_1} \ldots \frac{\partial S}{\partial b_p} \right).$$

Для вектора частных производных доказаны следующие формулы[1] **(§ 12.10)**:

$$\frac{\partial}{\partial b}(b'c) = c, \quad \frac{\partial}{\partial b}(b'Ab) = 2Ab,$$

где b и c — вектор-столбцы; A — симметрическая матрица, в которой элементы, расположенные симметрично относительно главной диагонали, равны.

---

[1] Справедливость приведенных формул проиллюстрируем на примере.

Пусть $b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$, $c = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$, $A = \begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix}$. Так как $b'c = (b_1 b_2)\begin{pmatrix} 3 \\ 4 \end{pmatrix} = 3b_1 + 4b_2$ и $b'Ab = (b_1 b_2)\begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix}\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = 2b_1^2 + 6b_1b_2 + 5b_2^2$, то $\frac{\partial}{\partial b}(b'c) = \frac{\partial}{\partial b}(3b_1 + 4b_2) = \begin{pmatrix} 3 \\ 4 \end{pmatrix} = c$ и $\frac{\partial}{\partial b}(b'Ab) = \frac{\partial}{\partial b}(2b_1^2 + 6b_1b_2 + 5b_2^2) = \begin{pmatrix} 4b_1 + 6b_2 \\ 6b_1 + 10b_2 \end{pmatrix} = 2\begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix}\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = 2Ab.$

Поэтому, полагая $c = X'Y$, а матрицу $A = X'X$ (она является симметрической — см. **(4.6)**), найдем

$$\frac{\partial S}{\partial b} = -2X'Y' + 2X'Xb = 0,$$

откуда получаем систему нормальных уравнений в матричной форме для определения вектора b:

$$X'Xb = X'Y. \qquad \textbf{(4.5)}$$

Найдем матрицы, входящие в это уравнение[1]. Матрица $X'X$ представляет матрицу сумм первых степеней, квадратов и попарных произведений n наблюдений объясняющих переменных:

$$X'X = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{n1} \\ \dots & \dots & \dots & \dots \\ x_{1p} & x_{2p} & \dots & x_{np} \end{pmatrix} \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{np} \end{pmatrix} = \qquad \textbf{(4.6)}$$

$$= \begin{pmatrix} n & \sum x_{i1} & \dots & \sum x_{ip} \\ \sum x_{i1} & \sum x_{i1}^2 & \dots & \sum x_{i1}x_{ip} \\ \dots & \dots & \dots & \dots \\ \sum x_{ip} & \sum x_{i1}x_{ip} & \dots & \sum x_{ip}^2 \end{pmatrix}.$$

Матрица $X'Y$ есть вектор произведений n наблюдений объясняющих и зависимой переменных:

$$X'Y = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{n1} \\ \dots & \dots & \dots & \dots \\ x_{1p} & x_{2p} & \dots & x_{np} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum y_i x_{i1} \\ \dots \\ \sum y_i x_{ip} \end{pmatrix}. \qquad \textbf{(4.7)}$$

В частном случае из рассматриваемого матричного уравнения **(4.5)** с учетом **(4.6)** и **(4.7)** для одной объясняющей переменной (p=1) нетрудно получить уже рассмотренную выше

---

[1] Здесь под знаком $\sum$ подразумевается $\sum_{i=1}^{n}$.

систему нормальных уравнений **(3.5).** Действительно, в этом случае матричное уравнение **(4.5)** принимает вид:

$$\begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum y_i x_i \end{pmatrix},$$

откуда непосредственно следует система нормальных уравнений **(3.5).**

Для решения матричного уравнения **(4.5)** относительно вектора оценок параметров b необходимо ввести еще одну предпосылку *6* (см. с. 61) для множественного регрессионного анализа: *матрица $X'X$ является неособенной*, т. е. ее определитель не равен нулю. Следовательно, ранг матрицы $X'X$ равен ее порядку, т.е. $r(X'X)=p+1$. Из матричной алгебры известно (см. § **12.4**), что $r(X'X)=r(X)$, значит, $r(X)=p+1$, т. е. ранг матрицы плана X равен числу ее столбцов. Это позволяет сформулировать предпосылку *6* множественного регрессионного анализа в следующем виде:

6. *Векторы значений объясняющих переменных, или столбцы матрицы плана X, должны быть линейно независимыми, т. е. ранг матрицы X — максимальный* ($r(X)=p+1$).

Кроме того, полагают, что *число* имеющихся *наблюдений* (значений) *каждой из объясняющих и зависимой переменных превосходит ранг матрицы X*, т. е. $n>r(X)$ или $n>p+1$, ибо в противном случае в принципе невозможно получение сколько-нибудь надежных статистических выводов.

Ниже, в § **4.3**, рассматривается ковариационная матрица вектора возмущений $\sum_\varepsilon$, являющаяся многомерным аналогом дисперсии одной переменной. Поэтому в новых терминах[1] приведенные ранее (с. 61, 82 и здесь) *предпосылки для множественного регрессионного анализа* могут быть записаны следующим образом [2]:

1. В модели (**4.2**) ε — случайный вектор, X — неслучайная (детерминированная) матрица.

2. $M(\varepsilon)=\mathbf{0}_n$.

3,4. $\sum_\varepsilon = M(\varepsilon\varepsilon')=\sigma^2 E_n$.

---

[1] В случае одной объясняющей переменной отпадает необходимость в записи под символом x второго индекса, указывающего номер переменной.

[2] При первом чтении этот материал может быть опущен. E_n — единичная матрица n-го порядка; $\mathbf{0}_n$ — нулевой вектор размера n.

5. ε — нормально распределенный случайный вектор, т.е. $\varepsilon \sim N_n(\mathbf{0}; \sigma^2 E_n)$.

6. $r(X) = p+1<n$.

Как уже отмечено в § **4.1**, модель (**4.2**), удовлетворяющая приведенным предпосылкам *1–6*, называется *классической нормальной линейной моделью множественной регрессии*; если же среди приведенных не выполняется лишь предпосылка *5* о нормальном законе распределения вектора возмущений ε, то модель (**4.2**) называется просто *классической линейной моделью множественной регрессии*.

Решением уравнения **(4.5)** является вектор

$$b = (X'X)^{-1} X'Y, \qquad \textbf{(4.8)}$$

где $(X'X)^{-1}$ — *матрица, обратная матрице коэффициентов системы* **(4.5)**, $X'Y$ *— матрица-столбец,* или *вектор, ее свободных членов*.

Теорема Гаусса—Маркова, рассмотренная выше для парной регрессионной модели, оказывается верной и в общем виде для модели (**4.2**) множественной регрессии:

При выполнении предпосылок[1] *множественного регрессионного анализа оценка метода наименьших квадратов* $b = (X'X)^{-1}X'Y$ *является наиболее эффективной, т. е. обладает наименьшей дисперсией в классе линейных несмещенных оценок* (*Best Linear Unbiased Estimator,* или *BLUE*)[2].

Зная вектор b, выборочное уравнение множественной регрессии представим в виде:

$$\hat{y} = X_0' b, \qquad \textbf{(4.9)}$$

где $\hat{y}$ — групповая (условная) средняя переменной Y при заданном векторе значений объясняющей переменной

$$X_0' = (1 \quad x_{10} \quad x_{20} \quad \ldots \quad x_{p0}).$$

▶ **Пример 4.1.** Имеются следующие данные[3] (условные) о сменной добыче угля на одного рабочего Y (т), мощности пласта

---

[1] Не включая предпосылку **5** — требование нормальности закона распределения вектора возмущений ε, которая в теореме Гаусса—Маркова не требуется.

[2] Доказательство теоремы приведено в § **4.4**.

[3] В этом примере использованы данные примера **3.1** с добавлением результатов наблюдений над новой объясняющей переменной X_2, при этом старую переменную X из примера **3.1** обозначаем теперь X_1.

X_1 (м) и уровне механизации работ X_2(%), характеризующие процесс добычи угля в 10 шахтах.

Т а б л и ц а 4.1

i	x_{i1}	x_{i2}	y_i	i	x_{i1}	x_{i2}	y_i
1	8	5	5	6	8	8	6
2	11	8	10	7	9	6	6
3	12	8	10	8	9	4	5
4	9	5	7	9	8	5	6
5	8	7	5	10	12	7	8

Предполагая, что между переменными Y, X_1 и X_2 существует линейная корреляционная зависимость, найти ее аналитическое выражение (уравнение регрессии Y по X_1 и X_2).

Р е ш е н и е. Обозначим

$$Y = \begin{pmatrix} 5 \\ 10 \\ \dots \\ 8 \end{pmatrix}, \qquad X = \begin{pmatrix} 1 & 8 & 5 \\ 1 & 11 & 8 \\ \dots & \dots & \dots \\ 1 & 12 & 7 \end{pmatrix}$$

(напоминаем, что в матрицу плана X вводится дополнительный столбец чисел, состоящий из единиц).

Для удобства вычислений составляем вспомогательную таблицу.

Т а б л и ц а 4.2

i	x_{i1}	x_{i2}	y_i	x_{i1}^2	x_{i2}^2	y_i^2	$x_{i1}x_{i2}$	y_ix_{i1}	y_ix_{i2}	$\hat{y}_i$	$e_i^2 = (\hat{y}_i - y_i)^2$
1	8	5	5	64	25	25	40	40	25	5,13	0,016
2	11	8	10	121	64	100	88	110	80	8,79	1,464
3	12	8	10	144	64	100	96	120	80	9,64	1,127
4	9	5	7	81	25	49	45	63	35	5,98	1,038
5	8	7	5	64	49	25	56	40	35	5,86	0,741
6	8	8	6	64	64	36	64	48	48	6,23	0,052
7	9	6	6	81	36	36	54	54	36	6,35	0,121
8	9	4	5	81	16	25	36	45	20	5,61	0,377
9	8	5	6	64	25	36	40	48	30	5,13	0,762
10	12	7	8	144	49	64	84	96	56	9,28	1,631
Σ	94	63	68	908	417	496	603	664	445	—	6,329

Теперь

$$X'X = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 8 & 11 & \dots & 12 \\ 5 & 8 & \dots & 7 \end{pmatrix} \begin{pmatrix} 1 & 8 & 5 \\ 1 & 11 & 8 \\ \dots & \dots & \dots \\ 1 & 12 & 7 \end{pmatrix} = \begin{pmatrix} 10 & 94 & 63 \\ 94 & 908 & 603 \\ 63 & 603 & 417 \end{pmatrix}$$

(см. суммы в итоговой строке табл. **4.2**);

$$X'Y = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 8 & 11 & \dots & 12 \\ 5 & 8 & \dots & 7 \end{pmatrix} \begin{pmatrix} 5 \\ 10 \\ \dots \\ 8 \end{pmatrix} = \begin{pmatrix} 68 \\ 664 \\ 445 \end{pmatrix}.$$

Матрицу $A^{-1}=(X'X)^{-1}$ определим по формуле $A^{-1} = \frac{1}{|A|}\overline{A}$, где $|A|$ — определитель матрицы $X'X$; $\overline{A}$ — матрица, присоединенная к матрице $X'X$. Получим

$$A^{-1} = \frac{1}{3738} \begin{pmatrix} 15027 & -1209 & -522 \\ -1209 & 201 & -108 \\ -522 & -108 & 244 \end{pmatrix}$$

(рекомендуем читателю убедиться в этом самостоятельно).

Теперь в соответствии с **(4.8)** умножая эту матрицу на вектор

$$X'Y = \begin{pmatrix} 68 \\ 664 \\ 445 \end{pmatrix},$$

получим $b = \frac{1}{3738} \begin{pmatrix} -13230 \\ 3192 \\ 1372 \end{pmatrix} = \begin{pmatrix} -3{,}5393 \\ 0{,}8539 \\ 0{,}3670 \end{pmatrix}$.

С учетом **(4.9)** уравнение множественной регрессии имеет вид: $\hat{y} = -3{,}54 + 0{,}854x_1 + 0{,}367x_2$. Оно показывает, что при увеличении только мощности пласта X_1 (при неизменном X_2) на 1 м добыча угля на одного рабочего Y увеличивается в среднем

на 0,854 т, а при увеличении только уровня механизации работ X_2 (при неизменной X_1) — в среднем на 0,367 т.

Добавление в регрессионную модель новой объясняющей переменной X_2 изменило коэффициент регрессии b_1 (Y по X_1) с 1,016 для парной регрессии (см. пример **3.1**) до 0,854 — для множественной регрессии. В этом никакого противоречия нет, так как во втором случае коэффициент регрессии позволяет оценить прирост зависимой переменной Y при изменении на единицу объясняющей переменной X_1 в чистом виде, независимо от X_2. В случае парной регрессии b_1 учитывает воздействие на Y не только переменной X_1, но и косвенно корреляционно связанной с ней переменной X_2. ▶

На практике часто бывает необходимо сравнение влияния на зависимую переменную различных объясняющих переменных, когда последние выражаются разными единицами измерения. В этом случае используют *стандартизованные коэффициенты регрессии* b'_j и *коэффициенты эластичности* E_j ($j = 1,2,..., p$):

$$b'_j = b_j \frac{s_{x_j}}{s_y}; \qquad \textbf{(4.10)}$$

$$E_j = b_j \frac{\bar{x}_j}{\bar{y}}. \qquad \textbf{(4.11)}$$

Стандартизованный коэффициент регрессии b'_j *показывает, на сколько величин* s_y *изменится в среднем зависимая переменная* Y *при увеличении только* j*-й объясняющей переменной на* s_{x_j}, *а коэффициент эластичности* E_j *— на сколько процентов* (*от средней*) *изменится в среднем* Y *при увеличении только* X_j *на* 1%.

▶ **Пример 4.2.**

По данным примера **4.1** сравнить раздельное влияние на сменную добычу угля двух факторов — мощности пласта и уровня механизации работ.

Р е ш е н и е. Для сравнения влияния каждой из объясняющих переменных по формуле **(4.10)** вычислим стандартизованные коэффициенты регрессии:

$$b'_1 = 0{,}8539 \cdot \frac{1{,}56}{1{,}83} = 0{,}728; \quad b'_2 = 0{,}3670 \cdot \frac{1{,}42}{1{,}83} = 0{,}285,$$

а по формуле **(4.11)** — коэффициенты эластичности:

$$E_1 = 0{,}8539 \cdot \frac{9{,}4}{6{,}8} = 1{,}180;\ E_2 = 0{,}3670 \cdot \frac{6{,}3}{6{,}8} = 0{,}340.$$

(Здесь мы опустили расчет необходимых характеристик переменных:

$$\bar{x}_1 = 9{,}4;\ \bar{x}_2 = 6{,}3;\ \bar{y} = 6{,}8;\ s_{x_1} = 1{,}56;\ s_{x_2} = 1{,}42;\ s_y = 1{,}83.)$$

Таким образом, увеличение мощности пласта и уровня механизации работ только на одно s_{x_1} или на одно s_{x_2} увеличивает в среднем сменную добычу угля на одного рабочего соответственно на $0{,}728 s_y$ или на $0{,}285 s_y$, а увеличение этих переменных на 1% (от своих средних значений) приводит в среднем к росту добычи угля соответственно на 1,18% и 0,34%. Итак, по обоим показателям на сменную добычу угля большее влияние оказывает фактор «мощность пласта» по сравнению с фактором «уровень механизации работ». ▶

Преобразуем вектор оценок **(4.8)** с учетом **(4.2)**:

$$b = (X'X)^{-1}X'(X\beta + \varepsilon) = (X'X)^{-1}(X'X)\beta + (X'X)^{-1}X'\varepsilon = \\ = E\beta + (X'X)^{-1}X'\varepsilon,$$

или

$$b = \beta + (X'X)^{-1}X'\varepsilon, \qquad \textbf{(4.12)}$$

т.е. *оценки параметров* **(4.8)**, найденные по выборке, *будут содержать случайные ошибки*.

Так как математическое ожидание оценки b равно оцениваемому параметру β, т. е.

$$M(b) = M[\beta + (X'X)^{-1}X'\varepsilon] = M(\beta) + (X'X)^{-1}X'M(\varepsilon) = \beta,$$

ибо в силу (**3.23**) $M(\varepsilon)=0$, то, очевидно, что *вектор b есть несмещенная оценка параметра* β.

4.3. Ковариационная матрица и ее выборочная оценка

Вариации оценок параметров будут в конечном счете определять точность уравнения множественной регрессии. Для их

измерения в многомерном регрессионном анализе рассматривают так называемую ***ковариационную матрицу*** *вектора оценок параметров* $\sum_b$, являющуюся *матричным аналогом дисперсии одной переменной*:

$$\sum_b = \begin{pmatrix} \sigma_{00} & \sigma_{01} & \dots & \sigma_{0p} \\ \sigma_{10} & \sigma_{11} & \dots & \sigma_{1p} \\ \dots & \dots & \dots & \dots \\ \sigma_{p0} & \sigma_{p1} & \dots & \sigma_{pp} \end{pmatrix},$$

где элементы σ_{ij} — *ковариации* (или *корреляционные моменты*) оценок параметров β_i и β_j. Ковариация двух переменных определяется как математическое ожидание произведения отклонений этих переменных от их математических ожиданий (см. § **2.4**). Поэтому

$$\sigma_{ij} = M[(b_i - M(b_i))(b_j - M(b_j))]. \qquad \textbf{(4.13)}$$

Ковариация характеризует как степень рассеяния значений двух переменных относительно их математических ожиданий, так и взаимосвязь этих переменных.

В силу того, что оценки b_j, полученные методом наименьших квадратов, являются несмещенными оценками параметров β_j, т. е. $M(b_j) = \beta_j$, выражение (**4.13**) примет вид:

$$\sigma_{ij} = M[(b_i - \beta_i)(b_j - \beta_j)].$$

Рассматривая ковариационную матрицу $\sum_b$, легко заметить, что на ее главной диагонали находятся дисперсии оценок параметров регресии, ибо

$$\sigma_{jj} = M\left[(b_j - \beta_j)(b_j - \beta_j)\right] = M(b_j - \beta_j)^2 = \sigma_{b_j}^2. \qquad \textbf{(4.14)}$$

В сокращенном виде ковариационная матрица вектора оценок параметров $\sum_b$ имеет вид:

$$\sum_b = M\left[(b - \beta)(b - \beta)'\right]$$

(в этом легко убедиться, перемножив векторы $(b - \beta)$ и $(b - \beta)'$).

Учитывая (**4.12**), преобразуем это выражение:

$$\sum_b = M\left\{\left[(X'X)^{-1}X'\varepsilon\right]\left[(X'X)^{-1}X'\varepsilon\right]'\right\} =$$
$$= M\left[(X'X)^{-1}X'\varepsilon\varepsilon' X(X'X)^{-1}\right] =$$
$$= (X'X)^{-1}X'M(\varepsilon\varepsilon')X(X'X)^{-1}, \qquad \textbf{(4.15)}$$

ибо элементы матрицы X — неслучайные величины.

Матрица $M(\varepsilon\varepsilon')$ представляет собой ковариационную *матрицу вектора возмущений*

$$\sum_{\varepsilon} = M(\varepsilon\varepsilon') = \begin{pmatrix} M(\varepsilon_1^2) & M(\varepsilon_1\varepsilon_2) & \ldots & M(\varepsilon_1\varepsilon_n) \\ M(\varepsilon_2\varepsilon_1) & M(\varepsilon_2^2) & \ldots & M(\varepsilon_2\varepsilon_n) \\ \ldots & \ldots & \ldots & \ldots \\ M(\varepsilon_n\varepsilon_1) & M(\varepsilon_n\varepsilon_2) & \ldots & M(\varepsilon_n^2) \end{pmatrix},$$

в которой все элементы, не лежащие на главной диагонали, равны нулю в силу предпосылки *4* о некоррелированности возмущений ε_i и ε_j между собой (см. **(3.25)**), а все элементы, лежащие на главной диагонали, в силу предпосылок *2* и *3* регрессионного анализа (см. **(3.23)** и **(3.24)**) равны одной и той же дисперсии σ^2:

$$M(\varepsilon_i^2) = M(\varepsilon_i - 0)^2 = D(\varepsilon_i^2) = \sigma^2.$$

Поэтому матрица

$$M(\varepsilon\varepsilon') = \sigma^2 E_n,$$

где E_n — единичная матрица n-го порядка. Следовательно, в силу **(4.15)** ковариационная матрица вектора оценок параметров:

$$\sum_b = [(X'X)^{-1}X'(\sigma^2 E_n)]X(X'X)^{-1} = \sigma^2((X'X)^{-1}(X'E_nX)(X'X)^{-1},$$

или
$$\sum_b = \sigma^2(X'X)^{-1}. \qquad \textbf{(4.16)}$$

Итак, *с помощью обратной матрицы* $(X'X)^{-1}$ *определяется не только сам вектор* b *оценок параметров* **(4.8)**, *но и дисперсии и ковариации его компонент.*

4.4. Доказательство теоремы Гаусса—Маркова. Оценка дисперсии возмущений

Теперь мы имеем возможность привести **доказательство теоремы Гаусса—Маркова**, сформулированной в § **4.2**.

Выше (§ **4.2**) мы уже показали, что оценка метода наименьших квадратов $b = (X'X)^{-1}X'Y$ есть несмещенная оценка для вектора параметров β, т. е. $M(b) = \beta$. Любую другую оценку b_1 вектора β без ограничения общности можно представить в виде

$$b_1 = [(X'X)^{-1}X' + C]\,Y,$$

где C — некоторая матрица размера $(p+1)\times n$.

Так как рассматриваемые в теореме оценки относятся к классу несмещенных оценок, то и $M(b_1) = \beta$ или $M(b_1) = M[(X'X)^{-1}X' + C]\,Y = \beta$.

Учитывая, что матрица в квадратных скобках — неслучайная, а в силу предпосылки *2* регрессионного анализа $M(\varepsilon)=0$, получим

$$M(b_1) = [(X'X)^{-1}X' + C]\;M(Y) = [(X'X)^{-1}X' + C]X\beta =$$
$$= [(X'X)^{-1}X'X + CX]\,\beta = (E + CX)\beta = \beta,$$

откуда следует, что $CX = 0$.

Далее

$$b_1 - \beta = [(X'X)^{-1}X' + C]\,Y - \beta = [(X'X)^{-1}X' + C](X\beta + \varepsilon) - \beta =$$
$$= (X'X)^{-1}X'X\beta + CX\beta + [(X'X)^{-1}X' + C]\varepsilon - \beta = [(X'X)^{-1}X' + C]\,\varepsilon,$$

так как $CX = 0$, $(X'X)^{-1}X'X\,\beta = E\beta = \beta$.

Теперь с помощью преобразований, аналогичных проведенным при получении формул (**4.15**), (**4.16**), найдем, что ковариационная матрица вектора оценок b_1, т. е.

$$\sum_{b_1} = M[(b_1 - \beta)(b_1 - \beta)'] = \sigma^2(X'X)^{-1} + \sigma^2 CC',$$

или, учитывая (**4.16**),

$$\sum_{b_1} = \sum_{b} + \sigma^2 CC'.$$

Диагональные элементы матрицы CC' неотрицательны[1], ибо они равны суммам квадратов элементов соответствующих строк этой матрицы. А так как диагональные элементы матриц $\sum_{b_1}$ и $\sum_{b}$ есть дисперсии компонент векторов оценок b_{1i} и b_i, то дисперсия $\sigma^2_{b_{1i}} \geq \sigma^2_{b_i}$ (i=1,2,..., p+1). Это означает, что *оценки коэффициентов регрессии, найденных методом наименьших квадратов, обладают наименьшей дисперсией*, что и требовалось доказать.

Итак, мы доказали, что оцека b метода наименьших квадратов является «наилучшей» линейной оценкой параметра β. Перейдем теперь к **оценке** еще одного параметра — **дисперсии возмущений** σ^2.

Рассмотрим вектор остатков e, равный в соответствии с (**4.2**′) $e = Y - Xb$.

В силу (**4.2**) и (**4.8**)

$$e = (X\beta + \varepsilon) - X[(X'X)^{-1}X'(X\beta + \varepsilon)] =$$
$$= X\beta + \varepsilon - X(X'X)^{-1}(X'X)\beta - X(X'X)^{-1}X'\varepsilon =$$
$$= X\beta + \varepsilon - XE\beta - X(X'X)^{-1}X'\varepsilon,$$

или
$$e = \varepsilon - X(X'X)^{-1}X'\varepsilon$$

(учли, что произведение $(X'X)^{-1}(X'X)=E$, т. е. равно единичной матрице E_{p+1} (p+1)-го порядка).

Найдем транспонированный вектор остатков e'. Так как при транспонировании матрица $(X'X)^{-1}$ не меняется, т. е.

$$[(X'X)^{-1}]' = [(X'X)']^{-1} = (X'X)^{-1},$$

то

$$e' = [\varepsilon - X(X'X)^{-1}X'\varepsilon] = \varepsilon' - \varepsilon'X(X'X)^{-1}X'.$$

Теперь

$$M(e'e) = M[(\varepsilon' - \varepsilon'X(X'X)^{-1}X')(\varepsilon - X(X'X)^{-1}X'\varepsilon)] =$$
$$= M(\varepsilon'\varepsilon) - M[\varepsilon'X(X'X)^{-1}X'\varepsilon] - M[\varepsilon'X(X'X)^{-1}X'\varepsilon)] +$$
$$+ M[\varepsilon'X(X'X)^{-1}(X'X)(X'X)^{-1}X'\varepsilon].$$

Так как последние два слагаемых взаимно уничтожаются, то

---

[1] Матрица CC' так же, как и ковариационные матрицы $\sum_{b_1}$ и $\sum_{b_2}$, является *неотрицательно определенной* (см. §**12.8**). В этом смысле можно записать, что матрица $\sum_{b_1} \geq \sum_{b}$, т.е. вектор оценок $b = (X'X)^{-1}X'Y$, полученный методом наименьших квадратов, обладает меньшим рассеиванием относительно параметра β по сравнению с любым другим вектором несмещенных оценок.

$$M(e'e)=M(\varepsilon'\varepsilon)-M[\varepsilon' X\,(X' X)^{-1}\,X'\varepsilon]. \qquad \textbf{(4.17)}$$

Первое слагаемое выражения (**4.17**)

$$M\left(\varepsilon'\varepsilon\right)=M\left(\sum_{i=1}^{n}\varepsilon_i^2\right)=\sum_{i=1}^{n}M\left(\varepsilon_i^2\right)=\sum_{i=1}^{n}\sigma^2=n\sigma^2, \quad \textbf{(4.18)}$$

ибо в силу предпосылок *2,3* регрессионного анализа

$$M\left(\varepsilon_i^2\right)=M\left(\varepsilon_i-0\right)^2=D\left(\varepsilon_i\right)=\sigma^2.$$

Матрица $B=X(X' X)^{-1}\,X'$ симметрическая (§ **12.8**), так как

$$B'=[X\,(X' X)^{-1}X']'=X\,(X' X)^{-1}\,X', \text{ т. е. } B'=B.$$

Поэтому $\varepsilon' B\varepsilon$ представляет квадратическую форму $\sum_{i,j=1}^{n}b_{ij}\varepsilon_i\varepsilon_j$;

ее математическое ожидание

$$M\,(\varepsilon' B\varepsilon)=M\sum_{i,j=1}^{n}\left(b_{ij}\varepsilon_i\varepsilon_j\right)=\sum_{i,j=1}^{n}b_{ij}M\left(\varepsilon_i\varepsilon_j\right).$$

Последнюю сумму можно разбить на две составляющие суммы элементов на главной диагонали матрицы B и вне ее:

$$M(\varepsilon' B\varepsilon)=\sum_{i=1}^{n}b_{ii}\,M\left(\varepsilon_i^2\right)+\sum_{i,j=1(i\neq j)}^{n}b_{ij}M\left(\varepsilon_i\varepsilon_j\right).$$

Второе слагаемое равно нулю в силу предпосылки *4* регрессионного анализа, т.е. $M\left(\varepsilon_i\varepsilon_j\right)=0$. Сумма диагональных элементов матрицы B образует след матрицы $\mathrm{tr}(B)$ (§ **12.8**). Получим

$$M(\varepsilon' B\varepsilon)=\sum_{i=1}^{n}b_{ii}\sigma^2=\sigma^2\sum_{i=1}^{n}b_{ii}=\sigma^2\mathrm{tr}(B). \qquad \textbf{(4.19)}$$

Заменив матрицу B ее выражением, получим

$$M(\varepsilon' B\varepsilon)=\sigma^2\mathrm{tr}[X\left(X'X\right)^{-1}X']=\sigma^2\mathrm{tr}[\left(X'X\right)^{-1}\left(X'X\right)]=$$

$$=\sigma^2\mathrm{tr}\left(E_{p+1}\right)=\sigma^2\left(p+1\right),$$

так как след матрицы не меняется при ее транспонировании (см. § **12.14**), т. е. $\mathrm{tr}(AC)=\mathrm{tr}(CA)$, а след единичной матрицы (т.е. сумма ее диагональных элементов) равен порядку этой матрицы.

Теперь по формуле (**4.17**), учитывая (**4.18**) и (**4.19**), получим:

$$M(e'e)=n\sigma^2-\sigma^2(p+1)=(n-p-1)\,\sigma^2,$$

т. е.

$$M(e'e) = M\left(\sum_{i=1}^{n} e_i^2\right) = (n - p - 1)\,\sigma^2 . \qquad \textbf{(4.20)}$$

Равенство (**4.20**) означает, что несмещенная оценка s^2 параметра σ^2 или выборочная остаточная дисперсия s^2 определяется по формуле:

$$s^2 = \frac{e'e}{n - p - 1} = \frac{\sum_{i=1}^{n} e_i^2}{n - p - 1} . \qquad \textbf{(4.21)}$$

Полученная формула легко объяснима. В знаменателе выражения (**4.21**) стоит $n - (p + 1)$, а не $n-2$, как это было выше в (**3.26**). Это связано с тем, что теперь $(p+1)$ степеней свободы (а не две) теряются при определении неизвестных параметров, число которых вместе со свободным членом равно $(p+1)$.

Можно показать, что рассмотренные в этом параграфе оценки b и s^2 параметров β и σ^2 *при выполнении предпосылки 5 регрессионного анализа о нормальном распределении вектора возмущений* ε ($\varepsilon \sim N_n(\mathbf{0};\sigma^2 E_n)$) *являются независимыми.* Для этого в данном случае достаточно убедиться в некоррелированности оценок b и s^2.

4.5. Определение доверительных интервалов для коэффициентов и функции регрессии

Перейдем теперь к оценке значимости коэффициентов регрессии b_j и построению доверительного интервала для параметров регрессионной модели β_j (j=1,2,..., p).

В силу (**4.14**), (**4.16**) и изложенного выше оценка $s_{b_j}^2$ дисперсии $\sigma_{b_j}^2$ коэффициента регрессии b_j определится по формуле:

$$s_{b_j}^2 = s^2[(X'X)^{-1}]_{jj} ,$$

где s^2 — несмещенная оценка параметра σ^2;

$[(X'X)^{-1}]_{jj}$ — диагональный элемент матрицы $(X'X)^{-1}$.

Среднее квадратическое отклонение (стандартная ошибка) коэффициента регрессии b_j примет вид:

$$s_{b_j} = s\sqrt{[(X'X)^{-1}]_{jj}} . \qquad \textbf{(4.22)}$$

Значимость коэффициента регрессии b_j можно проверить, если учесть, что статистика $(b_j - \beta_j)/s_{b_j}$ имеет t-распределение Стьюдента с $k = n - p - 1$ степенями свободы. Поэтому b_j значимо отличается от нуля (иначе — гипотеза H_0 о равенстве параметра β_j нулю, т. е. H_0: $\beta_j = 0$, отвергается) на уровне значимости α, если $|t| = \dfrac{|b_j|}{s_{b_j}} > t_{1-\alpha;n-p-1}$, где $t_{1-\alpha;n-p-1}$ — табличное значение t-критерия Стьюдента, определенное на уровне значимости α при числе степеней свободы $k = n - p - 1$.

В общей постановке гипотеза H_0 о равенстве параметра β_j заданному числу β_{j0}, т. е. H_0: $\beta_j = \beta_{j0}$, отвергается, если

$$|t| = \frac{|b_j - \beta_{j0}|}{s_{b_j}} > t_{1-\alpha;n-p-1}. \qquad \textbf{(4.23)}$$

Поэтому доверительный интервал для параметра β_j есть

$$b_j - t_{1-\alpha;n-p-1}s_{b_j} \le \beta_j \le b_j + t_{1-\alpha;n-p-1}s_{b_j}. \qquad \textbf{(4.23′)}$$

Наряду с интервальным оцениванием коэффициентов регрессии по (**4.23′**) весьма важным для оценки точности определения зависимой переменной (прогноза) является построение **доверительного интервала для функции регрессии** или **для условного математического ожидания зависимой переменной** $M_x(Y)$, найденного в предположении, что объясняющие переменные X_1, X_2,..., X_p приняли значения, задаваемые вектором $X_0' = (1\ x_{10}\ x_{20}\ \ldots\ x_{p0})$. Выше такой интервал получен для уравнения парной регрессии (см. (**3.34**) и (**3.33**)). Обобщая соответствующие выражения на случай множественной регрессии, можно получить доверительный интервал для $M_x(Y)$:

$$\hat{y} - t_{1-\alpha;k}s_{\hat{y}} \le M(Y) \le \hat{y} + t_{1-\alpha;k}s_{\hat{y}}, \qquad \textbf{(4.24)}$$

где $\hat{y}$ — групповая средняя, определяемая по уравнению регрессии,

$$s_{\hat{y}} = s\sqrt{X_0'(X'X)^{-1}X_0} \qquad \textbf{(4.25)}$$

— ее стандартная ошибка.

При обобщении формул (**3.36**) и (**3.35**) аналогичный **доверительный интервал для индивидуальных значений зависимой переменной** y_0^* примет вид:

$$\hat{y}_0 - t_{1-\alpha;n-p-1}s_{\hat{y}_0} \le y_0^* \le \hat{y}_0 + t_{1-\alpha;\,n-p-1}s_{\hat{y}_0}, \qquad \textbf{(4.26)}$$

где

$$s_{\hat{y}_0} = s\sqrt{1 + X_0'\left(X'X\right)^{-1}X_0}\,. \qquad \textbf{(4.27)}$$

Доверительный интервал для параметра σ^2 в множественной регрессии строится аналогично парной модели по формуле (**3.39**) с соответствующим изменением числа степеней свободы критерия χ^2:

$$\frac{ns^2}{\chi^2_{\alpha/2;n-p-1}} \le \sigma^2 \le \frac{ns^2}{\chi^2_{1-\alpha/2;n-p-1}}. \qquad \textbf{(4.28)}$$

▶ **Пример 4.3.** По данным примера **4.1** оценить сменную добычу угля на одного рабочего для шахт с мощностью пласта 8 м и уровнем механизации работ 6%; найти 95%-ные доверительные интервалы для индивидуального и среднего значений сменной добычи угля на 1 рабочего для таких же шахт. Проверить значимость коэффициентов регрессии и построить для них 95%-ные доверительные интервалы. Найти интервальную оценку для дисперсии σ^2.

Р е ш е н и е. В примере **4.1** уравнение регрессии получено в виде $\hat{y} = -3{,}54 + 0{,}854x_1 + 0{,}367x_2$.

По условию надо оценить $M_x(Y)$, где $X_0' = (1\ 8\ 6)$. Выборочной оценкой $M_x(Y)$ является групповая средняя, которую найдем по уравнению регрессии:

$$\hat{y} = -3{,}54 + 0{,}854 \cdot 8 + 0{,}367 \cdot 6 = 5{,}49(\text{т}).$$

Для построения доверительного интервала для $M_x(Y)$ необходимо знать дисперсию его оценки — $s_{\hat{y}}^2$. Для ее вычисления обратимся к табл. 4.2 (точнее к ее двум последним столбцам, при составлении которых учтено, что групповые средние определяются по полученному уравнению регрессии).

Теперь по (**4.21**): $s^2 = \dfrac{6{,}329}{10-2-1} = 0{,}904$ и $s = \sqrt{0{,}904} = 0{,}951$ (т).

Определяем стандартную ошибку групповой средней $\hat{y}$ по формуле (**4.25**). Вначале найдем

$$X_0'\left(X'X\right)^{-1}X_0 = (1\ \ 8\ \ 6)\frac{1}{3738}\begin{pmatrix} 15027 & -1209 & -522 \\ -1209 & 201 & -108 \\ -522 & -108 & 244 \end{pmatrix}\begin{pmatrix} 1 \\ 8 \\ 6 \end{pmatrix} =$$

$$= \frac{1}{3738}\begin{pmatrix} 2223 & -249 & 78 \end{pmatrix}\begin{pmatrix} 1 \\ 8 \\ 6 \end{pmatrix} = \frac{1}{3738}\left(699\right) = 0{,}1870.$$

Теперь $s_{\hat{y}} = 0{,}951\sqrt{0{,}1870} = 0{,}411$ (т).

По табл. II приложений при числе степеней свободы $k = 10-2-1=7$ находим $t_{0{,}95;7}=2{,}36$. По (**4.24**) доверительный интервал для $M_x(Y)$ равен

$$5{,}49-2{,}36 \cdot 0{,}411 \le M_x(Y) \le 5{,}49 +2{,}36 \cdot 0{,}411,$$

или

$$4{,}52 \le M_x(Y) \le 6{,}46 \text{ (т)}.$$

Итак, с надежностью 0,95 средняя сменная добыча угля на одного рабочего для шахт с мощностью пласта 8 м и уровнем механизации работ 6% находится в пределах от 4,52 до 6,46 т.

Сравнивая новый доверительный интервал для функции регрессии $M_x(Y)$, полученный с учетом двух объясняющих переменных, с аналогичным интервалом с учетом одной объясняющей переменной (см. пример **3.3**), можно заметить уменьшение его величины. Это связано о тем, что включение в модель новой объясняющей переменной позволяет несколько повысить точность модели за счет увеличения взаимосвязи зависимой и объясняющей переменных (см. ниже).

Найдем доверительный интервал для индивидуального значения y_0^* при $X_0' = \left(1\ \ 8\ \ 6\right)$:

по (**4.27**): $s_{\hat{y}_0} = 0{,}951\sqrt{1+0{,}1870} = 1{,}036$ (т)

и по (**4.26**): $5{,}49-2{,}36 \cdot 1{,}036 \le y_0^* \le 5{,}49+2{,}36 \cdot 1{,}036,$

т. е. $3{,}05 \le y_0^* \le 7{,}93$ (т).

Итак, с надежностью 0,95 индивидуальное значение сменной добычи угля в шахтах с мощностью пласта 8 м и уровнем механизации работ 6% находится в пределах от 3,05 до 7,93 (т).

Проверим значимость коэффициентов регрессии b_1 и b_2. В примере **4.1** получены b_1= 0,854 и b_2=0,367. Стандартная ошибка s_{b_1} в соответствии с (**4.22**) равна

$$s_{b_1} = 0{,}951\sqrt{\frac{1}{3738}\cdot 201} = 0{,}221.$$

Так как $t = \dfrac{0{,}854}{0{,}221} = 3{,}81 > t_{0{,}95;7} = 2{,}36$, то коэффициент b_1 значим.

Аналогично вычисляем $s_{b_2} = 0{,}951\sqrt{\dfrac{1}{3738}\cdot 244} = 0{,}243$ и $t = \dfrac{0{,}367}{0{,}243} =$

$= 1{,}51 < t_{0{,}95;7} = 2{,}36$, т. е. коэффициент b_2 незначим на 5%-ном уровне.

Доверительный интервал имеет смысл построить только для значимого коэффициента регрессии b_1: по (**4.23′**):

$0{,}854 - 2{,}36 \cdot 0{,}221 \le \beta_1 \le 0{,}854 + 2{,}36 \cdot 0{,}221$, или $0{,}332 \le \beta_1 \le 1{,}376$.

Итак, с надежностью 0,95 за счет изменения на 1 м мощности пласта X_1 (при неизменном X_2) сменная добыча угля на одного рабочего Y будет изменяться в пределах от 0,332 до 1,376 (т).

Найдем 95%-ный доверительный интервал для параметра σ^2. Учитывая, что $\alpha = 1 - 0{,}95 = 0{,}05$, найдем по таблице III приложений $n - p - 1 = n - 2 - 1 = n - 3$ степенях свободы

$$\chi^2_{\alpha/2;n-p-1} = \chi^2_{0{,}025;7} = 16{,}01;$$

$$\chi^2_{1-\alpha/2;n-p-1} = \chi^2_{0{,}975;7} = 1{,}69.$$

По формуле (**4.28**) $\dfrac{10 \cdot 0{,}904}{16{,}01} \le \sigma^2 \le \dfrac{10 \cdot 0{,}904}{1{,}69}$,

или $0{,}565 \le \sigma^2 \le 5{,}349$ и $0{,}751 \le \sigma \le 2{,}313$.

Таким образом, с надежностью 0,95 дисперсия возмущений заключена в пределах от 0,565 до 5,349, а их стандартное отклонение — от 0,751 до 2,313 (т). ▶

Формально переменные, имеющие незначимые коэффициенты регрессии, могут быть исключены из рассмотрения. В экономических исследованиях исключению переменных из регрес-

сии должен предшествовать тщательный *качественный* анализ. Поэтому может оказаться целесообразным все же оставить в регрессионной модели одну или несколько объясняющих переменных, не оказывающих существенного (значимого) влияния на зависимую переменную.

4.6. Оценка значимости множественной регрессии. Коэффициенты детерминации R^2 и $\hat{R}^2$

Как и в случае парной регрессионной модели (см § **3.6**), в модели множественной регрессии общая вариация Q — сумма квадратов отклонений зависимой переменной от средней (**3.41**) может быть разложена на две составляющие:

$$Q = Q_R + Q_e,$$

где Q_R, Q_e — соответственно сумма квадратов отклонений, обусловленная регрессией, и остаточная сумма квадратов, характеризующая влияние неучтенных факторов.

Получим более удобные, чем (**3.40**), формулы для сумм квадратов Q, Q_R и Q_e, не требующие вычисления значений $\hat{y}_i$, обусловленных регрессией, и остатков e_i.

В соответствии с (**3.40**), (**3.42**)

$$Q = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} = Y'Y - n\bar{y}^2 \qquad \textbf{(4.29)}$$

(ибо $$\sum_{i=1}^{n} y_i^2 = y_1^2 + y_2^2 + \ldots + y_n^2 = \begin{pmatrix} y_1 & y_2 \ldots y_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{pmatrix} = Y'Y).$$

С учетом (**4.4**) имеем

$$Q_e = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = Y'Y - 2b'X'Y + b'X'Xb = Y'Y - b'X'Y \quad \textbf{(4.30)}$$

(ибо в силу (**4.5**) $b'X'Xb = b'X'Y$).

Наконец,

$$Q_R = Q - Q_e = Y'Y - n\bar{y}^2 - \left(Y'Y - b'X'Y\right) = b'X'Y - n\bar{y}^2. \quad \textbf{(4.31)}$$

Уравнение множественной регрессии значимо (иначе — гипотеза H_0 о равенстве нулю параметров регрессионной модели, т. е. H_0: $\beta_1 = \beta_2 = ... = \beta_p = 0$, отвергается), если (учитывая (**3.43**) при $m = p + 1$)

$$F = \frac{Q_R(n-p-1)}{Q_e p} > F_{\alpha;p;n-p-1}, \quad \textbf{(4.32)}$$

где $F_{\alpha;p;n-p-1}$ — табличное значение F-критерия Фишера—Снедекора, а Q_R и Q_e определяются по формулам (**4.31**) и (**4.30**).

В § **3.6** был введен ***коэффициент детерминации*** R^2 *как одна из наиболее эффективных оценок адекватности регрессионной модели, мера качества уравнения регрессии, характеристика его прогностической силы.*

Коэффициент детерминации (или *множественный коэффициент детерминации*) R^2 определяется по формуле (**3.47**) или с учетом (**4.31**), (**4.29**):

$$R^2 = \frac{Q_R}{Q} = \frac{b'X'Y' - n\bar{y}^2}{Y'Y - n\bar{y}^2}. \quad \textbf{(4.33)}$$

Отметим еще одну формулу для коэффициента детерминации:

$$R^2 = 1 - \frac{Q_e}{Q} = 1 - \frac{\left(Y - Xb\right)'\left(Y - Xb\right)}{\left(Y - \bar{Y}\right)'\left(Y - \bar{Y}\right)}, \quad \textbf{(4.33′)}$$

или

$$R^2 = 1 - \frac{e'e}{y'y}, \quad \textbf{(4.33″)}$$

где $e = Y - Xb$, $\bar{Y} = \left(\bar{y}, \bar{y}, ..., \bar{y}\right)$, $y = \left(Y - \bar{Y}\right)$ — n-мерные векторы;

$$e'e = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2,$$

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

Напомним, что R^2 характеризует *долю вариации зависимой переменной, обусловленной регрессией или изменчивостью объясняющих переменных; чем ближе* R^2 *к единице, тем лучше регрессия*

описывает зависимость между объясняющими и зависимой переменными.

Вместе с тем использование только одного коэффициента детерминации R^2 для выбора наилучшего уравнения регрессии может оказаться недостаточным. На практике встречаются случаи, когда плохо определенная модель регрессии может дать сравнительно высокий коэффициент R^2.

Недостатком коэффициента детерминации R^2 является то, что он, вообще говоря, увеличивается при добавлении новых объясняющих переменных, хотя это и не обязательно означает улучшение качества регрессионной модели. В этом смысле предпочтительнее использовать *скорректированный* (*адаптированный, поправленный* (*adjusted*)) *коэффициент детерминации* $\hat{R}^2$, определяемый по формуле

$$\hat{R}^2 = 1 - \frac{n-1}{n-p-1}\left(1 - R^2\right), \qquad \textbf{(4.34)}$$

или с учетом (**4.33**″)

$$\hat{R}^2 = 1 - \frac{\left(n-1\right)e'e}{\left(n-p-1\right)y'y}. \qquad \textbf{(4.34}'\textbf{)}$$

Из (**4.34**) следует, что чем больше число объясняющих переменных p, тем меньше $\hat{R}^2$ по сравнению с R^2. В отличие от R^2 скорректированный коэффициент $\hat{R}^2$ может уменьшаться при введении в модель новых объясняющих переменных, не оказывающих существенного влияния на зависимую переменную. Однако даже увеличение скорректированного коэффициента детерминации $\hat{R}^2$ при введении в модель новой объясняющей переменной не всегда означает, что ее коэффициент регрессии значим (это происходит, как можно показать, только в случае, если соответствующее значение t-статистики больше единицы (по абсолютной величине), т. е. $|t|>1$. Другими словами, увеличение $\hat{R}^2$ еще не означает улучшения качества регрессионной модели.

Если известен коэффициент детерминации R^2, то критерий значимости (**4.32**) уравнения регрессии может быть записан в виде:

$$F = \frac{R^2\left(n-p-1\right)}{\left(1-R^2\right)p} > F_{\alpha;k_1;k_2}, \qquad \textbf{(4.35)}$$

где $k_1=p$, $k_2=n-p-1$, ибо в уравнении множественной регрессии вместе со свободным членом оценивается $m = p+1$ параметров.

▶ **Пример 4.4.** По данным примера **4.1** определить множественный коэффициент детерминации и проверить значимость полученного уравнения регрессии Y по X_1 и X_2 на уровне $\alpha = 0{,}05$.

Р е ш е н и е. Вычислим произведения векторов (см. пример **4.1**):

$$b'X'Y = \begin{pmatrix} -3{,}54 & 0{,}854 & 0{,}367 \end{pmatrix} \begin{pmatrix} 68 \\ 664 \\ 445 \end{pmatrix} =$$

$$= -3{,}54 \cdot 68 + 0{,}854 \cdot 664 + 0{,}367 \cdot 445 = 489{,}65$$

и $Y'Y = \sum_{i=1}^{10} y_i = 496$ (см. итоговую строку табл. 4.2). Из табл. 4.2 находим также $\sum_{i=1}^{10} y_i = 68$, откуда $\bar{y} = \sum_{i=1}^{n} y_i / n = 68/10 = 6{,}8$ (т).

Теперь по (**4.33**) множественный коэффициент детерминации

$$R^2 = \frac{489{,}65 - 10 \cdot 6{,}8^2}{496 - 10 \cdot 6{,}8^2} = 0{,}811.$$

Коэффициент детерминации R^2=0,811 свидетельствует о том, что вариация исследуемой зависимой переменной Y — сменной добычи угля на одного рабочего на 81,1% объясняется изменчивостью включенных в модель объясняющих переменных — мощности пласта X_1 и уровня механизации работ X_2.

Проделав аналогичные расчеты по данным примера **3.1** для одной объясняющей переменной X_1, можно было получить R'^2=0,751 (заметим, что в случае одной объясняющей переменной коэффициент детерминации R'^2 равен квадрату парного коэффициента корреляции r^2). Сравнивая значения R^2 и R'^2, можно сказать, что добавление второй объясняющей переменной X_2 незначительно увеличило величину коэффициента детерминации, определяющего качество модели. И это понятно, так как выше, в примере **4.3**, мы убедились в незначимости коэффициента регрессии b_2 при переменной X_2.

По формуле (**4.34**) вычислим скорректированный коэффициент детерминации:

при $p = 1$ $\quad \hat{R}'^2 = 1 - \frac{9}{8}(1 - 0{,}751) = 0{,}720;$

при $p = 2$ $\quad \hat{R}'^2 = 1 - \frac{9}{7}(1 - 0{,}811) = 0{,}757.$

Видим, что хотя скорректированный коэффициент детерминации и увеличился при добавлении объясняющей переменной X_2, но это еще не говорит о значимости коэффициента b_2 (значение t-статистики, равное 1,51 (см. § **4.4**), хотя и больше 1, но недостаточно для соответствующего вывода на приемлемом уровне значимости).

Зная R^2=0,811, проверим значимость уравнения регрессии. Фактическое значение критерия по (**4.35**):

$$F = \frac{0{,}811(10-2-1)}{(1-0{,}811)\cdot 2} = 15{,}0$$

больше табличного $F_{0,05;2;7}$=4,74, определенного на уровне значимости α=0,05 при k_1=2 и k_2=10−2−1=7 степенях свободы (см. табл. IV приложений), т. е. уравнение регрессии значимо, следовательно, исследуемая зависимая переменная Y достаточно хорошо описывается включенными в регрессионную модель переменными X_1 и X_2. ▶

Упражнения

4.5. Имеются следующие данные о выработке литья на одного работающего X_1 (т), браке литья X_2 (%) и себестоимости 1 т литья Y (руб.) по 25 литейным цехам заводов:

i	x_{1j}	x_{2i}	y_i	i	x_{1i}	x_{2i}	y_i	i	x_{1i}	x_{2i}	y_i
1	14,6	4,2	239	10	25,3	0,9	198	19	17,0	9,3	282
2	13,5	6,7	254	11	56,0	1,3	170	20	33,1	3,3	196
3	21,5	5,5	262	12	40,2	1,8	173	21	30,1	3,5	186
4	17,4	7,7	251	13	40,6	3,3	197	22	65,2	1,0	176
5	44,8	1,2	158	14	75,8	3,4	172	23	22,6	5,2	238
6	111,9	2,2	101	15	27,6	1,1	201	24	33,4	2,3	204
7	20,1	8,4	259	16	88,4	0,1	130	25	19,7	2,7	205
8	28,1	1,4	186	17	16,6	4,1	251				
9	22,3	4,2	204	18	33,4	2,3	195				

Необходимо: *а*) найти множественный коэффициент детерминации и пояснить его смысл; *б*) найти уравнение множественной регрессии Y по X_1 и X_2, оценить значимость этого уравнения и его коэффициентов на уровне α=0,05; *в*) сравнить раздельное влияние на зависимую переменную каждой из объясняющих переменных, используя стандартизованные коэффициенты регрессии и коэффициенты эластичности; *г*) найти 95%-ные доверительные интервалы для коэффициентов регрессии, а также для среднего и индивидуальных значений себестоимости 1 т литья в цехах, в которых выработка литья на одного работающего составляет 40 т, а брак литья — 5%.

4.6. Имеются следующие данные о годовых ставках месячных доходов по трем акциям за шестимесячный период:

Акция	*Доходы по месяцам, %*					
A	5,4	5,3	4,9	4,9	5,4	6,0
B	6,3	6,2	6,1	5,8	5,7	5,7
C	9,2	9,2	9,1	9,0	8,7	8,6

Есть основания предполагать, что доходы Y по акции C зависят от доходов X_1 и X_2 по акциям A и B. Необходимо: *а*) составить уравнение регрессии Y по X_1 и X_2; *б*) найти множественный коэффициент детерминации R^2 и пояснить его смысл; *в*) проверить значимость полученного уравнения регрессии на уровне α=0,05; *г*) оценить средний доход по акции C, если доходы по акциям A и B составили соответственно 5,5 и 6,0%.

Глава 5

Некоторые вопросы практического использования регрессионных моделей

В предыдущих главах была изучена классическая линейная модель регрессии, приведена оценка параметров модели и проверка статистических гипотез о регрессии. Однако мы не касались некоторых проблем, связанных с практическим использованием модели множественной регрессии. К их числу относятся: мультиколлинеарность, ее причины и методы устранения; использование фиктивных переменных при включении в регрессионную модель качественных объясняющих переменных, линеаризация модели, вопросы частной корреляции между переменными. Изучению указанных проблем посвящена данная глава.

5.1. Мультиколлинеарность

Под ***мультиколлинеарностью*** *понимается высокая взаимная коррелированность объясняющих переменных.* Мультиколлинеарность может проявляться в функциональной (явной) и стохастической (скрытой) формах.

При *функциональной форме* мультиколлинеарности по крайней мере одна из парных связей между объясняющими переменными является линейной функциональной зависимостью. В этом случае матрица $X'X$ особенная, так как содержит линейно зависимые векторы-столбцы и ее определитель равен нулю, т. е. нарушается предпосылка *6* регрессионного анализа. Это приводит к невозможности решения соответствующей системы нормальных уравнений и получения оценок параметров регрессионной модели.

Однако в экономических исследованиях мультиколлинеарность чаще проявляется в *стохастической форме*, когда между хотя бы двумя объясняющими переменными существует тесная корреляционная связь. Матрица $X'X$ в этом случае является неособенной, но ее определитель очень мал.

В то же время вектор оценок b и его ковариационная матрица $\sum_b$ в соответствии с формулами (**4.8**) и (**4.16**) пропорциональны обратной матрице $(X'X)^{-1}$, а значит, их элементы обратно пропорциональны величине определителя $|X'X|$. В результате получаются значительные средние квадратические отклонения (стандартные ошибки) коэффициентов регрессии b_0, b_1,..., b_p и оценка их значимости по t-критерию не имеет смысла, хотя в целом регрессионная модель может оказаться значимой по F-критерию.

Оценки становятся очень чувствительными к незначительному изменению результатов наблюдений и объема выборки. Уравнения регрессии в этом случае, как правило, не имеют реального смысла, так как некоторые из его коэффициентов могут иметь неправильные с точки зрения экономической теории знаки и неоправданно большие значения.

Точных количественных критериев для определения наличия или отсутствия мультиколлинеарности не существует. Тем не менее имеются некоторые эвристические подходы по ее выявлению.

Один из таких подходов заключается в *анализе корреляционной матрицы между объясняющими переменными* X_1, X_2,..., X_p и выявлении пар переменных, имеющих высокие коэффициенты корреляции (обычно больше 0,8). Если такие переменные существуют, то говорят о мультиколлинеарности между ними.

Полезно также находить *множественные коэффициенты детерминации* между одной из объясняющих переменных и некоторой группой из них. Наличие высокого множественного коэффициента детерминации (обычно больше 0,6) свидетельствует о мультиколлинеарности.

Другой подход состоит в исследовании матрицы $X'X$. *Если определитель матрицы* $X'X$ *либо ее минимальное собственное значение* $\lambda_{\min}$ *близки к нулю* (например, одного порядка с накапливающимися ошибками вычислений), *то это говорит о наличии мультиколлинеарности.* О том же может свидетельствовать и значительное отклонение максимального собственного значения $\lambda_{\max}$ матрицы $X'X$ от ее минимального собственного значения $\lambda_{\min}$.

Для устранения или уменьшения мультиколлинеарности используется ряд методов. Самый простой из них (но далеко не всегда возможный) состоит в том, что *из двух объясняющих переменных, имеющих высокий коэффициент корреляции* (больше 0,8), *одну переменную исключают из рассмотрения.* При этом, какую переменную оставить, а какую удалить из анализа, решают в первую

очередь на основании экономических соображений. Если с экономической точки зрения ни одной из переменных нельзя отдать предпочтение, то оставляют ту из двух переменных, которая имеет больший коэффициент корреляции с зависимой переменной.

Другой метод устранения или уменьшения мультиколлинеарности *заключается в переходе от несмещенных оценок,* определенных по методу наименьших квадратов, *к с м е щ е н н ы м оценкам, обладающим,* однако, *меньшим рассеянием относительно оцениваемого параметра*, т. е. меньшим математическим ожиданием квадрата отклонения оценки b_j от параметра β_j или $M\,(b_j - \beta_j)^2$.

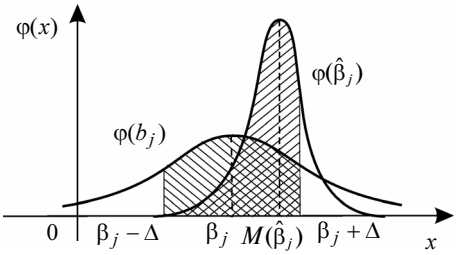


Рис. 5.1

Оценки, определяемые вектором (**4.8**), обладают в соответствии с теоремой Гаусса—Маркова минимальными дисперсиями в классе всех линейных *несмещенных* оценок, но при наличии мультиколлинеарности эти дисперсии могут оказаться слишком большими, и обращение к соответствующим *смещенным* оценкам может повысить точность оценивания параметров регрессии. На рис. 5.1 показан случай, когда смещенная оценка $\hat{\beta}_j$, выборочное распределение которой задается плотностью $\varphi(\hat{\beta}_j)$, «лучше» несмещенной оценки b_j, распределение которой представляет плотность $\varphi(b_j)$.

Действительно, пусть максимально допустимый по величине доверительный интервал для оцениваемого параметра β_j есть $(\beta_j - \Delta, \beta_j + \Delta)$. Тогда доверительная вероятность, или надежность оценки, определяемая площадью под кривой распределения на интервале $(\beta_j - \Delta, \beta_j + \Delta)$, как нетрудно видеть из рис. 5.1, будет в данном случае больше для оценки $\hat{\beta}_j$ по сравнению с b_j (на рис. 5.1 эти площади заштрихованы). Соответственно средний квадрат отклонения оценки от оцениваемого параметра будет меньше для смещенной оценки, т. е.

$$M\left(\hat{\beta}_j - \beta_j\right)^2 < M\left(b_j - \beta_j\right)^2.$$

При использовании «*ридж-регрессии*» (или «*гребневой регрессии*») вместо несмещенных оценок (**4.8**) рассматривают смещенные оценки, задаваемые вектором $\hat{\beta}_\tau = \left(X'X + \tau E_{p+1}\right)^{-1} X'Y$, где τ — некоторое положительное число, называемое «гребнем» или «хребтом», E_{p+1} — единичная матрица $(p+1)$-го порядка. *Добавление* τ к диагональным элементам матрицы $X'X$ *делает оценки параметров модели смещенными, но при этом увеличивается определитель матрицы* системы нормальных уравнений (**4.5**) — вместо $(X'X)$ он будет равен $\left|X'X + \tau E_{p+1}\right|$.

Таким образом, становится возможным исключение мультиколлинеарности в случае, когда определитель $\left|X'X\right|$ близок к нулю.

Для устранения мультиколлинеарности может быть использован *переход от исходных объясняющих переменных* $X_1, X_2, \ldots, X_n$, связанных между собой достаточно тесной корреляционной зависимостью, *к новым переменным*, представляющим линейные комбинации исходных. При этом новые переменные должны быть слабокоррелированными либо вообще некоррелированными. В качестве таких переменных берут, например, так называемые *главные компоненты вектора исходных объясняющих переменных*, изучаемые в компонентном анализе, и рассматривают регрессию на главных компонентах, в которой последние выступают в качестве обобщенных объясняющих переменных, подлежащих в дальнейшем содержательной (экономической) интерпретации.

Ортогональность главных компонент предотвращает проявление эффекта мультиколлинеарности. Кроме того, применяемый метод позволяет ограничиться малым числом главных компонент при сравнительно большом количестве исходных объясняющих переменных.

5.2. Отбор наиболее существенных объясняющих переменных в регрессионной модели

Еще одним из возможных методов устранения или уменьшения мультиколлинеарности является *использование пошаговых процедур отбора наиболее информативных переменных.* Например, на первом шаге рассматривается лишь одна объясняющая пере-

менная, имеющая с зависимой переменной Y наибольший коэффициент детерминации. На втором шаге включается в регрессию новая объясняющая переменная, которая вместе с первоначально отобранной образует пару объясняющих переменных, имеющую с Y наиболее высокий (скорректированный) коэффициент детерминации. На третьем шаге вводится в регрессию еще одна объясняющая переменная, которая вместе с двумя первоначально отобранными образует тройку объясняющих переменных, имеющую с Y наибольший (скорректированный) коэффициент детерминации, и т. д.

Процедура введения новых переменных продолжается до тех пор, пока будет увеличиваться соответствующий (скорректированный) коэффициент детерминации $\hat{R}^2$ (более точно — *минимальное* значение $\hat{R}^2_{\min}$).

▶ **Пример 5.1.** По данным $n = 20$ сельскохозяйственных районов области исследуется зависимость переменной Y — урожайности зерновых культур (в ц/га) от ряда переменных — факторов сельскохозяйственного производства:

X_1 — число тракторов (приведенной мощности на 100 га);
X_2 — число зерноуборочных комбайнов на 100 га;
X_3 — число орудий поверхностной обработки почвы на 100 га;
X_4 — количество удобрений, расходуемых на 1 га (т/га);
X_5 — количество химических средств защиты растений, расходуемых на 1 га (ц/га).

Исходные данные[1] приведены в табл. 5.1.

Таблица 5.1

i (*номер района*)	y_i	x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}
1	9,70	1,59	0,26	2,05	0,32	0,14
2	8,40	0,34	0,28	0,46	0,59	0,66
.....						
19	13,10	0,08	0,25	0,03	0,73	0,20
20	8,70	1,36	0,26	0,17	0,99	0,42

В случае обнаружения мультиколлинеарности принять меры по ее устранению (уменьшению), используя пошаговую процедуру отбора наиболее информативных переменных.

