

С. А. Айвазян

Учебник



Методы эконометрики

магистр

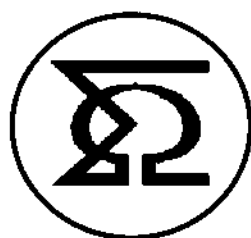




Рекомендовано Учебно-методическим объединением
по образованию в области математических методов
в экономике в качестве учебника для студентов высших учебных
заведений, обучающихся по специальности 080116
«Математические методы в экономике» и другим экономическим
специальностям

МОСКОВСКАЯ ШКОЛА ЭКОНОМИКИ
МГУ имени М. В. ЛОМОНОСОВА

С. А. Айвазян



Методы эконометрики

Учебник



Москва
магистр
ИНФРА-М
2010

УДК [31:33](075.8)
ББК 65.051я73-1
А36

Рецензенты:

кафедра статистических методов ГУ–ВШЭ
(зав. кафедрой д-р экон. наук, проф. *В. С. Мхитарян*);
д-р экон. наук, проф. РЭШ А. А. *Пересецкий*

*Печатается по решению Ученого совета
Московской школы экономики МГУ имени М. В. Ломоносова*

Айвазян С. А.

А36 Методы эконометрики : учебник / С. А. Айвазян. — М. : Магистр : ИНФРА-М, 2010. — 512 с.

ISBN 978-5-9776-0153-5 (в пер.)

ISBN 978-5-16-004050-9

Агентство СІР РГБ

Содержание учебника соответствует действующим образовательным стандартам и учебным программам высших учебных заведений экономического профиля по дисциплине «Эконометрика». Особенность данного издания заключается в том, что в нем в описание традиционных методов решения эконометрических задач впервые органично встроены (там, где это позволяет повысить точность и глубину анализа) современные методы многомерного статистического анализа, ранее не включавшиеся в инструментарий эконометрики (в частности, дискриминантный и кластер-анализы, метод главных компонент и др.).

Представленные в учебнике методы и модели регрессионного анализа, бинарного и множественного выбора, анализа временных рядов могут составить содержание одного или двух базовых семестровых курсов по эконометрике в рамках учебного плана бакалавриата.

Для студентов, аспирантов, преподавателей, а также специалистов по прикладной экономике и эконометрике.

УДК [31:33](075.8)
ББК 65.051я73-1

ISBN 978-5-9776-0153-5
ISBN 978-5-16-004050-9

© Айвазян С. А., 2010
© Издательство «Магистр», 2010

ОГЛАВЛЕНИЕ

Предисловие.....	9
Глава 1. Введение.....	13
1.1. Эконометрика: эволюция определения и реальность	13
1.2. Ободрение математического аппарата эконометрики.....	16
1.3. Место эконометрики в ряду математико-статистических и экономических дисциплин	19
1.4. Эконометрическая модель и проблемы эконометрического моделирования.....	22
Выводы	30
Глава 2. Введение в регрессионный анализ.....	33
2.1. Общая формулировка проблемы статистического исследования зависимостей.....	33
2.2. Какова конечная прикладная цель статистического исследования зависимостей?.....	42
2.3. Некоторые типовые задачи практики эконометрического моделирования.....	45
2.4. Основные типы зависимостей между количественными переменными	50
2.5. О выборе общего вида функции регрессии	55
Выводы	65
Глава 3. Введение в корреляционный анализ.....	67
3.1. Назначение и место корреляционного анализа в статистическом исследовании.....	67
3.2. Корреляционный анализ количественных признаков.....	69
3.3. Корреляционный анализ ранговых (ординальных) переменных: ранговая корреляция	96
3.4. Корреляционный анализ категоризованных переменных: таблицы сопряженности	111
Выводы	117
Глава 4. Классическая линейная модель множественной регрессии (КЛММР).....	121
4.1. Описание КЛММР. Основные допущения модели.....	121
4.2. Оценивание неизвестных параметров КЛММР: метод наименьших квадратов и метод максимального правдоподобия	126

4.3. Анализ вариации результирующего показателя y и выборочный коэффициент детерминации $\hat{R}_{y.X}^2$	140
4.4. Мультиколлинеарность и отбор наиболее существенных объясняющих переменных в КЛММР	145
4.5. КЛММР с линейными ограничениями на параметры	162
4.6. Общий подход к статистической проверке гипотез о наличии линейных связей между параметрами КЛММР	167
Выводы	176
Глава 5. Обобщенная линейная модель множественной регрессии	179
5.1. Описание обобщенной линейной модели множественной регрессии (ОЛММР)	179
5.2. Оценки параметров ОЛММР по обобщенному методу наименьших квадратов (ОМНК-оценки)	183
5.3. ОЛММР с гетероскедастичными остатками	188
5.4. ОЛММР с автокоррелированными остатками	198
5.5. Практически реализуемый ОМНК (общий подход)	207
Выводы	210
Глава 6. Прогнозирование, основанное на линейных моделях множественной регрессии	213
6.1. Анализ точности оцененной ЛММР (теоретическая база для решения задач прогноза)	214
6.2. Наилучший точечный прогноз $y(X)$ и $f(X) = E(y X)$, основанный на ОЛММР	216
6.3. Интервальный прогноз $y(X)$ и $f(X) = E(y X)$, основанный на ОЛММР	220
6.4. Анализ точности регрессионной модели и прогнозирование в условиях реалистической ситуации	226
Выводы	230
Глава 7. Линейные модели регрессии со стохастическими объясняющими переменными	233
7.1. Случайные остатки ε не зависят от предикторов X и оцениваемых коэффициентов регрессии Θ	235
7.2. Общий случай: стохастические предикторы X коррелированы с регрессионными остатками ε . Метод инструментальных переменных	238
7.3. Случайные ошибки в измерении значений объясняющих переменных	243
Выводы	249
Глава 8. Линейные регрессионные модели с переменной структурой	251
8.1. Проблема неоднородных (в регрессионном смысле) данных	251

8.2. Введение «манекенов» (фиктивных переменных) в линейную модель регрессии	254
8.3. Проверка регрессионной однородности двух групп наблюдений (критерий Г. Чоу)	263
8.4. Построение КЛММР по неоднородным данным в условиях, когда значения сопутствующих переменных неизвестны	265
Выводы	269
Глава 9. Модели с дискретными и дискретно-непрерыв- ными зависимыми переменными	271
9.1. Модели бинарного выбора	273
9.2. Модели множественного выбора	282
9.3. Связь моделей бинарного и множественного выбора с дискриминантным анализом	285
9.4. Модель с дискретно-непрерывной зависимой переменной (тобит-модель)	287
Выводы	291
Глава 10. Анализ одномерных временных рядов (модели и прогнозирование)	293
10.1. Временной ряд: определения, примеры, формулировка основных задач	295
10.2. Стационарные временные ряды и их основные характеристики	302
10.3. Неслучайная составляющая временного ряда и методы его сглаживания	314
10.4. Модели стационарных временных рядов и их идентификация	336
10.5. Модели нестационарных временных рядов и их идентификация	378
10.6. Прогнозирование экономических показателей, основанное на использовании моделей временных рядов	395
Выводы	409
Приложение 1. Таблицы математической статистики	413
Приложение 2. Необходимые сведения из матричной алгебры ..	433
Приложение 3. Многомерный статистический анализ	455
Литература	493
Алфавитно-предметный указатель	497

Предисловие

Дорогой читатель!

Вы держите в руках учебник по методам эконометрики — дисциплины, которая является одной из трех базовых дисциплин (наряду с микро- и макроэкономикой) высшего экономического образования. К сожалению, подобный статус эконометрики в России был признан с большим запозданием: лишь начиная с 1992 года эконометрика была введена в учебные планы экономического образования некоторых ведущих российских вузов. Такое позднее признание эконометрики сразу поставило российских студентов в невыгодное положение: к тому времени в России было издано лишь несколько относительно старых переводных книг по эконометрике, а первые отечественные учебники по этой дисциплине появились только в 1997–1998 гг. (см. [Магнус, Катышев, Пересецкий (2005)], [Айвазян (2001)]). Однако сейчас ситуация существенно выправилась: многократно переизданы упомянутые две книги, вышли в свет отечественные учебники под редакцией И.И. Елисеевой (2006), В.И. Сушкова (2005), переводы с английского прекрасных книг [Берндт (2005)], [Магнус, Нейдекер (2007)], [Вербик (2008)].

Значительно повысились возможности использования лучших образцов *англоязычной* эконометрической литературы (за счет повышения общего уровня владения английским языком нашими студентами и специалистами, а также — развития электронных средств связи, см., например, список англоязычной литературы в конце данного издания).

При таких обстоятельствах возникает естественный вопрос:

что побудило автора к созданию еще одного учебника по эконометрике?

Чтобы ответить на этот вопрос, прежде всего, должен заметить, что мое понимание сущности и назначения эконометрических методов *несколько отличается от общепринятого в североамериканском и западноевропейском эконометрическом сообществе*. Это понимание формировалось на базе теоретико-вероятностной и математико-статистической отечественной школы в процессе знакомства с лучшими образцами англоязычной эконометрической литературы, а также — личных научных контактов с коллегами из Гарвардского университета (США), Университета Париж-1/Сорбонна (Франция), Тилбургского и Роттердамского университетов (Голландия), Женевского университета (Швейцария) и других образовательных и научных центров мира. Сущность этих отличий кратко представлена в пп. 1.1 и 1.2 главы 1 (Введения) книги. К этому надо добавить, что со временем несколько трансформируются представления специалистов о багаже методов эконометрики, смещаются акценты в оценке областей их применения. Не со всеми такими представлениями, принятыми, скажем, в научных кругах США, я могу

согласиться. Так, например, принято включать в курсы (учебники) по эконометрике «Теорию больших выборок» (или «Асимптотическую теорию»), «Непараметрические и полупараметрические методы принятия статистических решений», развернутое изложение метода максимального правдоподобия. Но вся эта тематика традиционно представлена в качестве разделов в других самостоятельных научных дисциплинах — теории вероятностей и математической статистике. В то же время важнейшие для эконометрического анализа *прикладные методы многомерной статистики* (дискриминантный и кластер-анализы, метод главных компонент и др.) по непонятным причинам отсутствуют в эконометрических курсах и классических университетских учебниках Северной Америки и Западной Европы. Добавлю к этому, что за последние несколько лет серьезный импульс к развитию получили некоторые специальные методы многомерного статистического анализа, получен ряд важных результатов в области финансовой эконометрики, используемых при эконометрическом анализе финансовых данных в *задачах управления рисками*.

Все упомянутые обстоятельства и определили специфические отличия данного издания от традиционных учебников по эконометрике. Среди этих отличий, в первую очередь, следует выделить тот факт, что в описание традиционных методов решения эконометрических задач впервые, насколько мне известно, органично встроены (там, где это представляется объективно необходимым) процедуры многомерного статистического анализа, ранее не принимавшиеся в расчет (такие как кластер-анализ, дискриминантный анализ, метод главных компонент).

К особенностям книги следует отнести и факт включения в нее двух обширных **вводных** глав по регрессионному (глава 2) и корреляционному (глава 3) анализам. Многолетняя практика преподавания в ведущих российских вузах (Московской школе экономики МГУ им. М.В. Ломоносова, экономическом факультете МГУ, Государственном университете — Высшей школе экономики, Российской экономической школе, Московском государственном университете экономики, статистики и информатики) убедила меня в том, что, приступая к освоению эконометрики, студенты, как правило, имеют явный дефицит знаний и умений в основах этих двух разделов.

Возвращаясь к вопросу о мотивации создания учебника, следует признать, что существенное влияние на замысел и содержание книги оказал многолетний опыт исследовательской и педагогической работы автора в Московской школе экономики Московского государственного университета им. М.В. Ломоносова. Без постоянных рабочих контактов с коллегами по кафедре эконометрики и математических методов экономики, без главных критиков и генераторов вопросов — студентов МШЭ МГУ эта книга вряд ли появилась бы на свет.

Учебник охватывает весьма полный спектр методов математико-ста-

тистического инструментария эконометрики по всем его традиционным разделам, а именно:

- (1) *классическая линейная модель регрессии и классический метод наименьших квадратов* (гл. 4 и 6);
- (2) *обобщенная линейная модель регрессии и обобщенный метод наименьших квадратов* (гл. 5 и 6);
- (3) *линейная модель регрессии с переменной структурой* (гл. 8);
- (4) *регрессионные модели с дискретной зависимой переменной: модели бинарного и множественного выбора* (гл. 9);
- (5) *регрессионные модели в условиях цензурирования, усечения или выборочной селективности (the sample selection) зависимой переменной* (гл. 9);
- (6) *статистический анализ одномерных и многомерных временных рядов* (гл. 10).

В приложения вынесены, помимо математико-статистических таблиц, необходимые для усвоения материала книги сведения из матричной алгебры и многомерного статистического анализа. Предполагается, что читатель уже имеет необходимую подготовку по математической статистике в рамках базовых курсов, предусмотренных государственными стандартами для экономических специальностей вузов.

Следует, однако, подчеркнуть, что в предлагаемом издании представлены, конечно, далеко не все важнейшие разделы современной эконометрики. В нем нет, например, методов и моделей *анализа многомерных временных рядов, анализа панельных данных, обобщенного метода моментов*, не отражены последние достижения в области *финансовой эконометрики* (копула-функции, методы управления финансовыми рисками), не представлены *байесовский подход* к эконометрическому анализу и *методы измерения и анализа синтетических латентных категорий*, комплексно характеризующих качество или эффективность функционирования анализируемой системы. Вся эта проблематика будет представлена в **продвинутом** курсе эконометрики (предназначенном для *магистерского* уровня экономического образования), который готовится к изданию мной и моим итальянским коллегой (по МШЭ) *Деаном Фантаццини*.

Что касается **базового бакалаврского уровня** дисциплины «Эконометрика», то он обеспечен представленными в данном учебнике методами и моделями, которые могут составить содержание одного или двух (в зависимости от отведенного в учебном плане вуза времени) семестровых курсов по схеме: 2 часа лекций и 2 часа семинарских занятий

в неделю. Эти занятия должны быть, конечно, оснащены задачами и упражнениями (в том числе в компьютерных классах), для чего помимо примеров, приведенных в книге, можно рекомендовать, например, «Сборник задач к начальному курсу эконометрики» П. К. Катыхова, Я. Р. Магнуса и А. А. Пересецкого (издательство «Дело», 2008 г.).

Вычислительная реализация описываемых в книге методов основана на использовании статистических и эконометрических пакетов SPSS, EViews, R и STATA.

Автор старался следовать такому стилю изложения, который *помогал бы читателю понять, в первую очередь, основную идею и смысл описываемого метода*, избежать чисто формального, механистического восприятия материала. Правда, это неизбежно связано с увеличением объема книги.

В заключение хочу выразить признательность. Прежде всего я благодарен коллективам и администрации МШЭ МГУ и Центрального экономико-математического института РАН, плодотворная профессиональная среда которых существенно помогла в работе над учебником. Большую пользу я получил от общения с коллегами — преподавателями эконометрики и статистики различных вузов России, Литвы, Молдавии в ходе специально организованной серии семинаров (1997–2002 гг.), на которых отечественные и зарубежные специалисты (в их числе — и автор учебника) представляли свои циклы лекций в рамках общей программы повышения квалификации. Наконец, я благодарен Алле Павловне и Галине Юрьевне Грохотовым за самоотверженный и профессиональный труд по подготовке оригинал-макета книги.

Хочу обратить внимание читателя на следующий факт: для достижения успеха в приложениях эконометрических методов эконометристу приходится весьма тонко балансировать между экономической теорией, возможностями необходимого информационного обеспечения, формулировкой исходных допущений модели и самими методами. Другими словами, прикладная эконометрика — это *не только наука, но еще и искусство*, овладение которым постигается с опытом. Так что желаю читателю успеха в постижении науки и искусства владения тонким и эффективным инструментарием эконометрики!

С. А. Айвазян
Москва, ноябрь 2009 г.

Глава 1

Введение

До начала 90-х годов прошлого века *эконометрика* — одна из базовых (наряду с *микро- и макроэкономикой*) дисциплин экономического образования во всем мире — была предана в России, в силу известных исторических причин, забвению и в образовательном процессе, и в приложениях. Теперь мы как бы «наверстываем упущенное»: активно заимствуем зарубежный опыт в данной области (к сожалению, не всегда с нужной избирательностью и профессионализмом), создаем учебники, читаем студентам экономических специальностей базовые и продвинутые курсы по этой дисциплине, пытаемся проводить прикладные (эмпирические) исследования с использованием современного инструментария эконометрики. К сожалению, ряд тревожных симптомов характеризует все эти процессы, причем некоторые из них имеют и «интернациональный масштаб». Серьезные исследователи не могут проходить мимо этих симптомов, что приводит иногда и к радикальным позициям типа «*А не гол ли король?*» (см., например, [Тутубалин (2004)]).

Генетические корни этих тревожных симптомов мы видим в подмене двух понятий:

- а) понятие «*эконометрика*», как правило, подменяется де-факто понятием «*эконометрические методы*»;
- б) реальное содержание самих *эконометрических методов* подменяется (а точнее — искусственно ограничивается) сложившимися в 30-е — 40-е годы прошлого века каноническими рамками.

Остановимся кратко на развитии этих двух тезисов.

1.1 Эконометрика: эволюция определения и реальность

Эволюцию определения *эконометрики* как научной дисциплины можно проследить, например, сравнивая исходное и слишком узкое опре-

деление Павла Сиомпы с де факто принятым сейчас определением Рагнара Фриша. В своей малоизвестной (опубликованной в 1910 году в Германии) книге "Oekonometrie" П. Сиомпа определял, что цель эконометрики состоит «*в математическом описании рядов экономических данных и в их отображении в геометрической или графической форме*». Современное же понимание эконометрики соответствует позиции Рагнара Фриша, который в первом номере журнала «Эконометрика» (1933 г.) писал: «... Эконометрика не должна восприниматься как синоним применения математики в экономике. Опыт показывает, что и статистика, и экономическая теория, и математика, взятые по отдельности, являются необходимыми, но не достаточными для действительного понимания количественных отношений в современной экономической жизни. **Именно объединение всех трех частей дает мощный эффект. И именно это объединение и составляет эконометрику**».

Как мы видим, в определении П. Сиомпы речь идет, по существу, о *математических методах*, примененных к экономическим данным, и ни слова не говорится ни об экономической теории, ни о необходимости синтеза последней с экономическими измерениями и математическими методами. Сегодня практически все специалисты в области эконометрики **де факто** придерживаются определения Р. Фриша. Однако реальность свидетельствует, увы, о другом. Так, например, анализ имеющегося в мире учебно-методического обеспечения эконометрики свидетельствует: абсолютное большинство книг — учебников по эконометрике, включая высокопрофессиональные по методическому исполнению, широте и глубине охвата эконометрических методов (см., например, [Greene (2000)], [Hayashi (2000)], [Maddala (1988)], [Магнус, Катышев, Пересецкий (2005)], [Суслов и др. (2005)]), независимо от своего названия в действительности является учебниками **по методам эконометрики, а не по эконометрике!** Другими словами, из трех необходимых составляющих эконометрики во всех этих учебниках представлены в лучшем случае две — экономические измерения и математические методы. Но, заметьте, почти все упомянутые выше учебники озаглавлены «Эконометрика» или еще более обязывающе — «Эконометрический анализ» или «Эконометрическое моделирование» (лишь Джонстон и Ди Нардо честно назвали свой прекрасный учебник «Эконометрические методы», см. [Johnston, DiNardo (1997)]).

Это означает, в частности, что при изложении методологии эконометрического исследования и разборе конкретных эмпирических примеров авторы не уделяют (или уделяют явно недостаточно) внимания необходимым положениям **экономической теории** на важнейшем этапе спецификации модели, а также подробнейшему анализу генезиса и природы используемых **экономических измерений**, без которого невозможен адекватный подбор эконометрических методов и моделей. Подобная ситуация провоцирует справедливую критику существующего положения

дел в области эконометрического образования и эконометрических прикладных исследований. В качестве радикального примера такой критики сошлемся на книгу В.Н. Тутубалина «Эконометрика: образование, которое нам не нужно». В ней, в частности, автор с пониманием и доказательно говорит о невнимании к природе и генезису анализируемых статистических данных (с. 35 и 72), приводит примеры распространенных ошибок в спецификации и интерпретации эконометрических моделей (с. 27), дает подробный критический разбор примеров неудачного эконометрического моделирования (см., например, анализ модели «цен на квартиры», с. 76–81) и т.п.

Разделяя недоумение и огорчение автора этой книги по поводу выявленных им достаточно общих и типичных недостатков, мы не можем, конечно, присоединиться к его главному заключению (см. с. 163): «Что же можно предложить для общеэкономического образования?¹ Речь должна идти о применении достаточно простых приемов математической статистики для анализа экономической информации. . . ». В круг этих «достаточно простых приемов» автор не рекомендует включать даже метод наименьших квадратов, «из которого эконометрика сделала некий эзотерический культ» (завершающая фраза упомянутой книги).

Неадекватная экстремальность подобного вывода обусловлена неверным посылом автора в определении эконометрики как науки, «*которая представляет собой приложения вероятностно-статистических методов к анализу экономических данных*» (это определение дважды повторено на с. 3 и 62). По существу, это дословно повторяет ущербное определение П. Сиомпы столетней давности (см. выше). Для успешной деятельности в рамках этого определения, действительно, достаточно профессионализма математического статистика и его вдумчивого, внимательного отношения к содержательной сущности решаемой задачи. Но посмотрите, как решаются современные экономические проблемы с помощью эконометрического моделирования в книге [Берндт (2005)]! В ней прекрасно представлен органичный синтез трех базовых составляющих эконометрики:

- экономической теории;
- экономических измерений;
- эконометрического (математико-статистического) инструментария.

Реализовано это по следующей схеме. Начиная со второй главы (первая глава — вводная), каждая из десяти следующих глав начинается с подробной формулировки анализируемой экономической проблемы. Обсуждаются связанные с этой проблемой экономические понятия, определения, наконец, положения экономической теории. Именно последние

¹Взамен эконометрики, как это следует из контекста.

являются, как правило, главным подспорьем в решении задач спецификации модели, в частности, в определении наборов эндогенных и экзогенных переменных и в выборе общего вида модели. Затем обсуждаются проблемы реализации и анализа необходимых для эконометрического построения модели экономических измерений, имеющегося информационного обеспечения модели. И, наконец, анализируются вопросы необходимого эконометрического инструментария, выбираются и реализуются эконометрические методы спецификации и идентификации искомой модели. В заключение, как правило, обсуждаются и интерпретируются результаты моделирования, проводится сравнительный анализ прикладной дееспособности известных из литературы альтернативных вариантов рассматриваемой модели.

Это и есть, с моей точки зрения, идеальная схема полнокровного эконометрического анализа, которую, увы, не встретишь даже в лучших учебниках по эконометрическим методам! К сожалению, книга Берндта — лишь счастливое исключение не только в отечественной, но и в мировой учебно-методической литературе по эконометрике. И, конечно же, ни один математический статистик, даже самой высокой квалификации, не в состоянии провести эконометрическое исследование на профессиональном уровне примеров из этой книги.

А это значит, что речь должна идти лишь о наполнении эконометрики (и в образовательном процессе, и в приложениях) содержанием, соответствующим ее современному определению. И если в *прикладных исследованиях* наши западноевропейские и североамериканские коллеги имеют прекрасные образцы эконометрической продукции (некоторые из которых представлены в [Берндт (2005)]) и существенно опережают нас в этой области, то в *образовательном процессе* недостаток «эконометрические методы вместо эконометрики» почти одинаково присущ и тем, и другим.

1.2 Обеднение математического аппарата эконометрики

В.Л. Макаров писал в 2001 году во «Вступительном слове» к [Айвазян (2001)]: «Границы, отделяющие методы эконометрики, традиционно включаемые в канонические западные учебники, от некоторых других методов многомерного статистического анализа, являются формально жестко очерченными, а *по существу* — весьма размытыми и условными. К примеру, никто не сможет объяснить, почему так называемые модели дискретного выбора (логит-, пробит- модели и др.) являются неперенными составными частями эконометрических учебников, а решающие точно те же задачи методы дискриминантного анализа (кстати при определенных условиях *функционально связанные* с логит- и пробит- моде-

лями) к «территории» эконометрического инструментария относить не принято! А ведь такое неоправданное «отлучение» ряда эффективных многомерных статистических методов от эконометрики обедняет арсенал эконометриста, сужает возможности прикладных эконометрических исследований».

Приходится говорить о существовании в мировой исследовательской и педагогической практике **тенденции обеднения математического аппарата эконометрики**, выражающейся в недооценке и урезанности традиционно включаемого в этот аппарат инструментария многомерного статистического анализа (МСА). Эта тенденция, к сожалению, исторически узаконена принятыми в мировой педагогической и, в несколько меньшей степени, исследовательской литературе стандартами. В результате, такие актуальнейшие для эконометрического анализа разделы МСА как *дискриминантный анализ и кластерный анализ; модели и методы снижения размерности анализируемых многомерных наблюдений (включая главные компоненты, многомерное шкалирование и т.п.)* оказываются как бы на нелегальном положении среди традиционного математического инструментария эконометрики.

Приведем лишь несколько примеров задач эконометрического анализа, *полноценное* решение которых без привлечения упомянутых выше методов МСА мне представляется объективно невозможным.

Задача 1. Оценка функции регрессии по регрессионно-неоднородным исходным данным.

Это одна из классических задач эконометрики, но в рамках традиционного эконометрического инструментария её решение рассматривается только при условии, что моменты возможного нарушения регрессионной однородности наблюдений *известны исследователю* (при этом рекомендуется использовать технику введения так называемых фиктивных переменных). Однако в практике эконометрического анализа весьма широко распространены ситуации, когда исследователю предоставляются данные пассивного эксперимента *без какой-либо информации о моментах возможного нарушения регрессионной однородности этих данных*. В этих случаях необходимым этапом в построении искомой функции регрессии является разделение имеющихся данных на регрессионно однородные подвыборки с помощью подходящих методов **кластер-анализа**, (см. п. 8.4) о чем вы ни слова не найдете ни в одном из канонических учебников по эконометрике.

Задача 2. Эконометрический анализ моделей бинарного и множественного выбора.

В рамках традиционного инструментария эконометрики этот класс задач предлагается решать с помощью некоторой параметризации и оценки *апостериорных вероятностей* $p_j(X_i) = P(y_i = j | X = X_i)$ принадлежности объекта i к классу j ($j = 1, 2, \dots, k$) при условии, что зна-

чения характеризующих его объясняющих переменных $X = (x^{(1)}, \dots, x^{(m)})^T$ зафиксированы на уровнях $X_i = (x_i^{(1)}, \dots, x_i^{(m)})^T$. При этом и в педагогической литературе, и в исследовательской практике совершенно игнорируется возможность решения подобных задач с помощью разнообразных методов **дискриминантного анализа**, которые в одних обстоятельствах могут оказаться более эффективными, а в других — *просто функционально связанными с традиционными эконометрическими* (см. п. 9.3 в данном учебнике).

Задача 3. Измерение и анализ синтетических (латентных) категорий социально-экономического развития (качества жизни населения, уровня институционального развития или инвестиционной привлекательности региона и т.п.).

Если следовать общепринятому определению эконометрики, то прикладной эконометрический анализ — это эмпирическое исследование, основанное на синтезе экономической теории, экономических измерений и соответствующего математического инструментария. Именно так можно охарактеризовать исследования, посвященные измерению и анализу синтетических (латентных) категорий социально-экономического развития страны или отдельного региона, таких как

- *качество населения;*
- *уровень материального благосостояния;*
- *качество социальной сферы;*
- *уровень институционального развития;*
- *уровень инвестиционной привлекательности;*
- *уровень взяточничества и коррупции и т.д.*

Однако вы тщетно будете искать в канонических учебниках по эконометрике разделы, посвященные методам решения подобных задач. А среди инструментария, который приходится привлекать для решения этих задач, *методы главных компонент, факторного анализа, экстремальной группировки признаков, многомерного шкалирования и т.п.* Но все эти методы до сих пор не получили официального (легального) статуса в рамках дисциплины «эконометрика»!

Задача 4. Реализация двухшагового метода наименьших квадратов (2МНК) при идентификации систем одновременных уравнений с помощью метода главных компонент.

Примеры эффективного использования метода главных компонент при идентификации систем одновременных уравнений большой размерности были продемонстрированы еще в 1960-х годах Клоком и Меннесом (1960), Амения (1966), Клейном (1969)². Однако учебная литература по эконометрике до сих пор никак не среагировала на этот факт.

Итак, границы, отделяющие методы эконометрики, традиционно включаемые в канонические западные (а теперь и наши отечественные) учебники, от более широкого круга методов МСА, являются формально жестко очерченными, а по существу — весьма размытыми и условными. Неоправданное отлучение ряда эффективных многомерных статистических методов от эконометрики обедняет арсенал эконометриста, сужает возможности прикладных эконометрических исследований. Частичное объективное объяснение подобной (ненормальной, с позиций сегодняшнего дня) ситуации можно дать следующее. Становление дисциплины «эконометрика» относят к 20-м — 30-м годам прошлого столетия. В то время многомерный статистический анализ был *сугубо теоретическим* разделом математической статистики. Лишь за последние 30–40 лет он, благодаря бурному развитию вычислительных мощностей, превратился в эффективный прикладной инструмент исследования.

1.3 Место эконометрики в ряду математико-статистических и экономических дисциплин

Резюмируя сказанное в п. 1.1, мы приходим к следующему определению эконометрики.

Эконометрика — это самостоятельная научная дисциплина, объединяющая совокупность теоретических результатов, приемов, методов и моделей, предназначенных для того, чтобы на базе

- экономической теории,
- экономической статистики,
- математико-статистического инструментария

² Kloeck T. and Mennes L. (1960). Simultaneous Equation Estimation Based on Principal Components of Predetermined Variables. — “Econometrica”, vol. 28, p.45–61.

Ameniya T. (1966). On the Use of Principal Components of Independent Variables in 2SLS Estimation. — “Intern. Econ. Rev.”, vol.7, p.283–303.

Klein L. (1969). Estimation of Interdependent Systems in Macro-econometrics. — “Econometrica”, vol.37, p.171–192.

придавать конкретное количественное выражение общим (качественным) закономерностям экономической теории.

Именно это понимание и содержание эконометрики отражают сложившиеся к настоящему времени в рамках этой научной дисциплины институты и издания (международные научные общества; отделения, факультеты и кафедры в университетах; конференции; монографии, учебники, журналы и т.п.).

Таким образом, в соответствии с данным выше определением суть эконометрики — именно в синтезе экономической теории, экономических измерений и математики. Но при этом, говоря об **экономической теории** в рамках эконометрики, мы будем интересоваться не просто выявлением объективно существующих (на качественном уровне) экономических законов и связей между экономическими показателями, но и подходами к их формализации, включающими в себя *методы спецификации соответствующих моделей с учетом проблемы их идентифицируемости* то есть выяснение наличия принципиальной возможности статистики оценить неизвестные параметры анализируемой модели по имеющимся экономическим измерениям (обсуждение понятий спецификации и идентифицируемости модели см. ниже, в п. 1.4.3). При рассмотрении **экономических измерений** как составной части эконометрики нас будет интересовать лишь тот аспект, который непосредственно связан с *информационным обеспечением* анализируемой эконометрической модели, хотя в этих рамках специалисту по эконометрике зачастую приходится решать полный спектр соответствующих задач: выбор необходимых экономических показателей и обоснование способа их измерения, определение плана статистического обследования и т.п. И наконец, под **математико-статистическим инструментарием** эконометрики подразумевается, естественно, не математическая статистика в традиционном ее понимании, а лишь отдельные ее разделы (такие, как классическая и обобщенная линейные модели регрессионного анализа, анализ временных рядов, построение и анализ систем одновременных уравнений), снабженные определенными акцентами и дополненные некоторыми специальными сведениями (специальные типы моделей регрессии и временных рядов, подходы к решению проблем спецификации и идентифицируемости моделей и т.п.).

Именно «приземление» экономической теории на базу конкретной экономической статистики и извлечение из этого приземления с помощью подходящего математического аппарата *вполне определенных количественных взаимосвязей* являются ключевыми моментами в понимании сущности эконометрики, обеспечивают разграничение эконометрики с такими дисциплинами, как математическая экономика, описательная экономическая статистика и математическая статистика. Так, математическая экономика, которая на самом деле является математически сформулированной экономической теорией, изучает взаимосвязи меж-

ду экономическими переменными на общем (неколичественном) уровне. Она становится эконометрикой, когда символически представленные в этих взаимосвязях коэффициенты заменяются конкретными численными оценками, полученными на базе соответствующих экономических данных.

Из определения эконометрики следует, что ее происхождение и главное назначение — это экономические и социально-экономические приложения, а именно модельное описание *конкретных количественных взаимосвязей*, существующих между анализируемыми показателями.

При всем разнообразии спектра решаемых с помощью эконометрики задач их, тем не менее, было бы удобно расклассифицировать по трем параметрам: по конечным прикладным целям, по уровню иерархии и по профилю анализируемой экономической системы.

По конечным прикладным целям выделим две основные: (а) *прогноз* экономических и социально-экономических показателей (переменных), характеризующих состояние и развитие анализируемой системы; (б) *имитация* различных возможных сценариев социально-экономического развития анализируемой системы, когда статистически выявленные взаимосвязи между характеристиками производства, потребления, социальной и финансовой политики и т.п. используются для проследивания того, как планируемые (возможные) изменения тех или иных поддающихся управлению параметров производства или распределения скажутся на значениях интересующих нас «выходных» характеристик.

По уровню иерархии анализируемой экономической системы выделяются *макроуровень* (то есть страны в целом), *мезоуровень* (регионы, отрасли, корпорации) и *микроуровень* (индивидуумы, семьи, предприятия, фирмы).

В некоторых случаях должен быть определен **профиль** эконометрического моделирования: исследование может быть сконцентрировано на проблемах рынка, инвестиционной, финансовой или социальной политики, ценообразования, распределительных отношений, спроса и потребления, или на определенном *комплексе* проблем. Однако чем претенциознее по широте охвата анализируемых проблем эконометрическое исследование, тем меньше шансов провести его достаточно эффективно.

Представленная на рис. 1.1 схема, конечно, условна. Однако, мы надеемся, она поможет читателю лучше понять нашу точку зрения на эконометрику и на ее место в ряду математико-статистических и экономических дисциплин.

Заметим, что название нашей книги подчеркивает то обстоятельство, что акценты в ней сделаны *именно на методах*, в то время как приложения присутствуют в ней, как правило, лишь в виде иллюстративных примеров (исключения составляют главы, посвященные эконометрическому анализу финансовых данных в задачах управления рисками).

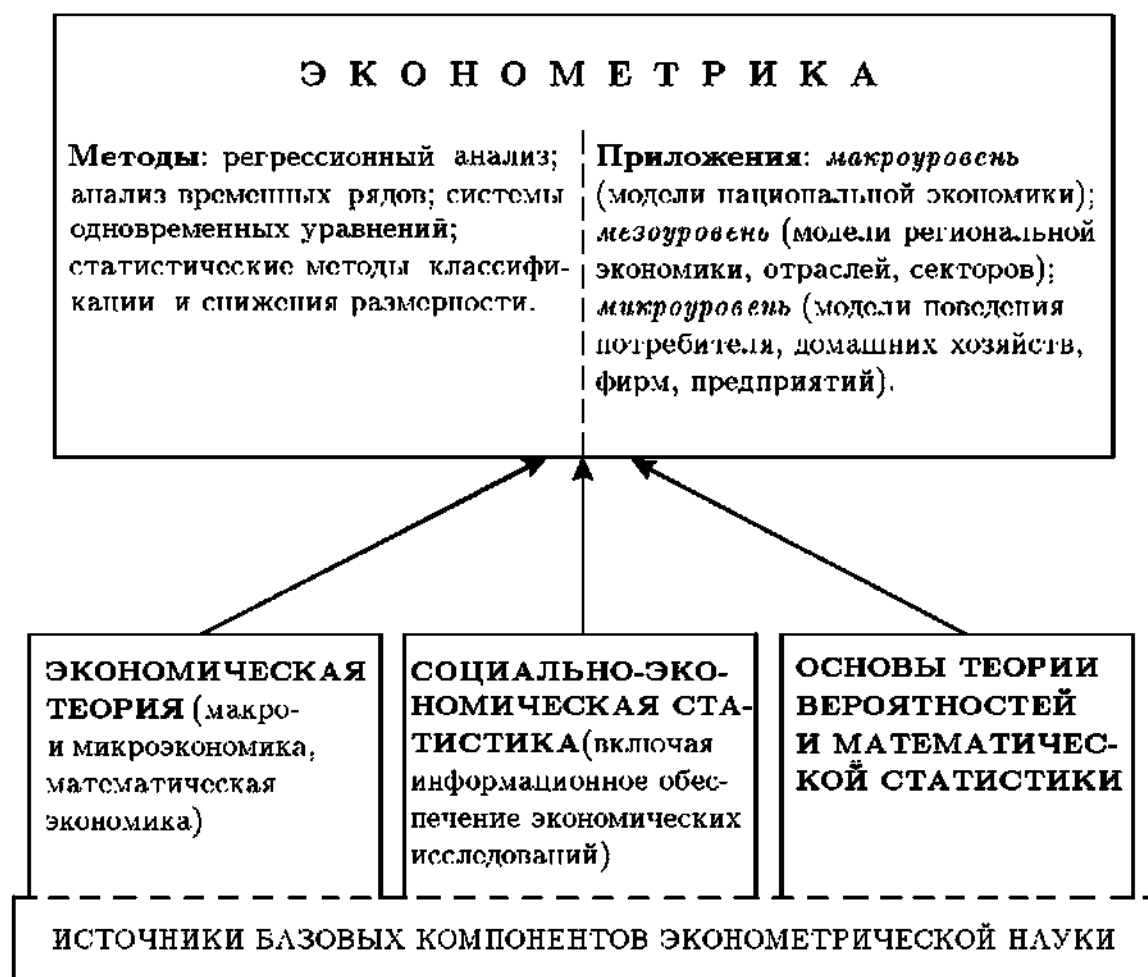


Рис. 1.1. Эконометрика и ее место в ряду других экономических и статистических дисциплин

1.4 Эконометрическая модель и проблемы эконометрического моделирования

1.4.1 От простых количественных взаимосвязей между экономическими переменными к эконометрической модели

Первая же принципиальная идея, с которой встречается каждый изучающий экономику — это *идея о взаимосвязях между экономическими переменными*. Формирующийся на рынке спрос на некоторый товар рассматривается как функция его цены; затраты, связанные с изготовлением какого-либо продукта, предполагаются зависящими от объема производства; потребительские расходы могут быть функцией дохода и т.д. Все это примеры связей между двумя переменными, одна из которых (спрос на товар, производственные затраты, потребительские расходы) играет роль *объясняемой переменной* (или *результатирующего*

показателя), а другие интерпретируются как *объясняющие переменные* (или *факторы-аргументы*). Однако для большей реалистичности в каждое такое соотношение приходится вводить *несколько* объясняющих переменных и остаточную случайную составляющую, отражающую влияние на результирующий показатель всех неучтенных факторов. Спрос на товар можно рассматривать как функцию его цены, потребительского дохода и цен на конкурирующие и дополняющие товары; производственные затраты будут зависеть от объема производства, от его динамики и от цен на основные производственные ресурсы; потребительские расходы можно определить как функцию дохода, ликвидных активов и предыдущего уровня потребления. При этом участвующая в каждом из этих соотношений случайная составляющая, отражающая влияние на анализируемый *результатирующий показатель* всех неучтенных факторов, обуславливает *стохастический характер* зависимости, а именно: даже зафиксировав на определенных уровнях значения объясняющих переменных, скажем, цены на сам товар и на конкурирующие с ним или дополняющие товары, а также потребительский доход, мы не можем ожидать, что тем самым *однозначно* определяется спрос на этот товар. Другими словами, переходя в своих наблюдениях спроса от одного временного или пространственного такта к другому, мы обнаружим *случайное варьирование величины спроса около некоторого уровня даже при сохранении значений всех объясняющих переменных неизменными*.

В прикладном статистическом анализе анализируются различные варианты формализации понятия **стохастической зависимости** между результирующим показателем y и объясняющими переменными $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ (см. ниже, п. 2.4).

Наиболее распространенной в эконометрических приложениях формой представления стохастической зависимости является *аддитивная линейная форма*, которая и будет главным предметом исследования в нашем изложении:

$$y_t = \theta_0 + \theta_1 x_t^{(1)} + \dots + \theta_p x_t^{(p)} + \delta_t. \quad (1.1)$$

Здесь y_t — значение результирующей (объясняемой) переменной, измеренное в t -м временном (или пространственном) такте, $x_t^{(1)}, x_t^{(2)}, \dots, x_t^{(p)}$ — значения участвующих в соотношении объясняющих переменных, полученные в том же t -м измерении, $\theta_0, \theta_1, \dots, \theta_p$ — некоторые параметры (как правило, не известные до проведения соответствующего статистического анализа), а δ_t — случайная составляющая, характеризующая *разницу между модельным и наблюдаемым значениями* анализируемой результирующей переменной, зафиксированную в t -м измерении и отражающая результат влияния на y_t факторов, не учтенных в рамках объясняющих переменных $x_t^{(1)}, x_t^{(2)}, \dots, x_t^{(p)}$. Под модельным значением результирующей переменной $\tilde{y}_t$ здесь и в дальнейшем мы будем понимать

ее значение, восстановленное по заданным величинам объясняющих переменных при условии, что коэффициенты $\theta_0, \theta_1, \dots, \theta_p$ нам известны, то есть

$$\tilde{y}_t = \theta_0 + \theta_1 x_t^{(1)} + \dots + \theta_p x_t^{(p)}. \quad (1.2)$$

При такой интерпретации модельного значения результирующей переменной случайную составляющую δ можно интерпретировать как **случайную ошибку прогноза y по заданным значениям $x^{(1)}, x^{(2)}, \dots, x^{(p)}$** , причем, чтобы исключить *систематическую* ошибку в оценке y_t по $\tilde{y}_t$, обычно полагают, что среднее значение случайной составляющей δ_t при всех значениях t равно нулю (то есть $E\delta_t \equiv 0$). Очевидно, чем больше информации заключено в значениях объясняющих переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ относительно величины y , тем надежнее будет прогноз и тем меньше будет ошибка прогноза δ . *А что значит малость случайной величины?* Это значит, что ее значения сосредоточены в окрестности нуля с малой дисперсией.

Следующий шаг в формализации положений экономической теории состоит в *группировке отдельных соотношений в модель*. Всякая математическая модель является лишь *упрощенным* формализованным представлением реального объекта (явления, процесса), и искусство ее построения состоит в том, чтобы совместить как можно большую лаконичность параметризации модели с достаточной адекватностью описания именно тех сторон моделируемой реальности, которые интересуют исследователя. Количество связей, включаемых в экономическую модель, зависит от условий, при которых эта модель конструируется, и от подробности объяснения, к которой мы стремимся. Например, традиционная модель спроса и предложения должна объяснять соотношения между ценой и объемом выпуска, характерные для некоторого определенного рынка. Она содержит три уравнения, а именно: уравнение спроса, уравнение предложения и уравнение реакции рынка. В эти уравнения, помимо интересующих нас объема выпуска и цены, будут входить и другие переменные; так, например, в уравнение спроса войдет потребительский доход, а в уравнение предложения — цена. Объяснение, достигнутое с помощью такой модели, обусловлено значениями некоторых «внешних» по отношению к модели переменных, и в этом смысле модель является *неполной*, или *условной*. Более претенциозные модели содержат гораздо больше уравнений, и с их помощью пытаются отразить поведение существенно большего числа переменных; однако и они остаются условными, поскольку тоже содержат переменные, не определяемые или не объясняемые моделью.

Все экономические модели, независимо от того, относятся они ко всему хозяйству или к его элементам (то есть к макроэкономике, отрасли, фирме или рынку), имеют некоторые общие особенности. Во-первых, они основаны на предположении, что поведение экономических переменных

определяется с помощью совместных и одновременных операций с некоторым числом экономических соотношений. Во-вторых, принимается гипотеза, в силу которой модель, допуская упрощение сложной действительности, тем не менее улавливает главные характеристики изучаемого объекта. В-третьих, создатель модели полагает, что на основе достигнутого с ее помощью понимания реальной системы удастся предсказать ее будущее движение и, возможно, управлять им в целях улучшения экономического благосостояния.

Чтобы проиллюстрировать сказанное и наметить пути для выяснения специфической роли эконометрики, рассмотрим пример весьма общей и приближенной макромоделли.

Пример 1.1. Предположим, что экономист-теоретик сформулировал следующие положения:

- потребление есть возрастающая функция от имеющегося в наличии дохода, но возрастающая, видимо, медленнее, чем рост дохода;
- объем инвестиций есть возрастающая функция национального дохода и убывающая функция характеристики государственного регулирования (например, нормы процента);
- национальный доход есть сумма потребительских, инвестиционных и государственных закупок товаров и услуг.

Наша первая задача — перевести эти положения на математический язык. И тут мы немедленно сталкиваемся с многообразием открывающихся перед нами возможных способов формализованного описания сформулированных априорных требований теоретика. Какие соотношения выбрать между переменными — линейные или нелинейные? Если остановиться на пелинейных, то какими они должны быть — логарифмическими, полиномиальными или какими-либо еще? Даже определив форму конкретного соотношения, мы оставляем еще нерешенной проблему выбора для различных уравнений запаздываний по времени. Будут ли, например, инвестиции текущего периода реагировать только на национальный доход, произведенный в последнем периоде, или же на них скажется динамика нескольких предыдущих периодов? Обычный выход из этих трудностей состоит в выборе при первоначальном анализе *наиболее простой* из возможных форм этих соотношений. Тогда появляется возможность записать на основе указанных выше положений следующую линейную относительно анализируемых переменных и аддитивную относительно случайных составляющих модель:

$$y_t^{(1)} = \alpha_0 + \alpha_1(y_t^{(3)} - x_t^{(1)}) + \delta_t^{(1)}, \quad (1.3)$$

$$y_t^{(2)} = \beta_1 y_{t-1}^{(3)} + \beta_2 \cdot x_t^{(2)} + \delta_t^{(2)}, \quad (1.4)$$

$$y_t^{(3)} = y_t^{(1)} + y_t^{(2)} + x_t^{(3)}, \quad (1.5)$$

где априорные ограничения выражены неравенствами

$$0 < \alpha_1 < 1; \quad \beta_1 > 0; \quad \beta_2 < 0.$$

Эти три соотношения вместе с ограничениями образуют модель. В ней $y_t^{(1)}$ обозначает потребление, $y_t^{(2)}$ — инвестиции, $y_t^{(3)}$ — национальный доход, $x_t^{(1)}$ — подоходный налог, $x_t^{(2)}$ — норму процента как инструмент государственного регулирования, $x_t^{(3)}$ — государственные закупки товаров и услуг, измеренные в момент времени t .

Присутствие в уравнениях (1.3) и (1.4) «остаточных» случайных составляющих $\delta_t^{(1)}$ и $\delta_t^{(2)}$ обусловлено необходимостью учесть влияние соответственно на $y_t^{(1)}$ и $y_t^{(2)}$ ряда неучтенных факторов. Действительно, нереалистично ожидать, что величина потребления $y_t^{(1)}$ будет *однозначно* определяться уровнями национального дохода ($y_t^{(3)}$) и подоходного налога ($x_t^{(1)}$); аналогично величина инвестиций $y_t^{(2)}$ зависит, очевидно, *не только* от достигнутого в предыдущий год уровня национального дохода ($y_{t-1}^{(3)}$) и от величины нормы процента ($x_t^{(2)}$), но и от ряда не учтенных в уравнении (1.4) факторов.

Полученная модель содержит два уравнения, объясняющие поведение потребителей и инвесторов, и одно тождество. Мы сформулировали ее для дискретных периодов времени и выбрали запаздывание (лаг) в один период для отражения воздействия национального дохода на инвестиции.

Мы привели здесь этот пример, чтобы пояснить общие черты одного из важнейших этапов эконометрического моделирования, в процессе которого исследователь математически формализует отдельные положения экономической теории и объединяет их в систему. В дальнейшем мы используем этот пример для пояснения ряда основных понятий эконометрического моделирования.

1.4.2 Основные понятия эконометрического моделирования

В любой эконометрической модели в зависимости от конечных прикладных целей ее использования все участвующие в ней переменные подразделяются на:

экзогенные, то есть задаваемые как бы «извне», автономно, в определенной степени управляемые (планируемые);

эндогенные, то есть такие переменные, значения которых формируются в процессе и *внутри* функционирования анализируемой социально-экономической системы в существенной мере под воздействием экзогенных переменных и, конечно, во взаимодействии друг с другом; в эконометрической модели они являются предметом объяснения;

предопределенные, то есть выступающие в системе в роли *факторов-аргументов*, или *объясняющих* переменных.

Из данных выше определений следует, что множество предопределенных переменных формируется из всех экзогенных переменных (которые могут быть «привязаны» к прошлым, текущему или будущим моментам времени) и так называемых *лаговых эндогенных переменных*, то есть таких эндогенных переменных, значения которых входят в уравнения анализируемой эконометрической системы измеренными в *прошлые* (по отношению к текущему) моменты времени, а следовательно, являются *уже известными, заданными*.

В нашем примере (1.3)–(1.5) из п. 1.4.1 потребление ($y_t^{(1)}$), инвестиции ($y_t^{(2)}$) и национальный доход ($y_t^{(3)}$) в текущий момент времени t являются *эндогенными* переменными; подоходный налог ($x_t^{(1)}$), норма процента как инструмент государственного регулирования ($x_t^{(2)}$) и государственные закупки товаров и услуг ($x_t^{(3)}$) — *экзогенные* переменные, которые вместе с национальным доходом в предшествующий момент времени ($y_{t-1}^{(3)}$) образуют множество *предопределенных* переменных.

Таким образом, можно сказать, что *эконометрическая модель служит для объяснения поведения эндогенных переменных в зависимости от значений экзогенных и лаговых эндогенных переменных*.

1.4.3 Основные проблемы эконометрического моделирования (общие формулировки)

Для пояснения сущности именно *эконометрической модели* и описания основных возникающих при ее построении и анализе проблем нам будет удобно разбить весь процесс моделирования на **шесть основных этапов**.

1-й этап (постановочный) — определение конечных целей моделирования, набора участвующих в модели факторов и показателей, их роли;

2-й этап (априорный) — предмодельный анализ экономической сущности изучаемого явления, формирование и формализация априорной информации, в частности, относящейся к природе и генезису исходных статистических данных и случайных остаточных составляющих;

3-й этап (параметризация) — собственно моделирование, то есть выбор **общего** вида модели, в том числе состава и формы входящих в нее связей;

4-й этап (информационный) — сбор необходимой статистической информации, то есть регистрация значений участвующих в модели факторов и показателей на различных временных или пространственных тактах функционирования изучаемого явления;

5-й этап (идентификация модели) — статистический анализ модели, и в первую очередь статистическое оценивание неизвестных параметров

модели;

6-й этап (верификация модели) — сопоставление реальных и модельных данных, проверка адекватности модели, оценка точности модельных данных.

Математическая модель, в том числе математическая модель *экономического* явления или процесса, может быть сформулирована на общем (качественном) уровне, без настройки на конкретные статистические данные, то есть она может иметь смысл и без 4-го и 5-го этапов. **Тогда она не является эконометрической.** Суть именно эконометрической модели заключается в том, что она, будучи представленной в виде набора математических соотношений, описывает функционирование **конкретной** экономической системы, а не системы вообще (именно экономики России или процесса «спрос — предложение» в данном конкретном месте и в данное время). Поэтому она обязательно «настраивается» на конкретных статистических данных, а значит предусматривает обязательную реализацию 4-го и 5-го этапов моделирования.

Обратимся теперь непосредственно к описанию основных проблем, которые приходится решать в процессе эконометрического моделирования.

Проблема спецификации модели. Эта проблема по существу решается на первых трех этапах моделирования и включает в себя:³

- (а) определение конечных целей моделирования (прогноз, имитация различных сценариев социально-экономического развития анализируемой системы, управление);
- (б) определение списка экзогенных и эндогенных переменных;
- (в) определение состава анализируемой системы уравнений и тождеств, их структуры и соответственно списка предопределенных переменных;
- (г) формулировку исходных предпосылок и априорных ограничений относительно:
 - стохастической природы остатков Δ_t (в классических вариантах моделей постулируются их взаимная статистическая независимость или некоррелированность, нулевые значения их средних величин и, иногда, сохранение постоянными в процессе наблюдения значений их дисперсий — *гомоскедастичность*);

³Ниже описывается содержание проблемы спецификации эконометрической модели в *предположении ее линейности*, то есть при условии, что в результате проведения 3-го этапа (этапа *параметризации или выбора общего вида модели*) эконометрист пришел к выводу о непротиворечивости гипотезы линейности анализируемых связей имеющимся в его распоряжении исходным статистическим данным.

- числовых значений отдельных параметров модели.

Итак, спецификация модели — это первый и, быть может, *важнейший* шаг эконометрического исследования. От того, насколько удачно решена проблема спецификации и, в частности, насколько реалистичны наши решения и предположения относительно состава эндогенных, экзогенных и предопределенных переменных, структуры самой системы уравнений и тождеств, стохастической природы случайных остатков и конкретных числовых значений части параметров модели, решающим образом зависит успех всего эконометрического исследования.

Спецификация опирается на имеющиеся экономические теории, специальные знания или на интуитивные представления исследователя об анализируемой экономической системе.

Проиллюстрируем сказанное на нашем примере из п. 1.4.1. Мы видим, что непосредственно из состава и смысла уравнений системы (1.3)–(1.5) (то есть из решения части вопросов проблемы спецификации модели) непосредственно следует, что нам предстоит статистически оценить лишь четыре параметра: $\alpha_0, \alpha_1, \beta_1$ и β_2).

Априорные сведения о системе находят свое отражение при спецификации модели не только в определении числа и смысла оцениваемых параметров, но и в выборе предположений относительно стохастической природы участвующих в уравнении переменных, в первую очередь относительно случайных остатков $\delta_t^{(j)}$. Обычно принимаются допущения о том, что все случайные остатки $\delta_t^{(j)}$ ($j = 1, 2, \dots, m_1$):

- имеют нулевые средние значения, то есть $E\delta_t^{(j)} \equiv 0, t = 1, \dots, n$;
- не коррелируют друг с другом, то есть $E(\delta_t^{(j)} \cdot \delta_t^{(l)}) = 0, (j \neq l)$;
- не имеют автокорреляций, то есть $E(\delta_{t_1}^{(j)} \cdot \delta_{t_2}^{(j)}) = 0, t_1 \neq t_2$;
- не коррелируют ни с одной из предопределенных переменных.

Как правило, подобные допущения, постулируемые в процессе спецификации модели, оказываются достаточно реалистичными.

Проблема идентифицируемости. Как было сказано выше, речь идет о необходимости получить ответ на вопрос: *существует ли принципиальная возможность состоятельно оценить неизвестные параметры модели по имеющимся у нас наблюдениям?*

Подробное рассмотрение проблемы идентифицируемости эконометрической модели, включающее в себя, в частности, *формулировки необходимых условий идентифицируемости отдельного параметра, целого уравнения и всей анализируемой системы уравнений*, можно найти, например, в [Айвазян (2001)], гл. 4.

Проблема идентификации. Решение этой проблемы предусматривает «настройку» анализируемой модели на реальные статистические

данные. Другими словами, речь идет о выборе и реализации методов статистического оценивания неизвестных параметров модели (1.3)–(1.5) по исходным статистическим данным. Описание необходимых методов статистического оценивания параметров в подобных моделях и в их отдельных фрагментах приводится в главах 4, 5, 7–10, а также в главе 4 учебника [Айвазян (2001)].

Проблема верификации модели. Эта проблема, так же как и проблема идентификации, является специфичной, связанной с построением именно *эконометрической* модели. Собственно построение эконометрической модели завершается ее идентификацией, то есть статистическим оцениванием участвующих в ней неизвестных коэффициентов (параметров). После этого, однако, возникают вопросы: (а) насколько удачно удалось решить проблемы спецификации, идентифицируемости и идентификации модели, то есть можно ли рассчитывать на то, что использование построенной модели в целях прогноза эндогенных переменных и в имитационных расчетах, определяющих варианты социально-экономического развития анализируемой системы, даст результаты, достаточно адекватные реальной действительности? (б) какова точность (абсолютная, относительная) прогнозных и имитационных расчетов, основанных на построенной модели? *Получение ответов на эти вопросы с помощью тех или иных математико-статистических методов и составляет содержание проблемы верификации эконометрической модели.*

Методы верификации модели описаны в главах 6, 10. Здесь же отметим лишь, что эти методы основаны на процедурах статистической проверки гипотез (при ответе на вопрос (а)) и на статистическом анализе характеристик точности различных приемов статистического оценивания параметров (при ответе на вопрос (б)). Отметим также, что наиболее распространенным и эффективным подходом к верификации эконометрической модели можно признать принцип так называемых *ретроспективных расчетов*.

Выводы

1. *Эконометрика* — самостоятельная экономико-математическая научная дисциплина, позволяющая на базе положений экономической теории и исходных данных экономической статистики, используя необходимый математико-статистический инструментарий, придавать конкретное количественное выражение общим (качественным) закономерностям, обусловленным экономической теорией.

2. Эконометрическое моделирование реальных социально-экономических процессов и систем обычно преследует *конечные прикладные цели двух типов* (или один из них): (а) *прогноз* экономических и социаль-

но-экономических показателей, характеризующих состояние и развитие анализируемой системы; (б) *имитацию* различных возможных сценариев социально-экономического развития анализируемой системы (многовариантные сценарные расчеты, ситуационное моделирование).

3. При постановке задач эконометрического моделирования следует определить их *иерархический уровень* и *профиль*. Анализируемые задачи могут относиться к *макро* (страна, межстрановой анализ), *мезо* (регионы внутри страны) и *микро* (индивидуумы, семьи, предприятия, фирмы) уровням и быть направленными на решение вопросов различного профиля инвестиционной, финансовой или социальной политики, ценообразования, распределительных отношений и т.п.

4. Эконометрическая модель содержит набор уравнений регрессионного типа, описывающих исследуемые стохастические связи между анализируемыми экономическими показателями, а также какое-то количество связывающих эти показатели тождеств, которые определяются экономическим смыслом проблемы. Наиболее распространенный математический вид исследуемых связей — *линейная* (относительно анализируемых переменных) и *аддитивная* формы. При этом возможны ситуации, когда одни и те же показатели в одних уравнениях модели играют роль объясняемых, а в других — объясняющих переменных (такие модели принято называть *системами одновременных уравнений*).

5. Эконометрическая модель служит для объяснения поведения *эндогенных* переменных в зависимости от значений *экзогенных* и *лаговых эндогенных* переменных. Под *экзогенными* переменными системы понимаются показатели, значения которых задаются как бы извне, автономно, они являются в определенной степени управляемыми или планируемыми. Значения *эндогенных* переменных формируются в процессе и внутри функционирования анализируемой социально-экономической системы в существенной мере под воздействием экзогенных переменных и во взаимодействии друг с другом. *Лаговые эндогенные* переменные — это такие эндогенные переменные, значения которых входят в уравнения анализируемой эконометрической модели измеренными в *прошлые моменты* времени, а следовательно, являются уже известными, заданными.

6. К основным проблемам эконометрического моделирования следует отнести:

- *спецификацию модели;*
- *выяснение факта ее идентифицируемости;*
- *идентификацию модели;*
- *верификацию модели.*

7. Содержание *основного математико-статистического инструментария* эконометрики в его традиционном понимании определяется

следующими семью разделами:

- (1) классическая модель регрессии и классический метод наименьших квадратов;
- (2) обобщенная линейная модель регрессии и обобщенный метод наименьших квадратов;
- (3) регрессионные модели с дискретной зависимой переменной (модели бинарного и множественного выбора);
- (4) регрессионные модели в условиях цензурирования, усечения или выборочной селективности (the sample selection) зависимой переменной;
- (5) статистический анализ одномерных и многомерных временных рядов.

Однако в современном контексте эти разделы должны быть дополнены широким спектром методов многомерного статистического анализа и, в первую очередь, методами и моделями распознавания социально-экономических образов, их типологизации и снижения размерности исследуемого факторного пространства.

Глава 2

Введение в регрессионный анализ

2.1 Общая формулировка проблемы статистического исследования зависимостей

Любой закон природы или общественного развития может быть выражен в конечном счете в виде описания характера или структуры взаимосвязей (зависимостей), существующих между изучаемыми явлениями или показателями (переменными величинами или просто *переменными*). Если эти зависимости: а) *стохастичны* по своей природе, то есть позволяют устанавливать лишь вероятностные логические соотношения между изучаемыми событиями A и B , а именно соотношения типа «из факта осуществления события A следует, что событие B должно произойти, но не обязательно, а лишь с некоторой (как правило, близкой к единице) вероятностью P »; б) выявляются на основании *статистического наблюдения* за анализируемыми событиями или переменными, осуществляемого по выборке из интересующей нас генеральной совокупности, — то мы оказываемся в рамках проблемы *статистического исследования зависимостей*. Соответствующий математический аппарат, будучи таким образом нацеленным в первую очередь на решение основной проблемы естествознания: как по отдельным, частным наблюдениям выявить и описать интересующую нас общую закономерность, — занимает, бесспорно, центральное место во всем прикладном математическом анализе.

Перед тем как перейти к формулировке общей и частных задач статистического исследования зависимостей, условимся описывать функционирование изучаемого реального объекта (системы, процесса, явления) набором переменных (рис. 2.1), среди которых:

$x^{(1)}, x^{(2)}, \dots, x^{(p)}$ — так называемые *входные* переменные, описывающие *условия функционирования* (часть из них, как правило, поддается

регулированию или частичному управлению); в соответствующих математических моделях их называют независимыми, факторами-аргументами, предикторными (или просто предикторами, то есть предсказателями), экзогенными, объясняющими (в книге мы будем использовать в основном два последних термина);

$y^{(1)}, y^{(2)}, \dots, y^{(m)}$ — *выходные* переменные, характеризующие *поведение или результат* (эффективность) *функционирования*; в математических моделях их называют зависимыми, откликами, эндогенными, результирующими или объясняемыми (в книге используются в основном три последних термина);

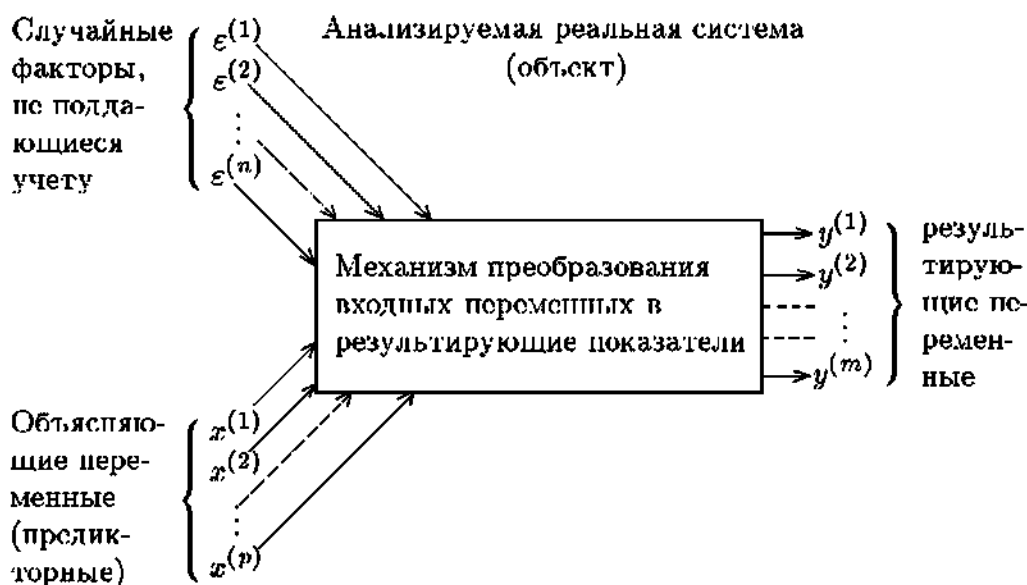


Рис. 2.1. Общая схема взаимодействия переменных при статистическом исследовании зависимостей

$\varepsilon^{(1)}, \varepsilon^{(2)}, \dots, \varepsilon^{(n)}$ — латентные (то есть скрытые, не поддающиеся непосредственному измерению) *случайные «остаточные» компоненты*, отражающие влияние (соответственно на $y^{(1)}, y^{(2)}, \dots, y^{(m)}$) не учтенных «на входе» факторов, а также случайные ошибки в измерении анализируемых показателей (в математических моделях мы их, как правило, будем именовать просто «остатками»).

Тогда общая задача статистического исследования зависимостей (в терминах изучаемых показателей) может быть сформулирована следующим образом:

по результатам n измерений

$$\{(x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)}; y_i^{(1)}, y_i^{(2)}, \dots, y_i^{(m)})\}_{i=1,2,\dots,n} \quad (2.1)$$

исследуемых переменных на объектах (системах, процессах) анализиру-

Таблица 2.1. Результаты обследования семей по среднедушевым доходам (x_i) и денежным сбережениям (y_i)

Средне- душевой доход (ден.ед.)	Средне- душевые сбереже- ния (ден.ед.)	Средние сбережения для семей данной группы (ден.ед.)	Оценка среднеквад- ратического отклонения $\bar{s}$ и коэффициента вари- ации $\hat{V}$ сбережений для семей данной доход- ной группы
1	2	3	4
$x_1 = x_2 =$ $\dots = x_{10} =$ $= x_1^0 = 80$	$y_1 = 15,2$ $y_2 = 10,7$ $y_3 = 18,5$ $y_4 = 14,9$ $y_5 = 24,1$ $y_6 = 10,3$ $y_7 = 14,2$ $y_8 = 31,0$ $y_9 = 20,4$ $y_{10} = 20,0$	$\bar{y}(x_1^0) =$ $= \frac{1}{10} \sum_{i=1}^{10} y_i = 17,9$	$\bar{s}(x_1^0) = \left[\frac{1}{9} \sum_{i=1}^{10} (y_i - \bar{y}(x_1^0))^2 \right]^{\frac{1}{2}} = 6,4$ $\hat{V}(x_1^0) = 36\%$
$x_{11} = x_{12} =$ $\dots = x_{20} =$ $= x_2^0 = 120$	$y_{11} = 70,1$ $y_{12} = 35,0$ $y_{13} = 43,0$ $y_{14} = 29,0$ $y_{15} = 17,0$ $y_{16} = 48,2$ $y_{17} = 18,9$ $y_{18} = 53,0$ $y_{19} = 39,4$ $y_{20} = 46,2$	$\bar{y}(x_2^0) =$ $= \frac{1}{10} \sum_{i=11}^{20} y_i = 40,0$	$\bar{s}(x_2^0) = \left[\frac{1}{9} \sum_{i=11}^{20} (y_i - \bar{y}(x_2^0))^2 \right]^{\frac{1}{2}} = 16,0$ $\hat{V}(x_2^0) = 40\%$
$x_{21} = x_{22} =$ $\dots = x_{30} =$ $= x_3^0 = 160$	$y_{21} = 49,6$ $y_{22} = 69,4$ $y_{23} = 77,8$ $y_{24} = 43,0$ $y_{25} = 31,8$ $y_{26} = 62,6$ $y_{27} = 100,2$ $y_{28} = 68,8$ $y_{29} = 78,0$ $y_{30} = 29,6$	$\bar{y}(x_3^0) =$ $= \frac{1}{10} \sum_{i=21}^{30} y_i = 61,1$	$\bar{s}(x_3^0) = \left[\frac{1}{9} \sum_{i=21}^{30} (y_i - \bar{y}(x_3^0))^2 \right]^{\frac{1}{2}} = 22,6$ $\hat{V}(x_3^0) = 37\%$

Продолжение таблицы 2.1

1	2	3	4
$x_{31} = x_{32} =$ $\dots = x_{40} =$ $= x_4^0 = 200$	$y_{31} = 125,5$ $y_{32} = 88,3$ $y_{33} = 62,0$ $y_{34} = 58,8$ $y_{35} = 84,0$ $y_{36} = 79,0$ $y_{37} = 95,5$ $y_{38} = 120,8$ $y_{39} = 98,1$ $y_{40} = 29,7$	$\bar{y}(x_4^0) =$ $= \frac{1}{10} \sum_{i=31}^{40} y_i = 84,2$	$\bar{s}(x_4^0) = \left[\frac{1}{9} \sum_{i=31}^{40} (y_i - \bar{y}(x_4^0))^2 \right]^{\frac{1}{2}} = 28,9$ $\hat{V}(x_4^0) = 34\%$

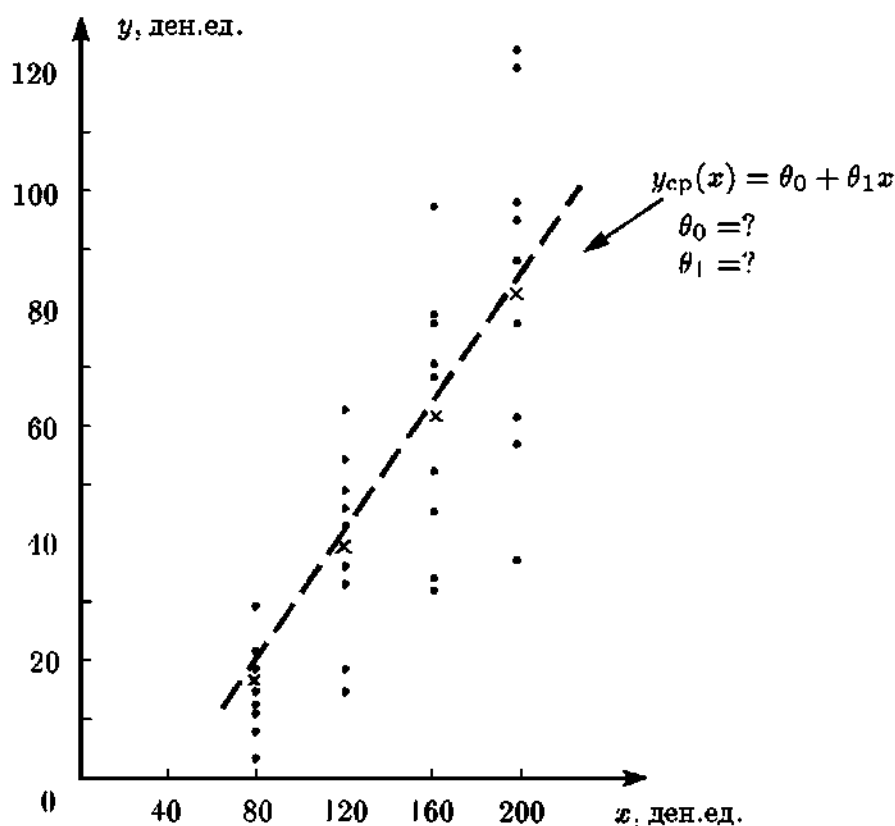


Рис. 2.2. Графическое представление результатов обследования 40 семей по их среднему доходу (x_i) и средним денежным сбережениям (y_i)

Мы видим, что даже в пределах каждой из этих групп величины средних сбережений семей подвержены некоторому неконтролируемому разбросу, обусловленному влиянием множества не поддающихся строгому учету и контролю факторов (то есть налицо упомянутый выше *стохастический характер* зависимости между x и y). Однако это

еще не значит, что расположение точек (x_i, y_i) , являющихся геометрическим изображением результатов обследования семей по доходу и сбережениям, должно быть совершенно хаотичным и не должно обнаруживать некоторой *вполне определенной тенденции*, характеризующей зависимость денежных сбережений в семье (η) от ее среднедушевого дохода (ξ). При исследовании подобных зависимостей встают следующие основные вопросы (в скобках после вопроса указываются главы и пункты настоящей книги, ему посвященные).