[1] Пример заимствован из [2]. Там же на с. 632 приведены полностью исходные данные.

Р е ш е н и е. По формуле (**4.8**) найдем вектор оценок параметров регрессионной модели $b = (3{,}515;\ -0{,}006;\ 15{,}542;\ 60{,}110;\ 4{,}475;\ -2{,}932)'$, так что в соответствии с (**4.9**) выборочное уравнение множественной регрессии имеет вид:

$$\hat{y} = 3{,}515 - 0{,}006X_1 + 15{,}542X_2 + 0{,}110X_3 + 4{,}475X_4 - 2{,}932X_5.$$
$$(5{,}41)\quad (0{,}60)\quad (21{,}59)\quad (0{,}85)\quad (1{,}54)\quad (3{,}09)$$

В скобках указаны средние квадратические отклонения (стандартные ошибки) s_{b_j} коэффициентов регрессии b_j, вычисленные по формуле (**4.22**). Сравнивая значения *t*-статистики (по абсолютной величине) каждого коэффициента регрессии β_j по формуле $t_{b_j} = \dfrac{b_j}{s_{b_j}}$ ($j = 0,1,2,3,4,5$), т. е. $t_{b_0} = 0{,}65$; $t_{b_1} = -0{,}01$; $t_{b_2} = 0{,}72$; $t_{b_3} = 0{,}13$; $t_{b_4} = 2{,}91$; $t_{b_5} = -0{,}95$ с критическим значением $t_{0{,}95;14} = 2{,}14$, определенным по табл. II приложений на уровне значимости α=0,05 при числе степеней свободы $k = n - p - 1 = 20 - 5 - 1 = 14$, мы видим, что значимым оказался только коэффициент регрессии b_4 при переменной X_4 — количество удобрений, расходуемых на гектар земли.

Вычисленный по (**4.33**) множественный коэффициент детерминации урожайности зерновых культур Y по совокупности пяти факторов (X_1—X_5) сельскохозяйственного производства оказался равным $R^2_{y.12345}$=0,517, т. е. 51,7% вариации зависимой переменной объясняется включенными в модель пятью объясняющими переменными. Так как вычисленное по (**4.35**) фактическое значение F=3,00 больше табличного $F_{0{,}05;5;14}$=2,96, то уравнение регрессии значимо по F-критерию на уровне α=0,05.

По формуле (**3.20**) была рассчитана матрица парных коэффициентов корреляции:

Переменные	Y	X_1	X_2	X_3	X_4	X_5
Y	1,00	0,43	0,37	0,40	0,58*	0,33
X_1	0,43	1,00	0,85*	0,98*	0,11	0,34
X_2	0,37	0,85*	1,00	0,88*	0,03	0,46*
X_3	0,40	0,98*	0,88*	1,00	0,03	0,28
X_4	0,58*	0,11	0,03	0,03	1,00	0,57*
X_5	0,33	0,34	0,46*	0,28	0,57*	1,00

Знаком* отмечены коэффициенты корреляции, значимые по *t*-критерию (**3.46**) на 5%-ном уровне.

Анализируя матрицу парных коэффициентов корреляции, можно отметить тесную корреляционную связь между переменными X_1 и X_2 (r_{12}= 0,85), X_1 и X_3 (r_{13} = 0,98), X_2 и X_3 (r_{23} = 0,88), что, очевидно, свидетельствует о мультиколлинеарности объясняющих переменных.

Для устранения мультиколлинеарности применим процедуру пошагового отбора наиболее информативных переменных.

1-й шаг. Из объясняющих переменных $X_1 - X_5$ выделяется переменная X_4, имеющая с зависимой переменной Y наибольший коэффициент детерминации $R^2_{y\cdot j}$ (равный для парной модели квадрату коэффициента корреляции r^2_{yj}). Очевидно, это переменная X_4, так как коэффициент детерминации $R^2_{y\cdot 4} = r^2_{y4} = 0{,}58^2 = 0{,}336$ — максимальный. С учетом поправки на несмещенность по формуле (**4.34**) скорректированный коэффициент детерминации $\hat{R}^2_{y\cdot 4} = 1 - \frac{19}{18}(1 - 0{,}336) = 0{,}299$.

2-й шаг. Среди всевозможных пар объясняющих переменных X_4, X_j, j =1,2,3,5, выбирается пара (X_4, X_3), имеющая с зависимой переменной Y наиболее высокий коэффициент детерминации $R^2_{y\cdot 4j} = R^2_{y\cdot 43} = 0{,}483$ и с учетом поправки по (**4.34**) $\hat{R}^2_{y\cdot 43} = 1 - \frac{19}{17}(1 - 0{,}483) = 0{,}422$.

3-й шаг. Среди всевозможных троек объясняющих переменных (X_4, X_3, X_j), j = 1,2,5, наиболее информативной оказалась тройка (X_4, X_3, X_5), имеющая максимальный коэффициент детерминации $R^2_{y.43j} = R^2_{y\cdot 435} = 0{,}513$ и соответственно скорректированный коэффициент $R^2_{y\cdot 435} = 0{,}422$.

Так как скорректированный коэффициент детерминации на 3-м шаге не увеличился, то в регрессионной модели достаточно ограничиться лишь двумя отобранными ранее объясняющими переменными X_4 и X_3.

Рассчитанное по формулам (**4.8**), (**4.9**) уравнение регрессии по этим переменным примет вид:

$$\hat{y} = 7{,}29 + 3{,}48X_3 + 3{,}48X_4.$$
$$(0{,}66)\ (0{,}13)\ \ (1{,}07)$$

Нетрудно убедиться в том, что теперь все коэффициенты регрессии значимы, так как каждое из значений t-статистики

$$t_{b_0} = \frac{7{,}29}{0{,}66} = 11{,}0\,; \quad t_{b_3} = \frac{3{,}48}{0{,}13} = 26{,}8\,; \quad t_{b_4} = \frac{3{,}48}{1{,}07} = 3{,}25$$

больше соответствующего табличного значения $t_{0,95;17} = 2{,}11$.

З а м е ч а н и е. Так как значения коэффициентов корреляции весьма высокие (больше 0,8): r_{12}=0,85, r_{13}=0,98, r_{23}=0,88, то, очевидно, из соответствующих трех переменных X_1, X_2, X_3 две переменные можно было сразу исключить из регрессии и без проведения пошагового отбора, но какие именно переменные исключить — следовало решать, исходя из качественных соображений, основанных на знании предметной области (в данном случае влияния на урожайность факторов сельскохозяйственного производства).▶

Кроме рассмотренной выше пошаговой процедуры *присоединения* объясняющих переменных используются также пошаговые процедуры *присоединения—удаления* и процедура *удаления* объясняющих переменных, изложенные, например, в [2]. Следует отметить, что какая бы пошаговая процедура ни использовалась, она не гарантирует определения оптимального (в смысле получения максимального коэффициента детерминации $\hat{R}^2$) набора объясняющих переменных. Однако в большинстве случаев получаемые с помощью пошаговых процедур наборы переменных оказываются оптимальными или близкими к оптимальным.

5.3. Линейные регрессионные модели с переменной структурой. Фиктивные переменные

До сих пор мы рассматривали регрессионную модель, в которой в качестве объясняющих переменных (регрессоров) выступали количественные переменные (производительность труда, себестоимость продукции, доход и т. п.). Однако на практике достаточно часто возникает необходимость исследования влияния качественных признаков, имеющих два или несколько уровней (градаций). К числу таких признаков можно отнести: пол (мужской, женский), образование (начальное, среднее, высшее), фактор сезонности (зима, весна, лето, осень) и т. п.

Качественные признаки могут существенно влиять на структуру линейных связей между переменными и приводить к скачкообразному изменению параметров регрессионной модели. В этом случае говорят об *исследовании регрессионных моделей с переменной структурой* или *построении регрессионных моделей по неоднородным данным*.

Например, нам надо изучить зависимость размера заработной платы Y работников не только от количественных факторов X_1, X_2,..., X_n, но и от качественного признака Z_1 (например, фактора «пол работника»).

В принципе можно было получить оценки регрессионной модели

$$y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip} + \varepsilon_i, \ i = 1,..., n \qquad \textbf{(5.1)}$$

для каждого уровня качественного признака (т. е. выборочное уравнение регрессии отдельно для работников-мужчин и отдельно — для женщин), а затем изучать различия между ними (см. § **5.4**).

Но есть и другой подход, позволяющий оценивать влияние значений количественных переменных и уровней качественных признаков с помощью *одного* уравнения регрессии. Этот подход связан с введением так называемых *фиктивных* (*манекенных*) *переменных*[1], или *манекенов* (*dummy variables*).

В качестве фиктивных переменных обычно используются *дихотомические* (*бинарные, булевы*) переменные, которые принимают всего два значения: «0» или «1» (например, значение такой переменной Z_1 по фактору «пол»: $Z_1 = 0$ для работников-женщин и $Z_1 = 1$ — для мужчин).

В этом случае первоначальная регрессионная модель (**5.1**) заработной платы изменится и примет вид:

$$y_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip} + \alpha_1 z_{i1} + \varepsilon_i, \ i = 1,..., n, \qquad \textbf{(5.2)}$$

где $z_{i1} = \begin{cases} 1, & \text{если } i\text{-й работник мужского пола;} \\ 0 & \text{если } i\text{-й работник женского пола.} \end{cases}$

Таким образом, принимая модель (**5.2**), мы считаем, что средняя заработная плата у мужчин на $\alpha_1 \cdot 1 = \alpha_1$ выше, чем у женщин, при неизменных значениях других параметров модели. А проверяя гипотезу H_0: $\alpha_1 = 0$, мы можем установить сущест-

[1] В отечественной литературе используется также термин *структурные переменные*.

венность влияния фактора «пол» на размер заработной платы работника.

Следует отметить, что в принципе качественное различие можно формализовать с помощью любой переменной, принимающей два разных значения, не обязательно «0» или «1». Однако в эконометрической практике почти всегда используются фиктивные переменные типа «0—1», так как при этом интерпретация полученных результатов выглядит наиболее просто. Так, если бы в модели (**5.2**) в качестве фиктивной выбрали переменную Z_1, принимающую значения $z_{i1}=4$ (для работников-мужчин) и $z_{i2}=1$ (для женщин), то коэффициент регрессии α_1 при этой переменной равнялся бы $1/(4-1)$, т. е. одной трети среднего изменения заработной платы у мужчин.

Если рассматриваемый качественный признак имеет несколько (k) уровней (градаций), то в принципе можно было ввести в регрессионную модель дискретную переменную, принимающую такое же количество значений (например, при исследовании зависимости заработной платы Y от уровня образования Z можно рассматривать $k=3$ значения: $z_{i1}=1$ при наличии начального образования, $z_{i2}=2$ — среднего и $z_{i3}=3$ при наличии высшего образования). Однако обычно так не поступают из-за трудности содержательной интерпретации соответствующих коэффициентов регрессии, а вводят $(k-1)$ бинарных переменных.

В рассматриваемом примере для учета фактора образования можно было в регрессионную модель (**5.2**) ввести $k - 1=3-1=2$ бинарные переменные Z_{21} и Z_{22}:

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_p x_{ip} + \alpha_1 z_{i1} + \alpha_{21} z_{i21} + \alpha_{22} z_{i22} + \varepsilon_i , \qquad \textbf{(5.3)}$$

где $z_{i21}=\begin{cases}1, & \text{если } i\text{-й работник имеет высшее образование;}\\ 0 & \text{во всех остальных случаях;}\end{cases}$

$z_{i22}=\begin{cases}1, & \text{если } i\text{-й работник имеет среднее образование;}\\ 0 & \text{во всех остальных случаях.}\end{cases}$

Третьей бинарной переменной X_{23}, очевидно, не требуется: если i-й работник имеет начальное образование, это будет отражено парой значений $z_{i21} = 0$, $z_{i22} = 0$.

Более того, вводить третью бинарную переменную Z_{23} (со значениями $z_{i13} = 1$, если i-й работник имеет начальное образование; $z_{i23}=0$ — в остальных случаях) н е л ь з я, так как при

этом для любого *i*-го работника $z_{i21}+z_{i22}+z_{i23}=1$, т. е. при суммировании элементов столбцов общей матрицы плана, соответствующих фиктивным переменным Z_{21}, Z_{22}, Z_{23}, мы получили бы столбец, состоящий из одних единиц. А так как в матрице плана такой столбец из единиц уже есть (напомним (**§4.1**), что это первый столбец, соответствующий свободному члену уравнения регрессии), то это означало бы линейную зависимость значений (столбцов) общей матрицы плана *X*, т. е. нарушило бы предпосылку *6* регрессионного анализа. Таким образом, мы оказались бы в условиях мультиколлинеарности в функциональной форме (**§ 5.1**) и как следствие — невозможности получения оценок методом наименьших квадратов.

Такая ситуация, когда сумма значений нескольких переменных, включенных в регрессию, равна постоянному числу (единице), получила название *«dummy trap»* или «ловушки». Чтобы избежать такие ловушки, число вводимых бинарных переменных должно быть на единицу меньше числа уровней (градаций) качественного признака.

Следует отметить не совсем удачный перевод на русский язык термина «*dummy variables*» как «фиктивная» переменная. Во-первых, в модели регрессионного анализа мы уже имеем фиктивную переменную *X* при коэффициенте β_0, всегда равную единице. Во-вторых, и это главное — все процедуры регрессионного анализа (оценка параметров регрессионной модели, проверка значимости ее коэффициентов и т. п.) проводятся при включении фиктивных переменных так же, как и «обычных», количественных объясняющих переменных. «Фиктивность» же переменных Z_i состоит только в том, что они количественным образом описывают качественный признак.

Рассматриваемые выше регрессионные модели (**5.2**) и (**5.3**) отражали влияние качественного признака (фиктивных переменных) только на значения переменной *Y*, т. е. на *свободный член* уравнения регрессии. В более сложных моделях может быть отражена также зависимость фиктивных переменных на *сами параметры* при переменных регрессионной модели. Например, при наличии в модели объясняющих переменных — количественной X_1 и фиктивных Z_{11}, Z_{12}, Z_{21}, Z_{22}, из которых Z_{11}, Z_{12} влияют только на значение коэффициента при X_1, а Z_{21}, Z_{22} — только на величину свободного члена уравнения, такая регрессионная модель примет вид:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_{11}(z_{i11} x_{i1}) + \beta_{12}(z_{i12} x_{i1}) + \\ + \alpha_{21} z_{i21} + \alpha_{22} z_{i22} + \varepsilon_i, \quad i = 1, \ldots, n. \qquad \textbf{(5.4)}$$

Модели типа (**5.4**) используются, например, при исследовании зависимости объема потребления Y некоторого продукта от дохода потребителя X, когда одни качественные признаки (например, фактор сезонности) влияют лишь на количество потребляемого продукта (свободный член уравнения регрессии), а другие (например, уровень доходности домашнего хозяйства) — на параметр β_1 при X, интерпретируемый как «склонность к потреблению».

▶ **Пример 5.2.** Необходимо исследовать зависимость между результатами письменных вступительных и курсовых (на I курсе) экзаменов по математике. Получены следующие данные о числе решенных задач на вступительных экзаменах X (задание — 10 задач) и курсовых экзаменах Y (задание — 7 задач) 12 студентов, а также распределение этих студентов по фактору «пол»:

№ студента	*Число решенных задач*		*Пол студента*	*№ студента*	*Число решенных задач*		*Пол студента*
i	x_i	y_i		i	x_i	y_i	
1	10	6	муж.	7	6	3	жен.
2	6	4	жен.	8	7	4	муж.
3	8	4	муж.	9	9	7	муж.
4	8	5	жен.	10	6	3	жен.
5	6	4	жен.	11	5	2	муж.
6	7	7	муж.	12	7	3	жен.

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

Р е ш е н и е. Вначале рассчитаем уравнение парной регрессии Y по X, используя формулы (**3.7**) — (**3.15**):

$$\hat{y} = -1{,}437 + 0{,}815x\,. \qquad \textbf{(5.5)}$$

По формуле (**3.47**) коэффициент детерминации $R^2_{y.x} = 0{,}530$, т. е. 53,0% вариации зависимой переменной Y обусловлено рег-

рессией. Уравнение регрессии значимо по F-критерию, так как в соответствии с (**3.48**) F=9,46>$F_{0,05;1;10}$ = 4,96.

Однако полученное уравнение не учитывает влияние качественного признака — фактора «пол».

Для ее учета введем в регрессионную модель фиктивную (бинарную) переменную Z_1 ,

где $$z_{i1}=\begin{cases}1, & \text{если } i\text{-й студент мужского пола,}\\ 0 & \text{если } i\text{-й студент женского пола.}\end{cases}$$

Полагая, что фактор «пол» может сказаться только на числе решенных задач (свободном члене) регрессии, имеем модель[1] типа (**5.2**):

$$y_i = -1{,}165+0{,}743x+0{,}466x_1. \qquad \textbf{(5.6)}$$
$$(2{,}410)\ (0{,}053)\ (0{,}405)$$

Таким образом, получили регрессионную модель

$$Y = X\beta + \varepsilon \qquad \textbf{(5.7)}$$

с общей матрицей плана [2]

$$X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 10 & 6 & 8 & 8 & 6 & 7 & 6 & 7 & 9 & 6 & 5 & 7 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \end{pmatrix}'.$$

По формуле (**4.8**) найдем вектор оценок параметров регрессии

$$b=(-1{,}165;\ 0{,}743;\ 0{,}466)'.$$

Так что в соответствии с (**4.9**) выборочное уравнение множественной регрессии примет вид:

$$\hat{y} = -1{,}165+0{,}743x+0{,}466z_1.$$
$$(2{,}410)\ (0{,}053)\ (0{,}405) \qquad \textbf{(5.8)}$$

Коэффициент детерминации $R^2_{y.xz} = 0{,}549$.

Уравнение регрессии значимо по F-критерию на 5%-ном уровне, так как в соответствии с (**4.35**)

---

[1] Здесь и далее в скобках под коэффициентами регрессии указываются их средние квадратические (стандартные) отклонения.

[2] Общая матрица плана X включает все значения переменных, в том числе значения фиктивных переменных Z_i.

$$F = 5{,}48 > F_{0{,}05;2;9} = 4{,}26.$$

Из (**5.8**) следует, что при том же числе решенных задач на вступительных экзаменах X, на курсовых экзаменах юноши решают в среднем на $0{,}466 \approx 0{,}5$ задачи больше. На рис. 5.2 показаны линии регрессии Y по X для юношей (при z_1=1, т. е. $\hat{y} = -1{,}165 + 0{,}743x + 0{,}466 \cdot 1$ или $\hat{y} = -0{,}699 + 0{,}743x$) и для девушек (при z_1=0, т. е. $\hat{y} = -1{,}165 + 0{,}743x$).

Эти уравнения отличаются только свободным членом, а соответствующие линии регрессии параллельны (см. рис. 5.2). Полученное уравнение множественной регрессии (**5.8**) по-прежнему значимо по F-критерию. Однако коэффициент регрессии α_1 при фиктивной переменной Z_1 незначим по t-критерию (ибо $t = \dfrac{0{,}466}{0{,}405} = 1{,}15$ и $t < t_{0{,}95;9} = 2{,}26$), возможно, из-за недостаточного объема выборки либо в силу того, что гипотеза H_0: $\alpha_1 = 0$ верна).

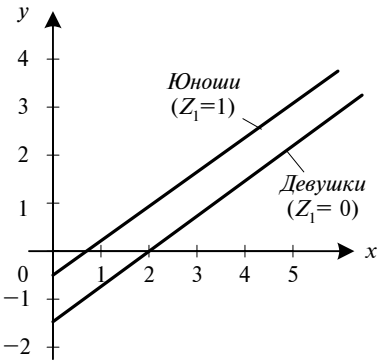


Рис. 5.2

В этом также легко убедиться, если вычислить по формуле (**4.34**) скорректированный коэффициент детерминации, который *уменьшился* от значения $\hat{R}^2_{y \cdot x} = 0{,}483$ для парной модели (**5.5**) до значения $R^2_{y \cdot xz} = 0{,}449$ для множественной модели (**5.6**). Следовательно, по имеющимся данным влияние фактора «пол» оказалось несущественным, и у нас есть основания считать, что регрессионная модель результатов курсовых экзаменов по математике в зависимости от вступительных одна и та же для юношей и девушек. ▶

З а м е ч а н и е. Если бы в регрессионной модели мы хотели учесть другие факторы с бóльшим, чем две, числом k_i градаций, то, как отмечено выше, следовало бы ввести в модель (k_i-1) бинарных переменных. Например, если было бы необходимо изучить влияние на результаты курсового экзамена фактора Z_2—«тип учебного заведения», оконченного студентом (школа, техникум, ПТУ), то в регрессионную модель (**5.6**) следовало ввести $k_i-1=3-1=2$ бинарные переменные Z_{21} и Z_{22}:

$$y_i = \beta_0 + \beta_1 x_{i1} + \alpha_1 z_{i1} + \alpha_{21} z_{i21} + \alpha_{22} z_{i22} + \varepsilon_i , \qquad \textbf{(5.9)}$$

где $z_{i21}=\begin{cases}1, & \text{если студент окончил школу;}\\ 0 & \text{в остальных случаях;}\end{cases}$

$z_{i22}=\begin{cases}1, & \text{если студент окончил техникум;}\\ 0 & \text{в остальных случаях.}\end{cases}$

Но при этом, конечно, следовало увеличить объем выборки n, так как надежность статистических выводов существенно зависит от отношения объема выборки n к общему числу всех параметров регрессионной модели: *чем больше величина отношения* $n/(p+1)$, *тем точнее соответствующие оценки, тем надежнее статистические выводы.*

5.4. Критерий Г. Чоу

В практике эконометриста нередки случаи, когда имеются две выборки пар значений зависимой и объясняющих переменных (x_i, y_i). Например, одна выборка пар значений переменных объемом n_1 получена при одних условиях, а другая, объемом n_2, — при несколько измененных условиях. Необходимо выяснить, действительно ли две выборки однородны в регрессионном смысле? Другими словами, можно ли *объединить* две выборки в одну и рассматривать единую модель регрессии Y по X?

При достаточных объемах выборок можно было, например, построить интервальные оценки параметров регрессии по каждой из выборок и в случае пересечения соответствующих доверительных интервалов сделать вывод о единой модели регрессии. Возможны и другие подходы.

В случае, если объем хотя бы одной из выборок незначителен, то возможности такого (и аналогичных) подходов резко су-

жаются из-за невозможности построения сколько-нибудь надежных оценок.

В критерии (*тесте*) *Г. Чоу* эти трудности в существенной степени преодолеваются. По каждой выборке строятся две линейные регрессионные модели:

$$y_i = \beta'_0 + \sum_{j=1}^{p} \beta'_j x_{ij} + \varepsilon'_i,\ i = 1,..., n_1;$$

$$y_i = \beta''_0 + \sum_{j=1}^{p} \beta''_j x_{ij} + \varepsilon''_i,\ i = n_1+1,..., n_1+n_2.$$

Проверяемая нулевая гипотеза имеет вид — H_0: $\beta'=\beta''$; $D(\varepsilon')=D(\varepsilon'')=\sigma^2$, где $\beta'=\beta''$ — векторы параметров двух моделей; $\varepsilon', \varepsilon''$ — их случайные возмущения.

Если нулевая гипотеза H_0 верна, то две регрессионные модели можно объединить в одну объема $n = n_1+n_2$:

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i,\ \ i = 1,2..., n.$$

Согласно критерию Г. Чоу нулевая гипотеза H_0 отвергается на уровне значимости α, если статистика

$$F = \frac{\left(\sum_{i=1}^{n} e_i^2 - \sum_{i=1}^{n_1} e_i^2 - \sum_{i=n_1}^{n} e_i^2\right)(n-2p-2)}{\left(\sum_{i=1}^{n_1} e_i^2 + \sum_{i=n_1}^{n} e_i^2\right)(p+1)} > F_{\alpha;p+1;n-2p-2}, \quad \textbf{(5.10)}$$

где $\sum_{i=1}^{n} e_{i,}^2 \quad \sum_{i=1}^{n_1} e_{i,}^2 \quad \sum_{i=n_1}^{n} e_i^2$ — остаточные суммы квадратов соответственно для объединенной, первой и второй выборок; $n=n_1+n_2$.

▶ **Пример 5.3.**

По данным примера **5.2**, используя критерий Г. Чоу, выяснить, можно ли считать одной и той же линейную регрессию Y по X для юношей и девушек.

Р е ш е н и е. По n_1= 6 парам наблюдений (x_i, y_i) для *юношей* — (10;6), (8;4), (7;7), (7;4), (9;7), (5;2) (1-ая выборка) и по n_2=6 парам наблюдений для *девушек* — (6;4), (8;5), (6;4), (6;3), (6;3), (7;3) (2-ая выборка) рассчитаем уравнения регрессии:

$$\hat{y} = -1{,}000+0{,}783x \text{ (для 1-й выборки);}$$

$$\hat{y} = -0{,}048 + 0{,}571x \text{ (для 2-й выборки).}$$

По всем $n = n_1 + n_2 = 12$ парам наблюдений рассчитаем уравнение регрессии для объединенной выборки (см. пример.**5.2**):

$$\hat{y} = -1{,}437 + 0{,}815x.$$

Так как вычисленное по (**5.10**) значение

$$F = 0{,}21 < F_{0{,}05;2;8} = 4{,}46,$$

то влияние фактора «пол» несущественно, и в качестве оценки регрессионной модели Y по X можно рассматривать уравнение регрессии, полученное по объединенной выборке. ▶

Критерий Г. Чоу может быть использован при построении регрессионных моделей при воздействии *качественных* признаков, когда имеется возможность разделения совокупности наблюдений по степени воздействия этого фактора на отдельные группы и требуется установить возможность использования единой модели регрессии.

Оценивание регрессии с использованием *фиктивных переменных* более информативно в том отношении, что позволяет использовать t-критерий для оценки существенности влияния каждой фиктивной переменной на зависимую переменную.

5.5. Нелинейные модели регрессии

До сих пор мы рассматривали л и н е й н ы е р е г р е с с и о н н ы е м о д е л и, в которых переменные имели первую степень (модели, *линейные по переменным*), а параметры выступали в виде коэффициентов при этих переменных (модели, *линейные по параметрам*). Однако соотношение между социально-экономическими явлениями и процессами далеко не всегда можно выразить линейными функциями, так как при этом могут возникать неоправданно большие ошибки.

Так, например, н е л и н е й н ы м и оказываются *производственные функции* (зависимости между объемом произведенной продукции и основными факторами производства — трудом, капиталом и т. п.), *функции спроса* (зависимость между спросом на товары или услуги и их ценами или доходом) и другие.

Для оценки параметров нелинейных моделей используются два подхода.

Первый подход основан на *линеаризации* модели и заключается в том, что с помощью подходящих преобразований исходных переменных исследуемую зависимость представляют в виде *линейного* соотношения между *преобразованными* переменными.

Второй подход обычно применяется в случае, когда подобрать соответствующее линеаризующее преобразование не удается. В этом случае применяются методы *нелинейной оптимизации* на основе исходных переменных.

Для л и н е а р и з а ц и и модели в рамках первого подхода могут использоваться как модели, не линейные по переменным, так и не линейные по параметрам.

Если модель **нелинейна по переменным**, то введением новых переменных ее можно свести к линейной модели, для оценки параметров которой использовать обычный метод наименьших квадратов.

Так, например, если нам необходимо оценить параметры регрессионной модели

$$y_i = \beta_0 + \beta_1 x_{i1}^2 + \beta_2 \sqrt{x_{i2}} + \varepsilon_i, \quad i = 1,..., n,$$

то вводя новые переменные $Z_1 = X_1^2$, $Z_2 = \sqrt{X_2}$, получим линейную модель

$$y_i = \beta_0 + \beta_1 z_{i1} + \beta_2 z_{i2} + \varepsilon_i, \quad i=1,..., n,$$

параметры которой находятся обычным методом наименьших квадратов по формуле (**4.8**).

Следует, однако, отметить и недостаток такой замены переменных, связанный с тем, что вектор оценок *b* получается не из условия минимизации суммы квадратов отклонений для *исходных* переменных, а из условия минимизации суммы квадратов отклонений для *преобразованных* переменных, что не одно и то же. В связи с этим необходимо определенное уточнение полученных оценок.

Более сложной проблемой является **нелинейность** модели **по параметрам**, так как непосредственное применение метода наименьших квадратов для их оценивания невозможно. К числу таких моделей можно отнести, например, *мультипликативную* (*степенную*) модель

$$y_i = \beta_0 x_{i1}^{\beta_1} x_{i2}^{\beta_2} \varepsilon_i, \; i = 1,..., n, \qquad \textbf{(5.11)}$$

экспоненциальную модель

$$y_i = e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}} \varepsilon_i, \quad i = 1,..., n \qquad \textbf{(5.12)}$$

и другие.

В ряде случаев путем подходящих преобразований эти модели удается привести к линейной форме. Так, модели (**5.11**) и (**5.12**) могут быть приведены к линейным логарифмированием обеих частей уравнений. Тогда, например, модель (**5.11**) примет вид:

$$\ln y_i = \ln \beta_0 + \beta_1 \ln x_{i1} + \beta_2 x_{i2} + \ln \varepsilon_i, \quad i = 1,..., n. \qquad \textbf{(5.13)}$$

К модели (**5.13**) уже можно применять обычные методы исследования линейной регрессии, изложенные в гл. **4**. Однако следует подчеркнуть, что критерии значимости и интервальные оценки параметров, применяемые для нормальной линейной регрессии, требуют, чтобы нормальный закон распределения в моделях (**5.11**), (**5.12**) имел логарифм вектора возмущений ε (т. е. $\ln \varepsilon \approx N_n$ (**0**, $\sigma^2 E_n$), а вовсе не ε. Другими словами, вектор возмущений ε должен иметь логарифмически нормальное распределение.

Заметим попутно, что к модели

$$y_i = \beta_0 x_{i1}^{\beta_1} x_{i2}^{\beta_2} + \varepsilon_i, \quad i = 1,..., n, \qquad \textbf{(5.14)}$$

рассматриваемой в качестве альтернативной по отношению к модели (**5.11**), изложенные выше методы исследования линейной регрессии уже непригодны, так как модель (**5.14**) нельзя привести к линейному виду. В этом случае используются специальные (итеративные) процедуры оценивания параметров.

В качестве примера использования линеаризирующего преобразования регрессии рассмотрим *производственную функцию Кобба—Дугласа*

$$Y = AK^{\alpha} L^{\beta}, \qquad \textbf{(5.15)}$$

где Y — объем производства, K — затраты капитала, L — затраты труда.

Показатели α и β являются *коэффициентами частной эластичности*[1] объема производства Y соответственно по затратам

---

[1] *Коэффициентом частной эластичности* $E_{x_i}(y)$ функции $y = f(x_1, x_2, ..., x_n)$ относительно переменной x_i ($i = 1,2,...,n$) называется предел отношения относительного частного приращения функции к относительному приращению этой переменной при $\Delta x_i \to 0$, т.е. $E_{x_i}(y) = \lim\limits_{\Delta x_i \to 0} \left(\dfrac{\Delta_{x_i} y}{y} : \dfrac{\Delta x_i}{x_i} \right) = \dfrac{x_i}{y} y'_{x_i}$. Нетрудно убедиться в том, что для функции Кобба—Дугласа $E_k(Y) = \alpha$, $E_L(Y) = \beta$.

капитала K и труда L. Это означает, что при увеличении одних только затрат капитала (труда) на 1% объем производства увеличится на α% (β%).

Учитывая влияние случайных возмущений, присущих каждому экономическому явлению, функцию Кобба—Дугласа (**5.15**) можно представить в виде

$$Y = AK^{\alpha} L^{\beta} \varepsilon . \qquad \textbf{(5.16)}$$

Полученную мультипликативную (степенную) модель легко свести к линейной путем логарифмирования обеих частей уравнения (**5.16**). Тогда для i-го наблюдения получим

$$\ln_{yi} = \ln A + \alpha \ln K_i + \beta \ln L_i + \ln \varepsilon_i , \; i = 1,..., n. \qquad \textbf{(5.17)}$$

Если в модели (**5.16**) $\alpha+\beta=1$ (т. е. модель такова, что при расширении масштаба производства — увеличении затрат капитала K и труда L в некоторое число раз — объем производства возрастает в то же число раз) функцию Кобба—Дугласа представляют в виде

$$Y = AK^{\alpha} L^{1-\alpha} \varepsilon$$

или

$$\frac{Y}{L} = A\left(\frac{K}{L}\right)^{\alpha} \varepsilon . \qquad \textbf{(5.18)}$$

Таким образом, получаем зависимость производительности труда (Y/L) от его капиталовооруженности (K/L). Для оценки параметров модели (**5.18**) путем логарифмирования приводим ее к виду (для i-го наблюдения)

$$\ln(Y/L)_i = \ln A + \alpha \ln(K/L)_i + \ln \varepsilon_i , \; i = 1,..., n. \qquad \textbf{(5.19)}$$

Функция Кобба—Дугласа с учетом технического прогресса имеет вид:

$$Y = AK^{\alpha} L^{\beta} e^{\theta t} \varepsilon , \qquad \textbf{(5.20)}$$

где t — время; параметр θ — темп прироста объема производства благодаря техническому прогрессу. Модель (**5.20**) приводится к линейному виду аналогично модели (**5.16**).

▶ **Пример 5.4.** По данным $n = 50$ предприятий легкой промышленности оценить производственную функцию Кобба—Дугласа в виде (**5.18**).

Р е ш е н и е. От исходных значений переменных K/L и Y/L перейдем к их натуральным логарифмам и, используя метод наименьших квадратов, рассчитаем оценки параметров модели (**5.19**)[1]. Получим

$$\ln\left(\hat{Y/L}\right)=0{,}43+0{,}25\ln\,(K/L),$$
$$(0{,}32)\quad(0{,}04)$$

или $$\hat{Y/L}=1{,}537(K/L)^{0{,}25}$$

(в скобках указаны стандартные отклонения коэффициентов регрессии). Коэффициент регрессии, равный 0,25 (он же коэффициент эластичности в модели (**5.18**)), говорит о том, что при изменении капиталовооруженности труда на 1% производительность труда на предприятиях отрасли увеличивается в среднем на 0,25%. С помощью t-критерия легко убедиться в том, что коэффициент регрессии, а значит, и уравнение парной регрессии, значимы. ▶

5.6. Частная корреляция

Выше, в § **3.3**, для оценки тесноты связи между переменными был введен выборочный коэффициент линейной корреляции. Если переменные коррелируют друг с другом, то на значении коэффициента корреляции частично сказывается влияние других переменных. В связи с этим часто возникает необходимость исследовать *частную корреляцию* между переменными при исключении (элиминировании) влияния одной или нескольких переменных.

Выборочным частным коэффициентом корреляции (или просто ***частным коэффициентом корреляции***) *между переменными* X_i и X_j *при фиксированных значениях остальных* $(p-2)$ *переменных называется выражение*

$$r_{ij.1,2\ldots p}=\frac{-q_{ij}}{\sqrt{q_{ii}q_{jj}}}, \qquad \textbf{(5.21)}$$

где q_{ii} и q_{jj} — алгебраические дополнения элементов r_{ij} и r_{jj} матрицы выборочных коэффициентов корреляции

[1] Исходные данные и соответствующие расчеты (полностью идентичные приведенным в гл. **2**) здесь не приводятся.

$$q_p=\begin{pmatrix} 1 & r_{12} & \dots & r_{1p} \\ r_{21} & 1 & \dots & r_{2p} \\ \dots & \dots & \dots & \dots \\ r_{p1} & r_{p2} & \dots & 1 \end{pmatrix},$$

а r_{ij} определяются по формулам (**3.18**)—(**3.20**). В частности, в случае трех переменных (n=3) из (**5.21**) следует, что

$$r_{ij.k}=\frac{r_{ij}-r_{ik}r_{jk}}{\sqrt{\left(1-r_{ik}^2\right)\left(1-r_{jk}^2\right)}}\ . \qquad \textbf{(5.22)}$$

Поясним полученную формулу (**5.22**). Предположим, что имеется обычная регрессионная модель $x_i=\beta_0+\beta_1x_j+\beta_2x_k+\varepsilon_i$ и необходимо оценить корреляцию между зависимой переменной X_i и объясняющей переменной X_j при исключении (элиминировании) влияния другой объясняющей переменной X_k. С этой целью найдем уравнения парной регрессии X_i по X_k ($\hat{x}_i = b_0+b_1x_k$) и X_j по X_k ($\hat{x}_i = b_0' + b_1'x_k$), а затем удалим влияние переменной X_k, взяв остатки $e_{x_i} = x_i - \hat{x}_i$ и $e_{x_j} = x_j - \hat{x}_j$. Очевидно, что коэффициент корреляции между остатками e_{x_i} и e_{x_j} будет отражать тесноту частной корреляции между переменными X_i и X_j при исключении влияния переменной X_k. Можно показать, что найденный по формуле (**3.18**) обычный *коэффициент корреляции между остатками* e_{x_i} и e_{x_j} равен частному коэффициенту корреляции $r_{ij.k}$, определенному по формуле (**5.22**).

Частный коэффициент корреляции $r_{ij.12\dots p}$, как и парный коэффициент r_{ij}, может принимать значения от -1 до $+1$. Кроме того, $r_{ij.12\dots p}$, вычисленный на основе выборки объема n, имеет такое же распределение, как и r_{ij}, вычисленный по $n'=n-p+2$ наблюдениям. Поэтому значимость частного коэффициента корреляции $r_{ij.12\dots p}$ оценивают так же, как и обычного коэффициента корреляции r (см. § **3.6**), но при этом полагают $n'=n-p+2$.

▶ **Пример 5.5.** Для исследования зависимости между производительностью труда (X_1), возрастом (X_2) и производственным стажем (X_3) была произведена выборка из 100 рабочих одной и той же специальности. Вычисленные парные коэффициенты

корреляции оказались значимыми и составили: r_{12}=0,20; r_{13}=0,41; r_{23}=0,82. Вычислить частные коэффициенты корреляции и оценить их значимость на уровне α=0,05.

Р е ш е н и е. По формуле (**5.22**) частные коэффициенты корреляции

$$r_{12.3} = \frac{0{,}20^2 - 0{,}41 \cdot 0{,}82}{\sqrt{(1-0{,}41)^2(1-0{,}82)^2}} = -0{,}26$$

и аналогично $r_{13.2}$=0,44; $r_{23.1}$=0,83.

Оценим значимость $r_{12.3}$. Значение статистики t-критерия по (**3.46**) при $n'=n-p+2=100-3+2=99$ (по абсолютной величине)

$$|t| = \frac{|-0{,}26|\sqrt{99-2}}{\sqrt{1-(-0{,}26)^2}} = 2{,}65$$

больше табличного $t_{0,95;97}$=1,99 (см. табл. II приложений), следовательно, частный коэффициент корреляции $r_{12.3}$ значим. Аналогично устанавливается значимость других частных коэффициентов корреляции.

Сравнивая частные коэффициенты корреляции $r_{ij\,k}$ с соответствующими парными коэффициентами, видим, что за счет «очищения связи» наибольшему изменению подвергся коэффициент корреляции между производительностью труда (X_1) и возрастом (X_2) рабочих (изменилось не только его значение, но и знак: r_{12}=0,20; $r_{12.3}$=−0,26, причем оба эти коэффициента значимы).

Итак, между производительностью труда (X_1) и возрастом (X_2) рабочих существует прямая корреляционная связь (r_{12}=0,20). Если же устранить (элиминировать) влияние переменной «производственный стаж» (X_3), то в чистом виде производительность труда (X_1) находится в обратной по направлению (и опять же слабой по тесноте) связи с возрастом рабочих (X_2) ($r_{12.3}$=−0,26). Это вполне объяснимо, если рассматривать возраст только как показатель работоспособности организма на определенном этапе его жизнедеятельности. Подобным образом могут быть интерпретированы и другие частные коэффициенты корреляции. ▶

Упражнения

5.6. Имеются следующие данные о потреблении некоторого продукта Y (усл. ед.) в зависимости от уровня урбанизации (до-

ли городского населения) X_1, относительного образовательного уровня X_2 и относительного заработка X_3 для девяти географических районов:

i (*номер района*)	x_{i1}	x_{i2}	x_{i3}	y_i	i (*номер района*)	x_{i1}	x_{i2}	x_{i3}	y_i
1	42,2	11,2	31,9	167,1	6	44,5	10,8	8,5	174,6
2	48,6	10,6	13,2	174,4	7	39,1	10,7	24,3	163,7
3	42,6	10,6	28,7	160,8	8	40,1	10,0	18,6	174,5
4	39,0	10,4	26,1	162,0	9	45,9	12,0	20,4	185,7
5	34,7	9,3	30,1	140,8					

Средние значения $\bar{x}_1 = 41{,}85$; $\bar{x}_2 = 10{,}62$; $\bar{x}_3 = 24{,}42$; $\bar{y} = 167{,}07$.

Стандартные отклонения $s_{x_1} = 4{,}176$; $s_{x_2} = 0{,}7463$; $s_{x_3} = 7{,}928$; $s_y = 12{,}645$.

Корреляционная матрица:

	X_1	X_2	X_3	Y
X_1	1	0,684	−0,616	0,802
X_2	0,684	1	−0,173	0,770
X_3	−0,616	−0,173	1	−0,629
Y	0,802	0,770	−0,629	1

Используя пошаговую процедуру отбора наиболее информативных объясняющих переменных, определить подходящую регрессионную модель, исключив при этом мультиколлинеарность. Оценить значимость коэффициентов регрессии полученной модели по t-критерию.

5.7. Имеются следующие данные о весе Y (в фунтах) и возрасте X (в неделях) 13 индеек, выращенных в областях *A*, *B*, *C*.

i	x_i	y_i	*Область происхождения*	i	x_i	y_i	*Область происхождения*
1	28	12,3	*A*	8	26	11,8	*B*
2	20	8,9	*A*	9	21	11,5	*C*
3	32	15,1	*A*	10	27	14,2	*C*
4	22	10,4	*A*	11	29	15,4	*C*
5	29	13,1	*B*	12	23	13,1	*C*
6	27	12,4	*B*	13	25	13,8	*C*
7	28	13,2	*B*				

Есть основание полагать, что на вес индеек оказывает влияние не только их возраст, но и область происхождения. Необходимо:

а) найти уравнение парной регрессии Y по X и оценить его значимость;

б) введя соответствующие фиктивные переменные, найти общее уравнение множественной регрессии Y по всем объясняющим переменным (включая фиктивные);

в) оценить значимость общего уравнения множественной регрессии по F-критерию и значимость его коэффициентов по t-критерию на уровне α=0,05;

г) проследить за изменением скорректированного коэффициента детерминации при переходе от парной к множественной регрессии;

д) оценить на уровне α=0,05 значимость различия между свободными членами уравнений, получаемых из общего уравнения множественной регрессии Y для каждой области.

5.8. При построении линейной зависимости расходов на одежду от располагаемого дохода по выборке для 10 женщин получены следующие суммы квадратов и произведений наблюдений:

$$\sum_{i=1}^{10} x_i = 110\,,\ \sum_{i=1}^{10} x_i^2 = 1540\,,\ \sum_{i=1}^{10} y_i = 60\,,\ \sum_{i=1}^{10} x_i y_i = 828\,,\ \sum_{i=1}^{10} y_i^2 = 448\,.$$

Аналогичные вычисления сумм по выборке из 5 мужчин дали:

$$\sum_{i=1}^{5} x_i = 35,\ \sum_{i=1}^{5} x_i^2 = 325,\ \sum_{i=1}^{5} y_i = 15,\ \sum_{i=1}^{5} x_i y_i = 140,\ \sum_{i=1}^{5} y_i^2 = 61\,.$$

По общей (объединенной) выборке оценена регрессия с использованием фиктивной переменной Z (Z= 1 для мужчины и Z= 0 для женщины), которая имеет вид:

$$\hat{y} = -0,06 + 0,438x + 0,46z + 0,072(zx).$$

На уровне значимости α=0,05 проверить гипотезу о том, что функция потребления одна и та же для мужчин и женщин, если выполнены все предпосылки классической нормальной линейной регрессии.

5.9. Решить задачу **5.8**, используя критерий Г. Чоу.

5.10. С целью исследования влияния факторов X_1 — среднемесячного количества профилактических наладок автоматической линии и X_2 — среднемесячного числа обрывов нити на показатель Y — среднемесячную характеристику качества ткани (в баллах) по данным 37 предприятий легкой промышленности были вычислены парные коэффициенты корреляции: r_{y1}=0,105, r_{y2}=0,024 и r_{12}=0,996. Определить частные коэффициенты корреляции $r_{y1.2}$ и $r_{y2.1}$ и оценить их значимость на 5%-ном уровне.

Глава 6

Временные ряды и прогнозирование

При рассмотрении классической модели регрессии характер экспериментальных данных, как правило, не имеет принципиального значения. Однако это оказывается не так, если условия классической модели нарушены.

Методы исследования моделей, основанных на данных пространственных выборок и временных рядов, вообще говоря, существенно отличаются. Объясняется это тем, что в отличие от пространственных выборок наблюдения во временных рядах, как правило, нельзя считать независимыми.

В этой главе мы остановимся на некоторых общих понятиях и вопросах, связанных с временными рядами, использованием регрессионных моделей временных рядов для прогнозирования. При анализе точности этих моделей и определении интервальных ошибок прогноза на их основе, будем полагать, что рассматриваемые в главе регрессионные модели временных рядов удовлетворяют условиям классической модели. Модели временных рядов, в которых нарушены эти условия, будут рассмотрены в гл. **7**, **8**.

6.1. Общие сведения о временных рядах и задачах их анализа

Под ***временным рядом*** (***динамическим рядом***, или ***рядом динамики***) в экономике подразумевается последовательность наблюдений некоторого признака (случайной величины) Y в последовательные моменты времени. Отдельные наблюдения называются *уровнями* ряда, которые будем обозначать y_t ($t = 1,2,..., n$), где n — число уровней.

В табл. 6.1 приведены данные, отражающие спрос на некоторый товар за восьмилетний период (усл. ед), т. е. временной ряд спроса y_t.

Таблица 6.1

Год, t	1	2	3	4	5	6	7	8
Спрос, y_t	213	171	291	309	317	362	351	361

В качестве примера на рис. 6.1 временной ряд y_t изображен графически.

В общем виде при исследовании экономического временного ряда y_t выделяются несколько составляющих:

$$y_t = u_t + v_t + c_t + \varepsilon_t \ (t = 1,2,..., n), \qquad \textbf{(6.1)}$$

где u_t — *тренд*, плавно меняющаяся компонента, описывающая чистое влияние долговременных факторов, т. е. длительную («вековую») тенденцию изменения признака (например, рост населения, экономическое развитие, изменение структуры потребления и т. п.);

v_t — *сезонная компонента,* отражающая повторяемость экономических процессов в течение не очень длительного периода (года, иногда месяца, недели и т. д., например, объем продаж товаров или перевозок пассажиров в различные времена года);

c_t — *циклическая компонента,* отражающая повторяемость экономических процессов в течение длительных периодов (например, влияние волн экономической активности Кондратьева, демографических «ям», циклов солнечной активности и т. п.);

ε_t — *случайная компонента,* отражающая влияние не поддающихся учету и регистрации случайных факторов.

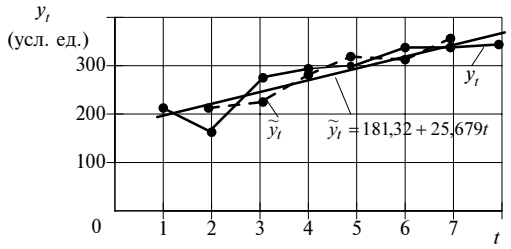


Рис 6.1

Следует обратить внимание на то, что в отличие от ε_t первые три составляющие (компоненты) u_t, v_t, c_t являются *закономерными, неслучайными.*

Важнейшей классической задачей при исследовании экономических временных рядов является выявление и статистическая оцен-

ка ***основной тенденции*** *развития изучаемого процесса и отклонений от нее.*

Отметим основные э т а п ы анализа временных рядов:

- графическое представление и описание поведения временного ряда;
- выделение и удаление закономерных (неслучайных) составляющих временного ряда (тренда, сезонных и циклических составляющих);
- сглаживание и фильтрация (удаление низко- или высокочастотных составляющих временного ряда);
- исследование случайной составляющей временного ряда, построение и проверка адекватности математической модели для ее описания;
- прогнозирование развития изучаемого процесса на основе имеющегося временного ряда;
- исследование взаимосвязи между различными временными рядами.

Среди наиболее распространенных методов анализа временных рядов выделим *корреляционный* и *спектральный анализ*, *модели авторегрессии* и *скользящей средней*. О некоторых из них речь пойдет ниже.

Если выборка y_1, y_2,..., y_i,..., y_n рассматривается как одна из реализаций случайной величины Y, временной ряд y_1, y_2,..., y_i,..., y_n рассматривается как одна из *реализаций* (*траекторий*) случайного процесса[1] $Y(t)$. Вместе с тем следует иметь в виду принципиальные отличия временного ряда y_t ($t = 1,2,..., n$) от последовательности наблюдений y_1, y_2,..., y_n, образующих случайную выборку. Во-первых, в отличие от элементов случайной выборки члены временного ряда, как правило, не являются статистически независимыми. Во-вторых, члены временного ряда не являются одинаково распределенными.

6.2. Стационарные временные ряды и их характеристики. Автокорреляционная функция

Важное значение в анализе временных рядов имеют с т а ц и о н а р н ы е временные ряды, вероятностные свойства ко-

[1] *Случайным процессом* (или *случайной функцией*) $Y(t)$ неслучайного аргумента t называется функция, которая при любом значении t является случайной величиной.

торых не изменяются во времени. Стационарные временные ряды применяются, в частности, при описании случайных составляющих анализируемых рядов.

Временной ряд y_t ($t = 1,2,..., n$) называется *строго стационарным* (или *стационарным в узком смысле*), если совместное распределение вероятностей n наблюдений y_1, y_2,..., y_n такое же, как и n наблюдений $y_{1+\tau}$, $y_{2+\tau}$,...., $y_{n+\tau}$ при любых n, t и τ. Другими словами, свойства строго стационарных[1] рядов y_t не зависят от момента t, т. е. закон распределения и его числовые характеристики не зависят от t. Следовательно, математическое ожидание $a_y(t) = a$, среднее квадратическое отклонение $\sigma_y(t) = \sigma$ могут быть оценены по наблюдениям y_t ($t = 1,2,..., n$) по формулам:

$$\bar{y}_t = \frac{\sum_{t=1}^{n} y_t}{n}; \qquad \textbf{(6.2)}$$

$$s_t^2 = \frac{\sum_{t=1}^{n} (y_t - \bar{y}_t)}{n}. \qquad \textbf{(6.3)}$$

Простейшим примером *стационарного временного ряда*, *у которого математическое ожидание равно нулю, а ошибки* ε_t *некоррелированы*, *является* «***белый шум***». Следовательно, можно сказать, что *возмущения* (*ошибки*) ε_t *в классической линейной регрессионной модели образуют белый шум*, а в случае их нормального распределения — *нормальный* (*гауссовский*) *белый шум*.

Степень тесноты связи между последовательностями наблюдений временного ряда y_1, y_2,..., y_n и $y_{1+\tau}$, $y_{2+\tau}$,..., $y_{n+\tau}$ (сдвинутых относительно друг друга на τ единиц, или, как говорят, с *лагом* τ) может быть определена с помощью коэффициента корреляции

$$\rho_{(\tau)} = \frac{M[(y_t - a)(y_{t+\tau} - a)]}{\sigma_x(t)\sigma_x(t+\tau)} = \frac{M[(y_t - a)(y_{t+\tau} - a)]}{\sigma^2}, \qquad \textbf{(6.4)}$$

ибо $M(y_t) = M(y_{t+\tau}) = a$, $\sigma_y(t) = \sigma_y(t+\tau) = \sigma$.

---

[1] Наряду со строго стационарными временными рядами (в узком смысле) в эконометрике рассматриваются *стационарные* ряды (*в широком смысле*), в которых требование неизменности при любых n, t и τ распространяется лишь на числовые характеристики указанного распределения.

Так как коэффициент $\rho(\tau)$ измеряет корреляцию между членами одного и того же ряда, его называют *коэффициентом автокорреляции*, а зависимость $\rho(\tau)$ — *автокорреляционной функцией*. В силу стационарности временного ряда y_t ($t = 1,2,..., n$) автокорреляционная функция $\rho(\tau)$ зависит только от лага τ, причем $\rho(-\tau) = \rho(\tau)$, т. е. при изучении $\rho(\tau)$ можно ограничиться рассмотрением только положительных значений τ.

Статистической оценкой $\rho(\tau)$ является *выборочный коэффициент автокорреляции* $r(\tau)$, определяемый по формуле коэффициента корреляции (**3.20**), в которой $x_i = y_t$, $y_i = y_{t+\tau}$, а n заменяется на $n - \tau$:

$$r(\tau) = \frac{(n-\tau)\sum_{t=1}^{n-\tau} y_t y_{t+\tau} - \sum_{t=1}^{n-\tau} y_t \sum_{t=1}^{n-\tau} y_{t+\tau}}{\sqrt{(n-\tau)\sum_{t=1}^{n-\tau} y_t^2 - \left(\sum_{t=1}^{n-\tau} y_t\right)^2}\sqrt{(n-\tau)\sum_{t=1+\tau}^{n} y_{t+\tau}^2 - \left(\sum_{t=1}^{n-\tau} y_{t+\tau}\right)^2}}. \qquad \textbf{(6.5)}$$

Функцию $r(\tau)$ называют *выборочной автокорреляционной функцией*, а ее график — *коррелограммой*.

При расчете $r(\tau)$ следует помнить, что с увеличением τ число $n - \tau$ пар наблюдений y_t, $y_{t+\tau}$ уменьшается, поэтому лаг τ должен быть таким, чтобы число $n - \tau$ было достаточным для определения $r(\tau)$. Обычно ориентируются на соотношение $\tau \le n/4$.

Для стационарного временного ряда с увеличением лага τ взаимосвязь членов временного ряда y_t и $y_{t+\tau}$ ослабевает и автокорреляционная функция $\rho(\tau)$ должна убывать (по абсолютной величине). В то же время для ее выборочного (эмпирического) аналога $r(\tau)$, особенно при небольшом числе пар наблюдений $n - \tau$, свойство монотонного убывания (по абсолютной величине) при возрастании τ может нарушаться.

Наряду с автокорреляционной функцией при исследовании стационарных временных рядов рассматривается *частная автокорреляционная функция* $\rho_{\text{част}}(\tau)$, где $\rho_{\text{част}}(\tau)$ есть *частный коэффициент корреляции между членами временного ряда* y_t и $y_{t+\tau}$, т. е. коэффициент корреляции между y_t и $y_{t+\tau}$ при устранении (элиминировании) влияния промежуточных (между y_t и $y_{t+\tau}$) членов.

Статистической оценкой $\rho_{\text{част}}(\tau)$ является *выборочная частная автокорреляционная функция* $r_{\text{част}}(\tau)$, где $r_{\text{част}}(\tau)$ — *выбороч-*

ный частный коэффициент корреляции, определяемый по формуле (**5.21**) или (**5.22**). Например, выборочный частный коэффициент автокорреляции 1-го порядка между членами временного ряда y_t и y_{t+2} при устранении влияния y_{t+1} может быть вычислен по формуле (**5.22**):

$$r_{\text{част}}(2) = r_{02.1} = \frac{r(2) - r(1)r(1,2)}{\sqrt{1-r^2(1)}\sqrt{1-r^2(1,2)}}, \qquad \textbf{(6.6)}$$

где $r(1)$, $r(1,2)$, $r(2)$ — выборочные коэффициенты автокорреляции между y_t и y_{t+1}, y_{t+1} и y_{t+2}, y_t и y_{t+2}, t=1,..., n.

▶ **Пример 6.1.** По данным табл. 6.1 для временного ряда y_t найти среднее значение, среднее квадратическое отклонение, коэффициенты автокорреляции (для лагов τ=1;2) и частный коэффициент автокорреляции 1-го порядка.

Р е ш е н и е. Среднее значение временного ряда находим по формуле (**6.2**):

$$\bar{y}_t = \frac{213+171+...+361}{8} = 296{,}88\,(\text{ед.}).$$

Дисперсию и среднее квадратическое отклонение можно вычислить по формуле (**6.3**), но в данном случае проще использовать соотношение

$$s_t^2 = \overline{y_t^2} - \bar{y}_t^2 = 92478{,}38 - 296{,}88^2 = 4343{,}61;$$

$$s_t = \sqrt{4343{,}61} = 65{,}31(\text{ед.}).$$

где $$\overline{y_t^2} = \frac{\sum_{t=1}^{n} y_t^2}{n} = \frac{213^2 + 171^2 + ... + 361^2}{8} = 92\,478{,}38.$$

Найдем коэффициент автокорреляции $r(\tau)$ временного ряда (для лага τ = 1), т. е. коэффициент корреляции между последовательностями семи пар наблюдений y_t и y_{t+1} (t = 1,2,...,7):

y_t	213	171	291	309	317	362	351
$y_{t+\tau}$	171	291	309	317	362	351	361

Вычисляем необходимые суммы:

$$\sum_{t=1}^{7} y_t = 213+171+...+351 = 2014;$$

$$\sum_{t=1}^{7} y_t^2 = 213^2+171^2+...+351^2 = 609\,506;$$

$$\sum_{t=1}^{7} y_{t+\tau} = 171+291+...+361=2162;$$

$$\sum_{t=1}^{7} y_{t+\tau}^2 = 171^2+291^2+...+361^2=694\ 458;$$

$$\sum_{t=1}^{7} y_t y_{t+\tau} = 213 \cdot 171+171 \cdot 291+...+351 \cdot 361=642\ 583.$$

Теперь по формуле (**6.5**) коэффициент автокорреляции

$$r(1) = \frac{7 \cdot 642\,583 - 2014 \cdot 2162}{\sqrt{7 \cdot 609\,506 - 2014^2}\sqrt{7 \cdot 694\,458 - 2162^2}} = 0{,}725.$$

Коэффициент автокорреляции $r(2)$ для лага $\tau = 2$ между членами ряда y_t и y_{t+2} ($t = 1,2,...,6$) по шести парам наблюдений вычисляем аналогично: $r(2)=0{,}842$.

Для определения частного коэффициента корреляции 1-го порядка $r_{\text{част}}(2) = r_{02.1}$ между членами ряда y_t и y_{t+2} при исключении влияния y_{t+1} вначале найдем (по аналогии с предыдущим) коэффициент автокорреляции $r(1,2)$ между членами ряда y_{t+1} и y_{t+2}: $r = (1,2)=0{,}825$, а затем вычислим $r_{\text{част}}(2)$ по формуле (**6.6**):

$$r_{\text{част}}(2) = r_{02.1} = \frac{0{,}842 - 0{,}725 \cdot 0{,}825}{\sqrt{1-0{,}725^2}\sqrt{1-0{,}825^2}} = 0{,}627. \blacktriangleright$$

Знание автокорреляционных функций $r(\tau)$ и $r_{\text{част}}(\tau)$ может оказать существенную помощь при подборе и идентификации модели анализируемого временного ряда и статистической оценке его параметров (см. об этом дальше).

6.3. Аналитическое выравнивание (сглаживание) временного ряда (выделение неслучайной компоненты)

Как уже отмечено выше, одной из важнейших задач исследования экономического временного ряда является выявление *основной тенденции* изучаемого процесса, выраженной неслучайной составляющей $f(t)$ (тренда либо тренда с циклической или (и) сезонной компонентой).

Для решения этой задачи вначале необходимо выбрать вид функции $f(t)$. Наиболее часто используются следующие функции:

линейная	—	$f(t) = b_0 + b_1 t$;
полиномиальная	—	$f(t) = b_0 + b_1 t + b_2 t^2 + ... + b_n t^n$;
экспоненциальная	—	$f(t) = e^{b_0 + b_1 t}$;
логистическая	—	$f(t) = \dfrac{a}{1 + be^{-ct}}$;
Гомперца	—	$\log_c f(t) = a - br^t$, где $0 < r < 1$.

Это весьма ответственный этап исследования. При выборе соответствующей функции $f(t)$ используют содержательный анализ (который может установить характер динамики процесса), визуальные наблюдения (на основе графического изображения временного ряда). При выборе полиномиальной функции может быть применен метод последовательных разностей (состоящий в вычислении разностей первого порядка $\Delta_t = y_t - y_{t-1}$, второго порядка $\Delta_t^{(2)} = \Delta_t - \Delta_{t-1}$ и т. д.), и порядок разностей, при котором они будут примерно одинаковыми, принимается за степень полинома.

Из двух функций предпочтение обычно отдается той, при которой меньше сумма квадратов отклонений фактических данных от расчетных на основе этих функций. Но этот принцип нельзя доводить до абсурда: так, для любого ряда из n точек можно подобрать полином $(n-1)$-й степени, проходящий через все точки, и соответственно с минимальной — нулевой — суммой квадратов отклонений, но в этом случае, очевидно, не следует говорить о выделении основной тенденции, учитывая случайный характер этих точек. Поэтому при прочих равных условиях предпочтение следует отдавать более простым функциям.

Для выявления основной тенденции чаще всего используется **метод наименьших квадратов**, рассмотренный в гл. **3**. Значения временного ряда y_t рассматриваются как зависимая переменная, а время t — как объясняющая:

$$y_t = f(t) + \varepsilon_t, \qquad \textbf{(6.7)}$$

где ε_t — возмущения, удовлетворяющие основным предпосылкам регрессионного анализа, приведенным в § **3.4**, т. е. представляющие независимые и одинаково распределенные случайные величины, распределение которых предполагаем нормальным.