1. Как исходя из конкретных прикладных целей исследования определить смысл, в котором понимается исследуемая зависимость? (п. 2.2. и 2.4).

2. Имеется ли вообще какая-либо связь между исследуемыми переменными (а в случае многих переменных какова структура этих связей?) и как измерить тесноту этой связи? (глава 3).

3. Каков *общий* математический вид искомой связи между η и ξ , то есть как определяется *общая структура* соответствующей математической модели? (п. 2.5).

4. Как, отпавляясь от принятой общей структуры модели, провести необходимую вычислительную обработку исходных данных (2.1) с целью получения *конкретного вида* зависимости η от ξ , что позволит в данном случае производить количественную оценку неизвестных денежных сбережений семьи по заданной величине ее среднедушевого дохода? (главы 4–6).

5. Поскольку наши выводы основаны на обработке *ограниченного ряда* наблюдений, то их количественные характеристики, естественно, подвержены (при повторениях соответствующих выборочных обследований) некоторому случайному разбросу. Как оценить *степень точности наших выводов*? (глава 6).

6. Как решать сформулированные выше вопросы в ситуациях, когда среди объясняющих (предикторных) переменных могут быть и неколичественные? (глава 3).

Вернемся к нашему примеру и попробуем ответить на некоторые из поставленных здесь вопросов, в том числе на принципиальные вопросы (а), (б) и (в), ответы на которые позволяют уточнить общую формулировку задачи статистического исследования зависимостей, данную выше.

Начнем «с конца», то есть с уточнения *конечных прикладных исследований* (см. вопросы 1, а также (а) и (в)). Известно, что из двух анализируемых характеристик материальной состоятельности семьи характеристика денежных сбережений (η) относится к категории статистически труднодоступных. Поэтому главной конечной целью нашего исследования (опирающегося, как мы будем всегда предполагать, на достоверную и репрезентативную выборку исходных данных) является возможность восстановления (прогноза):

- удельной (то есть в расчете на одного члена семьи за определенный отрезок времени) величины денежных сбережений в *конкретной семье* ($y(x)$) по заданному значению ее среднедушевого дохода x ;
- удельной величины *средних* денежных сбережений ($y_{\text{ср}}(x)$) в семьях данной группы x по доходам.

Этой цели мы сможем достигнуть, если сумеем математически описать закономерность изменения условных теоретических средних значений $y_{\text{ср}}(x) = \mathbf{E}(\eta \mid \xi = x)$ в зависимости от x , а также изучить характер случайного разброса денежных сбережений $y(x)$ *отдельных* семей данной группы x по доходам относительно своего среднего значения $y_{\text{ср}}(x)$ (при любом интересующем нас значении среднедушевого дохода x). Это естественным образом приводит нас к необходимости рассмотрения математической модели вида

$$\eta = f(x) + \varepsilon, \quad (2.3)$$

в которой остаточная компонента ε отражает случайное отклонение денежных сбережений наугад выбранной отдельной семьи с доходом $\xi = x$ от среднего значения $y_{\text{ср}}(x) = \mathbf{E}(\eta \mid \xi = x)$ этих сбережений, подсчитанного по всем семьям данной доходной группы, а функция $f(x)$ описывает характер изменения условного среднего $y_{\text{ср}}(x)$ (при $\xi = x$) в зависимости от изменения x , если дополнительно прийти к соглашению, что характер случайного разброса величин $y(x) = (\eta \mid \xi = x)$ относительно своих средних $y_{\text{ср}}(x)$ таков, что $\mathbf{E}(\varepsilon \mid \xi = x) = 0$ при всех x .

Таким образом, из (2.3) мы непосредственно получаем

$$y_{\text{ср}} = \mathbf{E}(\eta \mid \xi = x) = f(x). \quad (2.4)$$

Чтобы покончить с вопросами 1, (а) и (в), остается уточнить общую структуру модели, то есть определить, в *каком классе $\mathbf{F}$ функций $f(x)$ мы будем производить аппроксимацию искомой зависимости $y_{\text{ср}}(x)$* .

В нашем случае, учитывая однородный (по характеру потребительского поведения) состав исследуемой совокупности семей, естественно исходить из гипотезы об одинаковой (в среднем) склонности семей к сбережениям, выражающейся, в частности, в том, что все семьи начиная с некоторого «порогового» уровня дохода склонны отделять в сбережения в среднем одинаковую долю дохода. Математически, как легко понять, это выразится в виде

$$y_{\text{ср}}(x) = \theta_0 + \theta_1 x, \quad (2.5)$$

где θ_0 и θ_1 — некоторые константы (*неизвестные параметры модели*)¹.

¹ Может показаться неправомерным присутствие в уравнении (2.5) свободного члена θ_0 . Однако удельные денежные сбережения в семье (на определенный отрезок времени) складываются из остатка θ_0 удельных сбережений от предыдущего времени (он может быть как положительным, так и отрицательным, то есть долгом) и определенной доли θ_1 дохода x за настоящий отрезок времени.

Так что

$$\mathbf{F} = \{\theta_0 + \theta_1 x\}, \quad (2.6)$$

где под $\{f(x; \Theta)\}$ понимается *семейство* всех тех функций $f(x; \Theta)$, которые могут быть получены при подстановке вместо Θ ее различных конкретных значений (Θ — векторный параметр; в нашем случае $\Theta = (\theta_0, \theta_1)^T$).

Такой выбор «класса допустимых решений» $\mathbf{F} = \{f(x)\}$ подтверждается и характером расположения совокупности точек, являющихся геометрическим изображением исходных данных в нашем примере (см. на рис. 2.2 расположение «крестиков», ординаты которых определяются экспериментально подсчитанными, то есть вычисленными на основании имеющихся выборочных данных, условными средними $\bar{y}(x_i^0)$, $i = 1, 2, 3, 4$).²

И наконец, следует уточнить, в соответствии с каким именно критерием качества аппроксимации неизвестных величин среднедушевых семейных денежных сбережений $y(x)$ и $y_{\text{ср}}(x)$ с помощью функции $\theta_0 + \theta_1 x$ мы будем определять наилучший способ прогноза $y_{\text{ср}}(x)$ по x . Наиболее обоснованное и точное решение этого вопроса опирается на знание вероятностной природы (а именно типа закона распределения вероятностей) остатков ε в модели (2.3). Так, например, известно (см. ниже п. 4.2.2), что если предположить, что при любых значениях x распределение вероятностей остатков ε описывается $(0, \sigma^2)$ -нормальным законом и что остатки $\varepsilon(x_i)$, $i = 1, 2, \dots, n$, характеризующие различные наблюдения, статистически независимы, то наименьшая ошибка прогноза $y_{\text{ср}}(x)$ с помощью модели $f(x) \in \mathbf{F}$ (то есть функция $f(x)$ подбирается из класса $\mathbf{F}$) обеспечивается требованием метода наименьших квадратов, в соответствии с которым оценки $\hat{\theta}_0$ и $\hat{\theta}_1$ параметров θ_0 и θ_1 определяются из условия минимизации по θ_0 и θ_1 выражения

$$\sum_{i=1}^n (y_i - \theta_0 - \theta_1 x_i)^2. \quad (2.7)$$

В нашем примере явно нарушено условие постоянства дисперсии остатков (см. табл. 2.1), то есть условная дисперсия $\mathbf{D}(\varepsilon \mid \xi = x) = \sigma^2(x)$ существенно зависит от значения x . Можно устранить это нарушение, поделив все анализируемые величины, откладываемые по оси η , а следовательно и остатки $\varepsilon(x)$, на значения $\bar{s}(x)$ (являющиеся несмещенными статистическими оценками для $\sigma(x)$), то есть перейдя к анализу остатков $\tilde{\varepsilon}(x) = \varepsilon(x)/\bar{s}(x)$. Тогда можно показать, что гипотеза о

² Обращаем внимание читателя на разницу в смысле и обозначениях экспериментальных (выборочных) и теоретических условных средних соответственно $\bar{y}(x)$ и $y_{\text{ср}}(x)$. Строго говоря, на практике теоретических средних мы никогда знать не можем, однако мы опираемся в своем исследовании на тот факт, что в соответствии с законом больших чисел $\bar{y}(x) \rightarrow y_{\text{ср}}(x)$ (по вероятности), когда число наблюдений, по которым подсчитано $\bar{y}(x)$, стремится к бесконечности.

$(0; \sigma^2)$ -нормальном характере распределения остатков $\tilde{\varepsilon}(x)$ не противоречит имеющимся в нашем распоряжении данным (представленным в табл. 2.1) и, следовательно, требование (2.7) приводит к необходимости решения экстремальной задачи вида

$$\Delta_n(\theta_0, \theta_1) = \sum_{i=1}^n \left(\frac{y_i - \theta_0 - \theta_1 x_i}{s(x_i)} \right)^2 \rightarrow \min_{\theta_0, \theta_1}, \quad (2.7')$$

то есть к системе из двух линейных уравнений с двумя неизвестными (θ_0 и θ_1):

$$\begin{aligned} \frac{\partial \Delta_n(\theta_0, \theta_1)}{\partial \theta_0} &= -2 \sum_{i=1}^n \bar{s}^{-2}(x_i) \cdot (y_i - \theta_0 - \theta_1 x_i) = 0; \\ \frac{\partial \Delta_n(\theta_0, \theta_1)}{\partial \theta_1} &= -2 \sum_{i=1}^n \bar{s}^{-2}(x_i) \cdot x_i \cdot (y_i - \theta_0 - \theta_1 x_i) = 0. \end{aligned} \quad (2.7'')$$

Решение системы (2.7'') дает нам в качестве оценок $\hat{\theta}_0$ и $\hat{\theta}_1$ для неизвестных параметров соответственно θ_0 и θ_1 выражения:

$$\begin{aligned} \hat{\theta}_1 &= \frac{\left(\sum_{i=1}^n \lambda_i \right) \left(\sum_{i=1}^n \lambda_i x_i y_i \right) - \left(\sum_{i=1}^n \lambda_i x_i \right) \left(\sum_{i=1}^n \lambda_i y_i \right)}{\left(\sum_{i=1}^n \lambda_i \right) \left(\sum_{i=1}^n \lambda_i x_i^2 \right) - \left(\sum_{i=1}^n \lambda_i x_i \right)^2}, \\ \hat{\theta}_0 &= \frac{\sum_{i=1}^n \lambda_i y_i}{\sum_{i=1}^n \lambda_i} - \hat{\theta}_1 \frac{\sum_{i=1}^n \lambda_i x_i}{\sum_{i=1}^n \lambda_i}, \end{aligned} \quad (2.7''')$$

где $\lambda_i = \bar{s}^{-2}(x_i)$.

Расчет по этим формулам с использованием данных табл. 2.1 дает нам ответ на сформулированный выше вопрос 4:

$$\begin{aligned} \hat{\theta}_1 &= 0,685; \\ \hat{\theta}_0 &= -40,360, \end{aligned}$$

так что статистическая оценка искомой зависимости средней величины среднедушевых семейных сбережений $y_{\text{ср}}(x)$ от значения среднедушевого дохода семей доходной группы x имеет в этом случае вид

$$\hat{y}_{\text{ср}}(x) = -40,36 + 0,685 \cdot x.$$

При другой статистической природе остатков ε или при отсутствии достаточной информации о типе их вероятностного распределения возможен иной, чем по (2.7), выбор критерия качества аппроксимации Δ_n .

Отметим, однако, что наиболее широкое распространение в статистической практике получили именно различные варианты критерия наименьших квадратов (2.7) (см. ниже, главы 4 и 5).

Заканчивая обсуждение примера 2.1 и возвращаясь к общему описанию задач статистического исследования зависимостей, отметим, что функции $f(X) = E(\eta | \xi = X)$, описывающие поведение условных средних результирующего показателя η (вычисленных при значениях объясняющих переменных ξ , зафиксированных на уровне $\xi = X$) в зависимости от изменения X , принято называть *функциями регрессии*³.

2.2 Какова конечная прикладная цель статистического исследования зависимостей?

С этого вопроса должно начинаться любое статистическое исследование зависимостей. Ведь от ответа на этот вопрос существенно зависят план исследования, выбор общей структуры математической модели, интерпретация получаемых статистических характеристик и выводов и т. д.

Итак, для чего же строятся математические модели типа (2.3), описывающие статистические зависимости между исследуемыми переменными: результирующими показателями $Y = (y^{(1)}, y^{(2)}, \dots, y^{(m)})$, с одной стороны, и соответствующими объясняющими (экзогенными) переменными $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$, с другой стороны?

Выделим три основных типа конечных прикладных целей подобных исследований, расположив их как бы по нарастанию глубины проникновения в содержательную сущность анализируемой конкретной задачи.

Тип 1. Установление самого факта наличия (или отсутствия) статистически значимой связи между Y и X . При такой постановке задачи статистический вывод имеет двоичную (альтернативную) природу — «связь есть» или «связи нет» — и сопровождается обычно лишь численной характеристикой (измерителем) степени тесноты исследуемой зависимости. Выбор *формы связи* (то есть класса допустимых решений F и конкретного вида функции $f(X)$ в модели (2.3)) и состава объясняющих переменных X играет подчиненную роль и нацелен исключительно на максимизацию величины этого измерителя степени тесноты связи: исследователю часто не приходится даже «добираться» до конкретного вида функции $f(X)$, и тем более он не претендует на анализ причинных

³ Этимология этого названия восходит к исследованию шведских статистиков (XIX век), изучавших на основе статистических данных связь, существующую между ростом отца и сына. Было установлено, что хотя и существует положительная зависимость между ростом отца и сына (чем выше отец, тем, в среднем, выше сын), однако наблюдается явление регрессии, возврата к среднему: у очень высоких отцов сыновья, хоть и высокие, но уже не такие как отцы, а у очень низкорослых отцов сыновья тоже невысокие, но все-таки в среднем повыше, чем отцы.

влияний переменных X на результирующие показатели. Решению задач, подчиненных прикладным целям типа 1, посвящена гл. 3.

Тип 2. *Прогноз (восстановление) неизвестных значений интересующих нас индивидуальных ($Y(X) = (\eta \mid \xi = X)$) или средних ($Y_{cp}(X) = E(\eta \mid \xi = X)$) значений исследуемых результирующих показателей по заданным значениям X соответствующих объясняющих переменных.* При такой постановке задачи статистический вывод включает в себя описание интервала (области) вероятных значений прогнозируемого показателя $Y_{cp}(X)$ или $Y(X)$ и сопровождается величиной доверительной вероятности P , с которой гарантируется справедливость нашего прогноза. Как и в предыдущем случае, выбор формы связи (то есть класса допустимых решений F и конкретного вида функции $f(X)$ в модели (2.3)) и состава предикторов объясняющих переменных X играет подчиненную роль и нацелен исключительно на минимизацию ошибки получаемого прогноза. Однако в данном случае (в отличие от предыдущего) исследователь *существенно использует* значения функции $f(X)$, которые являются отправной точкой при построении прогнозных доверительных интервалов. Последние обычно определяются в форме множества всех тех значений Y , которые удовлетворяют неравенствам

$$f(X) - \varepsilon_P(X, n) \leq Y \leq f(X) + \varepsilon_P(X, n), \quad (2.8)$$

где $\varepsilon_P(X, n)$ — гарантируемая (с вероятностью, не меньшей заданного значения P) максимальная величина ошибки прогноза.⁴ Таким образом, исследователя интересуют в данном случае *лишь значения* функции $f(X)$, но не ее структура, определяющая, в частности, соотношение удельных весов влияния объясняющих переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ на каждый из результирующих показателей $y^{(k)} (k = 1, 2, \dots, m)$. Так, например, если при статистическом оценивании неизвестной истинной зависимости

$$f(X) = y_{cp}(x^{(1)}, x^{(2)}) = 1 + 3x^{(1)} + 5x^{(2)} \quad (2.9)$$

исследователю удалось получить оценку функции $f(X)$ в виде

$$\hat{f}(X) = 1 + 6x^{(1)} - x^{(2)}, \quad (2.9')$$

и при этом было установлено, что объясняющие переменные $x^{(1)}$ и $x^{(2)}$ связаны между собой «почти функциональной» линейной зависимостью⁵

$$x^{(1)} \approx 2x^{(2)}, \quad (2.10)$$

⁴ Напоминаем читателю, что f, ε и Y являются m -мерными векторами (см. (2.2)), так что запись (2.8) означает справедливость m соответствующих покомпонентных неравенств.

⁵ Говоря о «почти функциональной» линейной зависимости между $x^{(1)}$ и $x^{(2)}$, мы имеем в виду близость к единице (по абсолютной величине) коэффициента корреляции между этими переменными.

то функция $\hat{f}(X)$ будет обладать хорошими прогностическими свойствами, несмотря на существенное отличие ее коэффициентов при $x^{(1)}$ и $x^{(2)}$ от соответствующих коэффициентов истинной функции $f(X)$. (Обращаем внимание читателя на тот факт, что коэффициенты при $x^{(2)}$ в функциях $f(X)$ и $\hat{f}(X)$ отличаются даже по знаку!) При подстановке заданных значений объясняющих переменных $x^{(1)}$ и $x^{(2)}$ в правые части (2.9) и (2.9'), при условии, что эти значения связаны приближенным соотношением (2.10), мы будем получать совпадающие (или приближенно совпадающие) результаты $f(X)$ и $\hat{f}(X)$, характеризующие усредненную величину $y_{\text{ср}}(X)$ исследуемого результирующего показателя.

Тип 3. *Выявление причинных связей между объясняющими переменными X и результирующими показателями Y , частичное управление значениями Y путем регулирования величин объясняющих переменных X .* Такая постановка задачи претендует на проникновение в «физический механизм» изучаемых статистических связей, то есть в тот самый механизм преобразования «входных» переменных X и ϵ в результирующие показатели Y (см. рис. 2.1), который в большинстве случаев исследователь, не будучи в состоянии его конструктивно описать, вынужден именовать (следуя сложившейся кибернетической терминологии) «черным ящиком».

И при выявлении причинных связей, и при намерении исследователя использовать модели типа (2.3) или (2.4) для управления значениями результирующих показателей $Y_{\text{ср}}(X)$ или $Y(X)$ путем регулирования величин объясняющих переменных X на первый план выходит *задача правильного определения структуры модели* (то есть выбора общего вида функции $f(X)$), решение которой обеспечивает возможность количественного измерения эффекта воздействия на $Y(X)$ каждой из объясняющих переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ *в отдельности*. Однако как раз это место (правильный выбор общего вида функции $f(X)$) и является самым слабым во всей технике статистического исследования зависимостей: к сожалению, не существует стандартных приемов и методов, которые образовывали бы строгую теоретическую базу для решения этой важнейшей задачи (некоторые рекомендации по проведению этого этапа исследования содержатся в п. 2.5).

Заметим, что исследователи, пожалуй, чаще других ставят перед собой именно цели типа 3. И в таких прикладных задачах, как *управление качеством продукции* с помощью регулирования хода технологических процессов, *прогноз и анализ объемов произведенной продукции* по затратам на трудовые ресурсы и капитальные вложения, *построение интегральных целевых функций*, описывающих эффективность функционирования экономических единиц (предприятий, семей) по набору частных характеристик и др., это вполне оправданно. Однако, к сожалению, далеко не всегда целевые установки исследователей подкреплены объективными возможностями их реализации.

2.3 Некоторые типовые задачи практики эконометрического моделирования

Накопленный опыт практического использования аппарата статистического исследования зависимостей позволяет выделить те типы основных прикладных направлений исследований, в которых этот аппарат работает особенно часто и плодотворно. Если попытаться расщепить общую проблему оптимального управления сложной системой (то есть центральную проблему кибернетики) на основные составляющие (рис. 2.3), то в качестве этих составляющих как раз и фигурируют именно те направления прикладных исследований, в разработке которых существенную роль играет математический аппарат статистического исследования зависимостей.

Остановимся кратко на роли методов статистического исследования зависимостей в разработке каждого из упомянутых направлений.

1. Нормирование

Общая схема формирования нормативов с использованием методов статистического исследования зависимостей может быть представлена следующим образом. Нормативный показатель играет в моделях типа (2.3)–(2.4) роль результирующей (объясняемой) переменной y , а факторы, участвующие в расчете нормативного показателя, — роль объясняющих переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$. Предполагается, что привлечение для расчета норматива y полной системы определяющих его факторов, то есть такой системы, с помощью которой возможно *детерминированное* (однозначное) определение величины y , либо принципиально невозможно, либо нецелесообразно из-за чрезмерного усложнения расчетных формул. Поэтому анализируется связь между y и $(x^{(1)}, x^{(2)}, \dots, x^{(p)})$ вида

$$y = f(x^{(1)}, x^{(2)}, \dots, x^{(p)}; \Theta) + \varepsilon, \quad (2.11)$$

где ε — остаточная случайная компонента, обуславливающая возможную погрешность в определении норматива y по известным значениям факторов $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$, а $f(X; \Theta)$ — функция из некоторого известного параметрического семейства $F = \{f(X; \Theta)\}$, $\Theta \in A$, однако численное значение входящего в ее уравнение параметра Θ (вообще говоря, векторного) неизвестно. С целью подбора «подходящего» значения Θ проводится контрольный эксперимент (наблюдение), в результате которого исследователь получает исходные статистические данные вида (2.1). Далее на основании этих данных проводится необходимый статистический анализ модели (2.11) с целью получения оценки $\hat{\Theta}$ неизвестного параметра Θ и анализа точности полученной расчетной формулы $\hat{Y}_{cp}(X) = f(X; \hat{\Theta})$, в которой величина условной (экспериментальной) средней $\hat{Y}_{cp}(X)$ интерпретируется как средний нормативный показатель при значениях определяющих факторов, равных X .



Рис. 2.3. Основные направления практического использования аппарата статистического исследования зависимостей и центральная проблема кибернетики

Данный подход использовался, в частности, при разработке методик определения численности служащих (по различным их функциям) на промышленном предприятии отрасли по набору технико-экономических показателей, характеризующих предприятие, при построении автоматизированных систем нормирования ремонтных работ и в других областях.

II. Прогноз, планирование, диагностика

Отправляясь от общей формулировки задачи статистического исследования зависимостей (см. п. 2.1) и от ее модельной записи (2.11), определим в качестве результирующей переменной y интересующий нас прогнозируемый (планируемый, диагностируемый) показатель, а в качестве объясняющих переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ — сопутствующие факторы, значения которых содержат основную информацию о величине этого показателя⁶. Наличие остаточной случайной компоненты ε , как и прежде, отражает тот факт, что переменные $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ содержат не всю информацию об y , и обуславливает неизбежность погрешности в определении прогнозируемого (планируемого, диагностируемого) показате-

⁶ В моделях прогноза и планирования в качестве одного из объясняющих факторов $x^{(k)}$ вводится в явном виде «длина прогноза», или «горизонт планирования», t (в единицах времени).

ля по известным значениям объясняющих факторов $x^{(1)}, x^{(2)}, \dots, x^{(p)}$. Исходные статистические данные вида (2.1) исследователь получает, регистрируя одновременно значения y и $(x^{(1)}, \dots, x^{(p)})$ на анализируемых объектах в прошлом (в базовом периоде) или на других объектах, но однородных с анализируемыми.

III. Оценка труднодоступных для непосредственного наблюдения и измерения параметров системы

Восстановление возраста археологической находки по ряду косвенных признаков; *прочности бетона* с помощью косвенных (неразрушающих) методов контроля (например, по отношению диаметров отпечатков на поверхности испытуемого образца бетона и на воздействующем на него эталонном молотке); *денежных сбережений семьи* по ее доходу (в среднедушевом исчислении) — во всех этих ситуациях исследователь вынужден иметь дело с показателями, труднодоступными для непосредственного измерения (они выделены в тексте курсивом). Очевидно, для того чтобы иметь принципиальную возможность статистически выявить связь, существующую между труднодоступным показателем y и косвенно связанными с ним, но легко поддающимися наблюдению и измерению признаками $x^{(1)}, x^{(2)}, \dots, x^{(p)}$, исследователю необходимо располагать исходными статистическими данными вида (2.1), которые получают с помощью специально организованного контрольного эксперимента или наблюдения. После того как эта связь выявлена (и оценена степень ее точности), она используется для косвенного определения значений труднодоступных показателей лишь по значениям объясняющих переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$.

IV. Оценка эффективности функционирования (или качества) анализируемой системы

Пытаясь оценить (в целом) эффективность деятельности отдельного специалиста, подразделения или предприятия, проранжировать страны по некоторому интегральному качеству (например, по качеству жизни населения или по так называемому общему индексу человеческого развития), наконец, проставить балльные оценки спортсмену — участнику командных соревнований в игровых видах спорта за качество его игры в определенном цикле, мы каждый раз по существу решаем (на интуитивном уровне) одну и ту же задачу: отправляясь в своем анализе от набора частных показателей $x^{(1)}, x^{(2)}, \dots, x^{(p)}$, каждый из которых может быть измерен и характеризует какую-нибудь одну частную сторону понятия «эффективность», мы их как бы взвешиваем (то есть внутренне оцениваем удельный вес их влияния на общее, агрегированное, понятие эффективности) и выходим на некоторый скалярный агрегированный показатель эффективности y . Этот показатель — латентный (скрытый),

так как он принципиально не поддается непосредственному измерению (не существует или нам неизвестна объективная шкала, в которой он мог бы быть измерен). Но он с некоторой точностью восстанавливается по значениям частных показателей эффективности $x^{(1)}, x^{(2)}, \dots, x^{(p)}$. Это значит, что между латентным агрегированным показателем y и набором частных критериев эффективности $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ существует статистическая связь типа (2.11).

Главная особенность (и трудность) описываемой ситуации заключается в том, что при получении (сборе) исходной статистической информации вида (2.1) значения результирующего показателя y могут быть получены только с помощью специально организованного экспертного опроса (значения частных критериев эффективности $x^{(1)}, x^{(2)}, \dots, x^{(p)}$, как правило, поддаются непосредственному измерению). Форма экспертной информации о значениях y может быть различной (балльные оценки, упорядочения, парные сравнения). Но только располагая наряду со статистической информацией об $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$ одной из форм соответствующей экспертной информации об y , мы можем статистически построить некоторую аппроксимацию $\hat{y}_{\text{ср}}(X) = f(X; \hat{\Theta})$ для агрегированного критерия эффективности функционирования системы и использовать ее затем в качестве формализованного метода оценки интегрального понятия эффективности (то есть уже без привлечения экспертов, а лишь по частным критериям $x^{(1)}, x^{(2)}, \dots, x^{(p)}$). Такая модифицированная форма использования аппарата статистического исследования зависимостей предложена в [Айвазян (1974)] и носит название *экспертно-статистического метода построения неизвестной целевой функции* (см. развитие этого метода в [Айвазян (2010)]).

В описанную схему вкладывается широкий класс задач теории и практики измерения комплексного понятия «качество» сложной системы: в этих задачах y интерпретируется как агрегированный (комплексный) показатель качества системы, а $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ — как отдельные частные характеристики его качества (надежность, экономичность, удобство пользования, эстетический вид и т. п.). В качестве параметрических семейств $\mathbf{F} = \{f(X; \Theta)\}$, привлекаемых при статистическом анализе задач данного типа, чаще других используются функции *линейные*

$$f(X; \Theta) = \theta_0 + \theta_1 x^{(1)} + \dots + \theta_p x^{(p)} \quad (2.12)$$

и степенные

$$f(X; \Theta) = \theta_0 (x^{(1)})^{\theta_1} (x^{(2)})^{\theta_2} \dots (x^{(p)})^{\theta_p}. \quad (2.13)$$

V. Оптимальное регулирование параметров функционирования анализируемой системы, ситуационный анализ

Рассмотрим пример (заимствован из [Айвазян (1968)]). При анализе производительности мартеновских печей на одном из заводов исследо-

валась, в частности, зависимость между производительностью в тонно-часах (для исключения влияния задержек и простоев часовая производительность мартеновской печи определялась как частное от деления массы плавки на продолжительность периода от начала завалки до выпуска) и процентным содержанием углерода в металле по расплавлению ванны (пробу брали через час после первого скачивания шлака). Результаты замеров по 130 плавкам (то есть объем n обрабатываемой статистической выборки вида (2.1) равен 130) приведены на рис. 2.4. Очевидно, величины производительности (y_i) и процентного содержания углерода (x_i) подвержены некоторому неконтролируемому разбросу, обусловленному влиянием множества не поддающихся строгому учету и контролю факторов. Другими словами, последовательность пар чисел (x_i, y_i) , $i = 1, 2, \dots, 130$, представляет в данном случае результаты 130 независимых наблюдений двумерной случайной величины (ξ, η) . Однако сквозь кажущуюся хаотичность расположения точек (x_i, y_i) на рис. 2.4 просматривается вполне определенная закономерность зависимости условного среднего значения производительности $y_{\text{ср}}(x) = E(\eta | \xi = x)$ от величины процентного содержания углерода x . Поэтому, располагая статистической зависимостью $y_{\text{ср}}(x)$, мы можем дать рекомендации технологу по оптимальному (с точки зрения максимизации производительности) управлению процессом выплавки: поддерживать процентное содержание углерода в пределах 0,6 – 1,0%.

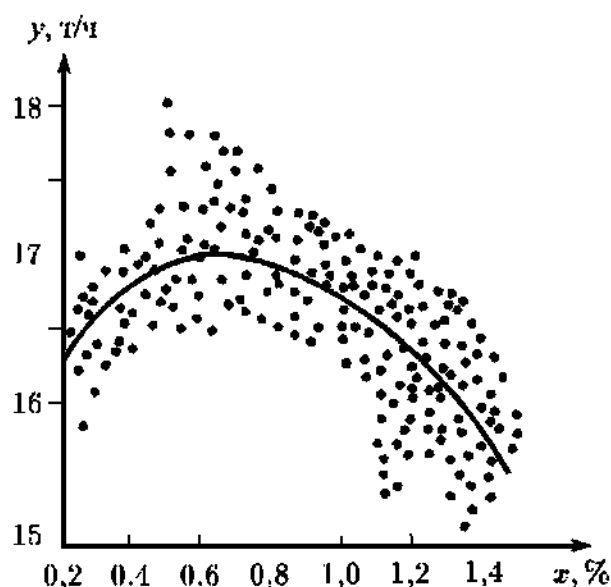


Рис. 2.4. Зависимость производительности (y , т/ч) от процентного содержания углерода (x , %) в металле до расплавления

Мы не случайно начали с этого примера. Использование методов статистического исследования зависимостей в задачах оптимального регулирования хода технологического процесса и построения соответству-

ющих автоматизированных систем управления технологическими процессами можно отнести к примерам грамотных и относительно распространенных актуальных приложений этого аппарата. Общая схема таких приложений предусматривает (в дополнение к приведенному выше частному примеру): а) одновременное рассмотрение нескольких результирующих показателей $y^{(1)}, y^{(2)}, \dots, y^{(p)}$ (производительность, качество продукции, расход сырья и энергии и т. п.) и многих регулируемых параметров технологического процесса $x^{(1)}, x^{(2)}, \dots, x^{(p)}$; б) возможность сбора исходной статистической информации вида (2.1).

Менее освоенным (но не менее правомерным и актуальным) является этот подход в задачах оптимального регулирования:

- характеристик социально-экономического поведения людей и целых коллективов в ситуациях, когда существует принципиальная возможность выявления статистических связей между этими характеристиками и набором объясняющих (и хотя бы частично регулируемых) факторов;

- структуры и объемов нагрузок и видов заданий в процессе профессиональной подготовки специалистов.

Особо выделим *макроуровень* подобного моделирования с целью оптимального регулирования параметров функционирования анализируемой системы. В этом случае речь идет об оптимальном регулировании тех макропараметров национальной экономики, которые поддаются хотя бы частичному управлению и планированию (институциональные и структурные преобразования, налоговая и социальная политика, инвестиционная активность государства и т. п.). Построив и оценив статистические связи, существующие между этими параметрами (так называемыми *экзогенными* переменными), с одной стороны, и результирующими (*эндогенными*, то есть формирующимися внутри и в ходе функционирования национальной экономики) переменными — с другой, исследователь может, придавая различные значения управляемым параметрам, отслеживать соответствующие реакции на это эндогенных переменных. То есть *происходит как бы многократная модельная «прогонка» различных сценариев социально-экономического развития* (такой способ исследования называют также *ситуационным анализом*). Реализуется этот подход, как правило, с помощью *систем регрессионных уравнений* типа (2.3)–(2.4), которые принято называть в эконометрике **системами одновременных уравнений** (см., например, главу 4 в [Айвазян (2002)]).

2.4 Основные типы зависимостей между количественными переменными

При изучении взаимосвязей между анализируемыми количественными показателями следует установить, к какому именно типу зависи-

мостей относится исследуемая схема. Под типом зависимости мы подразумеваем в данном случае не аналитический вид функции $Y_{\text{ср}}(X) = f(X; \Theta)$ в моделях вида (2.11) (о выборе общего аналитического вида функции $f(X; \Theta)$ см. п. 2.5), а природу анализируемых переменных (X, y) и соответственно интерпретацию функции $f(X; \Theta)$ в каждом конкретном случае.

Зависимость между неслучайными переменными (схема А). В этом случае результирующий показатель y детерминированно (то есть вполне определенно, однозначно) восстанавливается по значениям неслучайных объясняющих переменных $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$, то есть значения y зависят только от соответствующих значений X и полностью ими определяются. Это — обычная схема *чисто функциональной зависимости* между неслучайными переменными, когда y является некоторой функцией от p переменных X (то есть $y = f(X)$), что является вырожденным случаем зависимостей вида (2.11), когда остаточная случайная компонента ε равна нулю (с вероятностью единица).

Известно, например, что возраст дерева y (в годах) можно однозначно восстановить по числу колец x на срезе его ствола, а именно $y = x$. Примеры адекватного описания реальных зависимостей с помощью чисто функциональных (нестохастических) связей, к сожалению, крайне редки в практике исследований. Кроме того, при проведении их анализа нет необходимости использовать методы вероятностно-статистической теории. Поэтому в дальнейшем изложении мы не будем больше возвращаться к этому типу зависимостей.

Регрессионная зависимость случайного результирующего показателя η от неслучайных объясняющих переменных X (схема В). Природа такой связи может носить двойственный характер: а) регистрация результирующего показателя η неизбежно связана с некоторыми случайными ошибками измерения ε , в то время как предикторные (объясняющие) переменные $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$ измеряются без ошибок (или величины этих ошибок пренебрежимо малы по сравнению с соответствующими ошибками измерения результирующего показателя); б) значения результирующего показателя η зависят не только от соответствующих значений X , но и еще от ряда неконтролируемых факторов, поэтому при каждом фиксированном значении X^* соответствующие значения результирующего показателя $\eta(X^*) = (\eta | X = X^*)$ неизбежно подвержены некоторому случайному разбросу.

В этом случае объясняющие переменные X играют роль (векторного при $p > 1$) параметра, от которого зависит закон распределения вероятностей (в частности, среднее значение и дисперсия) исследуемого результирующего показателя η . Удобной математической моделью такого рода зависимостей является разложение вида

$$\eta(X) = f(X) + \varepsilon(X). \quad (2.14)$$

Модель (2.14) строится таким образом, что математическое ожидание случайного остатка $\varepsilon(X)$ равно нулю ($E\varepsilon(X) \equiv 0$) тождественно по X ; поэтому функция $f(X)$ описывает поведение условного среднего $y_{\text{ср}}(X) = E(\eta | X) = f(X)$ в зависимости от X . Предполагается обычно, что при всех X существует конечная дисперсия $\varepsilon(X)$ (то есть $D\varepsilon(X) < \infty$), причем величина этой дисперсии, вообще говоря, может зависеть от X (то есть $D\varepsilon(X) = \sigma^2(X)$). Подчеркнем то обстоятельство, что в описанной модели (2.14) ни природа случайной компоненты $\varepsilon(X)$, ни соответственно характеристики ее вероятностного распределения никак не связаны со структурой функции $f(X)$ и, в частности, не зависят от значений ее параметра Θ в параметрической записи модели (то есть когда вместо всех возможных функций $f(X)$ рассматривают какое-либо параметрическое семейство $f(X; \Theta)$, см., например, (2.12), (2.13)).

Если вернуться к примеру 2.1, то можно убедиться, что он хорошо укладывается в рамки модели (2.14). Для этого следует лишь заметить, что имевшаяся в этом примере возможность контролировать значения объясняющей (предикторной) переменной ξ по существу переводит эту переменную из категории случайных величин в категорию неслучайных (контролируемых) параметров модели. Дальнейший анализ примера 2.1 (см. табл. 2.1, формулу (2.5) и рис. 2.2) подсказал нам следующую конкретизацию допущений о природе составных частей модели (2.14):

$$\begin{aligned} y_{\text{ср}}(x) &= E(\eta | X) = f(x) = \theta_0 + \theta_1 x; \\ \sigma^2(x) &= D\varepsilon(x) = \sigma_0^2 \cdot (y_{\text{ср}}(x))^2, \end{aligned} \quad (2.15)$$

где σ_0 — константа, не зависящая от x .

Подчеркнем тот факт, что именно зависимость по схеме В, то есть регрессионная зависимость результирующего показателя от *неслучайных величин* объясняющих переменных, лежит в основе большинства эконометрических моделей и, в частности, используется при построении и анализе классической и обобщенной линейной модели регрессии (см. главы 4 и 5).

Корреляционно-регрессионная зависимость между случайными векторами η — результирующим показателем и ξ — объясняющей переменной (схема С). В данном типе моделей и компоненты вектора результирующего показателя η , и компоненты вектора объясняющих переменных ξ зависят от множества неконтролируемых факторов, так что являются случайными по своей физической сущности. Мы уже сталкивались с такой ситуацией в примере, в котором исследовалась связь между производительностью мартеновских печей и процентным содержанием углерода в металле (см. рис. 2.4). Зависимости такого типа вообще характерны для описания хода технологических процессов, реальные значения параметров которых $\xi = (\xi^{(1)}, \xi^{(2)}, \dots, \xi^{(p)})^T$, равно как и характеризующие их результирующие показатели $\eta = (\eta^{(1)}, \eta^{(2)}, \dots,$

$\eta^{(m)})^T$, как правило, флуктуируют случайным (но взаимосвязанным) образом около установленных номиналов.

В подобных ситуациях оказывается полезным рассмотреть разложение исследуемого результирующего показателя η на две случайные составляющие по формуле типа (2.3). Первая из них определяется некоторой (векторнозначной) функцией f от объясняющей переменной ξ , а вторая отражает остаточные влияния неучтенных случайных факторов на анализируемый результирующий показатель η . Итак,

$$\eta = f(\xi) + \varepsilon. \quad (2.16)$$

В частном случае единственного результирующего показателя ($m = 1$) и линейного вида функции $f(\xi)$ имеем:

$$\eta = \theta_0 + \sum_{k=1}^p \theta_k \xi^{(k)} + \varepsilon. \quad (2.17)$$

Подразумевая, как и прежде, под $y_{cp}(X) = E(\eta \mid \xi = X)$ условное математическое ожидание результирующего показателя η (при условии, что объясняющая переменная ξ приняла значение, равное X), мы от (2.17) приходим к линейному уравнению регрессии

$$y_{cp}(X) = \theta_0 + \sum_{k=1}^p \theta_k x^{(k)}. \quad (2.18)$$

Возможны случаи, когда вторая (остаточная) компонента в разложении (2.16) с полной мерой достоверности (то есть с вероятностью единицы) равна нулю. При этом исследуемые случайные величины η и ξ оказываются связанными *чисто функциональной зависимостью* $\eta = f(\xi)$, но ее следует отличать от функциональной зависимости неслучайных переменных (см. выше, схема А).

Пример 2.2. Рис. 2.5 иллюстрирует связь между вакуумом в печи для обжига стекла ξ и процентом брака η в стекольном производстве [Айвазян (1968)].

Случайные изменения свойств сырья, а также ряда неконтролируемых факторов приводят к случайным колебаниям обеих исследуемых переменных. Однако расположение точек на рис. 2.5 свидетельствует о том, что эти колебания взаимосвязаны, подчинены вполне определенной закономерности: «облако» рассеяния вытянуто вдоль некоторой прямой, не параллельной ни одной из координатных осей. Все это подтверждает целесообразность разложения случайной величины η по формуле (2.16) и исследования связи между η и ξ , которая в этом случае носит название корреляционной. К перечисленным вопросам регрессионного анализа (построение конкретного вида зависимости между переменными,

различные оценки ее точности) в этом случае присоединяется круг вопросов, связанных с исследованием степени тесноты связи между этими переменными. Совокупность методов, позволяющих решать эти вопросы, принято называть *корреляционным анализом* (см. главу 3).

Зависимости структурного типа, или зависимости по схеме конфлюэнтного анализа (схема D). В описываемой ниже схеме речь идет о восстановлении искомым зависимостей по искаженным наблюдениям анализируемых переменных, причем в отличие от регрессионной схемы В искаженными оказываются при наблюдении не только значения результирующего показателя, но и значения *объясняющих* (*предикторных*) переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$. Этот тип связей упоминается в специальной литературе как *структурные зависимости* [Кендалл, Стюарт (1973), с. 500–557] или как зависимости по схеме *конфлюэнтного анализа* [Айвазян (1979)].

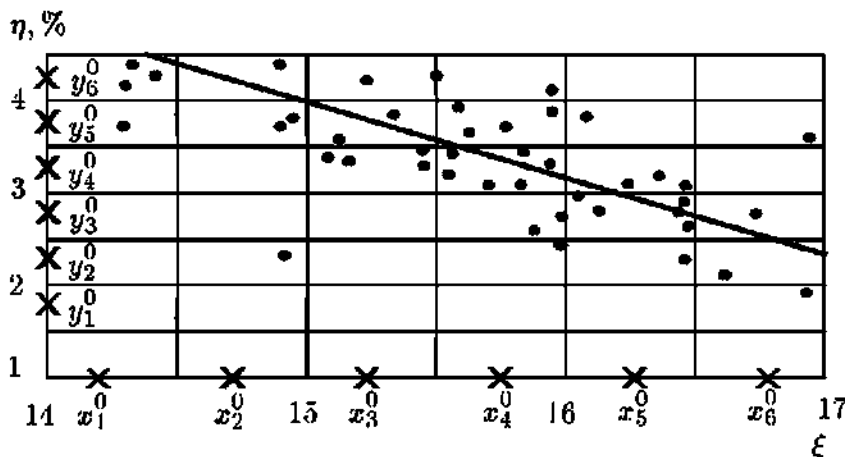


Рис. 2.5. Графическое представление данных по связи вакуума в печи для обжига стекла (ξ) и процента брака в стекольном производстве (η)

Таким образом, конфлюэнтный анализ предоставляет совокупность методов математико-статистической обработки данных, относящихся к анализу априори постулируемых функциональных связей между количественными (случайными или неслучайными) переменными $Y = (y^{(1)}, \dots, y^{(m)})^T$ и $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$ в условиях, когда наблюдаются не сами переменные, а случайные величины

$$\begin{aligned} \xi_i^{(k)} &= x_i^{(k)} + \varepsilon_{x_i}^{(k)}, \quad k = 1, 2, \dots, p; \\ \eta_i^{(j)} &= y_i^{(j)} + \varepsilon_{y_i}^{(j)}, \quad j = 1, 2, \dots, m; \quad i = 1, 2, \dots, n, \end{aligned} \quad (2.19)$$

где $\varepsilon_{x_i}^{(k)}$ и $\varepsilon_{y_i}^{(j)}$ — случайные ошибки измерений соответственно переменных $x^{(k)}$ и $y^{(j)}$ в i -м наблюдении, а n — общее число наблюдений. При

алгебраические полиномы степени $m \geq 2$:

$$f(X; \Theta) = \theta_0 + \sum_{k=1}^p \theta_k \cdot x^{(k)} + \sum_{k_1=1}^p \sum_{k_2=1}^p \theta_{k_1 k_2} x^{(k_1)} \cdot x^{(k_2)} + \dots + \sum_{k_1=1}^p \dots \sum_{k_m=1}^p \theta_{k_1 k_2 \dots k_m} \cdot x^{(k_1)} \cdot x^{(k_2)} \dots x^{(k_m)} \quad (10.21''')$$

и т. д.

Следует подчеркнуть, что этап исследования, посвященный выбору общего вида функции регрессии (параметризация модели), бесспорно, является *ключевым*: от того, насколько удачно он будет реализован, решающим образом зависит точность восстановления неизвестной функции регрессии $f(X)$. В то же время приходится признать, что этот этап находится, пожалуй, в самом невыгодном положении: к сожалению, не существует системы стандартных рекомендаций и методов, которые образовывали бы строгую теоретическую базу для его наиболее эффективной реализации.

Остановимся на некоторых рекомендациях, связанных с реализацией трех основных моментов, учет которых необходим при решении проблемы выбора общего вида функции регрессии: 1) максимальное использование априорной информации о *содержательной* (физической, экономической, социологической и т. п.) *сущности* анализируемой зависимости; 2) предварительный анализ *геометрической структуры* исходных данных вида (2.1), на основании которых конструируется искомая зависимость; 3) различные *статистические приемы* обработки исходных данных, позволяющие сделать наилучший выбор из нескольких сравниваемых вариантов.

2.5.1 Использование априорной информации о содержательной сущности анализируемой зависимости

Анализируя содержательную сущность изучаемой зависимости, исследователь еще *до обращения* к исходным статистическим данным может (и должен!) попытаться ответить на ряд вопросов по поводу характера искомой регрессионной связи:

а) будет ли искомая функция $f(X)$ монотонной или она должна иметь один экстремум (может быть, несколько)?

б) следует ли ожидать стремления (в процессе $x^{(k)} \rightarrow \infty$) $f(X)$ к асимптомам (по одной или нескольким предикторным переменным) и какова их содержательная интерпретация? Так, например, если $f(X)$ — средний объем благ определенного вида, потребляемых семьями группы X по доходам, то, очевидно, при $X \rightarrow \infty$ во многих случаях следует ожидать «насыщения», то есть $f(X)$ будет стремиться (снизу) к горизонтальной асимптоте;

в) какова принципиальная природа воздействия объясняющих переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ на формирование результирующего показателя y — аддитивная или мультипликативная? Так, например, многие схемы зависимостей в экономике характеризуются мультипликативной природой воздействия предикторов на y ;

г) не диктует ли содержательный смысл анализируемой зависимости обязательное прохождение графика искомой функции $f(X)$ через одну или несколько априори заданных точек в исследуемом факторном пространстве (X, y) ?

Поясним необходимость и возможность максимального извлечения информации об общем виде анализируемой функции регрессии $f(X)$ из соображений профессионально-теоретического характера на примере.

Пример 2.3. На рис. 2.6 представлены 63 результата специального эксперимента [Езекиэл, Фокс (1966), с. 57]. Расположение точек на рис. 2.6 не дает ответа на вопрос, описывать ли зависимость между скоростью автомобиля (x миль/ч) и расстоянием (y футов), пройденным им после поданного сигнала об остановке, линейной или параболической зависимостью.

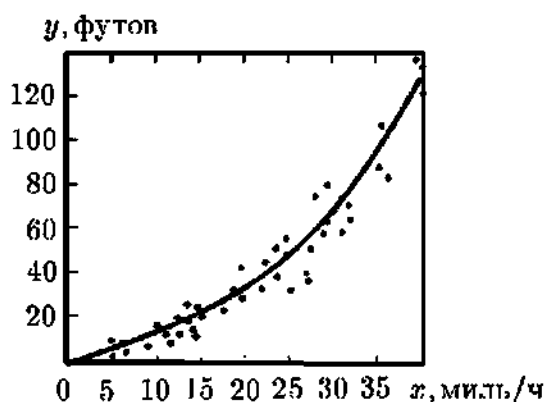


Рис. 2.6. График зависимости тормозного пути автомобиля (y) от скорости его движения (x)

Этот вопрос остается без ответа и после построения соответствующих кривых и применения известных статистических критериев, предназначенных решать, насколько хорошо согласуются кривые с экспериментальными данными. Однако несложные рассуждения профессионально-теоретического характера все-таки позволяют сделать этот выбор. Действительно, для каждого отдельного автомобиля и водителя расстояние, пройденное до остановки, определяется в основном тремя факторами: скоростью автомобиля (x) в момент подачи сигнала об остановке, временем реакции на этот сигнал водителя (θ_1 , ч) и тормозами автомобиля. Автомобиль успеет пройти путь $\theta_1 x$ до момента включения водителем тормозов и еще $\theta_1 \cdot x^2$ после этого момента, поскольку согласно элементарным физическим законам теоретическое расстояние, пройденное до остановки с момента торможения, пропорционально квадрату скорости.

Итак, $f(x) = \theta_1 x + \theta_2 x^2$, что после оценивания θ_1 и θ_2 с помощью метода наименьших квадратов (см. главу 4) дает $f(x) = 0,76x + 0,056x^2$.

2.5.2 Предварительный анализ геометрической структуры исходных данных

При выяснении вопроса о параметрическом виде исследуемой зависимости, как правило, идут от простого к сложному. Простейшей же аппроксимацией неизвестной функции регрессии $f(X) = E(\eta | \xi = X)$ является, естественно, *линейная модель*, то есть функция вида

$$f_{\text{л}}(X) = \theta_0 + \theta_1 x^{(1)} + \dots + \theta_p x^{(p)}. \quad (2.22)$$

Поэтому при *предварительном* анализе характера исследуемых зависимостей (то есть до проведения вычислительных процедур по оценке неизвестных значений параметров, входящих в гипотетичные уравнения связей) ограничиваются некоторыми приближенными эвристическими приемами, связанными в основном с изучением «геометрии» парных корреляционных полей и визуальной проверкой их линейности.

Содержание геометрического анализа парных корреляционных полей. Под *корреляционным полем* переменных (u, v) понимается графическое представление имеющихся измерений $(u_1, v_1), (u_2, v_2), \dots, (u_n, v_n)$ этих переменных в плоскости (u, v) . Мы уже неоднократно имели дело с корреляционными полями (см. рис. 2.2, 2.4, 2.5, 2.6).

Анализ парных корреляционных полей состоит обычно в следующем:

а) построение на основании имеющихся исходных данных вида (2.1) корреляционных полей для всевозможных пар переменных вида $(x^{(j)}, x^{(k)})$ и $(x^{(j)}, y)$, отобранных из набора всех $p + 1$ исследуемых признаков $(x^{(1)}, x^{(2)}, \dots, x^{(p)}; y)$; всего таких пар будет, очевидно, $p(p + 1)/2$, однако процесс этот легко автоматизируется с помощью современных вычислительных средств;

б) визуальное прослеживание характера вытянутости каждого корреляционного поля: эллипсоидально-линейное (см. рис. 2.5), нелинейно-монотонное (см. рис. 2.6), с наличием одного или нескольких экстремумов (см. рис. 2.4) и т. п.;

в) изучение поведения условных средних значений результирующего показателя при изменении величины переменной, откладываемой по оси абсцисс и играющей роль предикторной (см. рис. 2.2); для этого (если значения предикторной переменной *неконтролируемы* в ходе наблюдения или эксперимента) предварительно разбивают диапазон значений объясняющей переменной на *интервалы группирования* и подсчитывают средние значения ординат тех точек-наблюдений, которые попали в общий интервал группирования.

В результате такого анализа обычно получают формулировку нескольких рабочих гипотез об общем виде искомой зависимости, оконча-

тельная проверка которых и выбор наиболее адекватной из них осуществляются (при отсутствии априорных сведений содержательного характера) с помощью соответствующих *математико-статистических* методов. Описание некоторых приемов такого типа приводится в п. 2.5.3.

2.5.3 Статистические критерии проверки гипотез об общем виде функции регрессии

Подчеркнем сразу, что описанные ниже критерии проверки справедливости сделанного выбора общего вида искомой функции регрессии не могут ответить на вопрос: является ли проверяемый гипотетичный вид зависимости наилучшим, единственно верным? Они лишь подтверждают факт непротиворечивости проверяемого вида функции регрессии имеющимся у исследователя исходным данным (2.1) либо отвергают обсуждаемую гипотетичную форму зависимости как не соответствующую этим данным.

1. *Общий приближенный критерий, основанный на группированных данных* (или при наличии нескольких наблюдений при каждом фиксированном значении аргумента). Пусть высказана гипотеза об общем виде функции регрессии $H_0: E(\eta | \xi = X) = f_a(X; \theta_1, \theta_2, \dots, \theta_k)$ ($f_a(X; \Theta)$ — известная функция, $(\theta_1, \theta_2, \dots, \theta_k) = \Theta$ — неизвестные числовые параметры) и пусть вычислены (например, с помощью метода наименьших квадратов, см. главу 4) оценки $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_k$ неизвестных параметров, входящих в описание уравнения регрессии. При группировке данных (или при проведении эксперимента) мы должны соблюдать требование, в соответствии с которым число интервалов группирования (или число различных значений аргумента, в которых производились наблюдения) s должно обязательно превосходить число неизвестных параметров k , то есть $s - k \geq 1$.

Если высказанная гипотеза об общем виде зависимости является правильной, то статистика

$$v^2 = \frac{\frac{1}{s-k} \sum_{i=1}^s \nu_i |\bar{y}_i - f_a(X_i^0; \hat{\Theta})|^2}{\frac{1}{n-s} \sum_{i=1}^s \sum_{j=1}^{\nu_i} |y_{ij} - \bar{y}_i|^2} \quad (2.23)$$

должна приближенно подчиняться $F(m_1, m_2)$ -распределению с числом степеней свободы числителя $m_1 = s - k$ и знаменателя $m_2 = n - s$. Все величины в формуле (2.23) соответствуют ранее введенным обозначениям. В частности, X_i^0 — середина i -го гиперпараллелепипеда группирования (или i -е значение аргумента, в котором было проведено ν_i наблюдений); $f_a(X_i^0; \hat{\Theta})$ — значение гипотетической функции регрессии, вычисленное в точке $X = X_i^0$; $\bar{y}_i$ — условное среднее из ординат, попавших в i -й гиперпараллелепипед группирования (или из ординат, измеренных при

i -м фиксированном значении аргумента X_i^0 ; y_{ij} — j -е по счету значение ординаты из числа попавших в i -й интервал группирования (или из числа измеренных при i -м фиксированном значении аргумента X_i^0). Легко понять, что числитель в правой части (2.23) характеризует меру рассеивания экспериментальных данных вокруг аппроксимирующей выборочной регрессионной поверхности, а знаменатель — меру рассеивания экспериментальных данных около своих условных выборочных средних $\bar{y}_i$ (то есть меру, независимую от выбранного вида линии регрессии). Причем и числитель, и знаменатель являются практически независимыми (в некоторых частных случаях — *точно* независимыми) статистическими оценками одной и той же теоретической дисперсии $\sigma^2 = D(\eta | \xi = X)$.

Соответственно получаем следующее правило проверки гипотезы об общем виде функции регрессии. Задаемся, как обычно, достаточно малым уровнем значимости критерия α (например, $\alpha = 0,05$). С помощью таблицы П1.5 находим $100(1 - \frac{\alpha}{2})\%$ -ную точку $v_{1-\frac{\alpha}{2}}^2$ и $100\frac{\alpha}{2}\%$ -ную точку $v_{\frac{\alpha}{2}}^2$ $F(k - m, n - k)$ -распределения. Если окажется, что величина v^2 , подсчитанная по формуле (2.23), удовлетворяет неравенствам

$$v_{1-\frac{\alpha}{2}}^2 < v^2 < v_{\frac{\alpha}{2}}^2,$$

то высказанная нами гипотеза об общем виде функции регрессии признается не противоречащей экспериментальным данным (2.1). Если же эти неравенства оказались нарушенными, то гипотеза об общем виде функции регрессии отвергается с уровнем значимости α . При этом если v^2 «слишком мало» (то есть $v^2 < v_{1-\frac{\alpha}{2}}^2$), то, очевидно, при выборе общего вида регрессии мы неправомерно реагировали на случайные отклонения точек $(X_i^0, \bar{y}_i)$ от истинной функции регрессии и тем самым необоснованно завысили число параметров k , от которых зависит уравнение регрессии. Напротив, если v^2 «слишком велико» (то есть $v^2 > v_{\frac{\alpha}{2}}^2$), то «гибкость» аппроксимирующей функции регрессии $f_a(X; \Theta)$ следует признать недостаточной, поэтому целесообразно увеличить число неизвестных параметров регрессии (например, повысить порядок аппроксимирующего полинома).

Для случая, когда условная дисперсия зависимой переменной пропорциональна некоторой известной функции аргумента, то есть $D\eta(X) = \sigma^2 h^2(X)$, формула (2.23) преобразуется:

$$v^2 = \frac{\frac{1}{s-k} \sum_{i=1}^s w_i \nu_i |\bar{y}_i - f_a(X_i^0, \hat{\Theta})|}{\frac{1}{n-s} \sum_{i=1}^s w_i \sum_{j=1}^{\nu_i} (y_{ij} - \bar{y}_i)^2}, \quad (2.23')$$

где

$$w_i = 1/h^2(X_i^0).$$

Так, в примере 2.1: $n = 40$; $s = 4$; $\alpha = 0,05$; $k = 2$; дисперсионное отношение v'^2 , подсчитанное по формуле (2.23'), равно 1,04, в то время как 5%-ная точка $F(2,36)$ -распределения $v_{0,05}^2 = 3,26$. Это свидетельствует о том, что гипотеза о линейном виде регрессионной зависимости в данном случае не противоречит имеющимся в нашем распоряжении экспериментальным данным.

При проверке *линейности* регрессии (так же, впрочем, как и при проверке гипотезы о полиномиальном характере регрессии заданного порядка) в нормальных схемах зависимостей типа B и C (см. п. 2.4) описанный общий критерий является *точным*. При этом в линейном случае статистика v^2 , определенная соотношением (2.23), может быть выражена в более удобной форме, не требующей предварительного вычисления выборочной аппроксимирующей функции регрессии. Так, например, в случае *парной* линейной регрессии (то есть при $k = 2$) имеем:

$$v^2 = \frac{(n-s)(\hat{\rho}_{\eta\xi}^2 - \hat{r}^2)}{(s-2)(1 - \hat{\rho}_{\eta\xi}^2)}. \quad (2.24)$$

Здесь $\hat{\rho}_{\eta\xi}$ и $\hat{r}$ — соответственно выборочные корреляционное отношение (η по ξ) и коэффициент корреляции (см. главу 3). Логическая схема использования статистики (2.24) аналогична ранее изложенным критериям: задаются достаточно малым ($0,05 \sim 0,15$) уровнем значимости α ; находят по таблице 100 α %-ную точку v_{α}^2 распределения $F(s-2, n-s)$; сравнивают величину v^2 , определенную с помощью (2.24), с процентной точкой v_{α}^2 ; если оказывается, что $v^2 > v_{\alpha}^2$, то гипотезу о линейном виде регрессии считают статистически необоснованной.

Воспользуемся данным критерием для статистической проверки линейности регрессии в примере 2.2. Вычисления дают: $\hat{r}^2 = 0,429$, $\hat{\rho}_{\eta\xi}^2 = 0,459$, так что $v^2 = 0,513$. Принимая во внимание, что величина 5%-ной точки $F(4,37)$ -распределения равна $v_{0,05}^2 = 2,63$, делаем вывод о непротиворечивости гипотезы линейной регрессии и данных нашего эксперимента в данном примере ($0,513 < 2,63$).

2. *Общий приближенный критерий, основанный на негруппированных данных (при известной величине дисперсии остаточной случайной компоненты).*

Встречаются ситуации, когда в результате предварительных исследований или из других каких-либо соображений нам удастся заранее определить величину дисперсии σ^2 остаточной случайной компоненты ε в разложениях вида (2.14) и (2.16) (например, когда ε — ошибка измерения и нам известны характеристики точности используемого измерительного прибора). В этом случае можно отказаться от стеснительного требования группированности данных и для проверки гипотезы об общем виде функции регрессии воспользоваться фактом $\chi^2(n-k)$ -распределенности

статистики

$$\gamma = \frac{1}{\sigma^2} \cdot \frac{1}{n-k} \sum_{i=1}^n (y_i - f_a(X_i; \hat{\Theta}))^2 \quad (2.25)$$

(который имеет место при условии справедливости нашей гипотезы).

Задавшись уровнем значимости критерия α и найдя с помощью таблицы П1.4 величины $100(1 - \frac{\alpha}{2})\%$ -ных и $100\frac{\alpha}{2}\%$ -ных точек χ^2 -распределения с $n - k$ степенями свободы соответственно $\chi^2_{1-\frac{\alpha}{2}}(n - k)$ и $\chi^2_{\frac{\alpha}{2}}(n - k)$, проверяем выполнение неравенства

$$\chi^2_{1-\frac{\alpha}{2}}(n - k) < \gamma < \chi^2_{\frac{\alpha}{2}}(n - k),$$

где γ подсчитано по формуле (2.25). Если эти неравенства оказались нарушенными, то от гипотезы H_0 об общем виде функции регрессии следует отказаться. При этом если γ «слишком мало» (то есть $\gamma \leq \chi^2_{1-\frac{\alpha}{2}}(n - k)$), то, очевидно, при выборе общего вида мы неправильно реагировали на случайные отклонения экспериментальных точек (X_i, y_i) и тем самым необоснованно завысили число параметров k , от которых зависит уравнение регрессии.

Напротив, если γ «слишком велико» (то есть $\gamma \geq \chi^2_{\frac{\alpha}{2}}(n - k)$), то «гибкость» аппроксимирующей кривой регрессии $f_a(X; \Theta)$ следует признать недостаточной, поэтому целесообразно увеличить число неизвестных параметров регрессии (например, повысить порядок аппроксимирующего полинома).

Для случая, когда дисперсия зависимой переменной (или, что то же, дисперсия остаточной случайной компоненты) не остается постоянной при изменении X , а пропорциональна некоторой известной функции аргумента, то есть $D\eta(X) = \sigma^2 h^2(X)$, формула подсчета статистики γ несколько изменится:

$$\gamma = \frac{1}{\sigma^2} \cdot \frac{1}{n-k} \sum_{i=1}^n w_i (y_i - f_a(X_i; \hat{\Theta}))^2,$$

где $w_i = 1/h^2(X_i)$. В остальном схема проверки гипотезы об общем виде функции регрессии остается той же самой, что и в случае $D\eta(X) = \sigma^2 = \text{const.}$

2.5.4 Некоторые общие рекомендации

При выборе общего вида искомой функции регрессии $f(X)$, помимо соображений и приемов, описанных выше, полезно учитывать следующие общие рекомендации.

1) *Не следует гнаться за чрезмерной сложностью функции, описывающей поведение искомой функции регрессии*, руководствуясь исключительно соображениями оптимизации критерия качества аппроксимации. Дело в том, что если и оценки $\hat{\Theta}$ неизвестных параметров модели

Θ , и значение критерия качества аппроксимации *вычисляются на основании одной и той же выборки*, то за счет увеличения размерности k оцениваемого векторного параметра $\Theta = (\theta_1, \theta_2, \dots, \theta_k)^T$ можно добиться, на первый взгляд, идеального результата. Возможность эта основана на известном в математическом анализе результате, в соответствии с которым для любой заданной системы точек $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ (с неповторяющимися абсциссами) можно подобрать алгебраический полином степени $n - 1$, проходящий в точности через все точки этой системы. А это значит, что все «невязки» $y_i - f(x_i)$, на основании которых строится минимизируемый критерий, равны нулю, то есть «лучше» модели для описания поведения функции регрессии подобрать невозможно. На самом деле при таком подходе мы как бы заставляем функцию $f(x)$ реагировать на случайные флуктуации, объясняемые наличием в представлениях типа (2.14) остаточной случайной компоненты $\varepsilon(X)$. Поэтому, если мы попробуем применить полученный таким образом результат к другой выборке из той же самой генеральной совокупности, то увидим явное рассогласование модельных $(f(X_i))$ и наблюдаемых (y_i) значений результирующего показателя. Поэтому при подборе общего вида функции регрессии, как правило, *идут от простого к сложному*, то есть начиная с анализа возможности использовать простейшую линейную модель вида (2.22).

2) *Следует добиваться компромисса между сложностью регрессионной модели и точностью ее оценивания*. Из общих результатов математической статистики, относящихся к анализу точности оценивания исследуемой модели при ограниченных объемах выборки, следует, что с увеличением сложности модели (выраженной, например, размерностью k векторного параметра Θ , участвующего в ее уравнении) точность оценивания падает. Например, ширина доверительного интервала для неизвестного значения $y(X)$, при прочих равных характеристиках анализируемой схемы, увеличивается с ростом размерности параметра Θ , участвующего в вычислении функции регрессии $f(X; \Theta)$. Именно поэтому в ситуациях, когда исследователь располагает ограниченной исходной выборочной информацией вида (2.1), он вынужден искать компромисс между степенью общности привлекаемого класса допустимых решений $\mathbf{F}$ и точностью оценивания, которой возможно при этом добиться (пример подобного рода действий по поиску компромисса будет приведен в главе 4 в связи с задачей определения оптимального числа объясняющих переменных в линейной модели множественной регрессии).

3) *При обнаружении нелинейности в парных статистических связях анализируемых переменных $x^{(j)}$ и y ($j = 1, 2, \dots, p$) следует попытаться применить к этим переменным линеаризующие преобразования*. Простейший пример такого приема мы имеем, когда вместо анализа степенной зависимости вида

$$y = \theta_0(x)^{\theta_1}$$

исследователь рассматривает линейную зависимость между логарифмами исходных переменных, а именно:

$$\tilde{y} = \tilde{\theta}_0 + \theta_1 \tilde{x},$$

где $\tilde{y} = \ln y$, $\tilde{x} = \ln x$ и $\tilde{\theta}_0 = \ln \theta_0$. В зависимости от типа нелинейной связи, существующей между исходными переменными, подбираются и другие линеаризующие преобразования. Один из наиболее общих подходов к линеаризации анализируемых зависимостей реализуется с помощью так называемого *преобразования Бокса-Кокса* (см., например, [Айвазян (2001), п. 2.12.2]).

4) *Анализ регрессионных остатков.* Ряд статистических критериев проверки адекватности используемой аппроксимирующей модели регрессии основан на анализе регрессионных остатков (невязок) $\hat{\varepsilon}(X_i) = y_i - \hat{f}_a(X_i)$, $i = 1, 2, \dots, n$. В основе их конструирования — положение, в соответствии с которым правильный выбор модели $f_a(X)$ предопределяет в широком классе случаев асимптотическую (по $n \rightarrow \infty$) независимость остатков $\hat{\varepsilon}(X_i)$. Поэтому статистическая проверка правильности выбора общего вида функции регрессии сводится к проверке статистической независимости остатков, для чего могут быть использованы, например, критерии, описанные в главе 10. В выражении для невязок $\hat{\varepsilon}(X_i)$ под $\hat{f}_a(X_i)$ понимается значение оцененной аппроксимирующей функции регрессии в точке X_i , то есть $\hat{f}_a(X_i) = f_a(X_i; \hat{\Theta})$.

5) *Поиск модели, наиболее устойчивой к варьированию состава выборочных данных, на основании которых она оценивается.* Идея этого подхода к выбору общего вида исследуемой регрессионной зависимости основана на следующем простом соображении. Если общий параметрический вид зависимости $f(x^{(1)}, x^{(2)}, \dots, x^{(p)}; \Theta)$ «угадан» правильно, то результаты оценивания $\hat{\Theta}_1, \hat{\Theta}_2, \dots$ параметра Θ по различным подвыборкам выборки $\{x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)}; y_i\}_{i=\overline{1,n}}$ будут мало отличаться друг от друга (а следовательно, не сильно будут различаться между собой и соответствующие значения $f(x^{(1)}, x^{(2)}, \dots, x^{(p)}; \hat{\Theta}_1)$, $f(x^{(1)}, x^{(2)}, \dots, x^{(p)}; \hat{\Theta}_2), \dots$). И наоборот, при неудачном выборе общего вида искомой зависимости результаты ее восстановления по различным выборкам, как правило, будут сильно отличаться один от другого. Описание реализации этого подхода читатель может найти в [Айвазян, Енюков, Мешалкин (1985), п. 6.2.3].

Выводы

1. Аппарат статистического исследования зависимостей — составная часть инструментария эконометрики — нацелен на решение основной проблемы естествознания: как на основании частных результатов статистического наблюдения за анализируемыми событиями или показателями выявить и описать существующие между ними стохастические взаимосвязи.

2. Анализируемые переменные величины по своей роли в исследовании подразделяются на результирующие (прогнозируемые) Y и объясняющие (экзогенные, предикторные) X . Среди компонент векторов Y и X могут быть и количественные, и порядковые (ординальные), и классификационные (номинальные).

3. Центральным математическим объектом в процессе статистического исследования зависимостей является функция $f(X)$, называемая функцией регрессии Y по X и описывающая изменение условного среднего значения $Y_{\text{ср}}(X)$ результирующего показателя Y (вычисленного при фиксированных на уровне X значениях объясняющих переменных) в зависимости от изменения значений объясняющих переменных X .

4. Конечные прикладные цели статистического исследования зависимостей могут быть в основном трех типов: 1) установление самого факта наличия (или отсутствия) статистически значимой связи между Y и X , исследование структуры этих связей; 2) прогноз (восстановление) неизвестных значений индивидуальных или средних значений результирующего показателя по заданным значениям соответствующих объясняющих (экзогенных) переменных; 3) выявление причинных связей между объясняющими переменными X и результирующими показателями Y , частичное управление значениями Y путем регулирования величин объясняющих переменных X .

5. К основным типовым задачам практики, в которых использование аппарата статистического исследования зависимостей оказывается наиболее уместным и эффективным, следует отнести задачи: 1) нормирования; 2) прогноза, планирования и диагностики; 3) оценки труднодоступных (для непосредственного наблюдения и измерения) характеристик исследуемой системы; 4) оценки эффективности функционирования (или качества) анализируемой системы; 5) регулирования параметров функционирования анализируемой системы.

6. По своей природе исследуемые зависимости могут быть разделены на: 1) детерминированные (тип А), когда исследуется функциональная зависимость между неслучайными переменными; 2) регрессионные (тип В), когда исследуется зависимость случайного результирующего показателя от неслучайных объясняющих переменных — параметров системы; 3) корреляционные (тип С), когда исследуется зависимость между случайными переменными, причем объясняющие переменные могут быть

измерены без искажений; 4) конфлюэнтные (тип D), когда исследуется функциональная зависимость между случайными или неслучайными переменными в ситуации, когда те и другие могут быть измерены только с некоторой случайной ошибкой.

7. *Этап параметризации регрессионной модели*, то есть выбора параметрического семейства функций (класса допустимых решений), в рамках которого производится дальнейший поиск неизвестной функции регрессии, *является одновременно наиболее важным и наименее теоретически обоснованным* этапом регрессионного анализа.

8. Прежде всего исследователь должен сосредоточить свои усилия на *анализе содержательной сущности* искомой статистической зависимости, чтобы максимально использовать имеющиеся априорные сведения о «физическом» механизме изучаемой связи при выборе общего вида функции регрессии.