Напомним, что согласно методу наименьших квадратов параметры прямой[1] $\hat{y}_t = f(t) = b_0 + b_1 t$ находятся из системы нормальных уравнений (**3.5**), в которой в качестве x_i берем t:

$$\begin{cases} b_0 n + b_1 \sum\limits_{t=1}^{n} t = \sum\limits_{t=1}^{n} y_t, \\ b_0 \sum\limits_{t=1}^{n} t + b_1 \sum\limits_{t=1}^{n} t^2 = \sum\limits_{t=1}^{n} y_t t. \end{cases} \qquad \textbf{(6.8)}$$

Учитывая, что значения переменной $t = 1,2,...n$ образуют натуральный ряд чисел от 1 до n, суммы $\sum\limits_{t=1}^{n} t$, $\sum\limits_{t=1}^{n} t^2$ можно выразить через число членов ряда n по известным в математике формулам:

$$\sum_{t=1}^{n} t = \frac{n(n+1)}{2}; \quad \sum_{t=1}^{n} t^2 = \frac{n(n+1)(2n+1)}{6}. \qquad \textbf{(6.9)}$$

▶ **Пример 6.2**.

По данным табл. 6.1 найти уравнение неслучайной составляющей (тренда) для временного ряда y_t, полагая тренд линейным.

Р е ш е н и е. По формуле (**6.9**)

$$\sum_{t=1}^{8} t = \frac{8 \cdot 9}{1 \cdot 2} = 36; \quad \sum_{t=1}^{8} t^2 = \frac{8 \cdot 9 \cdot 17}{6} = 204.$$

Далее

$$\sum_{t=1}^{8} y_t = 213+171+...+361=2375;$$

$$\sum_{t=1}^{8} y_t^2 = 213^2+171^2+...+361^2=739\ 827;$$

$$\sum_{t=1}^{8} y_t t = 213 \cdot 1+171 \cdot 2+...+361 \cdot 8=11\ 766.$$

Система нормальных уравнений (**6.8**) имеет вид:

$$\begin{cases} 8b_0 + 36b_1 = 2375, \\ 36b_0 + 204b_1 = 11\ 766, \end{cases}$$

---

[1] В случае, если функция $f(t)$ нелинейная и не представляется возможным применить *методы линеаризации* модели (§ **5.5**), то параметры тренда находят из соответствующих (в зависимости от вида функции $f(t)$) систем нормальных уравнений, которые здесь не приводятся (см., например, [25]), либо с помощью специальных процедур оценивания.

откуда $b_0 = 181{,}32$; $b_1 = 25{,}679$ и уравнение тренда $\hat{y}_t = 181{,}32 + 25{,}679t$ (рис 6.1), т. е. спрос ежегодно увеличивается в среднем на 25,7 ед.

При решении задачи можно было не выписывать систему нормальных уравнений, а представить уравнение регрессии в виде (**3.12**), т. е. $\hat{y}_t - \bar{y} = b_1(t - \bar{t})$, где

$$\bar{t} = \frac{\sum_{t=1}^{n} t}{n} = \frac{1+n}{2}, \qquad \bar{y} = \frac{\sum_{t=1}^{n} y_t}{n},$$

а коэффициент регрессии b_1 найти по формуле (**3.13**):

$$b_1 = \frac{\overline{y_t t} - \bar{y}_t \bar{t}}{\overline{t^2} - \bar{t}^2},$$

где

$$\overline{y_t t} = \sum_{t=1}^{n} y_t t / n; \quad \bar{y}_t = \sum_{t=1}^{n} y_t / n;$$

$$\bar{t} = (1+n)/2; \quad \overline{t^2} = (n+1)(2n+1)/6.$$

Проверим значимость полученного уравнения тренда по F-критерию на 5%-ном уровне значимости. Вычислим с помощью формулы (**3.40**) суммы квадратов:

а) обусловленную регрессией —

$$Q_R = \sum_{t=1}^{n} \left(\hat{y}_t - \bar{y}_t\right)^2 = \sum_{t=1}^{n} b_1^2 \left(t - \bar{t}\right)^2 = b_1^2 \left(\sum_{t=1}^{n} t^2 - \frac{\left(\sum_{t=1}^{n} t\right)^2}{n} \right) =$$

$$= 25{,}679^2 \left(204 - \frac{36^2}{8} \right) = 27\,695{,}3;$$

б) общую —

$$Q = \sum_{t=1}^{n} \left(y_t - \bar{y}_t\right)^2 = \sum_{t=1}^{n} y_t^2 - \frac{\left(\sum_{t=1}^{n} y_t\right)^2}{n} =$$

$$= 739\,827 - \frac{2375^2}{8} = 34\,748{,}9;$$

в) остаточную—

$$Q_e = Q - Q_R = 34\,748{,}9 - 27\,695{,}3 = 7053{,}6.$$

Найдем по формуле (**3.44**) значение статистики:

$$F = \frac{Q_R(n-2)}{Q_e} = \frac{27\,695{,}3 \cdot 6}{7053{,}6} = 23{,}56.$$

Так как $F > F_{0,05;1;6}$ (см. табл. IV приложений), то уравнение тренда значимо. ▶

При применении метода наименьших квадратов для оценки параметров экспоненциальной, логистической функций или функции Гомперца возникают сложности с решением получаемой системы нормальных уравнений, поэтому предварительно, до получения соответствующей системы, прибегают к некоторым преобразованиям этих функций (например, логарифмированию и др.) (см. § **5.5**).

Другим методом выравнивания (сглаживания) временного ряда, т. е. выделения неслучайной составляющей, является **метод скользящих средних**. Он основан на переходе от начальных значений членов ряда к их средним значениям на интервале времени, длина которого определена заранее. При этом сам выбранный интервал времени «скользит» вдоль ряда.

Получаемый таким образом ряд скользящих средних ведет себя более гладко, чем исходный ряд, из-за усреднения отклонений ряда. Действительно, если индивидуальный разброс значений члена временного ряда y_t около своего среднего (сглаженного) значения a характеризуется дисперсией σ^2, то разброс средней из m членов временного ряда $(y_1 + y_2 + ... + y_m)/m$ около того же значения a будет характеризоваться существенно меньшей величиной дисперсии, равной σ^2/m. Для усреднения могут быть использованы средняя арифметическая (простая и с некоторыми весами), медиана и др.

▶ **Пример 6.3**. Провести сглаживание временного ряда y_t по данным табл. 6.1 методом скользящих средних, используя простую среднюю арифметическую с интервалом сглаживания $m = 3$ года.

Р е ш е н и е. Скользящие средние находим по формуле:

$$\widetilde{y}_t = \frac{\sum_{i=t-p}^{t+p} y_i}{m}, \qquad \textbf{(6.10)}$$

когда $m = (2p - 1)$— нечетное число; при $m = 3$ $p = 1$.

Например, при t=2 по формуле (**6.10**):

$$\widetilde{y}_2 = \frac{1}{3}(y_1 + y_2 + y_3) = \frac{1}{3}(213 + 171 + 291) = 225\ (\text{ед.});$$

при t =3

$$\widetilde{y}_3 = \frac{1}{3}(y_2 + y_3 + y_4) = \frac{1}{3}(171 + 291 + 309) = 241{,}0\ (\text{ед.})\ \text{и т. д.}$$

В результате получим сглаженный ряд:

t	1	2	3	4	5	6	7	8
$\widetilde{y}_t$	—	225,0	241,0	305,7	329,3	336,3	358,0	—

На рис. 6.1 этот ряд изображен графически в виде пунктирной линии. ▶

6.4. Прогнозирование на основе моделей временных рядов

Одна из важнейших задач (этапов) анализа временного (динамического) ряда, как отмечено выше, состоит в прогнозировании на его основе развития изучаемого процесса. При этом исходят из того, что тенденция развития, установленная в прошлом, может быть распространена (экстраполирована) на будущий период.

Задача ставится так: имеется временной (динамический) ряд $y_t(t = 1,2,...,n)$ и требуется дать прогноз уровня этого ряда на момент $n+\tau$.

Выше, в § **3.5, 4.2, 4.5**, мы рассматривали *точечный* и *интервальный прогноз* значений зависимой переменной *Y*, т. е. *определение точечных и интервальных оценок Y*, полученных для парной и множественной регрессий для значений объясняющих переменных *X*, расположенных вне пределов обследованного диапазона значений *X*.

Если рассматривать временной ряд как регрессионную модель изучаемого признака по переменной «время», то к нему могут быть применены рассмотренные выше методы анализа. Следует, однако,

вспомнить, что одна из основных предпосылок регрессионного анализа состоит в том, что возмущения $\varepsilon_t (t = 1,2,...,n)$ представляют собой независимые случайные величины с математическим ожиданием (средним значением), равным нулю. А при работе с временными рядами такое допущение оказывается во многих случаях неверным (см. об этом гл. **7**, **8**).

В данной главе мы полагаем, что возмущения ε_t ($t = 1,..., n$) удовлетворяют предпосылкам регрессионного анализа, т. е. условиям нормальной классической регрессионной модели (§ **3.4**).

▶ **Пример 6.4**. По данным табл. 6.1 дать точечную и с надежностью 0,95 интервальную оценки прогноза среднего и индивидуального значений спроса на некоторый товар на момент $t = 9$ (девятый год). (Полагаем, что тренд линейный, а возмущения удовлетворяют требованиям классической модели (см. дальше, пример **7.8**.)

Р е ш е н и е. Выше, в примере **6.2**, получено уравнение регрессии $\hat{y}_t = 181,32 + 25,679t$, т. е. ежегодно спрос на товар увеличивался в среднем на 25,7 ед. Надо оценить условное математическое ожидание $M_{t=9}(Y) = \bar{y}(9)$. Оценкой $\bar{y}(9)$ является групповая средняя

$$\hat{y}_{t=9} = 181,32 + 25,679 \cdot 9 = 412,4 (\text{ед.}).$$

Найдем по формуле (**3.26**) оценку s^2 дисперсии σ^2 (см. далее, табл. 7.1):

$$s^2 = \frac{\sum_{t=1}^{n} e_t^2}{n-2} = \frac{7059,2}{8-2} = 1176,5.$$

Вычислим оценку дисперсии групповой средней по формуле (**3.33**):

$$s^2_{\hat{y}_{t=9}} = 1176,5\left(\frac{1}{8} + \frac{(9-4,5)^2}{42}\right) = 714,3;$$

$$s_{\hat{y}_{t=9}} = \sqrt{714,3} = 26,73 \text{ (ед.)}$$

(здесь мы использовали данные, полученные в примере **6.2**:

$$\bar{t} = \frac{\sum_{t=1}^{n} t}{n} = \frac{36}{8} = 4,5;$$

$$\sum_{t=1}^{n}\left(t-\bar{t}\right)^2=\sum_{t=1}^{n}t^2-\frac{\left(\sum_{t=1}^{8}t\right)^2}{n}=204-\frac{36^2}{8}=42).$$

По табл. II приложений $t_{0,95;6}$=2,45. Теперь по формуле (**3.34**) интервальная оценка прогноза среднего значения спроса:

$$412,4-2,45\cdot 26,73\ \le \bar{y}(9)\le 412,4+2,45\cdot 26,73\,,$$

или $$346,9\le \bar{y}(9)\le 477,9\ (\text{ед.}).$$

Для нахождения интервальной оценки прогноза индивидуального значения $y^*(9)$ вычислим дисперсию его оценки по формуле (**3.35**):

$$s^2_{\hat{y}_{t_0=9}}=1176,5\left(1+\frac{1}{8}+\frac{(9-4,5)^2}{42}\right)=1890,8;$$

$$s_{\hat{y}_{t_0=9}}=43,48\ (\text{ед.}),$$

а затем по формуле (**3.36**) — саму интервальную оценку для $y^*(9)$:

$$412,4-2,45\cdot 43,48\le\ y*(9)\ \le 412,4+2,45\cdot 43,48,$$

или $$305,9\le\ y*(9)\ \le 518,9\ (\text{ед.}).$$

Итак, с надежностью 0,95 среднее значение спроса на товар на девятый год будет заключено от 346,9 до 477,9 (ед.), а его индивидуальное значение — от 305,9 до 518,9 (ед.).▶

Прогноз развития изучаемого процесса на основе экстраполяции временных рядов может оказаться эффективным, как правило, в рамках *краткосрочного*, в крайнем случае, *среднесрочного периода прогнозирования*.

6.5. Понятие об авторегрессионных моделях и моделях скользящей средней

Для данного временного ряда далеко не всегда удается подобрать ***адекватную*** *модель, для которой ряд возмущений* ε_t *будет удовлетворять основным предпосылкам регрессионного анализа*. До сих пор мы рассматривали модели вида (**6.7**), в которых в качестве регрессора выступала переменная t — «время». В эконометрике достаточно широкое распространение получили и другие

регрессионные модели, в которых регрессорами выступают ***лаговые переменные***, т. е. переменные, *влияние которых в эконометрической модели характеризуется некоторым запаздыванием.* Еще одним отличием рассматриваемых в этом параграфе регрессионных моделей является то, что представленные в них объясняющие переменные являются величинами *случайными.* (Подробнее об этих моделях см. гл. **8**.)

Авторегрессионная модель p-го порядка (или модель *AR* (*p*)) имеет вид:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + ... + \beta_p y_{t-p} + \varepsilon_t \quad (t = 1,2,...,n), \textbf{ (6.11)}$$

где $\beta_0, \beta_1, ..., \beta_p$ — некоторые константы.

Она описывает изучаемый процесс в момент t в зависимости от его значений в предыдущие моменты $t-1$, $t-2$,..., $t-p$.

Если исследуемый процесс y_t в момент t определяется его значениями только в предшествующий период $t-1$, то рассматривают *авторегрессионную модель* 1*-го порядка* (или *модель AR* (1) — *марковский случайный процесс*):

$$y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t, \quad (t = 1,2,..., n). \qquad \textbf{(6.12)}$$

▶ **Пример 6.5.** В таблице представлены данные, отражающие динамику курса акций некоторой компании (ден. ед.):

Таблица 6.2

t	1	2	3	4	5	6	7	8	9	10	11
y_t	971	1166	1044	907	957	727	752	1019	972	815	823
t	12	13	14	15	16	17	18	19	20	21	22
y_t	1112	1386	1428	1364	1241	1145	1351	1325	1226	1189	1213

Используя авторегрессионную модель 1-го порядка, дать точечный и интервальный прогноз среднего и индивидуального значений курса акций в момент $t = 23$, т. е. на глубину один интервал.

Р е ш е н и е. Попытка подобрать к данному временному ряду адекватную модель вида (**6.7**) с линейным или полиномиальным трендом оказывается бесполезной.

В соответствии с условием применим авторегрессионную модель (**6.12**). Получим (аналогично примеру **6.2**)

$$\hat{y}_t = 284,0 + 0,7503 y_{t-1}. \qquad \textbf{(6.13)}$$

Найденное уравнение регрессии значимо на 5%-ном уровне по F-критерию, так как фактически наблюдаемое значение статистики $F = 24,32 > F_{0,05;1;19} = 4,35$. Можно показать (например, с помощью критерия Дарбина—Уотсона) (см. далее, § **7.7**)), что возмущения (ошибки) ε_t в данной модели удовлетворяют условиям классической модели и для проведения прогноза могут быть использованы уже изученные нами методы.

Вычисления, аналогичные примеру **6.3**, дают точечный прогноз по уравнению (**6.13**):

$$\hat{y}_{t=23} = 284,0 + 0,7503 \cdot 1213 = 1194,1$$

и интервальный на уровне значимости 0,05 для среднего и индивидуального значений —

$$1046,6 \le \bar{y}_{t=23} \le 1341,6; \quad 879,1 \le y^*_{t_0=23} \le 1509,1.$$

Итак, с надежностью 0,95 среднее значение курса акций данной компании на момент $t = 23$ будет заключено в пределах от 1046,6 до 1341,6 (ден. ед.), а его индивидуальное значение — от 879,1 до 1509,1 (ден. ед.). ▶

Наряду с авторегрессионными моделями временных рядов в эконометрике рассматриваются также *модели скользящей средней*[1], в которой моделируемая величина задается линейной функцией от возмущений (ошибок) в предыдущие моменты времени.

Модель скользящей средней q-го *порядка* (или *модель*[2] *MA* (q)), имеет вид:

$$y_t = \varepsilon_t + \gamma_1 \varepsilon_{t-1} + \gamma_2 \varepsilon_{t-2} + \ldots + \gamma_q \varepsilon_{t-q}. \qquad \textbf{(6.14)}$$

В эконометрике используются также комбинированные модели временных рядов *AR* и *MA*.

---

[1] Термин «*скользящая средняя*» не следует путать с аналогичным термином, используемым в технике сглаживания временных рядов (§ **6.2**).

[2] «*MA*» — от английских слов «*moving average*» — скользящая средняя.

Авторегресссионная модель скользящей средней порядков p и *q соответственно* (или *модель ARMA* (*p*, *q*)) имеет вид:

$$y_t = \beta_0 + \beta_1 y_{t-1} + ... + \beta_p y_{t-p} + \varepsilon_t + \gamma_1 \varepsilon_{t-1} + ... + \gamma_q \varepsilon_{t-q}. \quad \textbf{(6.15)}$$

В заключение этой главы отметим, что использование соответствующих авторегрессионных моделей для прогнозирования экономических показателей, т. е. *автопрогноз* на базе рассмотренных моделей, может оказаться весьма эффективным (как правило, в краткосрочной перспективе).

Упражнения

В примерах **6.6—6.8** имеются следующие данные об урожайности озимой пшеницы y_t (ц/га) за 10 лет:

t	1	2	3	4	5	6	7	8	9	10
y_t	16,3	20,2	17,1	7,7	15,3	16,3	19,9	14,4	18,7	20,7

6.6. Найти среднее значение, среднее квадратическое отклонение и коэффициенты автокорреляции (для лагов $\tau = 1;2$) временного ряда.

6.7. Найти уравнение тренда временного ряда y_t, полагая, что он линейный, и проверить его значимость на уровне 0,05.

6.8. Провести сглаживание временного ряда y_t методом скользящих средних, используя простую среднюю арифметическую с интервалом сглаживания: *а*) $m = 3$; *б*) $m = 5$.

6.9. В таблице представлены данные, отражающие динамику роста доходов на душу населения y_t (ден. ед.) за восьмилетний период:

t	1	2	3	4	5	6	7	8
y_t	1133	1222	1354	1389	1342	1377	1491	1684

Полагая, что тренд линейный и условия классической модели выполнены:

а) найти уравнение тренда и оценить его значимость на уровне 0,05;

б) дать точечный и с надежностью 0,95 интервальный прогнозы среднего и индивидуального значений доходов на девятый год.

Глава 7

Обобщенная линейная модель. Гетероскедастичность и автокорреляция остатков

При моделировании реальных экономических процессов мы нередко сталкиваемся с ситуациями, в которых условия классической линейной модели регрессии оказываются нарушенными. В частности, могут не выполняться предпосылки *3* и *4* регрессионного анализа (см. (**3.24**) и (**3.25**)) о том, что случайные возмущения (ошибки) модели имеют постоянную дисперсию и не коррелированы между собой. Для линейной множественной модели эти предпосылки означают (см. § **4.2**), что ковариационная матрица вектора возмущений (ошибок) ε имеет вид:

$$\sum_{\varepsilon} = \sigma^2 E_n .$$

В тех случаях, когда имеющиеся статистические данные достаточно однородны, допущение $\sum_{\varepsilon} = \sigma^2 E_n$ вполне оправдано. Однако в других ситуациях оно может оказаться неприемлемым. Так, например, при использовании зависимости расходов на потребление от уровня доходов семей можно ожидать, что в более обеспеченных семьях вариация расходов выше, чем в малообеспеченных, т. е. дисперсии возмущений не одинаковы. При рассмотрении временных рядов мы, как правило, сталкиваемся с ситуацией, когда наблюдаемые в данный момент значения зависимой переменной коррелируют с их значениями в предыдущие моменты времени, т. е. наблюдается корреляция между возмущениями в разные моменты времени.

7.1. Обобщенная линейная модель множественной регрессии

Обобщенная линейная модель множественной регрессии (*Generalized Linear Multiple Regression model*)

$$Y = X\beta + \varepsilon , \qquad \textbf{(7.1)}$$

в которой переменные и параметры определены так же, как в § **4.1**, описывается следующей системой соотношений и условий:

1. ε — случайный вектор; X — неслучайная (детерминированная) матрица;

2. $M(\varepsilon)=\mathbf{0}_n$;

3. $\sum_{\varepsilon} = M(\varepsilon\varepsilon') = \Omega$, где Ω — положительно определенная матрица;

4. $r(X) = p+1<n$,

где p — число объясняющих переменных; n — число наблюдений.

Сравнивая обобщенную модель с классической (§ **4.2**), видим, что она отличается от классической только видом ковариационной матрицы: вместо $\sum_{\varepsilon} = \sigma^2 E_n$ для классической модели имеем $\sum_{\varepsilon} = \Omega$ для обобщенной. Это означает, что *в отличие от классической, в обобщенной модели ковариации и дисперсии объясняющих переменных могут быть произвольными*. В этом состоит суть обобщения регрессионной модели.

Для оценки параметров модели (**7.1**) можно применить обычный метод наименьших квадратов.

Оценка $b = (X'X)^{-1}X'Y$, полученная ранее и определенная соотношением (**4.8**), остается справедливой и в случае обобщенной модели. Оценка b по-прежнему несмещенная (доказательство точно такое же, как приведенное в § **4.2**) и состоятельная.

Однако полученная ранее формула для ковариационной матрицы вектора оценок $\sum_b$ оказывается неприемлемой в условиях обобщенной модели.

Действительно, учитывая (**4.15**), получим для обобщенной модели

$$\sum_{b^*} = (X'X)^{-1}X'M(\varepsilon\varepsilon')X(X'X)^{-1} = (X'X)^{-1}X'\Omega X(X'X)^{-1}, \qquad \textbf{(7.2)}$$

в то время как для классической модели имели по формуле (**4.16**)

$$\sum_b = \sigma^2(X'X)^{-1}. \qquad \textbf{(7.3)}$$

Найдем математическое ожидание остаточной суммы квадратов $\sum_{i=1}^{n} e_i^2 = e'e$. Используя преобразования, аналогичные приведенным в § **4.4**, можно показать[1], что для обобщенной модели

$$M(e'e) = \mathrm{tr}[(E_n - X(X'X)^{-1}X')\Omega], \qquad \textbf{(7.4)}$$

[1] См., например, [2] или [18].

т. е. в соответствии с (**4.21**)

$$M(s^2) = M\left(\frac{e'e}{n-p-1}\right) = \frac{\mathrm{tr}[(E_n - X(X'X)^{-1}X')\Omega]}{n-p-1}, \quad \textbf{(7.5)}$$

где символ tr означает след соответствующей матрицы (§ **11.2**).

Следовательно, если в качестве оценки ковариационной матрицы $\sum_b$ в соотношении (**7.3**) заменить σ^2 на s^2, т. е. взять матрицу $\sum_b = s^2(X'X)^{-1}$, то ее математическое ожидание

$$M\left(\hat{\Sigma}_b\right) = M\left(s^2\right)(X'X)^{-1} = \frac{\mathrm{tr}[(E_n - X(X'X)^{-1}X')\Omega]}{n-p-1}\left(X'X\right)^{-1} \quad \textbf{(7.6)}$$

в общем случае не совпадает с ковариационной матрицей, определенной соотношением (**7.2**). Это означает, что *обычный метод наименьших квадратов в обобщенной линейной регрессионной модели дает* ***смещенную*** *оценку ковариационной матрицы* $\sum_b$ *вектора оценок b.*

Оценка *b*, определенная по (**4.8**), хотя и будет *состоятельной,* но не будет *оптимальной* в смысле теоремы Гаусса—Маркова. Для получения наиболее *эффективной* оценки нужно использовать другую оценку, получаемую так называемым обобщенным методом наименьших квадратов.

7.2. Обобщенный метод наименьших квадратов

Вопрос об эффективности линейной несмещенной оценки вектора β для обобщенной регрессионной модели решается с помощью следующей теоремы.

Теорема Айткена. *В классе линейных несмещенных оценок вектора* β *для обобщенной регрессионной модели оценка*

$$b^* = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y \quad \textbf{(7.7)}$$

имеет наименьшую ковариационную матрицу.

Доказательство. Убедимся в том, что оценка b^* является несмещенной. Учитывая (**7.1**), представим ее в виде:

$$\begin{aligned} b^* &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\left(X\beta + \varepsilon\right) = \\ &= (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}X)\beta + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\varepsilon = \\ &= \beta + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\varepsilon. \end{aligned} \quad \textbf{(7.8)}$$

Математическое ожидание оценки b^*, т. е. $M(b^*)=\beta$, ибо $M(\varepsilon)=0$, т. е. оценка b^* есть *несмещенная* оценка β.

Для доказательства оптимальных свойств оценки b^* преобразуем исходные данные — матрицу X, вектор Y и возмущение ε к виду, при котором выполнены требования *классической* модели регрессии.

Из матричной алгебры известно, что всякая невырожденная симметричная ($n\times n$) матрица A допускает представление в виде $A=PP'$, где P — некоторая невырожденная ($n\times n$) матрица.

Поэтому существует такая невырожденная ($n\times n$) матрица P, что

$$\Omega = PP' \qquad \textbf{(7.8′)}$$

(представление матрицы Ω в виде (**7.8′**) не единственно, но для нас это не имеет значения).

Учитывая свойства обратных квадратных матриц, т. е. $(AB)^{-1}=B^{-1}A^{-1}$ и $(P')^{-1}=(P^{-1})'$, это означает, что

$$\Omega^{-1}=(P^{-1})'P^{-1}. \qquad \textbf{(7.9)}$$

Заметим, что если обе части равенства (**7.8′**) умножить слева на матрицу P^{-1}, а справа — на матрицу $(P')^{-1}=(P^{-1})'$, то в произведении получим единичную матрицу.

Действительно,

$$P^{-1}\Omega(P')^{-1}=P^{-1}(PP')(P')^{-1}=(P^{-1}P)P'(P')^{-1}=E_n,$$

т. е.

$$P^{-1}\Omega(P^{-1})'=E_n. \qquad \textbf{(7.10)}$$

Теперь, умножив обе части обобщенной регрессионной модели $Y=X\beta+\varepsilon$ на матрицу P^{-1} слева, получим

$$Y_*=X_*\beta+\varepsilon_*, \qquad \textbf{(7.11)}$$

где

$$Y_*=P^{-1}Y,\ X_*=P^{-1}X,\ \varepsilon_*=P^{-1}\varepsilon. \qquad \textbf{(7.12)}$$

Убедимся в том, что модель (**7.11**) удовлетворяет всем требованиям *классической линейной модели множественной регрессии* (§ **4.2**):

$$M(\varepsilon_*)=M(P^{-1}\varepsilon)=P^{-1}M(\varepsilon)=0,\ \text{ибо}\ M(\varepsilon)=0;$$

$$\sum_{\varepsilon_*}=M(\varepsilon_*\varepsilon_*')=M\left[(P^{-1}\varepsilon)(P^{-1}\varepsilon)'\right]=M\left[P^{-1}\varepsilon\varepsilon'(P^{-1})'\right]=$$

$$=P^{-1}M(\varepsilon\varepsilon')(P^{-1})'=P^{-1}\Omega(P^{-1})'=E_n$$

(учитывая (**7.10**));

$r(X)=p+1<n$ (так как матрица P — невырожденная).

Следовательно, на основании теоремы Гаусса—Маркова наиболее эффективной оценкой в классе всех линейных несмещенных оценок является оценка (**4.8**), т. е.

$$b^* = \left(X_*' X_*\right)^{-1} X_*' Y_* . \qquad \textbf{(7.13)}$$

Возвращаясь к исходным наблюдениям X и Y и учитывая (**7.9**), получим

$$b^* = [(P^{-1}X)'(P^{-1}X)]^{-1}(P^{-1}X)'P^{-1}Y =$$
$$= [X'(P^{-1})'P^{-1}X]^{-1}X'(P^{-1})'P^{-1}Y = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y,$$

т. е. выражение (**7.7**), что и требовалось доказать.

Нетрудно проверить, что в *случае классической модели*, т. е. при выполнении предпосылки $\sum_{\varepsilon} = \Omega = \sigma^2 E_n$, *оценка обобщенного метода наименьших квадратов* b^* (**7.7**) *совпадает с оценкой «обычного» метода* b (**4.8**).

При выполнении предпосылки *5* о нормальном законе распределения вектора возмущений ε можно убедиться в том, что *оценка* b^* *обобщенного метода наименьших квадратов для параметра* β *при* ***известной матрице*** Ω *совпадает с его оценкой, полученной методом максимального правдоподобия.*

В соответствии с (**4.3**) оценка $b^* = \left(X_*' X_*\right)^{-1} X_*' Y_*$ является точкой минимума по b остаточной суммы квадратов

$$S = \sum_{i=1}^{n} e_{i*}^2 = e_*' e_* = \left(Y_* - X_* b\right)' \left(Y_* - X_* b\right).$$

Переходя к исходным наблюдениям,

$$S = [P^{-1}(Y - Xb]'[P^{-1}(Y - Xb)] =$$
$$= (Y - Xb)'(P^{-1})'P^{-1}\left(Y - Xb\right) = \left(Y - Xb\right)'\Omega^{-1}\left(Y - Xb\right) =$$
$$= e'\Omega^{-1}e, \qquad \textbf{(7.14)}$$

т. е. *оценка* b^* *обобщенного метода наименьших квадратов может быть определена как точка минимума обобщенного критерия* $e'\Omega^{-1}e$ (**7.14**).

Следует отметить, что для обобщенной регрессионной модели, в отличие от классической, коэффициент детерминации, вычисленный по формуле (**4.33′**)

$$R^2 = 1 - \frac{(Y - Xb^*)'(Y - Xb^*)}{(Y - \overline{Y})'(Y - \overline{Y})} \qquad (7.15)$$

(где b^* — оценка обобщенного метода наименьших квадратов (**7.7**)), не является удовлетворительной мерой качества модели. В общем случае R^2 может выходить даже за пределы интервала [0;1], а добавление (удаление) объясняющей переменной не обязательно приводит к его увеличению (уменьшению).

Причина состоит в том, что разложение общей суммы квадратов Q на составляющие Q_R и Q_e (см. § **4.6**) выводилось в предположении наличия свободного члена в обобщенной модели. Однако, если в исходной модели (**7.1**) содержится свободный член, то мы не можем гарантировать его присутствие в преобразованной модели (**7.11**). Поэтому *коэффициент детерминации R^2 в обобщенной модели может использоваться лишь как весьма приближенная характеристика качества модели.*

В заключение отметим, что для применения обобщенного метода наименьших квадратов необходимо знание ковариационной матрицы вектора возмущений Ω, что встречается крайне редко в практике эконометрического моделирования. Если же считать все $n(n+1)/2$ элементов симметричной ковариационной матрицы Ω неизвестными параметрами обобщенной модели (в дополнении к $(p+1)$ параметрам β_i), то общее число параметров значительно превысит число наблюдений n, что сделает оценку этих параметров неразрешимой задачей. Поэтому для практической реализации обобщенного метода наименьших квадратов необходимо вводить дополнительные условия на структуру матрицы Ω. Так мы приходим к *практически реализуемому* (или *доступному*) *обобщенному* методу наименьших квадратов, рассматриваемому в § **7.11.**

Далее рассмотрим наиболее важные и часто встречающиеся виды структур матрицы Ω.

7.3. Гетероскедастичность пространственной выборки

Как уже отмечалось выше, равенство дисперсий возмущений (ошибок) регрессиии ε_i (*гомоскедастичность*) является существенным условием линейной классической регрессионной модели множественной регрессии, записываемым в виде $\sum_{\varepsilon} = \sigma^2 E_n$.

Однако на практике это условие нередко нарушается, и мы имеем дело с *гетероскедастичностью* модели.

Предположим, что необходимо изучить зависимость размера оплаты труда Y (в усл. ден. ед.) сотрудников фирмы от разряда X, принимающего значения от 1 до 10. Получены $n = 100$ пар наблюдений (x_i, y_i). График зависимости переменной Y от номеров наблюдений, упорядоченных по возрастанию уровня значений объясняющей переменной X, показан на рис. 7.1.

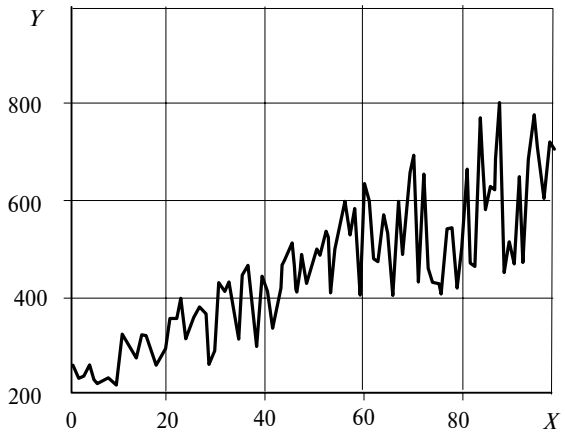


Рис. 7.1

Из рис. 7.1 отчетливо видно, что вариация размера оплаты труда сотрудников высоких уровней значительно превосходит его вариацию для сотрудников низких уровней. Следовательно, мы вправе предположить, что регрессионная модель получится гетероскедастичной, и условие $\sum_{\varepsilon} = \sigma^2 E_n$ не выполняется.

Мы еще вернемся к этому примеру, а пока обсудим, к каким последствиям приводит гетероскедастичность.

Предположим, что для оценки регрессионной модели Y по $X_1,.., X_n$ мы применили обычный метод наименьших квадратов и нашли оценку b параметра β по формуле (**4.8**). Тогда с учетом (**4.12**) будем иметь

$$b = (X'X)^{-1} X'Y = \beta + (X'X)^{-1} X'\varepsilon . \qquad \textbf{(7.16)}$$

Как было отмечено в § **7.1**, b — несмещенная и состоятельная оценка параметра β для *обобщенной* линейной модели множественной регрессии; следовательно, и в частном случае, когда мо-

дель гетероскедастична, оценка b — несмещенная и состоятельная. Эти свойства оценки b легко усматриваются из (**7.16**), если учесть, что $M(\varepsilon)=0$.

Таким образом, для определения неизвестных (*прогнозных*) значений зависимой переменной обычный метод наименьших квадратов, вообще говоря, применим и для гетероскедастичной модели.

Так, в нашем примере изучения зависимости размера оплаты Y от разряда X сотрудников фирмы регрессионная модель Y по X примет вид:

$$\hat{y} = 225,2 + 44,99x ,$$

которая вполне может быть использована для практических приложений.

Однако результаты, связанные с анализом *точности* модели, оценкой *значимости* и построением *интервальных оценок* ее коэффициентов, оказываются непригодными.

В самом деле, при построении t- и F-статистик, которые служат инструментом для проверки (тестирования) гипотез, существенное значение имеют оценки дисперсий и ковариаций параметров β_i ($i = 1,..., n$), т. е. ковариационная матрица $\sum_b$. Между тем, если модель не является классической, т. е. ковариационная матрица вектора возмущений $\sum_\varepsilon = \Omega \neq \sigma^2 E_n$, то, как показано в § **7.1**, ковариационная матрица вектора оценок параметров $\sum_b = (X'X)^{-1} X'\Omega X (X'X)^{-1}$ (**7.2**) существенно отличается от полученной для классической модели $\sum_b = \sigma^2 (X'X)^{-1}$ (**7.3**). А значит, использование матрицы $\sum_b$ (**7.2**) для оценки точности регрессионной модели (**7.1**) может привести к неверным выводам.

Напомним также (§ **7.1**), что оценка b (**7.16**), оставаясь несмещенной и состоятельной, не будет оптимальной в смысле теоремы Гаусса—Маркова, т. е. наиболее эффективной. Это означает, что при небольших выборках мы рискуем получить оценку b, существенно отличающуюся от истинного параметра β.

7.4. Тесты на гетероскедастичность

В примере, рассмотренном в § **7.3**, наличие гетероскедастичности не вызывает сомнения, — чтобы убедиться в этом, доста-

точно взглянуть на рис. 7.1. Однако в некоторых случаях гетероскедастичность визуально не столь очевидна.

Рассмотрим еще один пример, в котором исследуется зависимость дохода индивидуума (Y) от уровня его образования X_1, принимающего значения от 1 до 5, по данным $n = 150$ наблюдений. В число объясняющих переменных (регрессоров) включен также и возраст X_2.

На рис. 7.2 приведен график зависимости переменной Y от номеров наблюдений, упорядоченных по возрастанию уровня значений объясняющей переменной X_1.

Хотя диаграмма имеет локально расположенные пики, в целом подобный рисунок может соответствовать как гомо-, так и гетероскедастичной выборке.

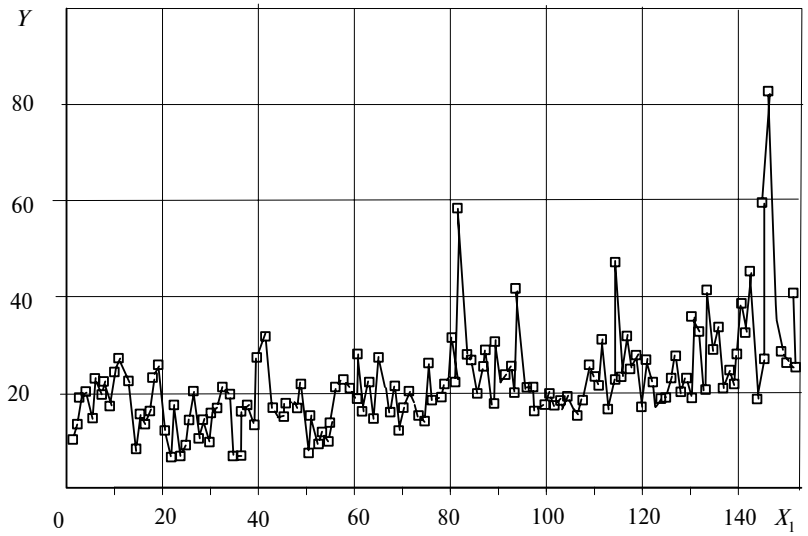


Рис. 7.2

Чтобы определить, какая же именно ситуация имеет место, используются *тесты на гетероскедастичность*. Все они используют в качестве нулевой гипотезы H_0 *гипотезу об отсутствии гетероскедастичности*.

Тест ранговой корреляции Спирмена использует наиболее общие предположения о зависимости дисперсий ошибок регрессии от значений регрессоров:

$$\sigma^2 = f_i(x_i), \ i = 1,..., n.$$

При этом никаких дополнительных предположений относительно вида функций f_i не делается. Не накладываются также ограничения на закон распределения возмущений (ошибок) регрессии ε_i.

Идея теста заключается в том, что абсолютные величины остатков регрессии e_i являются оценками σ_i, поэтому в случае гетероскедастичности абсолютные величины остатков e_i и значения регрессоров x_i будут коррелированы.

Для нахождения коэффициента ранговой корреляции $\rho_{x,e}$ (см. § **3.8**) следует ранжировать наблюдения по значениям переменной x_i и остатков e_i и вычислить $\rho_{x,e}$ по формуле (**3.49**):

$$\rho_{x,e} = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n^3 - n}, \qquad \textbf{(7.17)}$$

где d_i — разность между рангами значений x_i и e_i.

В соответствии с (**3.50**) коэффициент ранговой корреляции значим на уровне значимости α при $n > 10$, если статистика

$$|t| = \frac{|\rho_{x,e}|\sqrt{n-2}}{\sqrt{1-\rho_{x,e}^2}} > t_{1-\alpha;n-2}, \qquad \textbf{(7.18)}$$

где $t_{1-\alpha;n-2}$ — табличное значение t-критерия Стьюдента, определенное на уровне значимости α при числе степеней свободы (n−2).

Тест Голдфелда—Квандта. Этот тест применяется в том случае, если ошибки регрессии можно считать *нормально распределенными* случайными величинами.

Предположим, что средние квадратические (стандартные) отклонения возмущений σ_i пропорциональны значениям объясняющей переменной X (это означает постоянство часто встречающегося на практике *относительного* (а не абсолютного, как в классической модели) разброса возмущений ε_i регрессионной модели.

Упорядочим n наблюдений в порядке возрастания значений регрессора X и выберем m первых и m последних наблюдений.

В этом случае гипотеза о гомоскедастичности будет равносильна тому, что значения $e_1,\ldots, e_m$ и $e_{n-m+1},\ldots, e_n$ (т. е. остатки e_i регрессии первых и последних m наблюдений) представляют собой выборочные наблюдения *нормально распределенных* случайных величин, имеющих одинаковые дисперсии.

Гипотеза о равенстве дисперсий двух нормально распределенных совокупностей, как известно (см., например, [17]), проверяется с помощью критерия Фишера—Снедекора.

Нулевая гипотеза о равенстве дисперсий двух наборов по m наблюдений (т. е. гипотеза об отсутствии гетероскедастичности) отвергается, если

$$F = \frac{\sum_{i=1}^{m} e_i^2}{\sum_{i=n-m+1}^{n} e_i^2} > F_{\alpha;m-p;m-p}, \qquad \textbf{(7.19)}$$

где p — число регрессоров.

Заметим, что числитель и знаменатель в выражении (**7.19**) следовало разделить на соответствующее число степеней свободы, но в данном случае эти числа одинаковы и равны $(m - p)$.

Мощность теста, т. е. вероятность отвергнуть гипотезу об отсутствии гетероскедастичности, когда действительно гетероскедастичности нет, оказывается максимальной, если выбирать m порядка $n/3$.

При применении теста Голдфелда—Квандта на компьютере нет необходимости вычислять значение статистики F вручную, так как величины $\sum_{i=1}^{m} e_i^2$ и $\sum_{i=n-m+1}^{m} e_i^2$ представляют собой суммы квадратов остатков регрессии, осуществленных по «урезанным» выборкам.

▶ **Пример 7.1**. По данным $n = 150$ наблюдений о доходе индивидуума Y (рис. 7.2), уровне его образования X_1 и возрасте X_2 выяснить, можно ли считать на уровне значимости α=0,05 линейную регрессионную модель Y по X_1 и X_2 гетероскедастичной.

Р е ш е н и е. Возьмем по $m=n/3=150/3=50$ значений доходов лиц с наименьшим и наибольшим уровнем образования X_1.

Вычислим суммы квадратов остатков (само уравнение регрессии (**7.22**) приведено ниже)[1]:

$$\sum_{i=1}^{150} e_i^2 = 894{,}1; \ \sum_{i=101}^{150} e_i^2 = 3918{,}2; \ \ F=3918{,}2/894{,}1=4{,}38.$$

[1] Здесь и далее (если не приведено иное) исходные значения переменных и выполненные на их основе компьютерной программой стандартные расчеты, изложенные в гл. **3,4**, не приводятся.

Так как в соответствии с (**7.19**) $F=4{,}38>F_{0{,}05;48;48}=1{,}61$, то гипотеза об отсутствии гетероскедастичности регрессионной модели отвергается, т. е. доходы более образованных людей действительно имеют существенно бóльшую вариацию. ▶

Тест Уайта. Тест ранговой корреляции Спирмена и тест Голдфелда—Квандта позволяют обнаружить лишь само наличие гетероскедастичности, но они не дают возможности проследить количественный характер зависимости дисперсий ошибок регрессии от значений регрессоров и, следовательно, не представляют каких-либо способов устранения гетероскедастичности.

Очевидно, для продвижения к этой цели необходимы некоторые дополнительные предположения относительно характера гетероскедастичности. В самом деле, без подобных предположений, очевидно, невозможно было бы оценить n параметров (n дисперсий ошибок регрессии σ_i^2) с помощью n наблюдений.

Наиболее простой и часто употребляемый тест на гетероскедастичность — тест Уайта. При использовании этого теста предполагается, что дисперсии ошибок регрессии представляют собой о д н у и т у ж е ф у н к ц и ю от наблюдаемых значений регрессоров, т.е.

$$\sigma_i^2 = f(x_i),\ i = 1,\ldots, n. \qquad \textbf{(7.20)}$$

Чаще всего функция f выбирается квадратичной, что соответствует тому, что средняя квадратическая ошибка регрессии зависит от наблюдаемых значений регрессоров приближенно линейно. Гомоскедастичной выборке соответствует случай $f=\text{const}$.

Идея теста Уайта заключается в оценке функции (**7.20**) с помощью соответствующего уравнения регрессии для квадратов остатков:

$$e_i^2 = f(x_i)+u_i,\ i = 1,\ldots, n, \qquad \textbf{(7.21)}$$

где u_i — случайный член.

Гипотеза об отсутствии гетероскедастичности (условие $f=\text{const}$) принимается в случае незначимости регрессии (**7.21**) в целом.

В большинстве современных пакетов, таких, как «*Econometric Views*», регрессию (**7.21**) не приходится осуществлять вручную — тест Уайта входит в пакет как стандартная подпрограмма. В этом случае функция f выбирается *квадратичной,* регрессоры в (**7.21**) — это регрессоры рассматриваемой модели, их квадраты и, возможно, попарные произведения.

▶ **Пример 7.2.** Решить пример **7.1**, используя тест Уайта.

Р е ш е н и е. Применение метода наименьших квадратов дает следующее уравнение регрессии переменной Y (дохода индивидуума) по X_1 (уровню образования) и X_2 (возрасту):

$$\hat{y} = -3,06 + 3,25x_1 + 0,48x_2. \quad \textbf{(7.22)}$$
$$(-1,40)\ (5,96)\quad (8,35)$$

(В скобках указаны значения t-статистик коэффициентов регрессии.) Сравнивая их с табличным значением (**4.23**), т. е. $t_{0.95:147}$=1,98, видим, что константа оказывается незначимой.

Обращение к программе *White Heteroskedascity Test* (*Тест Уайта на гетероскедастичность*) дает следующие значения F-статистики: F= 7,12, если в число регрессоров уравнения (**7.21**) не включены попарные произведения переменных, и F = 7,78 — если включены. Так как в соответствии с (**4.32**) и в том и другом случае $F > F_{0.05:2:147}$=3,07, т. е. гипотеза об отсутствии гетероскедастичности отвергается. ▶

Заметим, что на практике применение теста Уайта с включением и невключением попарных произведений дают, как правило, один и тот же результат.

Тест Глейзера. Этот тест во многом аналогичен тесту Уайта, только в качестве зависимой переменной для изучения гетероскедастичности выбирается не квадрат остатков, а их абсолютная величина, т. е. осуществляется регрессия

$$|e_i| = f(x_i) + u_i,\ i=1,\ldots,\ n. \quad \textbf{(7.23)}$$

В качестве функций f обычно выбираются функции вида $f = \alpha + \gamma x^{\delta}$. Регрессия (**7.23**) осуществляется при разных значениях δ, затем выбирается то значение, при котором коэффициент γ оказывается наиболее значимым, т. е. имеет наибольшее значение t-статистики.

▶ **Пример 7.3.** По данным n = 100 наблюдений о размере оплаты труда Y (рис. 5.1) сотрудников фирмы и их разряде X выявить, можно ли считать на уровне значимости α линейную регрессионую модель Y по X гетероскедастичной. Если модель гетероскедастична, то установить ее характер, оценив уравнение $\sigma_i = f(x_i)$.

Р е ш е н и е. Предположим, что дисперсии ошибок σ_i связаны уравнением регрессии

$$\sigma_i = \alpha + \gamma x_i^{\delta}. \qquad \textbf{(7.24)}$$

Используя обычный метод наименьших квадратов, оценим регрессию Y по X, а затем — регрессию остатков e по X в виде функции (**7.24**) при различных значениях δ. Получим (в скобках указаны значения t-статистики коэффициента γ) при различных значениях δ:

$$\delta=1 \quad |\hat{e}_i| = 8,26 + 10,33x_i \quad (t = 7,18);$$

$$\delta=2 \quad |\hat{e}_i| = 30,75 + 0,89x_i^2 \quad (t = 6,90);$$

$$\delta=3 \quad |\hat{e}_i| = 39,89 + 0,08x_i^3 \quad (t = 6,32);$$

$$\delta=1/2 \quad |\hat{e}_i| = 32,89 + 43,38\sqrt{x_i} \quad (t = 6,99).$$

Так как все значения t-статистики больше $t_{0.95;98}$=1,99, то гипотеза об отсутствии гетероскедастичности отвергается. Учитывая, что наиболее значимым коэффициент регрессии γ оказывается в случае δ=1, гетероскедастичность можно аппроксимировать первым уравнением. ▶

7.5. Устранение гетероскедастичности

Пусть рассматривается регрессионная модель (**4.2**)

$$Y = X\beta + \varepsilon \qquad \textbf{(7.25)}$$

или

$$y_i = \beta_0 + \sum_{i=1}^{p} \beta_j x_{ij} + \varepsilon_i, \ i = 1,..., n. \qquad \textbf{(7.25′)}$$

Будем считать, что модель (**7.25**) г е т е р о с к е д а с т и ч н а, т. е. дисперсии возмущений (ошибок) $\sigma_i^2 \ (i = 1,..., n)$ не равны между собой, и сами возмущения ε_i и ε_k (k = 1,..., n) не коррелированы. Это означает, что ковариационная матрица вектора возмущений $\sum_{\varepsilon} = \Omega$ — *диагональная*:

$$\Omega = \begin{pmatrix} \sigma_1^2 & 0 & ... & 0 \\ 0 & \sigma_2^2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & \sigma_n^2 \end{pmatrix}. \qquad \textbf{(7.26)}$$

Если дисперсии возмущений σ_i^2 $(i=1,...,n)$ известны, то гетероскедастичность легко устраняется. В самом деле, будем рассматривать в качестве *i*-го наблюдения зависимой Y и объясняющих переменных $X_j\,(j=1,...,p)$ нормированные по σ_i переменные, т. е.

$$Z = Y/\sigma_i,\ \ V_j = X_j/\sigma_i,\ \ \ i = 1,...,\ n.$$

Тогда модель (**7.25**) примет вид:

$$z_i = \beta_0' + \sum_{j=1}^{P} \beta_j v_{ij} + \nu_i, \quad i = 1,...,\ n, \qquad \textbf{(7.27)}$$

где $\beta_0' = \beta_0/\sigma_i,\ \nu_i = \varepsilon_i/\sigma_i$.

Очевидно, дисперсия $D(\nu_i)=1$, т. е. модель (**7.27**) гомоскедастична. При этом ковариационная матрица $\sum_{\varepsilon}\Omega$ становится единичной, а сама модель (**7.27**) — классической.

Применяя к линейной регрессионной модели (**7.25**) теорему Айткена (§ **7.2**), наиболее эффективной оценкой вектора β является оценка (**7.7**):

$$b^* = \left(X'\Omega^{-1}X\right)^{-1}X'\Omega Y\,. \qquad \textbf{(7.28)}$$

Применение формулы (**7.28**) для отыскания параметра β, т. е. *обобщенный метод наименьших квадратов для модели с гетероскедастичностью,* когда ковариационная матрица возмущений $\sum_{\varepsilon} = \Omega$ есть диагональная матрица (**7.26**), *называется* ***взвешенным методом наименьших квадратов.***

Применяя *обычный* метод наименьших квадратов (§ **4.2**), неизвестные параметры регрессионной модели находим, минимизируя остаточную сумму квадратов $S = e'e = \sum_{i=1}^{n}(\hat{y}_i - y_i)^2$, используя *обобщенный* метод (§ **7.2**), — минимизируя $S = e'\Omega^{-1}e$, и, наконец, в частном случае, применяя *взвешенный* метод наименьших квадратов, — минимизируя $S = \sum_{i=1}^{n}\left[\frac{1}{\sigma_i}\left(\hat{y}_i - y_i\right)^2\right]$.

«Взвешивая» каждый остаток $e_i = \hat{y}_i - y_i$ с помощью коэффициента $1/\sigma_i$, мы добиваемся *равномерного вклада* остатков в общую сумму, что приводит в конечном счете к получению наиболее эффективных оценок параметров модели.

На практике, однако, значения σ_i почти никогда не бывают известны. В этом случае при нахождении переменных в формуле (**7.27**) значения σ_i следует заменить их состоятельными оценками $\hat{\sigma}_i$.

Если исходить из предположения (**7.20**), то состоятельными оценками σ_i^2 являются объясненные (прогнозные) значения $\hat{e}_i^2$ регрессии (**7.21**).

Оценка параметров регрессионной модели взвешенным методом наименьших квадратов реализована в большинстве компьютерных пакетов. Покажем ее проведение при использовании пакета «*Econometric Views*».

Сначала следует применить обычный метод наименьших квадратов к модели (**7.25**), затем надо найти регрессию квадратов остатков на квадратичные функции регрессоров, т. е. найти уравнение регрессии (**7.21**), где f — квадратичная функция, аргументами которой являются квадраты значений регрессоров и их попарные произведения. После чего следует вычислить прогнозные значения $\hat{e}_i^2$ по полученному уравнению регрессии и получить набор весов («*weight*»): $\hat{\sigma}_i = \sqrt{\hat{e}_i^2}$. Затем надо ввести новые переменные $X_{*j} = X_j / \hat{\sigma}_i \left(j = 1,\ldots,p\right)$, $Y_{*i} = Y_i / \hat{\sigma}_i$ и найти уравнение $\hat{Y}_* = X_* b$. Полученная при этом оценка b^* и есть оценка взвешенного метода наименьших квадратов исходного уравнения (**7.25**).

▶ **Пример 7.4.** По данным примера **7.1** оценить параметры регрессионной модели Y по X_1 и X_2 взвешенным методом наименьших квадратов.

Р е ш е н и е. В примере **7.2** к модели был применен обычный метод наименьших квадратов. При этом получен ряд остатков e_i.

Оценим теперь регрессию вида

$$\hat{e}_i^2 = \gamma_0 + \gamma_1 x_1^2 + \gamma_2 x_2^2 + \gamma_3 x_1 x_2 .$$

Применяя обычный метод наименьших квадратов, получим уравнение[1]:

$$\hat{e}_i^2 = 3{,}6 + 0{,}3x_1^2 + 0{,}1x_2^2 + 0{,}05x_1x_2 .$$
$$(0{,}1)\quad (0{,}07)\quad (0{,}1)$$

[1] Здесь и далее в скобках под коэффициентами регрессии указываются их средние квадратические (стандартные) отклонения.

Для применения взвешенного метода наименьших квадратов рассмотрим величины $\hat{\sigma}_i = \sqrt{\hat{e}_i^2}$ и введем новые переменные

$$X_{*j} = \frac{X_j}{\hat{\sigma}_i} \ \left(j = 1,2\right), \ Y_{i*} = \frac{Y_i}{\hat{\sigma}_i} \ \left(i = 1, ..., 150\right).$$

Оценивая регрессию Y_* по X_{*1} и X_{*2} получаем уравнение:

$$\hat{y}_* = -6,21 + 3,58x_{*1} + 0,53x_{*2},$$
$$(2,18)\ (0,58) \quad (0,08)$$

что и дает нам оценки взвешенного метода наименьших квадратов.

Если применить тест Уайта к последнему уравнению, получим $F = 0,76 < F_{0,05;2;147} = 3,06,$ откуда следует, что гетероскедастичность можно считать устраненной. ▶

На практике процедура устранения гетероскедастичности может представлять технические трудности. Дело в том, что реально в формулах (**7.26**) присутствуют не сами стандартные отклонения ошибок регрессии, а лишь их оценки. А это значит, что модель (**7.27**) вовсе не обязательно окажется гомоскедастичной.

Причины этого очевидны. Во-первых, далеко не всегда оказывается справедливым само предположение (**7.21**) или (**7.23**). Во-вторых, функция f в формуле (**7.21**) или (**7.23**), вообще говоря, не обязательно степенная (и уж тем более, не обязательно квадратичная), и в этом случае ее подбор может оказаться далеко не столь простым.

Другим недостатком тестов Уайта и Глейзера является то, что факт невыявления ими гетероскедастичности, вообще говоря, не означает ее о т с у т с т в и я. В самом деле, принимая гипотезу H_0, мы принимаем лишь тот факт, что отсутствует определенного вида зависимость дисперсий ошибок регрессии от значений регрессоров.

Так, если применить к рассматриваемой ранее модели зависимости дохода Y от разряда X взвешенный метод наименьших квадратов, используя уравнение (**7.23**) с линейной функцией f, то получим уравнение $\hat{y} = 196,47 + 50,6x$ и коэффициент детерминации $R^2 = 0,94$.

Если теперь использовать тест Глейзера для проверки отсутствия гетероскедастичности «взвешенного» уравнения, то соответствующая гипотеза подтвердится.

Однако, если для этой же цели применить тест Голдфелда—Квандта, то получим: $\sum_{i=1}^{33} e_i^2 = 26{,}49$, $\sum_{i=68}^{100} e_i^2 = 49{,}03$, $F = 1{,}85$.

Сравнивая с $F_{0.05:32:32}$=1,84, делаем вывод о том, что на 5%-ном уровне значимости гипотеза об отсутствии гетероскедастичности все же отвергается, хотя и вычисленное значение *F*-статистики очень близко к критическому.

Однако, даже если с помощью взвешенного метода наименьших квадратов не удается устранить гетероскедастичность, ковариационная матрица $\sum_{b*}$ оценок параметров регрессии β все же может быть состоятельна оценена (напомним, что именно несостоятельность стандартной оценки дисперсий и ковариаций β является наиболее неприятным последствием гетероскедастичности, в результате которого оказываются недостоверными результаты тестирования основных гипотез). Соответствующая оценка имеет вид:

$$\hat{\sum}_{b*} = n(X'X)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} e_i^2 x_i^2\right)(X'X)^{-1}.$$

Стандартные отклонения, вычисленные по этой формуле, называются *стандартными ошибками в форме Уайта.*

Так, для рассматриваемого примера зависимости дохода *Y* от разряда *X* стандартная ошибка в форме Уайта равна 2,87, в то время как ее значение, рассчитанное с помощью обычного метода наименьших квадратов, равно 2,96.

7.6. Автокорреляция остатков временного ряда. Положительная и отрицательная автокорреляция

Рассмотрим регрессионную модель *временного* (*динамического*) ряда.

Упорядоченность наблюдений оказывается существенной в том случае, если прослеживается механизм влияния результатов предыдущих наблюдений на результаты последующих. Математически это выражается в том, что случайные величины ε_i в регрессионной модели не оказываются независимыми, в частности, условие $r(\varepsilon_i, \varepsilon_j) = 0$ не выполняется. Такие модели называются *моделями с наличием автокорреляции* (*сериальной корреляции*), на практике ими оказываются именно

временные ряды (напомним, что в случае пространственной выборки отсутствие автокорреляции п о с т у л и р у е т с я).

Рассмотрим в качестве примера временной ряд y_t — ряд последовательных значений курса ценной бумаги *A*, наблюдаемых в моменты времени 1, ..., 100. Результаты наблюдений графически изображены на рис. 7.3.

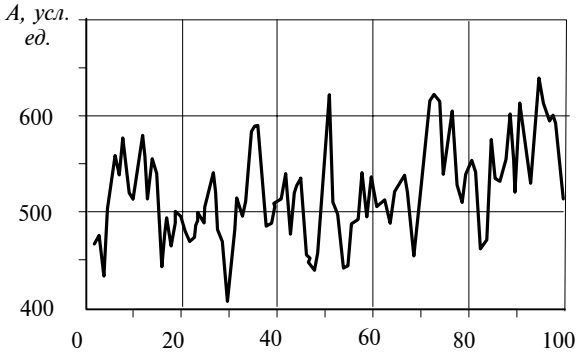


Рис. 7.3

Очевидно, курс ценной бумаги *A* имеет тенденцию к росту, что можно проследить на графике.

Оценивая обычным методом наименьших квадратов зависимость курса от номера наблюдений (т. е. от времени), получим следующие результаты:

$$y_t = \underset{(9,85)}{490,4247} + \underset{(0,17)}{0,7527t};\ R^2 = 0,17. \qquad \textbf{(7.29)}$$

Насколько достоверны эти уравнения? Очевидно, естественно предположить, что результаты предыдущих торгов оказывают влияние на результаты последующих: если в какой-то момент курс окажется завышенным по сравнению с реальным, то скорее всего он будет завышен на следующих торгах, т. е. имеет место *положительная автокорреляция*.

Графически положительная автокорреляция выражается в чередовании зон, где наблюдаемые значения оказываются выше объясненных (предсказанных), и зон, где наблюдаемые значения ниже.

Так, на рис. 7.4 представлены графики наблюдаемых значений y_t и объясненных, сглаженных $\hat{y}_t$.

Отрицательная автокорреляция встречается в тех случаях, когда наблюдения действуют друг на друга по принципу «маятника» — завышенные значения в предыдущих наблюдениях приводят к занижению их в наблюдениях последующих. Графически это выражается в том, что результаты наблюдений y_t «слишком часто» «перескакивают» через график объясненной части $\hat{y}_t$. Примерное поведение графика наблюдаемых значений временного ряда изображено на рис. 7.5.

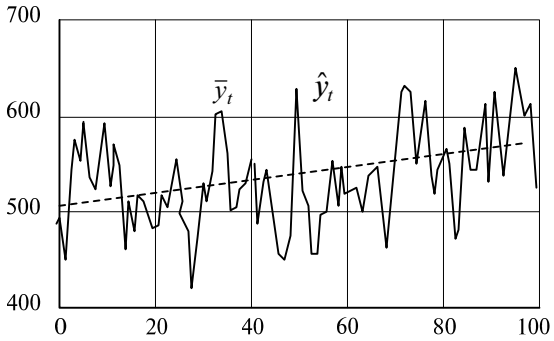


Рис. 7.4

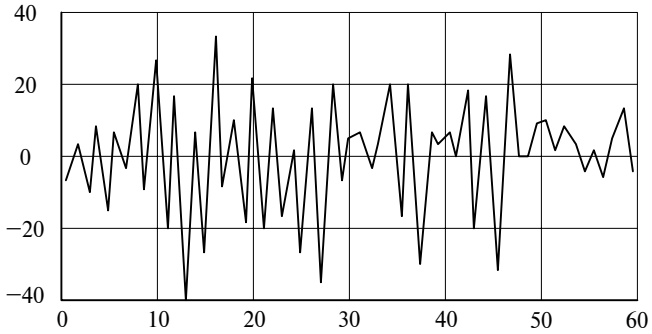


Рис. 7.5

Как и в случае любой обобщенной модели множественной регрессии, метод наименьших квадратов при наличии коррелированности ошибок регрессии дает несмещенные и состоятельные (хотя, разумеется, неэффективные) оценки коэффициентов

регрессии, однако, как уже было отмечено выше, оценки их дисперсий несостоятельные и смещенные (как правило, в сторону занижения), т. е. результаты тестирования гипотез оказываются недостоверными.

7.7. Авторегрессия первого порядка. Статистика Дарбина—Уотсона

Как правило, если автокорреляция присутствует, то наибольшее влияние на последующее наблюдение оказывает результат предыдущего наблюдения. Так, например, если рассматривается ряд значений курса какой-либо ценной бумаги, то, очевидно, именно результат последних торгов служит отправной точкой для формирования курса на следующих торгах.