9. Важную роль в правильном выборе параметрического класса допустимых решений играет *предварительный анализ геометрической структуры совокупности исходных данных* и в первую очередь анализ геометрии парных корреляционных полей, включающий в себя, в частности, учет и формализацию «гладких» свойств искомой функции регрессии, использование вспомогательных линеаризующих преобразований.

10. Сформулированные с помощью содержательного и геометрического анализа рабочие гипотезы об общем виде искомой функции регрессии могут быть проверены с привлечением соответствующих *математико-статистических критериев*. Среди фундаментальных идей, на которых базируются эти статистические критерии, следует выделить: а) идею компромисса между сложностью регрессионной модели («емкостью» класса допустимых решений) и точностью ее оценивания; б) идею поиска модели, наиболее устойчивой к варьированию состава выборочных данных, на основании которых она оценивается; в) идею проверки гипотез об общем виде функции регрессии на базе сравнения выборочных критериев адекватности и исследования статистических свойств получаемых при этом оценок размерности модели.

Глава 3

Введение в корреляционный анализ

3.1 Назначение и место корреляционного анализа в статистическом исследовании

При исследовании реальных социально-экономических явлений и систем статистику, эконометристу приходится сталкиваться, как правило, с необходимостью статистического анализа *многомерной* генеральной совокупности, то есть с ситуациями, когда на каждом из статистически обследуемых объектов этой совокупности регистрируются значения *целого набора* признаков $x^{(1)}, x^{(2)}, \dots, x^{(p)}$. В дальнейшем мы будем обозначать, как и прежде, этот набор признаков с помощью вектора-столбца $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^\top$, а результат регистрации значений этих признаков на i -м статистически обследованном объекте — i -е *многомерное наблюдение* — с помощью $X_i = (x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)})^\top$. Таким образом, «стартовая позиция» при статистическом анализе многомерной генеральной совокупности аналогична одномерному случаю: *исследователь по имеющейся у него случайной выборке*

$$X_1, X_2, \dots, X_n \tag{3.1}$$

значений анализируемой многомерной случайной величины $\xi = (\xi^{(1)}, \xi^{(2)}, \dots, \xi^{(p)})^\top$ должен сделать те или иные статистические выводы о ее «поведении».

Мы знаем, что *исчерпывающие* сведения о поведении анализируемой случайной величины ξ содержатся в ее законе распределения вероятностей (з.р.в.) $f_\xi(X)$, где под $f_\xi(X)$ понимается значение функции плотности вероятности в точке X , если случайная величина ξ непрерывна, и вероятность $P\{\xi = X\}$ того, что случайная величина примет значение, равное X , если ξ дискретна. Однако з.р.в. анализируемой случайной величины исследователю, как правило, неизвестен. И

если при описании поведения одномерных случайных величин исследователь еще имеет *практически реализуемые* возможности подбора и использования подходящих модельных законов распределения с последующей статистической оценкой участвующих в их записи параметров, то при исследовании признаков размерности $p \geq 2$ ему чаще всего приходится ограничиваться информацией, которую доставляют оценки моментов первых двух порядков, а именно оценками *вектора средних значений* $\hat{\mu} = (\hat{\mu}^{(1)}, \hat{\mu}^{(2)}, \dots, \hat{\mu}^{(p)})^T$ и *ковариационной матрицы* $\hat{\Sigma} = (\hat{\sigma}_{ij})$, $i, j = 1, 2, \dots, p$. Другими словами, можно сказать, что за редким исключением, все выводы *многомерного статистического анализа* строятся на базе оценок $\hat{\mu}$ и $\hat{\Sigma}$.

Оценка $\hat{\mu}$ вектора средних значений μ дает представление о *центре группирования* наблюдений анализируемого многомерного признака. Отличие от одномерного случая, когда мы располагаем оценкой $\hat{\mu}_1 = \bar{x}$ среднего значения анализируемой случайной величины, заключается лишь в том, что $\hat{\mu}$ в общем случае определяет точку в p -мерном пространстве, в то время как выборочное среднее $\bar{x}$ определяет точку на числовой прямой. По существу вся специфика многомерного случая сосредоточена в ковариационной матрице Σ , а при статистическом анализе — в ее оценке $\hat{\Sigma}$. Именно знание ковариационной матрицы позволяет исследователю строить и анализировать характеристики случайного рассеивания и статистической взаимосвязи (*коррелированности*) компонент анализируемого многомерного признака. Данная глава как раз и посвящена так называемому **корреляционному анализу** многомерной генеральной совокупности, назначение которого — получить (на основе имеющейся выборки (3.1)) ответы на следующие основные вопросы:

- как выбрать (с учетом специфики и природы анализируемых переменных) подходящий измеритель статистической связи между компонентами анализируемой случайной величины (коэффициент корреляции, корреляционное отношение, какую-либо информационную характеристику связи, ранговый коэффициент корреляции и т. п.)?
- как оценить (с помощью точечной и интервальной оценок) его числовое значение по имеющимся выборочным данным?
- как проверить гипотезу о том, что полученное числовое значение анализируемого измерителя связи действительно свидетельствует о наличии статистической связи (или, как говорят, проверить исследуемую корреляционную характеристику на статистически значимое ее отличие от нуля)?
- как определить структуру связей между компонентами исследуемого многомерного признака, сопоставив каждой паре компонент двоичный ответ («связь есть» или «связи нет»)?

Какое место занимает корреляционный анализ в общем прикладном эконометрическом исследовании? Центральное место в таком исследовании занимает проблема *статистического исследования зависимостей* (см. главу 2). Но при исследовании зависимостей между анализируемыми переменными мы должны дать в первую очередь ответ на вопрос: *а существует ли такая зависимость?* И только после утвердительного ответа на этот вопрос заняться выявлением вида и математической формы этой зависимости. Корреляционный анализ как раз и предоставляет средства, позволяющие ответить на первый вопрос. Выявление же вида и математической формы искомой зависимости производится с помощью разнообразных методов и моделей *регрессионного* анализа и анализа временных рядов, которым посвящены практически все остальные главы данного учебника. Добавим к этому, что рассматриваемые в рамках корреляционного анализа характеристики статистической связи (ковариации, различные коэффициенты корреляции и т. п.) используются в качестве «входной» (базовой) информации при решении задач других двух центральных проблем многомерного статистического анализа — *классификации объектов и признаков* и *снижения размерности анализируемого признакового пространства*, также играющих заметную роль в эконометрических исследованиях. Поэтому мы посчитали, что без главы, посвященной корреляционному анализу, не может обойтись изложение методов и моделей современной эконометрики.

3.2 Корреляционный анализ количественных признаков

Здесь речь идет об измерителях степени тесноты статистической связи между *количественными* компонентами вектора $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$. Напомним, что значение количественной случайной величины есть результат измерения степени проявления анализируемого свойства статистически обследованного объекта в числовой шкале определенного физического смысла (в штуках, метрах, денежных единицах, килограммах, единицах времени и т. д.).

В соответствии с изложенной в п. 2.1 общей логической схемой статистического исследования зависимостей анализируемые переменные по своей роли делятся на *результатирующие* и *объясняющие*. Поэтому, оставив за компонентами вектора X роль объясняющих переменных, введем в рассмотрение *результатирующую переменную* y . Тогда исходные статистические данные (3.1) пополнятся еще одним вектором-столбцом $Y = (y_1, y_2, \dots, y_n)^T$, так что в дальнейшем предметом нашего анализа будут исходные статистические данные вида

$$X^{(1)}, X^{(2)}, \dots, X^{(p)}, Y, \quad (3.1')$$

где $X^{(j)} = (x_1^{(j)}, x_2^{(j)}, \dots, x_n^{(j)})^T$ — вектор-столбец наблюдаемых значений j -й объясняющей переменной ($j = 1, 2, \dots, p$).

В п. 2.4 были описаны основные типы статистических зависимостей (B , C и D), которые могут наблюдаться между исследуемыми переменными. Умение правильно классифицировать каждую конкретную многомерную систему наблюдений играет решающую роль при правильном выборе подходящих математико-статистических методов анализа изучаемой зависимости *на стадии регрессионного анализа и при неформальной содержательной интерпретации выявленной связи*.

Однако при решении задач корреляционного анализа удобнее использовать *единый (унифицированный) подход*, при котором исследуемая *объясняющая переменная* (*случайная* при типах зависимостей C и D и *неслучайная* при зависимостях типа B) **интерпретируется как параметр, от которого зависит закон распределения результирующего показателя**. Его мы и будем придерживаться в данной главе.

3.2.1 Коэффициент детерминации как универсальная характеристика степени тесноты статистической связи

Основная идея, лежащая в основе определения коэффициента детерминации, состоит в следующем. Пусть нас интересует *степень тесноты статистической связи* (с.т.с.), существующей между результирующим показателем y и объясняющей переменной (вообще говоря, векторной) X . Очевидно, степень тесноты этой связи может считаться *максимальной*, если по заданному значению объясняющей переменной X можно *однозначно* (без всякой случайной ошибки) восстановить соответствующее значение результирующего показателя $y(X)$. И наоборот: *если значение величин X не несет никакой информации о значении результирующего показателя y* , то связь отсутствует вовсе, и соответствующий измеритель степени ее тесноты должен иметь *минимальное* (на выбранной шкале) возможное значение.

Для того чтобы математически формализовать эту естественную посылку в виде подходящего измерителя с.т.с., посмотрим, как она интерпретируется в терминах регрессионной модели типа (2.14).

Итак, мы рассматриваем статистическую зависимость вида

$$y(X) = f(X) + \varepsilon(X), \quad (3.2)$$

где $f(X) = E(y | X)$ — условное среднее значение результирующего показателя (при условии, указанном за прямой чертой, здесь и далее означает, что объясняющая переменная приняла заданное *фиксированное* значение X), то есть *функция регрессии y по X* , — а $\varepsilon(X)$ остаточная случайная компонента, з.р.в. которой, вообще говоря, может зависеть от X , но при этом предполагается:

$$\begin{aligned} \mathbf{E}(\varepsilon | X) &\equiv 0^1, \\ \mathbf{D}(\varepsilon | X) &= \sigma_\varepsilon^2(X), \end{aligned} \quad (3.2')$$

и $\varepsilon(X)$ не коррелирована с $f(X)$ в схемах зависимости типов C и D (в схемах типа B величина $f(X)$ не является случайной).

Отметим два важных соотношения, связывающих характеристики варьирования (условные и безусловные) составных частей модели (3.2)–(3.2'):

$$\mathbf{E}(y - y_0)^2 = \mathbf{E}(f - f_0)^2 + \mathbf{E}\varepsilon^2, \quad (3.3)$$

$$\mathbf{D}(y | X) = \mathbf{D}(\varepsilon | X) = \sigma_\varepsilon^2(X) \quad (3.4)$$

В соотношении (3.3), связывающем между собой безусловные характеристики варьирования, $y_0 = \mathbf{E}y$, $f_0 = \mathbf{E}f$, а символ $\mathbf{E}$ означает усреднение следующих за ним величин по всем возможным значениям как y , так и X . Нуждается в пояснении случай зависимости по схеме B (см. п. 2.4), в которой объясняющая переменная X не является случайной. В этом случае усреднение по X производится по всем наблюдаемым значениям $X_1, X_2, \dots, X_n$, так что, например:

$$\begin{aligned} f_0 = \mathbf{E}y &= \frac{1}{n} \sum_{i=1}^n f(X_i); \\ \sigma_f^2 = \mathbf{E}(f - f_0)^2 &= \frac{1}{n} \sum_{i=1}^n (f(X_i) - f_0)^2; \\ \mathbf{E}\varepsilon^2 &= \frac{1}{n} \sum_{i=1}^n \mathbf{E}\varepsilon^2(X_i). \end{aligned}$$

Что касается условных характеристик, участвующих в (3.4), то они были определены выше (здесь усреднение производится по всем возможным значениям y при заданном фиксированном значении X).

Вернемся непосредственно к обсуждению возможного способа измерения с.т.с.с. Из (3.2) видно, что однозначное восстановление y по X возможно только в ситуации «вырождения» (отсутствия) случайной остаточной компоненты $\varepsilon(X)$ при всех возможных значениях X , что равносильно (при нулевом среднем значении $\varepsilon(X)$) отсутствию какого-либо разброса значений $\varepsilon(X)$ относительно нуля. А это означает на языке вероятностных характеристик тождественное равенство нулю условных дисперсий $\varepsilon(X)$ и $y(X)$, то есть

$$\mathbf{D}(\varepsilon | X) = \mathbf{D}(y | X) = \sigma_\varepsilon^2(X) \equiv 0. \quad (3.5)$$

Другой крайний (с точки зрения с.т.с.с.) случай — это ситуация, в которой знание величины объясняющей переменной X не доставляет никакой информации о значении результирующего показателя y . Эта

¹Знак тождества « $\equiv$ » означает, что среднее значение ε равно нулю при всех возможных значениях объясняющей переменной X .

ситуация возможна, в рамках модели (3.2), лишь тогда (и только тогда), когда

$$f(X) \equiv c = \text{const.} \quad (3.6)$$

В этом случае поведение результирующего показателя y (определенного соотношением (3.2) с учетом (3.6)) как бы повторяет, с точностью до сдвига на постоянную величину c , поведение остаточной случайной компоненты ε и, в частности, в силу (3.3) и с учетом тождества $E(f - f_0)^2 \equiv 0$ имеет в точности ту же самую меру случайного рассеивания, то есть

$$E(y - y_0)^2 = E\varepsilon^2. \quad (3.7)$$

Теперь мы подготовлены к тому, чтобы ввести и проинтерпретировать достаточно универсальный измеритель с.т.с.с. между y и X — коэффициент детерминации y по X . Обозначим его с помощью $K_d(y; X)$ и определим соотношением

$$K_d(y; X) = 1 - \frac{D\varepsilon}{E(y - y_0)^2}, \quad (3.8)$$

где $D\varepsilon = E\varepsilon^2$ — безусловная дисперсия остаточной случайной компоненты, которая может быть подсчитана как усредненная по всем возможным значениям X условная дисперсия $D(\varepsilon | X)$, то есть

$$D\varepsilon = E_x(D(\varepsilon | X)) \quad (3.9)$$

(нижний индекс x у символа математического ожидания E означает, что усреднение проводится по всем возможным значениям X), а $E(y - y_0)^2$ — общая вариация результирующего показателя y , определенная соотношением (3.3).

Легко видеть, что введенный соотношением (3.8) коэффициент детерминации определяет в качестве шкалы измерения показателя с.т.с.с. отрезок $[0; 1]$, причем минимальное (нулевое) значение $K_d(y; X)$ соответствует полному отсутствию какой бы то ни было связи между y и X (поскольку в этом случае, как мы видели, см. (3.7), $D\varepsilon = E(y - y_0)^2$), а максимальное значение $K_d(y; X) = 1$ соответствует случаю чисто функциональной зависимости между y и X , когда значение y может быть в точности восстановлено по значению X с помощью формулы $y = f(X)$ (так как $D\varepsilon$, определенное формулой (3.9) с учетом тождества (3.5), равно нулю). При этом из определения коэффициента детерминации (3.8) с учетом соотношения (3.3) следует, что численное значение $K_d(y; x)$ отражает долю общей вариации результирующего признака y , объясненную изменением функции регрессии $f(x)$.

До сих пор мы оперировали теоретическими характеристиками анализируемых переменных. Однако в статистическом анализе их заменяют

соответствующими *выборочными* характеристиками. Вычисление выборочного (эмпирического) значения коэффициента детерминации y по X должно производиться по формуле

$$\hat{K}_d(y; X) = 1 - \frac{s_\varepsilon^2}{s_y^2}, \quad (3.8')$$

где

$$s_y^2 = \sum_{i=1}^n (y_i - \bar{y})^2 / n, \quad \bar{y} = \left(\sum_{i=1}^n y_i \right) / n,$$

а выборочное значение дисперсии «невязок» ε вычисляется по одной из двух формул:

$$s_\varepsilon^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{f}(X_i))^2, \quad (3.9')$$

если предварительный анализ показал, что условная дисперсия $\mathbf{D}(\varepsilon | X)$ не зависит от X (то есть $\mathbf{D}(\varepsilon | X) = \sigma_\varepsilon^2 = \text{const}$);

$$s_\varepsilon^2 = \frac{1}{n} \sum_{j=1}^s \sum_{i=1}^{\nu_j} (y_{ji} - \bar{y}_j)^2, \quad (3.9'')$$

если условная дисперсия $\mathbf{D}(\varepsilon | X)$ зависит от значения объясняющей переменной X , или во всех случаях, когда вычисления ведутся по *группированной* выборке.

В соотношении (3.9') $\hat{f}(X_i)$ есть статистически оцененное значение неизвестной функции регрессии $f(X)$ в точке $X = X_i$ (методы оценивания функции регрессии описаны в главах 4 и 5). В соотношении (3.9'') использованы *группированные* выборочные данные. В частности, s — это общее число интервалов группирования, ν_j — число выборочных данных, попавших в j -й интервал группирования, y_{ji} — значение результирующего признака в i -м по счету наблюдении, принадлежащем j -му интервалу группирования (нумерация наблюдений своя для каждого интервала группирования), а $\bar{y}_j$ — среднее значение результирующего признака, подсчитанное по наблюдениям, попавшим в j -й интервал группирования, то есть $\bar{y}_j = \left(\sum_{i=1}^{\nu_j} y_{ji} \right) / \nu_j$.

Пример 3.1. Подсчитаем коэффициент детерминации $\hat{K}_d(y, X)$ по данным примера 2.1 (см. таблицу 2.1 и рис. 2.2). Вычисления по форму-

лам (3.8'), (3.9') и (3.9'') дают:

$$\begin{aligned}\bar{y} &= 50,8; \quad s_y^2 = \frac{1}{40} \sum_{i=1}^{40} (y_i - \bar{y})^2 = 979,17; \\ s_\varepsilon^2 &= \frac{1}{40} \sum_{j=1}^4 9\bar{s}^2(x_j^0) = \\ &= \frac{9}{40} (40,96 + 256,00 + 510,76 + 835,21) = 369,96; \\ \hat{K}_d(y; x) &= 1 - \frac{369,96}{979,17} = 1 - 0,377 = 0,623.\end{aligned}$$

Такое значение коэффициента детерминации свидетельствует о наличии статистической связи между среднедушевыми денежными сбережениями семей y (η — в обозначениях примера 2.1) и их среднедушевым доходом x (ξ — в обозначениях примера 2.1), причем степень тесноты этой связи несколько выше среднего уровня. Более подробное обсуждение шкалы значений $\hat{K}_d$, так же как и вопросы проверки этих значений на статистически значимое отличие от нуля, отложим до описания некоторых частных версий этого важного индикатора.

3.2.2 Исследование линейной зависимости y от единственной объясняющей переменной x : парный коэффициент корреляции

Рассмотрение этой схемы зависимости между y и x разобьем на два случая: 1) переменные (x, y) «ведут себя» как двумерная нормальная случайная величина; 2) переменные x и y связаны *линейной* регрессионной зависимостью типа (3.2)-(3.2'), то есть $y(x) = \theta_0 + \theta_1 x + \varepsilon(x)$, где θ_0 и θ_1 — некоторые неизвестные параметры, $\varepsilon(x)$ и x взаимно некоррелированы, а з.р.в. случайных величин y, ε и переменной x не обязан быть нормальным.

1. Случай двумерной нормальности случайных величин (x, y)

Эта схема зависимости относится к типу C по классификации, описанной в п. 2.4.

В данном случае мы, в частности, предполагаем, что двумерная случайная величина $(x, y)^T$ подчиняется двумерному нормальному з.р.в. с вектором средних значений $\mu = (\mu^{(1)}, \mu^{(2)})^T$ и ковариационной матрицей $\Sigma = (\sigma_{jl})_{j,l=1,2}$ (см. Приложение П3.1), то есть

$$\begin{aligned}\mu^{(1)} &= \mathbf{E}x; \quad \mu^{(2)} = \mathbf{E}y; \quad \sigma_{11} = \mathbf{E}(x - \mu^{(1)})^2, \quad \sigma_{22} = \mathbf{E}(y - \mu^{(2)})^2, \\ \sigma_{12} &= \text{cov}(x, y) = \mathbf{E}[(x - \mu^{(1)})(y - \mu^{(2)})], \quad \text{а} \quad r(x; y) = \frac{\sigma_{12}}{(\sigma_{11} \cdot \sigma_{22})^{\frac{1}{2}}}.\end{aligned}$$

Нас интересуют вид функции регрессии y по x и другие вероятные характеристики модели (3.2).

Из свойств двумерного нормального з.р.в. (см. Приложение П2.1) следует:

$$\begin{aligned} \mathbf{E}(y|x) &= \theta_0 + \theta_1 \cdot x, \\ \text{где } \theta_1 &= r \left(\frac{\sigma_{22}}{\sigma_{11}} \right), \quad \theta_0 = \mu^{(2)} - \theta_1 \cdot \mu^{(1)}, \end{aligned} \quad (3.10)$$

а

$$D(y|x) = \sigma_{22}(1 - r^2). \quad (3.11)$$

Мы видим, что если анализируемые переменные $(x, y)^T$ подчинены двумерному нормальному з.р.в., то:

- функция $f(x)$ регрессии y по x линейна по x (см. (3.10));
- условная дисперсия y при фиксированном значении x не зависит от x и определяется формулой (3.11);
- коэффициент детерминации $K_d(y; x)$ связан с парным коэффициентом корреляции r простым соотношением:

$$K_d(y; x) = r^2(x, y) \quad (3.12)$$

(следует (3.8) с учетом (3.11)).

При этом в отличие от коэффициента детерминации (который по определению может быть только неотрицательным числом) коэффициент корреляции может быть как положительным, так и отрицательным. А поскольку он имеет одинаковый знак с угловым коэффициентом наклона θ_1 прямой $f(x) = \theta_0 + \theta_1 x$ (см. соотношение (3.10)), то *положительное значение r свидетельствует о положительном (монотонно возрастающем) характере парной связи y и x , а отрицательное значение r — об отрицательном (монотонно убывающем) характере этой связи.*

Сформулируем и докажем основные свойства парного коэффициента корреляции (справедливые в рамках рассматриваемой двумерной нормальной генеральной совокупности).

С в о й с т в о 3.1. Шкала возможных значений парного коэффициента корреляции r ограничена отрезком от -1 до $+1$, то есть $-1 \leq r \leq 1$, или $|r| \leq 1$.

Для доказательства этого свойства воспользуемся очевидными неравенствами

$$\mathbf{E} \left(\frac{x - \mu^{(1)}}{\sigma_x} \pm \frac{y - \mu^{(2)}}{\sigma_y} \right)^2 \geq 0. \quad (3.13)$$

где $\sigma_k = \sqrt{\sigma_{11}}$ и $\sigma_y = \sqrt{\sigma_{22}}$. Возводя в квадрат *сумму* в круглых скобках, имеем

$$\mathbf{E} \left(\frac{x - \mu^{(1)}}{\sigma_x} \right)^2 + 2\mathbf{E} \left[\frac{(x - \mu^{(1)})(y - \mu^{(2)})}{\sigma_x \sigma_y} \right] + \mathbf{E} \left(\frac{y - \mu^{(2)}}{\sigma_y} \right)^2 \geq 0. \quad (3.14)$$

Поскольку

$$\mathbf{E} \left(\frac{x - \mu^{(1)}}{\sigma_x} \right)^2 = \mathbf{E}(x - \mu^{(1)})^2 / \sigma_x^2 = 1$$

и

$$\mathbf{E} \left(\frac{y - \mu^{(2)}}{\sigma_y} \right)^2 = \mathbf{E}(y - \mu^{(2)})^2 / \sigma_y^2 = 1,$$

а

$$\mathbf{E} \left[\frac{(x - \mu^{(1)})(y - \mu^{(2)})}{\sigma_x \sigma_y} \right] = r \quad (\text{по определению}),$$

то из (3.14) непосредственно следует, что $r \geq -1$. Аналогично рассуждая при анализе выражения, получающегося от возведения в квадрат *разности* в круглых скобках левой части выражения (3.13), получаем, что $r \leq 1$. Объединение полученных двух неравенств дает нам доказательство свойства 1.

Свойство 3.2. Если случайные величины x и y статистически независимы, то $r(x, y) = 0$.

Доказательство этого свойства следует из определения парного коэффициента корреляции с учетом того, что для статистически независимых случайных величин их ковариация равна нулю, то есть

$$\begin{aligned} \mathbf{E}[(x - \mu^{(1)})(y - \mu^{(2)})] &= [\mathbf{E}(x - \mu^{(1)})][\mathbf{E}(y - \mu^{(2)})] = \\ &= (\mu^{(1)} - \mu^{(1)})(\mu^{(2)} - \mu^{(2)}) = 0. \end{aligned}$$

Свойство 3.3. Из факта $|r| = 1$ следует наличие чисто функциональной линейной связи между x и y и наоборот: если x и y связаны чисто функциональной линейной связью, то $|r| = 1$.

Для доказательства этого свойства следует обратить внимание на то, что неравенства $r \leq 1$ и $r \geq -1$ обращаются в точные равенства тогда и только тогда, когда обращаются в точные равенства неравенства (3.13) (по построению). А неравенства (3.13) обращаются в точные равенства тогда и только тогда, когда выражение в круглых скобках левой части этого соотношения тождественно равно нулю. Но из равенства

$$\frac{x - \mu^{(1)}}{\sigma_x} \pm \frac{y - \mu^{(2)}}{\sigma_y} = 0$$

следует, что x и y связаны линейной зависимостью.

Свойство 3.4. Коэффициент корреляции является *симметричной* характеристикой с.т.с.с. между x и y , то есть $r(x, y) = r(y, x)$.

Доказательство этого свойства следует непосредственно из определения $r(x, y)$.

Свойство 3.5. Из равенства нулю коэффициента корреляции (то есть из того, что $r(x, y) = 0$) следует статистическая независимость переменных x и y (доказательство см. в Приложении ПЗ.1).

Дав определение и познакомившись с основными свойствами парного коэффициента корреляции r , мы должны привести формулу, по которой вычисляется эмпирический (выборочный) аналог этой характеристики — $\hat{r}(x, y)$:

$$\hat{r}(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (3.15')$$

Однако для проведения полноценного статистического анализа свойств исследуемой генеральной совокупности, основанных на этой характеристике, *нам необходимо знать ее статистические свойства*. Это позволит судить о точности приближения (3.15) к неизвестному истинному значению, строить статистические критерии для проверки различных гипотез о численных значениях анализируемого коэффициента корреляции.

В частности, *какую величину выборочного коэффициента корреляции следует считать достаточной для статистически обоснованного вывода о наличии корреляционной связи между исследуемыми переменными?* Ведь надежность статистических характеристик, в том числе и $\hat{r}$, ослабевает с уменьшением объема соответствующей выборки, а потому принципиально возможны случаи, когда отклонение от нуля полученной величины выборочного коэффициента корреляции $\hat{r}$ оказывается *статистически незначимым*, то есть целиком обусловленным неизбежным случайным колебанием выборки, на основании которой он вычислен. Ответить на этот вопрос помогает знание закона вероятностного распределения $\hat{r}$. В случае совместной нормальной распределенности исследуемых переменных и при достаточно большом объеме выборки n (а именно при $n > 200$) распределение $\hat{r}$ можно считать приближенно нормальным со средним, равным своему теоретическому значению r , и дисперсией $\sigma_r^2 = \frac{(1-r^2)^2}{n}$.

Однако следует учитывать, что при малых значениях n и r , близких к ± 1 , это приближение оказывается очень грубым. Кроме того, при малых n следует принимать во внимание, что величина $\hat{r}$ является *смещенной*

оценкой своего теоретического значения r , в частности $E\hat{r} = r - [r(1-r^2)]/2n$.

Относительно хорошая степень приближения нормального распределения при малых значениях $|r|$ позволяет получить простой критерий проверки гипотезы

$$H_0: r = 0,$$

то есть гипотезы об отсутствии корреляционной связи между исследуемыми переменными x и y . При этом используется тот факт, что статистика

$$\gamma_n = \frac{\hat{r}\sqrt{n-2}}{\sqrt{1-\hat{r}^2}}$$

при условии справедливости гипотезы H_0 приблизительно распределена по з.р.в. Стьюдента с $n-2$ степенями свободы.

Поэтому если окажется, что

$$\frac{|\hat{r}|\sqrt{n-2}}{\sqrt{1-\hat{r}^2}} > t_{\frac{\alpha}{2}}(n-2),$$

то гипотеза H_0 об отсутствии корреляционной связи между x и y отвергается с вероятностью ошибиться, равной α (здесь, как и ранее, $t_q(m)$ — 100 q %-ная точка t -распределения с m степенями свободы).

Доверительные интервалы для истинного значения коэффициента корреляции r можно построить, используя следующее преобразование, предложенное Р. Фишером:

$$z = \frac{1}{2} \ln \frac{1+\hat{r}}{1-\hat{r}}. \quad (3.16)$$

Он показал, что величина z , определенная соотношением (3.16), уже при небольших n с хорошим приближением следует нормальному закону со средним $Ez \approx \frac{1}{2} \ln \frac{1+r}{1-r} + \frac{r}{2(n-1)}$ и дисперсией $Dz = \frac{1}{n-3}$. Это позволяет построить доверительный интервал $[z_1, z_2]$ для Ex по формуле

$$\begin{aligned} z_{1,2} &= \frac{1}{2} \ln \frac{1+\hat{r}}{1-\hat{r}} \mp \frac{u_{\frac{\alpha}{2}}}{\sqrt{n-3}} - \frac{\hat{r}}{2(n-1)} = \\ &= \operatorname{arcth} \hat{r} \mp \frac{u_{\frac{\alpha}{2}}}{\sqrt{n-3}} - \frac{\hat{r}}{2(n-1)}, \end{aligned}$$

откуда следует, что истинное значение коэффициента корреляции r с той же доверительной вероятностью $1-\alpha$ заключено в пределах

$$\operatorname{th} z_1 < r < \operatorname{th} z_2.$$

Здесь $\operatorname{th} z$ — это тангенс гиперболический от аргумента z (определяется с помощью соотношения $\operatorname{th} z = (e^z - e^{-z})/(e^z + e^{-z})$). Соответственно функция, определяющая величину z с помощью соотношения (3.16), —

это функция, обратная к тангенсу гиперболическому; так что часто вместо $z = \frac{1}{2} \ln \frac{1+\hat{r}}{1-\hat{r}}$ пишут $z = \operatorname{arcth} \hat{r}$ (или $z = \operatorname{th}^{-1} \hat{r}$). Нахождение z по данному значению $\hat{r}$ и, наоборот, определение $\hat{r}$ по заданной величине z производятся с помощью специальной таблицы П1.7 (знаки у аргумента и функции совпадают, так что если, например, $\hat{r}$ отрицателен, то и соответствующее значение $z = \operatorname{argth} \hat{r}$ также отрицательно).

Пример 3.2. По данным $n = 39$ предприятий получен коэффициент корреляции $\hat{r} = -0,654$, характеризующий тесноту связи между себестоимостью продукции (y) и производительностью труда (x). Построим интервальную оценку для r , задавшись 95%-й доверительной вероятностью.

По таблице П1.7 из приложения 1 для $\hat{r} = -0,654$ найдем $z = -0,7823$. Тогда

$$z_1 = -0,7823 - \frac{1,96}{\sqrt{36}} = -1,1090;$$

$$z_2 = -0,7823 + \frac{1,96}{\sqrt{36}} = -0,4556.$$

Теперь по таблице по найденным z_1 и z_2 найдем соответствующие значения $r_1 = -0,804$ и $r_2 = -0,426$.

Таким образом, можно утверждать, что с доверительной вероятностью $P = 0,95$ истинное значение коэффициента корреляции r между себестоимостью продукции y и производительностью труда x будет лежать в интервале от $-0,804$ до $-0,426$, то есть $-0,804 \leq r \leq -0,426$.

Влияние ошибок измерения анализируемых переменных на величину коэффициента корреляции r . Пусть мы хотим оценить степень тесноты корреляционной связи между компонентами двумерной нормальной случайной величины (x, y) , однако наблюдать мы их можем лишь с некоторыми случайными «ошибками измерения» соответственно ε_x и ε_y (см. схему зависимости D в п. 2.4). Поэтому экспериментальные данные (x_i, y_i) , $i = 1, 2, \dots, n$, — это практически выборочные значения искаженной двумерной случайной величины (x', y') , где $x' = x + \varepsilon_x$ и $y' = y + \varepsilon_y$. Если предположить, что ε_x и ε_y взаимно независимы, не зависят от x и y , нормальны, имеют нулевые математические ожидания и конечные дисперсии, соответственно σ_1^2 и σ_2^2 , то двумерная случайная величина (x', y') будет также подчиняться двумерному нормальному распределению. Однако, как легко подсчитать, параметры этого распределения и, в частности, коэффициент корреляции r' между x' и y' будут соответственно отличаться от параметров исходной двумерной схемы (x, y) . Действительно, в соответствии с основными правилами вычисления первых и

вторых моментов получаем

$$\begin{aligned} a_{x'} &= \mathbf{E}x' = \mathbf{E}x = a_x; \\ a_{y'} &= \mathbf{E}y' = \mathbf{E}y = a_y; \\ \sigma_{x'}^2 &= \sigma_x^2 + \sigma_1^2; \\ \sigma_{y'}^2 &= \sigma_y^2 + \sigma_2^2; \\ r' &= \frac{r}{\sqrt{\left(1 + \frac{\sigma_1^2}{\sigma_x^2}\right) \left(1 + \frac{\sigma_2^2}{\sigma_y^2}\right)}}. \end{aligned} \quad (3.17)$$

Из (3.17), в частности, следует, что коэффициент корреляции признаков, на которые наложены ошибки измерения, всегда меньше по абсолютной величине, чем коэффициент корреляции исходных признаков. Другими словами, ошибки измерения всегда ослабляют исследуемую корреляционную связь между исходными переменными, и это искажение тем меньше, чем меньше отношения дисперсий ошибок к дисперсиям самих исходных переменных. Формула (3.17) позволяет скорректировать искаженное значение коэффициента корреляции: для этого нужно либо знать «разрешающие» характеристики погрешностей измерений (и, следовательно, величины дисперсий ошибок σ_1^2 и σ_2^2), либо провести дополнительное исследование по их выявлению.

2. Общий случай парной линейной зависимости. Рассмотрим регрессионную модель вида (3.2) в случае $p = 1$ и линейной функции регрессии $f(X)$, то есть

$$y(x) = \theta_0 + \theta_1 x + \varepsilon(x), \quad (3.18)$$

где θ_0 и θ_1 — неизвестные параметры модели, единственная объясняющая переменная x может быть случайной величиной или неслучайной, $\varepsilon(x)$ и x взаимно некоррелированы, а $\mathbf{E}\varepsilon(x) = 0$ и $\mathbf{D}\varepsilon(x) = \sigma^2(x)$. В отличие от предыдущей схемы в данном случае не требуется двумерной нормальности анализируемой пары переменных (x, y) . Очевидно, рассматриваемая модель может относиться к любому из типов зависимостей B, C и D , описанных в п. 2.4 (все зависит от конкретизации природы исследуемых переменных).

Определенный соотношением (3.15) выборочный коэффициент корреляции может быть формально вычислен для любой двумерной системы **линейной** статистической связи между анализируемыми признаками.

Практически все свойства и рекомендации, касающиеся вычисления, использования статистических свойств и интерпретации парного коэффициента корреляции, остаются *приблизительно* справедливыми и в рамках более общей схемы (3.18) линейной зависимости между x и y . Так, например, из пяти сформулированных выше свойств коэффициента корреляции первые четыре остаются в силе и в рамках модели (3.18)

(«не работает» лишь свойство 5: в общем случае из некоррелированности x и y не следует их статистическая независимость).

Однако только в случае совместной нормальной распределенности исследуемых случайных величин x и y коэффициент корреляции r имеет четкий смысл как характеристика степени тесноты связи между ними. В частности, в этом случае соотношение $|r| = 1$ подтверждает чисто функциональную линейную зависимость между исследуемыми величинами, а уравнение $r = 0$ свидетельствует об их полной взаимной независимости. Кроме того, коэффициент корреляции вместе со средними и дисперсиями случайных величин x и y составляет те пять параметров, которые дают исчерпывающие сведения о стохастической зависимости исследуемых величин, так как однозначно определяют их двумерный закон распределения (см. Приложение П2.1). Во всех же остальных случаях (распределения x и y отклоняются от нормального, одна из исследуемых величин не является случайной и т. п.) коэффициент корреляции можно использовать лишь в качестве одной из возможных характеристик степени тесноты связи. При этом, несмотря на то, что в общем случае пока не предложено характеристики линейной связи, которая обладала бы очевидными преимуществами по сравнению с r , его интерпретация часто оказывается весьма ненадежной. Если же априори *допускается возможность отклонения от линейного вида зависимости, то можно построить примеры, когда, несмотря на $r = 0$, исследуемые переменные оказываются связанными чисто функциональным соотношением* (следовательно, $K_d(x, y) = 1$). Поэтому о величинах, для которых $r = 0$, обычно говорят, что они *некоррелированы*, и только после дополнительного статистического и профессионального анализа (исследование степени отклонения распределения рассматриваемых величин от нормального и т. п.) можно сказать, следует ли отсюда их независимость.

Замечания о необходимости известной осторожности при толковании корреляционной связи никоим образом не обесценивают желательность проверки значимости любого кажущегося соотношения. При этом следует использовать характеристики степени тесноты связи: коэффициента корреляции $\hat{r}$ и корреляционного отношения ρ (см. ниже). Но не всегда знание этих характеристик оказывается достаточным для получения информации о степени тесноты физической связи между исследуемыми переменными и тем более об их причинной взаимообусловленности.

3.2.3 Исследование парных нелинейных связей: корреляционное отношение

При отклонениях исследуемой зависимости от линейного вида, как уже отмечалось, коэффициент корреляции r теряет свой смысл как характеристика степени тесноты связи. В этих случаях исследователь должен воспользоваться имеющимися у него двумерными выборочными

данными $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ с целью построения оценок для определенной выше в некотором смысле универсальной теоретической характеристики степени тесноты связи — коэффициента детерминации $K_d(y; x)$ (см. (3.8')). Способ построения таких оценок выбирается в зависимости от природы имеющихся у нас выборочных данных и от характера некоторых дополнительных допущений.

Корреляционное отношение. Наиболее привлекательной в этом смысле является ситуация, в которой характер выборочных данных (их количество, «плотность» расположения на плоскости) допускает их группировку по оси объясняющей переменной и возможность подсчета так называемых «частных» средних ординат $\bar{y}_j$ внутри каждого (j -го) интервала группирования. Пусть такое группирование данных произведено. При этом, как обычно, s — число интервалов группирования по оси абсцисс; $\nu_j (j = 1, 2, \dots, s)$ — число выборочных точек, попавших в i -й интервал группирования; $\bar{y}_j = \left(\sum_{i=1}^{\nu_j} y_{ji} \right) / \nu_j$ — среднее значение ординат точек, попавших в j -й интервал группирования. Тогда, как легко понять, выборочным аналогом (оценкой) введенной ранее (см. п. 3.2.1) дисперсии σ_f^2 будет величина

$$s_{\bar{y}(x)}^2 = \frac{1}{n} \sum_{j=1}^s \nu_j (\bar{y}_j - \bar{y})^2, \quad (3.19)$$

где общее среднее $\bar{y} = \frac{1}{n} \sum_{j=1}^s \nu_j \bar{y}_j$. Соответственно получаем оценку для $K_d(y; x)$ в виде

$$\hat{\rho}_{yx}^2 = s_{\bar{y}(x)}^2 / s_y^2, \quad (3.20)$$

где выборочная дисперсия s_y^2 индивидуальных результатов наблюдения y_{ji} около общего среднего $\bar{y}$ вычисляется по формуле

$$s_y^2 = \frac{1}{n} \sum_{j=1}^s \sum_{i=1}^{\nu_j} (y_{ji} - \bar{y})^2.$$

Величину $\hat{\rho}_{yx}$ принято называть *корреляционным отношением зависимой переменной y по независимой переменной x* . Его вычисление не обременено никакими дополнительными допущениями относительно общего вида регрессионной зависимости (3.2). Однако в отличие от коэффициента корреляции *корреляционное отношение несимметрично по отношению к исследуемым переменным*, то есть, вообще говоря, $\rho_{yx} \neq \rho_{xy}$. Кроме того, корреляционное отношение, по определению, является величиной неотрицательной¹, так как под ним подразумевается

¹ Иногда, в частности при *монотонном* характере функции регрессии $f(x)$, корреляционному отношению приписывают знак, совпадающий со знаком первой производной этой функции.

результат извлечения арифметического значения корня квадратного из ρ^2 .

В остальном свойства корреляционного отношения во многом похожи на свойства коэффициента корреляции. Из (3.3) и (3.8), в частности, немедленно следует, что подобно коэффициенту корреляции корреляционное отношение не может быть больше единицы.

Из $|\rho| = 1$ следует наличие однозначной функциональной связи между y и x , и, наоборот, однозначная функциональная связь между y и x свидетельствует о том, что $|\rho| = 1$. Далее, отсутствие корреляционной связи между y и x означает, что условные средние $\bar{y}_j$ сохраняют постоянное значение, равное общему среднему $\bar{y}$, а потому $\rho_{yx} = 0$. Наоборот, если $\rho_{yx} = 0$, то $\bar{y}_j = \bar{y}$, и, следовательно, частные средние $\bar{y}_j$ не зависят от x , то есть соответствующая линия регрессии параллельна горизонтальной оси.

Отметим, что между ρ_{yx} и ρ_{xy} нет какой-либо простой зависимости. Некоррелированность y с x (то есть равенство нулю величины ρ_{yx}) не влечет за собой непосредственно некоррелированности x с y . Возможны ситуации, в которых один из этих показателей принимает нулевое значение, в то время как другой равен единице. Допустим, например, что $y = x^2$ и x принимает значения: $-1, 0$ и $+1$ с вероятностями $1/3$ каждое. В этом случае $\rho_{yx} = 1$, $\rho_{xy} = 0$ (в силу симметрии параболы относительно оси y и симметричности распределения x).

Можно показать, что корреляционное отношение ρ не может быть меньше абсолютной величины коэффициента корреляции r , характеризующего зависимость между теми же переменными. В случае линейной зависимости эти две характеристики связи совпадают. Это позволяет использовать величину разности $\hat{\rho}_{yx}^2 - \hat{r}^2$ в качестве меры отклонения регрессионной зависимости от линейного вида (см. соотношение (2.24) в п. 2.5).

И наконец, все замечания относительно смысловой интерпретации коэффициента корреляции r (в частности, о логическом соотношении понятий «корреляционная зависимость, связь между переменными, их причинная взаимообусловленность») остаются в силе и для корреляционного отношения.

Проверка гипотезы об отсутствии корреляционной связи. Какую величину корреляционного отношения можно признать статистически значимо отличающейся от нуля, то есть достаточной для статистически обоснованного вывода о наличии корреляционной связи между исследуемыми переменными? Ведь так же, как и в случае прямолинейного типа зависимости, принципиально возможны ситуации, когда отклонение от нуля полученной величины корреляционного отношения $\hat{\rho}$ является статистически незначимым, то есть обусловленным лишь неизбежными случайными колебаниями выборки. Для построения соответствующего критерия воспользуемся фактом приближенной $F(s-1, n-s)$ -

распределенности случайной величины

$$\hat{F}(0) = \frac{\hat{\rho}_{yx}^2}{1 - \hat{\rho}_{yx}^2} \cdot \frac{n - s}{s - 1}, \quad (3.21)$$

справедливым в предположении, что $K_d(y; x) = 0$ (или, что то же, $\rho_{yx} = 0$) и что условные распределения результирующей переменной $y(x)$ при любом фиксированном x описываются нормальным законом с постоянной дисперсией σ^2 .

Поэтому если окажется, что

$$\frac{\hat{\rho}_{yx}^2}{1 - \hat{\rho}_{yx}^2} \cdot \frac{n - s}{s - 1} > v_\alpha^2(s - 1, n - s),$$

то гипотеза $H: \rho_{yx} = 0$ об отсутствии корреляционной связи между y и x отвергается с вероятностью ошибки α (здесь, как и ранее, $v_\alpha^2(s - 1, n - s) - 100\alpha\%$ -ная точка F -распределения с числом степеней свободы числителя $s - 1$ и знаменателя $n - s$ находится из таблицы П1.5 Приложения 1). При выполнении обратного неравенства значение корреляционного отношения $\hat{\rho}_{yx}$ признается статистически незначимым, то есть делается вывод, что гипотеза об отсутствии корреляционной связи между y и x не противоречит наблюдениям.

Доверительные интервалы для истинного значения корреляционного отношения ρ_{yx} можно построить, опираясь на тот факт, что статистика

$$\gamma^{(n)} = \frac{(n - s)\hat{\rho}_{yx}^2}{(s - 1)(1 - \hat{\rho}_{yx}^2)} \cdot \frac{s - 1}{s - 1 + n\hat{\rho}_{yx}^2}$$

приближенно описывается F -распределением с числом степеней свободы числителя

$$\nu_1^* = \frac{(s - 1 + n\hat{\rho}_{yx}^2)^2}{s - 1 + 2n\hat{\rho}_{yx}^2} \quad (3.22)$$

и числом степеней свободы знаменателя $\nu_2 = n - s$ (см. [Айвазян, Енюков, Мешалкин (1985), п. 1.1.4]).

Таким образом, получаем следующее правило построения приближенных доверительных интервалов для истинного значения корреляционного отношения ρ_{yx} :

1) пользуясь формулой (3.20), вычисляем точечную оценку $\hat{\rho}_{yx}^2$ для истинного значения корреляционного отношения ρ_{yx}^2 ;

2) по формуле (3.22) подсчитываем вспомогательное число степеней свободы ν_1^* числителя для аппроксимирующего F -распределения;

3) задавшись уровнем доверия $P = 1 - 2\alpha$, с помощью таблиц приложения 1 находим $100(1 - \alpha)\%$ -ную точку $v_{1-\alpha}^2(\nu_1^*, n - s)$ и $100\alpha\%$ -ную

точку $v^2\alpha(\nu_1^*, n-s)$ F -распределения с числом степеней свободы числителя ν_1^* и знаменателя $n-s$;

4) утверждаем, что приблизительно с вероятностью $P = 1 - 2\alpha$ истинное значение корреляционного отношения ρ_{yx} удовлетворяет неравенствам

$$\frac{(n-s)\hat{\rho}_{yx}^2}{n(1-\hat{\rho}_{yx}^2)v_\alpha^2} - \frac{s-1}{n} < \rho_{yx}^2 < \frac{(n-s)\hat{\rho}_{yx}^2}{n(1-\hat{\rho}_{yx}^2)v_{1-\alpha}^2} - \frac{s-1}{n}. \quad (3.23)$$

Проиллюстрируем работоспособность описанного метода на следующем примере. Пусть в результате обработки 132 экспериментальных точек $(x_i, y_i) (i = 1, 2, \dots, 132)$ получено выборочное значение корреляционного отношения $\hat{\rho} = 0,60$. При этом мы воспользовались разбиением диапазона изменения независимой переменной на $s = 12$ равных интервалов группирования. Соответственно получаем в качестве вспомогательного числа степеней свободы числителя величину $\nu_1^* = \frac{(12-1+132 \cdot 0,36)^2}{12-1+2 \cdot 132 \cdot 0,36} \approx 27$ (частное округляем до целого числа). Задавшись доверительной вероятностью $P = 0,90$, из таблицы П1.5 находим (полагая $\alpha = 0,05$):

$$v_{0,05}^2(27, 120) = 1,58;$$

$$v_{0,95}^2(27, 120) = \frac{1}{v_{0,05}^2(120, 27)} = \frac{1}{1,73} \approx 0,58.$$

И наконец, в соответствии с формулой (3.23) находим левый ($\hat{\rho}_{\min}^2$) и правый ($\hat{\rho}_{\max}^2$) концы доверительного интервала для истинного значения ρ_{yx}^2 :

$$\hat{\rho}_{\min}^2 = \frac{120 \cdot 0,36}{132 \cdot 0,64 \cdot 1,58} - \frac{11}{132} = 0,24;$$

$$\hat{\rho}_{\max}^2 = \frac{120 \cdot 0,36}{132 \cdot 0,64 \cdot 1,58} - \frac{11}{132} = 0,87.$$

Таким образом, при точечной оценке $\hat{\rho}_{yx} = 0,6$ истинное значение заключено в пределах от $\sqrt{0,24}$ до $\sqrt{0,87}$ с вероятностью, приблизительно равной 0,9, то есть $0,49 < \rho_{yx} < 0,93$.

В этом примере хорошо видна *существенная несимметричность концов интервальной оценки* относительно точечной оценки (правый конец интервальной оценки отстоит от точечной оценки на 0,33, в то время как левый конец — всего лишь на 0,11).

Для значений точечных оценок $\hat{\rho}^2$, близких к нулю или к единице, левый или правый конец интервальной оценки может терять содержательный смысл, выходя за пределы отрезка $[0, 1]$. В этом случае в качестве левого или правого конца интервальной оценки следует брать соответствующее граничное значение — нуль или единицу (причина подобных нежелательных ситуаций — в аппроксимационном подходе к решению

данной задачи). Однако описанный прием все-таки следует признать гораздо более точным, чем применяемый иногда метод построения интервальных оценок для $\rho_{\eta\xi}$, необоснованно использующий приближительную $(\rho, \frac{1-\rho^2}{n})$ -нормальность статистики $\hat{\rho}_{\eta\xi}$.

3.2.4 Исследование линейной зависимости y от нескольких объясняющих переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$: множественный и частные коэффициенты корреляции

Аналогично той схеме, по которой излагался п. 3.2.2, разобьем рассмотрение данной проблемы на два случая: 1) «поведение» вектора переменных $(x^{(1)}, x^{(2)}, \dots, x^{(p)}; y)$ описывается $(p+1)$ -мерным нормальным законом (см. Приложение ПЗ.1; 2) результирующий показатель y связан с объясняющими переменными $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ линейной регрессионной зависимостью типа (3.2).

1. Случай многомерной нормальности векторного признака $(x^{(1)}, x^{(2)}, \dots, x^{(p)}; y)$. Из свойств многомерного нормального закона следуют результаты, аналогичные (3.10)–(3.12) (см. Приложение ПЗ.1). А именно доказано, что функция регрессии y по X имеет *линейный вид*:

$$E(y | X) = \theta_0 + \theta_1 x^{(1)} + \dots + \theta_p x^{(p)}, \quad (3.24)$$

где коэффициенты θ_j ($j = 0, 1, \dots, p$) явным образом выражаются в терминах компонент вектора средних значений μ и элементов ковариационной матрицы Σ анализируемого многомерного нормального распределения, условная дисперсия y (при условии, что объясняющие переменные зафиксированы на уровне X) *не зависит от X* и тоже выражается в явном виде через элементы ковариационной матрицы Σ , а коэффициент детерминации $K_d(y; X)$ равен квадрату множественного коэффициента корреляции $R_{y,X}^2$, который определяется как парный коэффициент корреляции между y и такой комбинацией $x^{(1)}, x^{(2)}, \dots, x^{(p)}$, на которой максимизируется это его значение (подробнее об $R_{y,X}$ см. ниже).

Однако в анализе *множественных* корреляционных связей (так называют статистические связи между более чем двумя переменными в отличие от парных связей, рассмотренных выше) есть своя специфика и возникают принципиально новые проблемы. Эта специфика связана в первую очередь с необходимостью уметь измерять степень тесноты связи между результирующей переменной y и *множеством объясняющих переменных* $x^{(1)}, x^{(2)}, \dots, x^{(p)}$, а также с возникающими трудностями в интерпретации парных коэффициентов корреляции между y и $x^{(j)}$, обусловленными возможным опосредованным влиянием на эту парную связь других (явно не учтенных в вычислении $r(y; x^{(j)})$) объясняющих переменных $x^{(i)}$ ($i \neq j$). Последнее обстоятельство, в частности, делает

необходимым введение таких измерителей статистической связи, которые были бы «очищены» от опосредованного влияния других переменных, давали бы оценку степени тесноты интересующей нас связи между переменными y и $x^{(j)}$ (или $x^{(i)}$ и $x^{(j)}$) при условии, что значения остальных переменных зафиксированы на некотором постоянном уровне. В этом случае говорят о статистическом анализе *частных* (или «очищенных») связей и используют соответственно *частные* («очищенные») коэффициенты корреляции или другие корреляционные характеристики.

Частные коэффициенты корреляции и их выборочные значения. Поставим в соответствие каждой из ранее введенных парных характеристик статистической связи между переменными $x^{(i)}$ и $x^{(j)}$ ($i, j = 0, 1, \dots, p$; $x^{(0)} \equiv y$) *частную* («очищенную») характеристику, определяемую по той же формуле, но только для *условного* распределения (см. п. 2.5, (2.13)) $\varphi(x^{(i)}, x^{(j)} | X^{(i,j)} = x)$. Здесь φ — это функция плотности вероятности переменных $x^{(i)}$ и $x^{(j)}$; $X^{(i,j)}$ — множество переменных, дополняющих пару $(x^{(i)}, x^{(j)})$ до *полного* набора рассматриваемых (наблюдаемых) переменных $X = (x^{(0)}, x^{(1)}, \dots, x^{(p)})$, а x — $(p-1)$ -мерный вектор, определяющий заданные уровни, на которых фиксируются значения «мешающих» переменных $X^{(i,j)}$. Есть два взаимосвязанных обстоятельства, которые препятствуют широкому практическому использованию частных характеристик статистической связи в *общем* (то есть негауссовском) *случае*:

- частные характеристики статистической связи, вообще говоря, зависят от заданных уровней x мешающих переменных (как их выбирать в каждом конкретном случае?);
- для подсчета *выборочных* значений частных характеристик статистической связи необходимо иметь выборку *специальной структуры*, обеспечивающей наличие хотя бы нескольких наблюдений при каждом из заданного ряда фиксированных значений x мешающих переменных.

Однако можно показать, что если исследуемые случайные переменные $(x^{(0)}, x^{(1)}, \dots, x^{(p)})$ подчиняются *многомерному нормальному закону*, то указанные неудобства автоматически исчезают, так как в этом случае частные коэффициенты корреляции не зависят от уровней мешающих переменных x , определяющих условие в соответствующем условном распределении. В частности, имеет место следующая формула (при условии невырожденности $(p+1)$ -мерного нормального закона):

$$r_{ij.X^{(i,j)}} = \frac{-R_{ij}}{(R_{ii}R_{jj})^{\frac{1}{2}}}, \quad (3.25)$$

где $r_{ij.X^{(i,j)}}$ — частный коэффициент корреляции между переменными $x^{(i)}$ и $x^{(j)}$ при фиксированных значениях всех остальных переменных

$X^{(i,j)}$, а R_{kl} — алгебраическое дополнение (см. Приложение П2) для элемента r_{kl} в определителе корреляционной матрицы $\mathbf{R}$ анализируемых признаков $x^{(0)} \equiv y, x^{(1)}, x^{(2)}, \dots, x^{(p)}$, то есть в определителе

$$\det \mathbf{R} = \begin{vmatrix} 1 & r_{01} & r_{02} & \dots & r_{0p} \\ r_{10} & 1 & r_{12} & \dots & r_{1p} \\ \dots & \dots & \dots & \dots & \dots \\ r_{p0} & r_{p1} & r_{p2} & \dots & 1 \end{vmatrix}, \quad \text{где } r_{ij} = r(x^{(i)}, x^{(j)}).$$

Формула (3.25), примененная к *трехмерному* признаку ($x^{(0)} \equiv y, x^{(1)}, x^{(2)}$), при $i = 0, j = 1$ и $X^{(i,j)} = x^{(2)}$ дает:

$$r_{01.x^{(2)}} = r_{01(2)} = \frac{r_{01} - r_{02}r_{12}}{\sqrt{(1 - r_{02}^2)(1 - r_{12}^2)}}. \quad (3.26)$$

Последовательно присоединяя к мешающим переменным все новые признаки из рассматриваемого набора, можно получить *рекуррентные* соотношения для подсчета частных коэффициентов корреляции $r_{01(2\dots k+1)}$ порядка k (то есть при исключении опосредованного влияния k мешающих переменных) по частным коэффициентам корреляции порядка $k - 1$ ($k = 1, 2, \dots, p - 1$):

$$r_{01(2,3,\dots,k+1)} = \frac{r_{01(2,3,\dots,k)} - r_{0k+1(2,\dots,k)} \cdot r_{1k+1(2,\dots,k)}}{\sqrt{(1 - r_{0k+1(2,\dots,k)}^2)(1 - r_{1k+1(2,\dots,k)}^2)}}. \quad (3.26')$$

Выборочные (эмпирические) значения частных коэффициентов корреляции вычисляются по тем же формулам (3.26) — (3.26') с заменой теоретических значений парных коэффициентов корреляции r_{ij} их выборочными аналогами $\hat{r}_{ij}$ (см. формулу (3.15)).

Если исследователь имеет дело лишь с тремя-четырьмя переменными ($p = 2, 3$), то удобно пользоваться рекуррентными соотношениями (3.26'). При больших размерностях анализируемого многомерного признака удобнее опираться на формулу (3.25), использующую расчет соответствующих определителей.

Статистические свойства выборочных частных коэффициентов корреляции (проверка на статистическую значимость их отличия от нуля, доверительные интервалы). При исследовании статистических свойств выборочного частного коэффициента корреляции порядка k (то есть при исключении опосредованного влияния k мешающих переменных) следует воспользоваться тем, что он распределен точно так же, как и обычный (парный) выборочный коэффициент корреляции между теми же переменными с единственной поправкой: объем выборки надо уменьшить на k единиц, то есть полагать его равным $n - k$, а не n . Поэтому при проверке статистически значимого отличия от нуля выборочного

частного коэффициента корреляции и при построении для него доверительных интервалов следует пользоваться рекомендациями п. 3.2.2 для парного коэффициента корреляции с заменой n на $n - k$.

Рассмотрим некоторые конкретные числовые примеры, демонстрирующие возможный характер искажающего опосредованного влияния «третьих факторов» на корреляцию между двумя анализируемыми переменными.

Пример 3.3. По итогам года для 37 однородных предприятий легкой промышленности были зарегистрированы следующие показатели их работы: $x^{(0)} \equiv y$ — среднемесячная характеристика качества ткани (в баллах); $x^{(1)}$ — среднемесячное количество профилактических наладок автоматической линии; $x^{(2)}$ — среднемесячное число обрывов нити.

По матрице исходных данных $(x_i^{(0)}, x_i^{(1)}, x_i^{(2)})_{i=1,37}$ были подсчитаны (с помощью (3.15)) выборочные *парные* коэффициенты корреляции $\hat{r}_{ij}(i, j = 0, 1, 2)$: $\hat{r}_{01} = 0,105$; $\hat{r}_{02} = 0,024$; $\hat{r}_{12} = 0,996$.

Проверка «на статистическую значимость», проведенная в соответствии с рекомендациями п. 3.2.2, свидетельствует об отсутствии статистически значимой парной корреляционной связи между качеством ткани, с одной стороны, и числом профилактических наладок и обрывов нити — с другой, что не согласуется с профессиональными представлениями технолога.

Однако расчет *частных* коэффициентов корреляции по формуле (3.26) дает значения $\hat{r}_{01(2)} = 0,907$; $\hat{r}_{02(1)} = -0,906$, которые вполне соответствуют нашим представлениям о естественном характере связей между изучаемыми показателями.

Доверительные интервалы для *истинных* значений $r_{01(2)}$ и $r_{02(1)}$ (в соответствии с рекомендациями п. 3.2.2) найдем с использованием z -преобразования Фишера для доверительной вероятности $P = 1 - \alpha$. Тогда

$$z_{1,2} = \frac{1}{2} \ln \frac{1 + \hat{r}}{1 - \hat{r}} \mp \frac{u_{\alpha/2}}{\sqrt{(n-1)-3}} - \frac{\hat{r}}{2[(n-1)-1]},$$

где u_q — q -квантиль стандартного нормального распределения (см. таблицу П1.3 и П1.7 Приложения 1).

В нашем примере $n = 37$, $\alpha = 0,05$. Подставляя поочередно в эту формулу значения $\hat{r}_{01(2)} = 0,907$ и $\hat{r}_{02(1)} = -0,906$ и пользуясь упомянутыми таблицами, найдем:

- для $\hat{r} = \hat{r}_{01(2)} = 0,907$: $z_1 = 1,16$ и $z_2 = 1,83$;
- для $\hat{r} = \hat{r}_{02(1)} = -0,906$: $z_1 = -1,83$ и $z_2 = -1,16$,

откуда, вновь воспользовавшись таблицей, окончательно получим:

$$\begin{aligned} 0,820 < r_{01(2)} < 0,950; \\ -0,950 < r_{02(1)} < -0,820. \end{aligned}$$

Пример 3.4. С целью исследования влияния погодных условий на урожайность кормовых трав Хукер (Journ. Roy. Stat. Soc., 1907, v. 65, p. 1) рассмотрел данные Министерства земледелия Англии за 20 лет, характеризующие урожайность $x^{(0)}$ (ц/акр), весеннее количество осадков $x^{(1)}$ (в дюймах) и накопленную за весну сумму «активных» (то есть выше $+5,5^\circ\text{C}$) температур $x^{(2)}$ (в градусах по Фаренгейту) однородной в метеорологическом отношении области Англии, включающей в себя группу восточных графств. По выборке $(x_i^{(0)}, x_i^{(1)}, x_i^{(2)})_{i=\overline{1,20}}$ были подсчитаны основные статистические характеристики изучаемой трехмерной случайной величины:

$$\begin{aligned}\hat{m}_1^{(0)} &= 28,02; & \hat{m}_1^{(1)} &= 4,91; & \hat{m}_1^{(2)} &= 594,0; \\ s_0^2 &= 19,54; & s_1^2 &= 1,21; & s_2^2 &= 7225; \\ \hat{r}_{01} &= 0,80; & \hat{r}_{02} &= -0,40; & \hat{r}_{12} &= -0,56.\end{aligned}$$

Действительно ли высокая температура в период созревания трав отрицательно влияет на их урожайность (ведь $\hat{r}_{02} = -0,40$) или здесь сказывается опосредованное влияние «мешающего» фактора — количества осадков $x^{(1)}$?

Вычисление частных коэффициентов корреляции по рекуррентной формуле (3.26) дает:

$$\hat{r}_{01(2)} = 0,759; \quad \hat{r}_{02(1)} = -0,097; \quad \hat{r}_{12(0)} = -0,436.$$

Как видим, если исключить одновременное влияние количества осадков $x^{(1)}$ на урожайность (с ростом $x^{(1)}$ она повышается) и на сумму активных температур (с ростом $x^{(1)}$ она понижается), то мы уже не обнаружим отрицательной корреляции между температурой и урожайностью ($\hat{r}_{02(1)} = 0,097$, в то время как $\hat{r}_{02} = -0,40$).

Построение доверительных интервалов для $r_{01(2)}$ и $r_{02(1)}$ (с уровнем доверия $P = 0,95$) с использованием z -преобразования Фишера дает в данном случае:

$$0,448 < r_{01(2)} < 0,890; \quad -0,419 < r_{02(1)} < 0,525.$$

Последнее неравенство свидетельствует о том, что у нас нет оснований считать положительную очищенную корреляционную связь между урожайностью и температурой ($r_{02(1)} = 0,097$) статистически значимой, так как нуль находится внутри доверительного интервала.

Измерение с.т.с.с. между результирующим показателем y и множеством объясняющих переменных $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$ в условиях многомерной нормальности (X^T, y) . Множественный коэффициент корреляции $R_{y.X}$. Из (3.8')–(3.9') следует, что вычисление выборочного значения коэффициента детерминации предусматривает проведение предварительных расчетов по статистическому оцениванию неизвестной функции регрессии $f(X)$, что противоречит обычной

хронологии исследований, в соответствии с которой сначала отвечают на вопросы корреляционного анализа и лишь после этого приступают к оценке функции регрессии. Сейчас мы увидим, что свойства многомерных *нормальных* совокупностей автоматически устраняют это неудобство, позволяя вычислять значение $K_d(y; X)$ до проведения регрессионного анализа.

Множественный коэффициент корреляции $R_{y.X}$ используется в качестве измерителя с.т.с.с. между результирующим показателем y и набором объясняющих переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ в моделях **линейной** регрессии. Он определяется как обычный парный коэффициент корреляции между y и линейной функцией регрессии y по X , то есть

$$R_{y.X} = r(y; f(X)), \quad (3.27)$$

где

$$f(X) = E(y | X) = \theta_0 + \theta_1 x^{(1)} + \dots + \theta_p x^{(p)}$$

(можно показать, что следующее определение $R_{y.X}$ является эквивалентным определению (3.27): $R_{y.X}$ — это парный коэффициент корреляции между y и такой линейной комбинацией $x^{(1)}, x^{(2)}, \dots, x^{(p)}$, для которой значение этого парного коэффициента корреляции достигает своего максимума).

При статистической обработке выборок, извлеченных из *нормальных* генеральных совокупностей, множественный коэффициент корреляции $R_{y.X}$ и его выборочное значение $\hat{R}_{y.X}$ обладают рядом удобных свойств (приведенные ниже формулы и свойства теоретического множественного коэффициента корреляции $R_{y.X}$ автоматически переносятся на выборочный $\hat{R}_{y.X}$ заменой участвующих в них теоретических характеристик соответствующими выборочными значениями).

1. *Вычисление $R_{y.X}$ по матрице парных коэффициентов корреляции.* Обозначая, как и прежде, $(p+1) \times (p+1)$ -корреляционную матрицу $(r_{ij})_{i,j=0,1,\dots,p}$ через $\mathbf{R}$, а алгебраическое дополнение элемента r_{kl} в ее определителе через R_{kl} , имеем

$$R_{y.X}^2 = 1 - \frac{\det \mathbf{R}}{R_{00}}. \quad (3.28)$$

2. *Вычисление $R_{y.X}$ по частным коэффициентам корреляции*

$$R_{y.X}^2 = 1 - (1 - r_{01}^2)(1 - r_{02(1)}^2)(1 - r_{03(12)}^2) \dots (1 - r_{0p(12\dots p-1)}^2). \quad (3.29)$$

3. *Множественный коэффициент корреляции мажорирует любой парный или частный коэффициент корреляции, характеризующий статистическую связь результирующего показателя, то есть*

$$R_{y.X} \geq |r_{0j(I_j)}|,$$

где $j = 1, 2, \dots, p$, а I_j — любое подмножество множества индексов $I_0 = \{1, 2, \dots, p\}$, не содержащее индекса j (это соотношение следует из (3.29)). Напоминаем, что $x^{(0)} \equiv y$.

4. Присоединение каждой новой предсказывающей переменной не может уменьшить величины R (независимо от порядка присоединения), то есть

$$R_{y,x(1)} \leq R_{y,(x^{(1)},x^{(2)})} \leq R_{y,(x^{(1)},x^{(2)},x^{(3)})} \leq \dots \leq R_{y,(x^{(1)},x^{(2)},\dots,x^{(p)})}. \quad (3.30)$$

5. Условная дисперсия результирующего показателя $D(y | X)$ не зависит от условия (то есть от значения X) и связана со значением множественного коэффициента корреляции $R_{y,X}$ соотношением

$$D(y | X) = (1 - R_{y,X}^2) Dy. \quad (3.31)$$

Последний результат позволяет связать между собой две характеристики степени тесноты статистической связи — коэффициент детерминации $K_d(y; X)$ и множественный коэффициент корреляции $R_{y,X}$. Действительно, учитывая тот факт, что $D(y | X) = D(\varepsilon | X)$ (см. (3.5)), а значит, и условная дисперсия остатков не зависит от X , получаем, что безусловная дисперсия $D\varepsilon$ (как результат усреднения по X значений условных дисперсий $D(\varepsilon | X)$, см. (3.9)) равна $D(\varepsilon | X)$. Но тогда мы можем подставить в (3.31) $D\varepsilon$ вместо $D(y | X)$ и получим

$$D\varepsilon = (1 - R_{y,X}^2) Dy, \quad (3.31')$$

откуда, учитывая формулу (3.8), имеем:

$$K_d(y; X) = R_{y,X}^2. \quad (3.32)$$

Это означает, что в рамках статистического анализа многомерной нормальной совокупности понятие введенного соотношением (3.8) коэффициента детерминации $K_d(y; X)$ совпадает с квадратом определенного в (3.27) множественного коэффициента корреляции $R_{y,X}^2$ и что коэффициент детерминации $K_d(y; X)$ может быть вычислен в данном случае до проведения регрессионного анализа (то есть до оценки функции регрессии $f(X)$) с помощью формул (3.28), (3.29).

Для проверки гипотезы $H_0: R_{y,X} = 0$, то есть для выяснения вопроса, можно ли считать выборочное значение множественного коэффициента корреляции $\hat{R}_{y,X}$ статистически значимо отличающимся от нуля, пользуются фактом $F(p, n - p - 1)$ -распределенности случайной величины

$$F(\hat{R}) = \frac{\hat{R}_{y,X}^2}{1 - \hat{R}_{y,X}^2} \cdot \frac{n - p - 1}{p},$$

справедливым в рамках рассматриваемой многомерной нормальной совокупности при условии, что истинное значение множественного коэффициента корреляции $R_{y,X}$ равно нулю. Если окажется, что $F(\hat{R}) >$

$> v_{\alpha}^2(p, n - p - 1)$, то гипотеза об отсутствии множественной корреляционной связи между y и $(x^{(1)}, x^{(2)}, \dots, x^{(p)})$ отвергается при уровне значимости критерия, равном α (здесь, как и ранее, $v_{\alpha}^2(p, n - p - 1) - 100\alpha\%$ -ная точка F -распределения с числом степеней свободы числителя p и знаменателя $n - p - 1$ находится из таблицы П1.5 Приложения 1).

Знание следующих статистических свойств оценок $\hat{R}_{y.X}$, определяемых по формулам (3.28), (3.29) с заменой участвующих в них парных и частных коэффициентов корреляции их выборочными аналогами, может оказаться полезным при проведении корреляционного и регрессионного анализов:

- $$E\hat{R}_{y.X}^2 \approx R_{y.X}^2 + \frac{p}{n-1}(1 - R_{y.X}^2), \quad (3.33)$$

что означает наличие положительного смещения (асимптотически устранимого) у оценки $\hat{R}_{y.X}^2$ коэффициента детерминации $R_{y.X}^2$;

- $$D\hat{R}_{y.X}^2 \approx \frac{2p(n-p-1)}{(n-1)(n^2-1)}(1 - R_{y.X}^2)^2 \quad (3.34)$$

(дисперсия $D\hat{R}_{y.X}^2$ характеризует *точность оценивания* коэффициента детерминации $R_{y.X}^2$ с помощью $\hat{R}_{y.X}^2$ и будет использована в регрессионном анализе при определении числа объясняющих переменных, которые следует включить в линейную регрессионную модель);

- подправленная на несмещенность оценка $\hat{R}_{y.X}^{*2}$ коэффициента детерминации $R_{y.X}^2$ имеет вид

$$\hat{R}_{y.X}^{*2} \approx 1 - (1 - \hat{R}_{y.X}^2) \frac{n-1}{n-p-1}. \quad (3.35)$$

Из последней формулы видно, что «подправленная» оценка $\hat{R}_{y.X}^{*2}$ всегда меньше смещенной оценки $\hat{R}_{y.X}^2$.