Ситуация, когда на значение наблюдения y_t оказывает основное влияние не результат y_{t-1}, а более ранние значения, является достаточно редкой. Чаще всего при этом влияние носит сезонный (циклический) характер, — например, на значение y_t оказывает наибольшее влияние y_{t-7}, если наблюдения осуществляются ежедневно и имеют недельный цикл (например, сбор кинотеатра). В этом случае можно составить ряды наблюдений отдельно по субботам, воскресеньям и так далее, после чего наиболее сильная корреляция будет наблюдаться между соседними членами.

Таким образом, отсутствие корреляции между соседними членами служит хорошим основанием считать, что корреляция отсутствует в целом, и обычный метод наименьших квадратов дает адекватные и эффективные результаты.

Тест Дарбина—Уотсона. Этот простой *критерий* (*тест*) определяет наличие автокорреляции между соседними членами.

Тест Дарбина—Уотсона основан на простой идее: если корреляция ошибок регрессии не равна нулю, то она присутствует и в остатках регрессии e_t, получающихся в результате применения обычного метода наименьших квадратов. В тесте Дарбина—Уотсона для оценки корреляции используется статистика вида

$$d = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}.$$

Несложные вычисления позволяют проверить, что статистика Дарбина—Уотсона следующим образом связана с выборочным коэффициентом корреляции между соседними наблюдениями:

$$d \approx 2(1-r). \qquad \textbf{(7.30)}$$

В самом деле

$$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} = \frac{\sum_{t=2}^{n} e_t^2 + \sum_{t=2}^{n} e_{t-1}^2 - 2\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2} =$$

$$= \frac{\sum_{t=1}^{n} e_t^2 - e_1^2 + \sum_{t=1}^{n} e_t^2 - e_n^2 - 2\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2} = 2\left[1 - \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2}\right] - \frac{e_1^2 + e_n^2}{\sum_{t=1}^{n} e_t^2}.$$

При большом числе наблюдений n сумма $e_1^2 + e_n^2$ значительно меньше $\sum_{t=1}^{n} e_t^2$ и

$$d \approx 2\left(1 - \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2}\right), \qquad \textbf{(7.30′)}$$

откуда и следует приближенное равенство (**7.30**), так как в силу условия $\sum_{t=1}^{n} e_t = 0$

$$\frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2} = r.$$

Естественно, что в случае отсутствия автокорреляции выборочный коэффициент r окажется не сильно отличающимся от нуля, а значение статистики d будет близко к двум. Близость наблюдаемого значения к нулю должна означать наличие положительной автокорреляции, к четырем — отрицательной. Хотелось бы получить соответствующие пороговые значения, кото-

рые присутствуют в статистических критериях и либо позволяют принять гипотезу, либо заставляют ее отвергнуть (см. § **2.8**). К сожалению, однако, такие пороговые (критические) значения однозначно указать невозможно.

Тест Дарбина—Уотсона имеет один существенный недостаток — распределение статистики d зависит не только от числа наблюдений, но и от значений регрессоров X_j ($j = 1, ..., p$). Это означает, что *тест Дарбина—Уотсона*, вообще говоря, *не представляет* собой *статистический критерий,* в том смысле, что нельзя указать критическую область, которая позволяла бы отвергнуть гипотезу об отсутствии корреляции, если бы оказалось, что в эту область попало наблюдаемое значение статистики d.

Однако существуют два пороговых значения $d_в$ и $d_н$, зависящие т о л ь к о от числа наблюдений, числа регрессоров и уровня значимости, такие, что выполняются следующие условия.

Если фактически наблюдаемое значение d:

а) $d_в < d < 4-d_в$, то гипотеза об отсутствии автокорреляции не отвергается (принимается);

б) $d_н < d < d_в$ или $4 - d_в < d < 4 - d_н$, то вопрос об отвержении или принятии гипотезы остается открытым (область неопределенности критерия);

в) $0 < d < d_н$, то принимается альтернативная гипотеза о положительной автокорреляции;

г) $4-d_н < d < 4$, то принимается альтернативная гипотеза об отрицательной автокорреляции.

Изобразим результат Дарбина—Уотсона графически:

Рис. 7.6

Для d-статистики найдены верхняя $d_в$ и нижняя $d_н$ границы на уровнях значимости $\alpha = 0{,}01$; 0,025 и 0,05.

В табл. V приложений приведены значения статистик $d_н$ и $d_в$ критерия Дарбина—Уотсона на уровне значимости $\alpha = 0{,}05$.

Недостатками критерия Дарбина—Уотсона является наличие области неопределенности критерия, а также то, что критиче-

ские значения d-статистики определены для объемов выборки не менее 15. Тем не менее тест Дарбина—Уотсона является наиболее употребляемым.

▶ **Пример 7.5.** Выявить на уровне значимости 0,05 наличие автокорреляции возмущений для временного ряда y_t по данным табл. 6.1.

Р е ш е н и е. В примере **6.2** получено уравнение тренда $\hat{y}_t = 181{,}32 + 25{,}679t$ (ед.). В табл. 7.1 приведен расчет сумм, необходимых для вычисления d-статистики.

Т а б л и ц а 7.1

t	y_t	$\tilde{y}_t$	$e_t = y_t - \tilde{y}_t$	e_{t-1}	$e_t e_{t-1}$	e_t^2
1	213	207,0	6,0	—	—	36,0
2	171	232,7	−61,7	6,0	−370,2	3806,9
3	291	258,4	32,6	−61,7	−2011,4	1062,8
4	309	284,0	25,0	32,6	815,0	625,0
5	317	309,7	7,3	25,0	182,5	53,3
6	362	335,4	26,6	7,3	194,2	707,6
7	351	361,1	−10,1	26,6	−268,7	102,0
8	361	386,8	−25,8	−10,1	260,6	665,6
$\sum_{t=1}^{8}$	—	—	—	—	−1198,0	7059,2

Теперь по формуле (7.**30′**) статистика

$$d \approx 2(1+1198{,}0/7059{,}2)=2{,}34.$$

По табл. V приложений при n=15 критические значения $d_н$=1,08; $d_в$=1,36, т. е. фактически найденное d=2,34 находится в пределах от $d_в$ до $4-d_в$ (1,36 < d < 2,64). Как уже отмечено, при n < 15 критических значений d-статистики в таблице нет, но судя по тенденции их изменений с уменьшением n, можно предполагать, что найденное значение останется в интервале ($d_в$; $4-d_в$), т. е. для рассматриваемого временного ряда спроса на уровне значимости 0,05 гипотеза об отсутствии автокорреляции возмущений не отвергается (принимается). ▶

При использовании компьютерных регрессионных пакетов значение статистики Дарбина—Уотсона приводится автоматически при оценивании модели методом наименьших квадратов.

Вернемся к рассмотренному в § **7.6** примеру зависимости курса ценной бумаги A от времени. Здесь статистика Дарбина—Уотсона $d = 0{,}993$, т. е. меньше единицы. Такое низкое ее значение выявляет наличие положительной корреляции между соседними наблюдениями. Заметим, что такой вывод предсказывался нами в § **7.6** на основании экономических соображений.

7.8. Тесты на наличие автокорреляции

Статистика Дарбина—Уотсона, безусловно, является наиболее важным индикатором наличия автокорреляции. Однако, как уже отмечалось, тест обладает и определенными недостатками. Это и наличие зоны неопределенности, и ограниченность результата (выявляется лишь корреляция между соседними членами). Ничего нельзя сказать и о характере автокорреляции.

Это приводит к необходимости использовать также и другие тесты на наличие автокорреляции. Во всех этих тестах в качестве основной гипотезы H_0 фигурирует гипотеза об отсутствии автокорреляции.

Некоторые из этих тестов мы рассмотрим в данном параграфе.

Тест серий (Бреуша—Годфри). Тест основан на следующей идее: если имеется корреляция между соседними наблюдениями, то естественно ожидать, что в уравнении

$$e_t = \rho e_{t-1} + \nu_t, \; t = 1, \ldots, n, \qquad \textbf{(7.31)}$$

(где e_t — остатки регрессии, полученные обычным методом наименьших квадратов), коэффициент ρ окажется значимо отличающимся от нуля.

Заметим, что уравнение (**7.31**) является авторегрессионным уравнением первого порядка (см. § **6.5**).

Практическое применение теста заключается в оценивании методом наименьших квадратов регрессии (**7.31**) (напомним, что временной ряд e_{t-1} представляет ряд e_t со сдвигом по времени на единицу: компьютерные регрессионные пакеты имеют команду, которая формирует по данному временному ряду e_t ряд e_{t-1}).

Преимущество теста Бреуша—Годфри по сравнению с тестом Дарбина—Уотсона заключается в первую очередь в том, что он проверяется с помощью статистического критерия, между тем как тест Дарбина—Уотсона содержит зону не-

определенности для значений статистики d. Другим преимуществом теста является возможность обобщения: в число регрессоров могут быть включены не только остатки с лагом 1, но и с лагом 2, 3 и т д., что позволяет выявить корреляцию не только между соседними, но и между более отдаленными наблюдениями.

Вернемся к примеру зависимости курса ценной бумаги A от времени и применим тест серий Бреуша—Годфри.

Рассмотрим авторегрессионную зависимость остатков от их предыдущих значений, используя авторегрессионную модель p-го порядка $AR(p)$. Применяя метод наименьших квадратов, получим следующее уравнение:

$$\underset{}{e_t} = \underset{(0{,}10)}{0{,}56e_{t-1}} - \underset{(0{,}12)}{0{,}12e_{t-2}} - \underset{(0{,}10)}{0{,}01e_{t-3}}. \qquad \textbf{(7.32)}$$

Как видно, значимым оказывается только регрессор e_{t-1}, т. е. существенное влияние на результат наблюдения e_t оказывает только одно предыдущее значение e_{t-1}. Положительность оценки соответствующего коэффициента регрессии указывает на положительную корреляцию между ошибками регрессии e_t и e_{t-1}. К такому же выводу приводит и значение статистики Дарбина—Уотсона, полученное в § **7.7**.

В большинстве современных компьютерных пакетов (например, в «*Econometric Views*») применение теста серий осуществляется специальной командой, и нет необходимости оценивать регрессию типа (**7.29**) непосредственно.

***Q*-тест Льюинга—Бокса.** Тест основан на рассмотрении выборочных автокорреляционной $r(\tau)$ и частной автокорреляционной $r_{\text{част}}(\tau)$ функций временного ряда (см. § **6.2**).

Если ряд стационарный, то, как можно доказать, *выборочный частный коэффициент корреляции* $r_{\text{част}}(p)$ *совпадает с оценкой обычного метода наименьших квадратов коэффициента* β_p *в авторегрессионной модели* $AR(p)$:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \ldots + \beta_p y_{t-p} + \varepsilon_t .$$

Это утверждение лежит в основе вычисления значений частной автокорреляционной функции.

Компьютерные регрессионные пакеты предусматривают возможность с помощью специальной команды получать для дан-

ного временного ряда выборочные автокорреляционные функции. Напомним, что график выборочной автокорреляционной функции называется *коррелограммой*. Коррелограмма является быстро убывающей функцией. (Если формально построенная коррелограмма не удовлетворяет этому свойству, это, скорее всего, означает, что ряд на самом деле нестационарный.)

Очевидно, что в случае отсутствия автокорреляции все значения автокорреляционной функции равны нулю. Разумеется, ее в ы б о р о ч н ы е значения $r(\tau)$ окажутся отличными от нуля, но в этом случае отличие не должно быть существенным. На этой идее основан еще один тест, проверяющий гипотезу об отсутствии автокорреляции, — Q-тест Льюинга—Бокса.

Тест Льюинга—Бокса. Статистика Льюинга—Бокса имеет вид:

$$Q_p = n(n+2)\sum_{\tau=1}^{p}\frac{r^2(\tau)}{n-\tau}. \qquad \textbf{(7.33)}$$

Можно доказать, что если верна гипотеза H_0 о равенстве нулю всех коэффициентов корреляции $\rho(e_t e_{t-\tau})$, где $\tau = 1,..., p$, то статистика Q_p имеет распределение χ^2 с p степенями свободы.

▶ **Пример 7.6.** Проверить гипотезу H_0 об отсутствии автокорреляции в модели зависимости курса ценной бумаги A от времени t (§ **7.6**).

Р е ш е н и е. Значение d-статистики Дарбина—Уотсона, примерно равное единице, дает оценку коэффициента корреляции между e_t и e_{t-1}, т. е. $r(1)=0,5$.

Отсюда по формуле (**7.33**)

$$Q_1 = \frac{100\cdot 102\cdot 0,25}{99} = 25,314.$$

Так как фактическое значение статистики больше критического $\chi^2_{0,05;1} = 3,84$, то гипотеза $Q_1 = 0$ отвергается.

Заметим, что гипотеза $Q_1 = 0$ и гипотеза $\rho=0$ о равенстве нулю коэффициента ρ в уравнении (**7.31**) представляют собой по сути одно и то же утверждение об отсутствии авторегрессии первого порядка. Результат тестирования этих гипотез должен совпадать с

выводом, к которому приводит значение статистики Дарбина—Уотсона.

Вычислить «вручную» значение Q_p-статистики при $p>1$, как правило, достаточно трудно (напомним еще раз, что значение статистики Дарбина—Уотсона дает информацию лишь о значении автокорреляционной функции п е р в о г о порядка), однако, в большинстве компьютерных пакетов тест Льюинга—Бокса выполняется специальной командой. В «*Econometric Views*» выдаются значения выборочной автокорреляционной функции $r(\tau)$, частной автокорреляционной функции $r_{\text{част}}(\tau)$, значения Q-статистики Льюинга—Бокса и вероятности $P\ (Q > Q_p)$ для всех порядков τ от 1 до некоторого p. Например, так выглядит соответствующий результат для рассматриваемого примера зависимости курса ценной бумаги A от номера наблюдения (табл. 7.2).

Т а б л и ц а 7.2

τ	*Автокорреляционная функция* $r(\tau)$	*Частная автокорреляционная функция* $r_{\text{част}}(\tau)$	*Статистика Льюинга—Бокса* Q_p	*Вероятность* $P(Q>Q_p)$
1	0,498	0,498	25,314	0,000
2	0,158	−0,120	27,875	0,000
3	0,024	−0,005	27,937	0,000
...	...	...	...	...
35	−0,063	−0,052	58,423	0,008
36	0,059	0,148	58,975	0,009

Как видно, все вероятности, приведенные в столбце 5, ниже уровня значимости, так что гипотезы $Q_p = 0$ об отсутствии автокорреляции отвергаются. ▶

Сделаем одно существенное замечание. Критические значения $\chi^2_{\alpha;p}$ статистики растут с увеличением p. Величины Q_p также растут, но, возможно, медленнее. Таким образом, формальное применение теста Льюинга—Бокса может привести к парадоксальным, на первый взгляд, выводам: например, отвергается гипотеза об отсутствии автокорреляции первого порядка, но не отвергается гипотеза об отсутствии автокорреляции всех порядков до 36-го включительно! На самом деле противоречия здесь

нет, — ведь тот факт, что гипотеза не отвергается, не означает, что она на самом деле верна, — можно лишь утверждать, что если она верна, то наблюдаемый результат возможен с вероятностью, большей, чем уровень значимости. Впрочем, подобного рода ситуации на практике встречаются достаточно редко.

7.9. Устранение автокорреляции. Идентификация временного ряда

Одной из причин автокорреляции ошибок регрессии является наличие «скрытых» регрессоров, влияние которых в результате проявляется через случайный член. Выявление этих «скрытых» регрессоров часто позволяет получить регрессионную модель без автокорреляции.

Наиболее часто «скрытыми» регрессорами оказываются *лаговые* объясняемые переменные (см. § **6.5**). В случае временного ряда вполне естественно предположить, что значения объясняемых переменных зависят не только от включенных уже регрессоров, но и от предыдущих значений объясняемой переменной. Рассмотренные тесты показывают, что это почти всегда имеет место в случае автокорреляции.

Другой механизм образования автокорреляции следующий. Случайные возмущения представляют собой белый шум ξ_t, но на результат наблюдения y_t влияет не только величина ξ_t, но (хотя обычно и в меньшей степени) несколько предыдущих величин $\xi_{t-1},\ldots, \xi_{t-p.}$

Например, рассматривая модель формирования курса ценной бумаги A, мы можем считать, что кроме временной тенденции на курс еще влияет конъюнктура рынка, которую в момент времени t можно считать случайной величиной ξ_t с нулевым средним и некоторой дисперсией. Будем предполагать, что величины ξ_t независимы. Естественно ожидать, что на формирование курса в момент времени t будет оказывать влияние в первую очередь конъюнктура ξ_t и (в меньшей степени) конъюнктуры в дни предыдущих торгов $\xi_{t-\tau}$ и т. д.

Наиболее распространенным приемом устранения автокорреляции во временных рядах является подбор соответствующей модели — *авторегрессионной* $AR(p)$, *скользящей средней* $MA(q)$ или

авторегрессионной модели скользящей средней ARMA(p,q) (см. § **6.5**) для случайных возмущений регрессии.

В число регрессоров в моделях временных рядов могут быть включены и константа, и временной тренд, и какие-либо другие объясняющие переменные. Ошибки регрессии могут коррелировать между собой, однако, мы предполагаем, что остатки регрессии образуют стационарный временной ряд.

Идентификацией *временного ряда* мы будем называть построение для ряда остатков ***адекватной*** *ARMA-модели*, т. е. такой *ARMA*-модели, в *которой остатки представляют собой белый шум, а все регрессоры значимы*. Такое представление, как правило, не единственное, например, один и тот же ряд может быть идентифицирован и с помощью *AR*-модели, и с помощью *MA*-модели. В этом случае выбирается наиболее простая модель.

Как подбирать для данного временного ряда модель *ARMA*? Безусловно, разумен (и часто применяется на практике) *метод элементарного подбора* — пробуются различные модели; при этом начинается поиск с самых простых моделей, которые постепенно усложняются, пока не будет достигнута идентификация.

Полезную информацию можно получить с *помощью выборочных автокорреляционной и частной автокорреляционной функций*. В самом деле, вспомним, что выборочная частная автокорреляционная функция $r_{\text{част}}(p)$ есть оценка параметра β_p в авторегрессионной модели p-го порядка. Отсюда делаем вывод:

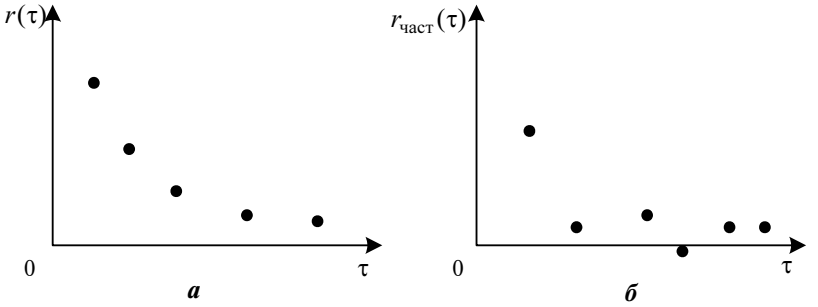
Если все значения выборочной частной автокорреляционной функции порядка выше p незначимо отличаются от нуля, временной ряд следует идентифицировать с помощью модели, порядок авторегрессии которой не выше p.

Теперь рассмотрим модель скользящей средней

$$y_t = \varepsilon_t + \gamma_1\varepsilon_{t-1} + \gamma_2\varepsilon_{t-2} + \ldots + \gamma_q\varepsilon_{t-q}\,.$$

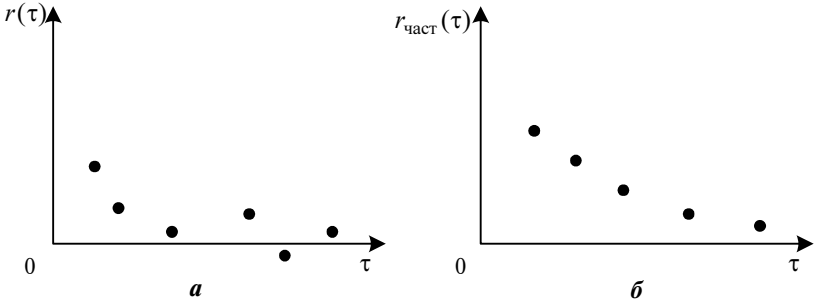
Из того, что величины ε_t при различных t не коррелируют, следует, что величины y_t и $y_{t-\tau}$ могут коррелировать только при $\tau<q$. Таким образом, если все значения выборочной автокорреляционной функции порядка выше q незначимо отличаются от нуля, временной ряд следует идентифицировать с помощью модели скользящей средней, порядок которой не выше q.

Типичные примеры коррелограмм временных рядов, идентифицируемых с помощью модели *ARMA*, изображены на рис. 7.7 (*AR*(1)) и 7.8 (*MA*(1)).



а *б*

Рис. 7.7



а *б*

Рис. 7.8

Вернемся к примеру формирования курса ценной бумаги *A*. Приведенные в § **7.8** значения автокорреляционной и частной автокорреляционной функций указывают на то, что модель *ARMA* остатков регрессии имеет порядок заведомо не выше второго.

Оценим ряд остатков с помощью *AR*(1)-модели. Получим следующее уравнение:

$$\hat{e}_t = \underset{(0,08)}{0,5} e_{t-1};\ d = 1,86.$$

Полученное значение *d*-статистики Дарбина—Уотсона достаточно близко к двум, гипотеза об отсутствии автокорреляции принимается. Тест Льюинга—Бокса также подтверждает гипотезу о равенстве нулю автокорреляционной функции всех порядков. Таким образом, достигнута идентификация ряда остатков с помощью модели *AR*(1).

Проведем теперь оценку модели *MA*(1). Уравнение получается следующим:

$$\hat{e}_t = \xi_t - 0{,}515\xi_{t-1};\ d{=}1{,}87.$$

И в этом случае как значение статистики Дарбина—Уотсона, так и тест Льюинга—Бокса подтверждают гипотезу об отсутствии автокорреляции, т. е. ряд остатков исходной модели может быть идентифицирован и с помощью модели *MA*(1). Причем сравнение полученных результатов показывает, что качество идентификации практически одинаково (возможно, модель *AR* (1) несколько предпочтительнее в силу того, что значение функции правдоподобия чуть выше: $\ln L{=}-504$ для модели *AR* и $\ln L{=}-508$ для модели *MA*. Впрочем, различие это несущественно).

Практически полное совпадение результатов идентификации может иметь и экономическое обоснование. Так, например, естественно считать, что конъюнктура рынка в момент предыдущих торгов — это в основном курс ценной бумаги *A* на этих торгах.

Если удастся построить *ARMA*-модель для ряда остатков, то можно получить эффективные оценки параметра β, а также несмещенные и состоятельные оценки дисперсий β с помощью обобщенного метода наименьших квадратов. Мы рассмотрим эту процедуру на простейшей (и в то же время наиболее часто встречающейся) авторегрессионной модели первого порядка.

7.10. Авторегрессионная модель первого порядка

Вновь возвратимся к регрессионной модели (**7.25**)

$$Y = X\beta + \varepsilon \qquad \textbf{(7.34)}$$

или

$$y_t = \beta_0 + \sum_{j=1}^{p} \beta_j x_{tj} + \varepsilon_t,\ t = 1,\ \ldots,\ n, \qquad \textbf{(7.34′)}$$

где y_t, x_{tj} — наблюдения переменных Y, X_j ($j = 1,\ldots, p$) в момент t.

Будем полагать, что случайные возмущения коррелированы и образуют наиболее простой процесс — *авторегрессионный процесс первого порядка*, т. е.

$$\varepsilon_t = \rho\varepsilon_{t-1} + \nu_t, \qquad \textbf{(7.35)}$$

где ν_t ($t = 1,..., n$) представляет *белый шум*, т. е. последовательность независимых нормально распределенных случайных величин с нулевой средней и дисперсией σ_0^2; ρ — параметр, называемый *коэффициентом авторегрессии*.

Учитывая, что в силу (**7.35**) ε_t и ν_t независимы, дисперсия возмущений

$$D(\varepsilon_t) = \rho^2 D(\varepsilon_{t-1}) + D(\nu_t). \qquad \textbf{(7.36)}$$

Так как процесс ε_t — стационарный, то $D(\varepsilon_t) = D(\varepsilon_{t-1}) = \sigma^2$. Таким образом, полагая $D(\nu_t) = \sigma_0^2$, имеем

$$\sigma^2 = \rho^2\sigma^2 + \sigma_0^2,$$

или

$$\sigma^2 = \frac{\sigma_0^2}{1-\rho^2}. \qquad \textbf{(7.37)}$$

Формально, чтобы использовать равенство (**7.35**) для значения $t = 1$, нам надо ввести случайную величину ε_0. Легко убедиться в том, что для сохранения равенства (**7.36**) случайную величину ε_0 надо определить как нормально распределенную с нулевой средней и дисперсией $D(\varepsilon_0) = \dfrac{\sigma_0^2}{1-\rho^2}$.

Из равенства (**7.37**), учитывая, что дисперсия — величина положительная, следует, что $|\rho| < 1$. *При* $|\rho| > 1$ *ряд оказывается нестационарным*.

Найдем автокорреляционную функцию процесса ε_t. Умножая (**7.35**) на ε_{t-1} и вновь учитывая независимость ε_{t-1} и ν_t, найдем

$$M(\varepsilon_t\varepsilon_{t-1}) = \mathrm{Cov}(\varepsilon_t, \varepsilon_{t-1}) = \rho D(\varepsilon_{t-1}) = \rho\sigma^2,$$

откуда коэффициент корреляции

$$\rho(\varepsilon_t, \varepsilon_{t-1}) = \frac{\mathrm{Cov}(\varepsilon_t, \varepsilon_{t-1})}{\sqrt{D(\varepsilon_t)}\sqrt{D(\varepsilon_{t-1})}} = \frac{\rho\sigma^2}{\sigma\cdot\sigma} = \rho,$$

т. е. *коэффициент авторегрессии* ρ *представляет собой коэффициент корреляции между соседними возмущениями* ε_t и ε_{t-1}, *или коэффициент автокорреляции* $\rho(1)$.

Аналогично получим

$$\mathrm{Cov}(\varepsilon_t, \varepsilon_{t-\tau}) = \rho^m \sigma^2 \qquad \textbf{(7.38)}$$

и

$$\rho(\varepsilon_t, \varepsilon_{t-\tau}) = \rho^m.$$

Учитывая (**7.37**), (**7.38**), ковариационную матрицу вектора возмущений ε для модели с автокорреляционными остатками можно представить в виде:

$$\sum_{\varepsilon} = \Omega = \frac{\sigma_0^2}{1-\rho^2} \begin{pmatrix} 1 & \rho & \ldots & \rho^{n-1} \\ \rho & 1 & \ldots & \rho^{n-2} \\ \ldots & \ldots & \ldots & \ldots \\ \rho^{n-1} & \rho^{n-2} & \ldots & 1 \end{pmatrix}.$$

Для получения наиболее эффективных оценок параметра β в такой модели, если **параметр ρ известен**, можно применить *обобщенный* метод наименьших квадратов.

Исключая $\varepsilon_t (t = 1,..., n)$ из равенств (**7.34**), (**7.35**), получим

$$y_t - \rho y_{t-1} = \beta_0 (1-\rho) + \sum_{j=1}^{p} \beta_j (x_{tj} - \rho x_{t-1,j}) + \nu_t, \ t = 1,..., n. \qquad \textbf{(7.39)}$$

Полученная модель (**7.39**) является *классической*, так как теперь случайные возмущения ν_t ($t = 1,..., n$) независимы и имеют постоянную дисперсию σ_0^2.

Равенство (**7.39**) имеет смысл только при $t \geq 2$, так как при $t = 1$ не определены значения лаговых переменных. Можно показать, что параметры модели сохранятся, если при $t = 1$ умножить обе части уравнения (**7.34′**) на $\sqrt{1-\rho^2}$:

$$\sqrt{1-\rho^2}\, y_1 = \sqrt{1-\rho^2}\, \beta_0 + \sqrt{1-\rho^2} \sum_{j=1}^{p} \beta_j x_{1j} + \sqrt{1-\rho^2}\, \varepsilon_1. \qquad \textbf{(7.40)}$$

Преобразование (**7.40**) называется *поправкой Прайса—Уинстона* для первого наблюдения. При большом количестве наблюдений поправка Прайса—Уинстона практически не изменяет результат, поэтому ее часто не учитывают, оставляя значение первого наблюдения неопределенным.

Таким образом, при *известном значении* ρ *автокорреляция легко устраняется*. На практике, однако, **значение ρ не** бывает **известно**, поэтому в равенстве (**7.39**) присутствует не точное значение ρ, а наблюдаемое значение его оценки $\hat{\rho}$.

Наиболее простой способ оценить ρ — применить обычный метод наименьших квадратов к регрессионному уравнению (**7.35**).

В качестве примера вновь рассмотрим модель формирования курса ценной бумаги *A*. В предыдущем параграфе была получена оценка $\hat{\rho} = 0{,}5$.

Перейдем от наблюдений y_t и t ($t = 1,..., n$)[1] к наблюдениям

$$w_t = y_t - \hat{\rho} y_{t-1} = y_t - 0{,}5 y_{t-1}, \qquad \textbf{(7.41)}$$

$$z_t = t_t - \hat{\rho} t_{t-1} = t_t - 0{,}5 y_{t-1}.$$

Применяя обычный метод наименьших квадратов, получим уравнение

$$\hat{w}_t = 245{,}32 + 0{,}7 z_t.$$

При этом значение статистики Дарбина—Уотсона d=1,864 достаточно близко к двум, так что можно считать, что автокорреляция устранена и получено оценочное значение параметра β: $\hat{\beta} = 0{,}7$.

Регрессия оказывается значимой несмотря на невысокое качество подгонки (низкое значение R^2=0,05). Заметим, что полученная оценка R^2 значительно отличается от оценки до устранения автокорреляции (см. (**7.29**)).

Двухшаговая процедура Дарбина. Как правило, более точную оценку параметра β дает двухшаговая процедура Дарбина, которая заключается в следующем. Исключая ε_t из уравнений (**7.34**)—(**7.35**), запишем регрессионную модель в виде

$$y_t = \beta_0 (1 - \rho) + \beta x_t + \rho y_{t-1} - \beta \rho x_{t-1} + \nu_t, \quad t = 1,..., n. \qquad \textbf{(7.42)}$$

Применим к уравнению (**7.42**) обычный метод наименьших квадратов, включая ρ в число оцениваемых параметров. Получим оценки r и θ величин ρ и $-\beta\rho$. Тогда оценкой Дарбина является величина

$$\hat{\beta} = -\frac{\theta}{r}. \qquad \textbf{(7.43)}$$

Например, в модели формирования курса ценной бумаги *A*, применяя двухшаговую процедуру Дарбина, получаем значение оценки параметра β, равное $\hat{\beta} = 0{,}698$ и близкое к полученному ранее $\hat{\beta} = 0{,}7$.

---

[1] Переменная время t здесь выступает в роли регрессора.

В большинстве компьютерных пакетов реализованы также *итеративные* процедуры, позволяющие оценивать значение параметра модели (**7.34**) при условии, что остатки модели образуют стационарный временной ряд, моделируемый как авторегрессионный процесс первого порядка, т. е. автокорреляция имеет характер (**7.35**).

Опишем две наиболее часто используемые процедуры.

Процедура Кохрейна—Оркатта. Указанная процедура заключается в том, что, получив методом наименьших квадратов оценочное значение $\hat{\rho}$ параметра ρ, от наблюдений y_t и t переходят к наблюдениям w_t, z_t по формулам (**7.41**) и, получив оценку параметра $\hat{\beta}_1$, образуют новый вектор остатков

$$e_1 = W - Z\hat{\beta}_1. \qquad \textbf{(7.44)}$$

Далее применяют метод наименьших квадратов к регрессионному уравнению

$$\left(\hat{e}_1\right)_t = \rho\left(e_1\right)_{t-1} + \left(\nu_1\right)_t, \qquad \textbf{(7.45)}$$

получают новую оценку $\hat{\rho}_1$ параметра ρ и от наблюдений w_t и z_t переходят к новым $(w_1)_t$ и $(z_1)_t$:

$$\left(w_1\right)_t = w_1 - \hat{\rho}_1 w_{t-1}, \ \ \left(z_1\right)_t = z_1 - \hat{\rho}_1 z_{t-1}.$$

Новую оценку $\hat{\beta}_2$ получают, применяя метод наименьших квадратов к модели (**7.45**), и процедура повторяется вновь.

7.11. Доступный (обобщенный) метод наименьших квадратов

В § **7.1** рассматривалась обобщенная модель множественной регрессии

$$Y = X\beta + \varepsilon. \qquad \textbf{(7.46)}$$

В случае, когда ковариационная матрица $\sum_{\varepsilon} = \Omega = \sigma^2\Omega_0$ известна, как показано в § **7.2**, наилучшей линейной несмещенной оценкой вектора β является оценка обобщенного метода наимень-

ших квадратов (совпадающая с оценкой максимального правдоподобия)

$$b^* = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y\,. \qquad \textbf{(7.47)}$$

Эта оценка получается из условия минимизации обобщенного критерия (**7.14**):

$$S=(Y-Xb)'\Omega^{-1}(Y-Xb).$$

При рассмотрении нормальной обобщенной модели регрессии, т. е. в случае нормального закона распределения возмущений, т. е. $\varepsilon \sim N_n(\mathbf{0};\sigma^2\Omega_0)$, оценка b^* также распределена нормально:

$$b^* \sim N(\beta;\sigma^2(X'\Omega_0^{-1}X)^{-1}).$$

На практике матрица возмущений Ω почти никогда неизвестна, и, как уже было отмечено в § **7.2**, оценить ее $n(n+1)/2$ параметров по n наблюдениям не представляется возможным.

Предположим, что задана структура матрицы Ω (и соответственно Ω_0), т. е. форма ее функциональной зависимости от относительно небольшого числа параметров θ_1, θ_2, ..., θ_m, т. е. матрица $\sum_\varepsilon = \sigma^2\Omega_0(\theta_1,\theta_2...,\theta_m)$. Например, в модели с автокоррелированными остатками (структура матрицы $\sum_\varepsilon$ определяется двумя параметрами σ^2 и $\theta_1=\rho$, так что матрица Ω_0 имеет вид (см. § **7.10**):

$$\Omega_0 = \begin{pmatrix} 1 & \rho & ... & \rho^{n-1} \\ \rho & 1 & ... & \rho^{n-2} \\ ... & ... & ... & ... \\ \rho^{n-1} & \rho^{n-2} & ... & 1 \end{pmatrix},$$

где ρ — неизвестный параметр, который необходимо оценить.

Идея оценки матрицы $\sum_\varepsilon = \sigma^2\Omega_0$ состоит в следующем. Вначале по исходным наблюдениям находят *состоятельные* оценки параметров $\theta=(\theta_1, \theta_2,..., \theta_n)'$. Затем получают оценку параметра σ^2. В соответствии с (**4.21**) такая оценка для классической модели находится делением минимальной остаточной суммы квадратов $\sum_{i=1}^{n} e_i^2 = e'e$ на число степеней свободы $(n - p - 1)$. Для обобщенной модели соответствующая оценка

$$s_*^2 = \frac{e'\Omega_0^{-1}e}{n-p-1} = \frac{\left(Y - Xb^*\right)' \Omega_0^{-1}\left(Y - Xb^*\right)}{n-p-1}. \qquad \textbf{(7.48)}$$

Зная s_*^2, вычисляем матрицу $\hat{\sum}_\varepsilon = s_*^2\hat{\Omega}_0$. Доказано, что при некоторых достаточно общих условиях использование полученных оценок в основных формулах обобщенного метода наименьших квадратов вместо неизвестных истинных значений σ^2 и $\sum_\varepsilon = \sigma^2\Omega_0$ даст также состоятельные оценки параметра β и ковариационной матрицы $\sum_\beta$. Описанный метод оценивания называется *доступным* (или *практически реализуемым*) *обобщенным методом наименьших квадратов*.

Таким образом, оценкой доступного обобщенного метода наименьших квадратов вектора β есть

$$\hat{b}^* = (X'\hat{\Omega}_0^{-1}X)^{-1}X'\hat{\Omega}_0^{-1}Y. \qquad \textbf{(7.49)}$$

Применяя метод максимального правдоподобия (см. § **2.7**, **3.4**) для оценки нормальной обобщенной линейной модели регрессии, можно показать, что оценки максимального правдоподобия $\hat{\beta}, \hat{\theta}, \hat{\sigma}^2$ находятся из системы уравнений правдоподобия:

$$\begin{cases} \hat{\beta} = (X'\hat{\Omega}_0^{-1}X)^{-1}X'\Omega_0^{-1}Y, & \textbf{(7.50)} \\ \hat{\sigma}^2 = \dfrac{1}{n}e'\Omega_0 e, & \textbf{(7.51)} \\ \dfrac{e'c_j e}{e'\Omega_0^{-1}e} = \dfrac{1}{n}\operatorname{tr}\left(c_j\hat{\Omega}_0\right), \quad j = 1,\ldots,m, & \textbf{(7.52)} \end{cases}$$

где $e = Y - X\hat{\beta}$, $c_j = \dfrac{\partial\hat{\Omega}_0^{-1}(\hat{\theta})}{\partial\theta_j}$.

Анализируя систему (**7.50**)—(**7.52**), видим, что *оценки* $\hat{\beta}$ и σ^2 *метода максимального правдоподобия совпадают с оценками* b^* и s_*^2 *обобщенного метода наименьших квадратов* (правда, если $\hat{\beta} = b^*$, то $\hat{\sigma}^2 \approx s_*^2$ с точностью до асимптотически устранимого смещения).

Для решения системы (**7.50**) — (**7.52**) иногда удается найти *точное* решение, однако чаще приходится прибегать к *итерационной процедуре*, например, двухшаговой.

1-й ш а г. Вначале находят оценку метода максимального правдоподобия $\hat{\beta}_0 = (X'X)^{-1} X'Y$, вычисляют остатки $e_1 = Y - X\hat{\beta}_1$ и решают полученную систему (**7.50**) — (**7.52**) при заданных остатках.

Находят $m \times 1$ вектор $\hat{\theta}_1$ и матрицу $\hat{\Omega}_{0(1)} = \Omega_0\left(\hat{\theta}_{(1)}\right)$.

2-й ш а г. Находят оценку вектора β по формуле (**7.49**):

$$\hat{\beta}_2 = \left(X'\hat{\Omega}_0^{-1}X\right)^{-1} X'\hat{\Omega}_0^{-1}Y .$$

Полученные таким образом оценки $\beta_{(2)}$ при некоторых, весьма слабых предположениях (как, например, состоятельность оценок $\hat{\theta}$) будут асимптотически (при большом n) *эквивалентны оценкам метода максимального правдоподобия*, а значит, будут *асимптотически эффективны* в широком классе случаев. Описанная двухшаговая итерационная процедура была фактически реализована в процедуре *Кохрейна—Оркатта* (§ **7.10**).

Упражнения

7.7. В таблице приведены данные по 18 наблюдениям модели пространственной выборки:

i	x_i	e_i^2	i	x_i	e_i^2
1	21,3	2,3	10	71,5	23,8
2	22,6	5,6	11	75,7	45,7
3	32,7	12,8	12	76,0	34,7
4	41,9	10,1	13	78,9	56,9
5	43,8	14,6	14	79,8	56,8
6	49,7	13,9	15	80,7	49,8
7	56,9	24,0	16	80,8	58,9
8	59,7	21,9	17	96,9	87,8
9	67,8	19,7	18	97,0	87,5

Предполагая, что ошибки регрессии представляют собой нормально распределенные случайные величины, проверить гипотезу о гомоскедастичности, используя тест Голдфелда—Квандта.

7.8. При оценивании модели пространственной выборки обычным методом наименьших квадратов получено уравнение:

$$\hat{y} = 3 + 0,6x_1 - 1,2x_2.$$

Уравнение регрессии квадратов остатков на квадраты регрессоров имеет вид:

$$\hat{e}^2 = 2 + 0,3x_1^2 + 0,1x_2^2;\ R^2=0,2.$$

Зная, что объем пространственной выборки n=200, проверить гипотезу Уайта о гомоскедастичности модели.

7.9. При оценивании модели пространственной выборки с помощью обычного метода наименьших квадратов получено следующее уравнение:

$$\hat{y} = 12 + 3,43x_1 - 0,45x_2;\ \ d\ =\ 1,2.$$
$$(0,1)\ \ (0,4)\quad (0,1)$$

При осуществлении регрессии квадратов остатков на квадраты регрессоров получено уравнение вида:

$$\hat{e}^2 = 4 + 2x_1^2 + 0,4x_2^2.$$
$$(0,7)\ \ (0,3)$$

Объем выборки n=3000. С какими из перечисленных ниже выводов следует согласиться:

а) полученные значения коэффициентов модели с большой вероятностью близки к истинным;

б) регрессор X_2 может быть незначимым;

в) так как значение статистики Дарбина—Уотсона d далеко от двух, следует устранить автокорреляцию остатков?

7.10. При оценивании модели временного ряда получены следующие результаты.

Уравнение модели имеет вид:

$$\hat{y}_t = 2 - 1,2t;\ \ d = 1,9.$$
$$(0,7)$$

С какими из перечисленных ниже выводов следует согласиться:

а) так как значение статистики Дарбина—Уотсона d близко к двум, автокорреляция остатков отсутствует;

б) коэффициент модели при t значим;

в) если объем выборки достаточно велик, значение коэффициента при t в любом случае с большой вероятностью близко к истинному;

г) применение теста Бреуша—Годфри может выявить автокорреляцию остатков между отдаленными наблюдениями?

7.11. При оценивании модели временного ряда методом наименьших квадратов получены следующие результаты:

$$\hat{y}_t = 12 - 0,2t; \;\; d = 1,8.$$
$$(0,01)$$

Известно, что количество наблюдений n=150. С помощью теста Льюинга—Бокса проверить гипотезу об отсутствии автокорреляции первого порядка.

7.12. Применение теста Льюинга—Бокса дает следующие результаты:

r (τ)	$r_{\text{част}}$ (τ)	Q_p	P ($Q > Q_p$)
0,34	0,33	21,45	0,00
0,12	0,11	24,12	0,00
0,01	0,00	24,13	0,00
0,00	0,00	24,13	0,00

Ряд остатков идентифицируется с помощью модели *ARMA* (p, q). Какие значения p и q целесообразно выбрать на основе полученных результатов тестирования?

Глава 8

Регрессионные динамические модели

8.1. Стохастические регрессоры

До сих пор мы предполагали, что в регрессионной модели (**4.2**)

$$Y = X\beta + \varepsilon \qquad \textbf{(8.1)}$$

объясняющие переменные X_j ($j = 1,..., p$), образующие матрицу X, не являются случайными. Напомним, что, по сути, это означает, что если бы мы повторили серию выборочных наблюдений, значения переменных X_j были бы теми же самыми (между тем, как значения Y изменились бы за счет случайного члена ε).

Подобное предположение, приводящее к значительным техническим упрощениям, может быть оправдано в том случае, когда экспериментальные данные представляют собой пространственную выборку. В самом деле, мы можем считать, что значения переменных X_j мы выбираем заранее, а затем наблюдаем получающиеся при этом значения Y (здесь имеется некоторая аналогия с заданием функции «по точкам» — значения независимой переменной выбираются произвольно, а значения зависимой вычисляются). В случае временного ряда, регрессоры которого представляют собой временной тренд, циклическую и сезонную компоненты, объясняющие переменные также, очевидно, не случайны.

Однако в тех случаях, когда среди регрессоров временного ряда присутствуют переменные, значения которых сами образуют временной ряд, предположение об их детерминированности неправомерно. Так что *в моделях временных рядов мы, как правило, должны считать наблюдения* x_{tj} ($t = 1,..., n$; $j = 1,..., p$) *случайными величинами.*

В этом случае естественно возникает вопрос о коррелированности между регрессорами и ошибками регрессии ε. Покажем, что от этого существенно зависят результаты оценивания — причем не только количественно, но и качественно.

Мы будем предполагать, что x_{tj} и y_t — с т а ц и о н а р н ы е временные ряды, т. е. все случайные величины x_t имеют одно и то же распределение (и аналогично y_t).

Предположим пока для простоты, что имеется всего один регрессор X, т. е. модель имеет вид:

$$y_t = \alpha + \beta x_t + \varepsilon_t . \qquad \textbf{(8.2)}$$

Рассмотрим ковариации обеих частей равенства (**8.2**) со случайной величиной x_t:

$$\mathrm{Cov}(x_t, y_t) = \beta \mathrm{Cov}(x_t, x_t) + \mathrm{Cov}(x_t, \varepsilon).$$

Отсюда (учитывая, что $\mathrm{Cov}(x_t, x_t) = D(x_t)$) получим:

$$\beta = \frac{1}{D(x_t)} \left[\mathrm{Cov}(x_t, y_t) - \mathrm{Cov}(x_t, \varepsilon_t) \right] \qquad \textbf{(8.3)}$$

(отметим, что так как x_t, y_t — стационарные ряды, то ковариации, входящие в правую часть равенства (**8.3**), для всех значений t совпадают). Рассмотрим выборочные оценки этих ковариаций:

$$\hat{D}(x_t) = \frac{1}{n} \sum_{i=1}^{n} x_t^2 - \left(\frac{\sum_{i=1}^{n} x_t}{n} \right)^2 ,$$

$$\hat{\mathrm{Cov}}(x_t, y_t) = \frac{1}{n} \sum_{t=1}^{n} x_t y_t - \frac{\sum_{t=1}^{n} x_t}{n} \cdot \frac{\sum_{t=1}^{n} y_t}{n} . \qquad \textbf{(8.4)}$$

Сравнивая (**8.4**) с (**3.13**), получаем следующее выражение для оценки параметра β:

$$b = \frac{\hat{\mathrm{Cov}}(x_t, y_t)}{\hat{D}(x_t)}$$

или, подставляя $y_t = \alpha + \beta x_t + \varepsilon_t$, находим:

$$b = \beta + \frac{\frac{1}{n} \sum_{t=1}^{n} x_t \varepsilon_t}{\hat{D}(x_t)} = \beta + \frac{1}{n} \left(\frac{\sum_{t=1}^{n} x_t}{\hat{D}(x_t)} \right) \varepsilon_t . \qquad \textbf{(8.5)}$$

Рассмотрим отдельно т р и с л у ч а я.

1. *Регрессоры X и ошибки регрессии* ε *не коррелируют,* т. е. г е н е р а л ь н а я ковариация $\mathrm{Cov}(x_s, \varepsilon_t) = 0$ для всех s, $t = 1,..., n$.

В этом случае из (**8.5**) следует:

$$M(b) = \beta + \frac{1}{n} M \left(\frac{\sum_{t=1}^{n} x_t \varepsilon_t}{\hat{D}(x_t)} \right).$$

Отсюда в силу некоррелированности X и ε и условия $M(\varepsilon)=0$, получаем:

$$M(b) = \beta + \sum_{t=1}^{n} M \left(\frac{x_t}{\hat{D}(x_t)} \right) M(\varepsilon_t) = \beta, \qquad \textbf{(8.6)}$$

т. е. оценка b оказывается в этом случае *несмещенной*.

Эта оценка является также *состоятельной*. В самом деле, при неограниченном увеличении объема выборки оценки ковариаций сходятся по вероятности[1] к их генеральным значениям. В то же время предел по вероятности отношения двух случайных величин равен отношению их пределов, т. е.

$$b = \frac{\hat{\mathrm{Cov}}(x_t, y_t)}{\hat{D}(x_t)} \xrightarrow[n\to\infty]{\mathscr{P}} \frac{\mathrm{Cov}(x_t, y_t)}{D(x_t)},$$

но если $\mathrm{Cov}(x_t,\varepsilon_t)=0$, то из (**8.3**) следует, что $\dfrac{\mathrm{Cov}(x_t, y_t)}{D(x_t)} = \beta$.

2. *Значения регрессоров X не коррелированы с ошибками регрессии* ε *в данный момент времени, но коррелируют с ошибками регрессии в более ранние моменты времени* $t - \tau$.

В этом случае оценка (**8.2**) остается состоятельной: доказательство дословно повторяет то, которое приведено в предыдущем пункте. Однако она уже не будет несмещенной. В самом деле, выборочная дисперсия $D(x_t)$ содержит значения во все моменты времени, т. е. $D(x_t)$ коррелирует с ε и, стало быть, равенство (**8.6**), используемое в предыдущем пункте для доказательства несмещенности β, неверно.

3. *Значения регрессоров* X_t *коррелированы с ошибками* ε_t.

В этом случае, очевидно, не проходят ни доказательство несмещенности, ни доказательство состоятельности. Состоятельность, вообще говоря, может сохраниться, если выборочная дисперсия $D(x_t)$ неограниченно возрастает с увеличением объема выборки, однако, в этом случае каждая конкретная задача требует отдельного анализа.

---

[1] См. сноску на с. 41.

Итак, если рассматривается модель

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

со стохастическими регрессорами, то о ц е н к и параметра β, полученные методом наименьших квадратов:

- *несмещенные и состоятельные*, если объясняющие переменные и ошибки регрессии не коррелируют;
- *состоятельные, но смещенные*, если объясняющие переменные коррелируют с ошибками регрессии в более ранние моменты времени, но не коррелируют в один и тот же момент времени;
- *смещенные и несостоятельные*, если объясняющие переменные и ошибки регрессии коррелируют в том числе и в одинаковые моменты времени.

В точности такие же результаты могут быть получены и в случае множественной регрессии, когда β — *векторный* параметр. Мы не будем здесь приводить соответствующие выкладки.

Таким образом, коррелированность регрессоров и ошибок регрессии оказывается значительно более неприятным обстоятельством, чем, например, гетероскедастичность или автокорреляция. Неадекватными оказываются не только результаты тестирования гипотез, но и сами оценочные значения параметров.

Мы рассмотрим д в е наиболее часто встречающиеся п р и ч и н ы коррелированности регрессоров и ошибок регрессии.

1. *На случайный член ε воздействуют те же факторы, что и на формирование значений регрессоров.*

Пусть, например, существует переменная U, такая, что в регрессионных моделях

$$y_t = \alpha + \beta x_t + \gamma u_t + \nu_t,$$
$$x_t = \lambda + \delta u_t + \xi_t$$

коэффициенты γ и δ значимо отличаются от нуля. (Будем здесь считать для простоты, что имеется лишь один регрессор X, рассматриваемый как случайная величина.)

Предположим, однако, что мы не имеем наблюдаемых значений U и рассматриваем модель

$$y_t = \alpha + \beta x_t + \varepsilon_t .$$

Естественно ожидать в этом случае коррелированность регрессоров X и ошибок регрессии ε ($\mathrm{Cov}(X, \varepsilon) = \gamma\delta\mathrm{Cov}(U, U)$).

Рассмотрим следующий модельный пример.

В поселке A производится сырье двух видов I и II. Сырье перевозится в город B, где на заводе производится субпродукт, ко-

торый продается фирме-производителю по цене X. Фирма изготовляет из субпродукта конечный товар, который перевозится в областной центр и реализуется по цене Y. Цены на сырье первого и второго типа меняются и образуют временные ряды Z_1 и Z_2.

Допустим, фирма-производитель хочет построить зависимость цены реализации Y от цены X и рассматривает регрессионную модель

$$y_t = \alpha + \beta x_t + \varepsilon_t .$$

Применение к экспериментальным данным метода наименьших квадратов дает следующий результат:

$$y_t = \underset{(4,82)}{13,64} + \underset{(0,01)}{1,415} x_t, \quad d=2,14, \quad R^2=0,96. \qquad \textbf{(8.7)}$$

Насколько достоверны полученные результаты? Очевидно, и доставка сырья в город B, и доставка конечного продукта в город C связаны с перевозками, а значит, такие факторы, как затраты на топливо, зарплата водителей, состояние дорог и т. д. будут влиять и на формирование цены X, и на конечную цену Y при з а д а н н о м X, т. е. на величину ошибок регрессии модели. Таким образом, регрессоры и ошибки регрессии оказываются коррелированными, и оценки, полученные методом наименьших квадратов, несостоятельны.

В следующем параграфе мы вернемся к этому модельному примеру, а пока отметим вторую типичную причину коррелированности объясняющих переменных и случайного члена.

2. *Ошибки при измерении регрессоров.* Пусть при измерении регрессора X_j допускается случайная ошибка U_j, удовлетворяющая условию $M(u_{ij})=0$, т. е. в обработку поступает не истинное наблюдаемое значение x_{tj}, а искаженное

$$\left(x_{tj}\right)^* = x_{tj} + u_{tj}.$$

При оценивании модели фактически рассматривается регрессия

$$Y = X\beta + \nu.$$

Именно из этого уравнения мы и получим оценку β .

Между тем, в действительности, имеем:

$$Y = X\beta + \varepsilon = (X_* - U)\beta + \varepsilon = X_*\beta + (\varepsilon - U\beta),$$

т. е. $\nu = \varepsilon - U\beta$, и, следовательно, в фактически рассматриваемой модели имеется корреляция между фиксируемыми значениями регрессоров X_* и случайным членом ν: $\mathrm{Cov}(X_*, \nu) = -\mathrm{Cov}(U, U)\beta$.

Отметим, что обе указанные причины коррелированности регрессоров и ошибок регрессии имеют один и тот же математический смысл; значения объясняемых переменных формируются не присутствующим в модели регрессором, а каким-то другим, и, стало быть, оценивается «не тот» параметр.

При рассмотрении конкретных регрессионных моделей временных рядов с коррелированностью регрессоров и ошибок приходится сталкиваться довольно часто. Мы рассмотрим примеры таких моделей в настоящей главе, а пока приведем наиболее часто используемый прием, применяемый в подобных случаях, — метод инструментальных переменных.

8.2. Метод инструментальных переменных

Идея метода заключается в том, чтобы подобрать новые переменные Z_j ($j = 1,..., l$), которые бы тесно коррелировали с X_j и не коррелировали с ε в уравнении (**8.1**). Набор переменных $\{Z_j\}$ может включать те регрессоры, которые не коррелируют с ε, а также другие величины. При этом, вообще говоря, количество переменных $\{Z_j\}$ может отличаться от исходного количества регрессоров (обычно набор $\{Z_j\}$ содержит большее число переменных).

Такие переменные $Z_1,..., Z_e$ называются *инструментальными*.

Они позволяют построить состоятельную оценку параметра β модели (**8.1**). Такая оценка имеет вид:

$$\widetilde{\beta}_{iv} = \left(Z'X\right)^{-1} Z'Y = \left(\frac{1}{n} Z'X\right)^{-1} \left(\frac{1}{n} Z'Y\right). \qquad \textbf{(8.8)}$$

Здесь Z, X, Y — матрицы наблюдаемых значений переменных.

Если рассматривается модель парной регрессии, и единственная переменная X заменяется на единственную инструментальную переменную Z, формула (**8.8**) примет вид:

$$\widetilde{\beta}_{iv} = \frac{\dfrac{1}{n}\sum_{t=1}^{n} z_t y_t - \dfrac{\sum_{t=1}^{n} z_t}{n} \dfrac{\sum_{t=1}^{n} y_t}{n}}{\dfrac{1}{n}\sum_{t=1}^{n} z_t x_t - \dfrac{\sum_{t=1}^{n} z_t}{n} \dfrac{\sum_{t=1}^{n} x_t}{n}}. \qquad \textbf{(8.9)}$$

Если ряд z_t — также стационарный и, следовательно, случайные величины z_t одинаково распределены, то формула (**8.9**) может быть записана в виде:

$$\tilde{\beta}_{iv} = \frac{\hat{\mathrm{Cov}}(z_t, y_t)}{\hat{\mathrm{Cov}}(z_t, x_t)} = \beta + \frac{\frac{1}{n}\sum_{t=1}^{n} z_t \varepsilon_t}{\hat{\mathrm{Cov}}(z_t, x_t)}. \qquad \textbf{(8.10)}$$

Так как Z и ε не коррелируют, равенство (**8.10**) означает, что оценка $\widetilde{\beta}_{iv}$ является состоятельной (очевидно, что коррелированность X и Z означает, что Cov (z_t,x_t) не стремится к нулю с увеличением объема выборки). В то же время из равенства (**8.10**) не следует несмещенность оценки $\widetilde{\beta}_{iv}$. Также эта оценка, вообще говоря, не обладает минимальной ковариацией. В самом деле $\widetilde{\beta}_{iv}$ явно зависит от Z (то они разные при разных наборах инструментальных переменных). Между тем оценка, обладающая минимальной ковариацией, очевидно, единственная.

Естественно, возникает вопрос: как выбрать «наилучшие» инструментальные переменные, т. е. такие переменные, при которых оценка $\widetilde{\beta}_{iv}$ имела бы наименьшую возможную ковариацию. Основой для решения этой задачи служит следующая теорема, которую мы примем без доказательства.

Теорема. *Распределение p-мерной случайной величины* $\widetilde{\beta}_{iv}$ *при неограниченном увеличении объема выборки стремится к нормальному с математическим ожиданием* β *и ковариационной матрицей, пропорциональной матрице* $(RR)^{-1}$*, где R — корреляционная матрица между инструментальными переменными Z и исходными переменными X.*

Здесь p — число исходных регрессоров, l — число инструментальных переменных ($l \geq p$); матрица R имеет размерность $l \times p$.

Таким образом, чем теснее коррелируют исходные переменные X и инструментальные Z, тем более эффективной будет оценка $\widetilde{\beta}_{iv}$. При этом, однако, должно выполняться условие $\mathrm{Cov}(Z,\varepsilon)=0$. Отсюда следует, что оптимальными инструментальными переменными являются переменные вида

$$Z_j^* = X_j - M_\varepsilon(X_j),\ j = 1,\ldots, p.$$

Здесь X, Z — p-мерные случайные величины.

Однако случайные величины Z^* н е н а б л ю д а е м ы. Таким образом, реально наилучшего набора инструментальных переменных не существует.

Пусть существует произвольный набор инструментальных переменных $\{Z_j\}$, имеющих реальный экономический смысл, причем их число, вообще говоря, может превосходить p — число исходных регрессоров X_j. Рассмотрим проекции $\widetilde{X}_j$ регрессоров X_j на пространство $\{Z\}$[1]. Для этого надо осуществить регрессию вида

$$X_j = Z\gamma + \nu$$

для всех регрессоров X_j и взять их объясненные (прогнозные) значения $\widetilde{X}_j$

$$\tilde{X}_j = Z\hat{\gamma} = Z(Z'Z)^{-1}Z'X_j \ . \qquad \textbf{(8.11)}$$

Переменные $\widetilde{X}_j$ не коррелируют с ошибками регрессии, так как линейно выражаются через инструментальные переменные $Z_1,\ldots, Z_e$. Рассмотрим $\widetilde{X}_j$ как новые инструментальные переменные.

По свойству проекции имеем $\widetilde{X}'X = \widetilde{X}'\widetilde{X}$, так что оценка $\widetilde{\beta}_{iv}$ (см. (**8.8**)) принимает вид:

$$\widetilde{\beta}_{iv} = (\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'Y \ , \qquad \textbf{(8.12)}$$

т. е. совпадает с оценкой, полученной обычным методом наименьших квадратов, модели $Y = \widetilde{X}\beta + \varepsilon$.

Описанная процедура называется *двухшаговым методом наименьших квадратов*. По сути метод наименьших квадратов применяется здесь дважды: сначала для получения набора регрессоров $\boldsymbol{X}$, затем для получения оценок параметра β.

Подставляя (**8.11**) в (**8.12**), получаем выражение оценки двухшагового метода наименьших квадратов через исходные инструментальные переменные Z:

$$\widetilde{\beta}_{iv} = \left(X'Z\left(Z'Z\right)'Z'X \right)^{-1} X'Z\left(Z'Z\right)^{-1}Z'Y \ . \qquad \textbf{(8.13)}$$

Процедура двухшагового метода наименьших квадратов реализована в большинстве компьютерных регрессионных пакетов.

---

[1] Напомним, что в § **3.7** мы рассматривали геометрическую интерпретацию регрессии и, в частности, проекцию вектора Y на пространство регрессоров.

Рассмотрим пример из § **8.1**. Строя модель зависимости цены конечного товара от стоимости субпродукта, в качестве инструментальных переменных можно выбрать цены на сырье I и II.

Применяя двухшаговый метод наименьших квадратов, получаем уравнение:

$$\hat{y} = 16,72 + 1,408x, \ d = 2,16, \ R^2 = 0,997.$$
$$(4,86) \quad (0,01)$$

Сравнивая полученные результаты с полученными обычным методом наименьших квадратов (см. (**8.7**)), можно заметить, что модельные уравнения различаются, хотя и оказываются довольно близкими друг к другу.

На практике, как правило, возможности выбора инструментальных переменных не столь широки. Так, в рассматриваемом примере мы имели, по сути, единственно возможный набор Z_1, Z_2. Нередко бывает, что нет и вовсе никаких наблюдаемых инструментальных переменных.

В следующем параграфе мы опишем еще некоторые возможные способы оценивания моделей, в которых регрессоры коррелируют с ошибками регрессии.

8.3. Оценивание моделей с распределенными лагами. Обычный метод наименьших квадратов

Как мы уже отмечали, в моделях временных рядов часто значения объясняемых переменных зависят от их значений в предыдущие моменты времени.

Авторегрессионной моделью с распределенными лагами порядков p и q (или моделью ADL (p, q) называется модель[1] вида

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + ... + \beta_p x_{t-p} + \gamma_1 y_{t-1} + ... + \gamma_q y_{t-q} + \varepsilon_t . \quad \textbf{(8.14)}$$

При этом, вообще говоря, имеет место корреляция между $y_{t-\tau}$ и ε_t.

Именно такого вида модели имеют наибольшее практическое значение, и именно такого вида механизм возникновения корреляции между регрессорами и ошибками регрессии наиболее часто встречается в экономических приложениях. На прак-

---

[1] «*ADL*» — от английских слов «*autogressive distributed lags*». Следует отметить, что наряду с авторегрессионной моделью $ADL(p,q)$ в эконометрике используется и «обычная» регрессионная *модель с распределенными лагами* p-го порядка (или модель $DL(p)$):

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + ... + \beta_p x_{t-p} + \varepsilon_t .$$

тике чаще всего возникают модели *ADL* порядка (0,1), т. е. модели вида[1]

$$y_t = \alpha + \beta x_t + \gamma y_{t-1} + \varepsilon_t \qquad \textbf{(8.15)}$$

с ошибками регрессии ε_t, подчиняющимися авторегрессионному процессу первого порядка *AR*(1) или закону скользящей средней первого порядка *MA*(1):

$$\varepsilon_t = \rho\varepsilon_{t-1} + \xi_t; \qquad \textbf{(8.16)}$$

$$\varepsilon_t = \xi_t + \rho\xi_{t-1}. \qquad \textbf{(8.17)}$$

В обоих случаях ξ_t — независимые, одинаково распределенные случайные величины с нулевыми математическими ожиданиями и дисперсиями σ^2 (белый шум).

В дальнейшем мы будем иметь дело почти исключительно с моделями вида (**8.15**).

Рассмотрим сначала результат применения к модели (**8.14**) обычного метода наименьших квадратов. Начнем с модели (**8.15**). Предположим сначала для простоты, что регрессор X в уравнении (**8.15**) отсутствует, т. е. лаговая переменная y_{t-1} является единственной объясняющей переменной, и модель имеет вид:

$$y_t = \gamma y_{t-1} + \varepsilon_t; \qquad \textbf{(8.18)}$$

$$\varepsilon_t = \rho\varepsilon_{t-1} + \xi_t. \qquad \textbf{(8.19)}$$

Тогда оценка γ, полученная обычным методом наименьших квадратов, имеет вид:

$$\hat{\gamma} = \frac{\sum_{t=1}^{n} y_t y_{t-1}}{\sum_{t=1}^{n} y_{t-1}^2} = \beta + \frac{\sum_{t=1}^{n} y_{t-1}\varepsilon_t}{\sum_{t=1}^{n} y_{t-1}^2}. \qquad \textbf{(8.20)}$$

Напомним, что все ряды предполагаются стационарными, т. е. имеют место равенства:

$$D(y_{t-1}) = D(y_t),$$

$$D(\varepsilon_{t-1}) = D(\varepsilon_t),$$

$$\mathrm{Cov}(y_{t-1}, \varepsilon_{t-1}) = \mathrm{Cov}(y_t, \varepsilon_t). \qquad \textbf{(8.21)}$$

Таким образом, оценка γ (**8.20**) при увеличении объема выборки сходится по вероятности к величине вида

[1] При записи модели *ADL*(0,1) удобнее коэффициент β_0 обозначать просто β.

$$\gamma + \frac{\mathrm{Cov}(y_{t-1}, \varepsilon_t)}{D(y_t)}. \qquad \textbf{(8.22)}$$

Найдем величину (**8.22**) в явном виде. Взяв дисперсии от обеих частей равенств (**8.18**) и (**8.19**) и используя (**8.21**), получим:

$$D(\varepsilon_t) = \frac{\sigma^2}{1-\rho^2}. \qquad \textbf{(8.23)}$$

$$D(y_t) = \gamma^2 D(y_t) + 2\gamma \mathrm{Cov}(y_{t-1}, \varepsilon_t) + D(\varepsilon_t). \qquad \textbf{(8.24)}$$

Далее:

$$\mathrm{Cov}(y_t, \varepsilon_t) = \gamma \mathrm{Cov}(y_{t-1}, \varepsilon_t) + D(\varepsilon_t), \qquad \textbf{(8.25)}$$

$$\begin{aligned}\mathrm{Cov}(y_{t-1}, \varepsilon_t) &= \mathrm{Cov}(y_{t-1}, \rho\varepsilon_{t-1} + \varepsilon_t) = \rho \mathrm{Cov}(y_{t-1}, \varepsilon_{t-1}) = \\ &= \rho \mathrm{Cov}(y_t, \varepsilon_t).\end{aligned} \qquad \textbf{(8.26)}$$

Исключая $\mathrm{Cov}(y_t, \varepsilon_t)$ из равенств (**8.25**), (**8.26**), получим, используя (**8.23**):

$$\mathrm{Cov}(y_{t-1}, \varepsilon_t) = \frac{\sigma^2 \rho}{(1-\rho^2)(1-\gamma\rho)}. \qquad \textbf{(8.27)}$$

Подставляя выражение (**8.23**) в (**8.24**), получаем с учетом (**8.27**):

$$D(y_t) = \frac{\sigma^2 (1+\gamma\rho)}{(1-\rho^2)(1-\gamma^2)(1-\gamma\rho)}. \qquad \textbf{(8.28)}$$

Подставляя (**8.27**), (**8.28**) в (**8.22**), получаем, что предел по вероятности оценки (**8.20**) равен

$$\frac{\gamma + \rho}{1 + \gamma\rho}.$$

Очевидно, что несостоятельность оценки (**8.20**) тем больше, чем сильнее автокорреляция ошибок ε. На практике, однако, часто выполняется условие $\rho << \gamma$. В этом случае *предел оценки наименьших квадратов будет близок к истинному значению параметра, хотя и не равен ему.*

Приведенные здесь выкладки можно почти дословно повторить и в том случае, если в модели присутствует объясняющая переменная *X*, не коррелирующая с ошибками регрессии. Приведем их окончательный результат. Оценка (**8.20**) сходится по вероятности к величине вида

$$\gamma + \frac{\sigma^2\rho\left(1-\gamma^2\right)}{\beta^2 D\left(x_t\right)\left(1-\rho^2\right)\left(1-\gamma\rho\right)+\sigma^2(1+\gamma\rho)}.$$

Если условие $\rho<<\gamma$ *не выполняется, обычный метод наименьших квадратов может давать существенное отклонение от истинного результата даже на выборках большого объема.*

Аналогичные выкладки можно проделать и для моделей (**8.15**), (**8,16**). Приведем конечный результат. При увеличении объема выборки оценка параметра γ сходится по вероятности к величине

$$\gamma + \frac{\rho\sigma^2\left(1-\gamma^2\right)}{\beta^2 D\left(x_t\right)+\sigma^2\left(1+2\gamma\rho+\rho^2\right)}. \qquad \textbf{(8.29)}$$

8.4. Оценивание моделей с распределенными лагами. Нелинейный метод наименьших квадратов

Рассмотрим модель (**8.15**) и запишем уравнение модели в момент времени $t-1$

$$y_{t-1} = \alpha + \beta x_{t-1} + \gamma y_{t-2} + \varepsilon_{t-1}. \qquad \textbf{(8.30)}$$

Подставим выражение (**8.30**) в (**8.15**). Получим:

$$y_t = \alpha\left(1+\gamma\right)+\beta x_t + \beta\gamma x_{t-1} + \gamma^2 y_{t-2} + \varepsilon_t + \gamma\varepsilon_{t-1}. \qquad \textbf{(8.31)}$$

Далее запишем уравнение (**8.15**) в момент времени $t-2$ и подставим полученное значение y_{t-2} в (**8.31**). Продолжая этот процесс до бесконечности, получим:

$$y_t = \frac{\alpha}{1-\gamma} + \beta x_t + \beta\sum_{k=1}^{\infty} y_1^k x_{t-k} + \sum_{k=0}^{\infty} y^k \varepsilon_{t-k}. \qquad \textbf{(8.32)}$$

Модель (**8.32**) называется моделью с *распределением Койка* лаговых объясняющих переменных. Ее еще иногда называют *моделью с геометрическим распределением*, имея в виду, что коэффициенты при лаговых переменных образуют геометрическую прогрессию со знаменателем γ_1 (напомним, что $\gamma_1<1$). Преобразование модели (**8.15**) к виду (**8.32**) называется *обратным преобразованием Койка*.

Заметим, что переменные X не коррелируют с ошибками ε, так что, применив обратное преобразование Койка, мы решили

проблему коррелированности регрессоров со случайными членами. Однако применение обычного метода наименьших квадратов к модели (**8.32**) оказывается на практике невозможным из-за бесконечно большого количества регрессоров. Разумеется, в силу того, что коэффициенты входящего в модель ряда убывают в геометрической прогрессии, и, стало быть, сам ряд быстро сходится, можно было бы ограничиться сравнительно небольшим числом лагов. Однако и в этом случае мы столкнулись бы по крайней мере с двумя трудно решаемыми проблемами. Во-первых, возникла бы сильная мультиколлинеарность, так как естественно ожидать, что лаговые переменные сильно коррелированы. Во-вторых, уравнение оказалось бы неидентифицируемым. В модели на самом деле присутствует всего четыре параметра. Между тем как, взяв всего лишь три лага, мы бы получили оценки пяти параметров.

Уравнение (**8.32**) может быть оценено с помощью процедуры, которая называется *нелинейным методом наименьших квадратов*. Опишем эту процедуру.

1. С достаточно мелким шагом (например, 0,01) перебираются все значения γ из возможной области значений этого параметра (если никакой априорной информации не имеется, то эта область — интервал (0, 1). Получается последовательность значений $\gamma^{(a)}$.

2. Для каждого значения $\gamma_1^{(a)}$ вычисляется значение

$$x_t^{(a)} = x_t + \sum_{k=1}^{k} \left(\gamma^{(a)}\right)^k x_{t-k} .$$

Предел суммирования K выбирается так, что дальнейшие члены ряда вносят в его сумму незначительный вклад (очевидно, чем меньше выбранное значение $\gamma^{(a)}$, тем меньше членов ряда приходится учитывать).

3. Для каждого $X^{(a)}$ методом наименьших квадратов оценивается уравнение

$$y_t = \alpha + \beta x_t^{(a)} + u_t . \qquad \textbf{(8.33)}$$

4. Выбирается то уравнение (**8.33**), которое обеспечивает наибольший коэффициент детерминации R^2. Соответствующее значение $\hat{\gamma}^{(a)}$ принимается за оценку параметра γ. Вычисляются оценки α, β.

5. Оценки исходных параметров находятся следующим образом:

$$\hat{\alpha}_1 = \hat{\alpha}(1-\gamma),\ \hat{\beta}_1 = \hat{\beta}\hat{\gamma}.$$

Процедура нелинейного метода наименьших квадратов реализована в большинстве компьютерных пакетов.

Обратим внимание на то, что хотя с помощью обратного преобразования Койка устранена коррелированность регрессоров с ошибками, но автокорреляция ошибок приобретает сложную структуру, и устранение ее может оказаться практически невозможным. Так что хотя получаемые таким образом оценки оказываются с о с т о я т е л ь н ы м и, они обладают всеми теми недостатками, о которых подробно говорилось в гл. **7**.

8.5. Оценивание моделей с лаговыми переменными. Метод максимального правдоподобия

Вновь обратимся к модели (**8.15**)—(**8.16**) и исключим из уравнений случайную величину ε. Запишем уравнение (**8.15**) в момент времени $t-1$, умножим его на ρ и вычтем из исходного уравнения (**8.15**). Учитывая, что $\varepsilon_t - \rho\varepsilon_{t-1} = \xi_t$, получим:

$$y_t = \alpha + \beta x_1 - \beta\rho x_{t-1} + (\gamma+\rho)y_{t-1} - \gamma\rho y_{t-2} + \xi_t. \qquad \textbf{(8.34)}$$

Если случайные величины ξ имеют нормальное распределение, то уравнение (**8.34**) может быть оценено методом максимального правдоподобия (см. § **2.7**). Так как в случае нормального распределения ошибок регрессии оценки максимального правдоподобия совпадают с оценками метода наименьших квадратов, на практике применение этого метода к модели (**8.15**) сводится к нелинейной задаче минимизации по α, β, γ и ρ функции

$$L(\alpha,\beta,\gamma,\rho) = \sum_{t=1}^{n}\left[y_t - \alpha - \beta x_t + \beta\rho x_{t-1} + (\gamma+\rho)y_{t-1} - \gamma\rho y_{t-2}\right]^2. \qquad \textbf{(8.35)}$$

Как известно, оценки, получаемые методом максимального правдоподобия, оказываются наиболее эффективными, однако применение этого метода требует знания в и д а р а с п р е д е л е н и я ошибок регрессии. Так, *минимизируя функцию*

(8.35), *мы получим наиболее точные оценки параметров, но лишь в том случае, если эта функция действительно является логарифмом функции правдоподобия, т. е. ошибки* ξ *действительно имеют нормальное распределение.*

В некоторых случаях нормальное распределение ошибок регрессии может следовать из некоторых теоретических соображений, — например, может использоваться центральная предельная теорема. Однако чаще всего нормальное распределение ошибок является модельным допущением, которое принимается в силу того лишь, что ему нет явных противоречий. В этом случае использование метода максимального правдоподобия требует определенной осторожности, и лучше использовать его в сочетании с другими методами.

Отметим также следующее обстоятельство. Если остатки ряда модели подчинены процессу скользящей средней, уравнение с нормально распределенными ошибками будет содержать бесконечное число лагов переменной Y. Коэффициенты при них убывают в геометрической прогрессии, и можно ограничиться несколькими первыми членами. В этом случае метод максимального правдоподобия практически равносилен нелинейному методу наименьших квадратов.

В последующих параграфах мы рассмотрим конкретные примеры моделей и применим к их оцениванию все перечисленные методы.

8.6. Модель частичной корректировки

В экономических приложениях часто встречается ситуация, когда под воздействием объясняющей переменной X формируется не сама величина Y, а ее идеальное, «желаемое» (*desired*) значение Y^{des}. Реальное же, наблюдаемое значение y_t представляет собой взвешенную сумму желаемого значения в момент времени t и наблюдаемого в предыдущий момент $t-1$.

Пусть, например, некая фирма выплачивает в момент времени t дивиденды y_t. Естественно считать, что сумма дивидендов представляет собой некоторую долю прибыли фирмы x_t:

$$y_t = \gamma x_t . \qquad \textbf{(8.36)}$$

На практике, однако, уравнение **(8.36)** подвергается частичной корректировке. Если прибыль окажется малой, на долю дивидендов уйдет бóльшая часть, чем γ, ибо известно, что уменьше-

ние дивидендов наносит серьезный удар по престижу фирмы. По этой же причине в случае большей прибыли доля дивидендов окажется меньшей — фирма проявит осторожность: возможно, в будущем периоде прибыль уменьшится, и тогда придется урезать дивиденды (другим сдерживающим рост дивидендов фактором может послужить желание фирмы инвестировать часть прибыли в расширение производства). В результате реальное изменение дивидендных выплат Δy_t формируется следующим образом:

$$\Delta y_t = \lambda(y_t^{\text{des}} - y_{t-1}) + \nu_t \quad (0 < \lambda < 1). \qquad \textbf{(8.37)}$$

Уравнение (**8.37**) называется уравнением *частичной корректировки* (соответствующая модель выплаты дивидендов была наиболее подробно рассмотрена Дж. Линтнером). Равенство (**8.37**) вместе с равенством (**8.36**) дают следующую модель:

$$y_t = \lambda(y_t^{\text{des}} + (1-\lambda)y_{t-1} + \nu_t, \qquad \textbf{(8.38)}$$
$$y_t^{\text{des}} = \alpha + \beta x_t.$$

Модель (**8.38**) называется *моделью частичной корректировки*. (Для модели выплаты дивидендов можно считать, что $\alpha = 0$, в общем же случае свободный член присутствует в уравнении формирования желаемого значения.) Здесь важно отметить, что величина Y^{des} является н е н а б л ю д а е м о й. (В уравнение ее формирования также можно добавить свободный член, но это ничего не изменит при рассмотрении модели в целом.)

Очевидно, регрессионное уравнение модели (**8.38**) может быть записано в виде:

$$y_t = \alpha\lambda + \beta\lambda x_t + (1-\lambda)y_{t-1} + \nu_t. \qquad \textbf{(8.39)}$$

Уравнение (**8.39**) представляет собой уравнение *ADL* порядка (0,1) и может быть оценено нелинейным методом наименьших квадратов после обратного преобразования Койка. Заметим, впрочем, что состоятельные оценки параметров уравнения (**8.39**) можно получить и *обычным методом наименьших квадратов*, так как в уравнении объясняющая переменная Y_{t-1} не коррелирует со значением случайного члена ε в момент времени t (см. § **8.1**).

8.7. Модель адаптивных ожиданий

Другим важным примером *регрессионной модели с распределенными лагами* является модель адаптивных ожиданий.

Пусть $Y_t = \log (M_t/P_t)$, где M — номинальное количество денег в обращении, P — уровень цен. Величина M/P называется *реальными денежными остатками*. Пусть Y_t^d — спрос на реальные денежные остатки.

Ф. Кейган, изучая динамику этой величины в периоды гиперинфляции, выдвинул предположение, что ее значение в момент t определяется *ожидаемым* уровнем инфляции в момент $t+1$. Модель, предложенная Кейганом, имеет вид:

$$y_t = \alpha + \beta x_{t+1}^w + \varepsilon_t , \qquad \textbf{(8.40)}$$

где x^w — ожидаемый уровень инфляции. Очевидно, это ненаблюдаемая величина. Кейган дополнил уравнение (**8.40**) уравнением

$$\Delta x_{t+1}^w = \lambda\left(x_t - x_t^w\right), \qquad \textbf{(8.41)}$$

где $\Delta x_{t+1}^w = x_{t+1}^w - x_t^w$.

Перепишем уравнения (**8.40**), (**8.41**) в виде:

$$x_{t+1}^w = \lambda x_t + (1-\lambda) x_t^w , \qquad \textbf{(8.41′)}$$

где $y_t = \alpha + \beta x_{t+1}^w + \xi_t$. **(8.42)**

Теперь становится понятным смысл уравнения (**8.41′**) — первого уравнения системы (**8.42**), состоящий в том, что ожидаемый уровень инфляции в момент $t+1$ представляет собой взвешенную сумму ожидаемого и реального уровня в момент t.

Модель (**8.42**) называется *моделью адаптивных ожиданий*. (Модель гиперинфляции Ф. Кейгана, по-видимому, представляет собой впервые рассмотренный пример такой модели.) Из уравнений (**8.40**), (**8.41**) может быть исключена ненаблюдаемая величина X^w. В самом деле, запишем уравнение (**8.41**) в момент времени $t-1$:

$$y_{t-1} = \alpha + \beta x_t^w + \xi_{t-1} .$$

В то же время имеем:

$$y_t = \alpha + \beta(\lambda x_t + (1-\lambda)\, x_t^w) + \xi_t =$$

$$= \alpha + \beta\lambda x_t + (1-\lambda)\beta x_t^w + \xi_t = \alpha + \beta\lambda x_t + (1-\lambda)\left(y_{t-1} - \alpha - \xi_{t-1}\right) + \xi_t .$$

Окончательно получим:

$$y_t = \alpha\lambda + \beta\lambda x_t + (1-\lambda) y_{t-1} + \varepsilon_t , \qquad \textbf{(8.43)}$$

где

$$\varepsilon_t = \xi_t - (1-\lambda)\xi_{t-1} . \qquad \textbf{(8.44)}$$

Модель (**8.43**) есть модель с распределенными лагами *ADL*(0,1), причем динамика случайного члена подчинена закону скользящей средней *MA*(1).

Явный вид (**8.44**) случайного члена показывает, что имеет место корреляция между лаговой переменной Y_{t-1} и ошибкой регрессии ε_t, то есть оценки метода наименьших квадратов не будут состоятельными.

Модель (**8.43**) можно оценить, применив обратное преобразование Койка и затем нелинейный метод наименьших квадратов.

Рассмотрим пример. Пусть X_t — цена на сырье в момент времени t, а Y_t — цена форвардной сделки на товар в момент времени t+1. Очевидно, Y_t зависит в большей степени от ожидаемого значения X_{t+1}^w на сырье в момент времени t+1, т. е. зависимость Y от X может быть описана моделью адаптивных ожиданий (**8.42**) или (в преобразованном виде) уравнением (**8.43**).

Имеется выборка из $n = 250$ наблюдений, данные о которой представлены на рис. 8.1, 8.2.

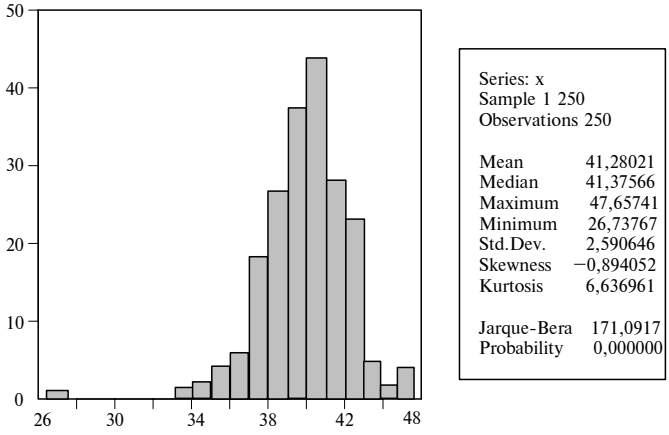
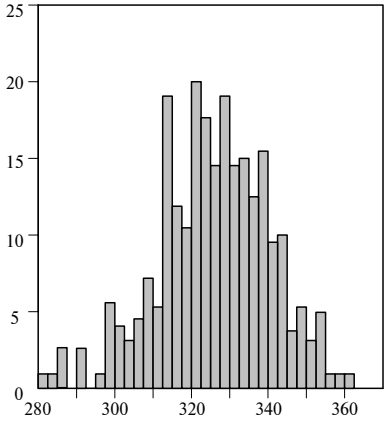


Рис. 8.1



Series Y	
Sample 4 249	
Observations 246	
Mean	325,4976
Median	325,7072
Maximum	362,2785
Minimum	280,5822
Std.Dev.	14,78411
Skewness	−0,245159
Kurtosis	3,106647
Jarque-Bera	2,580791
Probability	0,275162

Рис. 8.2

Применим сначала к уравнению (**8.43**) обычный метод наименьших квадратов. Получим уравнение вида:

$$y_t = -0{,}15 + 4{,}78x_t + 0{,}39y_{t-1}, \quad d = 2{,}55.$$
$$(3{,}32)\ \ (0{,}06)\ \ \ (0{,}01)$$

Из уравнения (**8.43**) следует, что значения оценок параметров α, β, λ находятся как решения системы

$$\begin{cases} \alpha\lambda = -0{,}15; \\ \beta\lambda = 4{,}78; \\ 1-\lambda = 0{,}39. \end{cases}$$

Таким образом, получаем следующие оценки:

$$\lambda=0{,}6; \quad \beta=7{,}88; \quad \alpha=-0{,}248. \qquad \textbf{(8.45)}$$

Насколько можно доверять полученным результатам? Оценим правую часть (**8.29**). Имеем $D(X_t)=2{,}59^2 = 6{,}7$. Выборочная дисперсия остатков ряда $D(e_t)=4{,}516$ — примем ее за оценку $D(\varepsilon)$. Тогда оценкой $D(\xi_t)$ будет величина $D(\varepsilon_t)/[1 + (1-\lambda)^2] = 3{,}89$. Полагаем также $\rho - (1-\lambda)=-\gamma=0{,}4$. Подставляя все эти значения в (**8.29**), получим оценку для предела γ по вероятности: $0{,}97576\gamma$.

Отсюда следует, что при больших выборках значение γ будет несколько занижено, впрочем, занижение это не очень значительно. Таким образом, мы можем ожидать, что оценки обычного

метода наименьших квадратов не очень значительно отличаются от истинных значений, хотя, безусловно, являются неточными.

Применим теперь к уравнению (**8.43**) обратное преобразование Койка:

$$y_t = \alpha + \beta\lambda\left[1 + \sum_{k=0}^{\infty}(1-\lambda)^k x_{t-k}\right] + \nu_t, \qquad \textbf{(8.46)}$$

где $\nu_t = \sum_{k=0}^{\infty}(1-\lambda)^k \varepsilon_{t-k}$, и оценим уравнение (**8.46**) нелинейным методом наименьших квадратов. Получим следующие значения оценок параметров α, β, λ (здесь ряд, стоящий в правой части уравнения (**8.46**), приближен первыми двадцатью членами):

$$\alpha = -2{,}2; \quad \beta = 7{,}93; \quad \lambda = 0{,}61. \qquad \textbf{(8.47)}$$

Наиболее заметное различие оценок (**8.47**) и (**8.45**) касается константы α, однако и в том, и в другом случае константа является незначимой. Оценки β и λ оказываются близкими, но все же различающимися. Очевидно, у нас больше оснований доверять оценкам (**8.47**).

8.8. Модель потребления Фридмена

Наиболее известным примером модели адаптивных ожиданий является модель потребления М. Фридмена. Рассмотрим ее подробнее.

Изучая зависимость между потреблением и доходом индивидуумов, М. Фридмен предположил, что пропорциональная зависимость должна строиться не между фактическими величинами, а между их постоянными составляющими. *Постоянный доход* — это сумма, на которую человек может рассчитывать в относительно долгосрочный период (заработная плата, стабильные гонорары, проценты с вкладов и т. д.). Фактический доход в рассматриваемый момент времени может значительно отличаться от постоянного. Аналогично, *постоянное потребление* — это, по сути, привычный уровень потребления. Его фактическое значение оказывается сильно отличающимся от постоянного в случае крупной покупки или непредвиденных расходов (подробнее о понятиях постоянного и временного дохода и потребления можно ознакомиться в [9]).

М. Фридмен исходил из предположения, что постоянное потребление индивида Y^c пропорционально его постоянному доходу X^c, т. е.

$$y_t^c = \beta x_t^c . \qquad \textbf{(8.48)}$$

В то же время фактическое потребление Y и фактический доход X представляют собой сумму постоянных и временных величин:

$$Y = Y^c + Y^T,\ X = X^c + X^T, \qquad \textbf{(8.49)}$$

причем временные величины X^T, Y^T являются случайными. Гипотеза Фридмена заключалась в том, что эти величины не коррелируют в различные моменты времени и между собой, их математические ожидания равны нулю, а дисперсии постоянны во времени.

Подставив выражение (**8.48**) в (**8.49**), получим регрессионную модель

$$y_t = \beta x_t^c + y_t^T . \qquad \textbf{(8.50)}$$

Проблема заключается в том, что постоянные доход и потребление являются субъективными, а следовательно, ненаблюдаемыми величинами. М. Фридмен применил к изучению зависимости (**8.50**) *модель адаптивных ожиданий*, предположив, что изменение постоянного дохода пропорционально разности между его реальным значением и предыдущим постоянным значением, т. е.

$$\Delta x_t^c = \lambda\left(x_t - x_{t-1}^c\right). \qquad \textbf{(8.51)}$$

Это означает, что при увеличении реального дохода индивиды корректируют свое представление о постоянном доходе, но не на полное значение прироста, а на некоторую его часть, понимая, что приращение может оказаться обусловленным временной, т. е. случайной, составляющей. Уравнение (**8.51**) может быть записано в стандартной форме модели адаптивных ожиданий:

$$x_t^c = \lambda x_t + (1-\lambda)x_{t-1}^c . \qquad \textbf{(8.52)}$$

Повторяя уже приведенные ранее выкладки, можно записать модель адаптивных ожиданий (**8.50**), (**8.51**) в виде следующего уравнения:

$$y_t = \beta\lambda x_t + (1-\lambda)x_{t-1} + \varepsilon_t , \qquad \textbf{(8.53)}$$

где $\varepsilon_t = \left[y_t^T - (1-\lambda)y_{t-1}^T\right]$.

Очевидно, в модели (**8.53**) имеет место коррелированность регрессора Y_{t-1} со случайным членом.

Модель (**8.53**) может быть оценена с помощью *нелинейного метода наименьших квадратов* (см. предыдущие примеры). Также к модели (**8.50**) может быть применен метод инструментальных переменных. Вероятно, впервые это было сделано Н. Левиатаном, который использовал в качестве инструментальных переменных для X^c фактические доход и потребление на другом временном отрезке (подробнее о различных аспектах модели М. Фридмена см. [9]).

8.9. Автокорреляция ошибок в моделях со стохастическими регрессорами

Внимательный читатель мог обратить внимание на одно, казалось бы, противоречивое обстоятельство: в рассматриваемых примерах значение статистики Дарбина—Уотсона оказывалось достаточно близким к двум, что, казалось бы, указывает на отсутствие автокорреляции. Лишь в модели адаптивных ожиданий из § **8.7** значение статистики d попадает в критическую область, да и то на самую ее границу.

Объяснение здесь очень простое: *тест Дарбина—Уотсона неприменим в том случае, если имеется корреляция между регрессорами и ошибками регрессии.* В самом деле, идея теста заключается в том, что корреляция ошибок регрессии имеет место в том и только том случае, когда она значимо присутствует в остатках регрессии. Но для того, чтобы это было действительно так, необходимо, чтобы набор значений остатков можно было бы интерпретировать как набор наблюдений ошибок. Между тем это не так, если регрессоры коррелируют с ошибками.

Можно показать, что в этом случае *значение статистики Дарбина—Уотсона будет часто попадать в область принятия гипотезы об отсутствии автокорреляции и в том случае, если на самом деле эта гипотеза неверна.* Это обстоятельство и делает тест Дарбина—Уотсона неприменимым и обусловливает необходимость других инструментов для обнаружения автокорреляции ошибок регрессии в моделях со стохастическими регрессорами.

В модели с распределенными лагами *ADL* (0,1) (заметим, что все рассматриваемые нами модели относились именно к этому

типу) для выявления автокорреляции ошибок можно применять *h-тест Дарбина*. Рассмотрим модель

$$y_t = \alpha + \beta x_t + \gamma y_{t-1} + \varepsilon_t . \qquad \textbf{(8.54)}$$

Предположим, что подозревается наличие авторегрессии первого порядка в динамике ε:

$$\varepsilon_t = \rho \varepsilon_{t-1} + \xi_t . \qquad \textbf{(8.55)}$$

Рассмотрим следующую статистику

$$h = (1 - 0{,}5d)\sqrt{\frac{n}{1 - nD(\hat{\gamma})}}, \qquad \textbf{(8.56)}$$

где d — значение статистики Дарбина—Уотсона уравнения (**8.54**), $\hat{\gamma}$ — оценка параметра γ, полученная применением к (**8.54**) обычного метода наименьших квадратов.

При справедливости гипотезы $\rho = 0$ распределение статистики h при увеличении объема выборки стремится к нормальному с математическим ожиданием, равным нулю, и дисперсией, равной единице. Таким образом, гипотеза об отсутствии автокорреляции ошибок отвергается, если наблюдаемое значение статистики h окажется больше, чем критическое значение стандартного нормального распределения.

Рассмотрим следующий модельный пример. Методом Монте-Карло (см. §**13.2**) была сымитирована модель

$$y_t = \alpha + \gamma y_{t-1} + \varepsilon_t , \qquad \textbf{(8.57)}$$

$$\varepsilon_t = \rho \varepsilon_{t-1} + \xi_t \qquad \textbf{(8.58)}$$

с модельными значениями $\alpha = 10$, $\gamma = 0{,}6$, $\rho = 0{,}3$.

Применив к уравнениям (**8.57**), (**8.58**) обычный метод наименьших квадратов, получили следующие результаты:

$\hat{y}_t = \underset{(1{,}15)}{5{,}257} + \underset{(0{,}045)}{0{,}79} y_{t-1}$, $d = 1{,}7$ для уравнения (**8.57**) и $\hat{\varepsilon}_t = \underset{(0{,}07)}{0{,}15} \varepsilon_{t-1}$, $d = 1{,}96$ для уравнения (**8.58**).

Таким образом, получены следующие оценки:

$$\hat{\alpha} = 5{,}257,\ \hat{\gamma} = 0{,}79,\ \hat{\rho} = 0{,}15.$$

Как мы знаем, эти оценки значительно отличаются от истинных значений параметров. Так как в рассматриваемом примере значения β и ρ вполне сравнимы, мы должны ожидать, что метод наименьших квадратов приведет к существенной несостоятельности оценок, что мы и наблюдаем в действительности.

Обратим особое внимание на достаточно близкое к двум значение статистики Дарбина—Уотсона для уравнения (**8.57**). На этот раз *значение d попадает в область принятия гипотезы*, что, очевидно, противоречит уравнению (**8.58**). Правильный результат можно получить с помощью *h*-теста Дарбина. Имеем:

$$d = 1{,}7;\ D(\hat{\gamma}) = 0{,}045;\ n = 185.$$

Подставляя эти значения в (**8.56**), получаем $h = 2{,}64$. Так как это значение больше критического $h_{0,05} = 1{,}96$, определяемого для нормального закона, гипотеза об отсутствии автокорреляции ошибок отвергается, имеет место авторегрессия ошибок первого порядка (еще раз заметим, что для рассматриваемой модели этот вывод был априорно очевиден).

В настоящем примере известно, что случайные величины ξ имеют нормальное распределение, так что наиболее точные оценки дает метод максимального правдоподобия. Применяя его, получим:

$$\hat{\alpha} = 9{,}338376;\ \hat{\beta} = 0{,}313694;\ \hat{\gamma} = 0{,}627373.$$

Эти значения уже достаточно близки к модельным. Большей точности можно было бы добиться, сымитировав модель на выборке большего объема.

8.10. *GARCH*-модели

Здесь мы отметим еще один тип моделей временных рядов со специфической зависимостью ошибок регрессии.

Пусть x_t и y_t — стационарные временные ряды, $t = 1,..., n$. Рассмотрим регрессионную модель

$$y_t = \beta x_t + \varepsilon_t, \qquad \textbf{(8.59)}$$

удовлетворяющую следующим условиям:

$$M_{\varepsilon_{t-1}}(\varepsilon_t) = 0; \qquad \textbf{(8.60)}$$

где $$D_{\varepsilon_{t-1}}(\varepsilon_t) = \alpha_0 + \alpha_1 \sigma_{t-1}^2, \qquad \textbf{(8.61)}$$

$$\sigma_{t-1}^2 = D(\varepsilon_{t-1}).$$

Очевидно, условие (**8.61**) означает, что большее отклонение от объясненного (прогнозируемого) значения в предыдущем наблюдении приводит к большей вероятности значительного отклонения также и в последующем наблюдении.

Исследования показывают, что подобные явления часто наблюдаются аналитиками финансового рынка: периоды «затишья», когда финансовые показатели лишь незначительно колеблются вокруг среднего, чередуются с периодами «всплеска», характеризующимися широким размахом значений тех же показателей. На рис. 8.3 приведен примерный график подобных наблюдений.

Обратимся к условиям (**8.60**), (**8.61**). Из (**8.60**) следует, что $r(\varepsilon_{t-1}, \varepsilon_t) = 0$, т. е. автокорреляция остатков отсутствует. Из этого же условия следует, что

$$D(\varepsilon_t) = M\left(D_{\varepsilon_{t-1}}(\varepsilon_t)\right),$$

или $$D(\varepsilon_t) = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2.$$

Если $$D(\varepsilon_t) = \frac{\alpha_0}{1 - \alpha_1},$$

то $$D(\varepsilon_t) = D(\varepsilon_{t-1})$$

и безусловная дисперсия ошибок регрессии постоянна, т. е. модель гомоскедастична.

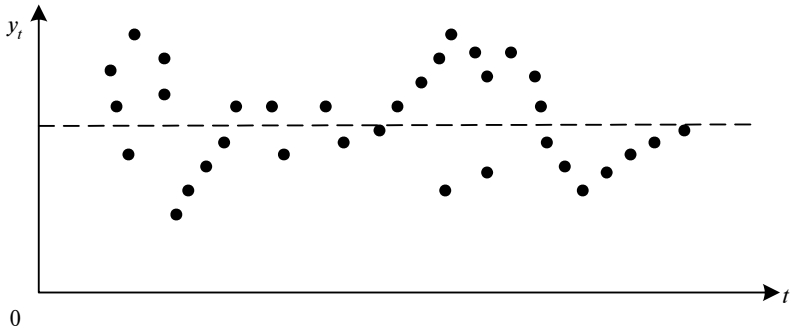


Рис. 8.3

В то же время соотношение (**8.61**) означает, что имеет место *условная гетероскедастичность* ошибок регрессии. Модель, удов-

летворяющая условиям (**8.60**), (**8.61**), называется *авторегрессионной условно гетероскедастичной моделью*, или *ARCH-моделью* (*AutoRegressive Conditional Heteroskedastic model*).

Эта модель допускает обобщения. Если вместо условия (**8.61**) вводится условие

$$D_{\varepsilon_{t-1}\ldots\varepsilon_{t-p}}(\varepsilon_t) = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \ldots + \alpha_p\varepsilon_{t-p}^2,$$

то модель называется *ARCH*(*p*) (арч-модель *p*-го порядка). Рассматриваются также более общие формы зависимости условной дисперсии ошибок, а именно зависимости следующего вида:

$$D_{\varepsilon_{t-1}\ldots\varepsilon_{t-p}}(\varepsilon_t) = \alpha_0 + \sum_{i=1}^{p}\alpha_i\varepsilon_{t-i}^2 + \sum_{i=1}^{q}\gamma_i\sigma_{t-i}^2 .$$

Соответствующая модель называется *обобщенной авторегрессионной условно гетероскедастичной моделью* (*Generalited AutoRegressive Conditional Heteroskedastic model*) порядков *p* и *q*, или *GARCH* (*p*, *q*).

Как же определить, имеется ли в модели условная гетероскедастичность? Как и в случае проверки гипотезы об отсутствии обычной гетероскедастичности, вместо ненаблюдаемых величин — ошибок регрессии — рассматриваются остатки. К модели (**8.59**) применяется обычный метод наименьших квадратов, выбирается порядок *p* и рассматривается регрессия

$$\hat{e}_t^2 = \alpha_0' + \alpha_1' e_{t-1}^2 + \ldots + \alpha_p' e_{t-p}^2.$$

Если при этом гипотеза о незначимости регрессии отвергается, то можно считать, что имеется *ARCH*(*p*)-модель.

Модели *ARCH* и *GARCH* удовлетворяют всем условиям классической модели, и метод наименьших квадратов позволяет получить оптимальные *линейные* оценки. В то же время можно получить более эффективные *нелинейные* оценки методом максимального правдоподобия. В отличие от модели с независимыми нормально распределенными ошибками регрессии в *ARCH*-модели оценки максимального правдоподобия отличаются от оценок, полученных методом наименьших квадратов.

Например, для *ARCH*(1)-модели логарифмическая функция правдоподобия имеет вид:

$$\ln L = -\frac{1}{2}\sum_{t=1}^{n}\ln\left(\alpha_0' + \alpha_1' e_{t-1}^2\right) - \frac{1}{2}\sum_{t=1}^{n}\frac{e_t^2}{\alpha_0' + \alpha_1' e_{t-1}^2}.$$

Ее минимизация с получением соответствующих оценок называется *оцениванием модели методом ARCH*. Соответствующая

процедура присутствует в эконометрических пакетах. При ее компьютерной реализации требуется указать порядок модели.

Использование *ARCH*- и *GARCH*-моделей оказывается в ряде случаев экономико-математического моделирования (например, процессов инфляции и внешней торговли, механизмов формирования нормы процента и т. п.) более адекватным действительности, что позволяет строить более эффективные оценки параметров рассматриваемых моделей по сравнению с оценками, полученными обычным и даже обобщенным методом наименьших квадратов.

8.11. Нестационарные временные ряды

До сих пор мы рассматривали регрессионные модели типа $Y=X\beta+\varepsilon$, в которых ряд остатков рассматривался как стационарный, а нестационарность самих рядов x_t и y_t обуславливалась наличием неслучайной компоненты (тренда). После выделения тренда все ряды оказывались стационарными, причем стационарность считалась заранее известной, априорной. На практике, однако, такая ситуация редко имеет место. Между тем, включение в модель нестационарных рядов может привести к совершенно неверным результатам. В частности, стандартный анализ с помощью метода наименьших квадратов модели

$$y_t = \alpha + \beta x_t + \varepsilon \qquad \textbf{(8.62)}$$

может показать наличие существенной значимости коэффициента β даже в том случае, если величины x_t и y_t являются независимыми. Такое явление носит название *ложной регрессии* и имеет место именно в том случае, когда в модели используются нестационарные временные ряды.

Таким образом, при моделировании какой-либо зависимости между величинами x_t и y_t естественно возникает вопрос: можно ли считать соответствующие временные ряды стационарными?

В настоящем параграфе мы рассмотрим некоторые вопросы, связанные с нестационарными временными рядами. При этом мы не будем ставить задачу обстоятельного изложения теории нестационарных рядов, так как это потребовало бы использования математического аппарата, существенно выходящего за рамки нашего рассмотрения. Поэтому мы ограничимся лишь тем, что затронем основные проблемы, возникающие при эконометрическом моделировании нестационарных временных рядов.

Итак, пусть имеется временный ряд y_t, при этом мы считаем, что в нем отсутствует неслучайная составляющая. Для простоты также будем считать, что среднее его значение равно нулю (очевидно, ряд остатков регрессионной модели удовлетворяет этим условиям).

Если ряд является стационарным, то в каждый следующий момент времени его значение «стремится вернуться к нулевому среднему». Иными словами, если мы будем объяснять значение y_t предыдущим значением y_{t-1}, то объясненная часть ξ_t будет находиться ближе к нулю, чем значение y_{t-1}.

Математически строго это условие можно сформулировать следующим образом: рассмотрим регрессионную модель

$$y_t = \rho y_{t-1} + \xi_t. \qquad \textbf{(8.63)}$$

Истинное значение параметра ρ должно удовлетворять условию $|\rho| < 1$. В случае, если $\rho=1$, мы имеем ситуацию, когда последующее значение одинаково легко может как приближаться к нулевому среднему, так и отдаляться от него. Соответствующий случайный процесс называется «*случайным блужданием*». Очевидно, дисперсия в этом случае растет. В самом деле, из равенства (**8.63**) имеем:

$$D(y_t) = D(y_{t-1}) + \sigma_t^2,$$

т. е. дисперсия $D(y_t)$ неограниченно возрастает, а значит, ряд y_t не является стационарным.

Разумеется, в случае $|\rho| > 1$ ряд тем более не будет стационарным, значения его стремительно нарастают. Соответствующий процесс иногда называется *взрывным*. Однако в реальных экономических задачах он никогда не возникает.

Практика показывает, что чаще всего в эконометрических исследованиях нестационарность рассматриваемого временного ряда носит именно характер случайного блуждания. Таким образом, вопрос о нестационарности ряда y_t, как правило, сводится к следующему: верно ли, что в регрессии $y_t = \rho y_{t-1} + \xi_t$ истинное значение параметра ρ равно единице? Соответствующая задача называется проблемой *единичного корня*.

Итак, пусть имеется временной ряд y_t. Рассмотрим модель авторегрессии

$$y_t = \rho y_{t-1} + \xi_t. \qquad \textbf{(8.64)}$$

Будем предполагать, что ошибки регрессии ξ_t *независимы и одинаково распределены, т. е. образуют белый шум*. Переходя к разностным величинам, перепишем соотношение (**8.64**) в виде:

$$\Delta y_t = \lambda y_{t-1} + \xi_t , \qquad \textbf{(8.65)}$$

где $\Delta y_t = y_t - y_{t-1}$, $\lambda = \rho - 1$.

Тогда проблема единичного корня сводится к следующей: верно ли, что в модели (**8.65**) истинное значение параметра λ равно нулю?

На первый взгляд кажется, что вопрос может быть решен тестированием гипотезы $\lambda=0$ с помощью статистики Стьюдента (§ **3.6**). Однако ситуация оказывается сложнее. В том случае, если ряд y_t на самом деле нестационарный, т. е. если на самом деле $\lambda=0$, стандартная *t*-статистика вида $t = \tilde{\lambda} / \hat{\sigma}_\lambda$ *не имеет распределения Стьюдента*!

Распределение *t*-статистики в этом случае описано Дики и Фуллером. Ими же получены критические значения для отвержения гипотезы о нестационарности ряда. Они существенно отличаются от критических значений распределения Стьюдента. В результате оказывается, что использование обычного *t*-теста приводит к тому, что гипотеза о нестационарности временного ряда отвергается слишком часто, в том числе и тогда, когда ряд действительно является нестационарным.

Таким образом, проблему единичного корня следует решать с помощью *теста Дики—Фуллера*, который реализован в большинстве современных регрессионных пакетов. В «*Econometric Views*» присутствует так называемый *пополненный тест Дики—Фуллера* (*Augmented Dickey—Fuller test — ADF*). Он является обобщением обычного теста Дики—Фуллера: в правую часть выражения (**8.65**) добавляются слагаемые вида $\Delta y_{t-1}, ..., \Delta y_{t-p}$, т. е. тестируется гипотеза $\lambda = 0$ для модели

$$\Delta y_t = \lambda y_{t-1} + \lambda_1 \Delta y_{t-1} + ... + \lambda_p \Delta y_{t-p} + \xi_t .$$

Это соответствует тому, что вместо уравнения (**8.64**) мы рассматриваем уравнение

$$y_t = \rho y_{t-1} + \rho_2 y_{t-2} + ... + \rho_{p+1} y_{t-p-1},$$

т. е. пытаемся идентифицировать ряд как авторегрессионный порядка $p+1$. Добавление приращений Δy_{t-1} производится для того, чтобы с возможно большей достоверностью избавиться от

автокорреляции ошибок. (Напомним, что распределение Дики—Фуллера t-статистики имеет место лишь в том случае, если ошибки являются белым шумом!) Однако добавление приращений в правую часть снижает мощность теста Дики—Фуллера.

Что делать в том случае, если ряд y_t оказался нестационарным? Часто при этом оказывается, что стационарным является ряд приращений Δy_t.

Если ряд Δy_t является стационарным, то исходный нестационарный ряд y_t называется *интегрируемым* (или *однородным*). В более общем случае нестационарный ряд y_t называется *интегрируемым* (*однородным*) *k-го порядка*, если после k-кратного перехода к приращениям

$$d^k y_t = d^{k-1} y_t - d^{k-1} y_{t-1},$$

где $d^1 y_t = \Delta y_t$, получается стационарный ряд $d^k y_t$.

Если при этом стационарный ряд $d^k y_t$ корректно идентифицируется как *ARMA*(p, q), то нестационарный ряд y_t обозначается *ARIMA*(p, q, k). Модель *ARIMA*(p, q, k) означает *модель авторегрессии — проинтегрированной скользящей средней* (*AutoRegressive Integrated Moving Average model*) порядков p, q, k и известна как *модель Бокса—Дженкинса*. Эта модель может достаточно успешно описывать поведение нестационарных временных рядов (в том числе содержащих сезонную и(или) циклическую компоненты), что позволяет эффективно использовать ее в задачах кратко- и среднесрочного автопрогноза. Процедура подбора модели *ARIMA* реализована во многих эконометрических пакетах.

Другим приемом устранения нестационарности является *коинтеграция* нескольких нестационарных рядов в некоторую стационарную линейную комбинацию. Опишем эту процедуру на простом примере двух нестационарных рядов.

Нестационарные ряды x_t и y_t называются *коинтегрируемыми*, если существуют числа λ_1 и λ_2 такие, что ряд $\lambda_1 x_t + \lambda_2 y_t$ является стационарным. Так как умножение на ненулевой множитель, очевидно, не влияет на стационарность, для коинтегрируемых рядов существует стационарный ряд вида $y_t - \beta x_t$.

Оказывается, если ряды x_t и y_t на самом деле являются коинтегрируемыми, то состоятельная оценка параметра β получается как оценка обычного метода наименьших квадратов, примененного к модели

$$y_t = \beta x_t + \varepsilon. \qquad \textbf{(8.66)}$$

Казалось бы, что раз так, то вопрос о наличии коинтеграции может быть решен следующим образом. Методом наименьших квадратов оценивается уравнение (**8.66**) и к ряду $y_t - \hat{\beta} x_t$ применяется тест Дики—Фуллера.

Однако оказывается, что тест Дики—Фуллера в этом случае неприменим! При его использовании гипотеза о нестационарности комбинации будет отвергаться слишком часто. На самом деле критические значения для *t*-статистики в этом случае другие. Они были оценены методом симуляции (методом Монте-Карло, см. гл. **13**). Сравнение наблюдаемого значения *t*-статистики с этими уточненными оценками критических значений составляет суть *коинтеграционного теста* (*Cointegration Test*), который также, как правило, реализуется в эконометрических пакетах.

Подробнее вопросы, связанные с нестационарными временными рядами, изложены в [1], [18].

Упражнения

8.1. Оценивается модель $Y = \beta X + \varepsilon$ обычным методом наименьших квадратов. Получается уравнение вида

$$\hat{y} = 1,2x.$$
$$(0,3)$$

Известно, однако, что измерения регрессора X производятся с ошибками, дисперсия которых оценивается как 10. Можно ли при этом считать, что истинное значение коэффициента β положительно, если: *а*) дисперсия X равна 100; *б*) дисперсия X равна 10?

8.2. Для оценки лагового уравнения $y_t = \gamma y_{t-1} + \varepsilon_t$ применен обычный метод наименьших квадратов и получено уравнение

$$\hat{y}_t = 0,9y_{t-1}, \quad d = 1,95.$$
$$(0,1)$$

Считается, что ошибки регрессии представляют собой стационарный авторегрессионный процесс первого порядка. Можно ли сделать вывод, что коэффициент λ: *а*) больше 0,5; *б*) больше 0,7? Объем выборки достаточно велик.

8.3. Рассматривается модель парной регрессии

$$Y = X\beta + \varepsilon. \qquad \textbf{(8.67)}$$

В следующей таблице приводятся данные наблюдений переменных X, Y и инструментальной переменной Z:

i	$\overline{x}_i$	y_i	z_i	i	x_i	y_i	z_i
1	10,0	20,5	11,0	7	11,5	23,2	11,5
2	10,1	20,6	10,1	8	11,1	22,4	11,1
3	10,2	20,6	10,2	9	11,5	23,9	11,5
4	10,3	21,5	10,3	10	10,2	20,7	10,2
5	11,0	22,2	11,0	11	10,1	21,1	10,1
6	11,6	23,4	11,6	12	11,9	24,3	11,9

Найти оценки параметра β, применяя к уравнению (**8.67**) обычный метод наименьших квадратов и метод инструментальных переменных.

8.4. После приведения модели адаптивных ожиданий к лагированной модели и ее оценки методом наименьших квадратов получено уравнение:

$$\hat{y}_t = 1,2 + 0,2x_t + 0,3y_{t-1}.$$

Найти оценки параметров исходной модели адаптивных ожиданий.

8.5. Методом наименьших квадратов получено следующее уравнение:

$$\hat{y}_t = 2 - 0,2x_t + 0,1y_{t-1}.$$
$$(0,03)\ \ (0,04)$$

Значение статистики Дарбина—Уотсона $d = 1,9$. Какой следует сделать вывод о наличии автокорреляции в модели?

Глава 9
Системы одновременных уравнений

9.1. Общий вид системы одновременных уравнений. Модель спроса и предложения

Одной из причин коррелированности регрессоров со случайными членами могут служить факторы, действующие одновременно и на сами регрессоры, и на объясняемые переменные при фиксированных значениях регрессоров. Иными словами, в рассматриваемой экономической ситуации значения объясняемых переменных и регрессоров формируются одновременно под воздействием некоторых внешних факторов. Это означает, что рассматриваемая модель не полна: ее следует дополнить уравнениями, в которых объясняемыми переменными выступали бы сами регрессоры. Таким образом, мы приходим к необходимости рассматривать *системы одновременных* или *регрессионных уравнений*.

Классическим примером является одновременное формирование спроса Q^d и предложения Q^s товара в зависимости от его цены P:

$$\begin{aligned} Q^d &= \beta_1 + \beta_2 P + \beta_3 I + \varepsilon_1; \\ Q^s &= \beta_4 + \beta_5 P + \varepsilon_2. \end{aligned} \qquad \textbf{(9.1)}$$

Здесь I — доход.

Если предположить, что рынок находится в состоянии равновесия, то в равенствах (**9.1**) следует положить $Q^d = Q^s = Q$. В этом случае наблюдаемое значение P — это цена равновесия, которая формируется одновременно со спросом и предложением. Таким образом, мы должны считать P и Q *объясняемыми* переменными, а величину дохода I — *объясняющей* переменной.

Разделение ролей между переменными в системе одновременных уравнений может быть проинтерпретировано следующим образом: переменные Q и P формируют свои значения, подчиняясь уравнениям (**9.1**), т.е. в н у т р и модели. Такие переменные называются *эндогенными.* Между тем переменная I считается в уравнениях (**9.1**) заданной, ее значения формируются в н е м о д е л и. Такие переменные называются *экзогенными.*

С математической точки зрения, главное отличие между экзогенными и эндогенными переменными заключается в том, что *экзогенные переменные не коррелируют с ошибками регрессии*, между тем как эндогенные могут коррелировать (и, как правило, коррелируют). Естественно предположить, что схожие случайные факторы действуют как на цену равновесия, так и на спрос на товар. Причинная зависимость между переменными и приводит, очевидно, к коррелированности их со случайными членами.

Набор экзогенных переменных может быть различным. Так, например, в модели спроса и предложения в качестве экзогенных переменных к доходу могут быть добавлены процентная ставка, временной тренд и т. д.

Приведем общий вид системы одновременных уравнений. Пусть $Y_1,\ldots, Y_m$ — эндогенные переменные, $X_1,\ldots, X_l$ — экзогенные переменные. Введем блочные матрицы B и Г вида:

$$\mathrm{B} = \begin{pmatrix} \beta_{11} \ldots \beta_{1m} \\ \ldots \ \ldots \ \ldots \\ \beta_{m1} \ldots \beta_{mm} \end{pmatrix}; \ \Gamma = \begin{pmatrix} \gamma_{11} \ldots \gamma_{1l} \\ \ldots \ \ldots \ \ldots \\ \gamma_{m1} \ldots \gamma_{ml} \end{pmatrix}.$$

Тогда общий вид системы одновременных уравнений представляется в матричной форме как

$$\mathrm{B}Y + \Gamma X = \varepsilon, \qquad \textbf{(9.2)}$$

где $$Y = \begin{pmatrix} Y_1 \\ \ldots \\ Y_m \end{pmatrix}; \quad X = \begin{pmatrix} X_1 \\ \ldots \\ X_l \end{pmatrix}; \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \ldots \\ \varepsilon_m \end{pmatrix}.$$

Кроме регрессионных уравнений (они называются также *поведенческими* уравнениями) модель может содержать *тождества*, которые представляют собой алгебраические соотношения между эндогенными переменными.

Например, для модели формирования спроса и предложения и цены равновесия имеем два поведенческих уравнения (**9.1**) и одно тождество $Q^s = Q^d$.

Тождества, вообще говоря, позволяют исключить некоторые эндогенные переменные и рассматривать систему регрессионных уравнений меньшей размерности. Так, в модели спроса и предложения можно положить $Q^s = Q^d = Q$ и рассматривать структурную форму (**9.2**), где[1]

$$Y = \begin{pmatrix} Q \\ P \end{pmatrix}; \quad X = \begin{pmatrix} 1 \\ I \end{pmatrix}; \quad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix};$$

$$B = \begin{pmatrix} 1 & -\beta_2 \\ 1 & -\beta_5 \end{pmatrix}; \quad \Gamma = \begin{pmatrix} -\beta_1 & -\beta_3 \\ -\beta_4 & 0 \end{pmatrix}.$$

В настоящей главе мы ограничимся случаем двух уравнений с двумя эндогенными переменными. Это не приведет ни к какой содержательной потере — все необходимые аспекты теории можно проследить на этом простейшем случае. В то же время такое ограничение позволит нам избежать излишней громоздкости в вычислениях.

Очевидно, что мы всегда можем выделить в левой части системы эндогенные переменные, т. е. записать уравнения в виде:

$$Y_1 = \alpha_1 + \beta_1 X_1 + \gamma_1 Y_2 + \varepsilon_1; \qquad \textbf{(9.3)}$$

$$Y_2 = \alpha_2 + \beta_2 X_2 + \gamma_2 Y_1 + \varepsilon_2. \qquad \textbf{(9.4)}$$

Наборы переменных X_1 и X_2 могут быть произвольными. Параметры β, вообще говоря, векторные. Если применить к уравнениям (**9.3**), (**9.4**) обычный метод наименьших квадратов, то, как показано в главе **8**, получатся несостоятельные оценки параметров α, β, γ. Таким образом, оценивание систем одновременных уравнений требует *специальных методов*, которым и посвящена настоящая глава.

9.2. Косвенный метод наименьших квадратов

В основе предлагаемого метода лежит простая идея. Поскольку препятствием к применению метода наименьших квад-

[1] В матрице-столбце X единица означает фиктивную переменную, умножаемую на свободные члены уравнений системы.

ратов является коррелированность эндогенных переменных со случайными членами, следует разрешить систему уравнений относительно Y, так, чтобы в правых частях уравнений оставались только экзогенные переменные X. Очевидно, что для уравнений (**9.3**), (**9.4**) это всегда можно сделать. Затем применить обычный метод наименьших квадратов к полученным уравнениям и получить оценки некоторых выражений от исходных параметров, из которых потом найти оценки и самих параметров.

Такая процедура называется *косвенным методом наименьших квадратов*. Продемонстрируем его на примере системы (**9.3**)—(**9.4**).

Разрешая уравнения (**9.3**)−(**9.4**) относительно Y_1, Y_2, запишем уравнения в виде:

$$\begin{aligned} Y_1 &= a_1 + b_1 X_1 + c_1 X_2 + \nu_1 , \\ Y_2 &= a_2 + b_2 X_1 + c_2 X_2 + \nu_2 . \end{aligned} \qquad \textbf{(9.5)}$$

где

$$a_1 = \frac{\alpha_1 + \gamma_1 \alpha_2}{1 - \gamma_1 \gamma_2};\ a_2 = \frac{\alpha_2 + \gamma_2 \alpha_1}{1 - \gamma_1 \gamma_2};\ b_1 = \frac{\beta_1}{1 - \gamma_1 \gamma_2};\quad b_2 = \frac{\gamma_2 \beta_1}{1 - \gamma_1 \gamma_2};$$
$$c_1 = \frac{\gamma_1 \beta_2}{1 - \gamma_1 \gamma_2};\quad c_2 = \frac{\beta_2}{1 - \gamma_1 \gamma_2};\quad \nu_1 = \frac{\gamma_1 \varepsilon_2 + \varepsilon_1}{1 - \gamma_1 \gamma_2};\ \nu_2 = \frac{\gamma_2 \varepsilon_1 + \varepsilon_2}{1 - \gamma_1 \gamma_2} \qquad \textbf{(9.6)}$$

(индекс t для простоты опущен).

Для дальнейшего упрощения будем считать, что переменные Y отцентрированы, т. е. $a = 0$. (При практическом применении метода это абсолютно несущественно.) Применив к (**9.5**) обычный метод наименьших квадратов, получим оценки параметров b, c.

$$\begin{aligned} \hat{b}_1 &= \frac{\langle X_2 X_2 \rangle \langle X_1 Y_1 \rangle - \langle X_1 X_2 \rangle \langle X_2 Y_1 \rangle}{\langle X_1 X_1 \rangle \langle X_2 X_2 \rangle - \langle X_1 X_2 \rangle^2}; \\ \hat{c}_1 &= \frac{\langle X_1 X_1 \rangle \langle X_2 Y_1 \rangle - \langle X_1 X_2 \rangle \langle X_1 Y_1 \rangle}{\langle X_1 X_1 \rangle \langle X_2 X_2 \rangle - \langle X_1 X_2 \rangle^2}; \\ \hat{b}_2 &= \frac{\langle X_2 X_2 \rangle \langle X_1 Y_2 \rangle - \langle X_1 X_2 \rangle \langle X_2 Y_2 \rangle}{\langle X_1 X_1 \rangle \langle X_2 X_2 \rangle - \langle X_1 X_2 \rangle^2}; \\ \hat{c}_2 &= \frac{\langle X_1 X_1 \rangle \langle X_2 Y_2 \rangle - \langle X_1 X_2 \rangle \langle X_1 Y_2 \rangle}{\langle X_1 X_1 \rangle \langle X_2 X_2 \rangle - \langle X_1 X_2 \rangle^2}, \end{aligned} \qquad \textbf{(9.7)}$$

где $\langle X_i X_j \rangle = \sum_{i=1}^{n} x_{ti} x_{tj}$, $\langle Y_i Y_j \rangle = \sum_{t=1}^{n} y_{ti} y_{tj}$, $\langle X_i Y_j \rangle = \sum_{t=1}^{n} x_{ti} y_{tj}$,

x_{ti}, x_{tj}, y_{ti}, y_{tj} — значения переменных X_i, X_j, Y_i, Y_j.

Между тем равенства (**9.6**) позволяют однозначно выразить исходные параметры α, β, γ через a, b, c:

$$\hat{\beta}_1 = \frac{\hat{b}_1\hat{c}_2 - \hat{b}_2\hat{c}_1}{\hat{c}_2},\ \hat{\beta}_2 = \frac{\hat{b}_1\hat{c}_2 - \hat{b}_2\hat{c}_1}{\hat{b}_1},$$
$$\hat{\gamma}_1 = \frac{\hat{c}_2}{\hat{c}_1};\ \hat{\gamma}_2 = \frac{\hat{b}_2}{\hat{b}_1}. \qquad \textbf{(9.8)}$$

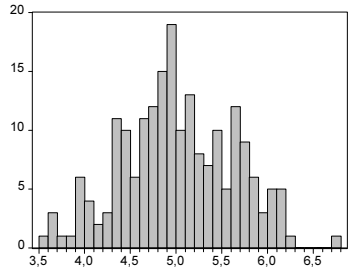
Таким образом, используя (**9.6**), получаем:

$$\hat{\beta}_1 = \frac{\langle X_1Y_1\rangle\langle X_2Y_2\rangle - \langle X_2Y_1\rangle\langle X_1Y_2\rangle}{\langle X_1X_1\rangle\langle X_2Y_2\rangle - \langle X_1X_2\rangle\langle X_1Y_2\rangle};$$
$$\hat{\beta}_2 = \frac{\langle X_1Y_1\rangle\langle X_2Y_2\rangle - \langle X_2Y_1\rangle\langle X_1Y_2\rangle}{\langle X_2X_2\rangle\langle X_1Y_1\rangle - \langle X_1X_2\rangle\langle X_2Y_1\rangle};$$
$$\hat{\gamma}_1 = \frac{\langle X_1X_1\rangle\langle X_2Y_1\rangle - \langle X_1X_2\rangle\langle X_1Y_1\rangle}{\langle X_1X_1\rangle\langle X_2Y_2\rangle - \langle X_1X_2\rangle\langle X_1Y_2\rangle}; \qquad \textbf{(9.9)}$$
$$\hat{\gamma}_2 = \frac{\langle X_2X_2\rangle\langle X_1Y_2\rangle - \langle X_1X_2\rangle\langle X_2Y_2\rangle}{\langle X_2X_2\rangle\langle X_2Y_1\rangle - \langle X_1X_2\rangle\langle X_2Y_1\rangle}.$$

Оценки (**9.8**) называются *оценками косвенного метода наименьших квадратов*. В отличие от оценок прямого применения метода наименьших квадратов оценки (**9.9**) состоятельны.