Отметим, что при малых истинных значениях $R_{y.X}^2$ и при «не слишком малых» величинах отношения p/n подправленные оценки, подсчитанные по формуле (3.35), могут принимать *отрицательные* значения. Можно устранить абсурдность отрицательных значений оценки, используя в качестве «еще раз подправленной» оценки величину

$$\hat{R}_{y.X}^{**2} = \max(\hat{R}_{y.X}^{*2}, 0)$$

(правда, $\hat{R}_{y.X}^{**2}$ уже не будет несмещенной оценкой).

Вернемся к ранее рассмотренным примерам и оценим в них степень тесноты множественной связи между результирующим показателем, с одной стороны, и набором объясняющих переменных — с другой. Будем пользоваться рекомендациями (а именно формулами (3.28), (3.29)),

правомерность которых, напомним, строго обоснована лишь в рамках многомерной нормальной генеральной совокупности.

Пример 3.4 (продолжение). Оценка $\hat{R}_{y.(x^{(1)}x^{(2)})}$ коэффициента множественной корреляции между характеристикой качества ткани y и совокупностью двух факторов: количеством профилактических наладок $x^{(1)}$ и числом обрывов нити $x^{(2)}$, подсчитанная с помощью формулы (3.29), дает:

$$\begin{aligned}\hat{R}_{y.(x^{(1)}x^{(2)})}^2 &= 1 - (1 - \hat{r}_{01}^2)(1 - \hat{r}_{02(1)}^2) = \\ &= 1 - [1 - (0,105)^2][1 - (0,906)^2] = \\ &= 1 - 0,989 \cdot 0,179 = 1 - 0,177 = 0,823.\end{aligned}$$

Отсюда $\hat{R}_{y.(x^{(1)}x^{(2)})} = \sqrt{0,823} = 0,9072$.

Пример 3.4 (продолжение). Оценка $\hat{R}_{y.(x^{(1)}x^{(2)})}$ коэффициента множественной корреляции между урожайностью кормовых трав ($y \equiv x^{(0)}$) и природными факторами — весенним количеством осадков ($x^{(1)}$) и накопленной суммой «активных» температур ($x^{(2)}$), подсчитанная по формуле (3.29), дает:

$$\begin{aligned}\hat{R}_{y.(x^{(1)}x^{(2)})}^2 &= 1 - (1 - \hat{r}_{01}^2)(1 - \hat{r}_{02(1)}^2) = \\ &= 1 - [1 - (0,80)^2][1 - (1,097)^2] = \\ &= 1 - 0,36 \cdot 0,99 = 0,6436.\end{aligned}$$

Отсюда $\hat{R}_{y.(x^{(1)}x^{(2)})} = \sqrt{0,6436} = 0,802$.

2. Исследование линейных множественных связей (общий случай). Речь идет о корреляционном анализе линейной модели вида

$$y = \theta_0 + \theta_1 x^{(1)} + \dots + \theta_p x^{(p)} + \varepsilon(X), \quad (3.36)$$

где $\Theta = (\theta_0, \theta_1, \dots, \theta_p)^T$ — неизвестные параметры модели, объясняющие переменные, могут быть как случайными (схемы зависимостей типов C и D), так и неслучайными (схема зависимости типа B , см. п. 2.4), $\varepsilon(X)$ и $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})$, взаимно некоррелированы, а $E\varepsilon(X) \equiv 0$, $D\varepsilon(X) = \sigma^2(X)$. В отличие от предыдущего случая в модели (3.36) не требуется совместной многомерной нормальности переменных $(y, x^{(1)}, x^{(2)}, \dots, x^{(p)})$.

В статистической практике свойства и рекомендации, справедливые в условиях многомерной нормальной совокупности (относящиеся к частным и множественным коэффициентам корреляции), обычно распространяют и на общий случай (3.36). И, как правило, частные коэффициенты корреляции, определенные соотношениями (3.25)–(3.26'), являются удовлетворительными измерителями очищенной линейной связи

между $x^{(i)}$ и $x^{(j)}$ при фиксированных значениях остальных переменных $X^{(i,j)}$ и в случае, когда распределение анализируемых показателей $(x^{(0)}, x^{(1)}, \dots, x^{(p)})$ отличается от нормального. При этом их можно интерпретировать как показатели тесноты очищенной связи, усредненные по всевозможным значениям фиксируемых на определенных уровнях «мешающих» переменных.

Точно так же множественный коэффициент корреляции $R_{y,X}$, определенный формулами (3.28)–(3.29), может быть формально вычислен для любой многомерной системы наблюдений. И в случае линейной зависимости (3.36) результаты вычисления коэффициента детерминации, полученные по этим формулам, относительно слабо отличаются от результата «прямого счета» по формуле (3.8'). Однако даже при незначительных отклонениях типа функции регрессии $f(X)$ от линейной доверять значению $\hat{R}_{y,X}^2$ как аппроксимации для величины коэффициента детерминации $K_d(y; X)$ не рекомендуется. В этом случае следует пользоваться непосредственно формулой (3.8'), требующей предварительной оценки функции регрессии $f(X)$ в точках $X_1, X_2, \dots, X_n$. Возникающие при этом неудобства (особенно в условиях отсутствия априорной информации об общем виде функции регрессии $f(X)$) преодолевают, в частности, следующим образом:

а) разбивают область возможных значений объясняющих переменных на многомерные аналоги интервалов группирования — гиперпараллелепипеды группирования $\Delta_1, \Delta_2, \dots, \Delta_s$;

б) подсчитывают условные средние $\bar{y}_j$ результирующего показателя по наблюдениям, попавшим в j -й гиперпараллелепипед группирования ($j = 1, 2, \dots, s$);

в) по наблюдениям, попавшим в Δ_j , оценивают условную дисперсию результирующего показателя s_j^2 по отклонениям этих наблюдений от своего условного среднего $\bar{y}_j$ ($j = 1, 2, \dots, s$);

г) с помощью взвешенного усреднения условных дисперсий s_j^2 ($j = 1, 2, \dots, s$) оценивают безусловную дисперсию остатков $D\varepsilon$, участвующую в формуле (3.8');

д) оценив общую дисперсию $s_y^2 = (\sum (y_i - \bar{y})^2)/n$ результирующего показателя, подставляют оценку $D\varepsilon$ и s_y^2 в формулу (3.8') и вычисляют $\hat{K}_d(y, X)$.

По существу эта процедура является многомерным аналогом процедуры вычисления парного корреляционного отношения.

3.3 Корреляционный анализ ранговых (ординальных) переменных: ранговая корреляция

Напомним, что *порядковая (ординальная)* переменная позволяет *упорядочивать* статистически обследованные объекты по степени проявления в них анализируемого свойства. Исследователь обращается к порядковым переменным в ситуациях, когда шкала непосредственного *количественного* измерения степени проявления этого свойства в объекте ему неизвестна (в том числе по причине объективного отсутствия таковой) или имеет условный смысл и интересуется его только как *вспомогательное средство для последующего ранжирования рассматриваемых объектов*. К подобным ситуациям относится рассмотрение таких переменных как «интегральный (сводный) показатель эффективности функционирования социально-экономической системы» (специалиста, предприятия, научно-производственного объединения и т. п.), «качество (мера оптимальности) структуры потребительского бюджета семьи», «качество жилищных условий семьи», «степень прогрессивности предлагаемого проекта решения социально-экономической, технической или другой проблемы» и т. п.

Таким образом, в отличие от статистического анализа k -го ($k = 0, 1, 2, \dots, p$) *количественного* признака $x^{(k)}$, когда в результате его измерения (наблюдения) на объектах мы могли каждому статистически обследованному объекту O_i поставить в соответствие некоторую *измеренную в физически интерпретируемой* шкале числовую характеристику $x_i^{(k)}$, результатом измерения *порядковой* переменной является приписывание каждому из обследованных объектов некоторой *условной числовой метки*. Эта метка может обозначать место объекта в ряду из всех n анализируемых объектов, упорядоченном по убыванию степени проявления в них k -го изучаемого свойства. В этом случае $x_i^{(k)}$ называют *рангом* i -го объекта по k -му признаку, а сам признак — **ранговой ординальной переменной** или просто — **ранговой переменной**.

Упомянутая условная метка может обозначать также *одну из упорядоченных градаций (категорий)*, на которые априори разбиваются анализируемые объекты по степени проявления в них изучаемого свойства. В этом случае сам признак называется **категоризованной ординальной переменной**, однако описание основанного на таких переменных корреляционного анализа мы откладываем до следующего пункта (см. п. 3.4).

3.3.1 Исходные статистические данные (таблица или матрица рангов типа «объект — свойство»)

Итак, в результате измерения $p + 1$ порядковых переменных $x^{(0)} \equiv y, x^{(1)}, \dots, x^{(p)}$ на каждом из n анализируемых объектов $O_1, O_2, \dots, O_n$ мы получаем таблицу (матрицу) исходных данных следующего вида (таблица 3.1).

Таблица 3.1. Матрица рангов типа «объект — свойство»

Порядковый номер объекта («объект»)	Порядковый номер исследуемой переменной («свойство»)						
	0	1	3	...	k	...	p
1	$x_1^{(0)}$	$x_1^{(1)}$	$x_1^{(2)}$	...	$x_1^{(k)}$	...	$x_1^{(p)}$
2	$x_2^{(0)}$	$x_2^{(1)}$	$x_2^{(2)}$	...	$x_2^{(k)}$	...	$x_2^{(p)}$
$\vdots$	...	...	...	...	...	...	...
i	$x_i^{(0)}$	$x_i^{(1)}$	$x_i^{(2)}$	...	$x_i^{(k)}$	...	$x_i^{(p)}$
...	...	...	...	...	...	...	...
n	$x_n^{(0)}$	$x_n^{(1)}$	$x_n^{(2)}$	...	$x_n^{(k)}$	...	$x_n^{(p)}$

В этой таблице элемент $x_i^{(k)}$ задает порядковое место (*ранг*), которое занимает объект O_i в ряду всех статистически обследованных объектов, упорядоченном по убыванию степени проявления k -го анализируемого свойства (то есть по переменной $x^{(k)}$).

Очевидно, если рассмотреть *столбец* с номером k этой таблицы ($k = 0, 1, \dots, p$), то он будет представлять перестановку из n элементов, а именно перестановку из n натуральных чисел $1, 2, \dots, n$, определяющую порядковые места объектов $O_1, O_2, \dots, O_n$ в ряду, упорядоченном по свойству $x^{(k)}$.

Замечание о случаях неразличимости рангов («объединенные ранги»). При упорядочении объектов по какому-либо свойству $x^{(k)}$ ($k = 0, 1, \dots, p$) могут встретиться ситуации, когда два объекта или целая группа их оказываются неразличимыми с точки зрения степени проявления в них этого свойства. Тогда каждому из объектов этой однородной группы приписывается ранг, равный среднему арифметическому значению тех мест, которые они делят, а полученные таким образом ранги принято называть «объединенными» (или «связными»). Так, например, упорядочивая семь альтернативных проектов A, B, C, D, E, F, G перспективного развития некоторой подотрасли с точки зрения их народнохозяйственной эффективности, эксперт поставил на 1-е место проект C , на 2-е — проект A , далее располагал проекты B, D и E , которые считал неразличимыми (равноценными) по эффективности, а последнее

место отвел проектам F и G . Тогда соответствующий столбец таблицы «объект — свойство» будет состоять из следующих компонент:

$$x_A = 2; \quad x_B = \frac{3 + 4 + 5}{3} = 4; \quad x_C = 1; \quad x_D = x_E = \frac{3 + 4 + 5}{3} = 4;$$

$$x_F = x_G = \frac{6 + 7}{2} = 6,5.$$

Мы видим, что появление объединенных рангов может привести к *дробным значениям* рангов, составляющих массив исходных статистических данных (значения рангов, соответствующие 6-му и 7-му проектам). При отсутствии объединенных рангов область возможных значений переменных $x^{(k)}$, очевидно, ограничивается множеством первых n чисел натурального ряда, где n — число сравниваемых объектов.

Мы увидим далее, что наличие объединенных рангов несколько усложняет вычислительные процедуры, связанные со статистическим анализом соответствующих корреляционных характеристик.

3.3.2 Понятие ранговой корреляции

Под *ранговой корреляцией* понимается статистическая связь между ранговыми порядковыми переменными. В статистической практике эта связь анализируется на основании исходных статистических данных, представленных упорядочениями (ранжировками) n рассматриваемых объектов по разным свойствам (см. столбцы таблицы 3.1). Есть ли хоть какая-то согласованность (или связь) между упорядочением анализируемых объектов по свойству $x^{(k)}$ и упорядочением тех же объектов по другому свойству $x^{(j)}$? Можно ли измерить и проанализировать совокупную статистическую связь, существующую между ранжировками одних и тех же объектов $O_1, O_2, \dots, O_n$, полученными в соответствии со степенью проявления в них сначала свойства $x^{(k_1)}$ (1-й способ упорядочения), затем — свойства $x^{(k_2)}$ (2-й способ упорядочения) ..., наконец свойства $x^{(k_m)}$ ($m \geq 2$)? Таким образом, речь идет о системе понятий и методов, позволяющих измерять и анализировать статистическую связь, существующую между двумя или несколькими ранжировками одного и того же конечного множества объектов $O_1, O_2, \dots, O_n$.

Система этих понятий и методов и составляет раздел математической статистики, который принято называть анализом *ранговых корреляций*. Методы ранговой корреляции широко используются, в частности, при организации и статистической обработке различного рода систем экспертных обследований.

3.3.3 Основные задачи статистического анализа связей между ранжировками

Предположим, мы ввели измерители парной и множественной ранговой статистической связи (см. ниже, пп. 3.3.4–3.3.7). Тогда, опираясь на эти характеристики, исследователь чаще всего пытается решить следующие три основные задачи статистического анализа структуры и характера связей, существующих между изучаемыми порядковыми переменными.

Задача А: *анализ структуры имеющейся совокупности упорядочений* $X^{(k)} = (x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)})^T$, $k = 0, 1, \dots, p$. Интерпретируя каждое упорядочение $X^{(k)}$ как точку в n -мерном пространстве, можно представить, например, три наиболее характерных типа такой структуры: 1) анализируемые точки *равномерно* разбросаны по всей области своих возможных значений (определяемой неравенствами $1 \leq x_i^{(k)} \leq n$, $i = 1, 2, \dots, n$), что означает отсутствие какой-либо связи или согласованности в представляемых ими ранжировках; 2) расположение $p + 1$ точек таково, что часть из них образует ядро из близко лежащих друг от друга точек («сгусток»), а остальные произвольно разбросаны относительно этого ядра. В этом случае существование ядра обеспечивает наличие подмножества согласованных переменных; 3) анализируемые точки — ранжировки располагаются в пространстве несколькими относительно далеко отстоящими друг от друга ядрами («сгустками»), что означает наличие нескольких подмножеств переменных таких, что переменные внутри одного подмножества обнаруживают высокую статистическую взаимосвязь, тогда как согласованности между переменными, взятыми из разных таких подсовокупностей, практически не существует.

Задача В: *анализ интегральной (совокупной) согласованности рассматриваемых переменных и их условная ранжировка по критерию степени тесноты связи каждой из них с остальными переменными*. Подобные задачи возникают, например, при исследовании степени согласованности мнений группы экспертов и при попытках условного упорядочения последних по их компетентности. В основе этого анализа лежит расчет коэффициента совокупной согласованности — *коэффициента конкордации* для различных комбинаций исследуемых переменных (см. п. 3.3.7).

Задача С: *построение единого группового упорядочения объектов на основе совокупности согласованных упорядочений «ядра»* (или нескольких групповых упорядочений — при наличии нескольких «ядер»). Решение этой задачи сводится к построению такого упорядочения, которое было бы, в определенном смысле, наиболее близким к каждому из упорядочений заданной совокупности — «ядра». Именно с такой задачей сталкивается, например, исследователь, желающий установить *неизвестное истинное упорядочение* заданной совокупности объектов по

имеющемуся в его распоряжении набору экспертных ранжировок тех же объектов. Для построения единого (группового) варианта упорядочения $X^{(0)}$ часто используют в качестве ранга $x^{(0)}$ объекта O_i среднее арифметическое или медиану имеющихся базовых рангов $x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)}$ этого объекта. Обоснование способа построения единого варианта упорядочения может быть получено, например, в рамках подхода, который опирается на меру близости между ранжировками (определяется ранжировка $X^{(0)}$, наименее удаленная, в смысле введенной меры близости, от всех ранжировок $X^{(1)}, X^{(2)}, \dots, X^{(p)}$ базовой совокупности). Задача C может быть сформулирована и как задача наилучшего (в определенном смысле) восстановления ранжировки $X^{(0)}$, связанной с результирующей переменной $y \equiv x^{(0)}$, по ранжировкам $X^{(1)}, X^{(2)}, \dots, X^{(p)}$, индуцируемым соответственно объясняющими переменными $x^{(1)}, x^{(2)}, \dots, x^{(p)}$. В такой формулировке ее называют также *задачей регрессии на порядковых (ординальных) переменных*.

3.3.4 Ранговый коэффициент корреляции Спирмэна

Для измерения степени тесноты связи между ранжировками $X^{(k)} = (x_1^{(k)}, x_2^{(k)}, \dots, x_n^{(k)})^T$ и $X^{(j)} = (x_1^{(j)}, x_2^{(j)}, \dots, x_n^{(j)})^T$ К. Спирмэн еще в 1904 г. предложил показатель

$$\hat{r}_{kj}^{(s)} = 1 - \frac{6}{n^3 - n} \sum_{i=1}^n (x_i^{(k)} - x_i^{(j)})^2, \quad (3.37)$$

названный впоследствии *ранговым коэффициентом корреляции Спирмэна*. Прямым подсчетом нетрудно убедиться, что для *совпадающих* ранжировок (то есть при $x_i^{(k)} = x_i^{(j)}$ для всех $i = 1, 2, \dots, n$) $\hat{r}_{kj}^{(s)} = 1$ а для *противоположных* (то есть при $x_i^{(k)} = n - x_i^{(j)} + 1$, $i = 1, 2, \dots, n$) — $\hat{r}_{kj}^{(s)} = -1$. Можно показать, что во всех остальных случаях $|\hat{r}_{kj}^{(s)}| < 1$.

Формула (3.37) пригодна лишь в случае отсутствия объединенных рангов в обеих исследуемых ранжировках. Для ее распространения на общий случай определим для каждой (k -й) ранжировки $X^{(k)}$ ($k = 0, 1, \dots, p$) величину

$$T^{(k)} = \frac{1}{12} \sum_{t=1}^{m^{(k)}} [(n_t^{(k)})^3 - n_t^{(k)}], \quad (3.38)$$

где $m^{(k)}$ — число групп неразличимых рангов у переменной $x^{(k)}$, а $n_t^{(k)}$ — число элементов (рангов), входящих в t -ю группу неразличимых рангов (в частном случае отсутствия объединенных рангов имеем $m^{(k)} = n$, $n_1^{(k)} = n_2^{(k)} = \dots = n_n^{(k)} = 1$ и соответственно $T^{(k)} = 0$; кроме того, группы неразличимых рангов, состоящие из единственного элемента, по существу, не участвуют в расчете величины $T^{(k)}$).

Тогда ранговый коэффициент корреляции Спирмэна между ранжировками $X^{(k)}$ и $X^{(j)}$ следует вычислять по формуле

$$\hat{\tau}_{kj}^{(s)} = \frac{\frac{1}{6}(n^3 - n) - \sum_{i=1}^n (x_i^{(k)} - x_i^{(j)})^2 - T^{(k)} - T^{(j)}}{\sqrt{\left[\frac{1}{6}(n^3 - n) - 2T^{(k)}\right] \left[\frac{1}{6}(n^3 - n) - 2T^{(j)}\right]}}. \quad (3.39)$$

Если $T^{(k)}$ и $T^{(j)}$ являются небольшими относительно $\frac{1}{6}(n^3 - n)$ величинами, то можно воспользоваться приближенным соотношением (а при $T^{(k)} = T^{(j)}$ оно точное)

$$\hat{\tau}_{kj}^{(s)} = 1 - \frac{\sum_{i=1}^n (x_i^{(k)} - x_i^{(j)})^2}{\frac{1}{6}(n^3 - n) - (T^{(k)} + T^{(j)})}. \quad (3.39')$$

Правда, при этом же условии (относительная малость $T^{(k)} + T^{(j)}$ по сравнению с $\frac{1}{6}(n^3 - n)$) и приближенная формула (3.37) дает хорошую точность.

Пример 3.5. Два эксперта проранжировали 10 предложенных им проектов реорганизации научно-производственного объединения (НПО) с точки зрения их эффективности (при заданных ресурсных ограничениях). Пронумеровав проекты в порядке ранжировки 1-го эксперта, получаем в качестве исходных данных: $X^{(1)} = (1; 2; 3; 4; 5; 6; 7; 8; 9; 10)^T$; $X^{(2)} = (2; 3; 1; 4; 6; 5; 9; 7; 8; 10)^T$.

Вычисления по формуле (3.37) дают:

$$\tau_{12}^{(s)} = 1 - \frac{6}{1000 - 10} (1 + 1 + 2^2 + 0 + 1^2 + 1^2 + 2^2 + 1 + 1 + 0) = 1 - \frac{6}{990} \cdot 14 = 0,915,$$

что свидетельствует о существенной положительной ранговой связи между исследуемыми переменными.

Пример 3.6. Десять однородных предприятий подотрасли были проранжированы вначале по степени прогрессивности их оргструктур (признак $x^{(1)}$), а затем — по эффективности их функционирования в отчетном году (признак $x^{(2)}$). В результате были получены следующие две ранжировки: $X^{(1)} = (1; 2,5; 2,5; 4,5; 4,5; 6,5; 6,5; 8; 9,5; 9,5)^T$; $X^{(2)} = (1; 2; 4,5; 4,5; 4,5; 4,5; 8; 8; 8; 10)^T$.

В первой ранжировке имеем четыре группы неразличимых рангов, число элементов в которых больше единицы, а во второй ранжировке — две такие группы. В соответствии с формулой (3.38) получаем:

$$T^{(1)} = \frac{1}{12} [(2^3 - 2) + (2^3 - 2) + (2^3 - 2) + (2^3 - 2)] = \frac{24}{12} = 2,00;$$

$$T^{(2)} = \frac{1}{12} [(4^3 - 4) + (3^3 - 3)] = 7,42.$$

Точная формула (3.39) дает $\hat{\tau}_{12}^{(s)} = 0,917$. Вычисление этого же коэффициента корреляции по приближенным формулам (3.37) и (3.39') дает, соответственно, значения 0,921 и 0,917. Все эти результаты оказываются совпадающими при округлении до второго десятичного знака.

3.3.5 Ранговый коэффициент корреляции Кендалла

Другой широко используемой характеристикой тесноты статистической связи между двумя упорядочениями является ранговый коэффициент корреляции Кендалла, определяемый соотношением

$$\hat{\tau}_{kj}^{(K)} = 1 - \frac{4\nu(X^{(k)}, X^{(j)})}{n(n-1)}, \quad (3.40)$$

где $\nu(X^{(k)}, X^{(j)})$ — минимальное число обменов соседних элементов последовательности $X^{(j)}$, необходимое для приведения ее к упорядочению $X^{(k)}$. Очевидно, величина $\nu(X^{(k)}, X^{(j)})$ симметрична относительно своих аргументов, так что с равным правом можно говорить о минимальном числе «соседских обменов» элементов последовательности $X^{(k)}$, необходимом для приведения к виду $X^{(j)}$.

Из (3.40) сразу следует, что при совпадающих ранжировках $X^{(k)}$ и $X^{(j)}$ $\hat{\tau}_{kj}^{(K)} = 1$ (так как $\nu(X^{(k)}, X^{(j)}) = 0$), а при противоположных (то есть при $x_i^{(k)} = n - x_i^{(j)} + 1$, $i = 1, 2, \dots, n$, так что $\nu(X^{(k)}, X^{(j)}) = \frac{1}{2}n \cdot (n-1)$), $\hat{\tau}_{kj}^{(K)} = -1$. Нетрудно показать, что во всех остальных случаях $|\tau_{kj}^{(K)}| < 1$.

Вычисление $\hat{\tau}_{kj}^{(K)}$ связано с необходимостью подсчета величины $\nu(X^{(k)}, X^{(j)})$ и, следовательно, является более *трудоемким*, чем вычисление $\tau_{kj}^{(S)}$. Однако, во-первых, коэффициент Кендалла обладает некоторыми преимуществами по сравнению с коэффициентом Спирмэна, главные из них: а) относительно большая продвинутость в исследовании его статистических свойств и, в частности, его выборочного распределения (см. ниже); б) возможность его использования и в частной («очищенной») корреляции рангов; в) большие удобства его пересчета при добавлении к n статистически обследованным объектам новых, то есть при удлинении анализируемых ранжировок: для вычисления нового значения рангового коэффициента корреляции приходится переранжировать значительную часть объектов, что в случае $\tau_{kj}^{(S)}$ означает необходимость пересчета разностей $x_i^{(k)} - x_i^{(j)}$; при вычислении же $\tau_{ij}^{(K)}$ значения рангов не играют никакой роли, важно лишь число необходимых «соседских обменов», которое при добавлении новых объектов подсчитывается рекуррентным способом (к старому значению $\nu(X^{(k)}, X^{(j)})$ может быть лишь дополнен некоторый «добавок»).

Во-вторых, можно воспользоваться рекомендациями, упрощающими подсчет числа $\nu(X^{(k)}, X^{(j)})$ как при ручном, так и при машинном счете.

Так, при ручном счете полезным оказывается известный факт тождественного совпадения величин $\nu(X^{(k)}, X^{(j)})$ и $I(X^{(k)}, X^{(j)})$, где *число инверсий* $I(X^{(k)}, X^{(j)})$ — это просто число расположенных в неординаковом порядке пар элементов последовательностей $X^{(k)}$ и $X^{(j)}$, являюще-

еся естественной мерой нарушения порядка объектов в одной последовательности относительно другой. Для удобства подсчета $I(X^{(k)}, X^{(j)})$ перенумеруем объекты в порядке, определяемом рангами последовательности $X^{(k)}$. Тогда анализируемые ранжировки $X^{(k)}, X^{(j)}$ соответствующим образом видоизменяются, то есть преобразуются к виду соответственно $\tilde{X}^{(k)}, \tilde{X}^{(j)}$, где $\tilde{X}^{(k)} = (1, 2, \dots, n)^\top$; $\tilde{X}^{(j)} = (\tilde{x}_1^{(j)}, \tilde{x}_2^{(j)}, \dots, \tilde{x}_n^{(j)})^\top$, а число инверсий $I(X^{(k)}, X^{(j)}) \equiv I(\tilde{X}^{(k)}, \tilde{X}^{(j)})$, а следовательно, и величина $\nu(X^{(k)}, X^{(j)})$ определяется по формуле

$$\nu(X^{(k)}, X^{(j)}) = I(X^{(k)}, X^{(j)}) = \sum_{q=1}^{n-1} \sum_{l=q+1}^n \nu_{ql}^{(j,k)}, \quad (3.41)$$

где

$$\nu_{ql}^{(j,k)} = \begin{cases} 1, & \text{если } \tilde{x}_q^{(j)} > \tilde{x}_l^{(j)} \quad (\text{то есть нарушен порядок} \\ & \text{последовательности } \tilde{X}^{(k)}); \\ 0 & - \text{в противоположном случае.} \end{cases}$$

Легко подсчитать, что число инверсий $I(X^{(k)}, X^{(j)})$ может меняться от 0 (что соответствует случаю совпадающих ранжировок) до $\frac{1}{2}n(n-1)$ (что соответствует случаю противоположных ранжировок).

Формулы (3.40)–(3.41) пригодны для подсчета $\hat{\tau}_{kj}^{(K)}$ лишь в случае отсутствия объединенных рангов в обеих исследуемых ранжировках. Соответствующее «подправленное» значение $\hat{\tau}_{kj}^{*(K)}$ при наличии объединенных рангов в анализируемых упорядочениях будет определяться соотношением

$$\hat{\tau}_{kj}^{*(K)} = \frac{\hat{\tau}_{kj}^{(K)} - \frac{2(U^{(k)} + U^{(j)})}{n(n-1)}}{\sqrt{\left(1 - \frac{2U^{(k)}}{n(n-1)}\right) \left(1 - \frac{2U^{(j)}}{n(n-1)}\right)}}, \quad (3.40')$$

в котором коэффициент $\hat{\tau}_{kj}^{(K)}$ вычисляется по формуле (3.40)–(3.41), а «поправочные» величины $U^{(l)}$ определяются соотношением

$$U^{(l)} = \frac{1}{2} \sum_{t=1}^{m^{(l)}} n_t^{(l)} (n_t^{(l)} - 1), \quad l = k, j \quad (3.42)$$

(смысл величин $m^{(l)}$ и $n_t^{(l)}$ определен выше, см. (3.38)).

Для пояснения работоспособности формул (3.40)–(3.42) вернемся к примерам 3.5 и 3.6.

Анализ степени согласованности ранжировок двумя экспертами десяти проектов реорганизации НПО (пример 3.5), осуществленный с использованием формул (3.40)–(3.41), дает:

$$\begin{aligned}
\nu_{12} &= 0; & \nu_{13} &= 1, \nu_{14} = \nu_{15} = \nu_{16} = \nu_{17} = \nu_{18} = \nu_{19} = \nu_{1.10} = 0; \\
\nu_{23} &= 1; & \nu_{24} &= \nu_{25} = \nu_{26} = \nu_{27} = \nu_{28} = \nu_{29} = \nu_{2.10} = 0; \\
\nu_{34} &= \nu_{35} = \nu_{36} = \nu_{37} = \nu_{38} = \nu_{39} = \nu_{3.10} = 0; \\
\nu_{45} &= \nu_{46} = \nu_{47} = \nu_{48} = \nu_{49} = \nu_{4.10} = 0; \\
\nu_{56} &= 1; & \nu_{57} &= \nu_{58} = \nu_{59} = \nu_{5.10} = 0; \\
\nu_{67} &= \nu_{68} = \nu_{69} = \nu_{6.10} = 0; \\
\nu_{78} &= 1; & \nu_{79} &= 1; & \nu_{7.10} &= 0; \\
\nu_{89} &= \nu_{8.10} = 0; \\
\nu_{9.10} &= 0.
\end{aligned}$$

Таким образом, $\nu(X^{(1)}, X^{(2)}) = 1 + 1 + 0 + 0 + 1 + 0 + 2 + 0 + 0 = 5$.

Соответственно

$$\hat{\tau}_{12}^{(K)} = 1 - \frac{4 \cdot 5}{10 \cdot 9} = 1 - 0,222 = 0,778$$

(напомним, что коэффициент Спирмэна в этом примере был равным 0,915).

При вычислении рангового коэффициента корреляции Кендалла в примере 3.6 следует воспользоваться формулой (3.40'), так как исследуемые ранжировки содержат объединенные ранги. Используя результаты расчета величин $m^{(1)} = 4$, $m^{(2)} = 2$, $n_1^{(1)} = n_2^{(1)} = n_3^{(1)} = n_4^{(1)} = 2$, $n_1^{(2)} = 4$, $n_2^{(2)} = 3$, получаем (в соответствии с (3.42)):

$$U^{(1)} = \frac{1}{2}(2 + 2 + 2 + 2) = 4; \quad U^{(2)} = \frac{1}{2}(4 \cdot 3 + 3 \cdot 2) = 9.$$

Обращаясь теперь к формуле (3.40'), имеем:

$$\hat{\tau}_{12}^{*(K)} = \frac{1 - \frac{26}{90}}{\sqrt{(1 - \frac{8}{90})(1 - \frac{18}{90})}} = 0,833$$

(напомним, что соответствующий коэффициент Спирмэна был равен 0,917).

Отметим, что существует обобщенная формула для парного коэффициента корреляции, из которой в качестве частных случаев могут быть получены *обычный парный коэффициент корреляции* (3.15), а также *ранговые коэффициенты корреляции Спирмэна и Кендалла* (см. например, [Айвазян, Мхитарян (2001), п. 11.3.6]). Заметим также, что значения ранговых коэффициентов корреляции Спирмэна и Кендалла, при условии, что абсолютные величины этих значений не слишком близки к единице, связывает следующее простое приближенное соотношение: $\hat{\tau}^{(S)} \approx 1,5 \hat{\tau}^{(K)}$.

3.3.6 Статистические свойства выборочных характеристик парной ранговой связи

До сих пор речь шла о *выборочных* характеристиках ранговой связи. Попробуем ответить на вопрос: как точно эти выборочные характеристики (определенные, в частности, формулами (3.37)–(3.42)) оценивают соответствующие истинные (*теоретические*) значения?

Для этого в первую очередь следует пояснить, что в данном случае понимается под теоретическими характеристиками.

Представим себе сначала *конечную* генеральную совокупность, состоящую из N объектов $O_1, O_2, \dots, O_N$, каждый из которых снабжен двумя порядковыми номерами: $O_i \leftrightarrow (x_i^{(k)}, x_i^{(j)})$, $i = 1, 2, \dots, N$, где $x_i^{(l)}$ означает место объекта O_i в общем ряду всех N объектов, упорядоченном по степени выраженности свойства $x^{(l)}$ ($l = k, j$). Будем полагать, что статистически обследованное множество объектов $O_{i_1}, O_{i_2}, \dots, O_{i_n}$ образуется как случайная выборка объема n , взятая из совокупности $O_1, O_2, \dots, O_N$ ($n \ll N$).

Определим *теоретические* (истинные) значения коэффициентов $\tau_{kj}^{(S)}$ и $\tau_{kj}^{(K)}$ соответственно теми же соотношениями (3.37) (или (3.39)) и (3.40) (или (3.40')), что и выборочные, с заменой объема выборки n объемом генеральной совокупности N . При работе с выборкой производится *естественная перенумерация* объектов и их рангов, не меняющая их упорядоченности в генеральной совокупности ни по одной из переменных.

В дальнейшем нас будет интересовать, как сильно могут отличаться выборочные значения $\hat{\tau}^{(S)}$ и $\hat{\tau}^{(K)}$ от соответствующих теоретических, в том числе в так называемых *асимптотических ситуациях*, то есть при $N \rightarrow \infty$ и $n(N) \rightarrow \infty$.

Проверка статистически значимого отличия от нуля ранговых корреляционных характеристик осуществляется при «не слишком малых» n ($n > 10$) и заданном уровне значимости критерия α с помощью неравенств

$$|\hat{\tau}^{(S)}| > t_{\frac{\alpha}{2}}(n-2) \sqrt{\frac{1 - (\hat{\tau}^{(S)})^2}{n-2}}; \quad (3.43)$$

$$|\hat{\tau}^{(K)}| > u_{1-\alpha/2}(n-2) \sqrt{\frac{2(2n+5)}{9n(n-1)}}, \quad (3.44)$$

в которых $t_q(\nu)$ и u_q , как и прежде, соответственно 100 q %-ная точка $t(\nu)$ - и q -квантиль стандартного нормального распределения (см. таблицы П1.3 и П1.6 в Приложении 1). Выполнение неравенств (3.43) и (3.44) сигнализирует о необходимости отвергнуть гипотезу об отсутствии статистически значимой ранговой корреляционной связи. В случае небольших объемов выборок ($4 \leq n \leq 10$) статистическая проверка гипотезы

об отсутствии ранговой корреляционной связи производится с помощью специальных таблиц П1.9 и П1.10 (см. Приложение 1).

Таблица П1.9 значений вспомогательной величины S_C позволяет при малых n ($n = 4, 5, \dots, 10$) построить то пороговое значение $\tau_{\max}^{(S)}$, при превышении которого (по абсолютной величине) коэффициентом Спирмена $\hat{\tau}^{(S)}$ следует признать наличие статистически значимой связи между анализируемыми переменными. Задавшись уровнем значимости критерия α и числом сравниваемых объектов n , определяем из таблицы величину $S_C = S_C(n, Q)$, соответствующую нашему n и значению $Q = \alpha/2$ (или *приблизительно* равному $\alpha/2$). Тогда

$$\tau_{\max}^{(S)} = \frac{2S_C(n, Q)}{K_n} - 1, \quad (3.45)$$

где $K_n = \frac{1}{3}(n^3 - n)$ (значения этой вспомогательной константы приведены в последней строке таблицы).

Так, в примере 3.5 для уровня значимости $\alpha = 0,06$ имеем: $n = 10$; $Q = 0,03$; $S_C = S_C(10; 0,3) = 268$; $K_{10} = 330$, так что в соответствии с (3.45)

$$\tau_{\max}^{(S)} = \frac{2 \cdot 268}{330} - 1 = 0,624.$$

Поскольку выборочное значение рангового коэффициента корреляции Спирмена $\tau^{(S)}$ в этом примере значительно превосходит пороговое значение ($\hat{\tau}^{(S)} = 0,915 > 0,624$), то гипотеза об отсутствии корреляционной связи отвергается.

И наконец, в таблице П1.10 значений вспомогательной величины S_K приведены значения величин S_K , позволяющие вычислить (при малых $n = 4, 5, \dots, 10$) то пороговое значение $\tau_{\max}^{(K)}$, при превышении которого (по абсолютной величине) коэффициентом Кендалла следует признать наличие статистически значимой связи между анализируемыми переменными. Для этого поступают следующим образом: задавшись объемом выборки n и уровнем значимости критерия α , находят в столбце, соответствующем данному n , величину, равную (или *приблизительно* равную) $\alpha/2$; затем находят значение $S_K = S_K(n, \alpha)$ в левом столбце *той же самой строки* и вычисляют $\tau_{\max}^{(K)}$ по формуле

$$\tau_{\max}^{(K)} = \frac{2S_K(n, \alpha)}{n(n-1)}. \quad (3.46)$$

Если окажется, что $\hat{\tau}^{(K)} > \tau_{\max}^{(K)}$, то гипотеза об отсутствии ранговой корреляционной связи отвергается (связь статистически значима).

Так, в примере 3.5 при уровне значимости $\alpha = 0,06$ имеем: $n = 10$; $0,23 < \frac{\alpha}{2} < 0,36$; следовательно, $S_K = 22$ (оно лежит между 21 и 23), так что

$$\tau_{\max}^{(K)} = \frac{2 \cdot 22}{10 \cdot 9} = \frac{44}{90} = 0,489.$$

Поскольку $\tau^{(K)} = 0,733 > 0,489$, делается вывод о наличии статистически значимой корреляционной связи между исследуемыми переменными в данном примере.

Построение доверительных интервалов для неизвестных истинных значений ранговых коэффициентов корреляции возможно лишь приближенно и только при измерении ранговой корреляции с помощью коэффициента Кендалла. При этом используют (при $n > 10$ и значениях $\tau^{(K)}$, не слишком близких по абсолютной величине к единице) приближенный факт нормальности распределения величины $\hat{\tau}^{(K)}$ со средним значением $E\hat{\tau}^{(K)} \approx \tau^{(K)}$ и с дисперсией $D\hat{\tau}^{(K)}$, не превышающей величины $\frac{2}{n}[1 - (\tau^{(K)})^2]$. Можно утверждать, что с доверительной вероятностью, не меньшей заданного уровня P , истинное значение коэффициента Кендалла $\tau^{(K)}$ заключено в пределах

$$\begin{aligned} \hat{\tau}^{(K)} - u_{\frac{1+P}{2}} \sqrt{\frac{2}{n}[1 - (\hat{\tau}^{(K)})^2]} < \tau^{(K)} < \hat{\tau}^{(K)} + \\ + u_{\frac{1+P}{2}} \sqrt{\frac{2}{n}[1 - (\hat{\tau}^{(K)})^2]}, \end{aligned} \quad (3.47)$$

где u_q — q -квантиль стандартного нормального распределения.

3.3.7 Коэффициент конкордации как измеритель статистической связи между несколькими ранговыми переменными

До сих пор мы рассматривали корреляцию между двумя порядковыми переменными. Однако при решении основных задач $A - C$ статистического анализа ранговых связей (см. п. 3.3.3) возникает необходимость уметь измерить статистическую связь между *несколькими* (более чем двумя) переменными. С этой целью Кендаллом был предложен показатель $\widehat{W}(m)$, названный *коэффициентом конкордации* (или *согласованности*), вычисляемый по формуле³

$$\widehat{W}(m) = \frac{12}{m^2(n^3 - n)} \sum_{i=1}^n \left(\sum_{j=1}^m x_i^{(k_j)} - \frac{m(n+1)}{2} \right)^2, \quad (3.48)$$

где m — число анализируемых порядковых переменных (сравниваемых упорядочений); n — число статистически обследованных объектов или длина ранжировки (объем выборки); $k_1, k_2, \dots, k_m$ — номера отобранных для анализа порядковых переменных (из исходной совокупности $x^{(0)}, x^{(1)}, x^{(2)}, \dots, x^{(p)}$, так что, очевидно, $m \leq p + 1$).

³Мы приводим здесь формулу для подсчета *выборочного* значения $\widehat{W}$ коэффициента конкордации W . Интерпретация и вычисление *теоретического* значения W непосредственно следуют из рассуждений, приведенных в п. 3.3.6 в связи с анализом статистических свойств выборочных парных ранговых коэффициентов корреляции.

Нетрудно устанавливаются следующие свойства коэффициента конкордации:

а) $0 \leq \widehat{W} \leq 1$;

б) $\widehat{W} = 1$ тогда и только тогда, когда все m анализируемых упорядочений совпадают;

в) если $m \geq 3$ и анализируемые ранжировки генерируются подобно случайному независимому m -кратному извлечению из множества всех n возможных упорядочений n объектов, то связи между ними нет и $W = 0$;

г) пусть $\bar{\tau}^{(S)}(m)$ — среднее значение коэффициента Спирмэна, подсчитанное по значениям $m(m-1)/2$ коэффициентов $\hat{\tau}_{k_i k_j}^{(S)}$ ($i, j = 1, 2, \dots, m; i \neq j$), характеризующих ранговую связь между всеми возможными парами переменных $(x^{(k_i)}, x^{(k_j)})$ из анализируемого набора $(x^{(k_1)}, x^{(k_2)}, \dots, x^{(k_m)})$; тогда

$$\bar{\tau}^{(S)}(m) = \frac{m\widehat{W}(m) - 1}{m - 1}; \quad (3.49)$$

в частности, из (3.49) следует для случая $m = 2$, что

$$\widehat{W}(2) = \frac{1}{2}(\hat{\tau}_{k_1 k_2}^{(S)} + 1), \quad (3.49')$$

то есть коэффициент конкордации, исчисленный для *двух* переменных, пропорционален введенному ранее парному ранговому коэффициенту корреляции Спирмэна.

То, что шкала измерения $W(m)$ не включает в себя отрицательных значений, объясняется следующим обстоятельством. В отличие от случая парных связей при анализе $m(m \geq 3)$ порядковых переменных противоположные понятия согласованности и несогласованности утрачивают прежнюю симметричность (относительно нуля); упорядочения, произведенные в соответствии с переменными $x^{(k_1)}, x^{(k_2)}, \dots, x^{(k_m)}$, могут полностью совпадать, но *не могут полностью не совпадать* в том смысле, который мы вкладывали в это понятие при $m = 2$.

Формула (3.48) получена (и справедлива) в предположении отсутствия объединенных рангов в каждом из анализируемых упорядочений. Если же таковые имеются, то формула должна быть модифицирована:

$$\widehat{W}(m) = \frac{\sum_{i=1}^n \left(\sum_{j=1}^m x_i^{(k_j)} - \frac{m(n+1)}{2} \right)^2}{\frac{1}{12}m^2(n^3 - n) - m \sum_{j=1}^m T^{(k_j)}}, \quad (3.48')$$

где поправочный коэффициент $T^{(k_j)}$ (соответствующий переменной $x^{(k_j)}$) подсчитывается по формуле (3.38).

3.3.8 Проверка статистически значимого отличия от нуля выборочного значения коэффициента конкордации

Как ведут себя выборочные значения $\widehat{W}(m)$ коэффициента конкордации при повторении выборок заданного объема n (из одной и той же генеральной совокупности) при отсутствии какой-либо связи между анализируемыми m переменными? Другими словами, нас интересует ответ на следующий вопрос. Предположим, что каждому объекту конечной генеральной совокупности (состоящей из N элементов) приписан какой-то определенный ранг по каждой из m рассматриваемых переменных. Так, например, если $m = 3$ и объекту O_i определены значения: $x_i^{(1)} = N$; $x_i^{(2)} = 1$; $x_i^{(3)} = 2$, то это означает, что по переменной $x^{(1)}$ он стоит на последнем (N -м) месте в упорядоченном ряду *всех* объектов генеральной совокупности, по переменной $x^{(2)}$ — на первом и по переменной $x^{(3)}$ — на втором. Тогда по исходным данным $\{(x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(m)})\}_{i=1, \overline{N}}$ с помощью формулы (3.48) может быть вычислен теоретический (генеральный) коэффициент конкордации $W(m)$, характеризующий степень тесноты ранговой связи между переменными $x^{(1)}, x^{(2)}, \dots, x^{(m)}$. Однако исследователю известны значения $(x^{(1)}, x^{(2)}, \dots, x^{(m)})$ *лишь для части объектов* генеральной совокупности, а именно для случайной выборки объектов объема n ($n < N$). После естественной перенумерации рангов, сохраняющей правило упорядочения объектов, но переводящей масштаб измерения рангов в шкалу $(1, 2, \dots, n)$ (для этого минимальный из оказавшихся в выборке рангов по каждой переменной объявляется рангом, равным 1, следующий по величине — рангом, равным 2, и т. д.), может быть вычислен (по той же формуле (3.48)) *выборочный* коэффициент конкордации $\widehat{W}(m)$. Извлекая другую выборку объема n из той же самой генеральной совокупности, мы получим, вообще говоря, другое значение выборочного коэффициента $\widehat{W}(m)$ и т. д.

Спрашивается, как сильно могут отклоняться от нуля выборочные значения коэффициента конкордации $\widehat{W}(m)$ в ситуации, когда значение теоретического коэффициента конкордации $W(m)$ свидетельствует о полном отсутствии ранговой связи между анализируемыми переменными $x^{(1)}, x^{(2)}, \dots, x^{(m)}$? Для малых значений m и n ($2 \leq m \leq 20$, $3 \leq n \leq 7$) ответ на этот вопрос может быть получен с помощью таблицы П1.11 (см. Приложение 1) значений величины S . Обозначенная в ней величина S есть не что иное, как

$$S = \sum_{i=1}^n \left(\sum_{j=1}^m x_i^{(k_j)} - \frac{m(n+1)}{2} \right)^2. \quad (3.50)$$

«Входами» в эту таблицу является тройка чисел (m, n, S) , «выходом» — вероятность того, что величина S может быть такой, какой она

является в нашей выборке, или большей в условиях отсутствия связи переменных в генеральной совокупности. Если окажется, что эта вероятность меньше принятой нами величины уровня значимости критерия α (например, $\alpha = 0,05$), то гипотезу об отсутствии связи следует отвергнуть, то есть признать статистическую значимость анализируемой связи. Таблица критических значений $W(m)$ построена несколько иначе. В ней при уровне значимости $\alpha = 0,05$ и в соответствии с «входами» (m, n) даны «критические» значения величины S , то есть такие значения, при превышении которых следует отвергать гипотезу об отсутствии связей (признавать их статистическую значимость).

При $n > 7$ для проверки статистической значимости анализируемой связи следует воспользоваться фактом приближенной $\chi^2(n-1)$ -распределенности величины $m(n-1) \times \widehat{W}(m)$, справедливым в условиях отсутствия связи в генеральной совокупности (значение $\widehat{W}(m)$, как и прежде, подсчитывается по формуле (3.48) или (3.48')). Поэтому, если окажется, что

$$m(n-1)\widehat{W}(m) > \chi^2_\alpha(n-1), \quad (3.51)$$

то гипотеза об отсутствии ранговой связи между переменными $x^{(k_1)}, x^{(k_2)}, \dots, x^{(k_m)}$ должна быть отвергнута (с уровнем значимости критерия, равным α); в (3.51) величина $\chi^2_\alpha(n-1)$ — это 100 α %-ная точка χ^2 -распределения с $(n-1)$ -й степенью свободы.

Строгих рекомендаций по построению доверительных интервалов для истинного значения W в условиях *наличия* ранговых связей в исследуемой генеральной совокупности, по-видимому, не существует.

Рассмотрим примеры, в которых реализуются приведенные выше рекомендации по статистическому анализу множественных ранговых связей.

Пример 3.7. Рассмотрим три порядковые переменные $(x^{(1)}, x^{(2)}, x^{(3)})$ и соответствующие им упорядочения десяти объектов:

$X^{(1)\top}$	1	4,5	2	4,5	3	7,5	6	9	7,5	10
$X^{(2)\top}$	2,5	1	2,5	4,5	4,5	8	9	6,5	10	6,5
$X^{(3)\top}$	2	1	4,5	4,5	4,5	4,5	8	8	8	10
Сумма	5,5	6,5	9	13,5	12	20	23	23,5	23,5	26,5

В соответствии с формулами (3.50), (3.38) имеем:

$$S = \sum_{i=1}^{10} \left(\sum_{j=1}^3 x_i^{(j)} - \frac{3 \cdot 11}{2} \right)^2 = (-11)^2 + (-10)^2 + (-7,5)^2 +$$

$$+ (-3)^2 + (-4,5)^2 + (3,5)^2 + (6,5)^2 + 7^2 + 9^2 + 10^2 = 591;$$

$$T^{(1)} = \frac{1}{12}(2^3 - 2)2 = 1;$$

$$T^{(2)} = \frac{1}{12}(2^3 - 2)2 = 1,5;$$

$$T^{(3)} = \frac{1}{12}(4^3 - 4 + 3^3 - 3) = 7.$$

Следовательно, в соответствии с (3.48')

$$\widehat{W}(3) = \frac{591}{\frac{1}{12}3^2(10^2 - 10) - 3(1 + 1,5 + 7)} = \frac{591}{742,5 - 28,5} = 0,828.$$

Пример 3.8. Требуется проверить статистическую значимость множественной ранговой связи 28 переменных ($m = 28$), характеризуемой величиной выборочного коэффициента конкордации $\widehat{W}(28) = 0,08$, подсчитанного по 13 объектам ($n = 13$).

Воспользуемся фактом $\chi^2(12)$ -распределенности случайной величины $m(n-1)\widehat{W}(m)$, который имеет место (приблизительно) в случае, если в исследуемой генеральной совокупности множественная ранговая связь отсутствует. Тогда критерий сводится к проверке неравенства (3.51). Задавшись уровнем значимости критерия $\alpha = 0,05$, находим из таблиц значение 5%-ной точки χ^2 -распределения с 12 степенями свободы $\chi_{0,05}^2(12) = 21,026$. В то же время $m(n-1)\widehat{W}(m) = 28 \cdot 12 \cdot 0,08 = 27$.

Поскольку $m(n-1)\widehat{W}(m) > \chi_{0,05}^2(12)$, то оказалось, что *даже такого маленького числа, как 0,08, «хватило» для того, чтобы объявить связь между 28 исследуемыми переменными статистически значимой.*

3.4 Корреляционный анализ категоризованных переменных: таблицы сопряженности

Признак называют *категоризованным*, если его возможные «значения» описываются конечным числом состояний или градаций. Слово «значения» взято в кавычки, так как речь идет, как правило, не о числовых значениях, а лишь об определенных условных метках возможных состояний. Так, категоризованный признак «пол индивидуума» имеет две градации, категоризованный признак «социальное положение респондента» — столько градаций, сколько установлено социальных страт (слов) в данном обществе, категоризованный признак «уровень жилищных

условий семьи» — столько градаций, сколько их определено в данном исследовании (например, при возможных ответах: «низкий», «удовлетворительный», «хороший» и «очень хороший» мы будем иметь четыре градации). Как мы видим, в класс категоризованных признаков попадают те номинальные⁴ и порядковые (ординальные) переменные, возможные значения которых описаны заданным (*известным*) набором градаций.

3.4.1 Исходные статистические данные (таблицы сопряженности)

Мы ограничимся здесь задачей измерения *парных* статистических связей между категоризованными переменными. Формально исходные данные для пары категоризованных переменных будут иметь уже знакомый нам вид таблицы «объект — свойство», в которой, правда, в качестве элементов $x_i^{(j)}$ ($j = 1, 2; i = 1, 2, \dots, n$) будут обозначены *условные метки (градации)* состояния объекта i по переменной j . Однако мы не можем работать с этими данными аналогично случаю количественных переменных: скажем, среднее значение или дисперсия «пола» или «социальной принадлежности» респондента лишены всякого смысла. Поэтому при статистическом анализе двух категоризованных переменных исходные данные преобразуют к виду таблицы перекрестных частот, называемой **двухвходовой таблицей сопряженности** признаков $x^{(1)}$ и $x^{(2)}$ (см. таблицу 3.2) или просто таблицей сопряженности.

Таблица 3.2. Двухвходовая таблица сопряженности

Градация признака $x^{(1)}$	Градация признака $x^{(2)}$						$n_{i\cdot}$
	1	2	...	j	...	m_2	
1	n_{11}	n_{12}	...	n_{1j}	...	n_{1m_2}	$n_{1\cdot}$
2	n_{21}	n_{22}	...	n_{2j}	...	n_{2m_2}	$n_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	...	$\vdots$	$\vdots$
i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{im_2}	$n_{i\cdot}$
$\vdots$	$\vdots$	$\vdots$	...	$\vdots$	...	$\vdots$	$\vdots$
m_1	n_{m_11}	n_{m_12}	...	n_{m_1j}	...	$n_{m_1m_2}$	$n_{m_1\cdot}$
$n_{\cdot j}$	$n_{\cdot 1}$	$n_{\cdot 2}$	...	$n_{\cdot j}$	...	$n_{\cdot m_2}$	n

В таблице 3.2 представлены результаты статистического обследования n объектов по признакам $x^{(1)}$ и $x^{(2)}$. В ней n_{ij} означает число объектов (из общего числа n обследованных), у которых «значение» признака

⁴Напомним, что случайная величина называется номинальной, если знание ее значений на статистически обследованных объектах позволяет разбить это множество на не поддающиеся упорядочению однородные по анализируемому свойству классы.

$x^{(1)}$ оказалось зафиксированным на уровне i -й градации, а «значение» признака $x^{(2)}$ — на уровне j -й градации; $n_{i.} = \sum_{j=1}^{m_2} n_{ij}$ — число объектов, значение признака $x^{(1)}$ у которых оказалось зарегистрированным на уровне i -й градации (при отсутствии каких бы то ни было условий на «значения» признака $x^{(2)}$), а $n_{.j} = \sum_{i=1}^{m_1} n_{ij}$ — число объектов, значение признака $x^{(2)}$ у которых оказалось зарегистрированным на уровне j -й градации ($i = 1, 2, \dots, m_1$; $j = 1, 2, \dots, m_2$).

Переход к задаче измерения степени тесноты связи *нескольких* (более двух) категоризованных переменных связан с необходимостью введения *многоходовых таблиц сопряженности* и более громоздких обозначений (количество нижних индексов у анализируемых частот приходится увеличивать до числа анализируемых переменных).

3.4.2 Основные измерители степени тесноты статистической связи между двумя категоризованными переменными

В статистической теории и практике в рамках данной проблемы существует *целый спектр характеристик с.т.с.с.* Разные меры связи акцентируют внимание на разных аспектах взаимоотношений между переменными, давая в целом многоаспектную информацию о природе изучаемой зависимости. Поэтому вряд ли целесообразно давать универсальные рекомендации по поводу выбора наиболее предпочтительной характеристики «на все случаи жизни». Выбор описанных ниже измерителей с.т.с.с. был основан на сочетании двух критериев: 1) *наибольшей распространенности* в практике статистических исследований; 2) *умении описать необходимые статистические свойства* этих измерителей.

Построение подавляющего большинства используемых в данной проблеме характеристик с.т.с.с. основано на одной общей идее: *характеристика должна принимать тем большее числовое значение, чем больше анализируемая ситуация отклоняется от гипотезы взаимной статистической независимости исследуемых переменных $x^{(1)}, x^{(2)}$* . Как мы знаем, формализуется понятие статистической независимости условием представления двумерного закона распределения векторной случайной величины $X = (x^{(1)}, x^{(2)})^T$ в виде произведения частных законов распределения компонент этого вектора, то есть

$$P\{x^{(1)} = x_i^{(1)0}; x^{(2)} = x_j^{(2)0}\} = P\{x^{(1)} = x_i^{(1)0}\} P\{x^{(2)} = x_j^{(2)0}\}, \quad (3.52)$$

где $x_i^{(1)0}$ и $x_j^{(2)0}$ — обозначения принятых в данном исследовании меток соответственно для i -й градации признака $x^{(1)}$ и j -й градации признака $x^{(2)}$.

Выборочный (эмпирический) эквивалент этого условия предусматривает замену участвующих в нем вероятностей соответствующими относительными частотами (с заменой, конечно, *точного* знака равенства на *приближенный*), то есть

$$\frac{n_{ij}}{n} \approx \frac{n_{i.}}{n} \cdot \frac{n_{.j}}{n} \quad (3.52')$$

или

$$D_{ij} = \frac{n_{ij}}{n} - \frac{n_{i.}}{n} \cdot \frac{n_{.j}}{n} \approx 0. \quad (3.52'')$$

Поэтому чем больше от нуля будет отличаться просуммированная по i и j левая часть соотношения (3.52''), вычисленная по данным таблицы сопряженности 3.2, тем более высокую с.т.с.с. между анализируемыми переменными должны обозначать используемые измерители.

Характеристика X^2 квадратичной сопряженности признаков $x^{(1)}$ и $x^{(2)}$ определяется соотношением

$$X^2 = n \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \frac{D_{ij}^2}{n_{i.} n_{.j}} = n \left(\sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \frac{n_{ij}^2}{n_{i.} n_{.j}} - 1 \right). \quad (3.53)$$

Его значение может меняться от нуля (при строгой статистической независимости переменных $x^{(1)}$ и $x^{(2)}$) до $+\infty$. Являясь, как и всякая функция от выборочных данных, величиной случайной, X^2 ведет себя (в предположении статистической независимости $x^{(1)}$ и $x^{(2)}$) приблизительно, — асимптотически по $n \rightarrow \infty$, — как χ^2 -распределенная случайная величина с числом степеней свободы, равным $(m_1 - 1)(m_2 - 1)$. Это дает нам право в случае

$$X^2 \geq \chi_\alpha^2((m_1 - 1)(m_2 - 1)) \quad (3.54)$$

заявить о том, что связь между анализируемыми переменными $x^{(1)}$ и $x^{(2)}$ является *статистически значимой* и она тем теснее, чем больше значение X^2 . Неудобство использования характеристики X^2 связано с тем, что ее верхняя граница стремится к бесконечности при возрастании объема выборки n . Поэтому вместо X^2 часто используют характеристику

$$C = \left[\frac{X^2}{n \min(m_1 - 1, m_2 - 1)} \right]^{1/2}, \quad (3.55)$$

которая называется **коэффициентом Крамера** и которая обладает более привычным для измерителей с.т.с.с. диапазоном изменения, а именно $0 \leq C \leq 1$. При этом нулевое значение C свидетельствует о строгой статистической независимости анализируемых признаков, а значение C , равное единице, — о возможности однозначного восстановления значения одной переменной по известному значению другой.

Информационная характеристика с.т.с.с. Y^2 признаков $x^{(1)}$ и $x^{(2)}$ также основана на мере отклонения от выполнения соотношения

независимости (3.52'), только вместо разности левых и правых частей этого соотношения в ней используется их отношение. А именно измеритель Y^2 определяется соотношением

$$Y^2 = 2 \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} n_{ij} \ln \left(\frac{n_{ij}}{n_{i.} n_{.j} / n} \right). \quad (3.56)$$

Правую часть (3.56) можно преобразовать к виду, более удобному для вычислений:

$$Y^2 = 2 \left(\sum_{i=1}^{m_1} \sum_{j=1}^{m_2} n_{ij} \ln n_{ij} - \sum_{i=1}^{m_1} n_{i.} \ln n_{i.} - \sum_{j=1}^{m_2} n_{.j} \ln n_{.j} + n \ln n \right). \quad (3.56')$$

Эта характеристика обладает теми же свойствами, что и X^2 : ее значения варьируют от 0 (в случае статистической независимости $x^{(1)}$ и $x^{(2)}$) до $+\infty$, и в условиях справедливости гипотезы о статистической независимости анализируемых признаков она приблизительно (асимптотически по $n \rightarrow \infty$) подчиняется з.р.в. χ^2 с $(m_1-1)(m_2-1)$ степенями свободы (для достижения удовлетворительной точности последнего утверждения требуется заполненность всех клеток таблицы сопряженности 3.2 ненулевыми элементами, удовлетворяющими условию $\min_{i,j} n_{ij} > 3$).

Таким образом, если мы хотим проверить гипотезу о статистически значимом отличии от нуля характеристики с.т.с.с. Y^2 (то есть гипотезу о наличии связи между $x^{(1)}$ и $x^{(2)}$), то необходимо убедиться в выполнении неравенства

$$Y^2 > \chi_\alpha^2((m_1-1)(m_2-1)), \quad (3.57)$$

где $\chi_\alpha^2(m)$, как всегда, — 100 α %-ная точка χ^2 -распределения с m степенями свободы (см. таблицу П1.4 в Приложении 1), а α — заданный уровень значимости критерия, то есть вероятность принять решение о наличии статистической связи между анализируемыми переменными в то время, как в действительности они являются статистически независимыми.

Если с помощью той или иной характеристики (X^2 или Y^2) мы убедились в том, что статистически значимая связь между $x^{(1)}$ и $x^{(2)}$ действительно существует, то возникает вопрос *интервальной оценки* характеристики этой связи. Для приближенного построения такой оценки мы должны уметь вычислить дисперсию соответствующей точечной оценки, то есть в нашем случае — дисперсии статистик X^2 и Y^2 в условиях, когда оцениваемый с их помощью теоретический показатель связи отличен от нуля.

Для характеристики X^2 и основанном на ней коэффициенте Крамера C существуют приближенные формулы⁵:

$$\sigma_{X^2}^2 = \mathbf{D}X^2 \approx 4 \left[X^2 + \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \left(\frac{\frac{n_{ij}}{n} - \frac{n_{i.}}{n} \frac{n_{.j}}{n}}{\frac{n_{i.}}{n} \frac{n_{.j}}{n}} \right)^2 - \frac{(X^2)^2}{n} \right], \quad (3.58)$$

$$\sigma_C^2 = \mathbf{D}C \approx \frac{1}{n \min(m_1 - 1, m_2 - 1)}. \quad (3.59)$$

Заметим, что в формуле (3.58) первое слагаемое в квадратных скобках (то есть X^2) является *главным членом*, так что два других слагаемых играют роль *остаточного члена*. Поэтому другой, более грубый (менее точный) вариант этой формулы имеет вид

$$\sigma_{X^2}^2 = \mathbf{D}X^2 \approx 4X^2. \quad (3.58')$$

Соответственно в качестве приближенных доверительных интервалов для истинных значений коэффициента квадратичной сопряженности и коэффициента Крамера используются (при доверительной вероятности $P = 1 - 2\alpha$):

$$[X^2 - u_{1-\alpha}\sigma_{X^2}; \quad X^2 + u_{1-\alpha}\sigma_{X^2}] \quad (3.60)$$

и

$$[C - u_{1-\alpha}\sigma_C; \quad C + u_{1-\alpha}\sigma_C], \quad (3.61)$$

где u_q , как обычно, q -квантиль стандартного нормального распределения (см. таблицу П1.3 в приложении 1), а вместо σ_{X^2} и σ_C подставляются их приближенные значения, вычисленные соответственно по формулам (3.58) (или (3.58')) и (3.59).

Пример 3.9. В таблице 3.3 (из [G i l b y W. H., Biometrika, 8, 94]) дано распределение 1725 школьников (то есть $n = 1725$) по значениям двух анализируемых категоризованных признаков: по качеству (опрятности) одежды ($x^{(1)}$) и по их умственным способностям ($x^{(2)}$). Были определены 4 градации (то есть $m_1 = 4$) по первому признаку (1: «*очень хорошо*»; 2: «*хорошо*»; 3: «*удовлетворительно*»; 4: «*плохо*») и 6 градаций (то есть $m_2 = 6$) по второму признаку (1: «*плохие*»; 2: «*ниже средних*»; 3: «*средние*»; 4: «*выше средних*»; 5: «*высокие*»; 6: «*превосходные*»).

⁵Точные формулы (они получены Холдуйном в 1939 г.) очень сложны, поэтому здесь приводится только приближенный вариант, полученный К. Пирсоном в 1915 г.

Таблица 3.3. Результаты обследования школьников Англии по опрятности одежды и умственным способностям

Градации признака $x^{(1)}$	Градации признака $x^{(2)}$						Σ
	1	2	3	4	5	6	
1	33	48	113	209	194	39	636
2	41	100	202	255	138	15	751
3	39	58	70	61	33	4	265
4	17	13	22	10	10	1	73
Σ	130	219	407	535	375	59	1725

Нас интересует, существует ли связь между манерой одеваться и способностями, и если «да», то какова степень тесноты этой связи.

Вычисление величины X^2 по формуле (3.53) дает $X^2 = 174,92$ и соответственно значение коэффициента Крамера (см. формулу (3.55)) $C = 0,184$.

Поскольку даже при уровне значимости $\alpha = 0,001$ соответствующее пороговое значение $\chi^2_{0,001}(15) = 37,697$ оказывается намного меньшим статистики X^2 , то связь есть и, по-видимому, характеризуется *достаточно высокой степенью тесноты*. Если сопоставить этот вывод с выборочным значением коэффициента Крамера ($C = 0,184$), то мы получаем основание относиться к этой шкале иначе, чем, например, к шкале абсолютных значений парного коэффициента корреляции $|\hat{r}|$: если для $|\hat{r}|$ величина 0,184 означает отсутствие или крайне слабую связь между $x^{(1)}$ и $x^{(2)}$, то для C это значение, как мы видим, свидетельствует о наличии достаточно тесной связи.

Величина σ_C , вычисленная по приближенной формуле (3.59), оказывается равной $\sqrt{1/3n} = 0,014$, так что в соответствии с (3.61) можем сделать вывод, что истинное (теоретическое) значение коэффициента Крамера заключено в пределах от $0,184 - 1,96 \cdot 0,014 = 0,157$ до $0,184 + 1,96 \cdot 0,014 = 0,211$.

Выводы

1. Трудности, стоящие на пути практического использования в статистическом исследовании моделей *многомерных* з.р.в., обусловили тот факт, что в подавляющем большинстве случаев все выводы многомерного статистического анализа строятся лишь на базе оценок вектора средних значений и ковариационной матрицы исследуемого векторного признака.

2. Корреляционный анализ составляет содержание начальных этапов исследования всех трех центральных проблем многомерного статисти-

ческого анализа. В *проблеме статистического исследования зависимостей* корреляционный анализ позволяет выявлять сам факт существования статистических связей между компонентами анализируемого векторного признака и оценивать степень тесноты этих связей. Для *проблем классификации объектов и признаков и снижения размерности анализируемого признакового пространства* он предлагает и оценивает подходящие характеристики парных отношений (ковариации, корреляции, разные виды парных сравнений), которые используются в дальнейшем в качестве базовой исходной информации.

3. Универсальной характеристикой степени тесноты статистической связи (с.т.с.с.) между результирующим *количественным* показателем y и объясняющими *количественными* переменными $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$ является коэффициент $K_d(y; X)$ детерминации y по X . Другие распространенные характеристики с.т.с.с. (парные, частные и множественные коэффициенты корреляции, корреляционное отношение) представляют собой те или иные частные версии коэффициента детерминации, реализованные в рамках различных конкретных схем зависимостей.

4. *Парные* корреляционные характеристики позволяют измерять степень тесноты статистической связи между парой переменных *без учета опосредованного или совместного влияния других показателей*, вычисляются (оцениваются) они по результатам наблюдений только анализируемой пары показателей.

5. Факт установления тесной статистической связи между переменными не является, вообще говоря, достаточным основанием для доказательства существования *причинно-следственной связи между этими переменными*.

6. Парные и частные *коэффициенты корреляции* являются измерителями степени тесноты *линейной* связи между переменными. В этом случае корреляционные характеристики могут оказаться как положительными, так и отрицательными в зависимости от одинаковой или противоположной тенденции взаимосвязанного изменения анализируемых переменных. При положительных значениях коэффициента корреляции говорят о наличии *положительной линейной статистической связи*, при отрицательных — об *отрицательной*.

7. При наложении случайных ошибок на значения исследуемой пары переменных (например, ошибок измерения) оценка статистической связи между исходными переменными, построенная по наблюдениям, оказывается искаженной. В частности, получаемые при этом оценки коэффициентов корреляции будут *заниженными*. Существуют методы, позволяющие учесть это искажение.

8. Измерителем степени тесноты связи *любой формы* является *корреляционное отношение*, для вычисления которого необходимо разбить область значений предсказывающей переменной X на интервалы (гиперпараллелепипеды) группирования. Возможна параметрическая мо-

дификация корреляционного отношения, при которой вычисление соответствующих выборочных значений не требует предварительного разбиения на интервалы группирования.

9. *Частный коэффициент корреляции* позволяет оценить степень тесноты линейной связи между двумя переменными, *очищенной от опосредованного влияния* других факторов. Для его расчета необходима исходная информация как по анализируемой паре переменных, так и по всем тем переменным, опосредованное («мешающее») влияние которых мы хотим элиминировать.

10. Анализ статистических связей между ранговыми порядковыми переменными сводится к статистическому анализу различных упорядочений (ранжировок) одного и того же конечного множества объектов и осуществляется с помощью методов *ранговой корреляции*. В зависимости от типа изучаемой ситуации (шкала измерения анализируемого свойства не известна исследователю или отсутствует вовсе; существуют косвенные или частные количественные показатели, в соответствии со значениями которых можно определить место каждого объекта в общем ряду всех объектов, упорядоченных по анализируемому основному свойству) процесс упорядочения объектов производится либо с привлечением экспертов, либо формализованно — с помощью перехода от исходного ряда наблюдений косвенного *количественного* признака к соответствующему вариационному ряду.

11. Исходные статистические данные для проведения рангового корреляционного анализа представлены *таблицей (матрицей) рангов* статистически обследованных объектов размера $n \times (p + 1)$ (число объектов на число анализируемых переменных). При формировании матрицы рангов допускаются случаи неразличимости двух или нескольких объектов по изучаемому свойству («объединенные» ранги).

12. К *основным задачам* теории и практики ранговой корреляции относятся: анализ структуры исследуемой совокупности упорядочений (задача А); анализ интегральной (совокупной) согласованности рассматриваемых переменных и их условная ранжировка по критерию степени тесноты связи каждой из них со всеми остальными переменными (задача В); построение единого группового упорядочения объектов на основе имеющейся совокупности согласованных упорядочений (задача С).

13. В качестве основных характеристик парной статистической связи между упорядочениями используются *ранговые коэффициенты корреляции Спирмена $\tau^{(S)}$ и Кендалла $\tau^{(K)}$* . Значения этих коэффициентов меняются в диапазоне от -1 до $+1$, причем экстремальные значения характеризуют связь соответственно пары прямо противоположных и пары совпадающих упорядочений, а нулевое значение рангового коэффициента корреляции получается при полном отсутствии статистической связи между анализируемыми порядковыми переменными.

14. В качестве основной характеристики статистической связи между *несколькими* (m) порядковыми переменными используется так называемый *коэффициент конкордации (согласованности) Кендалла* $W(m)$. Между значением этого коэффициента и значениями парных ранговых коэффициентов Спирмэна, построенных для каждой пары анализируемых переменных, существуют простые соотношения.

15. Если представить себе, что каждому объекту некоторой достаточно большой гипотетической совокупности (будем называть ее *генеральной совокупностью*) приписан какой-то ранг по каждой из рассматриваемых переменных и что статистическому обследованию подлежит *лишь часть* этих объектов (*выборка объема n*), то достоверность и практическая ценность выводов, основанных на анализе ранговой корреляции, существенно зависят от ответа на вопрос: как ведут себя выборочные значения интересующих нас ранговых корреляционных характеристик при повторениях выборок заданного объема, извлеченных из этой генеральной совокупности. Это и составляет *предмет исследования статистических свойств выборочных ранговых характеристик связи*. Результаты этого исследования относятся прежде всего к построению правил проверки статистической значимости анализируемой связи и к построению доверительных интервалов для неизвестных значений коэффициентов связи, характеризующих всю генеральную совокупность.

16. Основными измерителями степени тесноты *парной* статистической связи между *категоризованными* (ординальными и номинальными) переменными являются *коэффициент квадратичной сопряженности* X^2 и *информационная характеристика связи* Y^2 . Эти характеристики используют в ситуациях, когда анализируемыми переменными являются ординальные или номинальные признаки, шкала возможных «значений» которых определена заданным набором их состояний (*градаций*).

Глава 4

Классическая линейная модель множественной регрессии (КЛММР)

Регрессионный анализ занимает, бесспорно, центральное место во всем математико-статистическом инструментарии эконометрики. Общие постановки задач статистического исследования зависимостей и основные типы регрессионных зависимостей между количественными признаками обсуждались в вводной главе 2 (см. п. 2.1, 2.2, 2.4). Там же впервые дано определение *функции регрессии* (см. п. 2.1). И, наконец, в главе 3 достаточно подробно описан необходимый этап *предрегрессионного анализа* — так называемый *корреляционный анализ*, в процессе которого оценивается *степень тесноты статистической связи между анализируемыми переменными* (ведь именно от степени тесноты анализируемой связи зависит прогностическая сила конструируемой регрессионной модели).

Поэтому, приступая к этой главе, желательно освежить в памяти сведения, почерпнутые читателем из вводных глав 2 и 3.

4.1 Описание КЛММР. Основные допущения модели

Классическая линейная модель множественной регрессии (КЛММР) представляет собой простейшую версию конкретизации требований к общему виду функции регрессии $f(X)$, природе объясняющих переменных X и статистических регрессионных остатков $\varepsilon(X)$ в общих уравнениях

регрессионной связи (2.3) — (2.4)¹. В рамках КЛММП эти требования формулируются следующим образом:

$$\left\{ \begin{array}{l} y_i = \theta_0 + \theta_1 x_i^{(1)} + \dots + \theta_p x_i^{(p)} + \varepsilon_i, \quad i = 1, 2, \dots, n; \\ \mathbf{E}\varepsilon_i = 0, \quad i = 1, 2, \dots, n; \\ \mathbf{E}(\varepsilon_i \varepsilon_j) = \begin{cases} \sigma^2 & \text{при } i = j, \\ 0 & \text{при } i \neq j; \end{cases} \\ (x^{(1)}, x^{(2)}, \dots, x^{(p)}) - \text{нечисловые переменные;} \\ \text{ранг матрицы } \mathbf{X} = p + 1 < n, \end{array} \right. \quad (4.1)$$

где матрица $\mathbf{X}$ — это $n \times (p + 1)$ -матрица наблюдаемых значений объясняющих переменных, то есть

$$\mathbf{X} = \begin{pmatrix} 1 & x_1^{(1)} & x_1^{(2)} & \dots & x_1^{(p)} \\ 1 & x_2^{(1)} & x_2^{(2)} & \dots & x_2^{(p)} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_n^{(1)} & x_n^{(2)} & \dots & x_n^{(p)} \end{pmatrix} \quad (4.1a)$$

(присутствие первого столбца, состоящего из единиц, обусловлено чисто техническими причинами удобства представления модели в матричном виде, см. ниже соотношения (4.1')).

Из (4.1) следует, что в рамках КЛММП рассматриваются только *линейные* функции регрессии, то есть

$$f(X) = \mathbf{E}(y | X) = \theta_0 + \theta_1 x^{(1)} + \dots + \theta_p x^{(p)}, \quad (4.2)$$

где объясняющие переменные $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ играют роль *нечисловых параметров*, от которых зависит закон распределения вероятностей результирующей переменной y . Это, в частности, означает, что в повторяющихся выборочных наблюдениях $(x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)}; y_i)$ единственным источником случайных возмущений значений y_i являются случайные возмущения регрессионных остатков ε_i (подобную схему зависимости мы наблюдали в примере 2.1).

Кроме того, постулируется *взаимная некоррелированность* случайных регрессионных остатков ($\mathbf{E}(\varepsilon_i \varepsilon_j) = 0$ для $i \neq j$). Это требование к регрессионным остаткам $\varepsilon_1, \dots, \varepsilon_n$ относится к основным предположениям классической модели и оказывается вполне естественным в широком классе реальных ситуаций, особенно, если речь идет о *пространственных* выборках, то есть о ситуациях, когда значения анализируемых переменных регистрируются на различных объектах (индивидуумах, семьях,

¹ Процесс конкретизации подобных требований к структуре и характеру анализируемых моделей регрессионного типа обычно называют *спецификацией модели* (подробнее о спецификации модели см. выше, в п.1.4.3).

предприятиях, банках, регионах и т. п.). В этом случае данное предположение означает, что «возмущения» (регрессионные остатки), получающиеся при наблюдении одного какого-либо обследуемого объекта, не влияют на «возмущения», характеризующие наблюдения над другими объектами, и наоборот.

Тот факт, что для всех остатков $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ выполняется соотношение $E\varepsilon_i^2 = \sigma^2$, где величина σ^2 от номера наблюдения i не зависит, означает неизменность (постоянство, независимость от того, при каких значениях объясняющих переменных производятся наблюдения) дисперсий регрессионных остатков. Последнее свойство принято называть *гомоскедастичностью* регрессионных остатков.

Наконец, требуется, чтобы ранг матрицы X , составленной из наблюдаемых значений объясняющих переменных, был бы максимальным, то есть равнялся бы числу столбцов этой матрицы, которое в свою очередь должно быть меньше числа ее строк (то есть общего числа имеющихся наблюдений). Случаи $p + 1 \geq n$ не рассматриваются, поскольку при этом число n имеющихся в нашем распоряжении исходных статистических данных оказывается меньшим или равным числу оцениваемых параметров модели ($p + 1$), что исключает принципиальную возможность получения сколько-нибудь надежных статистических выводов. Что касается требования к рангу матрицы X , то оно означает, что не должно существовать строгой линейной зависимости между объясняющими переменными. Так, если, например, одна объясняющая переменная может быть линейно выражена через какое-то количество других, то ранг матрицы X окажется меньше $p + 1$, а следовательно, и ранг матрицы $X^T X$ будет тоже меньше $p + 1$ (см. Приложение 2). А это означает вырождение симметрической матрицы $X^T X$ (то есть $\det(X^T X) = 0$), что исключает существование матрицы $(X^T X)^{-1}$, которая, как мы увидим, играет важную роль в процедуре оценивания параметров анализируемой модели.

В дальнейшем нам удобнее будет оперировать с *матричной записью* модели (4.1). Для этого введем матрицы:

$$Y = (y_1, y_2, \dots, y_n)^T \quad - \quad (4.3)$$

вектор-столбец, наблюдаемых значений зависимой переменной;

$$I_n = \begin{pmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{pmatrix} \quad - \quad (4.4)$$

единичная матрица размерности $n \times n$;

$$\Theta = (\theta_0, \theta_1, \dots, \theta_p)^T \quad - \quad (4.5)$$

вектор-столбец неизвестных значений параметров;

$$\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^\top \quad - \quad (4.6)$$

вектор-столбец регрессионных остатков;

$$\mathbf{0}_n = (0, 0, \dots, 0)^\top \quad - \quad (4.7)$$

вектор-столбец высоты n , состоящий из одних нулей;

$$\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} = \mathbf{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top) = \begin{pmatrix} \mathbf{E}(\varepsilon_1^2) & \mathbf{E}(\varepsilon_1\varepsilon_2) & \dots & \mathbf{E}(\varepsilon_1\varepsilon_n) \\ \mathbf{E}(\varepsilon_2\varepsilon_1) & \mathbf{E}(\varepsilon_2^2) & \dots & \mathbf{E}(\varepsilon_2\varepsilon_n) \\ \dots & \dots & \dots & \dots \\ \mathbf{E}(\varepsilon_n\varepsilon_1) & \mathbf{E}(\varepsilon_n\varepsilon_2) & \dots & \mathbf{E}(\varepsilon_n^2) \end{pmatrix} \quad - \quad (4.8)$$

ковариационная матрица размерности $n \times n$ вектора остатков;

$$\hat{\boldsymbol{\Theta}} = (\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_p)^\top \quad - \quad (4.9)$$

вектор-столбец оценок неизвестных значений параметров;

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\Theta}}} = \mathbf{E}[(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta})(\hat{\boldsymbol{\Theta}} - \boldsymbol{\Theta})^\top] = (\sigma_{lj}(\hat{\boldsymbol{\Theta}})), \quad l, j = 0, 1, 2, \dots, p, \quad - \quad (4.10)$$

ковариационная матрица размерности $(p+1) \times (p+1)$ вектора несмещенных оценок $\hat{\boldsymbol{\Theta}}$ неизвестных параметров $\boldsymbol{\Theta}$ (в соотношении (4.10) $\sigma_{lj}(\boldsymbol{\Theta}) = \mathbf{E}[(\hat{\theta}_l - \theta_l)(\hat{\theta}_j - \theta_j)]$).

Тогда матричная форма записи КЛММР имеет вид:

$$\begin{cases} Y = \mathbf{X}\boldsymbol{\Theta} + \boldsymbol{\varepsilon}, \\ \mathbf{E}\boldsymbol{\varepsilon} = \mathbf{0}_n, \\ \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} = \sigma^2 \mathbf{I}_n, \\ (x^{(1)}, x^{(2)}, \dots, x^{(p)}) - \text{нечисловые переменные;} \\ \text{ранг матрицы } \mathbf{X} = p + 1 < n. \end{cases} \quad (4.1')$$

Когда дополнительно к условиям (4.1) (или (4.1')) постулируют *нормальный* характер распределения регрессионных остатков $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^\top$ (что записывается в виде $\boldsymbol{\varepsilon} \in N_n(\mathbf{0}; \sigma^2 \mathbf{I}_n)$), то говорят, что y и X связаны *нормальной* КЛММР.

Пример 4.1. Исследуется зависимость урожайности зерновых культур (y , ц/га) от ряда переменных, характеризующих различные факторы сельскохозяйственного производства, а именно:

$x^{(1)}$ — 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 гектар (т/га);

$x^{(5)}$ — количество химических средств защиты растений, расходуемых на гектар (ц/га).

Таблица 4.1. Исходные статистические данные для примера 4.1

i (номер района)	y_i	$x_i^{(1)}$	$x_i^{(2)}$	$x_i^{(3)}$	$x_i^{(4)}$	$x_i^{(5)}$
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
3	9,00	2,53	0,31	2,46	0,30	0,31
4	9,90	4,63	0,40	6,44	0,43	0,59
5	9,60	2,16	0,26	2,16	0,39	0,16
6	8,60	2,16	0,30	2,69	0,32	0,17
7	12,50	0,68	0,29	0,73	0,42	0,23
8	7,60	0,35	0,26	0,42	0,21	0,08
9	6,90	0,52	0,24	0,49	0,20	0,08
10	13,50	3,42	0,31	3,02	1,37	0,73
11	9,70	1,78	0,30	3,19	0,73	0,17
12	10,70	2,40	0,32	3,30	0,25	0,14
13	12,10	9,36	0,40	11,51	0,39	0,38
14	9,70	1,72	0,28	2,26	0,82	0,17
15	7,00	0,59	0,29	0,60	0,13	0,35
16	7,20	0,28	0,26	0,30	0,09	0,15
17	8,20	1,64	0,29	1,44	0,20	0,08
18	8,40	0,09	0,22	0,05	0,43	0,20
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

Исходные данные для 20 сельскохозяйственных районов области приведены в табл. 4.1.

Таким образом, в данном примере мы располагаем пространственной выборкой объема $n = 20$; число объясняющих переменных $p = 5$. Матрица X будет составлена из шести столбцов размерности 20 каждый, причем в качестве первого столбца используется вектор, состоящий из одних единиц, а столбцы со 2-го по 6-й представлены соответственно 3–7-м столбцами табл. 4.1. Вектор-столбец Y определяется 2-м столбцом табл. 4.1. Специальный анализ технологии сбора исходных статистических данных показал, что допущение о взаимной некоррелированности и гомоскедастичности регрессионных остатков ϵ может быть принято в качестве рабочей гипотезы. Поэтому мы можем записать уравнения статистической связи между y_i и $X_i = (x_i^{(1)}, x_i^{(2)}, x_i^{(3)}, x_i^{(4)}, x_i^{(5)})^T$ в виде (4.1').

4.2 Оценивание неизвестных параметров КЛММР: метод наименьших квадратов и метод максимального правдоподобия

Соотношения (4.1) и (4.1') определяют *специфицированные* уравнения статистической связи, существующей между результирующей переменной y и объясняющими переменными X . Однако значения участвующих в этих уравнениях параметров $\Theta = (\theta_0, \theta_1, \dots, \theta_p)^T$ и σ^2 нам не известны; их требуется определить (*статистически оценить*) по имеющимся в нашем распоряжении исходным статистическим данным вида (4.1a)–(4.3). Ниже описываются способы статистического оценивания параметров Θ и σ^2 в рамках КЛММР (метод наименьших квадратов) и в рамках *нормальной* КЛММР (метод максимального правдоподобия).

4.2.1 Метод наименьших квадратов (МНК)

В основе логики метода наименьших квадратов лежит стремление исследователя подобрать такие оценки $\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_p$ для неизвестных значений параметров функции регрессии соответственно $\theta_0, \theta_1, \dots, \theta_p$, при которых «*подогнанные*» (регрессионные) значения $\hat{\theta}_0 + \hat{\theta}_1 x_i^{(1)} + \dots + \hat{\theta}_p x_i^{(p)}$ результирующего показателя как можно меньше отличались бы от соответствующих *наблюдённых* значений y_i . Попробуем математически сформулировать этот принцип. Введем в качестве меры расхождения «подогнанного» и наблюдаемого (в i -м наблюдении) значений результирующего показателя разность

$$\hat{\varepsilon}_i = y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i^{(1)} - \dots - \hat{\theta}_p x_i^{(p)} \quad (4.11)$$

(будем в дальнейшем называть $\hat{\varepsilon}_i$ «*невязками*»). Очевидно значения $\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_p$ следует подбирать таким образом, чтобы минимизировать некоторую *интегральную* (по всем имеющимся наблюдениям) *характеристику невязок*. Примем за такую интегральную характеристику *подгонки* (выравнивания) значений y_i с помощью линейной функции от $x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)}$ ($i = 1, 2, \dots, n$) величину

$$Q(\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_p) = \sum_{i=1}^n (y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i^{(1)} - \dots - \hat{\theta}_p x_i^{(p)})^2 = \sum_{i=1}^n \hat{\varepsilon}_i^2. \quad (4.12)$$

Очевидно, величина Q будет определяться при заданной системе наблюдений (4.1a)–(4.3) конкретным выбором значений оценок параметров $\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_p$. *Оценки по методу наименьших квадратов (МНК-оценки) $\hat{\theta}_{0. \text{МНК}}, \hat{\theta}_{1. \text{МНК}}, \dots, \hat{\theta}_{p. \text{МНК}}$ как раз и подбираются таким образом, чтобы*

минимизировать величину Q , определенную соотношением (4.12), то есть:

$$Q(\hat{\theta}_{0. \text{мнк}}, \hat{\theta}_{1. \text{мнк}}, \dots, \hat{\theta}_{p. \text{мнк}}) = \min_{\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_p} Q(\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_p) \quad (4.13)$$

или

$$\hat{\Theta}_{\text{мнк}} = \arg \min_{\hat{\Theta}} Q(\hat{\Theta}). \quad (4.13')$$

Опишем процедуру решения оптимизационной задачи (4.13'). Начнем с простейшего частного случая, когда рассматривается зависимость y от *единственной* объясняющей переменной x (то есть $p = 1$). Этот случай в литературе обычно называют моделью **парной линейной регрессии**. Первое из уравнений связи (4.1) в данном случае имеет вид:

$$y_i = \theta_0 + \theta_1 x_i + \varepsilon_i, \quad i = 1, 2, \dots, n.$$

Критерий Q метода наименьших квадратов:

$$Q(\hat{\theta}_0, \hat{\theta}_1) = \sum_{i=1}^n (y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i)^2. \quad (4.12')$$

Необходимые условия экстремума по $\hat{\theta}_0$ и $\hat{\theta}_1$ функции $Q(\hat{\theta}_0, \hat{\theta}_1)$:

$$\begin{cases} \frac{\partial Q}{\partial \hat{\theta}_0} = -2 \sum_{i=1}^n (y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i) = 0, \\ \frac{\partial Q}{\partial \hat{\theta}_1} = -2 \sum_{i=1}^n x_i (y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i) = 0, \end{cases}$$

или, после раскрытия скобок и очевидных тождественных преобразований:

$$\begin{cases} n\hat{\theta}_0 + \hat{\theta}_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i, \\ \left(\sum_{i=1}^n x_i \right) \hat{\theta}_0 + \hat{\theta}_1 \cdot \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i. \end{cases} \quad (4.14)$$

Система (4.14) из двух линейных уравнений относительно $\hat{\theta}_0$ и $\hat{\theta}_1$ представляет так называемую *стандартную форму нормальных уравнений* (для случая $p = 1$). Ее решения легко выписываются в явном виде:

$$\hat{\theta}_{1. \text{мнк}} = \frac{n \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad (4.15)$$

$$\hat{\theta}_{0. \text{мнк}} = \bar{y} - \hat{\theta}_{1. \text{мнк}} \bar{x},$$

где $\bar{x} = \sum_{i=1}^n x_i/n$ и $\bar{y} = \sum_{i=1}^n y_i/n$.

Перейдем к случаю многих объясняющих переменных ($p > 1$). Как обычно, в этом случае более удобной оказывается матричная форма записи всех необходимых в данной задаче условий и соотношений:

$$\hat{\varepsilon} = Y - X\hat{\Theta} = (y_1 - \hat{\theta}_0 - \hat{\theta}_1 x_1^{(1)} - \dots - \hat{\theta}_p x_1^{(p)}, \dots, y_n - \hat{\theta}_0 - \hat{\theta}_1 x_n^{(1)} - \dots - \hat{\theta}_p x_n^{(p)})^T$$

— вектор-столбец невязок;

$$Q(\hat{\Theta}) = \sum_{i=1}^n (y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i^{(1)} - \dots - \hat{\theta}_p x_i^{(p)})^2 = (Y - X\hat{\Theta})^T (Y - X\hat{\Theta}) \quad (4.16)$$

— оптимизируемый (по $\hat{\Theta}$) критерий метода наименьших квадратов.

Перед тем как выписать необходимые условия экстремума функции $Q(\hat{\Theta})$ по $\hat{\Theta}$, преобразуем правую часть (4.16):

$$Q(\hat{\Theta}) = Y^T Y - 2\hat{\Theta}^T X^T Y + \hat{\Theta}^T X^T X \hat{\Theta}. \quad (4.16')$$

В этом преобразовании мы воспользовались правилом транспонирования произведения матриц (см. Приложение 2), а также тем, что $\hat{\Theta}^T X^T Y$ — число, а потому оно совпадает со своим транспонированным выражением $Y^T X \hat{\Theta}$.

Необходимые условия, которым удовлетворяют решения оптимизационной задачи (4.13'), получаются дифференцированием правой части (4.16') по $\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_p$. При выписывании получающейся при этом системы уравнений относительно $\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_p$ мы воспользуемся матричным обозначением производной

$$\frac{\partial Q(\hat{\Theta})}{\partial \hat{\Theta}} = \left(\frac{\partial Q(\hat{\Theta})}{\partial \hat{\theta}_0}, \frac{\partial Q(\hat{\Theta})}{\partial \hat{\theta}_1}, \dots, \frac{\partial Q(\hat{\Theta})}{\partial \hat{\theta}_p} \right)^T,$$

а также правилами записи матричного дифференцирования линейных и квадратичных функций от $\hat{\Theta}$ (см. п. П2.1 в Приложении 2):

$$\frac{\partial Q(\hat{\Theta})}{\partial \hat{\Theta}} = -2X^T Y + 2X^T X \hat{\Theta} = 0_{p+1}. \quad (4.17)$$

В (4.17) 0_{p+1} — это вектор-столбец размерности $p+1$, состоящий из одних нулей.

Разрешая систему уравнений (4.17) относительно $\hat{\Theta}$, получаем:

$$X^T X \hat{\Theta} = X^T Y \quad (4.18)$$

и следовательно:

В основной формуле метода наименьших квадратов (4.19) *подразумевается невырожденность матрицы* $X^T X$, которая следует из требования максимального ранга для матрицы X , входящего в описание КЛММР (см. (4.1) или (4.1')).

Применим общую формулу (4.18) к ранее рассмотренному частному случаю парной регрессии ($p = 1$). Очевидно, в этом случае

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad X^T X = \begin{pmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{pmatrix}, \quad X^T Y = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{pmatrix}.$$

Подставляя эти выражения в (4.18), получаем:

$$\begin{cases} n\hat{\theta}_0 + \hat{\theta}_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i, \\ \left(\sum_{i=1}^n x_i \right) \hat{\theta}_0 + \hat{\theta}_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i, \end{cases}$$

то есть ту же стандартную форму нормальных уравнений, с которой мы встретились в (4.14), анализируя эту же модель покомпонентно.

Пример 4.1 (продолжение). Применение формулы (4.19) к данным табл. 4.1 позволяет получить следующие МНК-оценки для параметров $\Theta = (\theta_0, \theta_1, \dots, \theta_5)$:

$$\begin{aligned} \hat{\theta}_{0. \text{МНК}} &= 3,515; \quad \hat{\theta}_{1. \text{МНК}} = -0,006; \quad \hat{\theta}_{2. \text{МНК}} = 15,542; \quad \hat{\theta}_{3. \text{МНК}} = 0,110; \\ \hat{\theta}_{4. \text{МНК}} &= 4,475; \quad \hat{\theta}_{5. \text{МНК}} = -2,932. \end{aligned}$$

Таким образом, оценка $\hat{f}(X)$ неизвестной функции регрессии $f(X)$ в данном случае имеет вид:

$$\hat{f}(X) = 3,515 - 0,006x^{(1)} + 15,542x^{(2)} + 0,110x^{(3)} + 4,475x^{(4)} - 2,932x^{(5)}.$$

Экономическое осмысление и интерпретацию полученных в данном примере результатов отложим до того момента, когда мы будем в состоянии охватить *весь комплекс* вопросов прикладного регрессионного анализа.

4.2.2 Метод максимального правдоподобия (ММП)

Метод максимального правдоподобия (ММП) может быть применен в тех случаях, когда с точностью до неизвестных значений параметров известен общий вид закона распределения вероятностей имеющих выборочных данных. Поэтому, если мы проводим регрессионный

анализ в рамках *нормальной* КЛММП, то есть если дополнительно к условиям (4.1) постулируется *нормальность регрессионных остатков* $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$, то, учитывая их взаимную некоррелированность (которая в нормальном случае влечет за собой их взаимную статистическую независимость, см. Свойство 5 в п. 3.2.2), можно выписать функцию правдоподобия в терминах остатков $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$:

$$\begin{aligned} L(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n \mid \Theta; \sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i - \theta_0 - \theta_1 x_i^{(1)} - \dots - \theta_p x_i^{(p)})^2} = \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp \left[-\frac{1}{2\sigma^2} (Y - \mathbf{X}\Theta)^\top (Y - \mathbf{X}\Theta) \right]. \end{aligned} \quad (4.20)$$

Оценки $\hat{\Theta}_{\text{ММП}}$ и $\hat{\sigma}_{\text{ММП}}^2$ максимального правдоподобия определяются как такие значения Θ и σ^2 , при которых функция правдоподобия L (или, что то же, логарифмическая функция правдоподобия $l = \ln L$) достигает своей максимальной величины. Соответствующие уравнения ММП получаются приравниванием к нулю производных функции l по Θ и σ^2 :

$$\begin{aligned} l(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n \mid \Theta; \sigma^2) &= -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} (Y - \mathbf{X}\Theta)^\top (Y - \mathbf{X}\Theta); \\ \begin{cases} \frac{\partial l}{\partial \Theta} = -\frac{1}{2\sigma^2} \frac{\partial}{\partial \Theta} [(Y - \mathbf{X}\Theta)^\top (Y - \mathbf{X}\Theta)] = \mathbf{0}_{p+1}, \\ \frac{\partial l}{\partial \sigma^2} = -\frac{n}{2} \cdot \frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2} (Y - \mathbf{X}\Theta)^\top (Y - \mathbf{X}\Theta) = 0. \end{cases} \end{aligned} \quad (4.21)$$

Первая строка (4.21) после сокращения левой и правой частей на $-1/2\sigma^2$ повторяет систему уравнений (4.17) метода наименьших квадратов. Следовательно;

$$\hat{\Theta}_{\text{ММП}} = \hat{\Theta}_{\text{МНК}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top Y. \quad (4.22)$$

Вторая строка системы (4.21) позволяет вычислить ММП-оценку для σ^2 :

$$\begin{aligned} \hat{\sigma}_{\text{ММП}}^2 &= \frac{1}{n} (Y - \mathbf{X}\hat{\Theta})^\top (Y - \mathbf{X}\hat{\Theta}) = \\ &= \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i^{(1)} - \dots - \hat{\theta}_p x_i^{(p)})^2, \end{aligned} \quad (4.23)$$

где $\hat{\Theta} = (\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_p)$ — оценки по методу наименьших квадратов (они же — оценки по методу максимального правдоподобия) неизвестных коэффициентов регрессии Θ . В дальнейшем оценки, полученные по формулам (4.22), мы будем обозначать просто $\hat{\Theta}$ (без индексирования метода).

4.2.3 Статистические свойства оценок параметров КЛММП. Проверка гипотез и интервальные оценки параметров

При повторениях выборок того же самого объема n из той же самой анализируемой генеральной совокупности и *при тех же самых значениях объясняющих переменных* X наблюдаемые значения результирующего показателя y будут случайным образом варьировать (за счет стохастического характера остатков ϵ), а следовательно, будут варьировать и зависящие от $y_1, y_2, \dots, y_n$ значения оценок $\hat{\Theta}$ и $\hat{\sigma}^2$, хотя каждый раз (то есть для каждой выборки (4.1a)–(4.3)) мы их будем вычислять по одним и тем же формулам, соответственно (4.22) и (4.23). Другими словами, *наши оценки «ведут себя» как случайные величины*. Поэтому естественно попытаться проанализировать их поведение и в первую очередь ответить на вопросы:

- *к чему стремятся (по вероятности) оценки $\hat{\Theta}$ и $\hat{\sigma}^2$ при неограниченном росте объема выборки?*
- *каковы средние значения этих оценок для выборок фиксированного объема n ?*
- *каковы основные характеристики случайного разброса в значениях этих оценок относительно истинных значений оцениваемых параметров?*

Состоятельность оценок $\hat{\Theta}$ и $\hat{\sigma}_{\text{ММП}}^2$. В данном случае свойство состоятельности наших оценок определяется структурой матрицы X . Существуют различные формулировки условий (в терминах элементов матрицы X), при которых оценки $\hat{\Theta}$ и $\hat{\sigma}_{\text{ММП}}^2$ являются состоятельными. Пожалуй, наиболее удобное для приложений условие состоятельности оценок $\hat{\Theta}$ и $\hat{\sigma}_{\text{ММП}}^2$ — следующее (см., например, [Себер Дж., п. 3.2]):

оценки $\hat{\Theta}$ и $\hat{\sigma}_{\text{ММП}}^2$ являются состоятельными тогда и только тогда, когда наименьшее собственное значение матрицы $X^T X$ стремится к бесконечности при $n \rightarrow \infty$.

Напомним, что наименьшее собственное значение $\lambda_{\min}$ матрицы $X^T X$ определяется как минимальный по величине корень уравнения $|X^T X - \lambda I_{p+1}| = 0$ (матрица $X^T X$ как симметричная и положительно определенная имеет $p + 1$ действительных положительных собственных значений $\lambda_1, \lambda_2, \dots, \lambda_p$, см. Приложение 2). Можно показать, что сформулированное условие состоятельности оценок $\hat{\Theta}$ и $\hat{\sigma}_{\text{ММП}}^2$ для случая, например,

парной линейной регрессионной зависимости (то есть при $p = 1$) равносильно требованию:

$$\sum_{i=1}^n (x_i - \bar{x})^2 \asymp n \quad \text{при} \quad n \rightarrow \infty, \quad (4.24)$$

где символ $\asymp$ означает, что величины, стоящие слева и справа от него, ведут себя асимптотически одинаково при $n \rightarrow \infty$, то есть:

$$\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} \rightarrow c > 0 \quad \text{при} \quad n \rightarrow \infty.$$

Так, например, условие (4.24) может быть нарушенным, если по мере роста объема выборки n наблюдаемые значения объясняющей переменной будут концентрироваться в пределах неограниченно сужающейся окрестности ее средней величины $\bar{x}(n)$. Грубо говоря, даже неограниченное добавление наблюдений в бесконечно малой окрестности среднего значения объясняющей переменной не несет полезной дополнительной информации относительно расположения анализируемой прямой регрессии.

Несмещенность оценок. Напомним, что оценка $\hat{\Theta}$ параметра Θ называется несмещенной, если $E\hat{\Theta} = \Theta$ (для векторного параметра это равенство понимается одновременно выполняющимся для *всех* компонент векторов $\hat{\Theta}$ и Θ).

Чтобы подсчитать среднее значение оценки (4.22), подставим в формулу (4.22) вместо Y его выражение из основного (первого) соотношения системы (4.1'):

$$\hat{\Theta} = (X^T X)^{-1} X^T (X\Theta + \varepsilon) = \Theta + (X^T X)^{-1} X^T \varepsilon. \quad (4.25)$$

В последнем равенстве мы воспользовались тем, что $(X^T X)^{-1} (X^T X) = I_{p+1}$. Таким образом, оценка $\hat{\Theta}$ представлена как сумма истинного (не известного нам) значения Θ и линейной комбинации случайных остатков $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$. Беря математические ожидания от левой и правой частей (4.25) с учетом того, что величины Θ и $(X^T X)^{-1} X^T$ неслучайны, а средние значения остатков равны нулю (то есть $E\varepsilon = 0_n$), получаем:

$$E\hat{\Theta} = E\Theta + (X^T X)^{-1} X^T E\varepsilon = \Theta. \quad (4.26)$$

Тем самым показано, что МНК-оценки (они же ММП-оценки) $\hat{\Theta}$ неизвестных параметров КЛММП являются *несмещенными*.

Покажем, что в отличие от $\hat{\Theta}$ оценка $\hat{\sigma}_{\text{ММП}}^2$ параметра σ^2 оказывается смещенной. Правда, проведенный ниже анализ позволяет «подправить» оценку $\hat{\sigma}_{\text{ММП}}^2$ таким образом, чтобы смещение было устранено.

Итак, наша цель — вычислить математическое ожидание оценки (4.23), то есть определить значение

$$\mathbf{E} \left[\frac{1}{n} (\mathbf{Y} - \mathbf{X}\hat{\Theta})^\top (\mathbf{Y} - \mathbf{X}\hat{\Theta}) \right],$$

где $\hat{\Theta}$ определено формулой (4.22). С этой целью выразим вектор невязок $\mathbf{Y} - \mathbf{X}\hat{\Theta}$ через остаточные случайные компоненты модели $\boldsymbol{\varepsilon}$:

$$\begin{aligned} \mathbf{Y} - \mathbf{X}\hat{\Theta} &= (\mathbf{X}\Theta + \boldsymbol{\varepsilon}) - \mathbf{X}[(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top (\mathbf{X}\Theta + \boldsymbol{\varepsilon})] = \\ &= \boldsymbol{\varepsilon} - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \boldsymbol{\varepsilon} = [\mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top] \boldsymbol{\varepsilon} = \mathbf{Z}\boldsymbol{\varepsilon}. \end{aligned} \quad (4.27)$$

В соотношении (4.27) использовано обозначение для $(n \times n)$ -матрицы

$$\mathbf{Z} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top. \quad (4.28)$$

Отметим, что матрица $\mathbf{Z}$ является симметричной и *идемпотентной* (см. Приложение 2), то есть $\mathbf{Z}^\top = \mathbf{Z}$ (симметричность) и $\mathbf{Z}^2 = \mathbf{Z}$ (идемпотентность). Оба свойства устанавливаются непосредственной проверкой. Действительно:

$$\mathbf{Z}^\top = (\mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top)^\top = \mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top = \mathbf{Z}$$

(при переходе к правой части использовались правило транспонирования произведения матриц, а также симметричность матрицы $(\mathbf{X}^\top \mathbf{X})^{-1}$);

$$\begin{aligned} \mathbf{Z}^2 &= \mathbf{Z}\mathbf{Z} = [\mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top] [\mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top] = \\ &= \mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top + [\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top] \times \\ &\times [\mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top] = \mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top = \mathbf{Z}. \end{aligned}$$

Теперь, воспользовавшись (4.27), а также симметричностью и идемпотентностью матрицы $\mathbf{Z}$, мы можем вычислить величину

$$\begin{aligned} \mathbf{E}[(\mathbf{Y} - \mathbf{X}\hat{\Theta})^\top (\mathbf{Y} - \mathbf{X}\hat{\Theta})] &= \mathbf{E}[(\mathbf{Z}\boldsymbol{\varepsilon})^\top (\mathbf{Z}\boldsymbol{\varepsilon})] = \\ &= \mathbf{E}(\boldsymbol{\varepsilon}^\top \mathbf{Z}^\top \mathbf{Z} \boldsymbol{\varepsilon}) = \mathbf{E}(\boldsymbol{\varepsilon}^\top \mathbf{Z} \boldsymbol{\varepsilon}) = \sigma^2 \operatorname{tr} \mathbf{Z} = \\ &= \sigma^2 \operatorname{tr} [\mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top] = \\ &= \sigma^2 [\operatorname{tr} \mathbf{I}_n - \operatorname{tr} \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top] = \\ &= \sigma^2 [n - \operatorname{tr} (\mathbf{X}^\top \mathbf{X})^{-1} (\mathbf{X}^\top \mathbf{X})] = \\ &= \sigma^2 (n - \operatorname{tr} \mathbf{I}_{p+1}) = \sigma^2 [n - (p + 1)]. \end{aligned} \quad (4.29)$$

В ходе преобразований мы воспользовались следующими двумя свойствами (см. Приложение 2):

- (а) $\mathbf{E}(\boldsymbol{\varepsilon}^\top \mathbf{A} \boldsymbol{\varepsilon}) = \sigma^2 \operatorname{tr} \mathbf{A}$,
- (б) $\operatorname{tr} (\mathbf{A} \mathbf{B} \mathbf{C}) = \operatorname{tr} (\mathbf{B} \mathbf{C} \mathbf{A}) = \operatorname{tr} (\mathbf{C} \mathbf{A} \mathbf{B})$.

Из соотношения (4.29) следует, что оценка $\hat{\sigma}_{\text{ММП}}^2$, определенная формулой (4.23), является смещенной, так как

$$\begin{aligned} E\hat{\sigma}_{\text{ММП}}^2 &= E\left[\frac{1}{n}(Y - \mathbf{X}\hat{\Theta})^\top(Y - \mathbf{X}\hat{\Theta})\right] = \\ &= \frac{1}{n}E[(Y - \mathbf{X}\hat{\Theta})^\top(Y - \mathbf{X}\Theta)] = \sigma^2\left(1 - \frac{p+1}{n}\right). \end{aligned} \quad (4.30)$$

Воспользуемся вместо $\hat{\sigma}_{\text{ММП}}^2$ оценкой

$$\begin{aligned} \hat{\sigma}^2 &= \frac{n}{n-p-1}\hat{\sigma}_{\text{ММП}}^2 = \\ &= \frac{1}{n-p-1}\sum_{i=1}^n(y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i^{(1)} - \dots - \hat{\theta}_p x_i^{(p)})^2 = \\ &= \frac{1}{n-p-1}(Y - \hat{\Theta}\mathbf{X})^\top(Y - \hat{\Theta}\mathbf{X}). \end{aligned} \quad (4.31)$$

Очевидно, такой способ оценивания неизвестной дисперсии остатков ε_i (так называемой *остаточной дисперсии* σ^2) уже будет *несмещенным*.

Пример 4.1 (продолжение). Оценка остаточной дисперсии σ^2 по данным табл. 4.1 по формуле (4.31) с использованием результатов оценивания функции регрессии дает

$$\hat{\sigma}^2 = 2,59. \quad (4.32)$$

Ковариационная матрица вектора $\hat{\Theta}$ МНК-оценок. Ковариационная матрица $\Sigma_{\hat{\Theta}}$ вектора $\hat{\Theta}$ была введена соотношением (4.10) при описании КЛММП в п. 4.1. С помощью ее элементов подсчитываются основные показатели случайного разброса оценок $\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_p$ около соответствующих истинных значений анализируемых параметров и одновременно характеристики взаимозависимости полученных оценок. Действительно, из определения $\Sigma_{\hat{\Theta}}$ следует, что ее диагональные элементы $E(\hat{\theta}_0 - \theta_0)^2, E(\hat{\theta}_1 - \theta_1)^2, \dots, E(\hat{\theta}_p - \theta_p)^2$ задают средние квадраты ошибок соответствующих оценок (а для *несмещенных* оценок, какими и являются МНК-оценки, это и есть дисперсии оценок!).

Наконец, степень тесноты парных линейных связей между компонентами $\hat{\theta}^{(l)}$ и $\hat{\theta}^{(j)}$ вектора оценок $\hat{\Theta}$ ($l, j = 0, 1, \dots, p; l \neq j$) определяется коэффициентом корреляции $r(\hat{\theta}^{(l)}, \hat{\theta}^{(j)})$ (см. п. 3.2.2), который в свою очередь выражается через элементы ковариационной матрицы $\Sigma_{\hat{\Theta}} = (\sigma_{lj}(\hat{\Theta})), -l, j = 0, 1, \dots, p$:

$$r(\hat{\theta}^{(l)}, \hat{\theta}^{(j)}) = \frac{\sigma_{lj}}{(\sigma_{ll}\sigma_{jj})^{1/2}}.$$

Все это говорит о том, что *ковариационная матрица $\Sigma_{\hat{\Theta}}$ вектора оценок $\hat{\Theta}$ содержит важнейшую информацию о качестве оценок $\hat{\Theta}$* . Определим ее в терминах параметров КЛММП и исходных статистических

данных (4.1a)–(4.3). Для этого при выражении разности $\hat{\Theta} - \Theta$ воспользуемся ранее полученным соотношением (4.25):

$$\begin{aligned}\Sigma_{\hat{\Theta}} &= E[(\hat{\Theta} - \Theta)(\hat{\Theta} - \Theta)^T] = \\ &= E\{[(X^T X)^{-1} X^T \varepsilon][(X^T X)^{-1} X^T \varepsilon]^T\} = \\ &= E[(X^T X)^{-1} X^T \varepsilon \varepsilon^T X (X^T X)^{-1}] = \\ &= (X^T X)^{-1} X^T E(\varepsilon \varepsilon^T) X (X^T X)^{-1} = \\ &= (X^T X)^{-1} X^T \sigma^2 I_n X (X^T X)^{-1} = \\ &= \sigma^2 (X^T X)^{-1} X^T X (X^T X)^{-1} = \sigma^2 (X^T X)^{-1}.\end{aligned}\quad (4.33)$$

При выводе соотношения (4.33) мы воспользовались:

- правилом транспонирования произведения матриц (см. Приложение 2);
- специальным видом ковариационной матрицы вектора остатков $\Sigma_{\varepsilon} = E(\varepsilon \varepsilon^T) = \sigma^2 I_n$ (см. (4.1'));
- тем, что при умножении на единичную любая «подходящая»² матрица не меняется (в нашем случае $X^T I_n = X^T$).

Заметим в заключение, что в статистической практике пользуются не самой матрицей $\Sigma_{\hat{\Theta}}$, а ее оценкой

$$\hat{\Sigma}_{\hat{\Theta}} = \hat{\sigma}^2 (X^T X)^{-1}, \quad (4.33')$$

где $\hat{\sigma}^2$ определяется по формуле (4.31). Кроме того, в большинстве случаев ограничиваются выводом на печать и использованием только среднеквадратических ошибок

$$s_l = \hat{\sigma}_l^{1/2}(\hat{\Theta}) = \hat{\sigma} \sqrt{a_{ll}} \quad (4.34)$$

оценок $\hat{\theta}_l$ ($l = 0, 1, \dots, p$); в соотношении (4.34) $\hat{\sigma}_{ll}(\hat{\Theta})$ — l -й диагональный элемент матрицы $\hat{\Sigma}_{\hat{\Theta}}$ (см. (4.33')), $\hat{\sigma}$ определено формулой (4.31), а a_{ll} — l -й диагональный элемент матрицы $A = (X^T X)^{-1}$, $-l = 0, 1, \dots, p$.

Пример 4.1 (продолжение). По формулам (4.31), (4.33'), (4.34) подсчитаны среднеквадратические ошибки s_l в оценивании коэффициентов регрессии θ_l ($l = 0, 1, \dots, 5$). Стандартная форма компьютерной выдачи результатов счета, объединяющая информацию о значениях оценок регрессии $\hat{\theta}_l$ и их среднеквадратических ошибках s_l , как правило,

² Матрица A как сомножитель, стоящий слева от I_n , является «подходящей», если длина ее строки (то есть число ее столбцов) равна размерности n единичной матрицы I_n .

имеет следующий вид:

$$\hat{f}(X) = \underset{(5,41)}{3,515} - \underset{(0,60)}{0,006} x^{(1)} + \underset{(21,59)}{15,542} x^{(2)} + \underset{(0,85)}{0,110} x^{(3)} + \underset{(1,54)}{4,475} x^{(4)} - \underset{(3,09)}{2,932} x^{(5)}. \quad (4.35)$$

В скобках под значениями оцененных коэффициентов регрессии $\hat{\theta}_i$ указаны их среднеквадратические ошибки z_i . Анализируя полученный результат, читатель даже на этой стадии подготовленности отметит для себя, что среднеквадратические ошибки в оценке всех коэффициентов, кроме θ_1 , превышают значения самих коэффициентов (что говорит о статистической ненадежности последних). Позже мы вернемся к этому примеру, чтобы более детально разобраться в возникшей ситуации.

Оптимальность МНК-оценок. При сравнении различных способов оценивания решающей характеристикой качества оценки $\hat{\theta}$ неизвестного числового параметра θ оказывается средний квадрат ошибки $E(\hat{\theta} - \theta)^2$. Поэтому говорят, что оценка $\hat{\theta}_1$ точнее (лучше, эффективнее), чем оценка $\hat{\theta}_2$, если $E(\hat{\theta}_1 - \theta)^2 < E(\hat{\theta}_2 - \theta)^2$. Соответственно оценка $\hat{\theta}_{\text{опт}}$ является оптимальной в классе оценок M , если

$$E(\hat{\theta}_{\text{опт}} - \theta)^2 = \min_{\hat{\theta} \in M} E(\hat{\theta} - \theta)^2. \quad (4.36)$$

Анализируя КЛММР, мы имеем дело с вектором оценок $\hat{\Theta} = (\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_p)^T$. Как определить качество и оптимальность векторной оценки? По видимому, для подавляющего большинства практических задач было бы достаточно, если бы удалось показать, что для любой линейной функции $C^T \Theta = \theta_C$ от неизвестных параметров $\theta_0, \theta_1, \dots, \theta_p$ оценка $C^T \hat{\Theta}_{\text{МНК}}$ является оптимальной для параметра θ_C в смысле (4.36) в некотором достаточно широком классе оценок M (здесь $C = (c_0, c_1, \dots, c_p)^T$ — $(p+1)$ -мерный вектор-столбец произвольных констант c_j). Действительно, полагая $C^T = (0, 0, \dots, 0, 1, 0, \dots, 0)$, где единица стоит на j -м месте, получаем оптимальность $\hat{\Theta}$ с точки зрения точности оценивания параметра θ_j ($j = 0, 1, \dots, p$). Полагая $C^T = (1, x^{(1)}, \dots, x^{(p)})$, получаем оптимальность $\hat{\Theta}$ с точки зрения точности оценивания неизвестной функции регрессии $f(X) = \theta_0 + \theta_1 x^{(1)} + \dots + \theta_p x^{(p)}$ при любых заданных значениях объясняющих переменных X . Как видим, подобная интерпретация оптимальности оценок $\hat{\Theta}_{\text{МНК}}$ оказывается вполне уместной и естественной. Поэтому именно в таком смысле формулируется основной результат, обосновывающий оптимальные свойства МНК-оценок (теорема Гаусса-Маркова):

- Рассматривается задача статистического оценивания заданной линейной функции $\theta_C = C^T \Theta$ от неизвестных параметров регрессионной модели (4.1') ($C = (c_0, c_1, \dots, c_p)^T$ — вектор-столбец заданных констант). Пусть $M = \{B^T Y\}$ — класс линейных относительно $y_1, y_2, \dots, y_n$ и несмещенных оценок параметра θ_C . ($B =$

$= (b_1, b_2, \dots, b_n)^T$ — вектор-столбец коэффициентов, с помощью которых формируются оценки класса M). Тогда оценка $C^T \hat{\Theta}_{\text{мнк}}$ является оптимальной в классе M оценкой параметра θ_C в смысле (4.36), то есть

$$\mathbf{E}(C^T \hat{\Theta}_{\text{мнк}} - \theta_C)^2 \leq \mathbf{E}(B^T Y - \theta_C)^2 \quad (4.37)$$

для любой оценки $B^T Y$ из класса M .

Докажем это утверждение. Прежде всего отметим, что оценка $C^T \hat{\Theta}_{\text{мнк}}$ принадлежит классу M , так как, во-первых, она линейна относительно Y (см. формулу (4.22)), а во-вторых, она является несмещенной (так как $\mathbf{E}(C^T \hat{\Theta}_{\text{мнк}}) = C^T \mathbf{E} \hat{\Theta}_{\text{мнк}} = C^T \Theta$ в силу того, что $\hat{\Theta}_{\text{мнк}}$ является несмещенной оценкой параметра Θ , см. (4.26)). Наша цель — доказать справедливость неравенства (4.37), поэтому проанализируем и сравним левую и правую части этого соотношения. Заметим сначала, что требование несмещенности оценок $B^T Y$ из класса M обуславливает существование определенных ограничений на компоненты векторов B , а именно:

$$\mathbf{E}(B^T Y) = \mathbf{E}[B^T (X\Theta + \varepsilon)] = B^T X\Theta = \theta_C = C^T \Theta,$$

откуда следует, что

$$B^T X = C^T. \quad (4.38)$$

Теперь вычислим и сравним средние квадраты ошибок для оценок $B^T Y$ и $C^T \hat{\Theta}_{\text{мнк}}$:

$$\begin{aligned} \mathbf{E}(B^T Y - C^T \Theta)^2 &= \mathbf{E}[B^T (X\Theta + \varepsilon) - C^T \Theta]^2 = \\ &= \mathbf{E}(B^T X\Theta + B^T \varepsilon - C^T \Theta)^2 = \mathbf{E}(C^T \Theta + B^T \varepsilon - C^T \Theta)^2 \\ &= \mathbf{E}(B^T \varepsilon \varepsilon^T B) = B^T \mathbf{E}(\varepsilon \varepsilon^T) B = \sigma^2 (B^T B). \end{aligned} \quad (4.39)$$

При выводе соотношения (4.39) мы воспользовались равенством (4.38), специальным видом ковариационной матрицы вектора остатков ε ($\mathbf{E}(\varepsilon \varepsilon^T) = \sigma^2 I_n$) и возможностью представления квадрата числа $B^T \varepsilon$ в виде $B^T \varepsilon \varepsilon^T B$.

$$\begin{aligned} \mathbf{E}(C^T \hat{\Theta}_{\text{мнк}} - C^T \Theta)^2 &= \mathbf{E}[C^T (\hat{\Theta}_{\text{мнк}} - \Theta)]^2 = \\ &= \mathbf{E}[C^T (\hat{\Theta}_{\text{мнк}} - \Theta)(\hat{\Theta}_{\text{мнк}} - \Theta)^T C] = \\ &= C^T \mathbf{E}[(\hat{\Theta}_{\text{мнк}} - \Theta)(\hat{\Theta}_{\text{мнк}} - \Theta)^T] C = \\ &= \sigma^2 C^T (X^T X)^{-1} C = \\ &= \sigma^2 B^T X (X^T X)^{-1} X^T B. \end{aligned} \quad (4.40)$$

При выводе соотношения (4.40) мы воспользовались возможностью представления квадрата числа $C^T (\hat{\Theta}_{\text{мнк}} - \Theta)$ в виде $C^T (\hat{\Theta}_{\text{мнк}} - \Theta) \times$

$\times (\hat{\Theta}_{\text{МНК}} - \Theta)^T C$, знанием ковариационной матрицы вектора оценок $\hat{\Theta}_{\text{МНК}}$ (см. (4.33)) и выражением C^T через $B^T X$ (см. (4.38)).

Остается показать, что разность правых частей (4.39) и (4.40) всегда неотрицательна. Действительно:

$$\begin{aligned} E(B^T Y - C^T \Theta)^2 - E(C^T \hat{\Theta}_{\text{МНК}} - C^T \Theta)^2 &= \sigma^2 (B^T B) - \\ &- \sigma^2 B^T X (X^T X)^{-1} X^T B = \sigma^2 B^T [I_n - X (X^T X)^{-1} X^T] B = \\ &= \sigma^2 B^T Z B, \end{aligned} \quad (4.41)$$

где матрица $Z = I_n - X (X^T X)^{-1} X^T$ нам уже встречалась при анализе смещения оценки $\hat{\sigma}_{\text{ММП}}^2$ (см. (4.28) и (4.29)). Было установлено, что она — симметричная и идемпотентная. Кроме того, из хода преобразования в соотношении (4.29) следует, что матрица Z — *неотрицательно определенная, то есть для любого вектора $B = (b_1, b_2, \dots, b_n)^T$ выражение $B^T Z B$ всегда неотрицательно* (в ходе преобразований (4.29), показано, что сумма квадратов невязок, являющаяся по определению существенно неотрицательным числом, представима в виде $\epsilon^T Z \epsilon$, где вектор невязок $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)^T$ может быть составлен из произвольных действительных компонент). А неотрицательность правой части (4.41) как раз и доказывает справедливость неравенства (4.37).

Свойства оценок, справедливые только при дополнительном условии нормальности регрессионных остатков (то есть в рамках *нормальной КЛММП*). Если мы несколько сузим класс рассматриваемых моделей, добавив к условиям (4.1) КЛММП требование нормальности регрессионных остатков, то соответствующая *нормальная КЛММП* может быть кратко описана следующими условиями:

$$\begin{cases} Y = X\Theta + \epsilon, \\ \epsilon \in N_n(0; \sigma^2 I_n), \\ (x^{(1)}, x^{(2)}, \dots, x^{(p)}) - \text{нечслучайные переменные;} \\ \text{ранг матрицы } X = p + 1. \end{cases} \quad (4.1'')$$

В (4.1'') $N_k(a; \Sigma)$, как и прежде, обозначает k -мерный нормальный з.р.в. с вектором средних значений, равным a , и ковариационной матрицей, равной Σ , а запись $\xi \in N_k(a; \Sigma)$ свидетельствует о том, что многомерная случайная величина ξ распределена по этому нормальному закону.

Сформулированные выше статистические свойства оценок $\hat{\Theta}_{\text{МНК}}$ и $\hat{\sigma}^2$ справедливы и без предположения о нормальности остатков. Однако, даже располагая информацией о состоятельности, несмещенности и оптимальности оценок $\hat{\Theta}_{\text{МНК}}$ и $\hat{\sigma}^2$, мы не сможем решить задач с *построением точных доверительных интервалов* для истинных значений этих параметров, так же как и для неизвестных значений функции регрессии или индивидуального значения анализируемого результирующего показате-

ля при заданных величинах объясняющих переменных. *Необходимой базой для решения этих задач является знание законов распределения вероятностей используемых оценок.* Именно это знание доставляют нам следующие (справедливые в рамках нормальной КЛММП) результаты, являющиеся по существу многомерным обобщением известной теоремы Фишера о распределении выборочного среднего значения и выборочной дисперсии, построенных по выборке из *нормальной* генеральной совокупности:

- оценки $\hat{\Theta}_{\text{мнк}}$ подчиняются $(p+1)$ -мерному нормальному з.р.в. с вектором средних значений, равных истинным значениям анализируемых параметров Θ , и с ковариационной матрицей $\Sigma_{\hat{\Theta}}$, определяемой соотношением (4.33), то есть

$$\hat{\Theta}_{\text{мнк}} \in N_{p+1}(\Theta; \sigma^2(\mathbf{X}^T \mathbf{X})^{-1}); \quad (4.42)$$

- случайная величина $(n-p-1)\hat{\sigma}^2/\sigma^2$ подчиняется χ^2 -распределению с $n-p-1$ степенями свободы, то есть

$$(n-p-1)\frac{\hat{\sigma}^2}{\sigma^2} \in \chi^2(n-p-1); \quad (4.43)$$

- оценки $\hat{\Theta}_{\text{мнк}}$ и $\hat{\sigma}^2$ являются статистически независимыми. (4.44)

Доказательство справедливости (4.43) основано на следующих фактах: 1) среднее значение и ковариационная матрица вектора оценок $\hat{\Theta}_{\text{мнк}}$ были подсчитаны ранее (см. (4.26) и (4.33)); 2) оценки $\hat{\Theta}_{\text{мнк}}$ строятся как *линейные комбинации нормально распределенных независимых случайных величин* y_i или ε_i (это видно соответственно из (4.19) и (4.25)), а любая линейная комбинация независимых нормально распределенных случайных величин снова распределена по нормальному закону. Утверждение (4.43) и статистическая независимость оценок $\hat{\Theta}_{\text{мнк}}$ и $\hat{\sigma}^2$ следует, например, из теоремы 8.2.2 [Андерсон (1963)].

Сформулируем одно важное следствие из приведенных выше результатов, которое позволит нам в дальнейшем воспользоваться t -распределением для проверки гипотез о численных значениях каждого из регрессионных коэффициентов $\theta_0, \theta_1, \dots, \theta_p$ и для построения соответствующих доверительных интервалов.

Следствие. Пусть θ_l^0 — истинное (гипотетическое) значение l -го коэффициента регрессии модели (4.1''). Тогда статистика

$$\gamma_l = \frac{\hat{\theta}_{l, \text{мнк}} - \theta_l^0}{s_l} = \frac{\hat{\theta}_{l, \text{мнк}} - \theta_l^0}{\hat{\sigma} \sqrt{a_{ll}}}, \quad l = 0, 1, \dots, p, \quad (4.45)$$

распределена по закону Стьюдента (t -распределения) с $n - p - 1$ степенями свободы. В (4.45) среднеквадратическая ошибка s_l оценки $\hat{\theta}_{l.\text{мнк}}$ определяется по формуле (4.34), $\hat{\sigma}^2$ — по формуле (4.31) и a_{ll} — это l -й диагональный элемент матрицы $\mathbf{A} = (\mathbf{X}^\top \mathbf{X})^{-1}$.

$$\text{Правило проверки гипотез вида } H_0: \theta_l = \theta_l^0 \quad (4.46)$$

Теперь мы можем сформулировать основанное на (4.45) правило проверки гипотез H_0 . Пусть θ_l^0 — заданное гипотетическое значение l -го коэффициента регрессии. По заданному уровню значимости критерия α из таблиц П1.6 процентных точек t -распределения (см. Приложение 1) находим 100 α /2%-ную точку распределения Стьюдента с $n - p - 1$ степенью свободы $t_{\alpha/2}(n - p - 1)$. Если окажется, что абсолютная величина статистики γ_l , определенной соотношением (4.45), больше значения $t_{\alpha/2}(n - p - 1)$, то проверяемую гипотезу (4.46) следует отвергнуть (вероятность ошибиться при этом равна α). В противном случае (то есть при $|\gamma_l| \leq t_{\alpha/2}(n - p - 1)$) гипотеза H_0 не отвергается. В распространенном частном случае проверки гипотезы $\theta_l = 0$ (что интерпретируется как полное отсутствие влияния объясняющей переменной $x^{(l)}$ на результирующий показатель y в рамках анализируемой линейной модели и является поводом к обсуждению вопроса об исключении $x^{(l)}$ из этой модели) решение о принятии или отклонении проверяемой гипотезы, следовательно, принимается на основании сравнения величины $|\hat{\theta}_{l.\text{мнк}}|/s_l$ со значением $t_{\alpha/2}(n - p - 1)$: при $(|\hat{\theta}_{l.\text{мнк}}|/s_l) > t_{\alpha/2}(n - p - 1)$ гипотеза « $H_0: \theta_l = 0$ » отвергается.

Построение доверительного интервала (интервальной оценки) для неизвестного значения коэффициента регрессии θ_l

Приведенное выше следствие (4.45) дает нам возможность построить доверительный интервал для каждого «отдельно взятого» неизвестного коэффициента регрессии θ_l . В частности, из (4.45) следует, что с доверительной вероятностью $P = 1 - \alpha$ истинное значение θ_l l -го коэффициента регрессии должно удовлетворять неравенствам

$$\hat{\theta}_{l.\text{мнк}} - s_l t_{\alpha/2}(n - p - 1) \leq \theta_l \leq \hat{\theta}_{l.\text{мнк}} + s_l t_{\alpha/2}(n - p - 1). \quad (4.47)$$

4.3 Анализ вариации результирующего показателя y и выборочный коэффициент детерминации $\hat{R}_{y.X}^2$

Из структуры КЛММР (4.1) видно, что общая изменчивость (вариация) результирующей переменной y обусловлена, во-первых, изменением значений функции регрессии $f(X_i) = \theta_0 + \theta_1 x_i^{(1)} + \dots + \theta_p x_i^{(p)}$ (которое в

свою очередь определяется варьированием значений объясняющих переменных $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$ и, во-вторых, поведением случайных остатков

$$\varepsilon_i = y_i - f(X_i), \quad i = 1, 2, \dots, n.$$

Определим общую выборочную (наблюдаемую) вариацию результирующей переменной y величиной

$$\text{вар}(y) = \sum_{i=1}^n (y_i - \bar{y})^2 = (Y - \bar{Y})^T (Y - \bar{Y}),$$

где $\bar{y} = \sum_{i=1}^n y_i / n$ — выборочное среднее значение результирующей переменной y , $Y = (y_1, y_2, \dots, y_n)^T$, а $\bar{Y} = (\bar{y}, \bar{y}, \dots, \bar{y})^T$ — n -мерный вектор-столбец.

Обозначим значения оцененной функции регрессии

$$\hat{f}_i = \hat{f}(X_i) = \hat{\theta}_0 + \hat{\theta}_1 x_i^{(1)} + \dots + \hat{\theta}_p x_i^{(p)}, \quad i = 1, 2, \dots, n.$$

Тогда

$$\begin{aligned} \text{вар}(y) &= (Y - \bar{Y})^T (Y - \bar{Y}) = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{f}_i + \hat{f}_i - \bar{y})^2 = \\ &= \sum_{i=1}^n (y_i - \hat{f}_i)^2 + \sum_{i=1}^n (\hat{f}_i - \bar{y})^2 + 2 \sum_{i=1}^n (y_i - \hat{f}_i)(\hat{f}_i - \bar{y}) = \\ &= \sum_{i=1}^n \hat{\varepsilon}_i^2 + \text{вар}(\hat{f}) + 2 \sum_{i=1}^n \hat{\varepsilon}_i(\hat{f}_i - \bar{y}), \end{aligned} \quad (4.48)$$

где $\hat{\varepsilon}_i = y_i - \hat{f}_i = y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i^{(1)} - \dots - \hat{\theta}_p x_i^{(p)}$, а $\text{вар}(\hat{f}) = \sum_{i=1}^n (\hat{f}_i - \bar{y})^2$ — *вариация выборочной функции регрессии*; она характеризует степень случайного разброса значений функции регрессии около среднего значения результирующей переменной.

Чтобы получить требуемый результат, нужно показать, что последняя сумма в выражении (4.48) равна нулю. С этой целью докажем сначала, что невязки $\hat{\varepsilon}_i$ ортогональны ко всем объясняющим переменным (включая искусственную переменную $x_i^{(0)} \equiv 1$), то есть что

$$\sum_{i=1}^n \hat{\varepsilon}_i = 0 \quad \text{и} \quad \sum_{i=1}^n \hat{\varepsilon}_i x_i^{(j)} = 0 \quad \text{для всех} \quad j = 1, 2, \dots, p. \quad (4.49)$$

Действительно:

$$\begin{aligned} \mathbf{X}^\top \hat{\boldsymbol{\varepsilon}} &= \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1^{(1)} & x_2^{(1)} & \dots & x_n^{(1)} \\ \dots & \dots & \dots & \dots \\ x_1^{(p)} & x_2^{(p)} & \dots & x_n^{(p)} \end{pmatrix} \begin{pmatrix} \hat{\varepsilon}_1 \\ \hat{\varepsilon}_2 \\ \vdots \\ \hat{\varepsilon}_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n \hat{\varepsilon}_i \\ \sum_{i=1}^n \hat{\varepsilon}_i x_i^{(1)} \\ \vdots \\ \sum_{i=1}^n \hat{\varepsilon}_i x_i^{(p)} \end{pmatrix} = \\ &= \mathbf{X}^\top (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\theta}}) = \mathbf{X}^\top \mathbf{Y} - \mathbf{X}^\top \mathbf{X} \hat{\boldsymbol{\theta}} = \\ &= \mathbf{X}^\top \mathbf{Y} - (\mathbf{X}^\top \mathbf{X})(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} = \mathbf{0}_{p+1}, \end{aligned}$$

что и означает справедливость равенств (4.49) (при выводе этого соотношения МНК-оценка $\hat{\boldsymbol{\theta}}$ была заменена ее выражением по формуле (4.22); $\mathbf{0}_{p+1}$, как и прежде, означает вектор-столбец размерности $p+1$, состоящий из одних нулей).

Теперь мы можем показать, что

$$\sum_{i=1}^n \hat{\varepsilon}_i (\hat{f}_i - \bar{y}) = 0. \quad (4.50)$$

Действительно, учитывая равенства (4.49), имеем:

$$\begin{aligned} \sum_{i=1}^n \hat{\varepsilon}_i (\hat{f}_i - \bar{y}) &= \sum_{i=1}^n \hat{\varepsilon}_i (\hat{\theta}_0 + \hat{\theta}_1 x_i^{(1)} + \dots + \hat{\theta}_p x_i^{(p)} - \bar{y}) = \\ &= \hat{\theta}_0 \sum_{i=1}^n \hat{\varepsilon}_i + \hat{\theta}_1 \sum_{i=1}^n \hat{\varepsilon}_i x_i^{(1)} + \dots \\ &\quad \dots + \hat{\theta}_p \sum_{i=1}^n \hat{\varepsilon}_i x_i^{(p)} - \bar{y} \sum_{i=1}^n \hat{\varepsilon}_i = 0. \end{aligned} \quad (4.51)$$

Возвращаясь к (4.48) и учитывая (4.51), имеем:

$$\text{вар}(y) = \text{вар}(f) + \sum_{i=1}^n \hat{\varepsilon}_i^2. \quad (4.52)$$

Поделив обе части (4.52) на $\text{вар}(y)$, получаем:

$$\frac{\text{вар}(f)}{\text{вар}(y)} = 1 - \frac{\sum \hat{\varepsilon}_i^2}{\text{вар}(y)} \quad (4.53)$$

или

$$\hat{R}_{y,X}^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i^{(1)} - \dots - \hat{\theta}_p x_i^{(p)})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}. \quad (4.54)$$

Сравнив левые и правые части соотношений (4.53) и (4.54), а последнее соотношение — с разложением (3.3) и ранее введенным коэффициентом детерминации (3.8), с учетом результата (3.32) можно сделать следующие выводы:

- характеристика, определенная соотношением (4.54), является выборочным аналогом коэффициента детерминации (3.8) и одновременно квадратом выборочного множественного коэффициента корреляции между y и X (в рамках нормальной КЛММР);

- выборочный коэффициент детерминации $\hat{R}_{y,X}^2$, определенный соотношением (4.54), учитывая соотношение (4.53), характеризует долю общей вариации результирующего признака y , объясненной поведением (вариацией) выборочной функции регрессии $\hat{f}(X) = \hat{\theta}_0 + \hat{\theta}_1 x^{(1)} + \dots + \hat{\theta}_p x^{(p)}$.

- принимая во внимание результат (3.32), для вычисления выборочного коэффициента детерминации $\hat{R}_{y,X}^2$ в рамках КЛММР можно использовать (помимо формулы (4.54)) формулы (3.28) и (3.29) с заменой участвующих в них коэффициентов парной и частной корреляции их выборочными аналогами (оценками).

- соответственно определенный формулой (4.54) выборочный коэффициент детерминации $\hat{R}_{y,X}^2$ обладает всеми описанными в п. 3.2.4 (см., в частности, соотношения (11.33)–(3.35)) статистическими свойствами квадрата множественного коэффициента корреляции.

В заключение еще раз выпишем выведенное выше соотношение (4.52), но в *матричной форме*:

$$(Y - \bar{Y})(Y - \bar{Y})^T = (X\hat{\Theta} - \bar{Y})(X\hat{\Theta} - \bar{Y})^T + \hat{\varepsilon}\hat{\varepsilon}^T. \quad (4.52')$$

Проверка гипотезы об отсутствии какой бы то ни было линейной связи между y и совокупностью объясняющих переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$

В п. 3.2.4 при описании корреляционных характеристик линейной зависимости y от нескольких объясняющих переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ был приведен результат, в соответствии с которым в предположении справедливости гипотезы

$$H_0: \theta_1 = \theta_2 = \dots = \theta_p = 0$$

статистика

$$\gamma = \frac{\hat{R}_{y,X}^2}{1 - \hat{R}_{y,X}^2} \cdot \frac{n - p - 1}{p} \quad (4.55)$$

должна подчиняться F -распределению с числами степеней свободы числителя и знаменателя, соответственно равными p и $n - p - 1$. Следовательно, гипотезу H_0 можно проверить следующим образом. По заданному уровню значимости критерия α определяем из таблиц П1.5 Приложения 1 $100\alpha\%$ -ную точку $F(p, n - p - 1)$ -распределения $v_\alpha^2(p, n - p - 1)$.

Если окажется, что

$$\frac{\hat{R}_{y.X}^2}{1 - \hat{R}_{y.X}^2} \cdot \frac{n - p - 1}{p} > v_{\alpha}^2(p, n - p - 1),$$

то гипотеза об отсутствии линейной связи между y и X отвергается (с вероятностью ошибиться, равной α), и принимается — в противном случае.

Пример 4.1 (продолжение). Отдельная проверка статистически значимого отличия от нуля оценок коэффициентов $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_5$ в модели (4.35), то есть последовательное сравнение отношений $\hat{\theta}_l/s_l$ с 2,5%-ной точкой распределения Стьюдента с $n - p - 1 = 20 - 5 - 1 = 14$ степенями свободы ($t_{0.025}(14) = 2,145$), показывает, что гипотеза о нулевом значении коэффициента регрессии оказывается принятой для всех объясняющих переменных, кроме $x^{(4)}$. Другими словами, на этой стадии анализа только переменная $x^{(4)}$ — количество удобрений, расходуемых на гектар земли, — подтвердила свое право на включение в модель. В то же время вычисление оценки $\hat{R}_{y.X}^2$ для множественного коэффициента корреляции между урожайностью зерновых культур y и совокупностью пяти факторов сельскохозяйственного производства ($\hat{R}_{y.X}^2 = 0,517$) и проверка гипотезы об отсутствии какой бы то ни были линейной связи между y и X с помощью статистики (4.55) ($\gamma = 0,517 \cdot 14 / 0,483 \cdot 5 = 7,238 / 2,415 = 3,00$; $v_{0.05}^2(5; 20) = 2,96$) говорят за то, чтобы продолжить изучение линейной связи между y и $(x^{(1)}, x^{(2)}, \dots, x^{(5)})$ в рамках КЛММР. Пытаясь понять причины невыразительности проявления влияния на y факторами $x^{(1)}, x^{(2)}, x^{(3)}$ и $x^{(5)}$ и более внимательно анализируя с этой целью как их содержательный смысл, так и подсчитанную матрицу парных корреляций $\mathbf{R}$ (табл. 4.2), мы вынуждены прийти к выводу, что в данном случае наша КЛММР «отягощена» так называемым *явлением мультиколлинеарности*. К исследованию природы и причин этого явления и описанию способов преодоления возникающих при этом трудностей мы и переходим в следующем пункте.

Таблица 4.2. Матрица $\mathbf{R}$ парных коэффициентов корреляции

	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

4.4 Мультиколлинеарность и отбор наиболее существенных объясняющих переменных в КЛММР

Речь идет о мультиколлинеарности (то есть *совместной, или множественной, взаимозависимости*) объясняющих переменных модели $x^{(1)}, x^{(2)}, \dots, x^{(p)}$. Можно ли строго определить мультиколлинеарность, каковы внешние признаки ее присутствия, какие трудности она создает при анализе линейной модели регрессии и как преодолевать эти трудности? Попробуем ответить на эти вопросы.

4.4.1 Признаки и причины мультиколлинеарности

Мультиколлинеарность в строгом смысле (или полная мультиколлинеарность) определяется условием нарушения одного из требований КЛММР, а именно, требования к рангу матрицы $\mathbf{X}$ (см. (4.1) или (4.1')): говорят, что объясняющие переменные модели $(x^{(1)}, x^{(2)}, \dots, x^{(p)})$ характеризуются свойством полной мультиколлинеарности, если ранг матрицы их наблюдаемых значений (4.1а) меньше $p + 1$. Мы уже обсуждали эту ситуацию в начале главы, комментируя смысл условия «ранг матрицы $\mathbf{X} = p + 1$ ». Напомним, что при нарушении этого условия между анализируемыми объясняющими переменными существует линейная *функциональная* связь (то есть значения по меньшей мере одной из них могут быть выражены в виде линейной комбинации наблюдаемых значений остальных переменных), а матрица $\mathbf{X}^T \mathbf{X}$ *оказывается вырожденной*, то есть ее определитель равен нулю (а значит не существует обратная матрица $(\mathbf{X}^T \mathbf{X})^{-1}$, участвующая в основных соотношениях метода наименьших квадратов). В практике статистических исследований полная мультиколлинеарность встречается достаточно редко, так как ее несложно избежать уже на предварительной стадии анализа и отбора множества объясняющих переменных.

Реальная (или частичная) мультиколлинеарность возникает в случаях существования достаточно тесных линейных *статистических* связей между объясняющими переменными. Точных количественных критериев для определения наличия/отсутствия реальной мультиколлинеарности не существует. Тем не менее, существуют некоторые эвристические рекомендации по выявлению мультиколлинеарности.