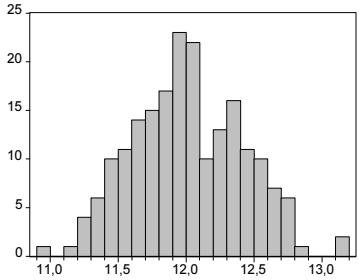
Рассмотрим пример исследования системы (**9.3**)–(**9.4**). Пусть имеются данные по $n = 200$ наблюдениям переменных.

На рис. 9.1 приведены гистограммы и основные числовые характеристики соответствующих выборок.



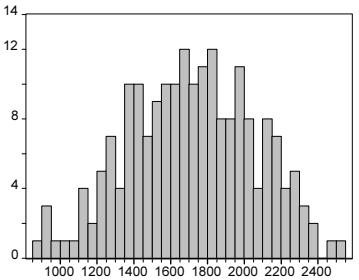
Series: X1	
Sample 1 200	
Observations 200	
Mean	5,019551
Median	4,967329
Maximum	6,715498
Minimum	3,589654
Std. Dev.	0,611454
Skewness	–0,001757
Kurtosis	2,589694
Jarque-Bera	1,403028
Probability	0,495834

а



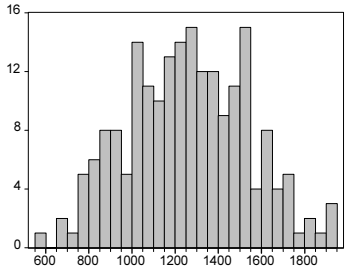
Series: X2	
Sample 1 200	
Observations 200	
Mean	12,00998
Median	11,98016
Maximum	13,12772
Minimum	10,95409
Std. Dev.	0,403707
Skewness	0,111117
Kurtosis	2,590569
Jarque-Bera	1,808515
Probability	0,404842

б



Series: Y1	
Sample 1 200	
Observations 200	
Mean	1712,685
Median	1708,201
Maximum	2545,273
Minimum	896,9183
Std. Dev.	346,7838
Skewness	–0,088008
Kurtosis	2,486183
Jarque-Bera	2,458246
Probability	0,292549

в



Series: Y2	
Sample 1 200	
Observations 200	
Mean	1255,944
Median	1257,273
Maximum	1916,141
Minimum	583,9797
Std. Dev.	279,9589
Skewness	0,092848
Kurtosis	2,482940
Jarque-Bera	2,515282
Probability	0,284324

г

Рис.9.1

Применим сначала *обычный* метод наименьших квадратов. Получим следующие результаты:

$$\hat{y}_1 = 3153{,}451 + 15{,}73x_1 - 1{,}2y_2;\ d = 1{,}894,\ R^2 = 0{,}9999;$$
$$(5{,}127)\quad (0{,}687)\ (0{,}001)$$
$$\hat{y}_2 = 2606{,}23 + 12{,}88x_2 - 0{,}83y_1;\ d = 1{,}893,\ R^2 = 0{,}9999.$$
$$(1{,}75)\quad (0{,}581)\quad (0{,}001)$$

В обоих случаях мы имеем практически стопроцентную подгонку: коэффициент детерминации равен единице с точностью до четвертого знака. Однако, как мы знаем, полученные оценки несостоятельны, и, следовательно, их значения могут заметно отклоняться от истинных значений параметров.

Применим теперь *косвенный* метод наименьших квадратов: оценим регрессионную зависимость Y_i по X_1 и X_2:

$$\hat{y}_1 = 2242{,}755 + 471{,}19x_1 - 241{,}07x_2;\ d = 1{,}96,\ R^2 = 0{,}778,$$
$$(361{,}2)\quad (19{,}04)\quad (28{,}83)$$
$$\hat{y}_2 = 727{,}7 - 376{,}37x_1 + 201{,}29x_2;\ d = 1{,}96,\ R^2 = 0{,}769,$$
$$(297{,}35)\ (15{,}67)\quad (23{,}74)$$

Таким образом, получаем:

$$b_1 = 471{,}19;\ c_1 = -241{,}07;\ b_2 = -376{,}37;\ c_2 = 201{,}29;\quad a_1 = 2242{,}755;\ a_2 = 727{,}7, \tag{9.10}$$

откуда, используя (**9.8**), получаем:

$$\alpha_1 = 3114{,}286;\ \alpha_2 = 2519{,}134;\ \beta_1 = 20{,}43;\ \beta_2 = 8{,}73;\quad \gamma_1 = -1{,}198;\ \gamma_2 = -0{,}8. \tag{9.11}$$

Как видно, полученные таким образом оценки заметно отличаются от полученных прямым методом наименьших квадратов.

9.3. Проблемы идентифицируемости

В рассмотренном примере уравнения (**9.6**) были однозначно разрешимы относительно исходных параметров, что позволило найти их состоятельные оценки. Очевидно, что такая ситуация имеет место не всегда. Рассмотрим эту проблему более подробно.

Форма (**9.2**) называется *структурной формой* системы уравнений. В случае двух уравнений с двумя неизвестными структурной формой будем называть также уравнения (**9.3**)−(**9.4**). Параметры структурной формы называются *структурными параметрами*. Форма (**9.5**) называется *приведенной формой* системы. Параметры приведенной формы оцениваются с помощью метода наименьших квадратов. Однако экономический смысл и интерес для анализа представляют *параметры структурной формы*. Именно структурная форма раскрывает экономический механизм формирования значений эндогенных переменных.

Структурный параметр называется ***идентифицируемым****, если он может быть однозначно оценен с помощью косвенного метода наименьших квадратов.*

Уравнение идентифицируемо, если идентифицируемы все входящие в него структурные параметры.

Структурный параметр называется ***неидентифицируемым****, если его значение невозможно получить, даже зная точные значения параметров приведенной формы.* Наконец, *параметр называется* ***сверхидентифицируемым****, если косвенный метод наименьших квадратов дает несколько различных его оценок.*

Пусть, например, в рассматриваемой модели мы предполагаем, что переменная Y_1 зависит от двух экзогенных переменных X_1, X_2, между тем как динамика Y_2 определяется только эндогенной переменной Y_1, т. е. система уравнений имеет вид:

$$\begin{aligned} Y_1 &= \alpha_1 + \beta_1 X_1 + \beta_2 X_2 + \gamma_1 Y_2 + \varepsilon_1, \\ Y_2 &= \alpha_2 \qquad\qquad\quad + \gamma_2 Y_1. \end{aligned} \tag{9.12}$$

В приведенной форме уравнения (**9.12**) имеют вид (**9.5**), где

$$\begin{aligned} b_1 &= \frac{\beta_1}{1-\gamma_1\gamma_2}, \quad b_2 = \frac{\beta_1\gamma_2}{1-\gamma_1\gamma_2}, \\ c_1 &= \frac{\beta_2}{1-\gamma_1\gamma_2}, \quad c_2 = \frac{\beta_2\gamma_2}{1-\gamma_1\gamma_2}, \end{aligned} \tag{9.13}$$

что может быть переписано в виде

$$\frac{\beta_1}{1-\gamma_1\gamma_2} = b_1, \quad \frac{\beta_2}{1-\gamma_1\gamma_2} = c_1, \tag{9.14}$$

$$\gamma_2 = \frac{b_2}{b_1} = \frac{c_2}{c_1}. \tag{9.15}$$

Очевидно, что три коэффициента β_1, β_2, γ_1 не могут быть найдены из двух уравнений (**9.14**). Это означает, что *существует бесконечное множество их возможных значений, приводящих к одной и той же приведенной форме*. Такие коэффициенты называются ***неидентифицируемыми*** и, соответственно, ***неидентифицируемым*** называется уравнение, содержащее эти параметры.

В то же время для определения γ_2 мы имеем две различные возможности, задаваемые соотношением (**9.15**). При этом заметим, что необходимо выполнение равенства

$$\frac{b_2}{b_1} = \frac{c_2}{c_1}.$$

Но хотя это равенство выполняется для истинных (неизвестных) значений параметров *c* и *b*, для их оценок оно, конечно, выполняться не будет.

В качестве примера рассмотрим модель (**9.12**) с данными из примера 1 (см. § **9.2**). Оценки параметров приведенной модели имеют значения (**9.10**).

Отсюда имеем:

$$\frac{b_2}{b_1} = -0{,}8\,;\ \frac{c_2}{c_1} = -0{,}835.$$

Параметр, для которого существует несколько способов выражения через коэффициенты приведенной формы, является ***сверхидентифицируемым***. Таков параметр γ_2 из рассматриваемого примера. Для сверхидентифицируемого параметра имеется несколько, вообще говоря, различных оценок.

Заметим, что проблема сверхидентифицируемости — это проблема количества наблюдений: с увеличением объема выборки все различные состоятельные оценки параметра стремятся к одному и тому же истинному значению. Между тем проблема неидентифицируемости — это проблема структуры модели. Неидентифицируемость не исчезает с ростом количества наблюдений и означает, что существует бесконечное число структурных моделей, имеющих одну и ту же приведенную форму.

Неидентифицируемость вовсе не является редким явлением. В самом деле для идентифицируемости, грубо говоря, надо, чтобы количество оцениваемых структурных параметров было бы равно количеству оцененных параметров приведенной формы. Очевидно, однако, что в общем случае структурных параметров больше.

Очевидно, неидентифицируемость модели означает, что косвенный метод наименьших квадратов неприменим. В после-

дующих параграфах мы рассмотрим другие методы оценивания систем одновременных уравнений.

9.4. Метод инструментальных переменных

Метод инструментальных переменных (см. главу **8**) — один из наиболее распространенных методов оценивания уравнений, в которых регрессоры коррелируют со свободными членами. Именно это явление оказывается характерным для систем одновременных уравнений. Мы рассмотрим отдельно два случая — идентифицируемой и неидентифицируемой системы.

1. Система идентифицируема.

Рассмотрим модель (**9.5**). Для ее коэффициентов метод наименьших квадратов дал оценки (**9.8**). Легко увидеть, что эти оценки совпадают с оценками, полученными методом инструментальных переменных для уравнений

$$\begin{aligned} Y_1 &= \alpha_1 + \beta_1 X_1 + \gamma_1 X_2 + \varepsilon_1; \\ Y_2 &= \alpha_2 + \beta_2 X_2 + \gamma_2 X_1 + \varepsilon_2 \ . \end{aligned} \qquad \textbf{(9.16)}$$

Таким образом, экзогенные переменные X_1 и X_2 используются как инструментальные для переменных Y_1, Y_2. Этот результат, полученный нами в § **9.2**, верен и в общем случае:

*Если при оценке **идентифицируемого** уравнения в качестве инструментальных переменных используются экзогенные переменные, то получаемые при этом оценки совпадают с оценками косвенного метода наименьших квадратов.*

Из этого следует, что *косвенный метод наименьших квадратов является частным случаем метода инструментальных переменных.* На практике метод инструментальных переменных применяется в форме двухшагового метода наименьших квадратов, подробно описанного в главе **8**. А именно, в качестве инструментальных переменных используются объясненные (прогнозные) значения $\widetilde{y}_1, \widetilde{y}_2$ переменных Y_1, Y_2, полученные при оценивании приведенной формы. Затем эти значения подставляются в правую часть структурной формы (**9.5**).

Если система идентифицируема, и количество экзогенных переменных X совпадает с количеством эндогенных переменных Y, оценки двухшагового метода совпадают с оценками косвенного метода наименьших квадратов.

Процедура двухшагового метода наименьших квадратов реализована в большинстве компьютерных пакетов. Так, при исследовании модели из § **9.2** при применении этого метода получились бы уравнения:

$$\hat{y}_1 = 3114{,}286 + 20{,}43x_1 - 1{,}198\tilde{y}_2;\ d = 1{,}96,\ R^2 = 0{,}778,$$
$$(462{,}53)\quad (57{,}53)\quad (0{,}14)$$
$$\hat{y}_2 = 2519{,}134 + 8{,}73x_2 - 0{,}8\tilde{y}_2;\ d = 1{,}96,\ R^2 = 0{,}769.$$
$$(326{,}42)\quad (25{,}2)\quad (0{,}08)$$

Как и следовало ожидать, полученные оценки совпадают с оценками (**9.11**), полученными косвенным методом наименьших квадратов.

2. Система неидентифицируема.

В этом случае метод инструментальных переменных, вообще говоря, тоже применим, однако для его использования необходимо располагать «внешними» инструментальными переменными — экзогенных переменных не хватает. (Очевидно, это не что иное, как другая интерпретация неидентифицируемости модели.)

Предположим, имеется избыток инструментальных переменных в количестве l, и имеется возможность использовать их различные наборы. В этом случае двухшаговый метод наименьших квадратов предоставляет оптимальный выбор. Пусть $\{Z\}$ — набор инструментальных переменных (как «внутренних», экзогенных, так и «внешних»). Пусть $\hat{Y}_i$ — проекции эндогенных переменных на пространство Z (для их получения следует осуществить регрессию

$$\hat{Y}_i = a_i + \sum_{j=1}^{l} b_{ij} Z_j$$

обычным методом наименьших квадратов). Очевидно, переменные $\hat{Y}_i$ представляют собой линейные комбинации инструментальных переменных, наиболее тесно коррелирующих с переменными Y_i.

Замену в структурной форме системы Y_i на $\hat{Y}_i$ иногда называют «*очищением*» эндогенной переменной. При этом удаляется та «часть» переменной, которая коррелирует с ошибками регрессии.

В качестве примера рассмотрим модель

$$Y_1 = \alpha_1 + \beta X + \gamma_1 Y_2 + \varepsilon_1;$$
$$Y_2 = \alpha_2 + \gamma_2 Y_1 + \varepsilon_2$$

с инструментальными переменными Z_1 и Z_2. На рис. 9.2 представлены графические изображения соответствующих временных рядов.

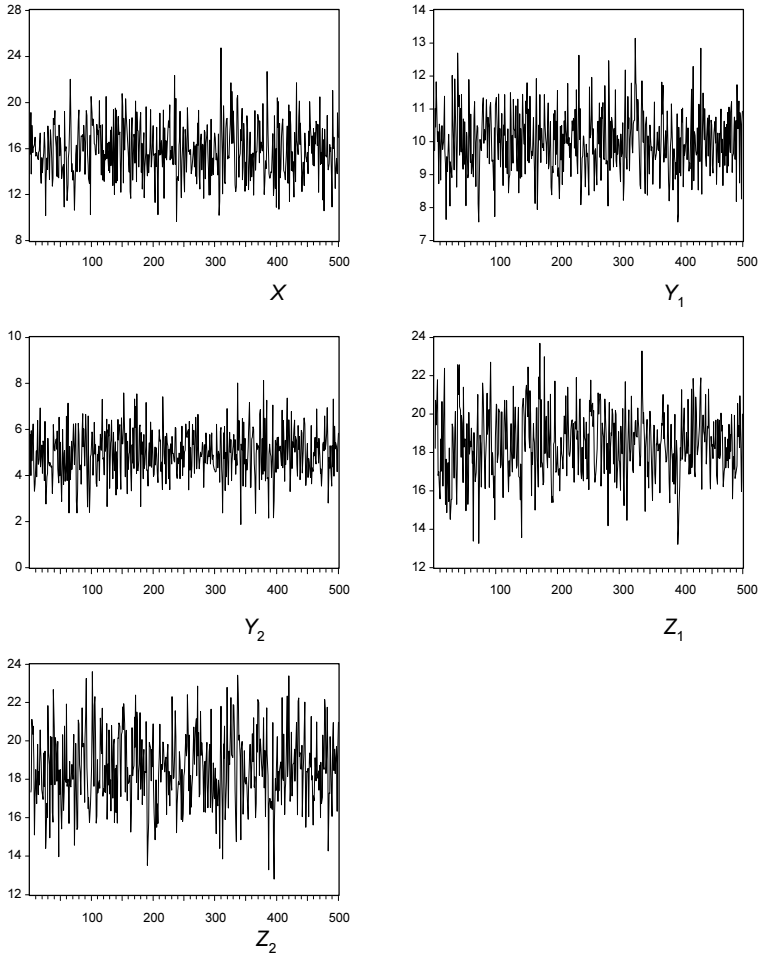


Рис. 9.2

Применяя к модели обычный метод наименьших квадратов, получаем оценки

$$\hat{\alpha}_1 = 4{,}837; \quad \hat{\beta} = 0{,}263; \quad \hat{\gamma}_1 = 0{,}206; \quad \hat{\alpha}_2 = 1{,}116; \quad \hat{\gamma}_2 = 0{,}385, \qquad \textbf{(9.17)}$$

которые, как известно, несостоятельны, и, следовательно, их значения могут существенно отличаться от истинных значений параметров. Применим теперь метод инструментальных переменных, выбрав в качестве инструментальных переменные X, Z_1, Z_2. Полученные при этом оценки имеют вид:

$$\hat{\alpha}_1 = 3,65; \quad \hat{\beta} = 0,239; \quad \hat{\gamma}_1 = 0,518; \qquad \hat{\alpha}_2 = -1,795; \qquad \hat{\gamma}_2 = 0,675 \qquad \textbf{(9.17′)}$$

и, как видно, значительно отличаются от оценок (**9.17**). Наилучшие оценки можно получить с помощью двухшагового метода наименьших квадратов. Они имеют вид:

$$\hat{\alpha}_1 = 3,632; \quad \hat{\beta} = 0,241; \quad \hat{\gamma}_1 = 0,522; \qquad \hat{\alpha}_2 = -1,783; \qquad \hat{\gamma}_2 = 0,675.$$

9.5. Одновременное оценивание регрессионных уравнений. Внешне не связанные уравнения

Косвенный метод наименьших квадратов по сути сводится к оцениванию по отдельности уравнений приведенной формы.

$$Y_1 = a_1 + b_1 X_1 + \nu_1, \qquad \textbf{(9.18)}$$

$$Y_2 = a_2 + b_2 X_2 + \nu_2. \qquad \textbf{(9.19)}$$

При этом, вообще говоря, $\mathrm{Cov}(\nu_1,\nu_2) \neq 0$. Отсюда следует, что эффективность оценивания можно повысить, если объединить уравнения (**9.18**), (**9.19**) в одно и применить к нему обобщенный метод наименьших квадратов.

Пусть

$$X = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}; \; Y = \begin{pmatrix} Y_1 \\ \dots \\ Y_2 \end{pmatrix}; \; \beta = \begin{pmatrix} \beta_1 \\ \dots \\ \beta_2 \end{pmatrix}; \; \nu = \begin{pmatrix} \nu_1 \\ \dots \\ \nu_2 \end{pmatrix}.$$

Тогда уравнения (**9.18**)—(**9.19**) можно записать в виде:

$$Y = X\beta + \nu. \qquad \textbf{(9.20)}$$

Пусть

$$\Sigma_{11} = \mathrm{Cov}(\nu_1, \nu_1), \quad \Sigma_{12} = \mathrm{Cov}(\nu_1, \nu_2), \; \Sigma_{22} = \mathrm{Cov}(\nu_2, \nu_2). \qquad \textbf{(9.21)}$$

Если уравнения (**9.18**), (**9.19**) по отдельности удовлетворяют условиям классической модели, матрицы Σ_{ij} — скалярные.

Тогда

$$\Sigma = \left[\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{12} & \Sigma_{22} \end{array}\right]$$

есть ковариационная матрица ошибок регрессии уравнения (**9.20**). Соответственно, оценка обобщенного метода наименьших квадратов уравнения (**9.20**) имеет вид (**7.7**):

$$b^* = (X' \Sigma^{-1} X)^{-1} X' \Sigma Y.$$

Для практического применения обобщенного метода наименьших квадратов следует оценить матрицу Σ. Это можно сделать, применив метод наименьших квадратов сначала к уравнениям (**9.18**), (**9.19**) по отдельности, найти остатки регрессии и принять в качестве оценок матриц Σ_{ij} выборочные ковариации $\hat{\mathrm{C}}\mathrm{ov}(e_i, e_j)$. Очевидно, эти оценки будут состоятельными.

Применяя метод одновременного оценивания, можно повысить эффективность косвенного метода наименьших квадратов. Заметим, однако, что *если наборы экзогенных переменных в обоих уравнениях совпадают, то оценка одновременного оценивания совпадает с оценкой метода наименьших квадратов, примененного к уравнениям по отдельности.* Так, для рассмотренного в § **9.4** примера одновременное оценивание не улучшит качество косвенного метода наименьших квадратов (или, что в данном случае то же самое, двухшагового метода наименьших квадратов).

Процедура одновременного оценивания регрессионных уравнений системы как внешне не связанных реализована в стандартных компьютерных пакетах. В западных эконометрических пакетах соответствующий метод оценивания называется *Seemingly Unreleased Regression* (*SUR*) (внешне не связанные уравнения).

Рассмотрим пример из гл. **8**. Там была рассмотрена модель вида

$$Y = \alpha + \gamma X + \varepsilon, \qquad \textbf{(9.22)}$$

где X — стоимость полуфабриката; Y — цена конечной продукции.

Применяя метод инструментальных переменных, получили следующее уравнение:

$$\hat{y} = 16{,}72 + 1{,}408x. \qquad \textbf{(9.23)}$$

В этой главе мы усложним модель, составив систему регрессионных уравнений. Будем считать, что стоимость полуфабриката X зависит от суммы цен на сырье, т.е. от величины $W = Z_1 + Z_2$ (предполагается, что оба вида сырья расходуются в равной пропорции — очевидно, это не есть ограничение, а лишь вопрос выбора единиц измерения). Пусть также Z — обобщенный фактор производства конечного продукта. Следующая диаграмма показывает выборочное распределение признака Z.

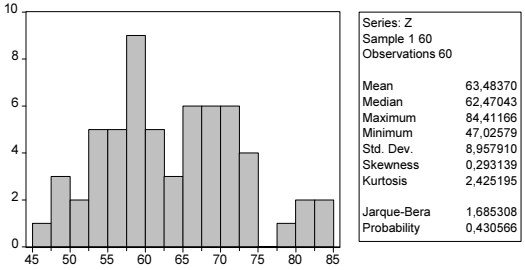


Рис. 9.3

Рассмотрим модель вида

$$X = \alpha_1 + \beta_1 W + \varepsilon_1,$$
$$Y = \alpha_2 + \beta_2 Z + \gamma X + \varepsilon_2. \qquad \textbf{(9.24)}$$

При этом ε_1, ε_2 коррелируют (на них действуют общие факторы, связанные со стоимостью перевозок), так что X — эндогенная переменная. Приведенная форма системы (**9.24**) имеет вид:

$$X = \alpha_1 + \beta_1 W + \varepsilon_1,$$
$$Y = (\alpha_2 + \gamma\alpha_1) + \beta_2 Z + \gamma\beta_1 W + (\gamma\varepsilon_1 + \varepsilon_2). \qquad \textbf{(9.25)}$$

Косвенный метод наименьших квадратов (уравнения (**9.25**) оцениваются по отдельности) дает следующие значения оценок:

$$\hat{\alpha}_1 = 19{,}31; \quad \hat{\beta}_1 = 1{,}77;$$
$$\hat{\alpha}_2 = 18{,}0; \quad \hat{\beta}_2 = 0{,}55; \qquad \textbf{(9.26)}$$
$$\hat{\gamma} = 1{,}325.$$

Теперь оценим уравнения (**9.25**) одновременно как внешне не связанные. Результатом оказываются следующие уравнения:

$$\hat{x} = \underset{(6{,}98)}{19{,}31} + \underset{(0{,}03)}{1{,}77W}, \; d = 1{,}9, \; R^2 = 0{,}984;$$
$$\hat{y} = \underset{(13{,}01)}{44} + \underset{(0{,}1)}{0{,}077Z} + \underset{(0{,}06)}{2{,}47W}, \; d = 2{,}0, \; R^2 = 0{,}97.$$

Отсюда получаем следующие значения оценок:

$$\hat{\alpha}_1 = 19{,}31; \quad \hat{\beta}_1 = 1{,}77;$$
$$\hat{\alpha}_2 = 17{,}0; \quad \hat{\beta}_2 = 0{,}08; \qquad \textbf{(9.27)}$$
$$\hat{\gamma} = 1{,}40.$$

Очевидно, мы должны считать оценки (**9.27**) более точными. Заметим при этом, что коэффициент $\hat{\beta}_2$ незначим.

9.6. Трехшаговый метод наименьших квадратов

Наиболее эффективная процедура оценивания систем регрессионных уравнений сочетает метод одновременного оценивания и метод инструментальных переменных. Соответствующий метод называется *трехшаговым методом наименьших квадратов*. Он заключается в том, что на первом шаге к исходной модели (**9.2**) применяется *обобщенный* метод наименьших квадратов с целью устранения корреляции случайных членов. Затем к полученным уравнениям применяется *двухшаговый* метод наименьших квадратов.

Очевидно, что если случайные члены (**9.2**) не коррелируют, трехшаговый метод сводится к двухшаговому, в то же время, если матрица B — единичная, трехшаговый метод представляет собой процедуру одновременного оценивания уравнений как внешне не связанных.

Применим трехшаговый метод к рассматриваемой модели (**9.24**):

$$\underset{(6{,}98)}{\alpha_1=19{,}31}; \quad \underset{(0{,}03)}{\beta_1=1{,}77}; \quad \underset{(4{,}82)}{\alpha_2=19{,}98}; \quad \underset{(0{,}08)}{\beta_2=0{,}05}; \quad \underset{(0{,}016)}{\gamma=1{,}4}.$$

Так как коэффициент β_2 незначим, то уравнение зависимости Y от X имеет вид:

$$\hat{y} = 16{,}98 + 1{,}4x.$$

Заметим, что оно практически совпадает с уравнением (**9.23**).

Как известно, очищение уравнения от корреляции случайных членов — процесс *итеративный*. В соответствии с этим при использовании трехшагового метода компьютерная программа запрашивает число итераций или требуемую точность. Отметим важное свойство трехшагового метода, обеспечивающего его наибольшую эффективность.

При достаточно большом числе итераций оценки трехшагового метода наименьших квадратов совпадают с оценками максимального правдоподобия.

Как известно, оценки максимального правдоподобия на больших выборках являются наилучшими.

9.7. Экономически значимые примеры систем одновременных уравнений

Здесь мы рассмотрим классические примеры систем уравнений, которые подробно изучаются в стандартных экономических курсах.

1. *Кейнсианская модель формирования доходов*:

$$C_t = \alpha + \beta Y_t + \varepsilon_t, \qquad \textbf{(9.28)}$$

$$Y_t = C_t + I_t, \qquad \textbf{(9.29)}$$

где Y, C, I соответственно представляют собой совокупный выпуск, объем потреблений и инвестиций. Здесь I рассматривается как экзогенная переменная, а Y — как эндогенная. Хорошо известно, что такая модель описывает закрытую экономику без государственного вмешательства.

Модель содержит одно поведенческое уравнение (**9.28**) и одно тождество (**9.29**).

Очевидно, модель (**9.28**)—(**9.29**) является идентифицируемой. Ее приведенная форма имеет вид:

$$Y = \frac{\alpha}{1-\beta} + \frac{1}{1-\beta} I + \frac{\varepsilon}{1-\beta}.$$

2. *Модель формирования спроса и предложения.*

В простейшем виде эта модель рассматривалась в § **9.1.** Здесь мы рассмотрим некоторые ее модификации.

Учет тренда. Если предположить, что привычки медленно меняются со временем, то в уравнение формирования спроса следует добавить временной тренд. Тогда модель будет иметь вид:

$$Q^d = \beta_1 + \beta_2 P + \beta_3 I + \rho t + \varepsilon_1; \qquad \textbf{(9.30)}$$

$$Q^s = \beta_4 + \beta_5 P + \varepsilon_2. \qquad \textbf{(9.31)}$$

Приведенная форма записывается в виде:

$$\begin{aligned} P &= \frac{\beta_1 - \beta_4}{\beta_5 - \beta_2} + \frac{\beta_3}{\beta_5 - \beta_2} I + \frac{\rho}{\beta_5 - \beta_2} t + \frac{\varepsilon_1 - \varepsilon_2}{\beta_5 - \beta_2} \\ Q &= \frac{\beta_1\beta_5 - \beta_2\beta_4}{\beta_5 - \beta_2} + \frac{\beta_3\beta_5}{\beta_5 - \beta_2} I + \frac{\rho\beta_5}{\beta_5 - \beta_2} t + \frac{\beta_5\varepsilon_1 - \beta_2\varepsilon_2}{\beta_5 - \beta_2}, \end{aligned} \qquad \textbf{(9.32)}$$

откуда следует, что система не является идентифицируемой. В то же время параметр β_5 оказывается сверхидентифицируемым. В самом деле, записав уравнения регрессии в виде

$$\hat{P} = a + bI + ct,$$
$$\hat{Q} = d + eI + ft,$$

легко заметить, что $\frac{e}{b}$ и $\frac{f}{c}$ дают оценку β_5.

У ч е т н а л о г а. Предположим теперь, что продавцы товара облагаются специальным налогом T. Величина налога меняется со временем и в выборке представлена временным рядом, т. е. является экзогенной переменной. Тогда уравнение спроса не меняется (спрос определяется лишь одной эндогенной переменной — рыночной ценой товара), а в уравнение предложения добавляется соответствующий член. Тогда модель примет вид:

$$Q^d = \beta_1 + \beta_2 P + \beta_3 I + \varepsilon_1;$$

$$Q^s = \beta_4 + \beta_5 P + \rho T + \varepsilon_2.$$

Очевидно, в этом случае модель будет идентифицируемой.

Предположим теперь, что доход I считается постоянным на протяжении длительного времени. Тогда в уравнении спроса следует исключить переменную I, и получатся уравнения:

$$Q^d = \beta_1 + \beta_2 P + \varepsilon_1; \qquad \textbf{(9.33)}$$

$$Q^s = \beta_4 + \beta_5 P + \rho T + \varepsilon_2. \qquad \textbf{(9.34)}$$

Система (**9.33**)−(**9.34**), очевидно, не является идентифицируемой. К ней может быть применен метод инструментальных переменных. При этом одна экзогенная переменная T, рассматриваемая как инструментальная, позволяет, вообще говоря, идентифицировать только уравнение (**9.33**), в которое она не

входит. Для идентификации (**9.34**) требуется «внешняя» инструментальная переменная.

Другим способом получить идентифицируемое уравнение формирования предложения оказывается ограничение на структурные коэффициенты: $\beta_5=-\rho$. Смысл этого ограничения очевиден: мы считаем, что продавцы исходят из суммы, которую они получают после уплаты налога, т.е. $P^*=P-T$. Тогда система может быть переписана в виде:

$$Q^d = \beta_1 + \beta_2 P + \varepsilon_1;$$
$$Q^s = \beta_4 + \beta_5 P^* + \varepsilon_2,$$

и экзогенная переменная T может быть использована как инструментальная для идентификации обоих уравнений.

Упражнения

9.1. Рассматривается система уравнений вида

$$\begin{cases} Y_1 = \beta X + \gamma Y_2 + \varepsilon_1; \\ Y_2 = \delta Y_1 + \varepsilon_2. \end{cases}$$

Проверить, является ли данная система идентифицируемой. Изменится ли ответ, если в число регрессоров второго уравнения включить: *а*) константу; *б*) переменную X ?

9.2. К системе двух уравнений вида

$$\begin{cases} Y_1 = \beta_1 X_1 + \gamma_1 Y_2 + \varepsilon_1; \\ Y_2 = \beta_2 X_2 + \gamma_2 Y_1 + \varepsilon_2 \end{cases} \qquad \textbf{(9.35)}$$

применен косвенный метод наименьших квадратов. Для коэффициентов приведенной формы

$$Y_1 = c_1 X_1 + c_2 X_2 + \nu_1;$$
$$Y_2 = c_3 X_1 + c_4 X_2 + \nu_2$$

получены следующие оценки c_1=2,2, c_2=0,4, c_3=0,08, c_4=−0,5.

Найти оценки двухшагового метода наименьших квадратов, примененного к системе (**9.35**).

9.3. При оценивании системы (**9.35**) двухшаговым и трехшаговым методом наименьших квадратов получены одинаковые оценки. Будут ли оценки, полученные обычным методом наименьших квадратов, состоятельными?

Глава 10
Проблемы спецификации модели

К проблемам *спецификации* традиционно относят д в а т и п а задач. *Первый* — это выбор структуры уравнения модели. *Второй* — это определение набора объясняющих переменных. В настоящей главе мы остановимся на проблемах второго типа, оставаясь в рамках линейных моделей.

Формально с проблемами спецификации приходится сталкиваться постоянно при анализе модели, например, при тестировании гипотез о значимости тех или иных регрессоров. Однако, как мы увидим здесь, принятие или отвержение гипотезы само по себе не тождественно принятию решения, какую именно модель использовать. В частности, мы увидим, что для максимально эффективного оценивания параметров при наиболее важных регрессорах вопрос о включении или невключении остальных регрессоров решается с помощью д р у г о г о критерия, нежели простая проверка гипотезы об их незначимости.

10.1. Выбор одной из двух классических моделей. Теоретические аспекты

В этом параграфе мы рассмотрим проблемы спецификации классической модели, удовлетворяющей предпосылкам *1—6* (см. § **4.2**). Соответствующая задача может быть сформулирована следующим образом. Пусть регрессоры разделены на две группы — X и Z, причем регрессоры X являются важными, и параметры при них требуется оценить с максимально возможной точностью, между тем, как регрессоры Z представляют значительно меньший интерес, и оценки соответствующих им параметров сами по себе для нас не важны. Следует выбрать одну из моделей:

$$Y = X\beta + Z\gamma + \varepsilon; \qquad \textbf{(10.1)}$$

$$Y = X\beta + \varepsilon. \qquad \textbf{(10.2)}$$

Модель (**10.1**) при решении задачи спецификации называют «длинной», а модель (**10.2**) — «короткой». При этом истинное значение коэффициента β в обеих моделях одно и то же. Именно этот параметр нам и требуется оценить.

Для того, чтобы сделать выбор между моделями (**10.1**) и (**10.2**), прежде всего надо определить критерии предпочтения. Таких критериев может быть два.

1. В случае равенства коэффициентов γ нулю истинное значение параметра β одно и то же, но значения оценок b, полученных с помощью метода наименьших квадратов из моделей (**10.1**), (**10.2**) будут различными. Следует предпочесть ту оценку, которая «ближе» к истинному значению. Такой характеристикой «близости» является средний квадрат отклонения

$$\mu(b) = M[(b - \beta)(b - \beta)'].$$

Заметим, что μ представляет собой квадратную матрицу порядка p, где p — число регрессоров X. (Напомним (см. § **12.8**), что квадратная матрица A называется «большей», чем квадратная матрица B, если их разность $A-B$ есть положительно определенная матрица.)

Таким образом, из двух моделей (**10.1**), (**10.2**) следует выбрать ту, для которой получается оценка β с меньшей величиной μ.

2. Пусть $\hat{y}_{n+1}$ — прогнозное значение, которое получается из модели для еще не наблюдаемых значений регрессоров. Величина $M(\hat{y}_{n+1} - y_{n+1})^2$ может рассматриваться как средняя ошибка прогноза. Следует выбрать ту модель, для которой эта ошибка меньше.

Можно показать, что критерии 1 и 2 (т.е. выбор модели по минимуму $\mu(b)$ или $M(\hat{y}_{n+1} - y_{n+1})^2$) равносильны. Их условия записываются с помощью одного и того же соотношения, полученного Я. Магнусом в работе [18]. В настоящей главе мы приведем соответствующие результаты в несколько упрощенной форме.

Пусть b_1, g — оценки параметров β и γ, полученные с помощью метода наименьших квадратов из модели (**10.1**); β_2 — оценка β из модели (**10.2**). Применение обычных формул (**4.8**) дает следующий вид этих оценок:

$$b_1 = (X'M_Z X)^{-1} X'M_Z Y; \qquad (\mathbf{10.3})$$

$$g = (Z'M_X Z)^{-1} Z'M_X Y; \qquad (\mathbf{10.4})$$

$$b_2 = (X'X)^{-1} X'Y, \qquad (\mathbf{10.5})$$

где M_X, M_Z — матрицы вида:

$$M_X = E - X(X'X)^{-1}X',\ M_Z = E - Z(Z'Z)^{-1}Z' \qquad \textbf{(10.6)}$$

(E — единичная матрица). Отметим, что матрицы M_X, M_Z являются *идемпотентными* (см. § **12.8**).

Имеет место соотношение

$$(X'M_ZX)^{-1}X'X - (X'X)^{-1}X'Z'\ (Z'M_XZ)^{-1}Z'X = E,$$

в справедливости которого легко убедиться непосредственно. Используя это соотношение, равенство (**10.3**) можно переписать в виде:

$$b_1 = (X'X)^{-1}X'Y - (X'X)^{-1}Z'(Z'M_XZ)^{-1}Z'M_XY,$$

т. е. (см. (**10.4**)):

$$b_1 = b_2 - (X'X)^{-1}X'Zg. \qquad \textbf{(10.7)}$$

Обозначим $(X'X)^{-1}(X'Z) = b_{zx}$. (Заметим, что вид матрицы b_{zx} такой же, как у оценки параметра эконометрической модели. При этом, однако, формально b_{zx} нельзя назвать оценкой, так как величины X, Z — н е с л у ч а й н ы е.)

Перепишем равенство (**10.7**) в виде:

$$b_2 = b_1 + b_{zx}g. \qquad \textbf{(10.8)}$$

Приведем теперь (опуская промежуточные вычисления) формулы для вычисления ковариационных матриц оценок b_1, b_2, g:

$$\begin{aligned} \mathrm{Cov}(b_1) &= \sigma^2[(X'X)^{-1} + (X'X)^{-1}X'Z(Z'M_XZ)^{-1}Z'X(X'X)^{-1}]; \\ \mathrm{Cov}(g) &= \sigma^2(Z'M_XZ)^{-1}; \qquad \textbf{(10.9)} \\ \mathrm{Cov}(b_2) &= \sigma^2(X'X)^{-1}, \end{aligned}$$

где σ^2 — дисперсия ε_i.

Используя формулы (**10.9**), получаем:

$$\mathrm{Cov}(b_1) = \mathrm{Cov}(b_2) + b_{zx}(\mathrm{Cov}(g)b'_{zx}. \qquad \textbf{(10.10)}$$

Проанализируем полученные результаты (**10.8**) и (**10.10**). В силу выполнения условий классической модели оценка b_1 в любом случае является несмещенной (и при $\gamma = 0$, и при $\gamma \neq 0$). Между тем равенство (**10.8**) показывает, что если истинное значение γ не равно нулю, оценка b_2 является смещенной со смещением $b_{zx}\gamma$. При этом явный вид смещения указывает на то, что оно будет положительным, если Z «одинаково направлены»

по отношению к X и Z, и отрицательно в противоположном случае.

В то же время равенство (**10.10**) означает, что оценка b_1 в любом случае имеет бóльшую ковариацию, чем оценка b_2.

Из формул (**10.8**)–(**10.10**) получаем:

$$\mu(b_1) - \mu(b_2) = b_{zx}\,\Theta\,b'_{zx}, \qquad \textbf{(10.11)}$$

где

$$\Theta = \mathrm{Cov}(g) - \gamma\gamma'. \qquad \textbf{(10.12)}$$

Таким образом, мы можем сделать вывод:

Если матрица (**10.12**) *является положительно определенной, т. е. имеет только положительные собственные значения, то модель* (**10.2**) *лучше оценивает параметр* β, *даже если на самом деле верна модель* (**10.1**).

Если матрица (**10.12**) имеет как положительные, так и отрицательные собственные значения, то вопрос не решается столь однозначно. Возможно, в этом случае имеет смысл предпочесть короткую модель (**10.2**), если след матрицы (**10.12**) положителен.

Обратимся теперь к критерию минимальности ошибки прогноза. Пусть имеется наблюдение $\tilde{x}_{n+1}$, т. е. (x_{n+1}, z_{n+1}) в случае модели (**10.1**) и x_{n+1} в случае модели (**10.2**). Прогнозные значения в этих случаях равны:

$$\hat{y}_1 = x_{n+1}b_1 + z_{n+1}\,g; \qquad \textbf{(10.13)}$$

$$\hat{y}_2 = x_{n+1}\,b_2. \qquad \textbf{(10.14)}$$

Заметим, что случайная величина (**10.13**) имеет то же математическое ожидание, что и наблюдаемое значение случайной величины y_{n+1}. Непосредственные вычисления дают следующие значения для средних ошибок прогноза:

$$M(\hat{y}_1 - y_{n+1})^2 = \sigma^2 + \sigma^2\,x_{n+1}(X'X)^{-1}x'_{n+1} + \nu\mathrm{Cov}(g)\nu',$$

$$M(\hat{y}_2 - y_{n+1})^2 = \sigma^2 + \sigma^2\,x_{n+1}(X'X)^{-1}x'_{n+1} + \nu\,\gamma\gamma'\nu',$$

где $\nu = x_{n+1}b_{zx} - z_{n+1}$.

Следовательно,

$$M(\hat{y}_1 - y_{n+1})^2 - M(\hat{y}_2 - y_{n+1}) = \nu\,\Theta\,\nu'. \qquad \textbf{(10.15)}$$

где Θ — матрица (**10.12**).

Таким образом, и здесь положительная определенность матрицы (**10.12**) означает бóльшую предпочтительность короткой модели (**10.2**) — даже, если истинное значение параметра γ не

равно нулю. Если же матрица (**10.12**) имеет как положительные, так и отрицательные собственные значения, то можно выбрать «короткую» модель (**10.2**), если положителен след матрицы (**10.12**).

Можно показать, что матрица вида (**10.12**) является положительно определенной в том и только том случае, если выполняется условие:

$$\theta = \gamma' \, (\mathrm{Cov}\,(g))^{-1} \gamma \; < 1. \qquad \textbf{(10.16)}$$

Таким образом, *модель* (**10.2**) оказывается с математической точки зрения *предпочтительней модели* (**10.1**), *если выполняется условие* (**10.16**).

10.2. Выбор одной из двух классических моделей. Практические аспекты

Величина θ, стоящая в левой части равенства (**10.16**), зависит от неизвестного параметра γ, т. е. является н е н а б л ю д а е м о й, так что полученный в § **10.1** критерий еще не дает ответа на вопрос, как осуществлять альтернативный выбор между моделями (**10.1**) и (**10.2**) на практике.

В реальности же мы располагаем лишь значением оценки:

$$\hat{\theta} = g'(\hat{\mathrm{Cov}}(g)^{-1} g\,.$$

Преобразуем величину $\hat{\theta}$. Для этого представим оценки g и $\hat{\mathrm{Cov}}(g)$ в удобной для нас форме. Имеем:

$$g = \left(Z' M_X Z\right)^{-1} Z' M_X Y = (Z' M_X Z)^{-1} Z' M_X (X\beta + Z\gamma + \varepsilon)\,.$$

Непосредственно перемножая матрицы, легко убедиться, что имеет место равенство $M_X X$=0. Таким образом, получаем:

$$g = \gamma + (Z' M_X Z)^{-1} Z' M_X \varepsilon\,. \qquad \textbf{(10.17)}$$

В то же время

$$\hat{\mathrm{Cov}}(g) = \frac{\hat{\sigma}^2}{\sigma^2}\mathrm{Cov}(g) = \hat{\sigma}^2 (Z' M_X Z)^{-1},$$

где $\hat{\sigma}^2 = \dfrac{e_1' e_1}{n - p - 1}$ — оценка параметра σ^2 (см. (**4.21**)). Здесь e_1 — столбец остатков регрессии (**10.1**). Таким образом, используя равенство (**10.17**), получаем:

$$\hat{\theta} = \frac{1}{\hat{\sigma}^2}[\gamma' + \varepsilon' M_X Z (Z' M_X Z)^{-1}](Z' M_X Z)[\gamma + (Z' M_X Z)^{-1} Z' M_X \varepsilon] =$$

$$= \frac{1}{\hat{\sigma}^2}\gamma' Z' M_X Z \gamma + \frac{1}{\hat{\sigma}^2}\left(\gamma' Z' M_X \varepsilon + \varepsilon' M_X Z \gamma\right) + \frac{1}{\hat{\sigma}^2}\varepsilon' M_X Z (Z' M_X Z)^{-1} Z' M_X \varepsilon,$$

или
$$\hat{\theta} = \frac{\sigma^2}{\hat{\sigma}}\theta + \xi + \frac{1}{\hat{\sigma}^2}\varepsilon' B \varepsilon, \qquad \textbf{(10.18)}$$

где
$$\xi = \frac{1}{\hat{\sigma}^2}\left(\gamma' Z' M_X \varepsilon + \varepsilon' M_X Z \gamma\right);$$

$$B = M_X Z (Z' M_X Z)^{-1} Z' M_X.$$

Можно показать, что величина ξ принимает с равной вероятностью как положительные, так и отрицательные значения, в то время, как величина $\frac{1}{\hat{\sigma}^2}\varepsilon' B \varepsilon$ принимает лишь положительные значения. Если число наблюдений n достаточно велико, значения σ^2 и оценки $\hat{\sigma}^2$, как правило, близки. Таким образом, используя равенство (**10.18**), мы можем считать малые значения наблюдаемой величины $\hat{\theta}$ достаточной предпосылкой малости и параметра θ. В частности:

Если $\hat{\theta} < 1$, *разумно предположить, что и* $\theta < 1$. **(10.19)**

Оказывается, величина $F = \hat{\theta}/l$ (где l — число регрессоров Z) есть не что иное, как наблюдаемое значение статистики при обычном тестировании гипотезы о равенстве нулю коэффициентов γ в модели (**10.1**). Если параметр γ на самом деле равен нулю, F имеет распределение Фишера—Снедекора).

В самом деле, явный вид этой статистики (см, например, [13])

$$F = \frac{\frac{1}{l}\left(e_2' e_2 - e_1' e_1\right)}{\hat{\sigma}^2}. \qquad \textbf{(10.20)}$$

В то же время, используя (**10.9**), получаем:

$e_2 = Y - Xb_2 = Xb_1 + Zg + e_1 - Xb_1 - Xb_{zx}\, g = e_1 - M_X Z g.$

Отсюда

$$e_2' e_2 - e_1' e_1 = \hat{\sigma}^2\, g' (\mathrm{Cov}(g))^{-1} g, \qquad \textbf{(10.21)}$$

и утверждение доказывается подстановкой (**10.21**) в (**10.20**).

Таким образом, может быть предложен следующий подход к альтернативному выбору модели: выбирается некоторое значение *c*. Если наблюдаемое значение *F*-статистики меньше *c*, то предпочитается модель (**10.2**), если больше *c* — то модель (**10.1**).

При этом пороговое значение *c* выбирается, вообще говоря, произвольно, как правило, в границах $\frac{1}{l} < c < F_{\alpha;l;n-p-l-1}$, где обычно $\alpha = 0{,}05$.

Рассмотрим более подробно случай $l = 1$, т. е. имеется один регрессор *Z,* относительно которого ставится вопрос о включении или невключении его в модель. В этом случае $F = \hat{\theta}$. Используя (**10.19**), получаем следующий критерий предпочтения:

В случае $l = 1$ (*наличие одного* «*спорного*» *регрессора*) *модель* (**10.2**) *оказывается предпочтительней, чем модель* (**10.1**), *если наблюдаемое значение F-статистики при тестировании гипотезы* $\gamma = 0$ *оказывается меньше* 1.

Получим еще одну формулировку приведенного критерия. Используем выражение (**4.34′**) для скорректированного коэффициента детерминации R^2.

Соответственно для регрессий (**10.1**), (**10.2**):

$$\hat{R}_1^2 = 1 - \frac{\frac{1}{n-p-1}e_1'e_1}{\frac{1}{n-1}y'y}, \quad \hat{R}_2^2 = 1 - \frac{\frac{1}{n-p}e_2'e_2}{\frac{1}{n-1}y'y},$$

где $y = Y - \overline{Y}$.

Отсюда

$$\hat{R}_1^2 - \hat{R}_2^2 = \frac{n-1}{y'y}\left[\frac{e_2'e_2}{n-p} - \frac{e_1'e_1}{n-p-1}\right] = \frac{(n-1)\,\sigma^2}{(n-p)\,y'y}(F-1). \quad \textbf{(10.22)}$$

Таким образом, *модель* (**10.2**) *оказывается предпочтительней модели* (**10.1**), *если скорректированный коэффициент детерминации при удалении регрессоров Z увеличивается* (заметим, что простой коэффициент детерминации модели (**10.1**) в с е г д а больше, чем модели (**10.2**)).

10.3. Спецификация модели пространственной выборки при наличии гетероскедастичности

В случае модели пространственной выборки показателем невключения в модель существенных переменных может служить *неустраняемая гетероскедастичность***.**

Вспомним, что наиболее часто употребляемые процедуры устранения гетероскедастичности так или иначе были основаны на предположении, что дисперсия ошибок регрессии σ^2 является функцией от каких-то регрессоров. Если σ^2 существенно зависит от регрессора Z, а при спецификации модели регрессор Z не был включен в модель, стандартные процедуры могут не привести к устранению гетероскедастичности.

Рассмотрим следующий пример. Пусть X, Z — регрессоры, Y — объясняемая величина, гистограммы которых и основные числовые характеристики распределения приведены на рис. **10.1**.

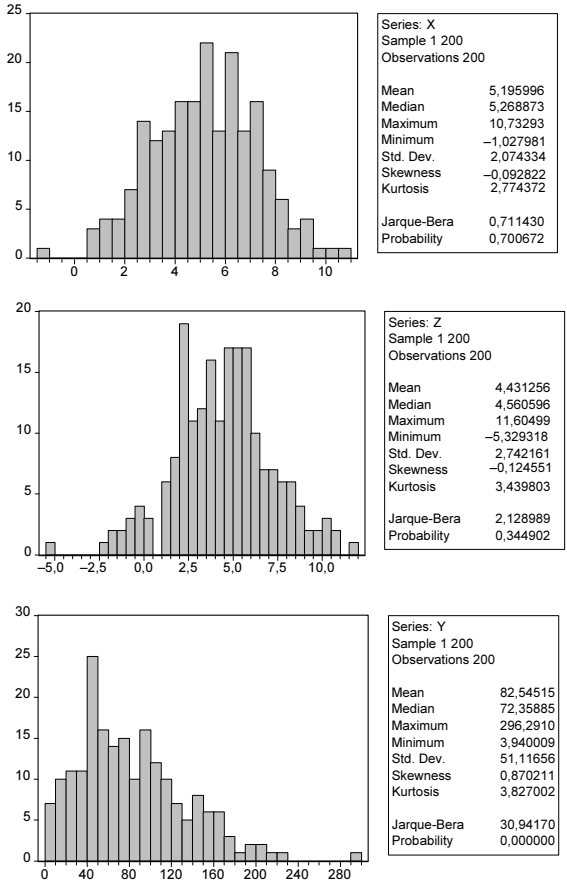


Рис. 10.1

Оценим линейную регрессию зависимости Y от X, т. е. рассмотрим модель

$$Y = \alpha + \beta X + \varepsilon. \qquad \textbf{(10.23)}$$

Применяя обычный метод наименьших квадратов, получим уравнение:

$$\hat{y} = 3{,}42 + 15{,}23x, \; R^2 = 0{,}38$$
$$(7{,}7) \quad (1{,}37)$$

(как видно, константа оказывается незначимой). Применение к полученному уравнению теста Уайта на гетероскедастичность дает следующий результат:

$$F = 9{,}62 > F_{0{,}05;\ 1;\ 99},$$

т. е. модель оказывается гетероскедастичной.

Применим к уравнению модели взвешенный метод наименьших квадратов. Тогда уравнение регрессии примет вид:

$$\hat{y} = 5{,}98 + 14{,}67x.$$
$$(4{,}92) \quad (1{,}11)$$

Применяя к нему тест Уайта, получаем:

$$F = 3{,}6 > F_{0{,}05;\ 1;\ 99},$$

т. е. гипотеза о гомоскедастичности вновь отвергается.

Попробуем исправить положение, включив в модель также и регрессор Z, т. е. оценим модель

$$Y = \alpha + \beta X + \gamma Z + \varepsilon.$$

Обычным методом наименьших квадратов получаем уравнение регрессии:

$$\hat{y} = 4{,}98 + 8{,}53x + 7{,}50z, \; R^2 = 0{,}47.$$
$$(7{,}15) \quad (1{,}73) \quad (1{,}31)$$

Эта модель также является гетероскедастичной, так как тест Уайта дает следующее значение F-статистики:

$$F = 10{,}74 > F_{0{,}05;\ 1;\ 99}.$$

Однако теперь, после применения взвешенного метода наименьших квадратов, мы получаем уравнение:

$$\hat{y} = 15{,}04 + 7{,}54x + 6{,}19z,$$

$$(3{,}64) \quad (1{,}08) \quad (1{,}03)$$

для которого гипотеза о гомоскедастичности уже принимается, так как использование теста Уайта дает значение F-статистики:

$$F = 1{,}17 < F_{0{,}05;\ 1;\ 99}.$$

10.4. Спецификация регрессионной модели временных рядов

В моделях временных рядов неверная спецификация может служить причиной автокорреляции ошибок регрессии.

Рассмотрим следующий пример. Пусть X — доход семьи, тыс. долл.; Y — ее расходы на отдых в зарубежных странах, тыс. долл. На рис. 10.2 представлены диаграммы распределения и основные количественные характеристики распределения величин X и Y.

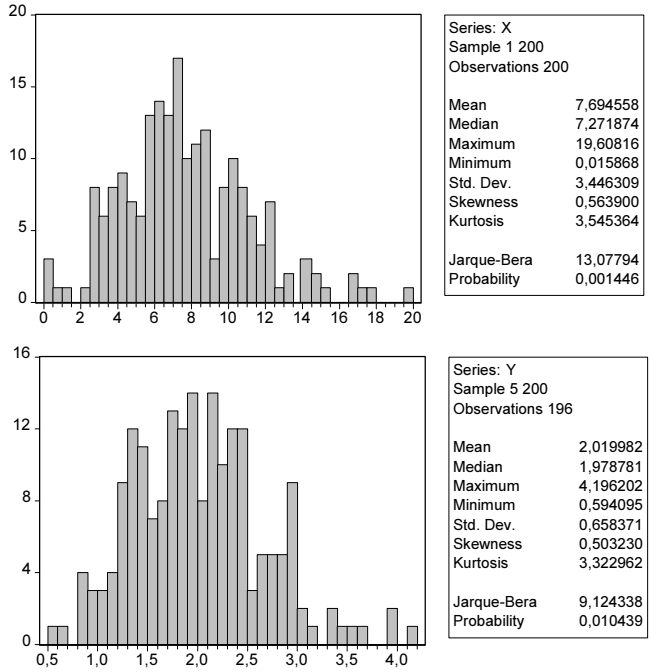


Рис. 10.2

При этом следует помнить, что X и Y — временные ряды, поэтому кроме диаграмм распределения важны и *коррелограммы*, приведенные в таблицах 10.1 и 10.2.

Таблица 10.1

Коррелограмма X

τ	$r(\tau)$	$r_{\text{част}}(\tau)$	Q_p	$P(Q>Q_p)$
1	0,750	0,750	114,34	0,000
2	0,547	−0,036	175,46	0,000
.....				
9	−0,097	−0,051	214,72	0,000
10	−0,126	0,068	218,11	0,000

Таблица 10.2

Коррелограмма Y

τ	$r(\tau)$	$r_{\text{част}}(\tau)$	Q_p	$P(Q>Q_p)$
1	0,813	0,813	131,55	0,000
2	0,629	−0,094	210,78	0,000
.....				
9	−0,191	−0,098	271,43	0,000
10	−0,227	0,056	282,20	0,000

Оценим модель вида

$$Y = \alpha + \beta X + \varepsilon \qquad \textbf{(10.24)}$$

зависимости расходов на отдых в зарубежье Y от доходов X. Применяя обычный метод наименьших квадратов, получим уравнение регрессии:

$$\hat{y} = 0,745 + 0,17x, \quad R^2 = 0,69, \; d = 1,28.$$
$$(0,066)\;(0,008)$$

Существенно отличающееся от двух значение статистики d Дарбина—Уотсона указывает на то, что имеется положительная автокорреляция ошибок регрессии. Одна из возможностей — попробовать идентифицировать ряд остатков как ряд модели *ARMA*(*p, q*). При этом самая простая модель *AR*(1) оказывается вполне адекватной:

$$\varepsilon_t = 0,35\varepsilon_{t-1}, \quad R^2 = 0,125, \; d = 2,03.$$

На этот раз значение статистики d Дарбина—Уотсона оказывается достаточно близким к двум. Таким образом, в качестве модели мы можем принять модель авторегрессии первого порядка

$$Y_t = \alpha + \beta X_t + \varepsilon_t, \qquad \varepsilon_t = \rho\varepsilon_{t-1} + \nu_t$$

с оценками значений параметров $\hat{\beta} = 0{,}17$, $\hat{\rho} = 0{,}35$.

Однако естественно предположить, что расходы на дорогостоящий товар — отдых на зарубежных курортах — зависят не только от текущих доходов, но и от доходов в предыдущие периоды. Изменим спецификацию модели, включив в качестве регрессоров лаговые переменные X.

Легко убедиться, что наиболее адекватной оказывается модель с четырьмя включенными лагами. Соответствующее уравнение регрессии имеет вид:

$$\hat{y}_t = \underset{}{0{,}32} + \underset{(0{,}044)}{0{,}08}\, x_t + \underset{(0{,}0066)}{0{,}07}\, x_{t-1} + \underset{(0{,}008)}{0{,}037}\, x_{t-2} + \underset{(0{,}008)}{0{,}017}\, x_{t-3} + \underset{(0{,}008)}{0{,}012}\, x_{t-4},$$

$$(0{,}006)$$

$$R^2 = 0{,}91,\ d = 2{,}09.$$

Как видно, значение статистики d Дарбина—Уотсона очень близко к двум, так что в новой модели проблема автокорреляции ошибок регрессии отсутствует. Отсюда следует, что ее причина была в неверной спецификации модели. Стоит также обратить внимание, что коэффициент регрессии при x_t уменьшился вдвое — на товары роскоши, подобные дорогому отдыху, расходы «рассредоточиваются» по нескольким ближайшим годам.

10.5. Важность экономического анализа

В любом случае при выборе спецификации модели следует в первую очередь руководствоваться экономическим анализом. Необходимо помнить, что выборочные данные — это всего лишь совокупность цифр, и, манипулируя ими, иногда можно получить чрезвычайно хорошую, с точки зрения математики, модель, лишенную, однако, какого-либо смысла.

Рассмотрим пример. X — производство бананов в Бразилии, Y — производство шин на Ярославском заводе. На рис. 10.3 изображены графики временных наблюдений этих величин.

Рассмотрим регрессионную модель зависимости Y от X и оценим ее обычным методом наименьших квадратов:

$$\hat{y} = -83{,}8 + 1{,}93x,\ R^2 = 0{,}97,\ \ d = 1{,}96.$$

$$(7{,}31)\ (0{,}035)$$

С математической точки зрения эта модель великолепна по всем параметрам! В то же время экономически очевидна бессмысленность такого результата. Объяснение здесь очень простое — в рассматриваемый период времени обе величины имели временной тренд, который и привел к высокому значению корреляции между X и Y.

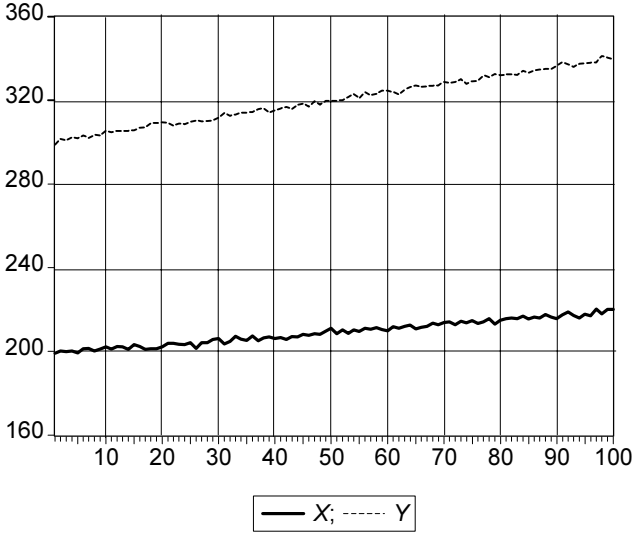


Рис. 10.3

Освободим рассматриваемые величины X и Y от влияния тренда. Для этого рассмотрим уравнения регрессии:

$$\hat{x}_t = 198{,}92 + 0{,}2t,\ \ R^2 = 0{,}97,\ d = 2{,}12$$

$$(0{,}186)\ (0{,}003)$$

$$\hat{y}_t = 300 + 0{,}4t,\ R^2 = 0{,}994,\ \ d = 1{,}885.$$

$$(0{,}175)\ (0{,}003)$$

Построим освобожденные от тренда величины:

$$X_{\text{detr}} = X - 0{,}2t,$$

$$Y_{\text{detr}} = Y - 0{,}4t.$$

Если теперь мы рассмотрим модель

$$Y_{\text{detr}} = \alpha + \beta X_{\text{detr}} + \varepsilon$$

и применим к ней обычный метод наименьших квадратов, то получим следующий результат:

$$\hat{y}_{\text{detr}} = 312,6 - 0,05x_{\text{detr}}, \quad R^2 = 0,004.$$
$$(0,09)$$

Теперь мы получаем экономически осмысленный результат — регрессия не значима.

Математически это означает, что частный коэффициент корреляции между величинами X и Y равен нулю, между тем как линейный коэффициент корреляции достаточно велик.

В рассмотренном примере мы отказались от вывода о значимости переменной X несмотря на то, что математические соображения (на первый взгляд!) свидетельствовали об обратном. Аналогично, если экономические соображения приводят к выводу о том, что зависимость от переменной существенна, возможно, ее следует включить в модель, даже если математические свойства модели ухудшаются, и поискать математически адекватную форму модели.

Таким образом, при спецификации модели должно быть найдено математическое решение в рамках, определенных экономическим анализом.