1) В первую очередь анализируют матрицу $\mathbf{R}$ парных коэффициентов корреляции, точнее, ту ее часть, которая относится к объясняющим переменным. Считается, что наличие значений коэффициентов корреляции, по абсолютной величине превосходящих 0,75–0,80, свидетельствует о присутствии мультиколлинеарности. Отметим, что анализ матрицы $\mathbf{R}$ в примере 4.1 (см. табл. 4.2) выявляет три коэффициента:

$r(x^{(1)}, x^{(3)}) = 0,98$; $r(x^{(2)}, x^{(3)}) = 0,88$ и $r(x^{(1)}, x^{(2)}) = 0,85$. Следует признать, что тесную связь по меньшей мере в одной из этих пар можно было легко спрогнозировать и до статистического обследования регионов: орудия поверхностной обработки почвы реализуются в подавляющем большинстве с помощью тракторов, а потому $x^{(3)}$ обязано быть достаточно высоко коррелированным с $x^{(1)}$.

2) Существование тесных линейных статистических связей между объясняющими переменными приводит к так пазываемой *слабой обусловленности* матрицы $X^T X$, то есть к близости к нулю ее определителя. Поэтому, если значение $\det(X^T X)$ оказывается близким к нулю (скажем, одного порядка с накапливающимися ошибками вычислений), то это тоже свидетельствует о наличии мультиколлинеарности.

3) Важную роль в анализе мультиколлинеарности играет и *минимальное собственное число* $\lambda_{\min}$ матрицы $X^T X$. Это объясняется двумя обстоятельствами. Во-первых, из близости к нулю $\lambda_{\min}$ следует близость к нулю величины $\det(X^T X)$, и наоборот. Во-вторых, можно показать, что среднеквадратическая ошибка оценки $\hat{\theta}_{j, \text{мнк}}$ обратно пропорциональна величине $\lambda_{\min}$ (см., например, [Джонстон (1980), п. 5.7]). Поэтому, наряду с величиной $\det(X^T X)$ (или вместо нее), вычисляют и сравнивают с накапливающимися ошибками от округлений значение $\lambda_{\min}$, то есть минимальный корень уравнения

$$|X^T X - \lambda I_{p+1}| = 0.$$

4) Анализ корреляционной матрицы R позволяет лишь в первом приближении (и относительно поверхностно) судить о наличии/отсутствии мультиколлинеарности в наших исходных данных (4.1а). Более внимательное изучение этого вопроса достигается с помощью расчета значений коэффициентов детерминации $\hat{R}_{x^{(j)}.X(j)}^2$ каждой из объясняющих переменных $x^{(j)}$ по всем остальным предикторам $X(j) = (x^{(1)}, \dots, x^{(j-1)}, x^{(j+1)}, \dots, x^{(p)})^T$. Это объясняется, в частности, тем, что среднеквадратические ошибки s_j оценок $\hat{\theta}_{j, \text{мнк}}$ связаны с величиной $R_{x^{(j)}.X(j)}^2$ соотношением $s_j^2 = \sigma^2/n(1 - R_{x^{(j)}.X(j)}^2)$.

5) Наконец, о присутствии явления мультиколлинеарности сигнализируют некоторые внешние признаки построенной модели, являющиеся его *следствиями*. К ним в первую очередь следует отнести такие:

- некоторые из оценок $\hat{\theta}_{j, \text{мнк}}$ имеют неправильные с точки зрения экономической теории знаки или неоправданно большие по абсолютной величине значения (см. пример в п. 2.2, соотношения (2.9)–(2.9')–(2.10));
- небольшое изменение исходных статистических данных (4.1а)–(4.3) (добавление или изъятие небольшой порции наблюде-

ний) приводит к существенному изменению оценок коэффициентов модели, вплоть до изменения их знаков;

- большинство или даже все оценки коэффициентов регрессии оказываются статистически незначимо отличающимися от нуля (при последовательной автономной проверке с помощью t -критерия (4.45)), в то время как в действительности многие из них имеют отличные от нуля значения, а модель в целом является значимой при проверке с помощью статистики (4.55) (именно с такой ситуацией мы столкнулись в примере 4.1).

Подобные не поддающиеся содержательной интерпретации особенности модели можно понять и предвидеть, если обратиться к основным соотношениям (4.22) и (4.33), определяющим правила вычисления соответственно МНК-оценок $\hat{\Theta}_{\text{МНК}}$ и их ковариационной матрицы. И в той и в другой формуле присутствует матрица $(X^T X)^{-1}$, элементы которой обратно пропорциональны величине определителя $\det(X^T X)$. Если эта величина достаточно мала, то изъятие или дополнение одной-двух строк в матрице X (что равносильно изъятию или дополнению одного-двух наблюдений в массив исходных данных (4.1a)–(4.3)) может радикально (во много раз) изменить значение $\det(X^T X)$, а следовательно, и зависящие от него значения $\hat{\Theta}_{\text{МНК}}$ и $\Sigma_{\hat{\Theta}}$. Одновременно малость величины $\det(X^T X)$ влечет за собой непомерно большие значения диагональных элементов матрицы $\Sigma_{\hat{\Theta}}$ (то есть дисперсий $s_j^2 = D(\hat{\theta}_j - \theta_j)^2$), что может приводить к статистически незначимым отклонениям от нуля критических значений $|\hat{\theta}_{j, \text{МНК}}|/s_j$.

4.4.2 Методы устранения мультиколлинеарности

I. Переход к смещенным методам оценивания. Как известно (см. в п. 4.2.3 формулировку теоремы Гаусса–Маркова, соотношение (4.37)), МНК-оценки являются оценками с минимальной среднеквадратической ошибкой в классе всех *несмещенных* и линейных относительно $Y = (y_1, y_2, \dots, y_n)^T$ оценок. Зададимся вопросом: может ли существовать *смещенная* оценка $\hat{\theta}_{\text{см}}$ неизвестного параметра θ , более точная (в смысле среднего квадрата ошибки $E(\hat{\theta} - \theta)^2$), чем наилучшая среди *несмещенных* оценок? Оказывается, да, может! Поясним это на следующей схеме (рис. 4.1). Пусть оценка $\hat{\theta}_{\text{МНК}}$ — наилучшая среди *несмещенных* оценок неизвестного значения θ анализируемого параметра, то есть $E(\hat{\theta}_{\text{МНК}} - \theta)^2 \leq E(\hat{\theta} - \theta)^2$ в классе всех несмещенных оценок $\hat{\theta}$. Повторяя выборки одного и того же объема n из анализируемой генеральной совокупности и подсчитывая значения $\hat{\theta}_{\text{МНК}}^{(l)}$ для каждой (l -й) выборки по одной и той же формуле построения наилучшей несмещенной оценки (при $l = 1, 2, \dots$), мы будем иметь *множество* значений $\hat{\theta}_{\text{МНК}}$, по которым можно оценить, например, плотность распределения оценки

$\hat{\theta}_{\text{МНК}}$ — функцию $f_{\hat{\theta}_{\text{МНК}}}(x)$. Поступая аналогичным образом с конкурирующей смещенной оценкой $\hat{\theta}_{\text{см}}$, мы получим плотность ее распределения $f_{\hat{\theta}_{\text{см}}}(x)$. Пусть далее Δ определяет допустимый предел погрешности в оценивании истинного значения θ анализируемого параметра, то есть если $|\hat{\theta} - \theta| \leq \Delta$, то оценка $\hat{\theta}$ считается «хорошей», а при $|\hat{\theta} - \theta| > \Delta$ — «плохой».

Простой визуальный анализ рис. 4.1 приводит нас к следующим очевидным выводам:

- доля «плохих» оценок $\hat{\theta}_{\text{см}}$ (а она определяется, в соответствии с вероятностным смыслом кривой плотности $f_{\hat{\theta}_{\text{см}}}(x)$, величиной заштрихованной площади под кривой плотности $f_{\hat{\theta}_{\text{см}}}(x)$ вне интервала $[\theta - \Delta, \theta + \Delta]$) в несколько раз меньше доли «плохих» оценок $\hat{\theta}_{\text{МНК}}$ (последняя аналогично определяется заштрихованной площадью под кривой плотности $f_{\hat{\theta}_{\text{МНК}}}(x)$ вне интервала $[\theta - \Delta, \theta + \Delta]$);
- средний квадрат ошибок при оценивании методом $\hat{\theta}_{\text{МНК}}$ (как результат интегрирования величин $(\hat{\theta}_{\text{МНК}} - \theta)^2$ с весами, определяемыми функцией плотности $f_{\hat{\theta}_{\text{МНК}}}(x)$, то есть $E(\hat{\theta}_{\text{МНК}} - \theta)^2 = \int_{-\infty}^{\infty} (x - \theta)^2 f_{\hat{\theta}_{\text{МНК}}}(x) dx$) будет превосходить средний квадрат ошибок, получаемых при оценивании с помощью смещенной оценки $\hat{\theta}_{\text{см}}$ (то есть величину $E(\hat{\theta}_{\text{см}} - \theta)^2 = \int_{-\infty}^{\infty} (x - \theta)^2 f_{\hat{\theta}_{\text{см}}}(x) dx$).

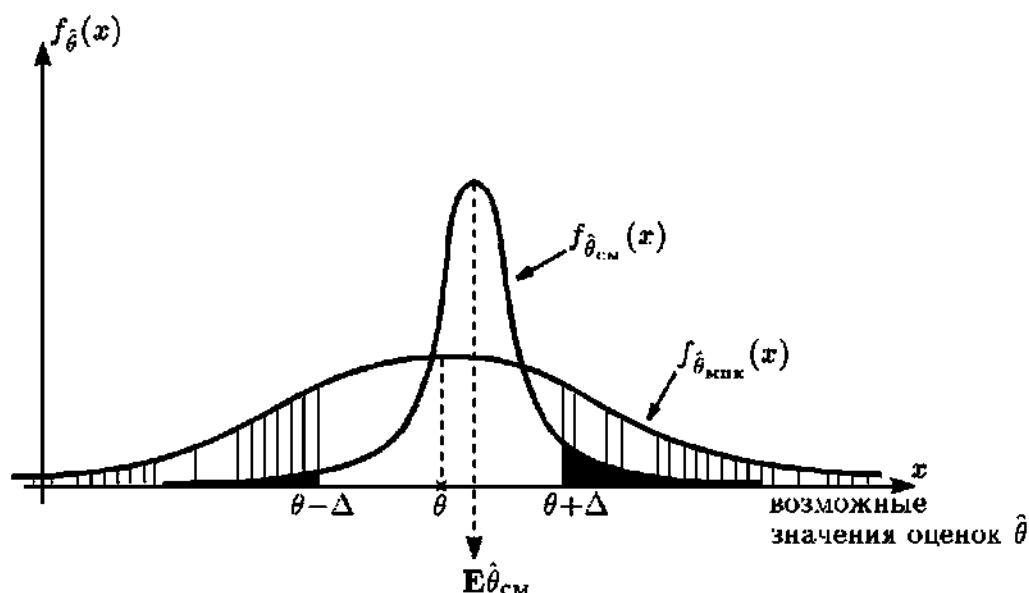


Рис. 4.1. Плотности распределения несмещенной ($f_{\hat{\theta}_{\text{МНК}}}(x)$) и смещенной ($f_{\hat{\theta}_{\text{см}}}(x)$) оценок истинного значения θ неизвестного параметра

Таким образом, учитывая, что в условиях мультиколлинеарности дисперсии даже наилучших несмещенных оценок могут быть слишком

большими, естественно попытаться отказаться от требования несмещенности, чтобы в более широком классе оценок найти те, которые будут обладать более высокой точностью.

Описание различных подходов к построению «хороших» *смещенных* оценок коэффициентов регрессии в условиях мультиколлинеарности можно найти в книге [Айвазян и др., (1985), пп. 8.3–8.5]. Приведем здесь лишь один из этих подходов, названный «*ридж-регрессией*» (или «*гребневой регрессией*»). Он основан на рассмотрении однопараметрического семейства несколько «подправленных» МНК-оценок, а именно оценок вида

$$\hat{\Theta}_\tau = (X^T X + \tau I_{p+1})^{-1} X^T Y. \quad (4.56)$$

Добавление к диагональным элементам матрицы $(X^T X)$ «гребня», «хребта» τ (τ — некоторое небольшое положительное число) с одной стороны, делает получаемые при этом оценки *смещенными*, а с другой, — превращает матрицу $X^T X$ из «плохо обусловленной» в «хорошо обусловленную». Соответственно в дальнейшем и, в частности, при вычислении средних квадратов ошибок для оценок $\hat{\Theta}_\tau$ мы не столкнемся с чрезмерно малыми значениями определителя матрицы $X^T X$ (теперь это будет уже определитель матрицы $X^T X + \tau I_{p+1}$) и связанными с этим неприятностями. Целесообразность использования оценок вида (4.56) основана на теореме (см. Hoerl A. E., Kennard R. W., Technometrics, 1970, vol. 12, №1, pp. 55–67), утверждающей, что в условиях мультиколлинеарности найдется такое значение τ_0 , при котором средние квадраты ошибок оценок $\hat{\Theta}_{\tau_0}$ окажутся меньше соответствующих характеристик для МНК-оценок $\hat{\Theta}_{\text{МНК}}$. Универсальных рекомендаций по поводу выбора конкретного значения τ нет (как правило, эта величина определяется в диапазоне значений от 0,1 до 0,4).

II. Переход к ортогонализированным объясняющим переменным с помощью метода главных компонент. Поскольку мультиколлинеарность связана с высокой степенью корреляции между объясняющими переменными, можно попытаться обойти эту трудность, используя в качестве новых объясняющих переменных некоторые линейные комбинации исходных, выбранные так, чтобы корреляции между вновь введенными переменными были малы или вообще отсутствовали. Если объясняющих переменных немного (как, скажем, в наших примерах 3.3 и 3.4) или имеется некоторая априорная информация, то выбор таких комбинаций может быть осуществлен из содержательных соображений. В более общей ситуации один из возможных подходов основывается на использовании *главных компонент* вектора исходных объясняющих переменных $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$ (см. Приложение П3.4). Соответствующий прием известен в литературе под названием «*построение регрессии на главных компонентах*». Ссылка на Приложение П3.4 требует пояснений. В Приложении П3.4 описывается метод построения глав-

ных компонент *многомерной случайной величины* $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^\top$, в то время как в нашей модели X — это неслучайный векторный параметр, от которого зависит закон распределения результирующего признака y . Поэтому, оставляя в силе все описанные в ПЗ.4 построения и результаты, мы должны сделать, тем не менее, ряд пояснений:

- Вероятностная терминология применительно к переменным $x^{(j)}$ (их математические ожидания, дисперсии, ковариации и т.п.) теряет смысл, а вместо их *вероятностных* характеристик мы будем рассматривать формально соответствующие им комбинации наблюдаемых значений. В частности, мы так же, как и в ПЗ.4, перейдем к *центрированным* наблюдениям $X_{\text{ц}i} = (x_i^{(1)} - \bar{x}^{(1)}, x_i^{(2)} - \bar{x}^{(2)}, \dots, x_i^{(p)} - \bar{x}^{(p)})^\top$, где $\bar{x}^{(j)} = \sum_{i=1}^n x_i^{(j)} / n$, а вместо теоретической ковариационной матрицы $\Sigma = \mathbf{E}[(X - \mathbf{E}X)(X - \mathbf{E}X)^\top]$, относительно которой решается задача определения собственных чисел и собственных векторов, будем рассматривать $(p \times p)$ -матрицу $\mathbf{X}_{\text{ц}}^\top \mathbf{X}_{\text{ц}}$, где $\mathbf{X}_{\text{ц}}$ — это $(n \times p)$ -матрица *центрированных* наблюдений, то есть

$$\mathbf{X}_{\text{ц}} = \begin{pmatrix} x_1^{(1)} - \bar{x}^{(1)} & x_1^{(2)} - \bar{x}^{(2)} & \dots & x_1^{(p)} - \bar{x}^{(p)} \\ x_2^{(1)} - \bar{x}^{(1)} & x_2^{(2)} - \bar{x}^{(2)} & \dots & x_2^{(p)} - \bar{x}^{(p)} \\ \dots & \dots & \dots & \dots \\ x_n^{(1)} - \bar{x}^{(1)} & x_n^{(2)} - \bar{x}^{(2)} & \dots & x_n^{(p)} - \bar{x}^{(p)} \end{pmatrix}$$

(заметим, что матрица $\mathbf{X}_{\text{ц}}^\top \mathbf{X}_{\text{ц}}$, поделенная на n , при случайных переменных X интерпретируется как выборочная ковариационная матрица вектора X).

- Уравнение регрессии Y по X в терминах центрированных переменных $Y_{\text{ц}} = Y - \bar{Y}$ и $X_{\text{ц}} = X - \bar{X}$ примет вид:

$$\mathbf{E}(Y - \bar{Y} \mid X - \bar{X}) = \theta_1(x^{(1)} - \bar{x}^{(1)}) + \theta_2(x^{(2)} - \bar{x}^{(2)}) + \dots + \theta_p(x^{(p)} - \bar{x}^{(p)}), \quad (4.57)$$

так что свободный член θ_0 из исходного уравнения (4.1) может быть выражен в виде

$$\theta_0 = \bar{y} - \sum_{j=1}^p \theta_j \bar{x}^{(j)}, \quad (4.58)$$

где $\bar{y} = \sum_{i=1}^n y_i / n$, а $\bar{Y} = (\bar{y}, \bar{y}, \dots, \bar{y})^\top$ — n -мерный вектор-столбец.

С учетом сделанных замечаний реализация метода построения регрессии Y на главные компоненты вектора X предусматривает выполнение следующих операций.

1) Определяются и упорядочиваются собственные числа λ_j и соответствующие им собственные векторы $l_j = (l_{j1}, l_{j2}, \dots, l_{jp})$ матрицы $X_{\Pi}^T X_{\Pi}$ ($\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$, см. ПЗ.4).

2) Составляется матрица коэффициентов преобразования

$$L = \begin{pmatrix} l_{11} & l_{12} & \dots & l_{1p} \\ l_{21} & l_{22} & \dots & l_{2p} \\ \dots & \dots & \dots & \dots \\ l_{p1} & l_{p2} & \dots & l_{pp} \end{pmatrix}, \quad (4.59)$$

в которой j -й собственный вектор l_j играет роль j -й строки. Отметим, что из построения следует, что матрица L ортогональна, то есть $L^T = L^{-1}$ и соответственно $L^T L = LL^T = I_p$.

3) С помощью матрицы L переходят к вектору главных компонент

$$Z = (z^{(1)}, z^{(2)}, \dots, z^{(p)})^T = LX_{\Pi}. \quad (4.60)$$

Соответственно i -е наблюдение вектора главных компонент определится соотношением

$$Z_i = (z_i^{(1)}, z_i^{(2)}, \dots, z_i^{(p)})^T = LX_{\Pi i},$$

а матрица наблюдений главных компонент — соотношением

$$Z = X_{\Pi} L^T. \quad (4.61)$$

Отметим, что из построения следует:

$$Z^T Z = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_p \end{pmatrix} \quad (4.62)$$

и соответственно

$$(Z^T Z)^{-1} = \begin{pmatrix} 1/\lambda_1 & & & 0 \\ & 1/\lambda_2 & & \\ & & \ddots & \\ 0 & & & 1/\lambda_p \end{pmatrix}. \quad (4.63)$$

Отметим также, что вектор центрированных исходных переменных X_{Π} может быть выражен из (4.60):

$$X_{\Pi} = X - \bar{X} = L^{-1} Z = L^T Z. \quad (4.64)$$

4) Возвращаясь к задаче регрессии, анализируем КЛММР Y_{Π} по Z :

$$E(Y_{\Pi} | Z) = c_1 z^{(1)} + c_2 z^{(2)} + \dots + c_p z^{(p)}. \quad (4.65)$$

В соответствии с основными соотношениями метода наименьших квадратов (4.22) и (4.33) получаем:

$$\begin{aligned}\hat{C}_{\text{мнк}} &= (\hat{c}_{1\text{ мнк}}, \hat{c}_{2\text{ мнк}}, \dots, \hat{c}_{p\text{ мнк}})^T = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T Y_{\text{ц}}, \\ \Sigma_{\hat{C}_{\text{мнк}}} &= \hat{\sigma}_{\text{гк}}^2 (\mathbf{Z}^T \mathbf{Z})^{-1},\end{aligned}$$

где

$$\hat{\sigma}_{\text{гк}}^2 = \frac{1}{n-p} \sum_{i=1}^n (y_{\text{ц}i} - \hat{c}_{1\text{ мнк}} z_i^{(1)} - \dots - \hat{c}_{p\text{ мнк}} z_i^{(p)})^2.$$

С учетом (4.63) можно сделать вывод, что *оценки $\hat{c}_{1\text{ мнк}}, \hat{c}_{2\text{ мнк}}, \dots, \hat{c}_{p\text{ мнк}}$ взаимно некоррелированы и*

$$\hat{c}_{j\text{ мнк}} = \frac{1}{\lambda_j} \sum_{i=1}^n z_i^{(j)} y_{\text{ц}i}, \quad (4.66)$$

$$\mathbf{E}(\hat{c}_{j\text{ мнк}} - c_j)^2 = \frac{\hat{\sigma}_{\text{гк}}^2}{\lambda_j}. \quad (4.67)$$

5) Наконец, пользуясь формулами (4.66) и (4.67), последовательно проверяем гипотезы (см. п. 4.2.3, (4.46))

$$H_{0j}: c_j = 0 \quad (j = 1, 2, \dots, p)$$

с помощью статистик

$$\gamma_j = \frac{|\frac{1}{\lambda_j} \sum_{i=1}^n z_i^{(j)} y_{\text{ц}i}|}{\hat{\sigma}_{\text{гк}} / \sqrt{\lambda_j}} = \frac{|\sum_{i=1}^n z_i^{(j)} y_{\text{ц}i}|}{\sqrt{\lambda_j} \hat{\sigma}_{\text{гк}}}.$$

Обозначим $J_{\text{сущ}}$ множество индексов тех главных компонент, для которых гипотеза H_{0j} оказалась отвергнутой. Тогда оценка функции регрессии y на главные компоненты может быть записана в виде

$$y - \bar{y} = \sum_{j \in J_{\text{сущ}}} \hat{c}_j z^{(j)}. \quad (4.68)$$

В данном случае исключение из модели ряда объясняющих переменных не приводит к необходимости пересчета оценок коэффициентов регрессии для оставшихся главных компонент, так как вид матрицы $(\mathbf{Z}^T \mathbf{Z})^{-1}$ (см. (4.63)) обеспечивает *независимость результатов вычисления оценок $\hat{c}_j$ от числа и состава включенных в модель главных компонент.*

Если удастся дать содержательную интерпретацию включенным в модель главным компонентам, то представление оценки функции регрессии в виде (4.68) может быть окончательным. В противном случае возвращаются к исходным переменным. Оценки $\hat{\theta}_{0\text{гк}}, \hat{\theta}_{1\text{гк}}, \dots, \hat{\theta}_{p\text{гк}}$ для

параметров $\theta_0, \theta_1, \dots, \theta_p$ исходной регрессионной модели (4.1) получают с учетом соотношений (4.57), (4.58), (4.60) и (4.68):

$$\hat{\theta}_{q\text{ГК}} = \sum_{j \in J_{\text{сущ}}} \hat{c}_{jljq}, \quad q = 1, 2, \dots, p,$$

$$\hat{\theta}_{0\text{ГК}} = \bar{y} - \sum_{q=1}^p \hat{\theta}_{q\text{ГК}} \bar{x}^{(q)}.$$

Вообще говоря, полученные таким образом оценки для коэффициентов регрессии θ_q ($q = 0, 1, \dots, p$) будут *смещенными*. Формулы для величин смещений и средних квадратов ошибок этих оценок приведены в книге [Айвазян и др., (1985), п. 8.5].

III. Отбор наиболее существенных объясняющих переменных в КЛММР. Среди основных предпосылок, обуславливающих возможность перехода от исходного числа p анализируемых показателей $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ к существенно меньшему числу p' наиболее информативных (в определенном смысле) переменных, упоминается обычно *дублирование информации, доставляемой сильно взаимозависимыми признаками*. Именно с этой ситуацией мы сталкиваемся при мультиколлинеарности объясняющих переменных в модели регрессии, а потому стремление исследователя *попытаться уменьшить общее число предикторов*, отобрать из априорного набора объясняющих переменных лишь самые существенные (с точки зрения влияния на y) представляется вполне естественной реакцией на данную ситуацию.

Каким бы способом ни проводился отбор объясняющих переменных, обусловленность матрицы $X^T X$ улучшается с уменьшением числа предикторов, включенных в модель. В действительности, *процедура отбора существенных переменных имеет самостоятельное (не «привязанное» к проблеме мультиколлинеарности) значение*. Она может рассматриваться как процедура выбора размерности линейной модели. Другими словами, она полезна и должна применяться и в ситуациях, когда матрица $X^T X$ хорошо обусловлена. Но особенно она необходима и эффективна в условиях мультиколлинеарности. Так, если две какие-либо объясняющие переменные сильно коррелированы с результирующим показателем y и друг с другом, то часто бывает достаточно включения в модель одной из них, а дополнительным вкладом от включения другой можно пренебречь.

В предположении, что матрица данных X является неслучайной, возможны две точки зрения на оценку уравнения регрессии, полученную после отбора существенных предсказывающих переменных.

Первая точка зрения исходит из того, что модель регрессии (4.1) является истинной, и несмещенная оценка коэффициентов регрессии получается с помощью метода наименьших квадратов, то есть путем решения системы уравнений (4.18) (в условиях мультиколлинеарности эта

оценка может быть неудовлетворительной, но тем не менее несмещенной). Тогда принудительное приравнивание части коэффициентов регрессионного уравнения к 0, что и происходит при отборе переменных, естественно, приводит, если матрица $X^T X$ недиагональна, к *смещенным* оценкам коэффициентов при оставшихся переменных, то есть мы приходим к классу смещенных оценок, рассмотренных в п. 4.4.2. Вопросы анализа возникающих при этом «*ошибок спецификации*» будут рассмотрены в следующем пункте.

С другой стороны, процесс отбора существенных переменных можно рассматривать как процесс *выбора истинной модели* из множества возможных линейных моделей, которые могут быть построены с помощью набора предсказывающих переменных, и тогда полученные после отбора оценки коэффициентов можно рассматривать как несмещенные. Этой точки зрения мы будем придерживаться в данном пункте.

Для случая, когда переменные $x^{(1)}, \dots, x^{(p)}, y$ — случайные величины, вопрос о правильности (истинности) модели не возникает. Все, что мы ищем в этом случае, — это модель, сохраняющую ошибку предсказания на разумном уровне, при ограниченном количестве переменных.

Существует несколько подходов к решению задачи отбора наиболее существенных предикторов в модели регрессии. Мы остановимся здесь на одном из них, с нашей точки зрения — наиболее распространенном и эффективном. Речь идет о процедуре *последовательного наращивания числа предикторов*, реализуемой в двух версиях: *версии «всех возможных регрессий»* и *версии «пошагового отбора переменных»*. Этот подход соответствует общей логике, которой следует придерживаться при построении и анализе регрессионных моделей, а именно — реализации пути «от простого к сложному».

Версия «всех возможных регрессий». Решается следующая задача: для заданного значения k ($k = 1, 2, \dots, p-1$) путем *полного перебора* всех возможных комбинаций из k объясняющих переменных, отобранных из исходного (априорного) набора $x^{(1)}, x^{(2)}, \dots, x^{(p)}$, состоящего из p предикторов, определить такие переменные $x^{(i_1^0(k))}, x^{(i_2^0(k))}, \dots, x^{(i_k^0(k))}$, для которых коэффициент детерминации с результирующим показателем y был бы максимальным, то есть

$$\hat{R}_{y, (x^{(i_1^0(k))}, \dots, x^{(i_k^0(k))})}^2 = \max_{\substack{1 \leq i_1, i_2, \dots, i_k \leq p \\ i_{q_1} \neq i_{q_2} \text{ при } q_1 \neq q_2}} \hat{R}_{y, (x^{(i_1)}, \dots, x^{(i_k)})}^2. \quad (4.69)$$

Таким образом, на *первом шаге* процедуры ($k = 1$) мы найдем *одну объясняющую переменную* $x^{(i_1^0(1))}$, которую можно назвать наиболее информативным предиктором при условии, что в регрессионную модель y по X мы можем включить только одну из априорного набора объясняющих переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$. На *втором шаге* реализация критерия (4.69) определит уже *наиболее информативную пару переменных*

$(x^{(i_1(2))}, x^{(i_2(2))})$, поскольку по построению среди всех возможных пар, отобранных из исходных p переменных, эта пара будет иметь наиболее тесную статистическую связь с результирующим показателем y . Отметим, что, вообще говоря, в состав этой пары может и не войти переменная $x^{(i_1(1))}$, объявленная наиболее информативной в классе моделей зависимости y от одной объясняющей переменной. На третьем шаге ($k = 3$) будет отобрана наиболее информативная тройка предикторов, на четвертом ($k = 4$) — наиболее информативная четверка объясняющих переменных и т. д. Строгих математических правил «останова» этой процедуры, то есть выбора оптимального числа k_0 предикторов, которые следует включить в модель, нет. Однако следующая приближенная рекомендация в большинстве случаев позволяет получить правильное решение этого вопроса.

Обратимся к рис. 4.2. Будем откладывать по оси абсцисс последовательно увеличивающееся на единицу число k предикторов в модели, а по оси ординат — соответствующие значения «подправленной на несмещенность» оценки $\hat{R}_{y, (x^{(i_1(k))}, \dots, x^{(i_k(k))})}^{*2}$ (см. формулу (3.35)). Одновременно нанесем на чертеж по оси ординат величину нижней доверительной границы $\hat{R}_{\min}^2(k)$ для $R_{y, (x^{(i_1(k))}, \dots, x^{(i_k(k))})}^2$, вычисленную по формуле

$$R_{\min}^2 = \hat{R}^{*2}(k) - 2\sqrt{\frac{2k(n-k-1)}{(n-1)(n^2-1)}}(1 - \hat{R}^2(k)). \quad (4.70)$$

В соотношении (4.70) для большей краткости в обозначениях положено $\hat{R}^{*2}(k) = \hat{R}_{y, (x^{(i_1(k))}, \dots, x^{(i_k(k))})}^{*2}$, $\hat{R}^2(k) = \hat{R}_{y, (x^{(i_1(k))}, \dots, x^{(i_k(k))})}^2$, а вычитаемое в правой части приближенно равно удвоенной среднеквадратической ошибке оценки $\hat{R}^2(k)$ (см. формулу (3.35)). Предлагается выбирать в качестве оптимального числа k_0 предикторов регрессионной модели значение k , при котором величина $R_{\min}^2(k)$ достигает своего максимума (см. рис. 4.2).

Реализация описанного метода всех возможных регрессий, предназначенного для отбора наиболее существенных объясняющих переменных регрессионной модели, требует больших объемов вычислений. Нетрудно подсчитать, что при каждом фиксированном значении k число различных наборов переменных будет равно числу сочетаний из p элементов по k (C_p^k), а полное число анализируемых наборов переменных при изменении k от 1 до $p-1$ будет равно $C_p^1 + C_p^2 + \dots + C_p^{p-1} = 2^p - 2$ (так, при $p = 20$ число всех возможных переборov составит величину, превосходящую 10^6). В практике используется два подхода к преодолению этих вычислительных трудностей. Первый связан с использованием приближенных (полуэвристических) методов оптимизации критерия

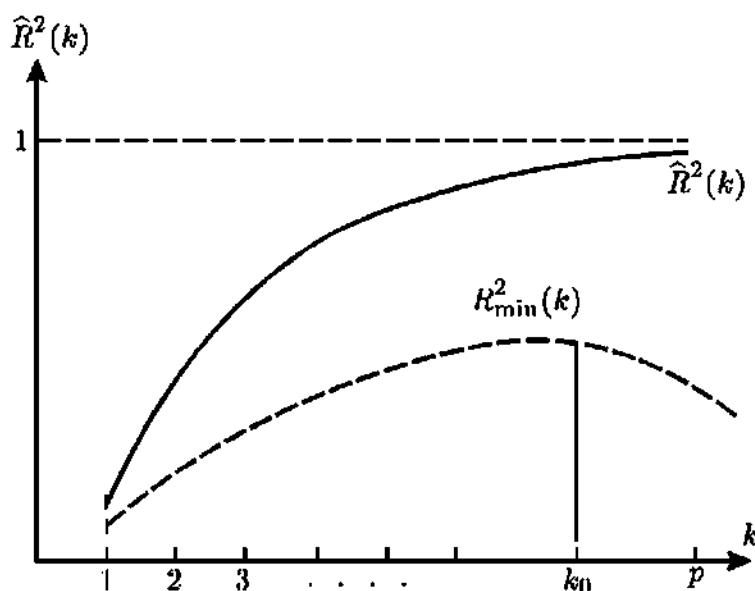


Рис. 4.2. Зависимость нижней доверительной границы коэффициента детерминации от числа предикторов (пунктирная кривая)

(4.69) (пожалуй, наиболее распространенным способом реализации этого подхода можно признать так называемый «метод ветвей и границ»; смысл этого метода состоит во введении некоторого грубого правила, которое позволяет отбросить большинство наборов объясняющих переменных, не вычисляя для них значения критерия в силу их очевидной бесперспективности). Другой подход основан на использовании так называемой пошаговой процедуры отбора, к описанию которой мы и переходим.

Версия пошагового отбора переменных. Эта версия отличается от предыдущей только тем, что на каждом следующем шаге (то есть при переходе от отбора k переменных к отбору $k + 1$ переменной) учитываются результаты предыдущего шага. Так, при переходе от $k = 1$ к $k = 2$ перебираются не все возможные пары $(x^{(i_1)}, x^{(i_2)})$, а лишь те, в которых непременно участвует переменная $x^{(i_1(1))}$, отобранная на 1-м шаге. Соответственно при переходе от шага « k » к шагу « $k + 1$ » первые k информативных переменных считаются уже определенными на предыдущем шаге, так что при оптимизации критерия (4.69) остается перебрать $p - k$ оставшихся переменных, поочередно присоединяя каждую из них к уже отобранной на предыдущем шаге группе переменных

$$x^{(i_1(1))}, x^{(i_2(2))}, \dots, x^{(i_k(k))}.$$

Все остальные характеристики процедуры (оптимизируемый критерий, правило выбора оптимального числа предикторов и т. д.) остаются теми же, что и в методе всех возможных регрессий. Нетрудно подсчитать, что при пошаговой реализации процедуры отбора наиболее существенных объясняющих переменных число необходимых переборов снижается с $2^p - 2$ до $p + (p - 1) + (p - 2) + \dots + 2 = (p + 2)(p - 1)/2$

(то есть при $p = 20$ вместо 1048574 переборov, необходимых в методе всех возможных регрессий, нам понадобится всего 209 переборov различных вариантов составов предикторов).

Следует, правда, признать, что пошаговые процедуры, вообще говоря, *не гарантируют* получения оптимального (в смысле критерия (4.69)) набора объясняющих переменных при каждом фиксированном $k \geq 2$. Однако в подавляющем большинстве ситуаций получаемые с помощью пошаговой процедуры наборы переменных оказываются оптимальными или близкими к оптимальным.

* * *

В пакетах программ по эконометрике и статистическому анализу данных реализованы различные варианты пошаговых процедур отбора существенных предикторов: наряду с описанной выше процедурой последовательного наращивания («присоединения») переменных имеются процедуры «присоединения-удаления», «удаления» и др. Их описание содержится в книге [Айвазян и др., 1985, п. 8.7].

Пример 4.1 (окончание). Напомним, что анализ матрицы парных корреляций (см. табл. 4.2) и проверка гипотез о статистически незначимом отличии от нуля коэффициентов регрессии y по $x^{(1)}, x^{(2)}, \dots, x^{(5)}$ в сочетании со знанием оценки коэффициента детерминации $\hat{R}_{y,x}^2$ урожайности y по всем пяти анализируемым факторам сельскохозяйственного производства ($\hat{R}_{y,x}^2 = 0,517$) привели нас к выводу о наличии мультиколлинеарности в рассматриваемой модели. Чтобы избавиться от нее, попробуем снизить размерность модели, то есть исключить из модели «несущественные» объясняющие переменные. Воспользуемся для этого пошаговой процедурой последовательного наращивания (присоединения) предикторов.

1-й шаг ($k = 1$). В классе моделей регрессии y по *единственной* объясняющей переменной выбирается наиболее информативный (в смысле критерия (4.69)) предиктор. Поскольку при $k = 1$ величина $R_{y,x}^2$ совпадает с квадратом обычного (парного) коэффициента корреляции $r(y, x)$, а $\max_{1 \leq j \leq 5} r^2(y, x^{(j)}) = r^2(y, x^{(4)}) = (0,58)^2 = 0,336$, то наиболее информативным предиктором в классе однофакторных (парных) регрессионных моделей оказывается переменная $x^{(4)}$ — количество удобрений, расходуемых на гектар площади (т/га). Подсчет подправленного (на несмещенность) значения $r^{*2}(y, x^{(4)}) = \hat{R}^{*2}(1)$ и его нижней доверительной границы $R_{\min}^{(1)}$ соответственно по формулам (3.35) и (4.70) (при $k = 1$) дает:

$$\hat{R}^{*2}(1) = 0,299 \quad \text{и} \quad R_{\min}^2(1) = 0,207.$$

2-й шаг ($k = 2$). Среди всевозможных *пар* объясняющих переменных $(x^{(4)}, x^{(j)})$, $j = 1, 2, 3, 5$, выбирается наиболее информативная (в смысле

критерия (4.69)) пара предикторов. Поскольку

$$\max_{\substack{1 \leq j \leq 5 \\ (j \neq 4)}} \hat{R}_{y.(x^{(4)}, x^{(j)})}^2 = \hat{R}_{y.(x^{(4)}, x^{(3)})}^2 = 0,483,$$

то наиболее информативной парой предикторов оказываются объясняющие переменные

$x^{(4)}$ — количество удобрений, расходуемых на гектар площади,

и

$x^{(3)}$ — число орудий поверхностной обработки почвы на 100 га

(коэффициенты детерминации на этом и последующем шаге подсчитываются как квадраты соответствующих множественных коэффициентов корреляции по формуле (3.28)).

Подсчет подправленного (на несмещенность) значения коэффициента детерминации $\hat{R}_{y.(x^{(4)}, x^{(3)})}^{*2} = \hat{R}^{*2}(2)$ и его нижней доверительной границы $R_{\min}^2(2)$ соответственно по формулам (3.35) и (4.70) дает:

$$\hat{R}^{*2}(2) = 0,422 \quad \text{и} \quad R_{\min}^2(2) = 0,324.$$

Оценка уравнения регрессии урожайности зерновых культур y (ц/га) по $x^{(3)}$ и $x^{(4)}$, подсчитанная по формуле (4.22), имеет вид:

$$\hat{f}(x^{(3)}, x^{(4)}) = \underset{(0,66)}{7,29} + \underset{(0,13)}{3,48} x^{(3)} + \underset{(1,07)}{3,48} x^{(4)}.$$

Заметим, что все три коэффициента регрессии статистически значимо отличаются от нуля при уровне значимости $\alpha = 0,05$ (см. выше о критериях проверки гипотез вида $H_0: \theta_j = 0$ с использованием статистик (4.45); напомним, что в скобках под значениями оценок коэффициентов регрессии приведены величины их среднеквадратических ошибок).

Целесообразность включения в модель в качестве второй объясняющей переменной предиктора $x^{(3)}$ подтверждается сравнением нижних доверительных границ $R_{\min}^2(1)$ и $R_{\min}^2(2)$, так как $R_{\min}^2(1) < R_{\min}^2(2)$.

3-й шаг ($k = 3$). Среди всевозможных троек объясняющих переменных $(x^{(4)}, x^{(3)}, x^{(j)})$, $j = 1, 2, 5$, наиболее информативной оказалась тройка $(x^{(4)}, x^{(3)}, x^{(5)})$, поскольку

$$\max_{j=1,2,5} \hat{R}_{y.(x^{(4)}, x^{(3)}, x^{(j)})}^2 = \hat{R}_{y.(x^{(4)}, x^{(3)}, x^{(5)})}^2 = 0,513$$

(напомним, что $x^{(5)}$ — это количество химических средств защиты растений, расходуемых на гектар, ц/га).

По схеме, полностью повторяющей предыдущий шаг, находим:

$$R^{*2}(3) = \hat{R}_{y.(x^{(4)}, x^{(3)}, x^{(5)})}^{*2} = 0,422 \quad \text{и} \quad R_{\min}^2(3) = 0,312.$$

Сравнение нижних доверительных границ $R_{\min}^2(2)$ и $R_{\min}^2(3)$ говорит, однако, о том, что третью объясняющую переменную $x^{(5)}$ в модель включать нецелесообразно, так как $R_{\min}^2(3) < R_{\min}^2(2)$ (см. также рис. 4.3). Этот вывод подтверждается также сравнением подправленных значений коэффициентов детерминации $\hat{R}^{*2}(3)$ и $\hat{R}^{*2}(2)$: добавление $x^{(5)}$ в качестве третьей объясняющей переменной в анализируемую регрессионную модель *практически не повышает значение $\hat{R}^{*2}$* .

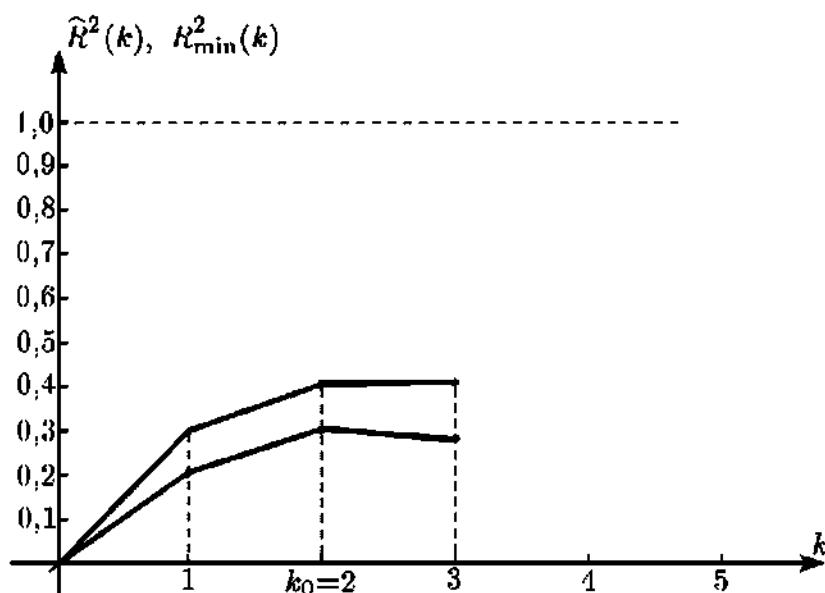


Рис. 4.3. График изменения $\hat{R}^2(k)$ (сплошная линия) и $R_{\min}^2(k)$ (пунктир) в примере 4.1

4.4.3 Ошибки спецификации модели

Выше (см. п. 1.4.3) было подробно описано содержание проблемы спецификации эконометрической модели. Напомним, что к вопросам спецификации модели относится, в частности, *выбор общего вида зависимости* между анализируемыми переменными (см. также п. 2.5), а в рамках выбранного линейного вида зависимости — *определение списка объясняющих переменных* и *спецификация случайных регрессионных остатков* («ошибок», где под спецификацией ошибок понимается формулировка исходных допущений и априорных ограничений относительно стохастической природы регрессионных остатков. В данном пункте мы кратко остановимся на ошибках спецификации (а не об упомянутой только что спецификации ошибок), причем ошибки спецификации понимаются в весьма узком смысле: речь идет только об ошибках той спецификации, которая связана с выбором списка (состава) объясняющих переменных

в рамках КЛММР. Другими словами, нас будут интересовать ошибки, возникающие при неправильной спецификации матрицы наблюдений X .

Итак, предположим, как и в (4.1'), что истинное уравнение связи между y и X имеет вид

$$Y = X\Theta + \varepsilon, \quad (4.71)$$

где X , как и прежде, матрица наблюденных значений объясняющих переменных размерности $n \times (p+1)$. Однако в процессе анализа исследователь вместо матрицы X ошибочно решил воспользоваться матрицей данных $\bar{X}$ размерности $n \times k$. При этом некоторые столбцы матриц X и $\bar{X}$ одинаковы, то есть часть объясняющих переменных истинной модели (4.71) исследователем включена верно. Однако в матрице $\bar{X}$ могут отсутствовать некоторые переменные, участвующие в истинной модели и одновременно в нее могут быть включены столбцы наблюдений по так называемым *несущественным* (то есть не участвующим в модели (4.71)) переменным. В этом и заключаются те ошибки спецификации модели, которые мы рассматриваем в данном пункте.

Соответственно, исследователь в качестве МНК-оценок неизвестных параметров Θ получит оценки

$$\bar{\Theta} = (\bar{X}^T \bar{X})^{-1} \bar{X}^T Y. \quad (4.72)$$

Подставим (4.71) в (4.72):

$$\bar{\Theta} = (\bar{X}^T \bar{X})^{-1} \bar{X}^T X \Theta + (\bar{X}^T \bar{X})^{-1} \bar{X}^T \varepsilon,$$

так что

$$E\bar{\Theta} = B\Theta, \quad (4.73)$$

где

$$B = (\bar{X}^T \bar{X})^{-1} \bar{X}^T X. \quad (4.74)$$

Из сравнения вида $k \times (p+1)$ -матрицы $B = (\beta_{ij})$, $i = 0, 1, \dots, k-1$; $j = 0, 1, \dots, p$ и формул для МНК-оценок коэффициентов регрессии следует, что матрица B состоит из столбцов $\beta_j = (\beta_{0j}, \beta_{1j}, \dots, \beta_{k-1j})^T$ ($j = 0, 1, \dots, p$), компоненты которых являются МНК-оценками коэффициентов регрессии j -й исходной (существенной) объясняющей переменной по всем k используемым исследователем предикторам $\bar{X}$.

Используем теперь соотношения (4.73)–(4.74) в целях анализа точности оценок $\bar{\Theta}$, получаемых при неправильной спецификации КЛММР.

1) Предположим сначала, что исследователь включил в свою модель только k (из $p+1$) существенных переменных ($k < p+1$), а остальные $p+1-k$ существенных переменных «выпали» из поля его зрения. Это означает, что матрица $\bar{X}$ отличается от матрицы X отсутствием $p+1-k$ последних столбцов, то есть

$$X = (X^{(0)} X^{(1)} \dots X^{(p)})$$

и

$$\bar{X} = (X^{(0)} X^{(1)} \dots X^{(k)}).$$

Здесь $X^{(j)}$ — вектор-столбец размерности n , состоящий из n наблюдений переменной j , то есть: $X^{(0)} = (1, 1, \dots, 1)^T$ и $X^{(j)} = (x_1^{(j)}, x_2^{(j)}, \dots, x_n^{(j)})^T$ для $1 \leq j \leq p$. Тогда $k \times (p+1)$ -матрица B будет представлена в виде:

$$B = \begin{pmatrix} 1 & 0 & \dots & 0 & \beta_{0k} & \dots & \beta_{0p} \\ 0 & 1 & \dots & 0 & \beta_{1k} & \dots & \beta_{1p} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & \beta_{k-1k} & \dots & \beta_{k-1p} \end{pmatrix}. \quad (4.75)$$

Действительно, если переменная $x^{(j)}$ входит в состав отобранных исследователем предикторов (а следовательно, наблюдения $X^{(j)}$ этой переменной входят в состав матрицы $\bar{X}$), то регрессия $x^{(j)}$ на множество векторов $\bar{X}$ дает коэффициент регрессии, равный единице при $x^{(j)}$ и нулю при остальных включенных в $\bar{X}$ объясняющих переменных. Последние $p+1-k$ столбцов матрицы B содержат коэффициенты регрессии по $\bar{X}$ для тех $x^{(k)}, x^{(k+1)}, \dots, x^{(p)}$, которые не вошли в матрицу $\bar{X}$. Подставим (4.75) в (4.73) и получим выражение для величины смещения в оценке любого из оцениваемых с помощью $\bar{\Theta}$ коэффициентов:

$$E\bar{\theta}_j = \theta_j + \beta_{jk}\theta_k + \beta_{jk+1}\theta_{k+1} + \dots + \beta_{jp}\theta_p. \quad (4.76)$$

Таким образом, оценки $\bar{\Theta}$ коэффициентов регрессии Θ окажутся смещенными по отношению к истинным значениям этих коэффициентов, а степень смещения зависит от статистических связей между включенными в анализ и «проигнорированными» объясняющими переменными, а также от значений коэффициентов регрессии при исключенных переменных $x^{(k)}, x^{(k+1)}, \dots, x^{(p)}$.

2) Теперь предположим, что исследователь включил в рассмотрение не только все «существенные» объясняющие переменные $x^{(0)} \equiv 1, x^{(1)}, \dots, x^{(p)}$, но и какое-то количество m «несущественных» (то есть, в действительности, не участвующих в истинной модели (4.71)). Тогда матрица $\bar{X}$ отличается от X дополнительными m столбцами, а матрица B имеет порядок $(p+1+m) \times (p+1)$ и принимает вид:

$$B = \begin{pmatrix} I_{p+1} \\ 0_{m,p+1} \end{pmatrix}, \quad (4.77)$$

где I_{p+1} — единичная матрица порядка $(p+1) \times (p+1)$, а $0_{m,p+1}$ — матрица порядка $m \times (p+1)$, состоящая из одних нулей. Подставляя (4.77) в (4.73), убеждаемся, что оценки $\bar{\Theta}$ в этом случае будут несмещенными.

Из проведенного выше анализа можно сделать следующие выводы.

- Соотношения (4.75) и (4.76) свидетельствуют о том, что неправомерное исключение из КЛММР объясняющих переменных может привести к серьезным ошибкам. При этом не только оценки коэффициентов регрессии окажутся смещенными, но и, в еще большей степени, будет смещена оценка (4.31) остаточной дисперсии σ^2 . А это значит, что неточность выводов, основанных на оценках $\bar{\Theta}$, еще более усугубится.

- Если позволяет объем выборки (n), которой мы располагаем, то лучше ошибиться, введя в модель «лишние» (несущественные) переменные, нежели исключив существенные. Как мы видели, полученные при этом оценки остаются несмещенными (правда, несколько ослабевает точность статистических выводов, зависящая от отношения p/n). Кроме того, увеличение числа предикторов может приводить к мультиколлинеарности, с которой связаны свои «неприятности» (см. п. 4.4.1).

- Рассмотренная схема служит предостережением против наивного заключения о том, что МНК автоматически дает наилучшие линейные несмещенные оценки. Следует помнить о том, что это справедливо лишь в случае выполнения всех условий КЛММР (см. соотношения (4.1) или (4.1')).

4.5 КЛММР с линейными ограничениями на параметры

4.5.1 Пример: оценка производственной функции Кобба–Дугласа для производства, функционирующего в режиме постоянной отдачи от масштаба

Как известно, производственная функция Кобба–Дугласа описывает зависимость объема выпуска продукции (V) от стоимости основных средств (K) и фонда оплаты труда (L) в форме:

- *детерминированная версия*

$$V = c \cdot K^{\theta_1} \cdot L^{\theta_2}; \quad (4.78)$$

- *стохастическая версия*

$$V = c \cdot K^{\theta_1} \cdot L^{\theta_2} e^{\varepsilon}, \quad (4.78')$$

где ε – стохастические регрессионные остатки, удовлетворяющие всем требованиям КЛММР.

Когда говорят, что предприятие функционирует в *режиме постоянства отдачи от масштаба*, то подразумевают, что изменение в m раз

стоимости основных факторов производства K и L приводит к изменению объема производства в той же пропорции, то есть в терминах (4.78);

$$mV = c(mK)^{\theta_1}(mL)^{\theta_2}$$

откуда с необходимостью следует что параметры θ_1 и θ_2 должны быть связаны соотношением:

$$\theta_1 + \theta_2 = 1. \quad (4.79)$$

Поэтому, если мы хотели бы оценить по наблюдениям $\{x_i^{(1)}, x_i^{(2)}, y_i\}_{i=1,2,\dots,n}$ параметры модели

$$y_i = \theta_0 + \theta_1 \cdot x_i^{(1)} + \theta_2 \cdot x_i^{(2)} + \varepsilon_i, \quad (4.78'')$$

где $y_i = \ln V_i$, $x_i^{(1)} = \ln K_i$, $x_i^{(2)} = \ln L_i$ и $\theta_0 = \ln c$, то мы должны сделать это с учетом ограничения (4.79).

Очевидно, это можно сделать двумя способами.

Способ 1 (специальный). Подставляя в правую часть (4.78'') вместо параметра θ_2 его выражение через θ_1 (то есть $\theta_2 = 1 - \theta_1$) и преобразуя полученное соотношение к виду

$$y_i - x_i^{(2)} = \theta_0 + \theta_1(x_i^{(1)} - x_i^{(2)}) + \varepsilon_i,$$

мы можем оценить параметры θ_0, θ_1 и $\sigma^2 = D\varepsilon_i$ как параметры уравнения парной регрессии $(y_i - x_i^{(2)})$ по $(x_i^{(1)} - x_i^{(2)})$.

Способ 2 (общий). Этот способ основан на решении условно-оптимизационной задачи вида:

$$\begin{cases} \sum_{i=1}^n (y_i - \theta_0 - \theta_1 \cdot x_i^{(1)} - \theta_2 \cdot x_i^{(2)})^2 \rightarrow \min_{\theta_0, \theta_1, \theta_2} \\ \theta_1 + \theta_2 - 1 = 0 \end{cases} \quad (4.80)$$

Функция Лагранжа $\varphi(\theta_0, \theta_1, \theta_2; \lambda)$ для этой задачи имеет вид:

$$\varphi(\theta_0; \theta_1; \theta_2; \lambda) = \sum_{i=1}^n (y_i - \theta_0 - \theta_1 x_i^{(1)} - \theta_2 x_i^{(2)})^2 - \lambda(\theta_1 + \theta_2 - 1),$$

так что оценки $\theta_0, \theta_1, \theta_2$ и $\hat{\lambda}$ определяются из системы уравнений:

$$\begin{cases} \frac{\partial \varphi(\theta_0; \theta_1; \theta_2; \lambda)}{\partial \theta_0} = 0 \\ \frac{\partial \varphi(\theta_0; \theta_1; \theta_2; \lambda)}{\partial \theta_1} = 0 \\ \frac{\partial \varphi(\theta_0; \theta_1; \theta_2; \lambda)}{\partial \theta_2} = 0 \\ \frac{\partial \varphi(\theta_0; \theta_1; \theta_2; \lambda)}{\partial \lambda} = 0 \end{cases}$$

Нетрудно проверить, что полученные этими двумя способами оценки параметров модели (4.78''), связанной ограничением (4.79), совпадут.

4.5.2 Общий вид линейных ограничений на параметры КЛММР

Обсудим ряд встречающихся в эконометрической практике линейных ограничений на параметры КЛММР и общий вид таких ограничений. Заметим, что эти ограничения могут формулироваться как в форме *априори жестко заданных условий* (как это было в нашем примере из п. 4.5.1), так и в форме *гипотез*, подлежащих статистической проверке (см. ниже, п. 4.6).

Итак, мы рассматриваем КЛММР вида (4.1'). В дальнейшем нам удобно будет разделить вектор коэффициентов регрессии $\Theta = (\theta_0; \theta_1, \dots, \theta_q; \theta_{q+1}, \dots, \theta_p)^\top$ на три части: θ_0 , $\Theta(1) = (\theta_1, \dots, \theta_q)^\top$ и $\Theta(2) = (\theta_{q+1}, \dots, \theta_p)^\top$.

(1) $\theta_j = 0$. Это ограничение исключает переменную $x^{(j)}$ из числа предикторов модели, и тогда нас может интересовать, как соотносятся МНК-оценки остальных параметров модели, подсчитанные «без оглядки» на это ограничение и с учетом него. Если же (1) интерпретируется как гипотеза, то нас будет интересовать процедура статистической проверки этой гипотезы (см. ниже, п. 4.6).

(2) $\Theta(2) = \mathbf{O}_{p-q}$. Это ограничение исключает целый ряд переменных (а именно, переменные $x^{(q+1)}, x^{(q+2)}, \dots, x^{(p)}$) из числа предикторов модели. Задачи, здесь возникающие, те же, что и в п. (1).

(3) $\begin{pmatrix} \Theta(1) \\ \Theta(2) \end{pmatrix} = \mathbf{O}_p$. Подобное ограничение имеет смысл, пожалуй, *только в форме гипотезы*, в соответствии с которой ни одна из объясняющих переменных модели, в действительности, не влияет на анализируемую зависимую переменную.

(4) $\theta_1 + \theta_2 = 1$ в модели (4.78). то есть речь идет об ограничении, проанализированном в примере п.4.5.1.

Здесь, конечно, возможна двойственность в интерпретации (4). Если нам из предыстории *уже известно*, что анализируемое предприятие функционирует в режиме постоянства отдачи от масштаба, то нас заботит идентификация модели (4.78'') по наблюдениям $\{x_i^{(1)}, x_i^{(2)}, y_i\}$, что сводится к решению условно-оптимизационной задачи (4.80). Если же мы должны дать ответ на вопрос, функционирует ли наблюдаемое предприятие в режиме постоянства отдачи от масштаба, то *наша задача сводится к статистической проверке гипотезы (4)*.

Все приведенные выше примеры линейных ограничений на параметры модели (4.1') являются частными случаями общей формы вида

$$\mathbf{R}\Theta = r, \quad (4.81)$$

где $\mathbf{R}$ — заданная $k \times (p+1)$ -матрица, а r — заданный $k \times 1$ -вектор-столбец (здесь k — общее число накладываемых, или статистически проверяемых, ограничений). Конечно, число ограничений (k) не может превос-

ходить число параметров $(p + 1)$, а сами ограничения не должны быть линейно взаимнозависимыми (то есть ранг $\mathbf{R} = k$). Другими словами, ограничения, задаваемые соотношением (4.81), не должны противоречить друг другу и не должны быть избыточными. Выпишем конкретные выражения матрицы $\mathbf{R}$ и вектор-столбца r для всех рассмотренных выше примеров (1)–(4).

- (1) $k = 1$; $\mathbf{R} = (0 \dots 010 \dots 0)$; $r = 0$, причем, единица в строке, определяющей матрицу $\mathbf{R}$, стоит на $(j + 1)$ -м месте;
- (2) $k = p - q$; $\mathbf{R} = (\Theta_{(p-q) \times (q+1)} \mathbf{I}_{(p-q)})$; $r = \mathbf{O}_{p-q}$, где $\mathbf{O}_{p-q}$ — вектор-столбец из $p - q$ нулей, $\mathbf{O}_{(p-q) \times (q+1)}$ — $(p - q) \times (q + 1)$ -матрица из нулей, а $\mathbf{I}_{p-q}$ — единичная матрица размерности $p - q$.
- (3) $k = p$; $\mathbf{R} = (\mathbf{O}_p \mathbf{I}_p)$; $r = \mathbf{O}_p$.
- (4) $k = 1$; $\mathbf{R} = (011)$; $r = 1$.

4.5.3 МНК-оценки КЛММР при наличии линейных ограничений на ее параметры

Итак, мы располагаем наблюдениями (4.1а)–(4.3) переменных y и $X = (1, x^{(1)}, \dots, x^{(p)})$ модели (4.1') при наличии ограничений (4.81) и хотим вычислить МНК-оценки параметров такой модели. Для этого необходимо решить условно-оптимизационную задачу вида

$$\begin{cases} (Y - \mathbf{X}\Theta)^T(Y - \mathbf{X}\Theta) \rightarrow \min_{\Theta} \\ \mathbf{R}\Theta - r = \mathbf{O}_k \end{cases}$$

Функция Лагранжа для этой условно-оптимизационной задачи:

$$\begin{aligned} \varphi(\Theta; \lambda) &= (Y - \mathbf{X}\Theta)^T(Y - \mathbf{X}\Theta) - \lambda^T(\mathbf{R}\Theta - r) = \\ &= Y^T Y - 2\Theta^T \mathbf{X}^T Y + \Theta^T (\mathbf{X}^T \mathbf{X}) \Theta - \Theta^T \mathbf{R}^T \lambda + \lambda^T r, \end{aligned}$$

так что, дифференцируя $\varphi(\Theta; \lambda)$ по Θ и по $\lambda = (\lambda_1, \dots, \lambda_k)^T$ и приравнявая полученные производные к нулю, получаем следующую систему уравнений относительно Θ и λ :

$$\begin{cases} \frac{\partial \varphi}{\partial \Theta} = -2\mathbf{X}^T Y + 2(\mathbf{X}^T \mathbf{X})\Theta - \mathbf{R}^T \lambda = \mathbf{O}_{p+1} \\ \frac{\partial \varphi}{\partial \lambda} = -(\mathbf{R}\Theta - r) = \mathbf{O}_k \end{cases} \quad (4.82)$$

(при дифференцировании φ по Θ и по λ использовались правила матричного дифференцирования, см. Приложение 3).

Получим решения системы (4.82). Из уравнений первого соотношения системы (4.82) имеем:

$$\tilde{\Theta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} + \frac{1}{2} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R}^T \lambda \quad (4.82a)$$

или

$$\tilde{\Theta} = \hat{\Theta}_{\text{МНК}} + \frac{1}{2} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R}^T \lambda \quad (4.82b)$$

Чтобы исключить из (4.82a) параметр λ , умножим соотношение (4.82b) слева на $\mathbf{R}$, из полученного соотношения определим λ и вставим его в (4.82a).

$$\lambda = 2 \left(\mathbf{R} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R}^T \right) (\mathbf{R} \tilde{\Theta} - \mathbf{R} \hat{\Theta}_{\text{МНК}}) = 2 \left(\mathbf{R} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R}^T \right)^{-1} (r - \mathbf{R} \hat{\Theta}_{\text{МНК}})$$

(поскольку $\mathbf{R} \tilde{\Theta} = r$ по построению). Вставляя эти значения λ в (4.82a), получаем:

$$\tilde{\Theta} = \hat{\Theta}_{\text{МНК}} + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R}^T \left(\mathbf{R} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R}^T \right)^{-1} (r - \mathbf{R} \hat{\Theta}_{\text{МНК}}) \quad (4.83)$$

Выведем теперь соотношение, связывающее между собой суммы квадратов остатков, оцененных в рамках КЛММР с ограничениями на параметры и без них.

$$\tilde{\varepsilon} = \mathbf{Y} - \mathbf{X} \tilde{\Theta} = \mathbf{Y} - \mathbf{X} \hat{\Theta}_{\text{МНК}} - \mathbf{X} (\tilde{\Theta} - \hat{\Theta}_{\text{МНК}}) = \hat{\varepsilon}_{\text{МНК}} - \mathbf{X} (\tilde{\Theta} - \hat{\Theta}_{\text{МНК}})$$

Транспонируя это выражение и умножая на $\tilde{\varepsilon}$, получаем:

$$\tilde{\varepsilon}^T \tilde{\varepsilon} = \hat{\varepsilon}_{\text{МНК}}^T \hat{\varepsilon}_{\text{МНК}} + (\tilde{\Theta} - \hat{\Theta}_{\text{МНК}})^T (\mathbf{X}^T \mathbf{X}) (\tilde{\Theta} - \hat{\Theta}_{\text{МНК}})$$

Подставляя в правую часть разность $\tilde{\Theta} - \hat{\Theta}_{\text{МНК}}$, выраженную из (4.83), имеем:

$$\tilde{\varepsilon}^T \tilde{\varepsilon} = \hat{\varepsilon}_{\text{МНК}}^T \hat{\varepsilon}_{\text{МНК}} + (\mathbf{R} \hat{\Theta}_{\text{МНК}} - r)^T \left(\mathbf{R} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R}^T \right)^{-1} (\mathbf{R} \hat{\Theta}_{\text{МНК}} - r), \quad (4.84)$$

или, после деления обеих частей (4.84) на n :

$$\tilde{\sigma}^2 = \hat{\sigma}_{\text{МНК}}^2 + \frac{1}{n} (\mathbf{R} \hat{\Theta}_{\text{МНК}} - r)^T \left(\mathbf{R} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R}^T \right)^{-1} (\mathbf{R} \hat{\Theta}_{\text{МНК}} - r), \quad (4.84')$$

где $\tilde{\sigma}^2 = \frac{1}{n} \tilde{\varepsilon}^T \tilde{\varepsilon}$ и $\hat{\sigma}_{\text{МНК}}^2 = \frac{1}{n} \hat{\varepsilon}_{\text{МНК}}^T \cdot \hat{\varepsilon}_{\text{МНК}}$.

З а м е ч а н и е. Нетрудно показать, что решение условно-оптимизационной задачи, возникающей при выводе оценок максимального правдоподобия $\hat{\Theta}_{\text{МП}}$ параметров Θ в нормальной КЛММР с ограничениями (4.81), совпадает с условными МНК-оценками $\tilde{\Theta}$, определяемыми соотношением (4.83) (оставляем доказательство этого факта в качестве упражнения читателю; более четкую постановку этой задачи см. ниже, соотношения (4.93)–(4.95)). Именно поэтому в дальнейшем с целью упрощения обозначений мы будем опускать нижний индекс МНК у рассматриваемых оценок ($\hat{\Theta}$ и $\tilde{\Theta}$) и у соответствующих векторов оцененных невязок ($\tilde{\varepsilon}$ и $\hat{\varepsilon}$).

4.6 Общий подход к статистической проверке гипотез о наличии линейных связей между параметрами КЛММР

В п. 4.2.3 мы уже касались статистических свойств МНК-оценок параметров нормальной КЛММР (см. (4.42)–(4.44)) и основанных на них возможностях статистической проверки гипотез о значениях параметров модели (см. (4.45)–(4.46)) и построения интервальных оценок (см. (4.47)). Однако существует общий подход к построению статистических критериев проверки гипотез вида (4.81), к описанию которого мы и приступаем.

4.6.1 Статистическая проверка гипотез о наличии линейных связей между параметрами нормальной КЛММР (F -критерий)

В основе подхода, используемого в рамках нормальной КЛММР и опирающегося только на обычные МНК-оценки коэффициентов регрессии, лежит соображение, что если гипотеза

$$H_0: R\Theta = r \quad (4.81a)$$

верна, то значения векторной случайной величины $R\hat{\Theta}_{\text{МНК}} - r$ должны концентрироваться около нулевого вектора соответствующей размерности.

Критическая статистика критерия проверки гипотезы (4.81a) строится с учетом свойств статистики $R\hat{\Theta}_{\text{МНК}}$, а именно:

- (а) среднее значение случайной величины $R\hat{\Theta}_{\text{МНК}}$: $E(R\hat{\Theta}_{\text{МНК}}) = R\Theta$;
- (б) ковариационная матрица случайной величины $R\hat{\Theta}_{\text{МНК}}$

$$\begin{aligned} \Sigma_{R\hat{\Theta}_{\text{МНК}}} &= E \left[(R\hat{\Theta}_{\text{МНК}} - R\Theta)(R\hat{\Theta}_{\text{МНК}} - R\Theta)^T \right] = \\ &= RE \left[(\hat{\Theta}_{\text{МНК}} - \Theta)(\hat{\Theta}_{\text{МНК}} - \Theta)^T \right] R^T = R \cdot \Sigma_{\hat{\Theta}_{\text{МНК}}} \cdot R^T = \\ &= \sigma^2 \cdot R(X^T X)^{-1} \cdot R^T \end{aligned}$$

Следовательно, с учетом известных свойств (4.42)–(4.46) МНК-оценок параметров модели (4.1''), а также того факта, что из (а; Σ)-многомерной нормальности k -мерной случайной величины ξ (при невырожденной ковариационной матрице Σ) вытекает $\chi^2(k)$ -распределенность случайной величины $\eta = (\xi - a)^T \Sigma^{-1}(\xi - a)$, имеем:

$$(в) (R\hat{\Theta}_{\text{МНК}} - r)^T (\sigma^2 \cdot R(X^T X)^{-1} R^T)^{-1} (R\hat{\Theta}_{\text{МНК}} - r) \sim \chi^2(k).$$

Левая часть этого выражения зависит от неизвестного значения σ^2 . Чтобы избавиться от этой зависимости, поделим обе части (в) на статистику

$$\frac{1}{n-p-1} \frac{\hat{\varepsilon}_{\text{МНК}}^T \cdot \hat{\varepsilon}_{\text{МНК}}}{\sigma^2},$$

которая, как известно (см. (4.43)–(4.44)), распределена по закону $\chi^2(n-p-1)$ и статистически независима с $\hat{\Theta}_{\text{МНК}}$, а значит, и со всей левой частью (в). Соответственно, получаем критическую статистику

$$\gamma = \frac{\frac{1}{k}(\mathbf{R}\hat{\Theta}_{\text{МНК}} - r)^T (\mathbf{R}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R}^T)^{-1} (\mathbf{R}\hat{\Theta}_{\text{МНК}} - r)}{\frac{1}{n-p-1} \hat{\varepsilon}_{\text{МНК}}^T \cdot \hat{\varepsilon}_{\text{МНК}}}, \quad (4.85)$$

которая в условиях справедливости гипотезы (4.81а) должна «вести себя» как $F(k; n-p-1)$ — распределенная случайная величина. При этом, именно большие отклонения разности $\mathbf{R}\hat{\Theta}_{\text{МНК}} - r$ от нулевого вектора будут свидетельствовать о нарушениях гипотезы (4.81а), так что основанный на γ статистический критерий имеет *односторонний характер*. А именно, если подсчитанное по (4.85) значение γ оказалось больше величины 100 α %-ной точки $F(k; n-p-1)$ -распределения, то гипотеза (4.81а) отвергается (с вероятностью ошибиться, равной α).

Посмотрим, как выглядит критическая статистика (4.85) в некоторых частных случаях.

1) $H_0: \theta_j = 0$. В этом случае $k = 1$ и $\mathbf{R} = (00 \dots 010 \dots 0)$ (единица определяет $(j+1)$ -ю компоненту вектора $\mathbf{R}$), $r = 0$, так что $\mathbf{R}\hat{\Theta} = \hat{\theta}_{j,\text{МНК}}$, $\mathbf{R}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R} = a_{jj}$, где a_{jj} — это j -й диагональный элемент матрицы $A = (a_{iq})_{i,q=0,1,\dots,p} = (\mathbf{X}^T \mathbf{X})^{-1}$. Соответственно:

$$\gamma = \frac{\hat{\theta}_{j,\text{МНК}}^2}{\hat{\sigma}_{\text{МНК}}^2 \cdot a_{jj}} \sim F(1; n-p-1),$$

а с учетом того, что $F(1; m) = t^2(n-p-1)$, получаем уже известный результат:

$$\frac{\hat{\theta}_{j,\text{МНК}}}{\sqrt{D\hat{\theta}_{j,\text{МНК}}}} \sim t(n-p-1) \quad (\text{в условиях справедливости гипотезы } H_0).$$

2) $H_0: \theta_j = \theta_j^0$ — некоторое заданное значение. Этот случай отличается от предыдущего только тем, что в данном случае $r = \theta_j^0$. Поэтому

$$\gamma = \frac{(\hat{\theta}_{j,\text{МНК}} - \theta_j^0)^2}{\hat{\sigma}_{\text{МНК}}^2 \cdot a_{jj}} \sim F(1; n-p-1),$$

так что

$$\frac{\hat{\theta}_{j,\text{МНК}} - \theta_j^0}{\hat{\sigma}_{\text{МНК}} \cdot \sqrt{a_{jj}}} \sim t(n-p-1) \quad (\text{в условиях справедливости гипотезы } H_0).$$

3) $H_0: \theta_1 + \theta_2 = 1$ в модели $y_i = \theta_0 + \theta_1 x_i^{(1)} + \theta_2 x_i^{(2)} + \varepsilon_i$. В этом случае $k = 1, \mathbf{R} = (011), r = 1$, так что

$$\mathbf{R}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top = \mathbf{R} \begin{pmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \\ a_{20} & a_{21} & a_{22} \end{pmatrix} \mathbf{R}^\top = a_{11} + 2a_{12} + a_{22}$$

(с учетом $a_{12} = a_{21}$).

Поэтому

$$\gamma = \frac{(\hat{\theta}_{1.\text{МНК}} + \hat{\theta}_{2.\text{МНК}} - 1)^2}{\hat{\sigma}_{\text{МНК}}^2 \cdot (a_{11} + 2a_{12} + a_{22})} \sim F(1; n - 3)$$

или

$$\frac{\hat{\theta}_{1.\text{МНК}} + \hat{\theta}_{2.\text{МНК}} - 1}{\hat{\sigma}_{\text{МНК}} \cdot \sqrt{a_{11} + 2a_{12} + a_{22}}} \sim t(n - 3)$$

(в условиях справедливости гипотезы H_0).

Заметим, что выражение, стоящее в знаменателе отношения γ , определяет дисперсию случайной величины $\hat{\theta}_{1.\text{МНК}} + \hat{\theta}_{2.\text{МНК}} - 1$. Действительно:

$$\begin{aligned} \mathbf{D}(\hat{\theta}_{1.\text{МНК}} + \hat{\theta}_{2.\text{МНК}} - 1) &= \mathbf{D}(\hat{\theta}_{1.\text{МНК}} + \hat{\theta}_{2.\text{МНК}}) = \\ &= \mathbf{D}\hat{\theta}_{1.\text{МНК}} + \mathbf{D}\hat{\theta}_{2.\text{МНК}} + 2\text{cov}(\hat{\theta}_{1.\text{МНК}}, \hat{\theta}_{2.\text{МНК}}) = \\ &= \hat{\sigma}_{\text{МНК}}^2 \cdot a_{11} + \hat{\sigma}_{\text{МНК}}^2 \cdot a_{22} + 2\hat{\sigma}_{\text{МНК}}^2 \cdot a_{12} = \hat{\sigma}_{\text{МНК}}^2 \cdot (a_{11} + 2a_{12} + a_{22}). \end{aligned}$$

4) $H_0: \Theta(2) = \mathbf{O}_{p-q}$ в модели $y_i = \theta_0 + \Theta^\top(1) \cdot X_i(1) + \Theta^\top(2) X_i(2) + \varepsilon_i$, где $X_i(1) = (x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(q)})^\top$, $X_i(2) = (x_i^{(q+1)}, \dots, x_i^{(p)})^\top$, то есть речь идет о проверке гипотезы, в соответствии с которой объясняющие переменные $x^{(q+1)}, x^{(q+2)}, \dots, x^{(p)}$ в модели (4.1'') не влияют на значения анализируемой зависимой переменной y (см. случай (2) в п. 4.5.2). Как мы уже видели, для представления гипотезы H_0 в *общей форме* (4.81a) следует положить:

$$\mathbf{R} = (\mathbf{O}_{(p-q) \times (p-q)} \mathbf{I}_{(p-q)}) \text{ и } r = \mathbf{O}_{p-q}.$$

С учетом этого, а также с помощью некоторых преобразований (см., например, [Магнус и др. (2005)], п. 3.5) можно показать, что критическая статистика (4.85) в данном случае имеет вид:

$$\gamma = \frac{\frac{1}{p-q} (\hat{\varepsilon}^\top(1) \cdot \hat{\varepsilon}(1) - \hat{\varepsilon}^\top \cdot \hat{\varepsilon})}{\frac{1}{n-p-1} \hat{\varepsilon}^\top \hat{\varepsilon}} \sim F(p-q; n-p-1), \quad (4.86)$$

где $\hat{\varepsilon}$ и $\hat{\varepsilon}(1)$ — это $n \times 1$ -векторы-столбцы МНК-оцененных «невязок» в регрессии y по **всем** $x^{(1)}, \dots, x^{(p)}$ и в регрессии y **только по** $x^{(1)}, \dots, x^{(q)}$

соответственно. Справедливо и равносильное выражение для статистики γ :

$$\gamma = \frac{\frac{1}{p-q} (\hat{R}_{y.X}^2 - \hat{R}_{y.X(1)}^2)}{\frac{1}{n-p-1} (1 - \hat{R}_{y.X}^2)} \sim F(p-q; n-p-1), \quad (4.86')$$

где $\hat{R}_{y.X}^2$ и $\hat{R}_{y.X(1)}^2$ — выборочные коэффициенты детерминации между y и $X = (1x^{(1)}x^{(2)} \dots x^{(p)})^T$ и между y и $X(1) = (1x^{(1)}x^{(2)} \dots x^{(q)})^T$ соответственно, определяемые соотношением (4.54) (см. п. 4.3).

5) $H_0: \begin{pmatrix} \Theta(1) \\ \Theta(2) \end{pmatrix} = O_p$ в модели $y_i = \theta_0 + \hat{\Theta}^T(1)X_i(1) + \hat{\Theta}^T(2)X_i(2) + \varepsilon_i$, то есть в данном случае речь идет о проверке гипотезы, в соответствии с которой ни одна из включенных в исследуемую модель объясняющих переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ не влияет на анализируемую зависимую переменную. В этом случае представление гипотезы H_0 в общей форме (4.81) использует матрицу $R = (O_p I_p)$ и вектор-столбец $r = O_p$. С помощью некоторых преобразований матрицы $X^T X$ можно показать (см., например, [Johnston, DiNardo (1997)], п. 3.4.5), что критическая статистика (4.85) сводится к виду:

$$\gamma = \frac{\frac{1}{p} \hat{R}_{y.X}^2}{\frac{1}{n-p-1} (1 - \hat{R}_{y.X}^2)} F(p; n-p-1). \quad (4.87)$$

Это, кстати, объясняет и обосновывает предложенный в п. 3.2.4 способ статистической проверки гипотезы $R_{y.X} = 0$.

4.6.2 Асимптотические критерии Вальда (W -тест), отношения правдоподобия (LR -тест) и множителей Лагранжа (LM -тест)

Описанные в данном разделе критерии также предназначены для статистической проверки общих линейных гипотез (4.81a) о параметрах КЛММР. С точки зрения теоретической рассмотрение этих критериев, наряду с описанным в предыдущем пункте F -критерием, вполне оправдано, так как каждый из них основан на своем принципе, имеет «свою идеологию».

Исходная посылка критерия Вальда (W -теста), по существу, совпадает с отправной точкой F -критерия: и тот, и другой строят свои критические статистики на соображении, в соответствии с которым значения разностей векторной статистики $R\hat{\Theta}_{\text{МНК}} - r$ должны концентрироваться (в предположении справедливости гипотезы H_0) около нулевого вектора соответствующей размерности. Однако Вальд не требует нормальности остатков анализируемой КЛММР и поэтому ему удается построить лишь асимптотический (по $n \rightarrow \infty$) критерий, так как он

использует лишь асимптотически точный вид ковариационной матрицы статистики $R\hat{\Theta}_{\text{МНК}} - r$.

Критерий отношения правдоподобия (LR -тест) и критерий множителей Лагранжа (ML -тест) эксплуатируют различные свойства функции правдоподобия наблюдений (4.1a)–(4.3)

$$L(y_1, \dots, y_n | X_1, \dots, X_n; \Theta; \sigma^2) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \times \times e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \theta_0 - \theta_1 x_i^{(1)} - \dots - \theta_p x_i^{(p)})^2}, \quad (4.88)$$

справедливые в условиях истинности гипотезы H_0 при подстановке в (4.88) обычных или условных, то есть полученных с учетом ограничений (4.81), оценок наименьших квадратов (ранее обозначенных, соответственно, $(\hat{\Theta}; \hat{\sigma}^2)$ и $(\tilde{\Theta}; \tilde{\sigma}^2)$). В частности, LR -тест исходит из того, что если гипотеза (4.81a) верна, то значения разницы между условным (при условиях (4.81)) и безусловным максимумом (по Θ и σ^2) функции (4.88) должны концентрироваться около нуля. А именно, критерий отношения правдоподобия проверяет, отличается ли статистика

$$LR = \ln \frac{L(y_1, \dots, y_n | X_1, \dots, X_n; \hat{\Theta}; \hat{\sigma}^2)}{L(y_1, \dots, y_n | X_1, \dots, X_n; \tilde{\Theta}; \tilde{\sigma}^2)} \quad (4.89)$$

статистически значимо от нуля. В критерии же множителей Лагранжа (в LM -тесте) исходная позиция заключается в том, что если гипотеза (4.81a) верна, то функция правдоподобия (4.88) должна достигать своего максимума при $\Theta = \tilde{\Theta}$ и $\sigma^2 = \tilde{\sigma}^2$, а следовательно, должны выполняться для функции L условия ее экстремума первого порядка при $\Theta = \tilde{\Theta}$ и $\sigma^2 = \tilde{\sigma}^2$, то есть производные

$$\left. \frac{\partial \ln L}{\partial \Theta} \right|_{\Theta=\tilde{\Theta}} \quad \text{и} \quad \left. \frac{\partial \ln L}{\partial \sigma^2} \right|_{\sigma^2=\tilde{\sigma}^2}$$

не должны статистически значимо отличаться от нуля.

Заметим, что при реализации упомянутых выше идей эксплуатируются асимптотические (по $n \rightarrow \infty$) свойства используемых статистик, поэтому все три теста носят асимптотический характер. При этом, асимптотические распределения всех трех статистик (W , LR и LM) совпадают. Однако на практике при конечных объемах выборок (n) значения этих трех статистик могут существенно различаться, причем, можно показать, что эти значения всегда упорядочены следующим образом:

$$LM \leq LR \leq W. \quad (4.90)$$

С точки зрения прикладной, практической, грубо говоря, для решения задач статистической проверки гипотез о значениях параметров

КЛММР хватило бы и одного из этих критериев (например, F -критерия). Однако возможность многократно, с *разных позиций*, статистически проверить имеющиеся гипотетические предположения относительно параметров модели, конечно, повышает нашу уверенность в справедливости полученного результата.

Теперь опишем кратко все три упомянутые критерия.

1) Критерий Вальда (W -тест)

Реализация этого критерия требует вычисления *только безусловных МНК-оценок* $\hat{\Theta}$ и $\hat{\sigma}^2$ (и в этом его предпочтительность в сравнении с критерием отношения правдоподобия, реализация которого требует вычисления как безусловных МНК-оценок $\hat{\Theta}$ и $\hat{\sigma}^2$, так и оценок $\tilde{\Theta}$ и $\tilde{\sigma}^2$, вычисленных с учетом ограничений (4.81)). При этом, используется выведенное ранее (см. п. 4.6.1) свойство (в). Если участвующую в этом соотношении дисперсию остатков σ^2 заменить на ее состоятельную МНК-оценку $\hat{\sigma}^2$, то можно воспользоваться асимптотически справедливым фактом:

$$W = \frac{1}{\hat{\sigma}^2} (\mathbf{R}\hat{\Theta} - r)^\top (\mathbf{R}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top)^{-1} (\mathbf{R}\hat{\Theta} - r) \stackrel{\text{ас.}}{\approx} \chi^2(k). \quad (4.91)$$

Так что гипотеза (4.81a) отвергается (с вероятностью ошибиться, равной α), если значение статистики W оказалось больше 100 α %-ной точки «хи-квадрат»-распределения с k степенями свободы (то есть если $W > \chi_\alpha^2(k)$).

Статистику W можно представить в терминах сумм квадратов «невязок» (условных $\tilde{\varepsilon}^\top \tilde{\varepsilon}$ и безусловных $\hat{\varepsilon}^\top \hat{\varepsilon}$). Действительно, учитывая выведенное ранее соотношение (4.84) и выражение для $\hat{\sigma}^2 = \frac{1}{n} \hat{\varepsilon}^\top \hat{\varepsilon}$, возвращаясь к (4.91), получаем:

$$W = \frac{n(\tilde{\varepsilon}^\top \tilde{\varepsilon} - \hat{\varepsilon}^\top \hat{\varepsilon})}{\hat{\varepsilon}^\top \hat{\varepsilon}} \stackrel{\text{ас.}}{\approx} \chi^2(k). \quad (4.91')$$

Правда, для вычисления W в форме (4.91') потребуются предварительно вычислить не только безусловные МНК-оценки $\hat{\Theta}$ и $\hat{\sigma}^2$, но и оценки $\tilde{\Theta}$ и $\tilde{\sigma}^2$, учитывающие ограничение (4.81).

2) Критерий отношения правдоподобия (LR -тест)

Поскольку этот критерий основан на сравнении значений функции правдоподобия $L(y_1, \dots, y_n \mid X_1, \dots, X_n; \Theta; \sigma^2) = L(\Theta; \sigma^2)$ (см. (4.88)) регрессионных наблюдений (4.1a)–(4.3) при подстановке в нее сначала обычных (безусловных) МНК-оценок $\hat{\Theta}$ и $\hat{\sigma}^2$ (которые одновременно являются и оценками максимального правдоподобия при нормальных остатках КЛММР, см. ниже, а затем МНК-оценок $\tilde{\Theta}$ и $\tilde{\sigma}^2$, учитывающих ограничение (4.81) (см. (4.89)), то его реализация требует предварительного вычисления и безусловных ($\hat{\Theta}, \hat{\sigma}^2$) и условных ($\tilde{\Theta}, \tilde{\sigma}^2$) МНК-оценок параметров анализируемой модели (4.1''). При построении критической

статистики LR этого критерия используется общий асимптотический результат, в соответствии с которым статистика

$$LR = 2 \left[\ln L(\hat{\Theta}; \hat{\sigma}^2) - \ln L(\tilde{\Theta}; \tilde{\sigma}^2) \right] \quad (4.92)$$

должна вести себя (в условиях справедливости гипотезы (4.81a) и при $n \rightarrow \infty$) как $\chi^2(k)$ случайная величина (для малых конечных n закон распределения статистики LR специально исследуется и моделируется).

Можно показать (упражнение читателю для самостоятельного решения), что:

- безусловные и условные (при условии (4.81)) оценки максимального правдоподобия $(\hat{\Theta}_{\text{мп}}; \hat{\sigma}_{\text{мп}}^2)$ и $(\tilde{\Theta}_{\text{мп}}; \tilde{\sigma}_{\text{мп}}^2)$ совпадают, соответственно, с безусловными и условными МНК-оценками параметров Θ и σ^2 анализируемой модели (то есть, соответственно, с $(\hat{\Theta}_{\text{мнк}}; \hat{\sigma}_{\text{мнк}}^2)$ и $(\tilde{\Theta}_{\text{мнк}}; \tilde{\sigma}_{\text{мнк}}^2)$);
- $\ln L(\hat{\Theta}; \hat{\sigma}^2) = -n \ln \sqrt{2\pi} - \frac{n}{2} \ln \frac{1}{n} - \frac{n}{2} \ln(\hat{\varepsilon}^T \hat{\varepsilon}) - \frac{n}{2}$ (с учетом того, что $\hat{\sigma}^2 = \frac{1}{n} \hat{\varepsilon}^T \hat{\varepsilon}$);
- $\ln L(\tilde{\Theta}; \tilde{\sigma}^2) = -n \ln \sqrt{2\pi} - \frac{n}{2} \ln \frac{1}{n} - \frac{n}{2} \ln(\tilde{\varepsilon}^T \tilde{\varepsilon}) - \frac{n}{2}$ (с учетом того, что $\tilde{\sigma}^2 = \frac{1}{n} \tilde{\varepsilon}^T \tilde{\varepsilon}$).

Поэтому, принимая во внимание эти факты, можем представить определенную соотношением (4.92) критическую статистику LR в форме:

$$LR = n \left(\ln(\tilde{\varepsilon}^T \tilde{\varepsilon}) - \ln(\hat{\varepsilon}^T \hat{\varepsilon}) \right) \sim \chi^2(k). \quad (4.92')$$

Так что, применяя *критерий отношения правдоподобия*, мы должны отвергнуть гипотезу (4.81a) (с вероятностью ошибки, равной заданному значению α), если $LR > \chi_{\alpha}^2(k)$, где $\chi_{\alpha}^2(k)$ — 100 α %-ная точка χ^2 -распределения с k степенями свободы.

3) Критерий множителей Лагранжа (LM-тест)

Название критерия обусловлено способом реализации его основной идеи, основанном на *описании поведения оценок множителей Лагранжа* в условно-оптимизационной задаче вычисления оценок максимального правдоподобия $\hat{\Theta}_{\text{мп}}$ и $\hat{\sigma}_{\text{мп}}^2$ параметров Θ и σ^2 при наличии заданных ограничений $R\Theta = r$.

Рассмотрим процедуру построения LM-теста подробнее.

Условно-оптимизационная задача вычисления оценок $\tilde{\Theta}_{\text{мп}}$ и $\tilde{\sigma}_{\text{мп}}^2$ при наличии ограничений $R\Theta = r$ имеет вид:

$$\begin{cases} \ln L(\Theta; \sigma^2) \rightarrow \max_{\Theta; \sigma^2}; \\ R\Theta - r = O_k. \end{cases} \quad (4.93)$$

Выпишем условия 1-го порядка существования решения задачи (4.93) в терминах соответствующей функции Лагранжа

$$\begin{aligned} \varphi(\Theta; \sigma^2; \lambda) &= \ln L(\Theta; \sigma^2; \lambda) - \lambda^\top \cdot (\mathbf{R}\Theta - r): \\ \begin{cases} \frac{\partial \varphi(\Theta; \sigma^2; \lambda)}{\partial \Theta} = \mathbf{O}_{p+1}; \\ \frac{\partial \varphi(\Theta; \sigma^2; \lambda)}{\partial \sigma^2} = 0; \\ \frac{\partial \varphi(\Theta; \sigma^2; \lambda)}{\partial \lambda} = \mathbf{O}_k \quad (\text{где } \lambda = (\lambda_1, \lambda_2, \dots, \lambda_k)^\top). \end{cases} \end{aligned} \quad (4.94)$$

Отправляясь от вида $L(\Theta; \sigma^2; \lambda)$ (см. (4.84)), имеем:

$$\begin{aligned} \varphi(\Theta; \sigma^2; \lambda) &= n \ln \left(\frac{1}{\sqrt{2\pi}} \right) - n \ln \sigma^2 - \frac{1}{2\sigma^2} (Y - \mathbf{X}\Theta)^\top (Y - \mathbf{X}\Theta) - \\ &\quad - \Theta^\top \mathbf{R}^\top \lambda + \lambda^\top r, \end{aligned}$$

или

$$\begin{aligned} \varphi(\Theta; \sigma^2; \lambda) &= n \ln \left(\frac{1}{\sqrt{2\pi}} \right) - n \ln \sigma^2 - \frac{1}{2\sigma^2} (Y^\top Y - 2\Theta^\top \mathbf{X}^\top Y + \\ &\quad + \Theta^\top (\mathbf{X}^\top \mathbf{X}) \Theta) - \Theta^\top \mathbf{R}^\top \lambda + \lambda^\top r, \end{aligned}$$

так что

$$\frac{\partial \varphi(\Theta; \sigma^2; \lambda)}{\partial \Theta} = \frac{1}{2\sigma^2} (2\mathbf{X}^\top Y - 2(\mathbf{X}^\top \mathbf{X})\Theta) - \mathbf{R}^\top \lambda = \mathbf{O}_{p+1}. \quad (4.95)$$

Отсюда

$$\tilde{\Theta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top Y - \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1} (\mathbf{R}^\top \lambda) = \hat{\Theta} - \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1} (\mathbf{R}^\top \lambda). \quad (4.96)$$

Поскольку нас интересует поведение оценок $\hat{\lambda}$ множителей Лагранжа λ в условиях справедливости гипотезы (4.81a), попытаемся выразить $\hat{\lambda}$ через случайный вектор $\mathbf{R}\hat{\Theta} - r$, закон распределения вероятностей которого (в условиях (4.81a)) мы уже выяснили (см. выше (а)–(б) в п. 4.6.1). С этой целью домножим левую и правую части соотношения (4.96) слева на матрицу $\mathbf{R}$:

$$\mathbf{R}\tilde{\Theta} = \mathbf{R}\hat{\Theta} - \sigma^2 \left(\mathbf{R}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top \right) \lambda. \quad (4.97)$$

Принимая во внимание, что, по построению, $\mathbf{R}\tilde{\Theta} = r$, и уединяя из (4.97) λ , получаем:

$$\hat{\lambda} = \left(\sigma^2 \mathbf{R}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top \right)^{-1} (\mathbf{R}\hat{\Theta} - r).$$

Но, как мы знаем (см. (а)–(б) в п.4.6.1):

$$\mathbf{R}\hat{\Theta} - r \in N_k \left(\mathbf{O}_k; \sigma^2 \mathbf{R}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top \right),$$

а значит $\hat{\lambda}$ также подчиняется (в условиях (4.81а)) k -мерному нормальному распределению, вектор средних значений и ковариационная матрица которого определяются соотношениями:

$$\mathbf{E}\hat{\lambda} = \mathbf{O}_k;$$

$$\begin{aligned} \Sigma_{\hat{\lambda}} &= \mathbf{E}(\hat{\lambda} \cdot \hat{\lambda}^\top) = \mathbf{E} \left[\left(\sigma^2 \mathbf{R}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top \right)^{-1} (\mathbf{R}\hat{\Theta} - r)(\mathbf{R}\hat{\Theta} - r)^\top \times \right. \\ &\quad \times \left. \left(\sigma^2 \mathbf{R}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top \right)^{-1} \right] = \left(\sigma^2 \mathbf{R}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top \right)^{-1} \times \\ &\quad \times \mathbf{E} \left[(\mathbf{R}\hat{\Theta} - r)(\mathbf{R}\hat{\Theta} - r)^\top \right] \left(\sigma^2 \mathbf{R}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top \right)^{-1} = \\ &= \left(\sigma^2 \mathbf{R}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top \right)^{-1}. \end{aligned}$$

Теперь, используя свойство (в) k -мерной нормальной случайной величины (см. п.4.6.1), можем выписать критическую статистику LM в форме:

$$LM = \hat{\lambda}^\top \cdot \left(\sigma^2 \mathbf{R}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{R}^\top \right) \hat{\lambda} \sim \chi^2(k). \quad (4.98)$$

Можно выразить $\mathbf{R}^\top \hat{\lambda}$ из соотношения (4.95) в форме:

$$\mathbf{R}^\top \hat{\lambda} = \frac{1}{\sigma^2} (\mathbf{X}^\top \mathbf{Y} - \mathbf{X}^\top \mathbf{X} \tilde{\Theta}) = \frac{1}{\sigma^2} \mathbf{X}^\top (\mathbf{Y} - \mathbf{X} \tilde{\Theta}) = \frac{1}{\sigma^2} \mathbf{X}^\top \tilde{\varepsilon}.$$

Подставив это выражение в (4.98), получаем (с учетом того, что $\tilde{\sigma}^2 = \frac{1}{n} \tilde{\varepsilon}^\top \tilde{\varepsilon}$):

$$LM = \frac{n \tilde{\varepsilon}^\top \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \tilde{\varepsilon}}{\tilde{\varepsilon}^\top \tilde{\varepsilon}} \stackrel{\text{ас.}}{\sim} \chi^2(k). \quad (4.98')$$

Наконец, критическая статистика LM критерия множителей Лагранжа, так же, как и статистики W и LR , может быть представлена только в терминах сумм квадратов условных и безусловных «невязок» $\tilde{\varepsilon}^\top \tilde{\varepsilon}$ и $\hat{\varepsilon}^\top \hat{\varepsilon}$. Действительно, воспользуемся тождеством:

$$\begin{aligned} \tilde{\varepsilon}^\top \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \tilde{\varepsilon} &= \tilde{\varepsilon}^\top \tilde{\varepsilon} - \tilde{\varepsilon}^\top \left(\mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \right) \tilde{\varepsilon} = \\ &= \tilde{\varepsilon}^\top \tilde{\varepsilon} - \tilde{\varepsilon}^\top \mathbf{M} \tilde{\varepsilon}. \end{aligned} \quad (4.99)$$

Но матрица $\mathbf{M} = \mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top$ — симметричная (то есть $\mathbf{M}^\top = \mathbf{M}$) и идемпотентная (то есть $\mathbf{M}^2 = \mathbf{M}$), причем, $\mathbf{M}\mathbf{X} = \mathbf{O}_{n \times (p+1)}$ (все

три свойства подтверждаются непосредственной проверкой). Поэтому $M\tilde{\varepsilon} = M(Y - X\hat{\Theta}) = MY - M(X\Theta + \varepsilon) = M\varepsilon = \hat{\varepsilon}$, так как

$$\begin{aligned}\hat{\varepsilon} &= Y - X\hat{\Theta} = X\Theta + \varepsilon - X \left[(X^T X)^{-1} X^T (X\Theta + \varepsilon) \right] = \\ &= \varepsilon - X(X^T X)^{-1} X^T \varepsilon = M\varepsilon.\end{aligned}$$

Но тогда, возвращаясь к (4.99), имеем: $\tilde{\varepsilon}^T X(X^T X)^{-1} X^T \tilde{\varepsilon} = \tilde{\varepsilon}^T \tilde{\varepsilon} - (M\tilde{\varepsilon})^T \times$
 $\times (M\tilde{\varepsilon}) = \tilde{\varepsilon}^T \tilde{\varepsilon} - \hat{\varepsilon}^T \hat{\varepsilon}$. Подставляя это выражение в (4.96'), имеем:

$$LM = \frac{n(\tilde{\varepsilon}^T \tilde{\varepsilon} - \hat{\varepsilon}^T \hat{\varepsilon})}{\tilde{\varepsilon}^T \tilde{\varepsilon}}. \quad (4.96'')$$

Выводы

1. Методы и модели регрессионного анализа занимают центральное место во всем математико-статистическом инструментарии эконометрики. Следуя от простого к сложному, эконометрист рассматривает в первую очередь *классическую линейную модель множественной регрессии* (КЛММР), то есть зависимость результирующего показателя y от неслучайных объясняющих переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ вида

$$y_i = \theta_0 + \theta_1 x_i^{(1)} + \dots + \theta_p x_i^{(p)} + \varepsilon_i, \quad i = 1, 2, \dots, n,$$

в которой ненаблюдаемые регрессионные остатки $\varepsilon_1, \dots, \varepsilon_n$ полагаются взаимнонекоррелированными случайными величинами, имеющими нулевые средние значения (то есть $E\varepsilon_i \equiv 0$) и одинаковые дисперсии (то есть $D\varepsilon_i = \sigma^2$), $(x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)}; y_i)$ — i -е наблюдаемые значения анализируемых переменных ($i = 1, 2, \dots, n$), а $\Theta = (\theta_0; \theta_1, \dots, \theta_p)$ — неизвестные (подлежащие, наряду с σ^2 , статистической оценке) параметры модели. Если дополнительно постулируется нормальность случайных остатков $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$, то анализируемая модель называется *нормальной КЛММР*.

2. Наиболее распространенный метод оценивания параметров КЛММР — *метод наименьших квадратов* (МНК) определяет статистические оценки $\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_p$ параметров $\theta_0, \theta_1, \dots, \theta_p$ как решения оптимизационной задачи вида

$$\sum_{i=1}^n \left[y_i - (\theta_0 + \theta_1 x_i^{(1)} + \dots + \theta_p x_i^{(p)}) \right]^2 \rightarrow \min_{\theta_0, \theta_1, \dots, \theta_p},$$

то есть минимизируется сумма квадратов «невязок» между *наблюдаемыми* (y_i) и соответствующими *модельными* («подогнанными») значе-

ниями результирующего показателя. МНК позволяет получить *состоятельные, несмещенные* и в определенном смысле наиболее точные (*эффективные*) оценки для неизвестных параметров КЛММР. Если остатки КЛММР нормальны, то для оценивания параметров модели может быть применен *метод максимального правдоподобия*, который определяет статистические оценки $(\hat{\theta}, \hat{\theta}, \dots, \hat{\theta}_p)_{\text{ММП}}$ как решения оптимизационной задачи, в которой максимизируется по Θ и σ^2 *функция правдоподобия имеющихся наблюдений* $(x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)}; y_i)_{i=\overline{1, n}}$ (см. (4.20)–(4.21)). При этом оказывается, что оба метода (МНК и ММП) приводят к одним и тем же оценкам.

3. Основной характеристикой *прогностической силы модели* является так называемый *коэффициент детерминации* R^2 , определяющий, какая доля общей вариации анализируемой результирующей переменной y обусловлена изменением объясняющих переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$. Так, например, из $R^2 = 1$ следует, что *вся вариация y обусловлена изменением объясняющих переменных*, а это значит: влияние регрессионных остатков ϵ на формирование значений y *сведено к нулю*, то есть *мы имеем возможность практически без ошибок предсказывать значения y по заданным значениям объясняющих переменных* $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$.

4. Говорят, что в анализируемой ЛММР присутствует свойство *мультиколлинеарности*, если включенные в модель объясняющие переменные $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ *характеризуются высокой степенью линейной взаимозависимости* (это может проявляться: в близости к единице, по абсолютной величине, значений парных и частных коэффициентов корреляции между объясняющими переменными; в близости к нулю определителя матрицы $X^T X$, где X — матрица наблюдаемых значений объясняющих переменных; в близости к единице хотя бы одного из коэффициентов детерминации $R^2_{x^{(j)}.X^{(j)}}$ каждой из объясняющих переменных по всем остальным объясняющим переменным). *Мультиколлинеарность приводит к ряду осложнений в построении и интерпретации ЛММР*, среди которых: слишком большие значения среднеквадратических ошибок в оценках коэффициентов регрессии; неустойчивость результатов оценивания модели по отношению к небольшим изменениям исходных выборочных данных; знаки оценок коэффициентов регрессии, не согласующиеся с содержательным смыслом влияния соответствующих объясняющих переменных на y , и т.п.

5. К основным *методам устранения мультиколлинеарности* в КЛММР относятся: модификация МНК-оценок в направлении перехода к смещенным методам оценивания параметров регрессии (например, к методу ридж-регрессии); предварительная ортогонализация объясняющих переменных (например, с помощью метода главных компонент); исключение из состава объясняющих переменных ряда признаков.

6. Важным этапом спецификации ЛММР является *отбор существенных предикторов* из априорного набора объясняющих переменных. Наиболее распространенными и эффективными процедурами такого отбора являются *метод всех возможных регрессий* и *метод пошаговой регрессии*. Неправильный отбор объясняющих переменных в ЛММР приводит к *ошибкам спецификации модели*. При этом неправомерное *ущемление* (сужение) истинного набора объясняющих переменных приводит к смещению в оценках коэффициентов регрессии при оставшихся предикторах, в то время как *избыточный* (по отношению к истинному) набор объясняющих переменных к смещению оценок не приводит (но может приводить к мультиколлинеарности).

7. Одна из возможных (и весьма распространенных) постановок задач оценивания параметров КЛММР предусматривает необходимость учета априорных линейных связей, существующих между их значениями. В подобных ситуациях МНК- и ММП-оценка параметров сводится к решению соответствующих *условно-оптимизационных задач* (см. (4.82) и (4.93)–(4.95)), однако и *условные* МНК- и ММП-оценки ($\tilde{\Theta}_{\text{мнк}}$ и $\tilde{\Theta}_{\text{мп}}$) совпадают. Существуют соотношения, связывающие безусловные и условные оценки параметров КЛММР (см. (4.83)).

8. Важной составной частью эконометрического анализа КЛММР является статистическая проверка гипотез о значениях ее параметров. Разработан *общий подход* к проверке гипотез вида (4.81а), на основе которого сконструирован ряд статистических критериев: *F*-критерий (см. (4.85)), критерий Вальда (см. (4.91)), критерий отношения правдоподобия (см. (4.92')), критерий множителей Лагранжа (см. (4.98')). Последние три критерия носят асимптотический характер (то есть их корректное использование обусловлено требованиями больших объемов имеющихся данных: $n \rightarrow \infty$). Тем не менее, они предоставляют исследователю возможность многократно, с *разных позиций*, статистически проверить имеющиеся гипотетические предположения относительно параметров КЛММР, что повышает уверенность в справедливости полученного результата.

Глава 5

Обобщенная линейная модель множественной регрессии

Естественно ожидать, что при моделировании многих реальных экономических или социально-экономических процессов мы можем столкнуться с ситуациями, в которых условия классической линейной модели множественной регрессии (4.1') оказываются нарушенными. Так, если в качестве исходных статистических данных (4.1a)–(4.3) мы используем временные или пространственно-временные выборки, то чрезмерно ограничительным, нереалистичным становится, как правило, условие *взаимной некоррелированности и гомоскедастичности регрессионных остатков*, которое в терминах ковариационной матрицы Σ_ϵ остатков выражается в (4.1') соотношением $\Sigma_\epsilon = \sigma^2 \mathbf{I}_n$. Даже при соблюдении условия взаимной некоррелированности остатков приходится часто анализировать регрессионные модели, в которых *разброс остатков около линии регрессии не остается постоянным*, а меняется, например, возрастая пропорционально значениям функции регрессии.

С этой точки зрения весьма актуальным представляется расширение КЛММР в направлении отказа от упомянутого условия.

5.1 Описание обобщенной линейной модели множественной регрессии (ОЛММР)

Формальная запись ОЛММР отличается от КЛММР только отказом от требования некоррелированности и гомоскедастичности регрессионных остатков. Пусть Σ_0 — некоторая симметричная, положительно определенная матрица порядка $n \times n$, где n , как и прежде, число исходных статистических данных (4.1a)–(4.3) (то есть объем имеющейся в нашем распоряжении выборки). И пусть ковариационная матрица Σ_ϵ регрессионных остатков $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)^T$ выражается через Σ_0 со-

отношением $\Sigma_\epsilon = \sigma^2 \Sigma_0$. Будем предполагать, что в этом соотношении число σ^2 неизвестно, а матрица Σ_0 известна. С точки зрения практической, прикладной последнее предположение в большинстве случаев нереалистично; однако в дальнейшем для некоторых частных случаев мы сможем отказаться от условия априорного знания матрицы Σ_0 .

Обобщенная линейная модель множественной регрессии описывается системой следующих соотношений и условий:

$$\begin{cases} Y = X\Theta + \epsilon, \\ E\epsilon = 0_n, \\ \Sigma_\epsilon = \sigma^2 \Sigma_0, \\ (x^{(1)}, x^{(2)}, \dots, x^{(p)}) - \text{нечисловые переменные}, \\ \text{ранг матрицы } X = p + 1 < n, \end{cases} \quad (5.1)$$

где участвующие в (5.1) векторы и матрицы определены ранее соотношениями (4.1a), (4.3)–(4.9).

Сравнение (5.1) с (4.1') показывает, что ОЛММР отличается от КЛММР только видом ковариационной матрицы остатков Σ_ϵ : в КЛММР мы предполагали, что матрица Σ_ϵ с точностью до неизвестной положительной константы σ^2 равна единичной матрице I_n (что обеспечивало некоррелированность и гомоскедастичность остатков ϵ), в то время как в ОЛММР мы допускаем, что ковариации (а следовательно, дисперсии и корреляции) остатков могут быть произвольными при сохранении, правда, условия невырожденности матрицы Σ_ϵ .

Именно в этом суть обобщения модели, именно поэтому модель (5.1) называется обобщенной ЛММР. Рассмотрим два типа распространенных в практике эконометрических исследований моделей, которые вкладываются в рамки ОЛММР, но не могут быть описаны в рамках КЛММР.

1) Линейная модель регрессии с гетероскедастичными регрессионными остатками.

Впервые с линейной моделью регрессии, регрессионные остатки которой не отвечают требованиям гомоскедастичности (то есть сохранения постоянного уровня величины их дисперсии при переходе от одного наблюдения к другому, а точнее — от одного значения объясняющей переменной к другому), мы познакомились в процессе анализа примера 2.1 (см. п. 2.1). Анализируя разброс значений удельных денежных сбережений в семье $y(x)$ при различных величинах семейных среднедушевых доходов x , мы обнаружили явную зависимость несмещенной оценки $\bar{s}^2(x)$ условной дисперсии остатков $D(\epsilon | x)$ от значения x (см. таблицу 5.1; данные составлены на основании информации, содержащейся в таблице 2.1).

Таблица 5.1. Зависимость разброса удельных денежных сбережений в семье от ее среднедушевого дохода (x_i^0)

i	1	2	3	4
x_i^0 (ден. ед.)	80	120	160	200
$\bar{s}(x_i^0) \approx \sqrt{D(\varepsilon x_i^0)}$	6,4	16,0	22,6	28,9
$\bar{s}^2(x_i^0) \approx D(\varepsilon x_i^0)$	40,96	256,00	510,76	835,21

Эти данные отражают, кстати, весьма естественную для практики социально-экономических исследований ситуацию, когда более реалистично постулировать постоянство *относительного*, а не абсолютного разброса регрессионных остатков $\varepsilon(X)$ около соответствующего регрессионного значения результирующего показателя $y_{\text{ср}}(X) = E(y | X)$ (относительный разброс измеряется соответствующим *коэффициентом вариации* $V(X) = \sqrt{D(y | X)}/y_{\text{ср}}(X)$).

Отметим, что при соблюдении условия взаимной некоррелированности регрессионных остатков ковариационная матрица вектора остатков ε в примере 2.1 может быть представлена в виде

$$\Sigma_{\varepsilon} = \sigma^2 \begin{pmatrix} 1/\lambda_1 & & & & & & & & & 0 \\ & \ddots & & & & & & & & \\ & & 1/\lambda_1 & & & & & & & \\ & & & 1/\lambda_2 & & & & & & \\ & & & & \ddots & & & & & \\ & & & & & 1/\lambda_2 & & & & \\ & & & & & & 1/\lambda_3 & & & \\ & & & & & & & \ddots & & \\ & & & & & & & & 1/\lambda_3 & \\ & & & & & & & & & 1/\lambda_4 \\ & & & & & & & & & & \ddots \\ 0 & & & & & & & & & & & 1/\lambda_4 \end{pmatrix},$$

где первые 10 диагональных элементов соответствуют первым десяти наблюдениям, сделанным при фиксированном значении среднедушевого семейного дохода x_1^0 , вторые 10 диагональных элементов соответствуют следующим десяти наблюдениям, сделанным при другом фиксированном значении среднедушевого дохода x_2^0 , и т. д.

Обобщая этот пример на случай зависимости результирующего показателя от *многих* объясняющих переменных, то есть рассматривая линейную модель *множественной* регрессии (ЛММР), условные дисперсии случайных остатков которой $D(\varepsilon | X)$ оказываются величинами, за-

висящими от значений объясняющих переменных, мы приходим к определению ЛММР с *гетероскедастичными* некоррелированными остатками. Очевидно, ковариационная матрица Σ_ϵ остатков ϵ может быть представлена в данном частном случае ОЛММР в виде

$$\Sigma_\epsilon = \sigma^2 \begin{pmatrix} \lambda_1^{-1} & & & 0 \\ & \lambda_2^{-1} & & \\ & & \ddots & \\ 0 & & & \lambda_n^{-1} \end{pmatrix}, \quad (5.2)$$

где в соответствии с описанием ОЛММР (см. (5.1)) величину σ^2 мы относим к неизвестным (оцениваемым по выборке) параметрам модели, а значения $\lambda_1, \lambda_2, \dots, \lambda_n$ пока считаются известными (в дальнейшем, при определенных формах параметризации зависимости $D(\epsilon | X)$ от X , будут определены подходы и к статистическому оцениванию *априори* неизвестных значений λ_i , $i = 1, 2, \dots, n$).

2) ЛММР с автокоррелированными остатками.

Как уже упоминалось в начале этой главы, в ситуациях, когда исходные наблюдения *регистрируются во времени* (тогда, очевидно, номер наблюдения « i » несет смысловую нагрузку времени регистрации наблюдения, а потому индекс « i » часто во временных выборках заменяется индексом « t »), регрессионные остатки ϵ_i ($i = 1, 2, \dots, n$) оказываются статистически взаимозависимыми, *коррелированными*, а значит, и их ковариационная матрица Σ_ϵ не может быть диагональной.

Естественным допущением относительно природы зависимости остатков является гипотеза, в соответствии с которой эта зависимость ослабевает по мере их взаимного удаления друг от друга во времени. Одной из простых и наиболее аналитически удобных (а потому и наиболее распространенных) математических форм реализации этого допущения (*при сохранении свойства гомоскедастичности остатков*) является следующая:

$$r(\epsilon_i, \epsilon_j) = \rho^{|i-j|}, \quad (5.3)$$

где $r(\epsilon_i, \epsilon_j)$ — коэффициент корреляции между ϵ_i и ϵ_j , а ρ — некоторое число, по модулю меньшее единицы (очевидно, из (5.3) следует, что по своему вероятностному смыслу ρ — это коэффициент корреляции между соседними по времени остатками). Соотношение (5.3), в частности, означает:

- корреляционная связь между регрессионными остатками зависит только от меры их «разнесенности» во времени, но не зависит от того, к каким именно «моментам» времени i и j они «привязаны» (то есть, скажем, $r(\epsilon_1, \epsilon_5) = r(\epsilon_3, \epsilon_7)$);

- корреляционная связь между ϵ_i и ϵ_j исчезает вовсе при $|i-j| \rightarrow \infty$, то есть при неограниченном удалении остатков ϵ_i и ϵ_j друг от друга во

времени;

- ковариации остатков $\sigma_{ij} = \mathbf{E}(\varepsilon_i \varepsilon_j)$ имеют вид

$$\sigma_{ij} = \sigma_\varepsilon^2 \rho^{|i-j|}, \quad (5.4)$$

где $\sigma_\varepsilon^2 = \mathbf{D}(\varepsilon \mid X_i)$ — условная дисперсия остатков (не зависящая от величины X_i в силу гомоскедастичности); так что ковариационная матрица остатков имеет в данном случае вид:

$$\Sigma_\varepsilon = \sigma_\varepsilon^2 \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \dots & \dots & \dots & \dots & \dots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-1} & \dots & 1 \end{pmatrix}. \quad (5.4')$$

Заметим, что соотношение (5.4) следует из формулы, связывающей $r(\varepsilon_i, \varepsilon_j)$ и $\sigma_{ij} = \mathbf{E}(\varepsilon_i \varepsilon_j)$: действительно, по определению, $r(\varepsilon_i, \varepsilon_j) = \sigma_{ij} / \sqrt{\mathbf{D}\varepsilon_i} \sqrt{\mathbf{D}\varepsilon_j} = \sigma_{ij} / \mathbf{D}\varepsilon$, откуда и следует (5.4).

5.2 Оценки параметров ОЛММР по обобщенному методу наименьших квадратов (ОМНК-оценки)

Итак, нам предстоит провести статистический анализ зависимости, описываемой ОЛММР (5.1). Это значит, что на основании исходных статистических данных вида (4.1a)–(4.3) мы должны уметь, в первую очередь, построить наилучшие (в определенном смысле) точечные оценки неизвестных значений параметров Θ и σ^2 . Выясним сначала, нельзя ли воспользоваться уже известными нам МНК-оценками $\hat{\Theta}_{\text{МНК}}$ и $\hat{\sigma}^2$, определяемыми соответственно соотношениями (4.22) и (4.33)?

Обычные МНК-оценки параметров ОЛММР. Можно показать, что определенные соотношениями (4.22) обычные МНК-оценки $\hat{\Theta}_{\text{МНК}}$ остаются и в рамках ОЛММР *состоятельными* (при тех же требованиях к матрице наблюдений X) и *несмещенными*. В частности, доказательство несмещенности оценок $\hat{\Theta}_{\text{МНК}}$ в задаче оценивания параметров ОЛММР в точности повторяет доказательство этого факта в условиях КЛММР (см. (4.25), (4.26)).

Однако полученные ранее формулы (4.33) и (4.33') для ковариационной матрицы $\Sigma_{\hat{\Theta}_{\text{МНК}}}$ оценок $\hat{\Theta}_{\text{МНК}}$ и для оценки этой ковариационной матрицы оказываются неработоспособными, неприменимыми в условиях ОЛММР. Действительно, поскольку (4.25) справедливо и для ОЛММР,

имеем:

$$\begin{aligned}
 \hat{\Theta}_{\text{МНК}} - \Theta &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \varepsilon, \\
 \Sigma_{\hat{\Theta}_{\text{МНК}}} &= \mathbf{E}[(\hat{\Theta}_{\text{МНК}} - \Theta)(\hat{\Theta}_{\text{МНК}} - \Theta)^T] = \\
 &= \mathbf{E}[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \varepsilon \varepsilon^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}] = \\
 &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{E}(\varepsilon \varepsilon^T) \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} = \\
 &= \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \Sigma_0 \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}.
 \end{aligned} \tag{5.5}$$

Сравнение правой части (5.5) с формулой (4.33), справедливой только в рамках КЛММП, позволяет установить факт существенного отличия *истинных* характеристик точности оценивания с помощью МНК параметров ОЛММП от тех, которые мы получили бы, воспользовавшись формулой (4.33) (ниже мы продемонстрируем существенность подобного расхождения на некоторых примерах). Мы не обсуждаем здесь свойств оценки $\hat{\sigma}^2$, поскольку в *общей* схеме ОЛММП параметр σ^2 не имеет четкой вероятностной интерпретации (в отличие от КЛММП (4.1') и ЛММП с автокоррелированными гомоскедастичными остатками, — см. (5.4'), где параметр σ^2 имеет смысл дисперсии случайных регрессионных остатков). Более детальный анализ (5.5) и ОЛММП показал, что *оценки $\hat{\Theta}_{\text{МНК}}$ в условиях ОЛММП теряют свои оптимальные свойства* и что можно предложить другие оценки — так называемые *оценки обобщенного метода наименьших квадратов $\hat{\Theta}_{\text{ОМНК}}$* , которые будут наиболее эффективными в смысле теоремы Гаусса–Маркова.

Оценки по обобщенному методу наименьших квадратов определяются соотношениями

$$\hat{\Theta}_{\text{ОМНК}} = (\mathbf{X}^T \Sigma_0^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma_0^{-1} \mathbf{Y}. \tag{5.6}$$

Можно доказать (*Aitken A. C. On Least-Squares and Linear Combinations of Observations. Proc. Royal Soc., Edinburgh, 1934, vol. 55, pp. 42–48*), что в *классе линейных несмещенных оценок параметров Θ модели (5.1) оценки $\hat{\Theta}_{\text{ОМНК}}$, определенные соотношениями (5.6), являются оптимальными в смысле теоремы Гаусса–Маркова (см. (4.37))*.

Проведем доказательство этого утверждения.

Несмещенность оценок (5.6) устанавливается по той же схеме, что и несмещенность $\hat{\Theta}_{\text{МНК}}$. В частности:

$$\begin{aligned}
 \hat{\Theta}_{\text{ОМНК}} &= (\mathbf{X}^T \Sigma_0^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma_0^{-1} (\mathbf{X} \Theta + \varepsilon) = \\
 &= (\mathbf{X}^T \Sigma_0^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma_0^{-1} \mathbf{X} \Theta + (\mathbf{X}^T \Sigma_0^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma_0^{-1} \varepsilon = \\
 &= \Theta + (\mathbf{X}^T \Sigma_0^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma_0^{-1} \varepsilon.
 \end{aligned} \tag{5.7}$$

Применение операции математического ожидания к левой и правой частям (5.7) с учетом $\mathbf{E} \varepsilon = \mathbf{0}_n$ дает:

$$\mathbf{E} \hat{\Theta}_{\text{ОМНК}} = \Theta. \tag{5.8}$$

С целью доказательства оптимальных свойств оценок (5.6) воспользуемся знанием матрицы Σ_0 для того, чтобы преобразовать исходные данные (4.1а)–(4.3) к виду, отвечающему требованиям *классической* (а не обобщенной) модели. Это значит, что после произведенного преобразования случайные остатки модели $\epsilon_{\text{пр}}$ должны удовлетворять свойствам гомоскедастичности и взаимной некоррелированности, то есть

$$\Sigma_{\epsilon_{\text{пр}}} = \sigma^2 \mathbf{I}_n.$$

Выполнить необходимое преобразование позволяет известный результат из матричной алгебры (см. Приложение 3), в соответствии с которым всякая положительно определенная симметричная $(m \times m)$ -матрица $\mathbf{A}$ допускает представление в виде

$$\mathbf{A} = \mathbf{C}\mathbf{C}^\top, \quad (5.9)$$

где $\mathbf{C}$ — некоторая невырожденная $(m \times m)$ -матрица (представление (5.9) не единственно, однако для нас это не имеет значения). Воспользуемся этим, чтобы представить в виде (5.9) матрицу Σ_0 . Итак, существует такая невырожденная $(n \times n)$ -матрица $\mathbf{C}$, что

$$\Sigma_0 = \mathbf{C}\mathbf{C}^\top. \quad (5.9')$$

Заметим, что тогда

$$\mathbf{C}^{-1}\Sigma_0(\mathbf{C}^{-1})^\top = \mathbf{I}_n, \quad (5.10)$$

$$(\mathbf{C}^{-1})^\top \mathbf{C}^{-1} = \Sigma_0^{-1} \quad (5.11)$$

(соотношение (5.10) получается из (5.9') домножением левой и правой частей последнего на матрицу $\mathbf{C}^{-1}$ слева и матрицу $(\mathbf{C}^\top)^{-1} = (\mathbf{C}^{-1})^\top$ справа, а (5.11) получается из (5.9') с учетом правила обращения произведения квадратных невырожденных матриц $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$).

Теперь вернемся к ОЛММП (5.1) и, домножив все члены первого соотношения этой модели слева на матрицу $\mathbf{C}^{-1}$, получим:

$$\mathbf{Y}_{\text{пр}} = \mathbf{X}_{\text{пр}}\Theta + \epsilon_{\text{пр}}, \quad (5.12)$$

где

$$\mathbf{Y}_{\text{пр}} = \mathbf{C}^{-1}\mathbf{Y}, \quad \mathbf{X}_{\text{пр}} = \mathbf{C}^{-1}\mathbf{X} \quad \text{и} \quad \epsilon_{\text{пр}} = \mathbf{C}^{-1}\epsilon. \quad (5.13)$$

Нетрудно проверить, что (5.12) удовлетворяет всем требованиям *классической* модели (4.1'). Для этого достаточно убедиться в том, что $\Sigma_{\epsilon_{\text{пр}}} = \mathbf{E}(\epsilon_{\text{пр}}\epsilon_{\text{пр}}^\top) = \sigma^2 \mathbf{I}_n$. Действительно,

$$\begin{aligned} \Sigma_{\epsilon_{\text{пр}}} &= \mathbf{E}(\epsilon_{\text{пр}}\epsilon_{\text{пр}}^\top) = \mathbf{E}(\mathbf{C}^{-1}\epsilon\epsilon^\top(\mathbf{C}^{-1})^\top) = \mathbf{C}^{-1}\mathbf{E}(\epsilon\epsilon^\top)(\mathbf{C}^{-1})^\top = \\ &= \mathbf{C}^{-1}\sigma^2\Sigma_0(\mathbf{C}^{-1})^\top = \sigma^2[\mathbf{C}^{-1}\Sigma_0(\mathbf{C}^{-1})^\top] = \sigma^2\mathbf{I}_n \end{aligned}$$

(при переходе к правой части соотношения было использовано равенство (5.10)).

Следовательно, в соответствии с ранее полученными для КЛММР результатами наилучшей в классе всех линейных (относительно Y) несмещенных оценок параметра Θ будет оценка

$$\hat{\Theta} = (X_{\text{пр}}^T X_{\text{пр}})^{-1} X_{\text{пр}}^T Y_{\text{пр}}, \quad (5.14)$$

ковариационная матрица которой определяется формулой

$$\Sigma_{\hat{\Theta}} = \sigma^2 (X_{\text{пр}}^T X_{\text{пр}})^{-1}. \quad (5.15)$$

Вернемся к *исходным* наблюдениям X и Y . Оценки (5.14), выраженные в исходных наблюдениях, назовем *оценками обобщенного метода наименьших квадратов* $\hat{\Theta}_{\text{омнк}}$:

$$\begin{aligned} \hat{\Theta}_{\text{омнк}} &= [X^T (C^{-1})^T (C^{-1} X)]^{-1} X^T (C^{-1})^T C^{-1} Y = \\ &= (X^T \Sigma_0^{-1} X)^{-1} X^T \Sigma_0^{-1} Y \end{aligned} \quad (5.14')$$

$$\Sigma_{\hat{\Theta}_{\text{омнк}}} = \sigma^2 (X^T \Sigma_0^{-1} X)^{-1}. \quad (5.15')$$

Как известно, МНК-оценки (5.14) являются по построению результатом минимизации по Θ критерия $Q(\Theta) = (Y_{\text{пр}} - X_{\text{пр}} \Theta)^T (Y_{\text{пр}} - X_{\text{пр}} \Theta)$ (см. (4.16)). Выпишем вид этого критерия в *исходных* наблюдениях ОЛММР:

$$\begin{aligned} Q(\Theta) &= [C^{-1}(Y - X\Theta)]^T [C^{-1}(Y - X\Theta)] = \\ &= (Y - X\Theta)^T (C^{-1})^T C^{-1} (Y - X\Theta) = \\ &= (Y - X\Theta)^T \Sigma_0^{-1} (Y - X\Theta). \end{aligned} \quad (5.16)$$

Поэтому ОМНК-оценки $\hat{\Theta}_{\text{омнк}}$ могут быть определены как «точка минимума» обобщенного критерия (5.16).

Анализ вариации результирующего показателя y и выборочный коэффициент детерминации $\hat{R}_{y.X}^2$ в ОЛММР. В п. 4.3 было получено разложение общей вариации результирующего показателя на две составляющие: вариацию, объясненную изменчивостью функции регрессии (то есть изменением значений объясняющих переменных), и вариацию случайных остатков, — справедливое в рамках *классической* ЛММР (см. (4.52) и (4.52')). Следовательно, в используемых в данном пункте обозначениях мы, *казалось бы*, можем записать это соотношение в виде

$$(Y_{\text{пр}} - \bar{Y}_{\text{пр}})^T (Y_{\text{пр}} - \bar{Y}_{\text{пр}}) = (X_{\text{пр}} \hat{\Theta}_{\text{омнк}} - \bar{Y}_{\text{пр}})^T (X_{\text{пр}} \hat{\Theta}_{\text{омнк}} - \bar{Y}_{\text{пр}}) + \hat{\varepsilon}_{\text{пр}}^T \hat{\varepsilon}_{\text{пр}}. \quad (5.17)$$

Тогда, возвращаясь с помощью (5.13) к *исходным* наблюдениям и параметрам *обобщенной* ЛММР, из (5.17) мы бы получили:

$$(Y - \bar{Y})^T \Sigma_0^{-1} (Y - \bar{Y}) = (X \hat{\Theta}_{\text{омнк}} - \bar{Y})^T \Sigma_0^{-1} (X \hat{\Theta}_{\text{омнк}} - \bar{Y}) + \hat{\varepsilon}^T \Sigma_0^{-1} \hat{\varepsilon}. \quad (5.17')$$

Откуда, определяя, как и прежде, выборочный коэффициент детерминации $\hat{R}_{y.X}^2$ как долю общей вариации признака y , объясненную изменением функции регрессии (или, что то же, изменением значений объясняющих переменных X), имеем:

$$\hat{R}_{y.X}^2 = 1 - \frac{\hat{\epsilon}^\top \Sigma_0^{-1} \hat{\epsilon}}{(Y - \bar{Y})^\top \Sigma_0^{-1} (Y - \bar{Y})}. \quad (5.18)$$

Однако в действительности мы не можем гарантировать справедливости соотношений (5.17) и, соответственно, (5.17'), а следовательно, — и ограниченности значений коэффициента $\hat{R}_{y.X}^2$ (определенного соотношением (5.18)) в пределах интервала $[0; 1]$. Причина в том, что базовое разложение (4.52') выводилось в предположении *наличия свободного члена* в анализируемой ЛММР. В то же время из построения видно, что мы не можем гарантировать присутствия свободного члена в модели (5.12). Поэтому определенное соотношением (5.18) коэффициент детерминации $\hat{R}_{y.X}^2$ иногда используется в ОЛММР лишь как приближенная эвристическая характеристика.

Оценка параметра σ^2 в ОЛММР. Выше упоминалось о том, что в отличие от КЛММР в рамках *обобщенной* модели мы уже не можем интерпретировать (без дополнительных предположений о структуре матрицы Σ_0) параметр σ^2 как величину дисперсии регрессионных остатков. Однако он остается и в рамках ОЛММР неизвестным параметром модели, который требуется статистически оценить.

С этой целью мы воспользуемся результатами решения этой задачи для *классической* модели (см. п. 4.2.3) и, в частности, соотношением (4.31) и свойствами (4.43)–(4.44). В обозначениях и терминах ОЛММР эти соотношения и свойства применимы по отношению к *преобразованным* данным $X_{\text{пр}}$ и $Y_{\text{пр}}$, что после обратного перехода к исходным данным X и Y *обобщенной* модели с помощью формул (5.13) дает:

- *оценка*

$$\hat{\sigma}^2 = \frac{1}{n - p - 1} (Y - X\hat{\Theta}_{\text{ОМНК}})^\top \Sigma_0^{-1} (Y - X\hat{\Theta}_{\text{ОМНК}}) \quad (5.19)$$

является несмещенной оценкой неизвестного параметра σ^2 модели (5.1);

- *статистика $(n - p - 1)\hat{\sigma}^2/\sigma^2$ подчиняется χ^2 -распределению с $n - p - 1$ степенями свободы, а оценки $\hat{\sigma}^2$ и $\hat{\Theta}_{\text{ОМНК}}$, определенные соотношениями соответственно (5.19) и (5.6), являются статистически независимыми.*

* * *

Завершая описание *общих* сведений об обобщенном методе наименьших квадратов, обратим внимание читателя на *проблему практической*

реализации этого метода. Ведь во всех основных формулах ОМНК (см. соотношения (5.6), (5.15'), (5.5)) мы исходим из того, что *ковариационная матрица остатков известна* (правда, с точностью до неизвестного постоянного множителя σ^2 , несмещенную оценку для которого мы строить умеем (см. (5.19)). Однако ситуации, когда матрица Σ нам известна заранее, крайне редки в практике эконометрического моделирования. Если же включать *формально все* элементы матрицы Σ_0 в множество параметров модели (5.1), оцениваемых по выборке (4.1a)–(4.3), то мы столкнемся с *неразрешимой* (без дополнительных априорных допущений о структуре матрицы Σ_0) статистической задачей, поскольку число неизвестных параметров только в матрице Σ_0 составляет $n(n+1)/2$, что превосходит объем располагаемых нами выборочных данных. Поэтому для того, чтобы от *общего* описания ОМНК перейти к *практически реализуемому* обобщенному методу наименьших квадратов, нам придется вводить дополнительные априорные ограничения на структуру матрицы Σ_0 , что и будет сделано в следующих двух параграфах (см. также п. 5.5).

5.3 ОЛММР с гетероскедастичными остатками

5.3.1 Общие сведения. Взвешенный МНК

В п. 5.1 описана линейная модель регрессии с гетероскедастичными регрессионными остатками как один из частных случаев ОЛММР (5.5), в котором ковариационная матрица остатков Σ_ε имеет вид (5.2). Это означает, что регрессионные остатки, как и в классической модели (4.1'), взаимно некоррелированы (то есть $E(\varepsilon_i \varepsilon_j) = 0$ для $i \neq j$), но, в отличие от условий, (4.1') не обладают свойством гомоскедастичности (то есть, вообще говоря, $D\varepsilon_i \neq D\varepsilon_j$ при $i \neq j$) или, как принято говорить, *они гетероскедастичны* (неоднородны по характеристике случайного разброса). Как было уже упомянуто в п. 5.1, гетероскедастичность остатков — вполне естественная ситуация для анализа *пространственных* выборок: для них чаще более оправданной является гипотеза постоянства *относительного* разброса остатков, выраженная, например, в виде

$$D(\varepsilon | X_i) = \sigma^2 [E(y | X_i)]^2 \quad (5.20)$$

или

$$D(\varepsilon | X_i) = \sigma^2 (a + bx_i^{(j)})^2, \quad (5.20')$$

где $x^{(j)}$ — та объясняющая переменная, пропорционально квадрату которой возрастает условная дисперсия регрессионных остатков.

Кстати, данные таблицы 5.1 и построенный на основании этих данных рис. 5.1 свидетельствуют о том, что именно соотношение (5.20') ока-

зывается подходящей формой допущения относительно поведения условной дисперсии $D(\varepsilon | x_i^0)$ в примере 2.1.

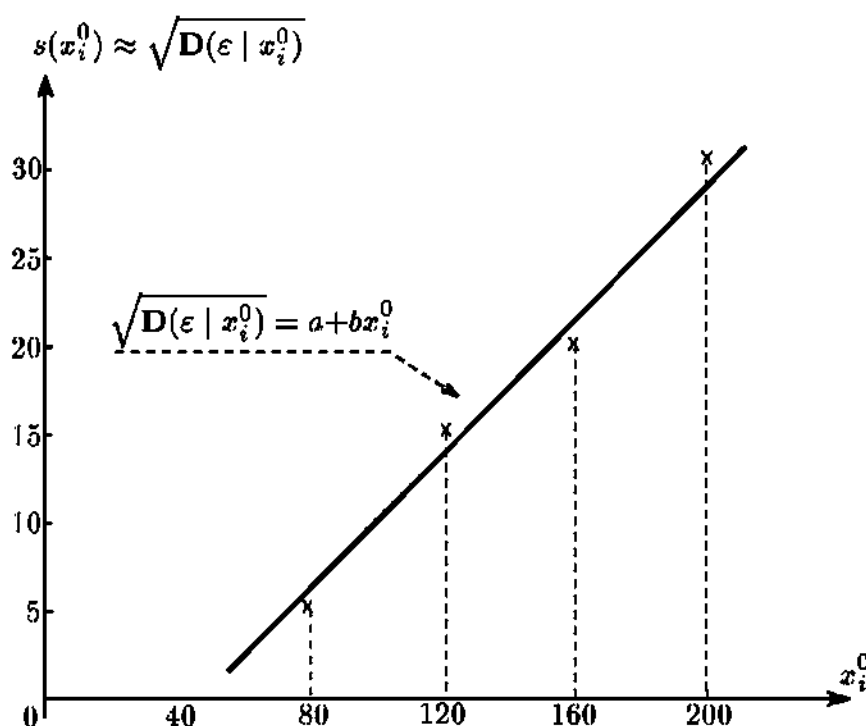


Рис. 5.1. Зависимость условного среднеквадратического отклонения остатков от значения объясняющей переменной в примере 2.1

Попытаемся конкретизировать общие формулы ОМНК для данного частного случая обобщенной линейной модели регрессии. С этой целью, во-первых, выпишем конкретный вид $(n \times n)$ -матрицы C вспомогательного преобразования (5.13). Отправляясь от вытекающего из (5.2) вида матрицы Σ_0 , определим

$$C = \begin{pmatrix} 1/\sqrt{\lambda_1} & & & 0 \\ & 1/\sqrt{\lambda_2} & & \\ & & \ddots & \\ 0 & & & 1/\sqrt{\lambda_n} \end{pmatrix}. \quad (5.21)$$

Легко проверить, что в этом случае выполняются соотношения (5.9'), (5.10) и (5.11). Кроме того, определенный соотношением (5.16) критерий ОМНК при матрице Σ_ε вида (5.2) представим в форме

$$Q(\Theta) = (Y - X\Theta)^\top \Sigma_0^{-1} (Y - X\Theta) = \sum_{i=1}^n \lambda_i (y_i - \theta_0 - \theta_1 x_i^{(1)} - \dots - \theta_p x_i^{(p)})^2. \quad (5.22)$$

Поэтому ОМНК в частном случае обобщенной ЛММР с гетероскедастичными остатками часто называют *методом взвешенных наименьших квадратов* или *взвешенным методом наименьших квадратов* (роль

«весов» в нем играют, как это видно из (5.22), диагональные элементы матрицы Σ_0^{-1} .

Соответственно несмещенная оценка $\hat{\sigma}^2$ неизвестного значения параметра σ^2 в выражении ковариационной матрицы остатков Σ_ε (см. (5.1)), определяемая формулой (5.19), в данном случае имеет вид

$$\hat{\sigma}^2 = \frac{1}{n - p - 1} \sum_{i=1}^n \lambda_i (y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i^{(1)} - \dots - \hat{\theta}_p x_i^{(p)})^2. \quad (5.19')$$

При определении ОМНК-оценок $\hat{\Theta}_{\text{ОМНК}}$ для параметров регрессии Θ и ковариационной матрицы $\Sigma_{\hat{\Theta}_{\text{ОМНК}}}$ следует воспользоваться, соответственно, формулами (5.6) и (5.15') с заменой в формуле (5.15') неизвестной величины σ^2 ее оценкой (5.19') и с учетом того, что диагональный вид матрицы Σ_0 (а следовательно, и Σ_0^{-1}) существенно упрощает сам процесс вычисления.

5.3.2 Сравнение ОМНК- и МНК-оценок в моделях регрессии с гетероскедастичными остатками

В общем виде ответ на вопрос, какие из оценок (ОМНК- и МНК-) лучше и почему, мы уже имеем: хотя и те, и другие оценки являются состоятельными и несмещенными, ОМНК-оценки оказываются более эффективными, более точными. Проиллюстрируем этот тезис на примере 2.1. Чтобы несколько упростить техническую сторону анализа, перенесем начало координат по оси абсцисс в точку $x = 40$ (ден. ед.), то есть будем измерять значения объясняющей переменной в новой шкале $O\tilde{x}$, где $\tilde{x}$ связано с x простым соотношением $\tilde{x} = x - 40$.

При таком преобразовании объясняющей переменной ее исходные значения автоматически уменьшаются на 40 ден. ед., а гипотетический характер зависимости условной дисперсии остатков $D(\varepsilon | x)$ от x , как это видно из рис. 5.1, может быть описан упрощенным вариантом соотношения (5.20'), а именно формулой

$$D(\varepsilon | x) = \sigma^2 x^2 \quad (5.23)$$

(в целях упрощения обозначений мы не будем с этого момента писать волну над x).

Для дальнейшего анализа этого примера выпишем основные формулы ОМНК в схеме гетероскедастичных остатков для случая одной объясняющей переменной.

Поскольку при $p = 1$

$$\mathbf{X}^\top \Sigma_0^{-1} \mathbf{X} = \begin{pmatrix} \sum_{i=1}^n \lambda_i & \sum_{i=1}^n \lambda_i x_i \\ \sum_{i=1}^n \lambda_i x_i & \sum_{i=1}^n \lambda_i x_i^2 \end{pmatrix},$$

то в соответствии с (5.6) получаем уравнения для определения ОМНК оценок параметров θ_0 и θ_1 :

$$\begin{cases} \theta_0 \sum_{i=1}^n \lambda_i + \theta_1 \sum_{i=1}^n \lambda_i x_i = \sum_{i=1}^n \lambda_i y_i; \\ \theta_0 \sum_{i=1}^n \lambda_i x_i + \theta_1 \sum_{i=1}^n \lambda_i x_i^2 = \sum_{i=1}^n \lambda_i x_i y_i. \end{cases}$$

Откуда, в частности, и следуют использованные нами ранее формулы (2.7''') для ОМНК-оценок $\hat{\theta}_{0.омнк}$ и $\hat{\theta}_{1.омнк}$.

Приведенный выше конкретный вид матрицы $X^T \Sigma_0^{-1} X$ и формула (5.15') позволяют выписать дисперсию оценки $\hat{\theta}_{1.омнк}$:

$$D\hat{\theta}_{1.омнк} = E(\hat{\theta}_{1.омнк} - \theta_1)^2 = \sigma^2 \frac{\sum_{i=1}^n \lambda_i}{\left(\sum_{i=1}^n \lambda_i\right) \left(\sum_{i=1}^n \lambda_i x_i^2\right) - \left(\sum_{i=1}^n \lambda_i x_i\right)^2}. \quad (5.24)$$

Общая формула (4.19) для МНК-оценок и ее частный случай для $p = 1$ (см. (4.15)) дают нам способ вычисления *обычных* МНК-оценок $\hat{\theta}_{0.мнк}$ и $\hat{\theta}_{1.мнк}$. Воспользовавшись формулой (5.5), мы можем выписать, например, дисперсию оценки $\hat{\theta}_{1.мнк}$ и затем сравнить ее с дисперсией оценки $\hat{\theta}_{1.омнк}$. Итак, из (5.5) для случая $p = 1$ следует:

$$\begin{aligned} D\hat{\theta}_{1.мнк} &= E(\hat{\theta}_{1.мнк} - \theta_1)^2 = \\ &= \sigma^2 \frac{\left(\sum_{i=1}^n \frac{1}{\lambda_i}\right) \left(\sum_{i=1}^n x_i\right)^2 - 2n \left(\sum_{i=1}^n \frac{x_i}{\lambda_i}\right) \left(\sum_{i=1}^n x_i\right) + n^2 \left(\sum_{i=1}^n \frac{x_i^2}{\lambda_i}\right)}{\left[n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2\right]^2}. \end{aligned} \quad (5.25)$$

Нам известен общий результат теории ОМНК, в соответствии с которым при известных значениях λ_i в условиях *обобщенной* линейной модели предпочтительнее оценивать Θ по формуле (4.15), чем по формуле (4.19), поскольку ОМНК-оценки, найденные по формуле (4.15), являются *наилучшими* линейными несмещенными оценками. Попробуем конкретно (численно) оценить потерю в эффективности оценивания при использовании оценок $\hat{\Theta}_{мнк}$ вместо $\hat{\Theta}_{омнк}$ в нашем примере 2.1 (см. пп. 2.1 и 5.1) при гипотезе (5.23) относительно характера гетероскедастичности регрессионных остатков. Эта гипотеза, в частности, означает, что числа λ_i , с помощью которых задается матрица Σ_0 (см. (5.1)–(5.2)), определяются соотношением:

$$\lambda_i = \frac{1}{x_i^2},$$

что в нашем конкретном примере 2.1 означает следующее:

$$\left\{ \begin{array}{l} \lambda_1 = \lambda_2 = \dots = \lambda_{10} = \frac{1}{(x_1^0)^2} = \frac{1}{40^2}; \\ \lambda_{11} = \lambda_{12} = \dots = \lambda_{20} = \frac{1}{(x_2^0)^2} = \frac{1}{80^2}; \\ \lambda_{21} = \lambda_{22} = \dots = \lambda_{30} = \frac{1}{(x_3^0)^2} = \frac{1}{(120)^2}; \\ \lambda_{31} = \lambda_{32} = \dots = \lambda_{40} = \frac{1}{(x_4^0)^2} = \frac{1}{(160)^2}. \end{array} \right.$$

Формулы (5.24) и (5.25) с учетом (5.23) принимают соответственно вид:

$$D\hat{\theta}_{1.омнк} = \sigma^2 \frac{\sum_{i=1}^n \left(\frac{1}{x_i^2}\right)}{n \sum_{i=1}^n \left(\frac{1}{x_i^2}\right) - \left[\sum_{i=1}^n \left(\frac{1}{x_i}\right)\right]^2}, \quad (5.24')$$

$$D\hat{\theta}_{1.мнк} = \sigma^2 \frac{\left(\sum_{i=1}^n x_i^2\right) \left(\sum_{i=1}^n x_i\right)^2 - 2n \left(\sum_{i=1}^n x_i^3\right) \left(\sum_{i=1}^n x_i\right) + n^2 \left(\sum_{i=1}^n x_i^4\right)}{\left[n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2\right]^2}. \quad (5.25')$$

Вычисление $D\hat{\theta}_{1.омнк}$ и $D\hat{\theta}_{1.мнк}$ с использованием формул (5.24') и (5.25') по данным примера 2.1 дает:

$$\begin{aligned} D\hat{\theta}_{1.омнк} &= 0,10\sigma^2, \\ D\hat{\theta}_{1.мнк} &= 0,17\sigma^2. \end{aligned}$$

Отсюда следует, что $D\hat{\theta}_{1.омнк}/D\hat{\theta}_{1.мнк} = 0,10\sigma^2/0,17\sigma^2 = 0,59$, то есть эффективность обыкновенного метода наименьших квадратов составляет в данном конкретном случае всего лишь 59% эффективности обобщенного метода наименьших квадратов.