Упражнения

10.1. Имеются два регрессора X и Z и одна объясняемая переменная Y. Обычным методом наименьших квадратов получены следующие регрессионные уравнения:

$$\hat{y} = 1,2 + 11,4x - 0,2z;$$
$$(0,2)\ (0,8)\quad (0,12) \qquad \textbf{(10.25)}$$

$$\hat{y} = 1,1 + 11,9x;$$
$$(0,2)\ (0,7) \qquad \textbf{(10.26)}$$

$$\hat{z} = 2,7 + 0,8x.$$
$$(0,3)\ (0,1)$$

Выбрано уравнение (**10.26**). В какую сторону смещена оценка при регрессоре X в том случае, если на самом деле верна мо-

дель, включающая регрессор Z? (Считать, что условия классической модели выполнены.)

10.2. Рассматриваются две альтернативные модели, удовлетворяющие условиям классической модели:

$$Y = X\beta + Z\gamma + \varepsilon; \qquad \textbf{(10.27)}$$

$$Y = X\beta + \varepsilon. \qquad \textbf{(10.28)}$$

При этом суммы квадратов остатков равны 1010 для модели (**10.27**) и 1075 — для модели (**10.28**), а количество наблюдений равно 2000. Какую из двух моделей следует предпочесть для более точного прогнозирования значений переменной Y?

10.3. Пусть справедлива следующая модель, удовлетворяющая условиям классической модели:

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

где экспериментальные данные представляют собой временные ряды, а третий член в правой части задает тренд. Пусть при этом оценивается «неправильная», короткая модель

$$Y = \alpha + X\beta + \nu \qquad \textbf{(10.29)}$$

без тренда.

Будут ли выполнены следующие условия для модели (**10.29**):

а) $M(\nu_t) = 0$;

б) сумма остатков регрессии равна нулю?

Глава 11

Модели с различными типами выборочных данных

Стандартные модели эконометрики с непрерывными данными в виде пространственной выборки или временного ряда не всегда позволяют адекватно оценить представляющие интерес параметры. В ряде случаев получаются неэффективные, иногда смещенные и несостоятельные оценки. В некоторых случаях применение стандартной модели и вовсе оказывается невозможным. В настоящей главе мы рассмотрим другие типы выборочных данных и связанные с ними экономические модели.

Основная часть этой главы посвящена *панельным данным*, которые в некотором роде оказываются объединением пространственных данных и временных рядов. Также мы рассматрим ограничения на возможные значения зависимой переменной и отбор наблюдений в выборку.

11.1. Статистические модели с панельными данными

В экономических исследованиях нередко приходится рассматривать большие массивы данных в последовательные моменты времени. Это могут быть индивидуумы, домашние хозяйства, регионы, страны, у которых отслеживается динамика некоторых количественных показателей. Такие данные, сочетающие в себе пространственные выборки и временные ряды, называются *панельными данными*.

Пусть имеется n объектов, T временных наблюдений. Рассматривается зависимая переменная Y со значениями y_{ti}, где $t = 1, ..., T$, $i = 1, ..., n$; и k регрессоров X со значениями X_{j,t_i}, $j = 1, ..., k$, $t = 1, ..., T$, $i = 1, ..., n$.

Во многих случаях допустимо предположение, что наблюдения в различные моменты времени не коррелируют друг с другом. Соответствующие модели называются *статическими*.

Возможны две крайние точки зрения на такую структуру наблюдений: их можно рассматривать как единый временной ряд с «толстым сечением», в котором наблюдения представляют собой n значений объясняемой переменной. В этом случае модель принимает вид

$$y_{ti} = \alpha + \sum_{j=1}^{k} \beta_j x_{j,ti} + \varepsilon_{ti} \qquad \textbf{(11.1)}$$

и содержат $k+1$ параметр. По сути эта модель является классической, и ее следует оценивать обычным методом наименьших квадратов. В этой модели объекты панели лишены какой-либо индивидуальности и представляют собой наблюдения однородной выборки.

Противоположная точка зрения заключается в том, чтобы рассматривать данные как n различных временных рядов. То есть все объекты считаются абсолютно индивидуальными, и действия регрессоров на них различными. Тогда уравнения модели имеет вид

$$y_{ti} = \alpha_i + \sum_{j=1}^{k} \beta_{ji} x_{j,ti} + \varepsilon_{ti}$$

и содержат $kn+n$ параметров.

Модели с панельными данными позволяют сочетать оба этих подхода: объекты считаются индивидуальными, но изучается общее действие регрессоров на эти объекты. Один из возможных подходов заключается в следующем: регрессоры действуют на все объекты одинаково, но кроме рассматриваемых величин X_j, есть еще — как правило, ненаблюдаемые — факторы Q_a, которые, собственно, и обуславливают различия между объектами. Если считать факторы Q_a постоянными по времени, можно рассматривать уравнения

$$y_{ti} = \alpha + \sum_{j=1}^{k} \beta_j x_{j,ti} + \sum_{a=1}^{A} \gamma_a q_{ai} + \varepsilon_{ti}\,.$$

Обозначив $$\alpha + \sum_{a=1}^{A} \gamma_a q_{ai} = \alpha_i\,,$$

получим *модель с панельными данными* в виде

$$y_{ti} = \alpha_i + \sum_{j=1}^{k} \beta_j x_{j,ti} + \varepsilon_{ti}. \qquad \textbf{(11.2)}$$

11.2. Межгрупповые и внутригрупповые оценки модели с панельными данными

Панельные данные характеризуются двумя индексами: t (момент времени) и i (номер объекта).

Рассмотрим следующие усреднения:

$$\overline{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{ti};\ \ \overline{x}_{ji} = \frac{1}{T}\sum_{t=1}^{T} x_{j,ti}; \qquad \textbf{(11.3)}$$

$$\overline{y}_t = \frac{1}{n}\sum_{i=1}^{n} y_{ti};\ \ \overline{x}_{jt} = \frac{1}{n}\sum_{i=1}^{n} x_{j,ti}. \qquad \textbf{(11.4)}$$

Величины (**11.3**) называются *внутригрупповыми* средними (средними по времени для одного объекта), а (**11.4**) — *межгрупповыми* (средними по всем объектам в каждый момент времени).

Для средних (**11.3**), (**11.4**) уравнения принимают вид:

$$\overline{y}_i = \alpha_i + \sum_{j=1}^{k} \beta_j \overline{x}_{ji} + \overline{\varepsilon}_i, \qquad \textbf{(11.5)}$$

$$\overline{y}_t = \overline{\alpha} + \sum_{j=1}^{k} \beta_j \overline{x}_{jt} + \overline{\varepsilon}_t. \qquad \textbf{(11.6)}$$

Вычитая из уравнений (**11.2**) уравнения (**11.5**), получаем:

$$y_{ti} - \overline{y}_i = \sum_{j=1}^{k} \beta_j (x_{j,ti} - \overline{x}_{ji}) + (\varepsilon_{ti} - \overline{\varepsilon}_i). \qquad \textbf{(11.7)}$$

Уравнения (**11.6**), (**11.7**) имеют по сравнению с (**11.2**) то преимущество, что число оцениваемых параметров в них невелико — $k+1$ и k соответственно.

Оценки параметров β, получаемые применением к уравнениям (**11.6**), (**11.7**) обычного метода наименьших квадратов, называются соответственно *межгрупповыми* $\hat{\beta}_B$ и *внутригрупповыми* $\hat{\beta}_W$ оценками. Непосредственные вычисления (см. [18]) дают следующий их вид:

$$\hat{\beta}_w = \left(\sum_{i=1}^{n}\sum_{t=1}^{T} (x_{ti} - \overline{x}_i)(x_{ti} - \overline{x}_i)' \right)^{-1} \cdot \sum_{i=1}^{n}\sum_{t=1}^{T} (x_{ti} - \overline{x}_i)(y_{ti} - \overline{y}_i); \qquad \textbf{(11.8)}$$

$$\hat{\beta}_B = \left(\sum_{i=1}^{n} (x_i - \overline{x})(x_i - \overline{x})' \right)^{-1} \cdot \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}), \qquad \textbf{(11.9)}$$

где $\overline{y} = \frac{1}{n}\sum_{i=1}^{n}\overline{y}_i$; $\overline{x} = \frac{1}{n}\sum_{i=1}^{n}\overline{x}_i$;

x_i' — транспонированный вектор размерности $k+1$.

11.3. Модели с фиксированным и случайным эффектами

Индивидуальность объектов в панели данных может носить различный характер. Объекты могут быть существенно различными, и их нельзя рассматривать как выборку из некоторой генеральной совокупности. Таковыми, например, могут быть государства, регионы, крупные предприятия. В этом случае величины α_i следует считать постоянными параметрами. Такие модели называются *моделями* с *фиксированным эффектом.*

В других случаях индивидуальность объектов носит случайный характер, то есть объекты панели считаются выборкой из некоторой генеральной совокупности. В этом случае характеризующие индивидуальность величины α_i рассматриваются в виде

$$\alpha_i = \alpha + u_i,$$

где u_i — случайные величины, некоррелирующие с ε_{ti} и регрессорами, и $M(u_i) = 0$.

Такие модели называются *моделями со случайным эффектом.*

С математической точки зрения типы эффектов различаются следующим образом: в моделях со случайным эффектом считается, что индивидуальные факторы не коррелируют с регрессорами, т.е.

$$\rho(u_i, X_{j,i}) = 0, \qquad \textbf{(11.10)}$$

между тем как в моделях с фиксированным эффектом равенство **(11.10)** не выполняется.

Рассмотрим следующий пример (заимствован из [18]).

Оценивается модель производственной функции Кобба-Дугласа $Q = AK^{\alpha}L^{\beta}$ топливно-энергетических предприятий: Q — выпуск, K — капиталовложения, L — трудозатраты, α, β — соответствующие эластичности, которые и требуется оценить.

Для этого следует рассмотреть уравнения

$$\ln Q_{ti} = \ln A_i + \alpha \ln K_{ti} + \beta \ln L_{ti} \qquad \textbf{(11.11)}$$

как регрессию на объясняемую переменную $\ln Q$ и объясняющие переменные $\ln K$, $\ln L$. При этом панельные данные содержат данные $t = 1, ..., 8$, $i = 1, ..., 48466$.

Если считать A — постоянной для всех предприятий величиной, то модель **(11.11)** оценивается с помощью обычного метода наименьших квадратов. При этом получается следующее уравнение

$$\ln Q = -2,48807 + 0,32957 \ln K + 0,92838 \ln L. \qquad \textbf{(11.12)}$$

Все регрессоры значимы и $R^2 = 0,5805$.

Однако предприятия, очевидно, существенно различаются и по размерам, и по эффективности, и по качеству управления. Так что представление об однородности объектов панели скорее всего неправомочно, и модель следует оценивать с учетом индивидуальных эффектов. Мы продолжим рассмотрение этого примера в последующих параграфах.

11.4. Оценивание модели с фиксированным эффектом

Введем обозначения

$$Y = \begin{pmatrix} y_{11} \\ \dots \\ y_{1n} \\ y_{21} \\ \dots \\ y_{2n} \\ \dots \\ y_{T1} \\ \dots \\ y_{Tn} \end{pmatrix}; \quad X = \left(\begin{array}{c|c|c|c} x_{1,11} & x_{2,11} & \dots & x_{k,T1} \\ \dots & \dots & \dots & \dots \\ x_{1,1n} & x_{2,1n} & \dots & x_{k,Tn} \\ \hline x_{1,21} & x_{2,21} & \dots & x_{k,21} \\ \dots & \dots & \dots & \dots \\ x_{1,2n} & x_{2,2n} & \dots & x_{k,2n} \\ \hline \dots & \dots & \dots & \dots \\ \hline x_{1,T1} & x_{2,T1} & \dots & x_{k,T1} \\ \dots & \dots & \dots & \dots \\ x_{1,Tn} & x_{2,Tn} & \dots & x_{k,Tn} \end{array}\right); \qquad \textbf{(11.13.1)}$$

$$I = \left(\begin{array}{c|c|c|c} 1 & & & \\ \dots & 0 & 0 & 0 \\ 1 & & & \\ \hline & 1 & & \\ 0 & \dots & 0 & 0 \\ & 1 & & \\ \hline \dots & \dots & \dots & \dots \\ \hline & & & 1 \\ 0 & 0 & 0 & \dots \\ & & & 1 \end{array}\right); \quad \varepsilon = \begin{pmatrix} \varepsilon_{11} \\ \dots \\ \underline{\varepsilon_{1n}} \\ \varepsilon_{21} \\ \dots \\ \underline{\varepsilon_{2n}} \\ \dots \\ \overline{\varepsilon_{T1}} \\ \dots \\ \varepsilon_{Tn} \end{pmatrix}; \quad \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \dots \\ \underline{\alpha_n} \\ \beta_1 \\ \dots \\ \beta_n \end{pmatrix}. \qquad \textbf{(11.13.2)}$$

Здесь $Y(Tn \times 1)$, $X(Tn \times k)$, $I(Tn \times n)$, $\varepsilon(Tn \times 1)$, $\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ $(n+k, 1)$. Тогда *модель с фиксированным эффектом* может быть записана в виде

$$Y = \tilde{X}\begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \varepsilon, \qquad \textbf{(11.14)}$$

где $\tilde{X} = (I|X) - (nT \times (n+k))$ матрица.

Применяя метод наименьших квадратов к уравнению (**11.14**), получим:

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \tilde{X}'\left(\tilde{X}\tilde{X}\right)^{-1} Y\,. \qquad \textbf{(11.15)}$$

Соответствующие оценки называются *оценками с фиксированным эффектом*. Как правило, наибольший интерес представляют оценки $\hat{\beta}_{FE}$ параметров β.

Непосредственным вычислением можно проверить утверждение:

$$\hat{\beta}_{FE} = \hat{\beta}_{W}\,. \qquad \textbf{(11.16)}$$

Таким образом, использование внутригруппового усреднения позволяет получать оценки параметров модели с фиксированным эффектом. Отметим, что использование уравнения (**11.7**) существенно проще уравнения (**11.2**), так как содержит гораздо меньше параметров. Параметры α_i, в свою очередь, могут быть оценены следующим образом:

$$\hat{\alpha}_i = \overline{y}_i - \sum_{j=1}^{k} \hat{\beta}_{jFE}\overline{x}_{ji}.$$

Подчеркнем, однако, еще раз, что, как правило, наибольший интерес представляют именно оценки параметров β.

Оценим уравнения **(11.11)** примера из §**11.3**, используя модель с фиксированным эффектом. Получим:

$$\ln Q = 0,75995 + 0,11421 \ln K + 0,60393 \ln L . \qquad \textbf{(11.17)}$$

Очевидно, результат **(11.17)** существенно отличается от результата **(11.12)**.

11.5. Оценивание модели со случайным эффектом

Модель со случайным эффектом (см. §**11.3**) имеет следующий вид:

$$y_{ti} = \alpha + \sum_{j=1}^{k} \beta_j x_{j,ti} + u_i + \varepsilon_{ti} . \qquad \textbf{(11.18)}$$

Обозначим

$$U = \begin{pmatrix} u_1 \\ \dots \\ u_n \\ u_1 \\ \dots \\ u_n \\ \dots \\ u_1 \\ \dots \\ u_n \end{pmatrix} \qquad \text{— столбец } (nT \times 1).$$

Тогда в обозначениях **(11.13)** уравнение **(11.18)** записывается в виде:

$$Y = \alpha + X\beta + (u + \varepsilon). \qquad \textbf{(11.19)}$$

Уравнение **(11.19)** представляет собой *обобщенную модель* линейной регрессии: матрица ковариации ошибок регрессии $u + \varepsilon$ имеет вид:

$$\Omega = \begin{pmatrix} (\sigma_u^2 + \sigma_\varepsilon^2) E_T & \sigma_u^2 E_T & & \sigma_u^2 E_T \\ \sigma_u^2 E_T & (\sigma_u^2 + \sigma_\varepsilon^2) E_T & & \sigma_u^2 E_T \\ \dots & \dots & \dots & \dots \\ \sigma_u^2 E_T & \sigma_u^2 E_T & \dots & (\sigma_u^2 + \sigma_\varepsilon^2) E_T \end{pmatrix}. \qquad \textbf{(11.20)}$$

Здесь E_T — единичная матрица порядка T; $\Omega(nT \times nT)$. Для оценивания модели **(11.19)** можно применить метод наименьших квадратов (см. §**7.2**). Соответствующие вычисления можно найти в [18]. Приведем лишь окончательный качественный результат: оценка параметра β в модели со случайным эффектом имеет вид

$$\hat{\beta}_{RE} = W_{\hat{\beta}_B} + (E_k - W)\hat{\beta}_W, \qquad \textbf{(11.21)}$$

где W — матрица, пропорциональная обратной матрице ковариаций оценки $\hat{\beta}_B$.

Эта оценка представляет «средневзвешенное» межгрупповой и внутригрупповой оценок.

Для практического использования формулы **(11.21)** следует оценить матрицу Ω, т.е. параметры σ_u^2 и σ_ε^2. Как обычно, для оценок ошибок регрессии используются суммы квадратов остатков:

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{nT - n - k}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{ti} - \overline{y}_i - (x_{ti} - \overline{x})'\hat{\beta}_{FE})^2,$$

$$\hat{\sigma}_B^2 = \frac{1}{n - k - 1}\sum_{i=1}^{n}(\overline{y}_i - \hat{\alpha}_B - \overline{x}_{ti}'\hat{\beta}_B)^2,$$

$$\hat{\sigma}_u^2 = \hat{\sigma}_B^2 - \frac{1}{T}\hat{\sigma}_\varepsilon^2.$$

Оценим теперь уравнение **(11.11)** с помощью модели со случайным эффектом. Получим:

$$\ln Q = -1{,}13749 + 0{,}24671\ln K + 0{,}77544\ln L. \qquad \textbf{(11.22)}$$

Как видно, получается существенное различие как с уравнением **(11.17)**, так и с уравнением **(11.12)**.

11.6. Проблема выбора модели с панельными данными

Рассмотренный выше пример показывает, что результат оценивания в значительной мере зависит от выбора модели, т.е. от структуры выборки.

При выборе статической модели панельных данных исследователь должен определить, считать ли объекты существенно индивидуальными, случайно выбранными из однородной генеральной совокупности, либо же лишенными всякой индивидуальности. Неверный выбор может привести либо к искажению результатов, либо к значительной потере эффективности.

Сразу следует заметить, что оценки модели с фиксированным эффектом ($\hat{\beta}_{BE}$) в любом случае являются состоятельными и несмещенными. Поэтому в случае достаточно большой выборки (при большом T) их можно использовать всегда. Однако, если регрессоры не коррелируют с индивидуальными эффектами — так обстоит дело в моделях со случайным эффектом — оценки $\hat{\beta}_{RE}$ оказываются неэффективными, и это очень существенно для выборок небольшого объема. В то же время, если в моделях присутствует фиксированный эффект, оценки $\hat{\beta}_{RE}$ оказываются несостоятельными.

Проблему выбора модели исследователь решает индивидуально, часто на основании интуиции. При этом, однако, можно использовать статистические тесты.

Итак, пусть рассматривается модель

$$y_{ti} = \alpha_i + \sum_{j=1}^{k} \beta_j x_{j,ti} + \varepsilon_{ti}$$

и требуется сделать выбор между моделью с фиксированным и случайным эффектом (заметим, что классическую модель можно рассматривать и как частный случай модели с фиксированным эффектом, если все α_i равны, и как частный случай модели со случайным эффектом, если $\sigma_u^2 = 0$).

Выбор делается в пользу случайного эффекта, если принимается гипотеза H_0:

$$r\left(\alpha_i, X_{j,ti}\right) = 0.$$

Проверку гипотезы H_0 можно осуществить с помощью *теста Хаусмана.* Его идея основана на том, что в случае справедливости H_0 обе оценки $\hat{\beta}_{FE}$ и $\hat{\beta}_{RE}$ являются состоятельными, а, следовательно, не должны существенно различаться, т.е. матрица $Cov\left(\hat{\beta}_{FE} - \hat{\beta}_{RE}\right)$ не должна быть слишком большой. Статистика

$$\left(\hat{\beta}_{FE}-\hat{\beta}_{RE}\right)'\mathrm{Cov}\left(\hat{\beta}_{FE}-\hat{\beta}_{RE}\right)^{-1}\left(\hat{\beta}_{FE}-\hat{\beta}_{RE}\right) \qquad \textbf{(11.23)}$$

имеет χ^2 — распределение с k степенями свободы.

Для использования статистики **(11.23)** необходимо получить оценку матрицы $Cov\left(\hat{\beta}_{FE}-\hat{\beta}_{RE}\right)$. Можно показать, что при выполнении нулевой гипотезы выполняется асимптотическое равенство

$$\mathrm{Cov}\left(\hat{\beta}_{FE}-\hat{\beta}_{RE}\right)\approx\mathrm{Cov}\left(\hat{\beta}_{FE}\right)-\mathrm{Cov}\left(\hat{\beta}_{RE}\right).$$

При этом (см. [18]):

$$\hat{\mathrm{Cov}}\left(\hat{\beta}_{RE}\right)=\hat{\sigma}_{\varepsilon}^2\left(\sum_{i=1}^{n}\sum_{t=1}^{T}(x_{ti}-\bar{x}_i)(x_{ti}-\bar{x}_i)'+\frac{\hat{\sigma}_{\varepsilon}^2 T}{\hat{\sigma}_{\varepsilon}^2+T\hat{\sigma}_u^2}\sum_{i=1}^{n}(\bar{x}_i-\bar{x})(\bar{x}_i-\bar{x})'\right)^{-1},$$

$$\hat{\mathrm{Cov}}\left(\hat{\beta}_{FE}\right)=\hat{\sigma}_{\varepsilon}^2\left(\sum_{i=1}^{n}\sum_{t=1}^{T}(x_{ti}-\bar{x})(x_{ti}-\bar{x}_i)'\right)^{-1},$$

где

$$\hat{\sigma}_{\varepsilon}^2=\frac{1}{nT-n-k}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{ti}-\bar{y}_i-(x_{ti}-\bar{x})'\hat{\beta}_{FE})^2.$$

Тест Хаусмана реализован в большинстве современных эконометрических программ.

Применим тест Хаусмана в рассматриваемом в данной главе примере оценки эластичностей производственной функции (уравнения **(11.11)**). Вычисленное значение статистики **(11.23)** равно $\chi^2=141{,}01$. При этом $\chi^2_{0,05;2}(X)=5{,}99$, $\chi^2_{0,01;2}(X)=9{,}21$, т.е. гипотеза отвергается при любом разумном уровне значимости. Это означает, что при оценивании уравнения **(11.11)** следует выбирать модель с фиксированным эффектом и применять в качестве оценки уравнение **(11.17)**.

Выше мы рассматривали только статические модели. Динамические модели, учитывающие корреляцию между временными наблюдениями, оказываются значительно более сложными, и их рассмотрение выходит за рамки большинства стандартных курсов.

11.7. Бинарные модели с дискретными зависимыми переменными

Рассмотрим теперь зависимые переменные, принимающие конечное число значений. Как правило, без потери общности можно считать, что это значения 0, 1, ..., *m*. Подобные ситуации возникают в тех случаях, когда значения объясняемой переменной соответствуют выбору решения из набора $m + 1$ возможных решений.

Наиболее простыми моделями оказываются *модели бинарного выбора*. В этом случае объясняемая переменная принимает всего два значения — 0 и 1. Можно считать, что 1 соответствует положительному решению, а 0 — отрицательному.

Пусть Y_i — объясняемая величина, где $y_i = 0$ или $y_i = 1$. Величина Y принимает одно из своих возможных значений под воздействием факторов $X_1, ..., X_k$, которые могут принимать непрерывные значения.

Рассматривается регрессионная модель

$$y_i = F(x_{1i}, ..., x_{ki};\ \beta_1, ..., \beta_k) + \varepsilon_i, \qquad \textbf{(11.24)}$$

в которой $\beta_1, ..., \beta_k$ — параметры модели, соответствующие регрессорам $X_1, ..., X_k$.

Наиболее простой является линейная модель регрессии

$$y_i = \sum_{j=1}^{k} \beta_j x_{ji} + \varepsilon_i. \qquad \textbf{(11.25)}$$

Так как в уравнении **(11.24)** полагается, что $M(\varepsilon_i) = 0$, то $M(y_i) = F(x, \beta)$. В то же время

$$M(y_i) = 0 \cdot P(y_1 = 0) + 1 \cdot P(y_1 = 1) = P(y_1 = 1),$$

откуда следует, что

$$P(y_i = 1) = F(x, \beta). \qquad \textbf{(11.26)}$$

Уравнение **(11.26)** называется *уравнением модели бинарного выбора*.

В частности, для линейной модели **(11.25)** имеем:

$$P(y_i = 1) = \sum_{j=1}^{k} \beta_j x_{ji}. \qquad \textbf{(11.27)}$$

Использование уравнения **(11.27)** сопряжено со значительными трудностями. В первую очередь, это связано с тем, что величины $\sum_{j=1}^{n} \hat{\beta}_i x_{ji}$, получаемые оцениванием параметров β, могут не попадать в промежуток [0, 1]. Кроме этого, неправомочно предположение о том, что ошибки ε_i имеют нормальное распределение. Линейную модель можно использовать в ряде случаев при большом числе наблюдений и достаточно точной спецификации модели. В основном же она используется лишь как грубый инструмент первичной обработки данных.

11.8. Probit- и logit-модели

Перечисленные в предыдущем параграфе трудности легко преодолимы, если в качестве функции $F(X, \beta)$ в равенстве **(11.26)** выбирается функция распределения некоторой случайной величины.

Подобный выбор становится естественным, если предположить, что значение дискретной переменной Y скачкообразно изменяется в зависимости от значений некоторой — как правило, ненаблюдаемой — количественной переменной Y^*. Наиболее простой случай соответствует наличию порогового значения, т.е.

$$y_i = \begin{cases} 1, & \text{если } y_i^* \geq 0; \\ 0, & \text{если } y_i^* < 0. \end{cases}$$

Величину y_i^* можно интерпретировать как разность полезностей альтернативы 1 и 0. Пусть количественная переменная y_i^* удовлетворяет регрессионному уравнению

$$y_i^* = \sum_{j=1}^{k} \tilde{\beta}_j x_{ij} + \varepsilon_i, \qquad \textbf{(11.28)}$$

причем константа включена в число регрессоров **(11.28)**, а ошибки ε_i независимы и одинаково распределены с нулевым средним и дисперсией σ^2. Пусть функция $F(t)$ — функция распределения случайной величины $\frac{\varepsilon_i}{\sigma}$. Тогда

$$P(y_i = 1) = P(y_i^* \geq 0) = P\left(\sum_{j=1}^{k} x_{ij}\tilde{\beta}_j + \varepsilon_i \geq 0\right) =$$

$$= P\left(\varepsilon_i \geq -\sum_{j=1}^{k} x_{ij}\tilde{\beta}_j\right) = P\left(\varepsilon_i \leq \sum_{j=1}^{k} x_{ij}\tilde{\beta}_j\right) = F\left(\sum_{j=1}^{k} x_{ij}\frac{\tilde{\beta}_j}{\sigma}\right).$$

Обозначив $\beta = \frac{\tilde{\beta}}{\sigma}$, получим уравнение **(11.26)** в форме

$$P(y_i = 1) = F\left(\sum_{j=1}^{k} \beta_j x_{ij}\right). \qquad \textbf{(11.29)}$$

Уравнение **(11.29)** представляет собой стандартную форму модели бинарного выбора.

Наиболее естественно в качестве функции $F(t)$ выбрать функцию стандартного нормального распределения. Это соответствует тому, что **(11.28)** представляет собой классическую нормальную линейную модель. Соответствующая дискретная модель **(11.29)** в этом случае называется *probit-моделью.*

Также используется функция логистического распределения

$$F(t) = \Lambda(t) = \frac{e^t}{1 + e^t} \qquad \textbf{(11.30)}$$

и соответствующая модель **(11.29)** при таком выборе называется *logit—моделью.*

Выбор функции **(11.30)** объясняется тем, что ее вид существенно проще, чем у функции стандартного нормального распределения. При этом для небольших (по модулю) t эти функции достаточно близки друг к другу, и качественные выводы, сделанные по probit- и logit-моделям, в основном совпадают. В то же время для больших значений регрессоров возможны и значимые различия.

Важно отметить то обстоятельство, что модель **(11.29)** нелинейная. Это приводит к тому, что предельный эффект каждого фактора X_j (производная $\frac{\partial P(y=1)}{\partial x}$) является переменным и зависящим от значений всех остальных факторов. Так что для получения представления о «среднем» предельном эффекте следует вычислить производные

$$\frac{\partial P(y=1)}{\partial x_i} = F'\left(\sum_{j=1}^{k} \beta_j x_{ij}\right)\beta_i$$

для средних по выборке значений независимых переменных X.

Для оценивания модели используется метод максимального правдоподобия. Если наблюдения $y_1, ..., y_n$ независимы, то функция правдоподобия имеет вид:

$$L = \prod_{y_i=0}\left(1 - F\left(\sum_{j=1}^{k}\beta_j x_{ij}\right)\right) \prod_{y_i=1} F\sum_{j=1}^{k}\beta_j x_{ij} =$$

$$= \prod_i F\left(\sum_{j=1}^{k}\beta_j x_{ij}\right)^{y_i}\left(1 - F\sum_{j=1}^{k}\beta_j x_{ij}\right)^{1-y_i}.$$

Тогда

$$\ln L = \sum_{i=1}^{n}\left[y_i \ln F\left(\sum_{j=1}^{k}\beta_j x_{ij}\right) + (1-y_i)\ln\left(1 - F\sum_{j=1}^{k}\beta_j x_{ij}\right)\right].$$

Условия правдоподобия имеют вид:

$$\frac{\partial \ln L}{\partial \beta} = 0.$$

Отсюда получаем следующее уравнение для определения β:

$$\sum_{i=1}^{n}\left(\frac{y_i \varphi\left(\sum_{j=1}^{k}\hat{\beta}_j x_{sj}\right)}{F\left(\sum_{j=1}^{k}\hat{\beta}_j x_{sj}\right)} - \frac{(1-y_i)\varphi\left(\sum_{j=1}^{k}\tilde{\beta}_j x_{sj}\right)}{1 - F\left(\sum_{j=1}^{k}\hat{\beta}_j x_{sj}\right)}\right)x_i = 0, \quad \textbf{(11.31)}$$

где $\varphi(t)$ — плотность вероятности, соответствующая функции распределения $F(t)$.

Для logit-модели уравнения **(11.31)** приобретают более простой вид

$$\sum_{i=1}^{n}\left(y_i - \Lambda\left(\sum_{j=1}^{k}\hat{\beta}_j x_{sj}\right)\right)x_i = 0,$$

что и является одной из причин ее популярности.

Можно доказать, что для probit- и logit-моделей функции правдоподобия $\ln L$ вогнуты, и следовательно, уравнения **(11.31)** дают оценки максимального правдоподобия.

11.9. Дискретные модели с панельными данными

Пусть рассматриваются панельные данные (n объектов и T моментов времени) с индивидуальными эффектами α_i. Модель бинарного выбора имеет вид

$$P(y_{ti} = 1) = F(x_{j,ti},\ \beta_j,\ \alpha_i). \qquad \textbf{(11.32)}$$

Подход к построению уравнений модели, используемый для обычных пространственных данных (см. § **11.3**), может быть применен и для панельных данных. Пусть y^*_{ti} — ненаблюдаемая переменная, обуславливающая скачкообразное изменение y_{ti}:

$$y_{ti} = 1 \text{ при } y^*_{ti} \geq 0,$$

$$y_{ti} = 0 \text{ при } y^*_{ti} < 0$$

и

$$y^*_{ti} = \alpha_i + \sum_{j=1}^{k} \beta_j x_{j,ti} + \varepsilon_{ti}.$$

Уравнение **(11.32)** может быть оценено методом максимального правдоподобия, если функция $F(t)$ является функцией распределения некоторой случайной величины. Проблема, однако, в том, что для получения адекватных оценок параметров требуется, чтобы T было велико (состоятельность имеет место при $T \to \infty$). Однако на практике панельные данные бывают с большим количеством объектов n и не слишком большим числом временных изменений T. Между тем, с ростом n увеличивается и количество оцениваемых параметров α_i, т.е. при $n \to \infty$ оценки не оказываются состоятельными.

В отличие от линейных моделей внутригрупповое и межгрупповое усреднение здесь не дает возможности исключить параметры α_i.

Преодоление этих трудностей может представлять собой сложную задачу. Некоторые методы описаны в [18]. При этом оказывается, что для модели с фиксированным эффектом удобнее использовать функцию распределения логистического закона (logit-модель), а для модели со случайным эффектом — функцию стандартного нормального распределения (probit-модель).

11.10. Выборки с ограничениями

Помимо ограничений на тип данных в моделях встречаются также ограничения на выборочные наблюдения. Чаще всего используются ограничения в виде урезания и цензурирования выборки.

Выборка называется *урезанной*, если отбрасываются наблюдения, не удовлетворяющие некоторым априорным ограничениям — чаще всего если значения зависимой переменной меньше некоторой заданной величины. Например, при исследовании зависимости затрат на отдых от дохода можно исключить тех индивидуумов, доход которых меньше прожиточного минимума.

Выборка называется *цензурированной*, если для части наблюдений фиксируется не истинное значение зависимой переменной, а некоторое ее усеченное значение. Основу изучения подобных моделей заложил Дж. Тобин (1958). Он исследовал зависимость эластичности спроса на автомобили от дохода. При этом для некоторых семей (как правило, с низким доходом) расходы на автомобили равнялись нулю.

Тобин предположил модель вида:

$$y_t = \begin{cases} y_t^*, & \text{если } y_t^* > 0; \\ 0, & \text{если } y_t^* \le 0, \end{cases} \qquad \textbf{(11.33)}$$

где y_t^* — ненаблюдаемая величина, удовлетворяющая стандартному уравнению регрессии;

$$y_t^* = x_t'\beta + \varepsilon_t. \qquad \textbf{(11.34)}$$

Модель **(11.33)**, **(11.34)** называется *tobit- моделью.*

Так как стандартный метод наименьших квадратов в моделях с урезанными и цензурированными выборками дает смешанные и несостоятельные оценки параметров, для их оценивания используется метод максимального правдоподобия. Более подробно эти вопросы освещены в [18].

В заключение главы, отметим, что возможности оценивания моделей с панельными данными, дискретными данными, а также с урезанными и цензурированными выборками имеются в современных компьютерных программах по эконометрике. Выбор logit-, probit- и tobit-моделей присутствует в таких программах как самостоятельные опции.

Упражнения

11.1. Исследуется зависимость изменения спортивного результата y (в условных баллах) и тренировочных часов в месяцах x. В таблице приведены данные наблюдений по 10 спортсменам в течение 5 месяцев.

x \ y	1	2	3	4	5	6	7	8	9	10
1	102 / 13	110 / 10	88 / 8	96 / 14	124 / 10	108 / 8	112 / 6	108 / 4	110 / 4	105 / 4
2	95 / 7	136 / 12	96 / 12	114 / 8	130 / 8	120 / −2	126 / −1	120 / 10	100 / 6	120 / 12
3	98 / 3	146 / −2	120 / −4	90 / 10	146 / −6	100 / 10	105 / 10	120 / 3	100 / 2	140 / −2
4	120 / −1	130 / 8	100 / 18	96 / −2	120 / 12	100 / 2	100 / 6	130 / −1	100 / 1	120 / 6
5	100 / 2	138 / 10	100 / 3	98 / 3	120 / 3	102 / 4	105 / 8	118 / 2	100 / −2	110 / 3

Оценить линейную модель панельных данных $y = \alpha + \beta x$, используя: а) обычный метод наименьших квадратов; б) модель с фиксированным эффектом; в) модель со случайным эффектом.

Какая модель оказывается более адекватной в случае, если спортсмены: а) мастера спорта; б) разрядники.

11.2. Исследуется зависимость отказа женщины от работы:

$y = 1$, если отказ произошел из-за рождения ребенка;

$d = 1$, если ребенок родился.

Рассматривается модель $P(y_t = 1) = F(\alpha + \beta d)$. Ниже приведены результаты 100 измерений:

d \ y	0	1
0	50	6
1	18	26

Оценить параметры α, β, используя probit- и logit-модели.

11.3 Рассматривая модель бинарного выбора с панельными данными:

$y_{ti}^{*} = \alpha_i + \beta x_{ti} + \varepsilon_{ti}$;

$y_{ti} = 1$, если $y_{ti}^{*} \geq 0$;

$y_{ti} = 0$, если $y_{ti}^{*} < 0$,

где x_{ti} — доход i-го домашнего хозяйства $(i = 1, \ldots, 10)$ в период наблюдения t $(t = 1, \ldots, 5)$, y_{ti} — участие в паевом фонде ($y_{ti} = 1$, в случае участия). Данные исследования приведены в таблице:

x \ y	1	2	3	4	5	6	7	8	9	10
1	300 / 0	180 / 0	1200 / 1	1400 / 1	280 / 0	130 / 0	1500 / 1	1200 / 1	300 / 0	560 / 0
2	500 / 1	200 / 0	1300 / 1	1500 / 1	300 / 0	140 / 0	1600 / 1	1200 / 1	300 / 0	680 / 0
3	520 / 1	300 / 0	0 / 1	1500 / 1	380 / 0	160 / 0	1500 / 1	1200 / 1	300 / 0	700 / 1
4	500 / 1	500 / 0	0 / 0	1500 / 1	380 / 0	180 / 0	1500 / 1	1200 / 1	300 / 0	700 / 1
5	400 / 1	500 / 1	800 / 0	1500 / 0	380 / 0	200 / 0	1400 / 1	1200 / 1	300 / 0	650 / 1

Построить функцию правдоподобия для модели.

11.4. Имеются следующие данные наблюдений по затратам индивидуумов на приобретение акций и получаемых по ним доходам:

Затраты	100	85	90	500	0	0	300	0	200	300
Доход	1050	980	960	5600	300	280	3200	340	2500	3600

Составить tobit-модель и записать для нее функцию правдоподобия.

Приложения

Глава 12
Элементы линейной алгебры

Аппарат линейной, и в частности, матричной алгебры является необходимым инструментом для компактного и эффективного описания и анализа эконометрических моделей и методов. Ниже (как правило, без доказательств) приводят краткие сведения из линейной алгебры, необходимые для изучения эконометрики. Более подробно с методами линейной алгебры можно познакомиться, например, в пособиях [4]—[6].

12.1. Матрицы

Матрицей размера ($m \times n$) или *mn-матрицей называется таблица чисел, содержащая m строк* и *n столбцов*. Матрицы обозначаются прописными латинскими буквами, например,

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2j} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{in} \\ a_{m1} & a_{m2} & \dots & a_{mj} & \dots & a_{mn} \end{pmatrix} \qquad \textbf{(12.1)}$$

или $A_{m\times n} = (a_{ij})$, где a_{ij} — элемент матрицы. Наряду с круглыми скобками используются и другие обозначения: [ ], || ||.

Виды матриц

Если $m = 1$,то $A_{1\times n} = (a_{11} a_{12} \dots a_{1n})$ – *матрица* (или *вектор*)-*строка* размера *n*.

Если $n = 1$, то $A_{m\times 1} = \begin{pmatrix} a_{11} \\ a_{21} \\ \dots \\ a_m \end{pmatrix}$ — *матрица* (или *вектор*)-*столбец* размера *m*.

Если $m = n$, то $A_{n\times n}=A_n$ — *квадратная матрица n-го порядка.*

Если $a_{ij} = 0$ при $i\neq j$, то $A_n=D$ — *диагональная матрица*:

$$D=\begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix} = \operatorname{diag}(a_{11}\ a_{22}\dots a_{nn}).$$

Если $\begin{cases} a_{ij}=0 \text{ при } i\neq j; \\ a_{ij}=1 \text{ при } i=j, \end{cases}$ то $A_n=E_n$ (или E), *единичная матрица n-го порядка*:

$$E_n=\begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}.$$

Если $a_{ij}=0$ ($i=1,\dots, m$; $j=1,\dots, n$), то $A_{m\times n}=\mathbf{0}_{m\times n}$ (или $\mathbf{0}$) — *нулевая матрица*, или *нуль-матрица*:

$$\mathbf{0}_{m\times n}=\begin{pmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{pmatrix}.$$

Равенство матриц

Матрицы $A_{m\times n}=(a_{ij})$ и $B_{m\times n}=(b_{ij})$ равны, т. е. $A = B$, если $a_{ij} = b_{ij}$, $i = 1,\dots, m$; $j = 1,\dots, n$.

Операции над матрицами

1. *Произведение матрицы* $A_{m\times n}=(a_{ij})$ *на число* λ есть матрица $B_{m\times n}=(b_{ij})$:

$$B = A\lambda, \text{ если } b_{ij} = a_{ij}\,\lambda,\ i = 1,\dots, m;\ j = 1,\dots, n. \qquad \textbf{(12.2)}$$

В частности, $A\cdot 0=\mathbf{0}$.

2. *Сумма двух матриц* $A_{m\times n}=(a_{ij})$ и $B_{m\times n}=(b_{ij})$ есть матрица $C_{m\times p}=(c_{ij})$:

$$C = A+B, \text{ если } c_{ij} = a_{ij}+b_{ij}, \quad i = 1,\dots, m;\ j = 1,\dots, n. \qquad \textbf{(12.3)}$$

В частности, $A+\mathbf{0}=A$.

3. *Произведение матрицы* $A_{m\times n}=(a_{ij})$ *на матрицу* $B_{n\times p}=(b_{ij})$ есть матрица $C_{m\times p}=(c_{ij})$:

$$C = AB, \text{ если } c_{ij}=\sum_{k=1}^{n} a_{ik}b_{kj},\ i = 1,\dots, m;\ j = 1,\dots, n. \qquad \textbf{(12.4)}$$

Из определения следует, что для умножения матрицы A и B должны быть с о г л а с о в а н н ы м и: *число n столбцов матрицы A должно быть равно числу строк n матрицы B.*

▶ **Пример 12.1.** Даны матрицы:

$$A=\begin{pmatrix}3 & 4\\ 1 & 2\\ 5 & 6\end{pmatrix},\ B=\begin{pmatrix}2 & 7 & 8\\ 9 & 2 & 0\end{pmatrix}.$$

Найти AB и BA.

Р е ш е н и е. Размер матрицы произведения $A_{3\times 2}\cdot B_{2\times 3}=C_{3\times 3}$;

$$C_{3\times 3}=AB=\begin{pmatrix}3\cdot 2+4\cdot 9 & 3\cdot 7+4\cdot 2 & 3\cdot 8+4\cdot 0\\ 1\cdot 2+2\cdot 9 & 1\cdot 7+2\cdot 2 & 1\cdot 8+2\cdot 0\\ 5\cdot 2+6\cdot 9 & 5\cdot 7+6\cdot 2 & 5\cdot 8+6\cdot 0\end{pmatrix}=\begin{pmatrix}42 & 29 & 24\\ 20 & 11 & 8\\ 64 & 47 & 40\end{pmatrix}.$$

Аналогично $B_{2\times 3}\cdot A_{3\times 2}=C_{2\times 2}$;

$$C_{2\times 2}=BA=\begin{pmatrix}2\cdot 3+7\cdot 1+8\cdot 5 & 2\cdot 4+7\cdot 2+8\cdot 6\\ 9\cdot 3+2\cdot 1+0\cdot 5 & 9\cdot 4+2\cdot 2+0\cdot 6\end{pmatrix}=\begin{pmatrix}53 & 70\\ 29 & 40\end{pmatrix}.$$

Получили, что произведения матриц AB и BA существуют, но являются матрицами разных размеров (порядков). ▶

В частном случае

$$AE=EA=A. \qquad \textbf{(12.5)}$$

4. *Целая положительная степень* A^m ($m>1$) квадратной матрицы есть $A^m=\underbrace{A\cdot A...A}_{m\text{ раз}}$.

По определению,

$$A^0=E,\ \ A'=A,\ \ A^m\cdot A^k=A^{m+k},$$

$$\left(A^m\right)^k=A^{mk}. \qquad \textbf{(12.6)}$$

О с о б е н н о с т и о п е р а ц и й н а д м а т р и ц а м и.

В общем случае $AB\neq BA$, т. е. умножение матриц некоммутативно. Если $A^m=\mathbf{0}$ или $AB=\mathbf{0}$, то это еще не означает, что $A=\mathbf{0}$ или $B=\mathbf{0}$.

Транспонирование матрицы *— переход от матрицы* A *к матрице* A' (или A^T), *в которой строки и столбцы поменялись местами с сохранением их порядка.*

Если $A_{m\times n}=(a_{ij})$, то $A'_{n\times m}=(a_{ji})$. Например, если

$$A=\begin{pmatrix}2 & 3 & 4\\ 5 & 2 & 0\end{pmatrix},\ \text{то}\ A'=\begin{pmatrix}2 & 5\\ 3 & 2\\ 4 & 0\end{pmatrix}.$$

Свойства операции транспонирования:

1. $(A')'=A$.
2. $(\lambda A)'=\lambda A'$.
3. $(A+B)'=A'+B'$.
4. $(AB)'=B'A'$. **(12.7)**

Заметим, если $A_{m\times n}=(a_{ij})$, то AA' и $A'A$ есть *квадратные* матрицы соответственно m-го и n-го порядков.

В частности, если $A = (a_1\ a_2\ ...\ a_n)'$ есть матрица (вектор)-столбец, то $A'A=\sum_{i=1}^{n}a_i^2$ — квадратная матрица 1-го порядка, т. е. *число*; а $AA'=\begin{pmatrix}a_1^2 & a_1a_2 & ... & a_1a_n\\ a_2a_1 & a_2^2 & ... & a_2a_n\\ a_na_1 & a_na_2 & ... & a_n^2\end{pmatrix}$ — квадратная матрица n-го порядка.

12.2. Определитель и след квадратной матрицы

Определителем (или ***детерминатом***) *квадратной матрицы* n-го *порядка* (или ***определителем n-го порядка***) $A_{n\times n}=A_n=(a_{ij})$ *называется число, обозначаемое* $|A_n|$ (или Δ_n det$A$) и *определяемое по следующим правилам*:

при $n = 1$ $$\Delta_1=|A_1|=|a_{11}|=a_{11};$$ **(12.8)**

при $n = 2$ $$\Delta_2=|A_2|=\begin{vmatrix}a_{11} & a_{12}\\ a_{21} & a_{22}\end{vmatrix}=a_{11}a_{22}-a_{21}a_{12};$$ **(12.9)**

при $n = 3$ $$\Delta_3=|A_3|=\begin{vmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{vmatrix}=$$

$$= a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - \\ - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32}. \qquad \textbf{(12.10)}$$

▶ **Пример 12.2**. Вычислить определители:

а) $\Delta_2 = \begin{vmatrix} 4 & 7 \\ 3 & -8 \end{vmatrix}$; *б*) $\Delta_3 = \begin{vmatrix} 5 & -8 & 7 \\ -3 & 4 & 5 \\ -6 & 3 & 2 \end{vmatrix}$.

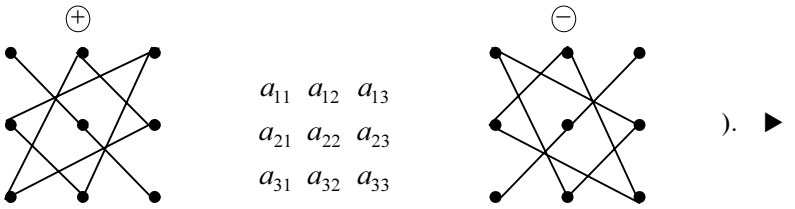
Р е ш е н и е:
а) По формуле (**12.9**)

$$\Delta_2 = \begin{vmatrix} 4 & 7 \\ 3 & -8 \end{vmatrix} = 4 \cdot (-8) - 3 \cdot 7 = -53;$$

б) По формуле (**12.10**)

$$\Delta_3 = \begin{vmatrix} 5 & -8 & 7 \\ -3 & 4 & 5 \\ -6 & 3 & 2 \end{vmatrix} = 5 \cdot 4 \cdot 2 + (-8) \cdot 5(-6) + 7 \cdot (-3) \cdot 3 -$$
$$- 7 \cdot 4(-6) - (-8)(-3) \cdot 2 - 5 \cdot 5 \cdot 3 = 262.$$

(При вычислении определителя 3-го порядка Δ_3 использовали *правило треугольников*, согласно которому соответствующие произведения трех элементов матрицы берутся со знаками «+» и «−»:



). ▶

Определитель квадратной матрицы n-го порядка (или ***определитель n-го порядка***) при *любом n* определяется более сложно. Он может быть вычислен с помощью *разложения по элементам строки или столбца* (*теоремы Лапласа*):

$$|A| = \sum_{\substack{i=1 \\ (j=1)}}^{n} a_{ij} A_{ij},\ j = 1,..., n\ (i = 1,..., n), \qquad \textbf{(12.11)}$$

где a_{ij} — элементы любой строки (столбца),

A_{ij} — алгебраическое дополнение элемента a_{ij}:

$$A_{ij} = (-1)^{i+j} M_{ij}; \qquad \textbf{(12.12)}$$

M_{ij} — минор элемента a_{ij} — определитель матрицы $(n-1)$-го порядка, полученной из матрицы A вычеркиванием i-й строки и j-го столбца.

▶ **Пример 12.3.** Вычислить определитель Δ_3 матрицы из примера **12.2**, разложив его по элементам строки (столбца).

Р е ш е н и е.

Раскладывая по элементам, например, 1-ой строки, получим по формуле (**12.11**) с учетом (**12.12**):

$$\begin{vmatrix} 5 & -8 & 7 \\ -3 & 4 & 5 \\ -6 & 3 & 2 \end{vmatrix} = 5A_{11} - 8A_{12} + 7A_{13} =$$

$$= 5(-1)^{1+1} \begin{vmatrix} 4 & 5 \\ 3 & 2 \end{vmatrix} + (-8)(-1)^{1+2} \begin{vmatrix} -3 & 5 \\ -6 & 2 \end{vmatrix} + 7(-1)^{1+3} \begin{vmatrix} -3 & 4 \\ -6 & 3 \end{vmatrix} =$$

$$= 5(-7) + (-8)(-24) + 7 \cdot 15 = 262\ .\ \blacktriangleright$$

С в о й с т в а о п р е д е л и т е л е й:

1. $|A| = |A'|$.

2. При перестановке любых строк матрицы меняется только знак определителя матрицы.

3. $|A| = 0$, если элементы двух строк (или столбцов) пропорциональны (в частном случае — равны).

4. За знак определителя матрицы можно выносить общий множитель элементов любой строки (столбца).

5. Определитель матрицы не изменится, если к элементам любой строки (или столбца) прибавить элементы другой строки (или столбца), умноженные на одно и то же число.

6. $|AB| = |BA| = |A| \cdot |B|$, где A, B — квадратные матрицы.

7. $|\lambda A| = \lambda^n |A|$, где λ — число, n — порядок матрицы A.

8. $|\mathrm{diag}\,(a_{11}\ a_{22}...a_{nn})| = a_{11}a_{22}...a_{nn}$.

9. $|E_n| = 1$.

Следом *квадратной матрицы A n-го порядка* (обозначается $\mathrm{tr}(A)$ (от английского слова «*trace*»)) *называется сумма ее диагональных элементов*:

$$\mathrm{tr}(A) = a_{11} + a_{22} + ... + a_{nn} = \sum_{i=1}^{n} a_{ii}\ . \qquad \textbf{(12.13)}$$

Свойства следа матриц:

1. $\mathrm{tr}(E_n) = n$.
2. $\mathrm{tr}(\lambda A) = \lambda \mathrm{tr}(A)$.
3. $\mathrm{tr}(A') = \mathrm{tr}(A)$.
4. $\mathrm{tr}(A + B) = \mathrm{tr}(A) + \mathrm{tr}(B)$. **(12.14)**
5. $\mathrm{tr}(AB) = \mathrm{tr}(BA)$.

В частности, если A — $(n \times 1)$ вектор-столбец, $B = A'$, то

$$\mathrm{tr}(AA') = \mathrm{tr}(A'A), \qquad \textbf{(12.15)}$$

где, напомним, AA' и $A'A$ — соответственно квадратные матрицы n-го и 1-го порядков.

12.3. Обратная матрица

Матрица A называется ***невырожденной*** (***неособенной***), *если* $|A| \neq 0$. *В противном случае* (при $|A| = 0$) *A —* ***вырожденная*** (***особенная***) матрица.

Матрица A^{-1} называется ***обратной по отношению к квадратной матрице*** *A, если*

$$A^{-1}A = AA^{-1} = E. \qquad \textbf{(12.16)}$$

Для существования обратной матрицы A^{-1} необходимо и достаточно, чтобы $|A| \neq 0$, т. е. матрица A была невырожденной.

Обратная матрица может быть найдена по формуле

$$A^{-1} = \frac{1}{|A|}\widetilde{A}\,, \qquad \textbf{(12.17)}$$

где $\widetilde{A}$ — присоединенная матрица:

$$\widetilde{A}=\begin{pmatrix} A_{11}\ A_{12}\ldots A_{1n}\\ A_{21}\ A_{22}\ldots A_{2n}\\ \ldots\ \ldots\ \ldots\ \ldots\\ A_{n1}\ A_{n2}\ldots A_{nn}\end{pmatrix}'=\begin{pmatrix} A_{11}\ A_{21}\ldots A_{n1}\\ A_{21}\ A_{22}\ldots A_{n2}\\ \ldots\ \ldots\ \ldots\ \ldots\\ A_{1n}\ A_{n2}\ldots A_{nn}\end{pmatrix},$$

т. е. матрица, элементы которой есть алгебраические дополнения A_{ij} элементов матрицы A', транспонированной к A.

▶ **Пример 12.4**.

Дана матрица

$$A=\begin{vmatrix} 4 & 3 & 6\\ -1 & 0 & -3\\ 3 & -1 & 2\end{vmatrix}.$$

Найти A^{-1}.

Р е ш е н и е.

1. $|A|=-27$ (вычисляем по формуле **(12.10)** или **(12.11))**. Так как $|A|\neq 0$, то A^{-1} существует.

2. $A'=\begin{vmatrix} 4 & -1 & 3\\ 3 & 0 & -1\\ 6 & -3 & 2\end{vmatrix}$.

3. $\widetilde{A}=\begin{vmatrix} -3 & -12 & -9\\ -7 & -10 & 6\\ 1 & 13 & 3\end{vmatrix}$,

где элементы есть алгебраические дополнения A_{ij} элементов матрицы A', определяемые по (**12.12**).

4. По формуле (**12.17**)

$$A^{-1}=\frac{1}{-27}\begin{pmatrix} -3 & -12 & -9\\ -7 & -10 & 6\\ 1 & 13 & 3\end{pmatrix}=\begin{pmatrix} 1/9 & 4/9 & 1/3\\ 7/27 & 10/27 & -2/9\\ -1/27 & -13/27 & -1/9\end{pmatrix}.\ ▶$$

С в о й с т в а о б р а т н о й м а т р и ц ы:

1. $A^{-1}=\dfrac{1}{|A|}$.

2. $(A^{-1})^{-1}=A$.

3. $\left(A^m\right)^{-1}=\left(A^{-1}\right)^m$.

4. $\left(AB\right)^{-1}=B^{-1}A^{-1}$, где A, B — квадратные матрицы.

5. $\left(A^{-1}\right)'=\left(A'\right)^{-1}$. **(12.18)**

6. $\left(\mathrm{diag}\left(a_{11}\ a_{22.}\ldots a_{nn}\right)\right)^{-1}=\mathrm{diag}\left(a_{11}^{-1}\ a_{22}^{-1}\ldots a_{nn}^{-1}\right)$.

12.4. Ранг матрицы и линейная зависимость ее строк (столбцов)

***Рангом** матрицы A* (обозначается rang A или $r(A)$) *называется наивысший порядок ее миноров, отличных от нуля.*

▶ **Пример 12.5.**

Найти ранг матрицы

$$A=\begin{pmatrix}4 & 0 & 2 & 0\\ 2 & 0 & 1 & 0\\ 6 & 0 & 3 & 0\\ 8 & 0 & 4 & 0\end{pmatrix}.$$

Р е ш е н и е. Легко убедиться в том, что все миноры 2-го порядка равны нулю (например, $\begin{vmatrix}4 & 0\\ 2 & 0\end{vmatrix}=0$, $\begin{vmatrix}0 & 0\\ 0 & 0\end{vmatrix}=0$, $\begin{vmatrix}4 & 2\\ 6 & 3\end{vmatrix}=0$ и т. д.), а следовательно, равны нулю миноры 3-го порядка и минор 4-го порядка.

Так как среди миноров 1-го порядка есть не равные нулю (например, $|4|\neq 0$ и т. д.), то $r(A)=1$. ▶

С в о й с т в а р а н г а м а т р и ц ы:

1. $0\le r(A)\le \min(m,n)$.

2. $r(A)=0$ тогда и только тогда, когда A — нулевая матрица, т. е. $A=\mathbf{0}$.

3. Для квадратной матрицы A n-го порядка $r(A)=n$ тогда и только тогда, когда A — невырожденная матрица.

4. $r(A+B)\le r(A)+r(B)$.

5. $r(A+B)\ge \left|r(A)-r(B)\right|$.

6. $r(AB) \le \min\{r(A), r(B)\}$.

7. $r(AB) = r(A)$, если B — квадратная матрица n-го порядка ранга n.

8. $r(BA) = r(A)$, если B — квадратная матрица m-го порядка ранга m.

9. $r(AA') = r(A'A) = r(A)$, где AA' и $A'A$ — квадратные матрицы соответственно m-го и n-го порядков.

10. $r(AB) \ge r(A) + r(B) - n$, где n — число столбцов матрицы A или строк матрицы B.

Понятие ранга матрицы тесно связано с понятием линейной зависимости (независимости) ее строк (или столбцов).

Пусть строки матрицы A:

$$e_1 = (a_{11}\ a_{12} \ldots a_{1n}),\ \ e_2 = (a_{21}\ a_{22} \ldots a_{2n}),\ \ldots,\ e_m = (e_{m1}\ e_{m2} \ldots e_{mn}).$$

Строка e называется линейной комбинацией строк e_1, e_2..., e_m матрицы, если

$$e = \lambda_1 e_1 + \lambda_2 e_2 + \ldots + \lambda_m e_m, \qquad \textbf{(12.19)}$$

где $\lambda_1, \lambda_2, \ldots, \lambda_m$ — *какие-то* числа. (Равенство (**12.19**) понимается в смысле п о э л е м е н т н о г о сложения строк, т. е. равенству (**12.19**) соответствует n равенств для элементов строк с номерами 1,2,..., n.)

Строки матрицы e_1, e_2..., e_m *называются* ***линейно зависимыми****, если существуют такие числа* $\lambda_1, \lambda_2, \ldots, \lambda_m$, *из которых хотя бы одно отлично от нуля, что линейная комбинация строк равна нулевой строке*:

$$\lambda_1 e_1 + \lambda_2 e_2 + \ldots + \lambda_m e_m = \mathbf{0}, \qquad \textbf{(12.20)}$$

где $\mathbf{0}$ = (0 0... 0).

Если равенство (**12.20**) выполняется тогда и только тогда, когда $\lambda_1 = \lambda_2 = \ldots = \lambda_m = 0$, то строки матрицы называются ***линейно независимыми.***

Ранг матрицы равен максимальному числу ее линейно независимых строк или столбцов, через которые линейно выражаются все остальные ее строки (*столбцы*).

12.5. Система линейных уравнений

Система m линейных уравнений с n переменными имеет вид:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1j}x_j + \ldots + a_{1n}x_n = b_1; \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2j}x_j + \ldots + a_{2n}x_n = b_2; \\ \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \\ a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{ij}x_j + \ldots + a_{in}x_n = b_i; \\ \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mj}x_j + \ldots + a_{mn}x_n = b_m, \end{cases} \qquad \textbf{(12.21)}$$

или в краткой записи с помощью знаков суммирования:

$$\sum_{j=1}^{n} a_{ij}A_{ij} = b_i \ , \ i = 1,\ldots, \ m. \qquad \textbf{(12.22)}$$

В матричной форме система (**12.21**) имеет вид:

$$AX = B, \qquad \textbf{(12.23)}$$

где

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}; \ X = \begin{pmatrix} x_1 \\ x_2 \\ \ldots \\ x_n \end{pmatrix}; \ B = \begin{pmatrix} b_1 \\ b_2 \\ \ldots \\ b_m \end{pmatrix}.$$

Если число уравнений равно числу переменных, т.е. $m = n$, и *квадратная матрица A — невырожденная* ($|A| \neq 0$), *то система* (**12.21**) *имеет единственное решение*:

$$X = A^{-1}B. \qquad \textbf{(12.24)}$$

Это же решение, т. е. переменные x_1, x_2,..., x_j,..., x_n при $m = n$, могут быть найдены по *формулам Крамера*

$$x_j = \frac{\Delta_j}{\Delta} \ , \ j = 1,\ldots, \ n, \qquad \textbf{(12.25)}$$

где Δ — определитель матрицы A, т. е. $\Delta = |A|$;

Δ_j — определитель матрицы A_j, получаемой из матрицы A заменой j-го столбца столбцом свободных членов, т. е. $\Delta = |A_j|$.

▶ **Пример 12.6.** Решить методом обратной матрицы систему уравнений

$$\begin{cases} 4x_1 + 3x_2 + 6x_3 = 12; \\ -x_1 \qquad - 3x_3 = 0; \\ 3x_1 - x_2 + 2x_3 = 5. \end{cases}$$

Р е ш е н и е. Пусть

$$A = \begin{pmatrix} 4 & 3 & 6 \\ -1 & 0 & -3 \\ 3 & -1 & 2 \end{pmatrix}; \; X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}; \; B = \begin{pmatrix} 12 \\ 0 \\ 5 \end{pmatrix}.$$

Уравнение в матричной форме примет вид $AX=B$. Найдем обратную матрицу:

$$A^{-1} = -\frac{1}{27}\begin{pmatrix} -3 & -12 & -9 \\ -7 & -10 & 6 \\ 1 & 13 & 3 \end{pmatrix}$$ (см. пример **12.4**).

По формуле (**12.22**)

$$X = -\frac{1}{27}\begin{pmatrix} -3 & -12 & -9 \\ -7 & -10 & 6 \\ 1 & 13 & 3 \end{pmatrix}\begin{pmatrix} 12 \\ 0 \\ 5 \end{pmatrix} = -\frac{1}{27}\begin{pmatrix} -81 \\ -54 \\ 27 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \\ -1 \end{pmatrix},$$

т. е. $x_1=3$, $x_2=2$, $x_3=-1$.▶

Система линейных однородных уравнений, т. е. система $AX = \mathbf{0}$ с нулевыми свободными членами, *имеет ненулевое решение тогда и только тогда, когда матрица A — вырожденная*, т. е. $|A| = 0$.

12.6. Векторы

n-мерным вектором называется упорядоченная совокупность n действительных чисел, записываемых в виде $x = (x_1, x_2,..., x_n)$, где x_i — i-я компонента вектора x.

Равенство векторов

Векторы x и y равны, т. е. $x = y$, если $x_i = y_i$, $i = 1,..., n$.

Операции над векторами

1. *Умножение вектора на число*:

$$u = x\lambda, \text{ если } u_i = \lambda x_i,\ i = 1,..., n.$$

2. *Сложение двух векторов*:

$$z = x + y, \text{ если } z_i = x_i + y_i,\ i = 1,..., n.$$

Векторным (линейным) пространством *называется множество векторов (элементов) с действительными компонентами, в котором определены операции сложения векторов и умножения вектора на число, удовлетворяющие следующим* с в о й с т в а м :

1. $x + y = y + x$.
2. $(x+y)+z = x+(y+z)$.
3. $\alpha(\beta x)=(\alpha\beta)x$.
4. $\alpha(x+y) = \alpha x+\alpha y$.
5. $(\alpha+\beta)x = \alpha x+\beta x$.
6. Существует нулевой вектор $\mathbf{0}=(0\ 0...0)$ такой, что $x+\mathbf{0} = x$ для любого вектора x.
7. Для любого вектора x существует противоположный вектор $(-x)$ такой, что $x+(-x)=\mathbf{0}$.
8. $1 \cdot x = x$ для любого вектора x.

Понятие ***линейной комбинации, линейной зависимости и независимости векторов*** e_1, e_2,..., e_m аналогичны соответствующим понятиям для строк матрицы e_1, e_2,..., e_m (§ **12.5**).

Линейное пространство R^n *называется* ***n-мерным****, если в нем существует n линейно независимых векторов, а любые из* $(n+1)$ *векторов уже являются зависимыми.* Иначе, ***размерность пространства*** *— это максимальное число содержащихся в нем линейно независимых векторов*, т. е. $\dim(R^n) = n$.

Совокупность n линейно независимых векторов n-мерного пространства R^n *называется* ***базисом***.

Каждый вектор x линейного пространства R можно представить единственным способом в виде линейной комбинации векторов базиса:

$$x = x_1e_1+x_2e_2+...+x_ne_n, \qquad \textbf{(12.26)}$$

где x_1, x_2,..., x_n — координаты вектора x относительно базиса e_1, e_2,..., e_n.

Скалярным произведением *двух векторов* $x = (x_1, x_2,..., x_n)$ и $y = (y_1, y_2,..., y_n)$ *называется число*

$$(x, y) = x_1y_1+x_2y_2+...+x_ny_n=\sum_{i=1}^{n} x_iy_i\ . \qquad \textbf{(12.27)}$$

Евклидовым пространством *называется векторное* (*линейное*) *пространство, в котором задано скалярное произведение векторов, удовлетворяющее следующим свойствам:*

1. $(x, y) = (y, x)$.
2. $(x, y+z)=(x, y)+(x, z)$.
3. $(\alpha,x, y) = \alpha(x, y)$.
4. $(x, x) \geq 0$, если $x \neq \mathbf{0}$; $(x, x)=\mathbf{0}$, если $x = \mathbf{0}$.

Длиной (нормой) *вектора x в евклидовом пространстве называется корень квадратный из его скалярного квадрата*:

$$|x| = \sqrt{(x, x)} = \sqrt{x_1^2 + x_2^2 + ... + x_n^2}\ . \qquad \textbf{(12.28)}$$

Два вектора называются ***ортогональными****, если их скалярное произведение равно нулю.*

Векторы e_1, e_2,..., e_n *n-мерного евклидова пространства образуют* ***ортонормированный базис,*** *или* ***ортонормированную систему векторов****, если эти векторы попарно ортогональны и длина каждого из них равна* 1, т. е. если $(e_i, e_j) = 0$ при $i \neq j$ и $|e_i| = 1$, $i = 1,..., n$.

12.7. Собственные векторы и собственные значения квадратной матрицы

Вектор $x \neq \mathbf{0}$ *называется* ***собственным вектором квадратной матрицы*** *A*, если найдется такое число λ, что

$$Ax = \lambda x\ . \qquad \textbf{(12.29)}$$

Число λ *называется* ***собственным значением*** (или ***собственным числом***) матрицы *A*, *соответствующим вектору x.*

Собственный вектор *x* определен с точностью до коэффициента пропорциональности.

Для существования ненулевого решения ($x \neq \mathbf{0}$) уравнения (**12.29**) или равносильного ему уравнения

$$(A - \lambda E)x = \mathbf{0} \qquad \textbf{(12.30)}$$

необходимо и достаточно, чтобы определитель системы (**12.30**)

$$|A - \lambda E| = 0\ . \qquad \textbf{(12.31)}$$

Определитель $|A - \lambda E|$ называется *характеристическим многочленом матрицы A*, а уравнение (**12.31**) — ее *характеристическим уравнением.*

▶ **Пример 12.7.**

Найти собственные значения и собственные векторы матрицы

$$A=\begin{pmatrix}1 & 4\\ 9 & 1\end{pmatrix}.$$

Р е ш е н и е. Составим характеристическое уравнение (**12.31**)

$$\left|A-\lambda E\right|=\begin{vmatrix}1-\lambda & 4\\ 9 & 1-\lambda\end{vmatrix}=0,$$

или $\lambda^2-2\lambda-35=0$, откуда собственные значения матрицы A: $\lambda_1=-5$, $\lambda_2=7$.

При $\lambda_1=-5$ уравнение (**12.30**) примет вид:

$$\left(A-\lambda_1 E\right)\begin{pmatrix}x_1\\ x_2\end{pmatrix}=\mathbf{0},$$

или $\begin{pmatrix}6 & 4\\ 9 & 6\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix}=\begin{pmatrix}0\\ 0\end{pmatrix}$, откуда $x_2=-1{,}5x_1$. Положив $x_1=c$, получим, что векторы $x^{(1)}=(c;-1{,}5c)$ при любом $c\neq 0$ являются собственными векторами матрицы A с собственным значением $\lambda_1=-5$.

Аналогично можно показать, что векторы $x^{(2)}=\left(\frac{2}{3}c_1;c_1\right)$ при любом $c_1\neq 0$ являются собственными векторами матрицы A с собственным значением $\lambda_2=7$. ▶

Разным собственным значениям матрицы соответствуют линейно независимые собственные векторы.

12.8. Симметрические, положительно определенные, ортогональные и идемпотентные матрицы

Квадратная матрица A называется ***симметрической*** **(*симметричной*)**, если $A'=A$, т. е. $a_{ij}=a_{ji}$, $i=1,\ldots,n$; $j=1,\ldots,n$.

Симметрическая матрица A n-го порядка называется ***положительно (неотрицательно) определенной****, если для любого ненулевого вектора* $x=(x_1, x_2,\ldots, x_n)'$ *выполняется неравенство*

$$x'Ax>0 \quad (x'Ax\geq 0). \qquad \textbf{(12.32)}$$

Например, матрица $A'A$ неотрицательно определена, так как для любого вектора xx' $(A'A)x = (x'A')$ $Ax = (Ax)'Ax = y'y \geq 0$, ибо $y'y$ представляет скалярный квадрат вектора $y = Ax$.

Понятие положительно (неотрицательно) определенной симметрической матрицы A тесно связано с понятием *положительно определенной* (*полуопределенной*) *квадратичной формы*.

$$L(x_1, x_2, ..., x_n) = x'Ax = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_i x_j \left(L > 0 \text{ или } L \geq 0\right).$$

Для положительно (неотрицательно) определенных матриц используется запись $A > 0 (A \geq 0)$.

Соотношение $A > B$ $(A \geq B)$ означает, что матрица $A - B$ положительно (неотрицательно) определена.

Свойства положительно (неотрицательно) определенных матриц.

1. Если $A > B$, то $a_{ii} > b_{ii}$, $i = 1,..., n$, т. е. диагональные элементы матрицы A более соответствующих диагональных элементов матрицы B.

2. Если $A > B$, $C \geq 0$, то $A + C > B$.

3. Если $A > B$, где A и B — невырожденные матрицы, то $B^{-1} > A^{-1}$.

4. Если $A > \mathbf{0}$ $(A \geq \mathbf{0})$, то все собственные значения матрицы A положительны (неотрицательны), т. е. $\lambda_i > 0$ $(\lambda_i \geq 0)$, $i = 1,..., n$.

Свойства симметрической положительно определенной матрицы A n-го порядка.

1. Если $n \geq m$, rang $(B_{n\times m}) = m$, то $B'AB$ — положительно определенная матрица.

2. Матрица A^{-1}, обратная к A, также симметрическая и положительно определенная.

3. Определитель $|A| > 0$, а значит, и все *главные миноры* матрицы A (получаемые для подматриц, образованных из матрицы A вычеркиванием строк и столбцов с одинаковыми номерами) положительны.

4. След матрицы A равен сумме ее собственных значений:

$$\mathrm{tr}(A) = \lambda_1 + \lambda_2 + ... + \lambda_n = \sum_{i=1}^{n} \lambda_i . \qquad \textbf{(12.33)}$$

Квадратная матрица C называется ***ортогональной****, если*

$$C^{-1} = C' . \qquad \textbf{(12.34)}$$

Свойства ортогональной матрицы C:

1. $C'C=E$.

2. Определитель $|C|=1$ или $|C|=-1$.

3. В ортогональной матрице как строки, так и столбцы образуют ортонормированную систему векторов (§ **12.6**).

4. С помощью ортогональной матрицы C симметричная матрица A может быть приведена к диагональному виду

$$C'AC = \Lambda, \qquad \textbf{(12.35)}$$

где $\Lambda=\text{diag}\ (\lambda_1,\lambda_2,\ldots,\lambda_n)$;

$\lambda_1,\lambda_2,\ldots,\lambda_n$ — собственные значения матрицы A.

5. Симметричная матрица A может быть представлена через ортогональную и диагональную матрицу в виде

$$A = C\Lambda C', \qquad \textbf{(12.36)}$$

где $\Lambda = \text{diag}\ (\lambda_1,\lambda_2,\ldots,\lambda_n)$;

$\lambda_1,\lambda_2,\ldots,\lambda_n$ — собственные значения матрицы A.

Симметрическая[1] *матрица A называется* ***идемпотентной****, если она совпадает со своим квадратом*, т. е.

$$A = A^2. \qquad \textbf{(12.37)}$$

Свойства идемпотентных матриц:

1. $A^k=A$, где k — натуральное число.

2. Собственные значения *идемпотентной* матрицы A равны либо нулю, либо единице: $\lambda= 0$ или $\lambda = 1$.

3. Все идемпотентные матрицы неотрицательно определены.

4. Ранг идемпотентной матрицы равен ее следу, т. е. числу ненулевых собственных значений.

12.9. Блочные матрицы. Произведение Кронекера

В ряде случаев исходную матрицу $A_{m\times n}$ удобно разбить на *блоки*, например,

$$A=\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \qquad \textbf{(12.38)}$$

где A_{11} — $(m_1\times n)$-матрица; A_{12} — $(m_1\times n_2)$-матрица; A_{21} — $(m_2\times n_1)$-матрица, A_{22}— $(m_2\times n_2)$-матрица.

---

[1] Вообще говоря, требование симметричности матрицы A не строго обязательно для определения идемпотентной матрицы, но именно симметрические идемпотентные матрицы встречаются в эконометрике.

В этом случае сама матрица A называется ***блочной***.

Операции сложения и умножения блочных матриц проводятся по правилам соответствующих операций над матрицами, если заменить их элементы блоками:

$$A+B=\begin{pmatrix} A_{11}\ A_{12} \\ A_{21}\ A_{22} \end{pmatrix}+\begin{pmatrix} B_{11}\ B_{12} \\ B_{21}\ B_{22} \end{pmatrix}=\begin{pmatrix} A_{11}+B_{11} & A_{12}+\ B_{12} \\ A_{21}+B_{21} & A_{22}+\ B_{22} \end{pmatrix};$$

$$AB=\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}\begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}=\begin{pmatrix} A_{11}B_{11}+A_{12}B_{21} & A_{11}B_{12}\ +A_{12}B_{22} \\ A_{21}B_{11}+A_{22}B_{21} & A_{21}B_{12}+A_{22}B_{22} \end{pmatrix}.$$

Пусть матрица A разбита на блоки такие, что A_{11} и A_{22} — квадратные матрицы.

Тогда определитель матрицы A:

$$|A|=\begin{vmatrix} A_{11}\ A_{12} \\ A_{21}\ A_{22} \end{vmatrix}=|A_{11}|\,|A_{22}-A_{21}A_{11}^{-1}A_{12}|=|A_{22}|\,|A_{11}-A_{12}A_{22}^{-1}A_{21}|\,.$$

В частности, для *блочно-диагональной матрицы* A:

$$|A|=\begin{vmatrix} A_{11} & \mathbf{0} \\ \mathbf{0} & A_{22} \end{vmatrix}=|A_{11}|\,|A_{22}|\,. \qquad \textbf{(12.39)}$$

Матрица, *обратная блочно-диагональной*:

$$A^{-1}=\begin{pmatrix} A_{11} & \mathbf{0} \\ \mathbf{0} & A_{22} \end{pmatrix}^{-1}=\begin{pmatrix} A_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & A_{22}^{-1} \end{pmatrix}. \qquad \textbf{(12.40)}$$

Произведением Кронекера *двух матриц* $A_{m\times n}$ и $B_{k\times l}$ *называется блочная матрица* $A\otimes B$ *размера* $km\times ln$:

$$A\otimes B=\begin{pmatrix} a_{11}B & a_{12}B & \ldots & a_{1n}B \\ a_{21}B & a_{22}B & \ldots & a_{2n}B \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1}B & a_{m2}B & \ldots & a_{mn}B \end{pmatrix}. \qquad \textbf{(12.41)}$$

▶ **Пример 12.8**.

Дано $A=\begin{pmatrix}3 & 4\\ 2 & 0\end{pmatrix}$, $B=\begin{pmatrix}1 & 0\\ 6 & 7\end{pmatrix}$.

Найти $A\otimes B$.

Р е ш е н и е .

$$A\otimes B=\begin{pmatrix}3\begin{pmatrix}1 & 0\\ 6 & 7\end{pmatrix} & 4\begin{pmatrix}1 & 0\\ 6 & 7\end{pmatrix}\\ 2\begin{pmatrix}1 & 0\\ 6 & 7\end{pmatrix} & 0\begin{pmatrix}1 & 0\\ 6 & 7\end{pmatrix}\end{pmatrix}=\begin{pmatrix}3 & 0 & 4 & 0\\ 18 & 21 & 24 & 28\\ 2 & 0 & 0 & 0\\ 12 & 14 & 0 & 0\end{pmatrix}.$$ ▶

С в о й с т в а п р о и з в е д е н и я К р о н е к е р а:

1. Если A и B — невырожденные матрицы, то

$$\left|A\otimes B\right|^{-1}=A^{-1}\otimes B^{-1}.$$

2. Если A и B — квадратные матрицы соответственно m-го и n-го порядков, то $\left|A\otimes B\right|=\left|A\right|^{m}\left|B\right|^{n}$.

3. $\left(A\otimes B\right)'=A'\otimes B'$.

4. $\mathrm{tr}\left(A\otimes B\right)=\mathrm{tr}\left(A\right)\cdot\mathrm{tr}\left(B\right)$.

12.10. Матричное дифференцирование

Производной скалярной функции $\varphi(x)$ *от векторного аргумента* (*вектора-столбца*) $x=(x_1, x_2,...,x_n)'$ *называется вектор* (*вектор-строка*)

$$\frac{\partial\varphi(x)}{\partial x'}=\left(\frac{\partial\varphi(x)}{\partial x_1},\frac{\partial\varphi(x)}{\partial x_2},...,\frac{\partial\varphi(x)}{\partial x_n}\right). \qquad \textbf{(12.42)}$$

Производной векторной $(m\times 1)$ ***функции*** $f(x)$ *от векторного* $(n\times 1)$ *аргумента* $x=(x_1, x_2,..., x_n)'$ *называется* mn*-матрица:*

$$\frac{\partial f(x)}{\partial x'}=\begin{pmatrix}\frac{\partial f_1(x)}{\partial x_1} & \frac{\partial f_1(x)}{\partial x_2} & ... & \frac{\partial f_1(x)}{\partial x_n}\\ ..... & & & \\ \frac{\partial f_m(x)}{\partial x_1} & \frac{\partial f_m(x)}{\partial x_2} & ... & \frac{\partial f_m(x)}{\partial x_n}\end{pmatrix} \qquad \textbf{(12.43)}$$

(*матрица Якоби*, определитель которой — *якобиан*).

Определение (**12.42**) является частным случаем (**12.43**) при $m = 1$.

Ч а с т н ы е с л у ч а и:

1. Если $\varphi(x) = a'x$, где $a = (a_1, a_2, ..., a_n)'$ и $x = (x_1, x_2, ..., x_n)'$ — векторы-столбцы, то

$$\frac{\partial \varphi(x)}{\partial x'} = \frac{\partial (a'x)}{\partial x'} = a'.$$

2. Если $\varphi(x) = x'Ax$, где A — симметрическая квадратная матрица n-го порядка, то

$$\frac{\partial \varphi(x)}{\partial x'} = \frac{\partial (x'Ax)}{\partial x'} = 2x'A.$$

3. Если $f(x) = Ax$, где A — mn-матрица, то

$$\frac{\partial f(x)}{\partial x'} = \frac{\partial (Ax)}{\partial x'} = A.$$

Упражнения

12.9. Вычислить матрицу $D = (AB)' - C^2$, где

$$A = \begin{pmatrix} 3 & 4 & 2 \\ 1 & 0 & 5 \end{pmatrix}, \quad B = \begin{pmatrix} 2 & 0 \\ 1 & 3 \\ 0 & 5 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 3 \\ 0 & 4 \end{pmatrix}.$$

12.10. Вычислить произведение и найти след матриц AB и BA, если:

$$A = \begin{pmatrix} 2 & -3 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 4 \\ 3 \\ 1 \end{pmatrix}.$$

12.11. Вычислить определители матриц:

а) $\begin{vmatrix} 1 & 1 & 1 \\ 2 & -3 & 1 \\ 4 & -1 & -5 \end{vmatrix}$; *б)* $\begin{vmatrix} 0 & 1 & 2 & 3 \\ 1 & 0 & 1 & 2 \\ 2 & 1 & 0 & 1 \\ 3 & 2 & 1 & 0 \end{vmatrix}$.

12.12. Найти матрицу, обратную данной:

а) $\begin{pmatrix} 5 & -4 \\ -8 & 6 \end{pmatrix}$; *б*) $\begin{pmatrix} 4 & -8 & -5 \\ -4 & 7 & -1 \\ -3 & 5 & 1 \end{pmatrix}$.

12.13. Найти ранг матрицы:

а) $\begin{pmatrix} 1 & 2 & 1 & 4 \\ 0 & 5 & -1 & 4 \\ -1 & 3 & 4 & 6 \end{pmatrix}$; *б*) $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \end{pmatrix}$.

12.14. Решить систему уравнений:

а) $\begin{cases} x_1 + 2x_2 + x_3 = 8, \\ -2x_1 + 3x_2 - 3x_3 = -5, \\ 3x_1 - 4x_2 + 5x_3 = 10; \end{cases}$ *б*) $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ $X = \begin{pmatrix} 1 & 0 & 7 \\ 8 & 1 & 2 \end{pmatrix}$.

12.15. Решить матричное уравнение $AXB = C$, если

$$A = \begin{pmatrix} 2 & 3 \\ 5 & 8 \end{pmatrix}, \quad B = \begin{pmatrix} 5 & 4 & 1 \\ 1 & 1 & 7 \\ 6 & 5 & 9 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}.$$

12.16. Являются ли линейно зависимыми векторы a_1=(2; — 1;3), a_2=(1;4; −1), a_3=(0; −9;5)?

12.17. Найти собственные значения и собственные векторы

а) $\begin{pmatrix} 2 & 4 \\ -1 & -3 \end{pmatrix}$, *б*) $\begin{pmatrix} 1 & 2 & -2 \\ 1 & 0 & 3 \\ 1 & 3 & 0 \end{pmatrix}$.

12.18. Найти собственные значения и собственные векторы матрицы

$$A = (1 - \alpha)E_n + \alpha SS',$$

где $S = (1,1,...,1)' - n \times 1$-вектор.

Известно, что матрица A идемпотентная. Убедиться в том, что матрица $B = E - A$ также идемпотентная и $BA = \mathbf{0}$.

Глава 13

Эконометрические компьютерные пакеты

13.1. Оценивание модели с помощью компьютерных программ

На практике совокупность выборочных данных содержит большое количество наблюдений (иногда несколько тысяч). Разумеется, «вручную» производить вычисления такого объема невозможно.

Компьютерные эконометрические (их также называют регрессионными) пакеты содержат в качестве готовых подпрограмм основные, наиболее часто встречающиеся вычисления, которые приходится выполнять исследователю.

В последние годы появились десятки эконометрических пакетов[1], в том числе и разработанные в России. В настоящем учебнике мы рассмотрим в качестве примера пакет «*Econometric Views*», который используется, например, для обучения в Российской Экономической Школе.

Перечислим основные возможности, предоставляемые пакетом «*Econometric Views*».

1. Анализ данных.

Первым этапом эконометрического моделирования является анализ экспериментальных (выборочных) данных, которые предварительно должны быть введены в рабочий файл («*Workfile*»). Там они хранятся в виде последовательностей чисел, причем каждая выборка имеет свое имя.

Имеется возможность представить данные графически, получить значения основных количественных характеристик выборочных данных.

В качестве примера рассмотрим моделирование формирования цены на автомобили на вторичном рынке в зависимости от года выпуска машины и ее пробега. Пусть X_1 — срок эксплуата-

[1] О возможностях некоторых из них см., например, [18], [22].

ции автомобиля — разность между текущим годом и годом выпуска автомобиля, X_2 — пробег (в тыс. км), Y — цена (в у.е.). Рабочий файл пакета содержит основные числовые данные — выборочные значения всех перечисленных переменных. Чтобы получить экспериментальные данные в виде столбцов чисел, надо выполнить команду «*Open*». Экспериментальные данные могут быть получены на экране в виде столбцов:

Obs	X_1	X_2	Y
1	6,000000	91,00000	14800,00
2	7,000000	116,0000	14200,00
3	2,000000	34,00000	17000,00
.....			
48	5,000000	90,00000	15300,00
49	14,00000	217,0000	10400,00
50	4,000000	64,00000	15800,00

Графическое изображение данных представлено на рис. 13.1.

При желании можно получить все графики на одном рисунке.

Регрессионные программы допускают преобразования экспериментальных данных. Так, например, имеется команда, позволяющая упорядочить выборочные наблюдения по возрастанию какой-либо переменной. Обычно это имеет смысл в случае пространственной выборки. Также имеется возможность «редактировать» экспериментальные данные: например, исследователь может удалить слишком «нетипичные» наблюдения, которые могут неадекватно исказить модель.

Более подробное представление о распределении выборочных данных дают гистограммы.

В «*Econometric Views*» имеются команды, с помощью которых сразу получаются основные количественные характеристики выборочных распределений. Вот как выглядит соответствующий результат для нашего примера:

	X_1	X_2	Y
Mean	7,560000	117,9400	13918,00
Median	6,500000	103,5000	14450,00
Maximum	24,00000	373,0000	17500,00
Minimum	1,000000	15,00000	5000,000
Std. Dev.	5,107657	78,26470	2783,530
Skewness	1,127798	1,157630	−1,123340
Kurtosis	3,994277	4,115747	3,973437
Jarque-Bera	12,65896	13,76109	12,48990
Probability	0,001783	0,001028	0,001940
Observations	50	50	50

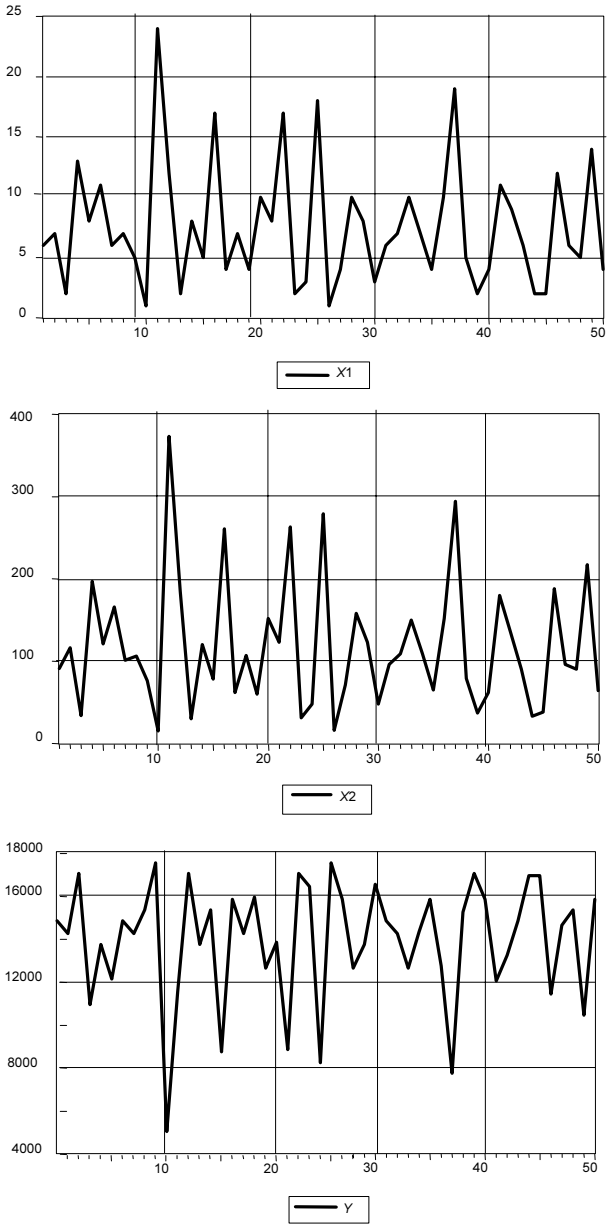

25
20
15
10
5
0
10
20
30
40
50
X1
400
300
200
100
0
10
20
30
40
50
X2
18000
16000
12000
8000
4000
10
20
30
40
50
Y

Рис. 13.1

Выдается математическое ожидание (*mean*), максимум, минимум и медиана распределения, стандартное отклонение (*Std. Dev.*), асимметрия (*Skewness*), эксцесс (*Kurtosis*). Статистика *Jarque-Bera* — это статистика проверки гипотезы о том, что соответствующая выборка взята из нормально распределенной совокупности. Ниже приводится соответствующее значение вероятности (*Probability*). В строке *Observations* (наблюдения) указывается число наблюдений n,

2. Оценивание модели.

После того, как анализ данных завершен, можно приступать к собственно моделированию. Допустим, мы хотим оценить зависимость Y от X_1 и X_2 с помощью метода наименьших квадратов. Для этого в «*Econometric Views*» имеется специальная команда, результат выполнения которой имеет следующий вид:

LS // Dependent Variable is Y
Sample: 1 50
Included observations: 50

Variable	Coefficient	Std. Error	*t*-statistic	Prob.
C	18046,09	14,98877	1203,974	0,0000
X_1	−483,3643	32,93229	−14,67752	0,0000
X_2	−4,017735	2,149205	−1,869406	0,0678

R-squared	0,999614	Mean dependent var	13918,00
Adjusted *R*-squared	0,999598	S. D. dependent var	2783,530
S.E. of regression	55,83376	Akaike info criterion	8,102882
Sum squared resid	146518,2	Schwarz criterion	8,217603
Log likelihood	−270,5190	*F*-statistic	60869,03
Durbin—Watson stat.	1,924701	Prob. (*F*-statistic)	0,000000

В столбце «*Coefficient*» находятся значения оценок соответствующих параметров регрессии, в столбце «*Std. Error*» — их стандартные отклонения, в столбце «*t-statistic*» — значения *t-критерия* Стьюдента при проверке гипотезы о незначимости соответствующих регрессоров. В последнем столбце приводится вероятность $P\,(t > t_{\text{набл}})$.

Ниже основной таблицы помещаются некоторые количественные характеристики регрессионной модели: коэффициент детерминации R^2 (*R-squared*), скорректированный коэффициент

детерминации $\hat{R}^2$ (*Adjusted R-squared*), стандартная ошибка регрессии — оценка σ в предположении, что выполняются условия классической модели Гаусса—Маркова (*S.E. of regression*), сумма квадратов остатков (*sum squared resid*), логарифм функции правдоподобия (*log likelihood*).

Регрессионная программа имеет дело с набором чисел, она не может понимать их природу. Поэтому при использовании метода наименьших квадратов выдаются результаты, имеющие смысл, вообще говоря, только для временных рядов. Таковы статистика Дарбина—Уотсона (*Durbin—Watson stat.*) и еще некоторые характеристики, которые мы не рассматривали в данном учебнике.

В конце выдается значение *F*-статистики при проверке гипотезы о незначимости регрессии в целом и значение соответствующей вероятности.

Как видно, на пятипроцентном уровне значимости регрессор X_2 (пробег автомобиля) оказался незначимым (хотя он значим на десятипроцентном уровне, ибо $t_{0,95;\ 48} = 2{,}01$; $t_{0,9;\ 48} = 1{,}67$). Причиной этого может быть высокая коррелированность между двумя величинами X_1 и X_2 и малый объем выборки.

Обычный метод наименьших квадратов является наиболее распространенным, но, как известно, далеко не всегда наилучшим методом оценивания. Регрессионная программа позволяет выбрать метод наиболее отвечающий характеру экспериментальных данных и их взаимозависимости. При этом в одном меню на выбор предлагаются как методы, специфические, как правило, для пространственной выборки (например, взвешенный метод наименьших квадратов), так и применимые исключительно для временных рядов — например, *ARMA*. Отметим еще раз, что программа не различает характера экспериментальных данных, и ее неосознанное использование может привести к абсолютно бессмысленному результату.

Для оценивания систем регрессионных уравнений предлагается отдельное меню, в которое входят обычный метод наименьших квадратов, двухшаговый и трехшаговый методы наименьших квадратов, а также метод одновременного оценивания уравнений как внешне не связанных. При выборе двухшагового или трехшагового метода программа запросит также ввести имена инструментальных переменных.

3. Анализ модели.

После получения адекватного регрессионного уравнения, удовлетворяющего исследователя на данный момент, наступает этап анализа модели. Обычно при этом проверяются некоторые предположения — гипотезы, описанные в основном тексте пособия. При этом применяются основные тесты — такие, как тест Уайта, тест Чоу, линейный тест и другие. Большинство из них (что также было отмечено в основном тексте) выполняются специальной командой.

Вот как выглядит, например, применение теста Уайта для рассматриваемого примера формирования цены автомобиля:

White Heteroskedasticity Test:

F-statistic	0,146329	Probability	0.963694
Obs**R*-squared	0,642001	Probability	0.958284

Test Equation:
LS // Dependent Variable is RESID^2
Sample: 1 50
Included observations: 50

Variable	Coefficient	Std. Error	t-statistic	Prob.
C	2027,021	2099,532	0,965463	0,3395
X_1	−3481,328	6934,929	−0,501999	0,6181
$X_{1\wedge 2}$	192,2855	449,3960	0,427875	0,6708
X_2	238,8914	451,6586	0,528920	0,5995
$X_{2\wedge 2}$	0,846622	1,871088	−0,452476	0,6531

R-squared	0,012840	Mean dependent var	930,364
Adjusted *R*-squared	−0,074908	S.D. dependent var	4618,038
S.E. of regression	4787,878	Akaike info criterion	17,04232
Sum squared resid	1,03E+09	Schwarz criterion	17,23353
Log likelihood	−492,0050	*F*-statistic	0,146329
Durbin—Watson stat.	2,006108	Prob. (*F*-statistic)	0,963694

Как видно, низкое значение *F*-статистики и соответствующее высокое значение вероятности позволяет принять гипотезу о гомоскедастичности.

Приведем также пример проверки гипотезы о выполнении некоторого условия, задаваемого линейным равенством.

Некоторые дилеры считают, что год «жизни» не слишком дорогого автомобиля уменьшает его стоимость примерно на 500 долларов. Проверим гипотезу о том, что истинное значение (по абсолютной величине) параметра β_1 равно 500 (разумеется, значение оценки будет при каждом выборочном наблюдении несколько другим).

Вот как выглядит тестирование гипотезы в компьютерном пакете «*Econometric Views*».

Null Hypothesis: $C(2)=-500$

F-statistic	0,255174	Probability	0,615816
Chi-square	0,255174	Probability	0,613455

Гипотезу о выполнении линейного условия можно проверять с помощью *F*-теста и χ^2-теста. Компьютерная программа приводит значения соответствующих статистик и значения соответствующих вероятностей. Как видно, результаты тестирования позволяют принять соответствующую гипотезу.

13.2. Метод Монте-Карло

Эксперимент по методу Монте-Карло — это эксперимент, основанный на компьютерном моделировании случайных величин.

Регрессионные программы (в частности, «*Econometric Views*») позволяют генерировать выборки равномерно распределенной случайной величины, а также величины, распределенной нормально с произвольным математическим ожиданием и дисперсией. Так как основные распределения (χ^2, *t*, *F*) определяются через нормальное, то компьютерные программы позволяют генерировать выборки и этих распределений.

Суть метода Монте-Карло заключается в том, что с помощью компьютера можно многократно наблюдать случайную величину с заранее известным распределением. Это позволяет получить (или проверить) статистические результаты экспериментально.

Метод широко применим не только в эконометрическом моделировании, но и вообще в статистическом исследовании. Так, с его помощью можно оценивать вероятности событий, связанных со случайными величинами.

Приведем простейший пример. Пусть *X* имеет равномерное распределение на отрезке [−1; 1], *Y* — нормальное распределе-

ние с математическим ожиданием, равным нулю, и дисперсией, равной единице. Требуется оценить вероятность того, что случайная величина $Z=XY^2$ примет значение на отрезке [0;1].

Регрессионная программа позволит сгенерировать серию выборочных наблюдений величины произвольного, заранее выбранного объема. Вот как, например, выглядит гистограмма распределения Z, полученная с помощью программы «*Econometric Views*»:

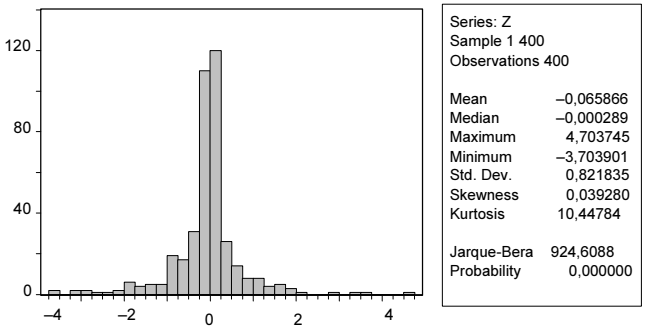


Рис. 13.2

В серии из четырехсот наблюдений величина Z приняла значение на интервале [0; 1] 173 раза. Таким образом, экспериментальной оценкой вероятности можно считать 173/400 = 0,4325.

В эконометрическом моделировании значение метода Монте-Карло особенно велико. С его помощью можно построить модель с заранее известными параметрами (отметим еще раз, что в реальных моделях параметры никогда не бывают известны).

Метод Монте-Карло позволяет проверить экспериментально результаты, полученные теоретически. В качестве примера рассмотрим задачу выбора спецификации модели. Пусть имеются фиксированные выборки переменных X, Z, а случайные выборки переменной Y генерируются по формуле

$$Y = 12 + 8X + Z + \varepsilon,$$

где ε — нормально распределенная случайная величина с математическим ожиданием, равным нулю, и стандартным отклонением, равным пяти.

Методом наименьших квадратов оцениваются модели, включающие переменную Z и не включающие ее (см. (**10.1**) и (**10.2**)).

Эксперимент повторяется четыре раза. Получаются следующие регрессионные уравнения:

$$\hat{y} = 12{,}3 + 7{,}8x + 0{,}7z; \qquad \hat{y} = 12{,}1 + 7{,}9x;$$
$$(0{,}09) \qquad (0{,}08)$$
$$\hat{y} = 11{,}6 + 8{,}1x + 1{,}2z; \qquad \hat{y} = 11{,}8 + 8{,}15x;$$
$$(0{,}09) \qquad (0{,}08)$$
$$\hat{y} = 11{,}4 + 8{,}08x + 1{,}3z; \qquad \hat{y} = 11{,}8 + 8{,}12x;$$
$$(0{,}09) \qquad (0{,}08)$$
$$\hat{y} = 12{,}8 + 7{,}93x + 0{,}8z; \qquad \hat{y} = 12{,}1 + 7{,}99x.$$
$$(0{,}09) \qquad (0{,}08)$$

Как видно, оценка параметра при переменной x в короткой модели оказывается смещенной вверх. В то же время точность оценок близка в обоих случаях.

Разумеется, в реальной практике эксперимент должен повторяться не четыре раза, а значительно большее количество раз.

С помощью метода Монте-Карло можно наглядно демонстрировать результаты применения тестов, а также экспериментально оценить последствия нарушения тех или иных условий.

Наконец, отметим особенно значимую роль экспериментов по методу Монте-Карло в процессе *обучения*. Именно с помощью таких экспериментов можно увидеть различия между методами оценивания моделей, наблюдать эффекты, вызванные нарушением тех или иных условий и т. д.

Упражнения

13.1. С помощью метода Монте-Карло построить следующие величины:

$$x_i = 10 + 0{,}5i, \qquad y_i = 2 + 3x_i + i\xi_i,$$

где ξ — нормально распределенная случайная величина с математическим ожиданием, равным нулю, и дисперсией, равной единице (белый шум). Оценить модель

$$y_i = \alpha + \beta x_i + \varepsilon_i.$$

С помощью теста Уайта убедиться в наличии гетероскедастичности.

Сравнить оценки обычного и взвешенного метода наименьших квадратов при регрессоре X. Повторить эксперимент несколько раз. Убедиться, что оценки взвешенного метода наименьших квадратов в основном ближе к значению 3.

13.2. Методом Монте-Карло сгенерировать следующие временные ряды:

$$X_1=2+\varepsilon_1;$$

$$X_2=1+2\varepsilon_2;$$

$$Y_1=0{,}5+1{,}2X_1-0{,}3X_2+3\varepsilon_3; \qquad \textbf{(13.1)}$$

$$Y_2=1{,}5+2{,}0X_1+0{,}7X_2+0{,}5\varepsilon_4, \qquad \textbf{(13.2)}$$

где ε_i — нормально распределенные случайные величины с нулевым математическим ожиданием и единичной дисперсией.

Рассмотреть систему одновременных уравнений:

$$\begin{cases} Y_1 = \alpha_1 + \beta_1 X_1 + \gamma_1 Y_1, \\ Y_2 = \alpha_2 + \beta_2 X_2 + \gamma_2 Y_2. \end{cases}$$

Применить к уравнениям системы обычный и косвенный методы наименьших квадратов. Сравнить полученные оценки. Сравнить полученные регрессионные уравнения с модельными **(13.1)**−**(13.2)**. Повторить эксперимент несколько раз.

Литература

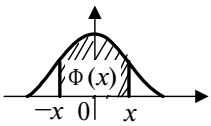
1. *Айвазян С.А.* Основы эконометрики. Т. 2. — М.: ЮНИТИ-ДАНА, 2001.
2. *Айвазян С.А.*, *Мхитарян В.С.* Прикладная статистика и основы эконометрики. — М.: ЮНИТИ, 1998.
3. *Бернд Э.Р.* Приактика эконометрики. Классика и современность / Пер. с англ. под ред. С.А. Айвазяна. — М.: ЮНИТИ-ДАНА, 2005.
4. *Высшая* математика для экономистов / Под ред. Н.Ш. Кремера. — М.: ЮНИТИ-ДАНА, 2008.
5. *Высшая* математика для экономических специальностей / Под ред. Н.Ш. Кремера. — М.: Высшее образование, 2009.
6. *Головина Л.И.* Линейная алгебра и некоторые ее приложения. — М.: Наука, 1985.
7. *Дайитбегов Д.М.* Компьютерные технологии анализа данных в эконометрике. — М.: Инфра-М — Вузовский учебник, 2008.
8. *Джонстон Дж.* Эконометрические методы: Пер. с англ. — М.: Статистика, 1980.
9. *Доугерти К.* Введение в эконометрику: Пер. с англ. — М.: Инфра-М, 1997.
10. *Дрейпер И.*, *Смит Г.* Прикладной регрессионный анализ: Пер. с англ. — Кн. 1, 2. — М.: Финансы и статистика, 1986, 1987.
11. *Дубров А.М.*, *Мхитарян В.С.*, *Трошин Л.И.* Многомерные статистические методы. — М.: Финансы и статистика, 1998.
12. *Канторович Г.Г.* Эконометрика //Методические материалы по экономическим дисциплинам для преподавателей средних школ и вузов. Экономическая статистика. Эконометрика. Программы, тесты, задачи, решения /Под ред. Л.С. Гребнева. — М.: ГУ-ВШЭ, 2000.
13. *Карасев А.И.*, *Кремер Н.Ш.*, *Савельева Т.И.* Математические методы и модели в планировании. — М.: Экономика, 1987.
14. *Крамер Г.* Математические методы статистики: Пер. с англ. — М.: Мир, 1975.
15. *Кремер Н.Ш.* Математическая статистика. — М.: Экономическое образование, 1992.
16. *Кремер Н.Ш., Путко Б.А., Тришин И.М.* Математика для экономистов: от Арифметики до Эконометрики / Под ред. Н.Ш. Кремера. — М.: Высшее образование, 2009.

17. *Кремер Н.Ш.* Теория вероятностей и математическая статистика. — М.: ЮНИТИ-ДАНА, 2007.
18. *Магнус Я.Р.*, *Катышев Л.К.*, *Пересецкий А.А.* Эконометрика. Начальный курс. — М.: Дело, 2004.
19. *Магнус Я.Р.*, *Нейдеккер Х.* Матричное диференциальное исчисление с приложениями к статистике и эконометрике. — М.: Физматлит, 2002.
20. *Маленво Э.* Статистические методы эконометрии. — М.: Статистика, 1975—1976. — Вып. 1, 2.
21. *Тихомиров Н.П., Дорохина Е.Ю.* Эконометрика. — М.: Экзамен, 2003.
22. *Тюрин Ю.Н.*, *Макаров А.А.* Статистический анализ данных на компьютере / Под ред. В.Э. Фигурнова. — М.: Инфра-М, 2003.
23. *Уотшем Т. Дж., Паррамоу К.* Количественные методы в финансах: Пер. с англ. — М.: ЮНИТИ, 1999.
24. *Ферстер Э.*, *Ренц Б.* Методы корреляционного и регрессионного анализа: Пер. с нем. — М.: Финансы и статистика, 1982.
25. *Четыркин Е.М.*, *Калихман И.Л.* Вероятность и статистика. — М.: Финансы и статистика, 1982.
26. *Эконометрика* / Под ред. В.С. Мхитаряна. — М.: Проспект, 2008.
27. *Эконометрика* /Под ред. Н.И. Елисеевой. — М.: Финансы и статистика, 2001.
28. *Экономико*-математические методы и прикладные модели / Под ред. В.В. Федосеева. — М.: ЮНИТИ, 1999.

Математико-статистические таблицы

Таблица I

Значения функции Лапласа $\Phi(x)=\frac{2}{\sqrt{2\pi}}\int_0^x e^{-t^2/2}\,dt$



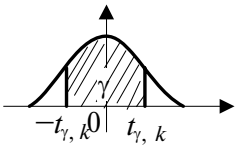
Целые и десятые доли x	Сотые доли x									
	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0080	0,0160	0,0239	0,0319	0,0399	0,0478	0,0558	0,0638	0,0717
0,1	0,0797	0,0876	0,0955	0,1034	0,1113	0,1192	0,1271	0,1350	0,1428	0,1507
0,2	0,1585	0,1633	0,1741	0,1819	0,1897	0,1974	0,2051	0,2128	0,2205	0,2282
0,3	0,2358	0,2434	0,2510	0,2586	0,2661	0,2737	0,2812	0,2886	0,2960	0,3035
0,4	0,3108	0,3182	0,3255	0,3328	0,3401	0,3473	0,3545	0,3616	0,3686	0,3759
0,5	0,3829	0,3899	0,3969	0.4039	0,4108	0.4177	0.4245	0,4313	0,4381	0,4448
0,6	0,4515	0,4581	0,4647	0,4713	0,4778	0,4843	0,4907	0,4971	0,5035	0,5098
0,7	0,5161	0,5223	0,5285	0,5346	0,5407	0,5467	0,5527	0,5587	0,5646	0,5705
0,8	0,5763	0,5821	0,5878	0,5935	0,5991	0,6047	0,6102	0,6157	0,6211	0,6265
0,9	0,6319	0,6372	0,6424	0,6476	0,6528	0,6579	0,6629	0,6679	0,6729	0,6778
1,0	0,6827	0,6875	0,6923	0,6970	0,7017	0,7063	0,7109	0,7154	0,7199	0,7243
1,1	0,7287	0,7330	0,7373	0,7415	0,7457	0,7499	0,7540	0,7580	0,7620	0,7660
1,2	0,7699	0,7737	0,7775	0,7813	0,7850	0,7887	0,7923	0,7959	0,7984	0,8029
1,3	0,8064	0,8098	0,8132	0,8165	0,8198	0,8230	0,8262	0,8293	0,8324	0,8355
1,4	0,8385	0,8415	0,8444	0,8473	0,8501	0,8529	0,8557	0,8584	0,8611	0,8688
1,5	0,8664	0,8690	0,8715	0,8740	0,8764	0,8789	0,8812	0,8836	0,8859	0,8882
1,6	0,8904	0,8926	0,8948	0,8969	0,8990	0,9011	0,9031	0,9051	0,9070	0,9090
1,7	0,9109	0,9127	0,9146	0,9164	0,9181	0,9199	0,9216	0,9233	0,9249	0,9265
1,8	0,9281	0,9297	0,9312	0,9327	0,9342	0,9357	0,9371	0,9385	0,9392	0,9412
1,9	0,9426	0,9439	0,9451	0,9464	0,9476	0,9488	0,9500	0,9512	0,9523	0,9533
2,0	0,9545	0,9556	0,9566	0,9576	0,9586	0,9596	0,9606	0,9616	0,9625	0,9634
2,1	0,9643	0,9651	0,9660	0,9668	0,9676	0,9684	0,9692	0,9700	0,9707	0,9715
2,2	0,9722	0,9729	0,9736	0,9743	0,9749	0,9756	0,9762	0,9768	0,9774	0,9780
2,3	0,9786	0,9791	0,9797	0,9802	0,9807	0,9812	0,9817	0,9822	0,9827	0,9832
2,4	0,9836	0,9841	0,9845	0,9849	0,9853	0,9857	0,9861	0,9865	0,9869	0,9872
2,5	0,9876	0,9879	0,9883	0,9886	0,9889	0,9892	0,9895	0,9898	0,9901	0,9904

Окончание табл. I

Целые и десятые доли x	Сотые доли x									
	0	1	2	3	4	5	6	7	8	9
2,6	0,9907	0,9910	0,9912	0,9915	0,9917	0,9920	0,9922	0,9924	0,9926	0,9928
2,7	0,9931	0,9933	0,9935	0,9937	0,9939	0,9940	0,9942	0,9944	0,9946	0,9947
2,8	0,9949	0,9951	0,9952	0,9953	0,9955	0,9956	0,9958	0,9959	0,9960	0,9961
2,9	0,9963	0,9964	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972
3,0	0,9973	0,9974	0,9975	0,9976	0,9976	0,9977	0,9978	0,9979	0,9979	0,9980
3,1	0,9981	0,9981	0,9982	0,9983	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986
3,2	0,9986	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,3	0,9990	0,9991	0,9994	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,4	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995	0,9995
3,5	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997	0,9997
3,6	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998	0,9998	0,9998
3,7	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
4,0	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999

Таблица II

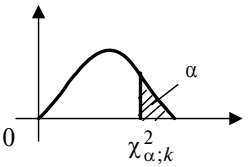
Значения $t_{\gamma,k}$-критерия Стьюдента



Число степеней свободы k	Вероятность γ											
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	0,98	0,99
1	0,16	0,32	0,51	0,73	1,00	1,38	1,96	3,08	6,31	12,71	31,82	63,66
2	14	29	44	62	0,82	06	34	1,89	2,92	4,30	6,96	9,92
3	14	28	42	58	76	0,98	25	64	35	3,18	4,54	5,84
4	13	27	41	57	74	94	19	53	13	2,78	3,75	4,60
5	13	27	41	56	73	92	16	48	01	57	36	03
6	0,13	0,26	0,40	0,55	1,72	1,91	1,13	1,44	1,94	2,45	3,14	3,71
7	13	26	40	55	71	90	12	41	89	36	00	50
8	13	26	40	55	70	89	11	40	86	31	2,90	35
9	13	26	40	54	70	88	10	38	83	26	82	25
10	13	26	40	54	70	88	09	37	81	23	76	17
11	0,13	0,26	0,40	0,54	0,70	0,88	1,09	1,36	1,80	2,20	2,72	3,11
12	13	26	39	54	69	87	08	36	78	18	68	05
13	13	26	39	54	69	87	08	35	77	16	65	01
14	13	26	39	54	69	87	08	34	76	14	62	2,98
15	13	26	39	54	69	87	07	34	75	13	60	95
16	0,13	0,26	0,39	0,53	0,69	0,86	1,07	1,34	1,75	2,12	2,58	2,92
17	13	26	39	53	69	86	07	33	74	11	57	90
18	13	26	39	53	69	86	07	33	73	10	55	88
19	13	26	39	53	69	86	07	33	73	09	54	86
20	13	26	39	53	69	86	06	32	72	09	53	84
21	0,13	0,26	0,39	0,53	0,69	0,86	1,06	1,32	1,72	2,08	2,52	2,83
22	13	26	39	53	69	86	06	32	72	07	51	82
23	13	26	39	53	68	86	06	32	71	07	50	81
24	13	26	39	53	68	86	06	32	71	06	49	80
25	13	26	39	53	68	86	06	32	71	06	48	79
26	0,13	0,26	0,39	0,53	0,68	0,86	1,06	1,31	1,71	2,06	2,48	2,78
27	13	26	39	53	68	85	06	31	70	05	47	77
28	13	26	39	53	68	85	06	31	70	05	47	76
29	13	26	39	53	68	85	05	31	70	04	46	76
30	13	26	39	53	68	85	05	31	70	04	46	75
40	0,13	0,25	0,39	0,53	0,68	0,85	1,05	1,30	1,68	2,02	2,42	2,70
60	13	25	39	53	68	85	05	30	67	00	39	66
120	0,13	0,25	0,39	0,53	0,68	0,84	1,04	1,29	1,66	1,98	2,36	2,62
∞	13	25	38	52	67	84	04	28	64	96	33	58

Таблица III

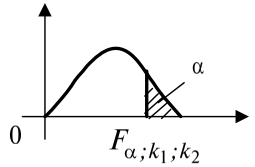
Значения $\chi^2_{\alpha;k}$ критерия Пирсона



Число степеней свободы k	Вероятность α												
	0,99	0,98	0,95	0,90	0,80	0,70	0,50	0,30	0,20	0,10	0,05	0,02	0,01
1	0,00	0,00	0,00	0,02	0,06	0,15	0,45	1,07	1,64	2,71	3,84	5,41	6,64
2	0,02	0,04	0,10	0,21	0,45	0,71	1,39	2,41	3,22	4,60	5,99	7,82	9,21
3	0,11	0,18	0,35	0,58	1,00	1,42	2,37	3,66	4,64	6,25	7,82	9,84	11,3
4	0,30	0,43	0,71	1,06	1,65	2,20	3,36	4,88	5,99	7,78	9,49	11,7	13,3
5	0,55	0,75	1,14	1,61	2,34	3,00	4,35	6,06	7,29	9,24	11,1	13,4	15,1
6	0,87	1,13	1,63	2,20	3,07	3,83	5,35	7,23	8,56	10,6	12,6	15,0	16,8
7	1,24	1,56	2,17	2,83	3,82	4,67	6,35	8,38	9,80	12,0	14,1	16,6	18,5
8	1,65	2,03	2,73	3,49	4,59	5,53	7,34	9,52	11,0	13,4	15,5	18,2	20,1
9	2,09	2,53	3,32	4,17	5,38	6,39	8,34	10,7	12,2	14,7	16,9	19,7	21,7
10	2,56	3,06	3,94	4,86	6,18	7,27	9,34	11,8	13,4	16,0	18,3	21,2	23,2
11	3,05	3,61	4,58	5,58	6,99	8,15	10,3	12,9	14,6	17,3	19,7	22,6	24,7
12	3,57	4,18	5,23	6,30	7,81	9,03	11,3	14,0	15,8	18,5	21,0	24,1	26,2
13	4,11	4,76	5,89	7,04	8,63	9,93	12,3	15,1	17,0	19,8	22,4	25,5	27,7
14	4,66	5,37	6,57	7,79	9,47	10,8	13,3	16,2	18,1	21,1	23,7	26,9	29,1
15	5,23	5,98	7,26	8,55	10,3	11,7	14,3	17,3	19,3	22,3	25,0	28,3	30,6
16	5,81	6,61	7,96	9,31	11,1	12,6	15,3	18,4	20,5	23,5	26,3	29,6	32,0
17	6,41	7,26	8,67	10,1	12,0	13,5	16,3	19,5	21,6	24,8	27,6	31,0	33,4
18	7,02	7,91	9,39	10,9	12,9	14,4	17,3	20,6	22,8	26,0	28,9	32,3	34,8
19	7,63	8,57	10,1	11,6	13,7	15,3	18,3	21,7	23,9	27,2	30,1	33,7	36,2
20	8,26	9,24	10,8	12,4	14,6	16,3	19,3	22,8	25,0	28,4	31,4	35,0	37,6
21	8,90	9,92	11,6	13,2	15,4	17,2	20,3	23,9	26,2	29,6	32,7	36,3	38,9
22	9,54	10,6	12,3	14,0	16,3	18,1	21,3	24,9	27,3	30,8	33,9	37,7	40,3
23	10,2	11,3	13,1	14,8	17,2	19,0	22,3	26,0	28,4	32,0	35,2	39,0	41,6
24	10,9	12,0	13,8	15,7	18,1	19,9	23,3	27,1	29,6	33,2	36,4	40,3	43,0
25	11,5	12,7	14,6	16,5	18,9	20,9	24,3	28,2	30,7	34,4	37,7	41,7	44,3
26	12,2	13,4	15,4	17,3	19,8	21,8	25,3	29,2	31,8	35,6	38,9	42,9	45,6
27	12,9	14,1	16,1	18,1	20,7	22,7	26,3	30,3	32,9	36,7	40,1	44,1	47,0
28	13,6	14,8	16,9	18,9	21,6	23,6	27,3	31,4	34,0	37,9	41,3	45,4	48,3
29	14,3	15,6	17,7	19,8	22,5	24,6	28,3	32,5	35,1	39,1	42,6	46,7	49,6
30	14,9	16,3	18,5	20,6	23,4	25,5	29,3	33,5	36,2	40,3	43,8	48,0	50,9

Таблица IV

Значения $F_{\alpha;k_1;k_2}$ -критерия Фишера—Снедекора



k_1	α=0,05																		
k_2	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161	200	216	225	230	234	237	239	240	242	244	246	248	249	250	251	252	253	254
2	18,5	19,0	19,2	19,2	19,3	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55	8,53
4	7,71	6,94	6.59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93	1,88

Продолжение табл. IV

k_2 \ k_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
									α=0,05										
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,13	2,05	2,01	1,96	1,91	1,86	1,81	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79	1,73
25	4,24	3,39	2,99	2,76	2,60	4,49	2,40	2,34	2,28	2,24	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,77	1,71
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,07	1,99	1,95	1,90	1,85	1,80	1,75	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,13	2,06	1,97	1,93	1,88	1,84	1,79	1,73	1,67
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,12	2,04	1,96	1,91	1,87	1,82	1,77	1,71	1,65
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,10	2,03	1,94	1,90	1,85	1,81	1,75	1,70	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58	1,51
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,91	1,83	1,75	0,66	1,61	1,55	1,50	1,43	1,35	1,25
∞	3,84	3,00	3,60	2,37	2,21	2,10	2,01	1,94	1,83	1,83	1,75	1,67	1,57	1,52	1,46	1,39	1,32	1,22	1,00
									α=0,01										
1	4052	4999,5	5403	5625	5764	5859	5928	5982	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39	99,40	99,42	99,43	99,45	99,46	99,47	99,47	99,48	99,49	99,50
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,35	27,23	27,05	26,87	26,69	26,60	26,50	26,41	26,32	26,22	26,13
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55	14,37	14,20	14,02	13,93	13,84	13,75	13,65	13,56	13,46
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,89	9,72	9,55	9,47	9,38	9,29	9,20	9,11	9,02
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,72	7,56	7,40	7,31	7,23	7,14	7,06	6,97	6,88
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,47	6,31	6,16	6,07	5,99	5,91	5,82	5,74	5,65
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,67	5,52	5,36	5,28	5,20	5,12	5,03	4,95	4,86
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,11	4,96	4,81	4,73	4,65	4,57	4,48	4,40	4,31
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,71	4,56	4,41	4,33	4,25	4,17	4,08	4,00	3,91
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,40	4,25	4,10	4,02	3,94	3,86	3,78	3,69	3,60
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,16	4,01	3,86	3,78	3,70	3,62	3,54	3,45	3,36
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	3,96	3,82	3,66	3,59	3,51	3,43	3,34	3,25	3,17
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,80	3,66	3,51	3,43	3,35	3,27	3,18	3,09	3,00
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,67	3,52	3,37	3,29	3,21	3,13	3,05	2,96	2,87

Окончание табл. IV

k_2 \ k_1	α=0,01																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,55	3,41	3,26	3,18	3,10	3,02	2,93	2,84	2,75
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,46	3,31	3,16	3,08	3,00	2,92	2,83	2,75	2,65
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,37	3,23	3,08	3,00	2,92	2,84	2,75	2,66	2,57
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,30	3,15	3,00	2,92	2,84	2,76	2,67	2,58	2,49
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,23	3,09	2,94	2,86	2,78	2,69	2,61	2,52	2,42
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31	3,17	3,03	2,88	2,80	2,72	2,64	2,55	2,46	2,36
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,12	2,98	2,83	2,75	2,67	2,58	2,50	2,40	2,31
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,07	2,93	2,78	2,70	2,62	2,54	2,45	2,35	2,26
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17	3,03	2,89	2,74	2,66	2,58	2,49	2,40	2,31	2,21
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13	2,99	2,85	2,70	2,62	2,54	2,45	2,36	2,27	2,17
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09	2,96	2,81	2,66	2,58	2,50	2,42	2,33	2,23	2,13
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06	2,93	2,78	2,63	2,55	2,47	2,38	2,29	2,20	2,10
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03	2,90	2,75	2,60	2,52	2,44	2,35	2,26	2,17	2,06
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00	2,87	2,73	2,57	2,49	2,41	2,33	2,23	2,14	2,03
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,84	2,70	2,55	2,47	2,39	2,30	2,21	2,11	2,01
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,66	2,52	2,37	2,29	2,20	2,11	2,02	1,92	1,80
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,50	2,35	2,20	2,12	2,03	1,94	1,84	1,73	1,60
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47	2,34	2,19	2,03	1.95	1,86	1,76	1,66	1,53	1,38
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32	2,18	2,04	1,88	1,79	1,70	1,59	1,47	1,32	1,00

k_1 — число степеней свободы для большой дисперсии, k_2 — для меньшей дисперсии.

Таблица V

Значения $d_н$ и $d_в$ критерия Дарбина—Уотсона на уровне значимости $\alpha = 0{,}05$ (n — число наблюдений, p — число объясняющих переменных)

n	p=1		p=2		p=3		p=4		p=5	
	$d_н$	$d_в$	$d_н$	$d_в$	$d_н$	$d_в$	$d_н$	$d_в$	$d_н$	$d_в$
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	1,78	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	0,85	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	1,99
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
31	1,36	1,50	1,30	1,57	1,23	1,65	1,16	1,74	1,09	1,83
32	1,37	1,50	1,31	1,57	1,34	1,65	1,18	1,73	1,11	1,82
33	1,38	1,51	1,32	1,58	1,26	1,65	1,19	1,73	1,13	1,81
34	1,39	1,51	1,33	1,58	1,27	1,65	1,21	1,73	1,15	1,81
35	1,40	1,52	1,34	1,58	1,28	1,65	1,22	1,73	1,16	1,80
36	1,41	1,52	1,35	1,59	1,29	1,65	1,24	1,73	1,18	1,80
37	1,42	1,53	1,36	1,59	1,31	1,66	1,25	1,72	1,19	1,80
38	1,43	1,54	1,37	1,59	1,32	1,66	1,26	1,72	1,21	1,79
39	1,43	1,54	1,38	1,60	1,33	1,66	1,27	1,72	1,22	1,79
40	1,44	1,54	1,39	1,60	1,34	1,66	1,29	1,72	1,23	1,79
45	1,48	1,57	1,43	1,62	1,38	1,67	1,34	1,72	1,29	1,78
50	1,50	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,77
55	1,53	1,60	1,49	1,64	1,45	1,68	1,41	1,72	1,38	1,77
60	1,55	1,62	1,51	1,65	1,58	1,69	1,44	1,73	1,41	1,77
65	1,57	1,63	1,54	1,66	1,50	1,70	1,47	1,73	1,44	1,77
70	1,58	1,64	1,55	1,67	1,52	1,70	1,49	1,74	1,46	1,77
75	1,60	1,65	1,57	1,68	1,54	1,71	1,51	1,74	1,49	1,77
80	1,61	1,66	1,59	1,69	1,56	1,72	1,53	1,74	1,51	1,77
85	1,62	1,67	1,60	1,70	1,57	1,72	1,55	1,75	1,52	1,77
90	1,63	1,68	1,61	1,70	1,59	1,73	1,57	1,75	1,54	1,78
95	1,64	1,69	1,62	1,71	1,60	1,73	1,58	1,75	1,56	1,78
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78

Предметный указатель

Оглавление