5.3.3 Статистическая проверка гетероскедастичности остатков. Практически реализуемый МНК

Приступая к анализу ОЛММР, исследователь должен, во-первых, ответить на все вопросы, связанные со спецификацией модели, и, во-вторых, найти пути к определению ковариационной матрицы остатков $\Sigma_\varepsilon = \sigma^2 \Sigma_0$ (либо на основе априорных теоретических сведений, что бывает достаточно редко, либо с помощью некоторых специальных способов статистического оценивания на базе исходных данных X, Y).

Из вопросов, связанных со спецификацией модели, на стадии, когда уже определены конкретные прикладные цели моделирования, анализируемые переменные и общий вид зависимости (в нашем случае, то есть в рамках ОЛММР, он постулируется *линейным по X*), *остаются только вопросы, касающиеся вероятностной природы регрессионных остатков*. Это значит, что мы должны статистически проверить, находимся ли мы, действительно, в условиях гетероскедастичности остатков ε .

В специальной литературе [Johnston, DiNardo (1997)], [Greene (2000)], [Pindyck, Rubinfeld (1976)] описываются различные методы проверки гипотезы

$$H_0: D(\varepsilon | X) = \text{const} \quad (5.26)$$

при альтернативе, состоящей в том, что остатки $\varepsilon_1, \dots, \varepsilon_n$ гетероскедастичны. Опишем три таких критерия.

Первый из них (*тест Уайта*) выбран нами потому, что он является наиболее общим в том смысле, что не требует никаких дополнительных предположений относительно природы гипотетичной гетероскедастичности остатков, и в частности, относительно параметрической структуры ковариационной матрицы остатков Σ_ε . Однако он является *неконструктивным*, поскольку при отклонении гипотезы о гомоскедастичности остатков *он не предоставляет средств для построения практически реализуемого ОМНК*.

Два других критерия, как раз, позволяют получать практически реализуемый ОМНК в условиях гетероскедастичности остатков. Правда, *критерий Бреуша–Пагана* требует, при этом, определенной параметрической спецификации зависимости дисперсии $D(\varepsilon | X_i)$ от X_i , а *критерий Бартлетта* приспособлен только для сгруппированных (по X) данных и корректен при *нормально* распределенных (внутри каждой группы) остатках.

1) **Общий тест Уайта** (см. White H. F. *Heteroscedasticity — Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity*. — *Econometrica*, vol. 48 (1980) pp. 817 — 838). Тест называется *общим*, поскольку в качестве альтернативы гипотезе гомоскедастичности (5.26) предполагает регрессионную зависимость квадратов МНК-оцененных невязок $\hat{\varepsilon}_i^2$ самого общего вида. А именно, в качестве объясняющих переменных в этой вспомогательной регрессии выступают все объясняющие переменные исходной (анализируемой) ЛММР, их квадраты и всевозможные их попарные произведения, то есть:

$$\hat{\varepsilon}_i^2 = \beta_0 + \sum_{j=1}^p \beta_j x_i^{(j)} + \sum_{q=1}^p \sum_{l=1}^p \beta_{ql} \cdot x_i^{(q)} \cdot x_i^{(l)} + \delta_i. \quad (5.27)$$

$(q \leq l)$

Поэтому предлагается следующая последовательность действий для реализации теста:

- 1) Вычисляются обычные МНК-оценки $\hat{\Theta}_{\text{МНК}}$ параметров Θ модели (5.1):

$$\hat{\Theta}_{\text{МНК}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}.$$

- 2) Вычисляются МНК-оценки $\hat{\varepsilon}_{\text{МНК}}$ остатков ε модели (5.1): $\hat{\varepsilon}_{\text{МНК}} = \mathbf{Y} - \mathbf{X}\hat{\Theta}_{\text{МНК}}$.

- 3) Строятся МНК-оценки параметров $\beta_0, \beta_1, \dots, \beta_p, \beta_{ql}$ ($q, l = 1, 2, \dots, p; q \leq l$) модели (5.27).

- 4) Вычисляется коэффициент детерминации $\hat{R}_{\varepsilon^2 Z}^2$ (см. формулу (4.54)), где N -мерный вектор объясняющих переменных Z объединяет в себе все регрессоры модели (5.27), не считая свободного члена β_0 (очевидно $N = p(p+3)/2$).

- 5) Доказано, что в условиях справедливости гипотезы гомоскедастичности остатков (5.26) критическая статистика $\gamma = n\hat{R}_{\varepsilon^2 Z}^2$ должна вести себя асимптотически (то есть при $n \rightarrow \infty$) как $\chi^2(N)$ — случайная величина; «ненормально большие», в условиях справедливости гипотезы (5.26), значения статистики γ сигнализируют о наличии статистически значимой связи в регрессии (5.27), то есть — о гетероскедастичности остатков.

- 6) Поэтому, если $n\hat{R}_{\varepsilon^2 Z}^2 > \chi_\alpha^2(N)$, то гипотеза (5.26) отвергается с вероятностью ошибки, равной заданному значению α (здесь $\chi_\alpha^2(N)$, как обычно, 100 α %-ная точка $\chi^2(N)$ -распределения).

2) **Критерий Бреуша–Пагана** (см. *Breusch T. and Pagan A. A Simple Test for Heteroscedasticity and Random Coefficient Variation. — Econometrica, vol. 47 (1979), pp. 1287 — 1294*). Реализация этого критерия требует более определенной (чем в критерии Уайта) спецификации зависимости дисперсий остатков $D\varepsilon_i$ от ряда вспомогательных переменных $z_i^{(1)}, z_i^{(2)}, \dots, z_i^{(k)}$, которые, в частности, могут комплектоваться и из числа объясняющих переменных анализируемой ЛММР (что обычно и делается в эконометрической практике). Итак, основываясь на определенных содержательных соображениях и анализе поведения МНК-оцененных остатков $\hat{\varepsilon}_i$ в зависимости от $y_i, x_i^{(j)}$ и, если есть такая возможность, от значений некоторых поддающихся измерению экзогенных переменных $z_i^{(l)}$, не являющихся объясняющими переменными рассматриваемой ЛММР, формируется состав вспомогательных переменных $Z = (1, z^{(1)}, \dots, z^{(k)})^\top$ и выбирается общий вид функции $h(u)$, определяющих характер гипотетической зависимости

$$D\varepsilon_i = \sigma^2 \cdot h(\beta_0 + \beta_1 z_i^{(1)} + \dots + \beta_k \cdot z_i^{(k)}). \quad (5.28)$$

Тогда гипотеза о гомоскедастичности остатков (5.26) принимает вид:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0, \quad (5.26')$$

так как в этом случае $D\varepsilon_i = \sigma^2 \cdot h(\beta_0)$ (правая часть от i не зависит). Функцию $h(u)$ выбирают таким образом, чтобы она была непрерывно дифференцируемой, принимала бы только положительные значения и удовлетворяла бы условию нормировки: $h(0) = 1$.

Построение критерия множителей Лагранжа (см. п. 4.6) для проверки гипотезы (5.26') приводит к следующей процедуре:

- 1) Вычисляются МНК-оценки $\hat{\varepsilon}_{\text{МНК}}$ анализируемой ЛММР (5.1) (см. первые два пункта процедуры Уайта).
- 2) Вычисляются значения $\tilde{y}_i = \hat{\varepsilon}_i^2 / \hat{\sigma}^2$ ($i = 1, 2, \dots, n$), где $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2$.
- 3) Оцениваются (с помощью МНК) параметры $\beta_0, \beta_1, \dots, \beta_k$ регрессии (вообще говоря, – нелинейной)

$$\tilde{y}_i = h(\beta_0 + \beta_1 z_i^{(1)} + \dots + \beta_k z_i^{(k)}) + \delta_i,$$

где δ_i – остатки, удовлетворяющие всем требованиям КЛММР, то есть определяются значения $(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k)$ как решения оптимизационной задачи

$$(\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k) = \arg \min_{\beta_0, \dots, \beta_k} \sum_{i=1}^n (\tilde{y}_i - h(\beta_0 + \beta_1 z_i^{(1)} + \dots + \beta_k z_i^{(k)}))^2.$$

- 4) Бреуш и Паган показали, что LM-статистика в задаче проверки гипотезы (5.26') имеет вид

$$LM = \frac{1}{2} \sum_{i=1}^n (h_i - \bar{h})^2 \quad (5.29)$$

(где $h_i = h(\hat{\beta}_0 + \hat{\beta}_1 z_i^{(1)} + \dots + \hat{\beta}_k z_i^{(k)})$, а $\bar{h} = \frac{1}{n} \sum_{i=1}^n h_i$) и должна подчиняться (асимптотически по $n \rightarrow \infty$, в условиях справедливости гипотезы (5.26)) $\chi^2(k)$ -распределению.

Поэтому, если оказалось, что $LM \leq \chi_{\alpha}^2(k)$, то гипотеза о гомоскедастичности остатков модели (5.1) не отвергается и последующий анализ проводится в рамках КЛММР.

Если же $LM > \chi_{\alpha}^2(k)$, то гипотеза H_0 отвергается и с моделью работают в рамках ОЛММР с гетероскедастичными остатками, в

которой в качестве матрицы Σ_0 используют ее оценку вида

$$\hat{\Sigma}_0 = \begin{pmatrix} h_1 & & \Theta \\ & h_2 & \\ & & \ddots \\ \Theta & & & h_n \end{pmatrix}$$

Это и есть практически реализуемый ОМНК!

З а м е ч а н и е 1. Иногда используют простейший вариант теста, в котором $h(u) = u$, то есть $D\varepsilon_i = \sigma^2 \cdot (\beta_0 + \beta_1 z_i^{(1)} + \dots + \beta_k z_i^{(k)})$. В этом случае $LM = \frac{1}{2}[\tilde{Y}^T Z(Z^T Z)^{-1} X \times X^T \tilde{Y}]$, где $\tilde{Y}$ - $n \times 1$ - вектор-столбец значений $\tilde{y}_i$, а Z - $(k+1)$ - матрица наблюдаемых значений вспомогательных объясняющих переменных. Однако при таком выборе функции $h(u)$ некоторые из значений h_i могут оказаться отрицательными.

З а м е ч а н и е 2. Весьма подробно исследован случай *мультипликативной гетероскедастичности*, при котором $h(u) = e^u$ (см. Godfrey L. Testing for Multiplicative heteroscedasticity. - Journal of Econometrics, vol. 8 (1978), pp. 227-236).

3) Случай группированных данных и больших выборок. Это — наиболее простой случай. Предполагается, что объем n имеющихся исходных данных (4.1a)–(4.3) достаточно велик, и при этом, выборка может быть разбита на определенное количество (k) подвыборок объемов, соответственно, $n_1, n_2, \dots, n_k$ ($n_1 + n_2 + \dots + n_k = n$) таким образом, что внутри каждой из подвыборок значения объясняющих переменных либо совпадают (как в нашем примере 2.1, в котором $k = 4$, а $n_1 = n_2 = n_3 = n_4 = 10$), либо принадлежат одному интервалу (гиперпараллелепипеду) группирования. Если среднюю точку j -го интервала группирования (или общее значение объясняющей переменной в j -й подвыборке) обозначить X_j^0 , то проверка гипотезы (5.26) сведется к построению статистического критерия проверки гипотезы об однородности дисперсий

$$H_0: D(\varepsilon | X_1^0) = D(\varepsilon | X_2^0) = \dots = D(\varepsilon | X_k^0)$$

по величинам соответствующих выборочных дисперсий

$$s_j^2 = \frac{1}{n_j - 1} \sum_{i=1}^{n_j} (y_{ji} - \bar{y}_j)^2, \quad j = 1, 2, \dots, k,$$

где y_{ji} — значение результирующей переменной в i -м наблюдении j -й подвыборки, а $\bar{y}_j = \sum_{i=1}^{n_j} y_{ji} / n_j$.

В качестве такого критерия может быть использован, например, критерий Бартлетта, который позволяет проверять гипотезу об однородности дисперсий в нескольких (k) *нормальных* генеральных совокупностях

(это означает, что мы должны постулировать нормальность регрессионных остатков в нашей модели). Критическая статистика предложенного Бартлеттом теста имеет вид

$$\gamma = q \sum_{j=1}^k (n_j - 1) \ln \left(\frac{s_j^2}{s^2} \right),$$

где

$$s^2 = \frac{1}{n_1 + n_2 + \dots + n_k - k} \sum_{j=1}^k (n_j - 1) \cdot s_j^2,$$

$$q = \left[1 + \frac{1}{3(k-1)} \left(\sum_{j=1}^k \frac{1}{n_j - 1} - \frac{1}{n_1 + \dots + n_k - k} \right) \right]^{-1}.$$

При $\min(n_1, n_2, \dots, n_k) > 3$ в условиях справедливости проверяемой гипотезы H_0 статистика γ должна подчиняться приблизительно закону χ^2 -распределения с $k - 1$ степенями свободы. Поэтому если оказалось, что величина $\gamma > \chi_{\alpha}^2(k - 1)$ (здесь $\chi_{\alpha}^2(k - 1)$ — 100%-ная точка распределения $\chi^2(k - 1)$, а α — уровень значимости критерия), то гипотезу о гомоскедастичности остатков следует отвергнуть. В этом случае значения s_j^2 могут быть использованы как диагональные элементы матрицы $\widehat{\Sigma}_{\varepsilon}$, что позволяет нам получить *практически реализуемый ОМНК*!

З а м е ч а н и е 3. В некоторых ситуациях при наличии гетероскедастичности в остатках исследователь ограничивается вычислением *обычных* МНК-оценок $\widehat{\Theta}_{\text{МНК}}$, поскольку *и в условиях гетероскедастичности они, как известно, остаются состоятельными*. Но тогда приходится решать проблему оценки ковариационной матрицы $\Sigma_{\widehat{\Theta}_{\text{МНК}}}$ вектора $\widehat{\Theta}_{\text{МНК}}$, так как классическая формула (4.33') в условиях гетероскедастичности «не работает», но справедлива формула (5.5), использование которой требует знания матрицы Σ_0 или ее оценки. Обойти эту проблему помогает результат Уайта (содержащийся в упомянутой выше работе 1980-го года), утверждающий: если мы находимся в условиях модели с некоррелированными и гетероскедастичными остатками, то матрица

$$\widehat{\Sigma}_{\widehat{\Theta}_{\text{МНК}}} = (\mathbf{X}^T \mathbf{X})^{-1} \left(\sum_{i=1}^n \widehat{\varepsilon}_i^2 \tilde{X}_i \tilde{X}_i^T \right) (\mathbf{X}^T \mathbf{X})^{-1}, \quad (5.30)$$

где $\widehat{\varepsilon}_i = y_i - \widehat{\Theta}_{\text{МНК}}^T \tilde{X}_i$, а $\tilde{X}_i = (1, x_i^{(1)}, \dots, x_i^{(p)})^T$, является состоятельной оценкой матрицы ковариаций обычных МНК-оценок коэффициентов регрессии $\theta_0, \theta_1, \dots, \theta_p$.

Стандартные отклонения $\sqrt{\mathbf{E}(\widehat{\theta}_{j, \text{МНК}} - \theta_j)^2}$, рассчитанные по формуле (5.30), иногда называют *стандартными ошибками в форме Уайта*.

5.4 ОЛММР с автокоррелированными остатками

5.4.1 Общее описание модели и ее основных характеристик

Мы уже познакомились в общих чертах в п. 5.1 с еще одним частным случаем обобщенной линейной модели множественной регрессии — регрессионной *моделью с автокоррелированными остатками*. Там же обсуждались реальные ситуации, естественно приводящие к необходимости рассмотрения такого рода моделей (в первую очередь к ним относятся так называемые «временные выборки», то есть ситуации, в которых исходные статистические данные модели *регистрируются во времени*). Наконец, была сформулирована *базовая идея* подобного рода зависимостей между остатками: эта зависимость неограниченно ослабевает по мере удаления остатков друг от друга по времени.

Рассмотрим теперь более систематически и подробно один из вариантов математической формализации этой идеи — *модель линейной регрессии, в которой регрессионные остатки $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ связаны автокорреляционной зависимостью 1-го порядка*¹. Подобного рода зависимость между остатками описывается соотношениями

$$\varepsilon_i = \rho \varepsilon_{i-1} + \delta_i, \quad (5.31)$$

где ρ — некоторое число, по абсолютной величине меньшее единицы (то есть $|\rho| < 1$), а случайные величины δ_i удовлетворяют требованиям, предъявляемым к регрессионным остаткам *классической модели*, то есть

$$E\delta_i \equiv 0, \quad E(\delta_i \delta_j) = \begin{cases} \sigma_0^2 & \text{при } i = j; \\ 0 & \text{при } i \neq j. \end{cases} \quad (5.32)$$

При этом полагается, что соотношения (5.31) справедливы для *любого* «момента времени» i , сколь угодно удаленного в прошлое или будущее, то есть i может «пробегать» все целочисленные значения от $-\infty$ до $+\infty$.

Отправляясь от (5.31) и (5.32), определим основные числовые характеристики (средние значения $E\varepsilon_i$ и ковариационную матрицу $\Sigma_\varepsilon = E(\varepsilon\varepsilon^T)$) вектора регрессионных остатков ε .

¹Регрессионные модели с более широким спектром типов взаимозависимостей, существующих между регрессионными остатками $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$, рассмотрены, например, в книге: М. Песаран, Л. Слейтер. Динамическая регрессия: теория и алгоритмы (пер. с англ.). М.: Финансы и статистика, 1984. — 312 с. См. также [Greene (2000), Ch.13].

Из (5.31) следует:

$$\begin{aligned}\varepsilon_i &= \rho\varepsilon_{i-1} + \delta_i = \rho(\rho\varepsilon_{i-2} + \delta_{i-1}) + \delta_i = \rho^2\varepsilon_{i-2} + \rho\delta_{i-1} + \delta_i = \\ &= \rho^2(\rho\varepsilon_{i-3} + \delta_{i-2}) + \rho\delta_{i-1} + \delta_i = \rho^3\varepsilon_{i-3} + \rho^2\delta_{i-2} + \rho\delta_{i-1} + \delta_i = \dots = \\ &= \delta_i + \rho\delta_{i-1} + \rho^2\delta_{i-2} + \dots = \sum_{k=0}^{\infty} \rho^k \delta_{i-k}.\end{aligned}\quad (5.33)$$

Из (5.33) с учетом (5.32) непосредственно следует:

$$\begin{aligned}\mathbf{E}\varepsilon_i &\equiv 0, \\ \sigma^2 &= \mathbf{D}\varepsilon_i = \mathbf{E}\varepsilon_i^2 = \mathbf{E}\delta_i^2 + \rho^2\mathbf{E}\delta_{i-1}^2 + \rho^4\mathbf{E}\delta_{i-2}^2 + \dots = \\ &= \sigma_0^2(1 + \rho^2 + \rho^4 + \dots) = \frac{\sigma_0^2}{1 - \rho^2}.\end{aligned}\quad (5.34)$$

Для вычисления ковариаций $\mathbf{E}(\varepsilon_i \varepsilon_{i-k})$ ($i = 1, 2, \dots, n$; $k = 1, 2, \dots, i - 1$) представим произведение $\varepsilon_i \varepsilon_{i-k}$ с учетом соотношения (5.33) в виде:

$$\begin{aligned}\varepsilon_i \varepsilon_{i-k} &= [(\delta_i + \rho\delta_{i-1} + \dots + \rho^{k-1}\delta_{i-(k-1)}) + \rho^k(\delta_{i-k} + \rho\delta_{i-k-1} + \dots)] \times \\ &\times (\delta_{i-k} + \rho\delta_{i-k-1} + \dots).\end{aligned}$$

Тогда, поскольку из взаимной некоррелированности δ_i следует взаимная некоррелированность случайных величин $(\delta_i + \rho\delta_{i-1} + \dots + \rho^{k-1}\delta_{i-(k-1)})$ и $(\delta_{i-k} + \rho\delta_{i-k-1} + \dots)$, получаем:

$$\text{cov}(\varepsilon_i, \varepsilon_{i-k}) = \mathbf{E}(\varepsilon_i \varepsilon_{i-k}) = \mathbf{E}[\rho^k(\delta_{i-k} + \rho\delta_{i-k-1} + \dots)^2] = \rho^k \mathbf{E}\varepsilon_{i-k}^2 = \sigma^2 \rho^k, \quad (5.35)$$

где дисперсия σ^2 определена соотношением (5.34).

Таким образом, мы пришли к тому, что автокорреляционная зависимость 1-го порядка (5.31), связывающая между собой регрессионные остатки ОЛМНР, в терминах ковариаций этих остатков эквивалентна соотношениям (5.34)–(5.35). Это позволяет записать обобщенную линейную модель множественной регрессии с автокоррелированными остатками в виде:

$$\left\{ \begin{array}{l} Y = \mathbf{X}\Theta + \varepsilon, \\ \mathbf{E}\varepsilon = \mathbf{0}_n, \\ \Sigma_\varepsilon = \mathbf{E}(\varepsilon\varepsilon^\top) = \frac{\sigma_0^2}{1 - \rho^2} \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-2} & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-3} & \rho^{n-2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & \rho & 1 \end{pmatrix} = \sigma^2 \Sigma_0, \\ (x^{(1)}, x^{(2)}, \dots, x^{(p)}) - \text{нечисловые переменные;} \\ \text{ранг матрицы } \mathbf{X} = p + 1 < n. \end{array} \right. \quad (5.36)$$

Таким образом, $(n \times n)$ -матрица Σ_0 определяется в данной модели значением *единственного* параметра ρ по формуле

$$\Sigma_0 = \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-2} & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-3} & \rho^{n-2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & \rho & 1 \end{pmatrix}. \quad (5.37)$$

Понадобится нам и матрица Σ_0^{-1} , участвующая во всех основных формулах ОМНК. Поэтому приведем здесь ее вид:

$$\Sigma_0^{-1} = \frac{1}{1 - \rho^2} \begin{pmatrix} 1 & -\rho & 0 & 0 & \dots & 0 & 0 & 0 \\ -\rho & 1 + \rho^2 & -\rho & 0 & \dots & 0 & 0 & 0 \\ 0 & -\rho & 1 + \rho^2 & -\rho & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & -\rho & 1 + \rho^2 & -\rho \\ 0 & 0 & 0 & 0 & \dots & 0 & -\rho & 1 \end{pmatrix}. \quad (5.38)$$

Становится понятным теперь и вероятностный смысл формально введенного в (5.31) параметра ρ . Действительно, из определения коэффициента корреляции и соотношений (5.34)–(5.35) следует, что коэффициент корреляции $r(\varepsilon_i, \varepsilon_{i \pm k})$ между регрессионными остатками, отстоящими друг от друга по времени на k единиц, равен:

$$r(\varepsilon_i, \varepsilon_{i \pm k}) = \rho^k, \quad k = 0, 1, 2, \dots \quad (5.39)$$

Именно в такой математической форме реализуется в модели (5.36) идея ослабления корреляционных связей между регрессионными остатками по мере их взаимного удаления во времени.

Как мы знаем (см. п. 5.2), общий прием, с помощью которого обобщенная модель регрессии сводится к классической, заключается во временном переходе к преобразованным исходным данным ($X_{\text{пр}}, Y_{\text{пр}}$) с помощью некоторым специальным образом подобранной $(n \times n)$ -матрицы преобразований C^{-1} . Для модели с автокоррелированными регрессионными остатками в качестве такой матрицы может быть использована матрица

$$C^{-1} = \begin{pmatrix} \sqrt{1 - \rho^2} & 0 & 0 & \dots & 0 & 0 & 0 \\ -\rho & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -\rho & 1 & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -\rho & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & -\rho & 1 \end{pmatrix} \cdot \frac{1}{\sqrt{1 - \rho^2}}. \quad (5.40)$$

Путем непосредственного перемножения соответствующих матриц нетрудно убедиться в справедливости равенств (5.9'), (5.10) и (5.11) при

матрице C^{-1} вида (5.40), а соответственно и в справедливости соотношения

$$E(C^{-1}\epsilon\epsilon^T(C^{-1})^T) = \sigma^2 I_n.$$

Перейдем теперь к *статистическому анализу* модели (5.36). При этом, мы должны уметь отвечать на два вопроса:

- (а) действительно ли остатки $\epsilon_1, \epsilon_2, \dots, \epsilon_n$ связаны моделью авто-регрессии 1-го порядка (5.31) или они являются взаимнонекоррелированными?
- (б) если они автокоррелированы, то как построить практически реализуемый ОМНК, то есть как состоятельно оценить параметр ρ ?

Описанию одного из возможных способов решения вопроса (а) посвящен следующий пункт.

5.4.2 Проверка гипотезы о наличии/отсутствии автокоррелированности регрессионных остатков

Как будет продемонстрировано ниже (см. п. 5.4.4), игнорирование автокорреляции регрессионных остатков создает серьезные трудности для применения обыкновенного МНК. Поэтому важно владеть методами, позволяющими устанавливать ее присутствие. Приведем здесь описание одного из них, — известного в специальной литературе под названием *критерия Дёрбина–Уотсона* (Durbin J., Watson G.S. *Testing for Serial Correlation in Least-Squares Regression. Biometrika*, vol. 37 (1950), pp. 409–428; and vol. 38 (1951), pp. 159–178).

Критическая статистика Дёрбина–Уотсона имеет вид

$$DW = \frac{\sum_{i=2}^n (\hat{\epsilon}_i - \hat{\epsilon}_{i-1})^2}{\sum_{i=1}^n \hat{\epsilon}_i^2}, \quad (5.41)$$

где $\hat{\epsilon}_i = y_i - \hat{\theta}_{0, \text{мнк}} - \hat{\theta}_{1, \text{мнк}} x_i^{(1)} - \dots - \hat{\theta}_{p, \text{мнк}} x_i^{(p)}$ — невязки обыкновенного метода наименьших квадратов. В упомянутой выше работе исследовано распределение статистики DW в предположении справедливости гипотезы

$$H_0: \rho = 0. \quad (5.42)$$

Посмотрим сначала как должна «вести себя» критическая статистика (5.41) в условиях справедливости гипотезы (5.42) на уровне средних

значений, приближенно. Представляя статистику DW в виде

$$\begin{aligned} DW &= \frac{\sum_{i=2}^n (\hat{\varepsilon}_i - \hat{\varepsilon}_{i-1})^2}{\sum_{i=1}^n \hat{\varepsilon}_i^2} = \frac{\sum_{i=1}^n \hat{\varepsilon}_i^2 - \hat{\varepsilon}_1^2 + \sum_{i=1}^n \hat{\varepsilon}_i^2 - \hat{\varepsilon}_n^2 - 2 \sum_{i=2}^n \hat{\varepsilon}_i \hat{\varepsilon}_{i-1}}{\sum_{i=1}^n \hat{\varepsilon}_i^2} = \\ &= 2 \left(1 - \frac{\sum_{i=1}^n \hat{\varepsilon}_i \hat{\varepsilon}_{i-1}}{\sum_{i=1}^n \hat{\varepsilon}_i^2} \right) - \frac{\hat{\varepsilon}_1^2 + \hat{\varepsilon}_n^2}{\sum_{i=1}^n \hat{\varepsilon}_i^2} \end{aligned}$$

и принимая во внимание, что объясняющие переменные модели неслучайны, модель содержит свободный член (а потому $\sum_{i=1}^n \hat{\varepsilon}_i = 0$, см. (4.49)), можно приближенно утверждать:

$$DW \approx 2(1 - \hat{r}(\varepsilon_i, \varepsilon_{i-1})) \approx 2(1 - r(\varepsilon_i, \varepsilon_{i-1})) = 2(1 - \rho). \quad (5.43)$$

Поэтому в условиях справедливости гипотезы (5.42) (то есть при $\rho = 0$) значения статистики DW должны группироваться в некоторой окрестности своего среднего, то есть *в окрестности числа два*.

Более тщательный анализ свидетельствует, что распределение статистики DW, к сожалению, зависит от наблюдаемых значений $\mathbf{X}$, но, тем не менее при $\rho = 0$

$$E(DW) = 2 + \frac{2p}{n - p - 1}. \quad (5.44)$$

Сравнивая (5.43) и (5.44), приходим к выводу, что среднее значение статистики DW в условиях справедливости гипотезы (5.42) (то есть при $\rho = 0$) *лишь асимптотически* не зависит от числа p объясняющих переменных модели.

Поскольку, как было отмечено, распределение статистики DW в условиях (5.42) зависит от $\mathbf{X}$, то точно подсчитать 100α -процентные точки $d(\alpha; n; p)$ этого распределения «на все случаи жизни» (то есть справедливые при всех $\mathbf{X}$) не представляется возможным. Однако Дёрбину и Уотсону удалось для двух значений уровня значимости критерия (для $\alpha = 0,05$ и $\alpha = 0,01$) и для набора значений p и n в условиях, когда $DW < 2$ (то есть имеет место видимая *положительная* корреляция между ε_i и ε_{i-1} , см. (5.43)) вычислить нижнюю ($d_L(\alpha; n; p)$) и верхнюю ($d_U(\alpha; n; p)$) границы для истинного значения искомой процентной точки, то есть гарантируется, что

$$d_L(\alpha; n; p) \leq d(\alpha; n; p) \leq d_U(\alpha; n; p)$$

(данные значения приведены в таблице П1.12^а Приложения 1).

Поэтому в случае $DW < 2$ при проверке гипотезы (5.42) при альтернативе

$$H_1: \rho > 0$$

решение принимается следующим образом:

- если $DW \leq d_L(\alpha; n; p)$, то гипотеза H_0 отвергается (с вероятностью ошибки, не превосходящей α);
- если $DW \geq d_U(\alpha; n; p)$, то гипотеза H_0 не отвергается;
- если $d_L(\alpha; n; p) < DW < d_U(\alpha; n; p)$, то сделать определенный вывод об отклонении или принятии гипотезы H_0 невозможно («зона неопределенности»).

Если имеет место видимая отрицательная корреляция между ε_i и ε_{i-1} (то есть при $DW > 2$), то альтернативой гипотезы (5.42) является гипотеза

$$H'_1: \rho < 0.$$

Факт симметричности распределения DW относительно своего среднего значения (в условиях справедливости гипотезы (5.42)) позволяет определить правило принятия решения:

- если $DW \geq 4 - d_L(\alpha; n; p)$, то гипотеза H_0 отвергается (с вероятностью ошибки, не превосходящей α);
- если $DW \leq 4 - d_U(\alpha; n; p)$, то гипотеза H_0 не отвергается;
- если $4 - d_U(\alpha; n; p) < DW < 4 - d_L(\alpha; n; p)$, то сделать определенный вывод об отклонении или принятии гипотезы H_0 невозможно («зона неопределенности»).

Так что, грубо говоря, в случае отсутствия автокорреляции в регрессионных остатках и при достаточном объеме наблюдений (скажем, при $n \geq 10p$) значение критической статистики DW должно «не слишком отличаться» от 2. При этом, обычные значения d_L и d_U варьируют, соответственно, на уровне 1,2 ~ 1,3 и 1,6 ~ 1,7. Однако, при сопоставимых величинах n и p «зона неопределенности» может быть весьма широкой, а значения d_U могут «залезать» даже правее двойки! Так, например, при $\alpha = 0,05$ имеем:

при $n = 15$ и $p = 5$ $d_L = 0,56$ и $d_U = 2,22$;

при $n = 15$ и $p = 9$ $d_L = 0,175$ и $d_U = 3,22$;

В заключение подчеркнем, что данный критерий выведен лишь в предположении неслучайности объясняющих переменных и наличия свободного члена в анализируемой модели регрессии. Методы проверки автокоррелированности регрессионных остатков в рамках более широкого

класса моделей (например, в ситуациях, когда среди объясняющих переменных могут присутствовать лагированные значения зависимой переменной) читатель найдет в [Johnston, DiNardo (1997)], [Greene (2000)].

5.4.3 Практически реализуемый ОМНК

В ситуациях, в которых значение коэффициента корреляции ρ между соседними по времени регрессионными остатками *известно*, исследователь не должен испытывать затруднений в практической реализации основных формул ОМНК (5.6), (5.5) и (5.19).

Поэтому остановимся на ситуации (гораздо более реалистичной), когда *значение параметра ρ априори не известно исследователю*. Практически все процедуры, предложенные для реализации ОМНК в модели регрессии с автокоррелированными остатками при неизвестном значении ρ , имеют итерационный характер (см. [Pindyck, Rubinfeld, (1976), п. 6.2.1]). Приведем здесь описание одной из наиболее распространенных процедур подобного типа, известной в литературе под названием *процедуры Кокрейна–Оркатта* (Cochrane D., Orcutt G. H. Application of Least-Squares Regressions to Relationships Containing Autocorrelated Error Terms. «Journ. of the Amer. Stat. Assoc.», vol. 44 (1949), pp. 32–61):

- 1) вычисляются обычные МНК-оценки $\hat{\Theta}_{\text{МНК}}$ по формуле (4.22);
- 2) подсчитываются невязки 1-й итерации $\hat{\varepsilon}_i^{(1)} = y_i - \hat{\theta}_{0,\text{МНК}} - \hat{\theta}_{1,\text{МНК}}x_i^{(1)} - \dots - \hat{\theta}_{p,\text{МНК}}x_i^{(p)}$;
- 3) первое приближение $\hat{\rho}^{(1)}$ оценки $\hat{\rho}$ неизвестного параметра ρ определяется в качестве МНК-оценки коэффициента регрессии ρ в модели $\hat{\varepsilon}_i^{(1)} = \rho\hat{\varepsilon}_{i-1}^{(1)} + \delta_i^{(1)}$, где остатки $\delta_i^{(1)}$ удовлетворяют условиям классической модели (4.1);
- 4) вычисляются ОМНК-оценки $\hat{\Theta}_{\text{ОМНК}}(\hat{\rho}^{(1)})$ по формуле (5.6) с матрицей $\Sigma_0(\hat{\rho}^{(1)})$, определенной соотношением (5.37), в котором вместо ρ подставлены значения $\hat{\rho}^{(1)}$;
- 5) подсчитываются невязки 2-й итерации

$$\hat{\varepsilon}_i^{(2)} = y_i - \hat{\theta}_{0,\text{ОМНК}}(\hat{\rho}^{(1)}) - \hat{\theta}_{1,\text{ОМНК}}(\hat{\rho}^{(1)})x_i^{(1)} - \dots - \hat{\theta}_{p,\text{ОМНК}}(\hat{\rho}^{(1)}) \cdot x_i^{(p)};$$

- 6) второе приближение $\hat{\rho}^{(2)}$ оценки $\hat{\rho}$ неизвестного параметра ρ определяется в качестве МНК-оценки коэффициента регрессии ρ в модели $\hat{\varepsilon}_i^{(2)} = \rho\hat{\varepsilon}_{i-1}^{(2)} + \delta_i^{(2)}$, где остатки $\delta_i^{(2)}$ удовлетворяют условиям классической модели (4.1);

- 7) вычисляются ОМНК-оценки $\hat{\Theta}_{\text{ОМНК}}(\hat{\rho}^{(2)})$ по формуле (5.6) с матрицей $\Sigma_0(\hat{\rho}^{(2)})$, определенной соотношением (5.37), в котором вместо ρ подставлены значения $\hat{\rho}^{(2)}$ — и т. д.

Процедуру заканчивают при стабилизации получаемых значений $\hat{\rho}^{(k)}$, то есть на стадии, когда очередное приближение $\hat{\rho}$ мало отличается от

предыдущего. «Тонкое место» метода определяется типовым недостатком подобных процедур, заключающимся в возможности «скатиться» в ходе итераций в *локальный*, а не *глобальный* минимум критерия наименьших квадратов. В этом случае значение параметра ρ может быть определено с большой ошибкой.

5.4.4 Искажение характеристик точности МНК-оценок, обусловленное игнорированием автокоррелированности регрессионных остатков

Под игнорированием автокоррелированности регрессионных остатков мы понимаем ситуацию, в которой исследователь вместо ОМНК-оценок $\hat{\Theta}_{\text{ОМНК}}$, определяемых формулой (5.6), и соотношений (5.5) и (5.15'), позволяющих оценивать точность соответственно МНК- и ОМНК-оценок, пользуется соответствующими формулами *обычного* метода наименьших квадратов — (4.22) для $\hat{\Theta}_{\text{МНК}}$ и (4.33) для ковариационной матрицы оценок $\hat{\Theta}_{\text{МНК}}$.

В этом пункте мы хотим проанализировать меру искажения характеристик точности обычных МНК-оценок при использовании формулы (4.33') (позволяющей оценивать ковариационную матрицу МНК-оценок $\hat{\Theta}_{\text{МНК}}$ только в условиях *классической* ЛММР) вместо действующей в условиях *обобщенной* ЛММР формулы (5.5).

Заметим, что правомерность использования обычного МНК для построения оценок $\hat{\Theta}_{\text{МНК}}$ неизвестных параметров Θ подкрепляется сохранением свойств несмещенности и (в широком классе ситуаций) состоятельности оценок $\hat{\Theta}_{\text{МНК}}$ и в условиях *обобщенной* ЛММР. Однако неустраняемые искажения при этом возникают в оценивании их характеристик точности, то есть ковариационной матрицы $\Sigma_{\hat{\Theta}_{\text{МНК}}}$.

С целью анализа этих искажений рассмотрим модель (5.36) зависимости y от *единственной* объясняющей переменной x без *свободного* члена. Очевидно, в этом случае $X = x, \Theta = \theta$, то есть $y_i = \theta x_i + \varepsilon_i$, а матрица наблюдаемых значений объясняющей переменной x имеет вид

$$X = (x_1, x_2, \dots, x_n)^T.$$

Обычная оценка МНК, подсчитанная по общей формуле (4.22), в данном случае имеет вид

$$\hat{\theta}_{\text{МНК}} = (X^T X)^{-1} X^T Y = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}. \quad (5.45)$$

Подсчитаем дисперсию (средний квадрат ошибки) этой оценки параллельно двумя способами: 1) по формуле (4.33), справедливой только

в рамках условий *классической* модели, то есть игнорирующей автокоррелированность регрессионных остатков, и 2) по формуле (5.5), учитывающей эту автокоррелированность в рамках условий обобщенной модели (5.36).

Использование формулы (4.33) дает:

$$(D\hat{\theta}_{\text{мнк}})_{\text{прибл}} = \sigma^2(\mathbf{X}^T \mathbf{X})^{-1} = \frac{\sigma^2}{\sum_{i=1}^n x_i^2}. \quad (5.46)$$

Проведем расчет величины $D\hat{\theta}_{\text{мнк}}$ по *правильной* формуле (5.5):

$$\begin{aligned} D\hat{\theta}_{\text{мнк}} &= \sigma^2(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \Sigma_0 \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} = \\ &= \sigma^2 \frac{1}{\sum_{i=1}^n x_i^2} (x_1, x_2 \dots x_n) \times \\ &\times \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-2} & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-3} & \rho^{n-2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & \rho & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \frac{1}{\sum_{i=1}^n x_i^2} = \\ &= \frac{\sigma^2}{\sum_{i=1}^n x_i^2} \left(1 + 2\rho \frac{\sum_{i=1}^{n-1} x_i x_{i+1}}{\sum_{i=1}^n x_i^2} + 2\rho^2 \frac{\sum_{i=1}^{n-2} x_i x_{i+2}}{\sum_{i=1}^n x_i^2} + \dots + 2\rho^{n-1} \frac{x_1 x_n}{\sum_{i=1}^n x_i^2} \right). \end{aligned} \quad (5.47)$$

Возьмем для сравнения результатов (5.47) и (5.46) их отношение:

$$\frac{D\hat{\theta}_{\text{мнк}}}{(D\hat{\theta}_{\text{мнк}})_{\text{прибл}}} = 1 + 2\rho \frac{\sum_{i=1}^{n-1} x_i x_{i+1}}{\sum_{i=1}^n x_i^2} + \dots + 2\rho^{n-1} \frac{x_1 x_n}{\sum_{i=1}^n x_i^2}. \quad (5.48)$$

Очевидно, правая часть (5.48) может существенно превышать единицу, например, при $\rho > 0$ и при положительной коррелированности значений объясняющей переменной x . При этом мы имеем в виду, что коррелированность значений x измеряется формально вычисленным коэффициентом корреляции (в предположении нулевого среднего значения x), то есть

$$r(k) = r(x_i, x_{i+k}) = \frac{\sum_{i=1}^{n-k} x_i x_{i+k}}{\sum_{i=1}^n x_i^2}, \quad k = 1, 2, \dots, n-1.$$

Тогда, придав определенное числовое значение ρ и положив, например, $r(k) = \rho^k$, можно численно оценить правую часть (5.48). Так, при $\rho = 0,8$ имеем (при достаточно больших n):

$$1 + 2\rho \frac{\sum_{i=1}^n x_i x_{i+1}}{\sum_{i=1}^n x_i^2} + \dots + 2\rho^{n-1} \frac{x_1 x_n}{\sum_{i=1}^n x_i^2} = 1 + 2\rho^2 + 2\rho^4 + \dots + 2\rho^{2(n-1)}$$

$$= 2(1 + \rho^2 + \rho^4 + \dots + \rho^{2(n-1)}) - 1 \approx 1 \frac{1}{1 - \rho^2} - 1 = \frac{1 + \rho^2}{1 - \rho^2} \approx 4,56.$$

Мы видим, что игнорирование автокоррелированности остатков в обобщенной линейной модели регрессии вида (5.36) при подсчете среднего квадрата ошибки МНК-оценки $\hat{\theta}_{\text{МНК}}$ приводит к неправомерному занижению этой ошибки более, чем в 4,5 раза!

Обратим внимание читателя еще на одно обстоятельство. Выбранный нами способ сравнения $(D\hat{\theta}_{\text{МНК}})_{\text{прибл.}}$ и $D\hat{\theta}_{\text{МНК}}$ позволил обойти еще одно «тонкое место» в оценивании характеристик обобщенной модели (5.36) методами, игнорирующими автокорреляцию регрессионных остатков. Речь идет об оценивании параметра σ^2 (беря отношения сравниваемых характеристик (5.47) и (5.46), мы сумели элиминировать этот параметр из нашего рассмотрения). Проведенный аналогично только что выполненному анализ результатов правильного (с помощью формулы (5.19)) и некорректного (с помощью формулы (4.31)) оценивания параметра σ^2 показывает, что расхождения здесь могут быть еще более существенными, чем те, которые получаются при сравнении оценок для $D\hat{\theta}_{\text{МНК}}$ (см., например, [Джонстон (1980), с. 246–248]).

5.5 Практически реализуемый ОМНК (общий подход)

Если ковариационная матрица Σ_ϵ регрессионных остатков ОЛММР априори не известна, и если ее удастся состоятельно оценить по исходным наблюдениям (4.1а)–(4.3) с тем, чтобы использовать в основных формулах ОМНК, то говорят о **практически реализуемом ОМНК**.

В замечании, завершающем п. 5.2, мы уже обращали внимание читателя на *проблемы практической реализации* ОМНК: в подавляющем большинстве ситуаций практики эконометрического моделирования ковариационная матрица регрессионных остатков Σ_ϵ не известна исследователю заранее. В то же время ее значение необходимо для реализации основных формул ОМНК: (5.6), (5.5), (5.15'), (5.18) и (5.19). Если элементы матрицы Σ_ϵ не связывать дополнительно никакими априорными ограничениями и условиями, то мы приходим к практически неразрешимой задаче статистического оценивания $n(n+1)/2$ неизвестных параметров (элементов матрицы Σ_ϵ) всего по n исходным наблюдениям. Именно

поэтому основной подход к статистическому анализу обобщенной линейной модели множественной регрессии с помощью ОМНК в ситуации, когда ковариационная матрица регрессионных остатков неизвестна, базируется на предположении, что задана *параметрическая структура матрицы* Σ_ϵ (и соответственно Σ_0), то есть форма ее функциональной зависимости от сравнительно небольшого количества параметров $\sigma^2, a^{(1)}, a^{(2)}, \dots, a^{(m)}$. Это значит, что матрица Σ_0 рассматривается как функция *известного вида* от *неизвестных* параметров $a = (a^{(1)}, \dots, a^{(m)})$ и соответственно: $\Sigma_\epsilon = \sigma^2 \Sigma_0(a^{(1)}, \dots, a^{(m)}) = \sigma^2 \Sigma_0(a)$. Так, в примере 2.1 удалось построить подобную схему с *единственным* неизвестным параметром σ^2 , так как в нем была использована структура матрицы Σ_ϵ , определяемая соотношением (5.2), где $\lambda_i = 1/x_i^2$.

В примере ОЛММР с автокоррелированными остатками (см. (5.4')) структура матрицы Σ_ϵ определяется двумя параметрами σ^2 и $a^{(1)} = \rho$, так что элементы матрицы Σ_0 восстанавливаются по единственному параметру ρ . Наконец, в описанном выше критерии Бреуша-Пагана, используемом при анализе ОЛММР с гетероскедастичными (но взаимно некоррелированными) регрессионными остатками, матрица Σ_0 определяется $(m+1)$ параметрами $a^{(0)}, a^{(1)}, \dots, a^{(m)}$, участвующими в описании поведения дисперсий

$$\sigma_i^2 = E\epsilon_i^2 = \sigma^2 \cdot h(a^{(0)} + a^{(1)} \cdot x_i^{(j_1)} + \dots + a^{(m)} \cdot x_i^{(j_m)}),$$

где $x^{(j_l)}$ ($l = 1, 2, \dots, m$) те объясняющие переменные модели, от которых предположительно зависит характер случайного разброса остатков. Главная идея общего подхода заключается в следующем: по исходным наблюдениям (X, Y) сначала строятся состоятельные оценки $\hat{a} = (\hat{a}^{(1)}, \dots, \hat{a}^{(m)})$ параметров $a = (a^{(1)}, \dots, a^{(m)})$. Затем вычисляются оценки $\hat{\Sigma}_0 = \Sigma_0(\hat{a})$, $\hat{\sigma}^2$ (по формуле (5.19)) и $\hat{\Sigma}_\epsilon = \hat{\sigma}^2 \hat{\Sigma}_0$. При некоторых достаточно общих условиях² удастся показать, что использование полученных таким образом оценок $\hat{\Sigma}_0$ и $\hat{\Sigma}_\epsilon$ в основных формулах ОМНК (5.5), (5.6) и (5.15') вместо неизвестных истинных значений Σ_0 и Σ_ϵ дает также состоятельные оценки соответственно для $\Sigma_{\hat{\Theta}_{\text{МНК}}}$, Θ и $\Sigma_{\hat{\Theta}_{\text{ОМНК}}}$. Подобный способ оценивания и составляет основную идею общего подхода к построению *практически реализуемого обобщенного метода наименьших квадратов*.

Опишем общую схему процедуры практически реализуемого ОМНК, которая используется в рамках *нормальной* ОЛММР, то есть в рамках условий модели (5.1), дополненных условием $(0_n; \sigma^2 \Sigma_0(a))$ — нормальности вектора регрессионных остатков ϵ .

²В качестве таких условий используются, например, условия сходимости по вероятности (при $n \rightarrow \infty$): а) матрицы $\frac{1}{n} X^T \hat{\Sigma}_0^{-1} X$ — к некоторой конечной невырожденной матрице и б) вектора-столбца $\frac{1}{n} X^T \hat{\Sigma}_0^{-1} \epsilon$ — к вектору-столбцу 0_{p+1} , состоящему из одних нулей.

В этом случае функция правдоподобия остатков $\varepsilon = Y - X\Theta$ имеет вид (сравнить с (4.20)):

$$L(\varepsilon | \Theta; \sigma^2; \Sigma_0(a)) = \frac{1}{(2\pi)^{\frac{n}{2}} (\sigma^2)^{\frac{n}{2}} |\Sigma_0(a)|^{\frac{1}{2}}} \times \\ \times \exp \left[-\frac{1}{2\sigma^2} (Y - X\Theta)^\top \Sigma_0^{-1}(a) (Y - X\Theta) \right].$$

Переходим к *логарифмической функции правдоподобия* $l = \ln L$ и, приравнявая ее производные по Θ , σ^2 и a к нулю, составляем систему уравнений для отыскания оценок максимального правдоподобия $\hat{\Theta}_{\text{ММП}}$, $\hat{\sigma}_{\text{ММП}}^2$ и $\hat{a} = (\hat{a}_1, \dots, \hat{a}_m)$.

$$l(\varepsilon | \Theta; \sigma^2; \Sigma_0(a)) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) + \frac{1}{2} \ln |\Sigma_0^{-1}(a)| - \\ - \frac{1}{2\sigma^2} (Y - X\Theta)^\top \Sigma_0^{-1}(a) (Y - X\Theta);$$

$$\begin{cases} \frac{\partial l}{\partial \Theta} = -\frac{1}{2\sigma^2} \frac{\partial}{\partial \Theta} \left[(Y - X\Theta)^\top \Sigma_0^{-1}(a) (Y - X\Theta) \right] = \mathbf{0}_{p+1}, \\ \frac{\partial l}{\partial \sigma^2} = -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2} (Y - X\Theta)^\top \Sigma_0^{-1}(a) (Y - X\Theta) = 0, \\ \frac{\partial l}{\partial a^{(j)}} = -\frac{(Y - X\Theta)^\top \omega_j (Y - X\Theta)}{(Y - X\Theta)^\top \Sigma_0^{-1}(a) (Y - X\Theta)} + \frac{1}{n} \text{tr}(\omega_j \Sigma_0(a)) = 0, \end{cases} \quad (5.49)$$

где $\omega_j = \frac{\partial \Sigma_0^{-1}(a)}{\partial a^{(j)}}$, $j = 1, 2, \dots, m$.

Анализ системы уравнений (5.49) позволяет сделать следующие выводы:

1) Решение первого уравнения системы доставляет минимум критерию ОМНК (см. (5.16)), а следовательно, *даже не производя дифференцирования по Θ* , можно заключить, что оценки параметров Θ в схеме ОЛММП по обобщенному методу наименьших квадратов совпадают с оценками по методу максимального правдоподобия (то есть $\hat{\Theta}_{\text{ОМНК}} = \hat{\Theta}_{\text{ММП}}$, где оценка $\hat{\Theta}_{\text{ОМНК}}$ определена формулой (5.6)).

2) Решение второго уравнения системы дает в качестве оценки $\hat{\sigma}_{\text{ММП}}^2$ ту же оценку $\hat{\sigma}^2$ (с точностью до асимптотически устранимого смещения), которая используется в ОМНК, то есть

$$\hat{\sigma}_{\text{ММП}}^2 = \frac{1}{n} (Y - X\Theta)^\top \Sigma_0^{-1} (Y - X\Theta) = \frac{n - p - 1}{n} \hat{\sigma}^2,$$

где оценка $\hat{\sigma}^2$ определена формулой (5.19).

3) Сложность решения системы уравнений относительно оценок вспомогательных параметров $a = (a^{(1)}, \dots, a^{(m)})$ (эта система описана третьим уравнением в системе (5.49), протиражированным для $j = 1, 2, \dots, m$)

зависит от принятой формы зависимости элементов матрицы Σ_0 (или Σ_0^{-1}) от этих параметров и от их общего числа m . При экономных и относительно простых способах параметризации матрицы Σ_0 решение этой системы обычно не представляет трудностей. Что касается участвующих в этой системе регрессионных остатков $\epsilon = Y - X\Theta$, то они заменяются соответствующими «невязками», $\hat{\epsilon} = Y - X\hat{\Theta}$ в режиме обычной (для ОМНК) итерационной процедуры: сначала в качестве Θ в выражение $\epsilon = Y - X\Theta$ подставляется обычная МНК-оценка $\hat{\Theta}_{\text{МНК}}$; решают систему относительно a и получают первое приближение оценок $\hat{a}^{(1)}, \hat{a}^{(2)}, \dots, \hat{a}^{(m)}$ и, соответственно, первое приближение для $\hat{\Sigma}_0 = \Sigma_0(\hat{a}^{(1)}, \dots, \hat{a}^{(m)})$; затем используют эту матрицу Σ_0 для получения первого приближения для ОМНК-оценок $\hat{\Theta}_{\text{ОМНК}}$, по ним вычисляют «невязки», $\hat{\epsilon} = Y - X\hat{\Theta}_{\text{ОМНК}}$ и т.д. Получаемые, в конечном счете, таким образом оценки при весьма слабых предположениях оказываются асимптотически эквивалентными оценками максимального правдоподобия, а следовательно, в достаточно широком классе случаев, — асимптотически эффективными. Заметим, что описанная выше итерационная процедура Кохрейна–Оркатта (см. п. 5.4.3) относится к этому же типу методов оценивания.

Выводы

1. Если при рассмотрении линейной модели множественной регрессии выполняются все условия КЛММР кроме хотя бы одного из двух условий, — *гомоскедастичности и взаимной некоррелированности регрессионных остатков* (то есть ковариационная матрица Σ_ϵ регрессионных остатков имеет вид $\Sigma_\epsilon = \sigma^2 \cdot \Sigma_0$, где *симметричная положительно определенная матрица Σ_0 не является единичной*), то такая модель называется *обобщенной линейной моделью множественной регрессии* (ОЛММР).

2. Наилучшими (несмещенными, состоятельными и эффективными в классе линейных несмещенных оценок) оценками параметров ОЛММР являются оценки *обобщенного метода наименьших квадратов* (ОМНК-оценки), реализация которого требует знания матрицы Σ_0 из соотношения $\Sigma_\epsilon = \sigma^2 \cdot \Sigma_0$. Правда, и оценки по обычному методу наименьших квадратов (МНК-оценки) параметров ОЛММР остаются состоятельными и несмещенными, но перестают быть эффективными.

3. В большинстве ситуаций практики эконометрического моделирования ковариационная матрица регрессионных остатков Σ_ϵ не известна исследователю заранее, а следовательно, он не имеет *прямой возможности* воспользоваться формулами ОМНК (в которых эта матрица присутствует). Поэтому существует специальная итерационная процедура (*называемая практически реализуемым ОМНК*), позволяющая получать,

тем не менее, именно ОМНК-оценки неизвестных параметров Θ . Главная идея, на которой основан практически реализуемый ОМНК, заключается в следующем: по исходным наблюдениям $\{(x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)}; y_i)\}_{i=1,2,\dots,n}$ сначала строятся *обычные* МНК-оценки $\hat{\Theta}_{\text{МНК}} = (\hat{\theta}_{0,\text{МНК}}, \hat{\theta}_{1,\text{МНК}}, \dots, \hat{\theta}_{p,\text{МНК}})^T$ параметров Θ ; затем вычисляются «невязки», (первое приближение) $\hat{\varepsilon}_i^{(1)} = y_i - \hat{\theta}_{0,\text{МНК}} - \hat{\theta}_{1,\text{МНК}}x_i^{(1)} - \dots - \hat{\theta}_{p,\text{МНК}}x_i^{(p)}$ и по ним строится первое приближение $\hat{\Sigma}_\varepsilon^{(1)}$ оценки для неизвестной матрицы Σ_ε ; затем матрица $\hat{\Sigma}_\varepsilon^{(1)}$ используется для вычисления *обобщенных* МНК-оценок $\hat{\Theta}_{\text{ОМНК}}^{(1)}$. После этого пересчитываются невязки $\hat{\varepsilon}_i^{(2)} = y_i - \hat{\theta}_{0,\text{ОМНК}}^{(1)} - \hat{\theta}_{1,\text{ОМНК}}^{(1)}x_i^{(1)} - \dots - \hat{\theta}_{p,\text{ОМНК}}^{(1)}x_i^{(p)}$ (то есть вычисляется их второе приближение) и т. д. При этом, как правило, неизвестная ковариационная матрица остатков Σ_ε каким-то образом *параметризуется*, то есть выписывается в форме, в которой все $n(n+1)/2$ ее элементов являются функциями заданного общего вида от небольшого числа (двух-трех) одних и тех же параметров.

4. Одним из распространенных частных случаев ОЛММП является ЛММП с гетероскедастичными регрессионными остатками. В этом случае нарушается единственное из требований КЛММП — требование гомоскедастичности остатков (то есть матрица Σ_0 остается диагональной, но имеет, вообще говоря, разные значения по диагонали). Критерием, минимизация которого дает оценки ОМНК параметров такой модели, является взвешенная сумма квадратов разностей между наблюдаемым и модельным значениями зависимой переменной. Поэтому ОМНК в условиях гетероскедастичности регрессионных остатков часто называют *взвешенным методом наименьших квадратов*.

5. Существуют различные подходы к *статистической проверке гипотезы гомоскедастичности остатков при альтернативе их гетероскедастичности*. Один из распространенных методов решения этой задачи — **критерий Бреуша–Пагана**, который, в случае отклонения гипотезы гомоскедастичности, позволяет построить *практически реализуемый* ОМНК оценивания параметров ОЛММП с гетероскедастичными остатками.

6. Если при оценивании параметров ОЛММП с гетероскедастичными остатками ограничиваются обычным МНК, то для получения состоятельной оценки ковариационной матрицы таких МНК-оценок можно воспользоваться, например, стандартными ошибками в форме Уайта (см. (5.30)).

7. В ситуациях, когда значения анализируемых переменных регистрируются *во времени* (то есть номер наблюдения i несет смысловую нагрузку времени регистрации наблюдения), регрессионные остатки оказываются часто *взаимозависимыми во времени*, а значит их ковариационная матрица Σ_ε (а следовательно, и Σ_0) не может быть диагональной. Распространенным способом математической формализации такого ро-

да зависимостей является соотношение (5.31), связывающее соседние по времени остатки *моделью авторегрессии 1-го порядка*. В этих случаях говорят об **ОЛММР с автокоррелированными остатками**, причем, анализируемая ЛММР может быть представлена в рамках ОЛММР с матрицей Σ_0 , зависящей от *единственного* (неизвестного) параметра ρ – коэффициента корреляции между соседними остатками.

8. Наиболее распространенный подход к статистической проверке гомоскедастичности регрессионных остатков *при альтернативе их автокоррелированности* – это критерий Дёрбина–Уотсона, основанный на критической статистике DW (см. (5.41)). Процедура построения *практически реализуемого ОМНК* основана в данном случае на итерационных оценках параметра ρ , среди которых к одной из наиболее распространенных можно отнести процедуру Кохрейна–Оркатта (см. п. 5.4.3).

9. Если при оценивании параметров ОЛММР с автокоррелированными остатками ограничиваются обычными МНК-оценками, то использование формулы (4.22) для оценки ковариационной матрицы таких оценок (формулы, выведенной в условиях соблюдения требований к КЛММР!) является некорректным и *дает заниженные значения среднеквадратических ошибок оценивания* (см. п. 5.4.4). Альтернатива – использование корректной (в условиях ОЛММР) формулы (5.15') с подстановкой в нее вместо σ^2 и ρ их оценок, соответственно $\hat{\sigma}^2$ и $\hat{\rho}$.

Глава 6

Прогнозирование, основанное на линейных моделях множественной регрессии

В главе 2 (п. 2.3) мы обсуждали основные типовые задачи практики, требующие изучения статистических связей между результирующей переменной y и набором объясняющих переменных $X = (1, x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$. И, как мы видим, каждый из типов этих задач (и «нормирование», и «прогноз-планирование-диагностика», и «оценка эффективности функционирования анализируемой системы», и «оптимальное регулирование параметров функционирования анализируемой системы – ситуационный анализ», и «оценка трудно доступных для непосредственного измерения параметров системы») требует от исследователя умения предсказывать значение результирующей переменной $y(X)$ (или его условного среднего $f(X) = E(y|X)$) по заданным значениям объясняющих переменных X . Эту задачу мы и будем называть *задачей прогнозирования* независимо от того, в рамках какого из перечисленных выше типов конечных прикладных целей исследования она решается. При этом заметим, что прогноз $\hat{y}(X)$ может быть **точечным** (то есть определенным в виде числа) или **интервальным** (определенным в виде интервала значений, в который гарантируется попадание неизвестного прогнозируемого значения $y(X)$ с заданной доверительной вероятностью P_0 , как правило, близкой к единице).

В данной главе описывается решение задачи построения прогноза (точечного и интервального) для значения $y(X)$ или $f(X) = E(y|X)$ по заданным значениям объясняющих переменных $X = (1, x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$ в рамках нормальной обобщенной линейной модели множествен-

ной регрессии (НОЛММР)

$$\begin{cases} y(X_i) = \Theta^T X_i + \varepsilon(X_i), & i = 1, 2, \dots, n; \\ \varepsilon \in N_n(\Theta; \sigma^2 \Sigma_0), \end{cases} \quad (6.1)$$

где $\Theta = (\theta_0, \theta_1, \dots, \theta_p)^T$ и σ^2 — оцениваемые параметры модели, $\varepsilon = (\varepsilon(X_1), \varepsilon(X_2), \dots, \varepsilon(X_n))^T$ — вектор регрессионных остатков, Σ_0 — некоторая симметричная положительно-определенная $n \times n$ -матрица (как правило, тоже относящаяся к оцениваемым параметрам модели), а n — общее число наблюдений, имеющихся в распоряжении исследователя.

Очевидно, построение точечного и интервального прогноза для $y(X)$ и/или $f(X) = E(y|X)$, определенных моделью (6.1), должно опираться на анализ точности оценивания этой модели, то есть — на свойства оценок $\hat{\Theta}$, $\hat{\sigma}^2$ и регрессионных остатков $\varepsilon(X)$. При этом, пока мы говорим о построении лишь **точечных** оценок для Θ , σ^2 , $y(X)$ или $f(X) = E(y|X)$, нет необходимости требовать нормальности остатков $\varepsilon(X)$. Однако для обоснования корректности построения соответствующих **интервальных** оценок нормальность регрессионных остатков нужна, хотя полученные рекомендации используются как *приближенные* (но асимптотически, по $n \rightarrow \infty$, точные) и в рамках «не нормальной» ОЛММР.

6.1 Анализ точности оцененной ЛММР (теоретическая база для решения задач прогноза)

В главе 5 были выведены ОМНК-оценки $\hat{\Theta}_{\text{омнк}}$ и $\hat{\sigma}_{\text{омнк}}^2$ для параметров, соответственно, Θ и σ^2 модели (6.1) (см. формулы (5.14') и (5.19)). Оставляем читателю в качестве самостоятельного упражнения вывод оценок максимального правдоподобия $\hat{\Theta}_{\text{мп}}$ и $\hat{\sigma}_{\text{мп}}^2$ для тех же параметров, откуда будет следовать доказательство факта: $\hat{\Theta}_{\text{омнк}} = \hat{\Theta}_{\text{мп}} = \hat{\Theta}$ и $\hat{\sigma}_{\text{омнк}}^2 = \frac{n}{n-p-1} \hat{\sigma}_{\text{мп}}^2 = \hat{\sigma}^2$. Кроме того, мы выяснили, что оценки $\hat{\Theta}$ и $\hat{\sigma}^2$ являются статистически взаимонезависимыми, *несмещенными* и знаем выражение для *ковариационной матрицы* $\Sigma_{\hat{\Theta}}$ вектора оценок $\hat{\Theta}$ (см. (5.15')).

Однако при построении *интервальных* прогнозов для $y(X)$ и $f(X)$, а также при описании *доверительных областей* для Θ нам понадобятся более глубокие сведения о поведении оценок $\hat{\Theta}$ и $\hat{\sigma}^2$ параметров модели (6.1). В частности, мы будем использовать следующие факты:

$$\hat{\Theta} \in N_{p+1}(\Theta; \sigma^2(\mathbf{X}^T \Sigma_0^{-1} \mathbf{X})^{-1}); \quad (6.2)$$

$$\frac{(n-p-1)}{\sigma^2} \sim \chi^2(n-p-1); \quad (6.3)$$

$$\frac{\frac{1}{p+1}(\hat{\Theta} - \Theta)^T (X^T \Sigma_0^{-1} X)(\hat{\Theta} - \Theta)}{\hat{\sigma}^2} \in F(p+1; n-p-1). \quad (6.4)$$

Доказательство этих фактов основано на известных свойствах многомерного нормального распределения (см. Приложение П3.1). В частности, в соответствии с (5.14') ОМНК-оценка $\hat{\Theta}$ является линейной комбинацией компонент $(X\Theta; \sigma^2 \Sigma_0)$ -нормально распределенного вектора Y , а именно: $\hat{\Theta} = CY$, где $C = (X^T \Sigma_0^{-1} X)^{-1} X^T \Sigma_0^{-1}$ — $(p+1) \times n$ -матрица. А поскольку из построения следует, что матрица C имеет *полный ранг* (то есть $\text{ранг } C = p+1$, напомним, что $p+1 < n$), то в соответствии с общим результатом, проведенным в П3.1, $\hat{\Theta} \in N_{p+1}(C \cdot EY; C \cdot \Sigma_Y \cdot C^T)$, откуда следует (6.2).

О факте (6.3) и взаимной статистической независимости оценок $\hat{\Theta}$ и $\hat{\sigma}^2$ уже упоминалось в главе 5 (см. формулу (5.19) и текст в конце п. 5.2). Доказательство можно найти, например, в книге [Магнус, Катышев, Пересецкий (2005), п. 3.3].

Доказательство факта (6.4) легко получить, опираясь на следующие два свойства многомерного нормального закона и связанных с ним распределений:

- (i) *если вектор $\xi \in N_k(a; \Sigma_\xi)$, причем, ковариационная матрица Σ_ξ невырождена, то*

$$(\xi - a)^T \Sigma^{-1} (\xi - a) \sim \chi^2(k);$$

- (ii) *если $\eta \sim \chi^2(k_1)$ и $\xi \sim \chi^2(k_2)$, причем, случайные величины η и ξ статистически независимы, то, по определению:*

$$\frac{\frac{1}{k_1} \eta}{\frac{1}{k_2} \xi} \in F(k_1; k_2)$$

В роли ξ в нашем случае выступает вектор оценок $\hat{\Theta}$, а в роли ξ — выражение левой части (6.3) (заметим, что свойствами (i)~(ii) мы уже пользовались в п. 4.6.1 при выводе свойства (в)).

Основываясь на фактах (6.2)~(6.4), можно построить следующие интервальные оценки и доверительные области для параметров θ_j ($j = 0, 1, 2, \dots, p$), σ^2 и Θ :

$$\hat{\theta}_j - t_\alpha(n-p-1) \cdot \hat{\sigma} \cdot \sqrt{\hat{a}_{jj}} < \theta_j < \hat{\theta}_j + t_\alpha(n-p-1) \cdot \hat{\sigma} \cdot \sqrt{\hat{a}_{jj}} \\ \text{(с доверительной вероятностью } P_0 = 1 - 2\alpha); \quad j = 0, 1, 2, \dots, p; \quad (6.5)$$

$$\frac{n-p-1}{\chi_\alpha^2(n-p-1)} \cdot \hat{\sigma}^2 < \sigma^2 < \frac{n-p-1}{\chi_{1-\alpha}^2(n-p-1)} \cdot \hat{\sigma}^2 \\ \text{(с доверительной вероятностью } P_0 = 1 - 2\alpha); \quad (6.6)$$

$$\Theta \in \left\{ \Theta: \frac{(\hat{\Theta} - \Theta)^T (\mathbf{X}^T \hat{\Sigma}_0^{-1} \mathbf{X}) (\hat{\Theta} - \Theta)}{(p+1)\hat{\sigma}^2} < F_{\alpha}(p+1; n-p-1) \right\}$$

(с вероятностью $P_0 = 1 - \alpha$). (6.7)

В соотношениях (6.5)~(6.7) $\hat{a}_{jj}$ — это j -й диагональный элемент матрицы $\hat{A} = (\mathbf{X}^T \hat{\Sigma}_0^{-1} \mathbf{X})^{-1}$ ($\hat{\Sigma}_0$ -оценка матрицы Σ_0 , полученная в ходе практически реализуемого ОМНК, см. § 5.4); величина $\hat{\sigma}^2$ определена формулой (5.19) с подстановкой в нее вместо матрицы Σ_0 ее оценки $\hat{\Sigma}_0$, а $t_q(k)$, $\chi_q^2(k)$ и $F_q(k_1; k_2)$ — это 100 q %-ные точки, соответственно, $t(k)$ -, $\chi^2(k)$ - и $F(k_1, k_2)$ -распределений. Выражение в фигурных скобках правой части (6.7) задает доверительную область для неизвестного Θ , определенную как множество всех тех значений векторов Θ , которые удовлетворяют данному неравенству.

6.2 Наилучший точечный прогноз $y(X)$ и $f(X) = E(y|X)$, основанный на ОЛММР

Сформулируем строго задачу, решаемую в данном параграфе.

Пусть мы располагаем исходными статистическими данными (X, Y) (см. (4.1a)–(4.3)) и нам известно, что статистическая связь между y и $X = (1, x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$ может быть описана регрессионной моделью вида (5.1), то есть удовлетворяет условиям обобщенной линейной модели множественной регрессии (ОЛММР). Соответственно известна (или оценена методами, описанными в предыдущей главе) ковариационная матрица регрессионных остатков S_{ε} . Помимо упомянутых исходных данных (X, Y) нам заданы также значения объясняющих переменных $X_{n+1} = (1, x_{n+1}^{(1)}, \dots, x_{n+1}^{(p)})^T$, то есть еще одна $(n+1)$ -я строка матрицы X , однако соответствующее значение результирующего показателя $y_{n+1} = y(X_{n+1})$ нам не известно. Известен, правда, характер ковариационных связей регрессионного остатка ε_{n+1} из уравнения $y_{n+1} = X_{n+1}^T \Theta + \varepsilon_{n+1}$ со всеми остальными регрессионными остатками $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$, описываемый вектор-столбцом

$$\sigma_{\varepsilon}^{(n+1)} = (E(\varepsilon_1 \varepsilon_{n+1}), E(\varepsilon_2 \varepsilon_{n+1}), \dots, E(\varepsilon_n \varepsilon_{n+1}))^T,$$

а также то, что $E\varepsilon_{n+1} = 0$ и $E\varepsilon_{n+1}^2 = \Delta^2$. Требуется построить наилучший (в смысле среднего квадрата ошибки) линейный относительно $y_1, y_2, \dots, y_n$ и несмещенный прогноз для неизвестного значения y_{n+1} .

Итак, по описанным выше исходным данным мы должны найти такой вектор-столбец коэффициентов $C_0 = (c_0^{(1)}, c_0^{(2)}, \dots, c_0^{(n)})^T$, который обладал бы следующим свойством:

$$E(y_{n+1} - C_0^T Y)^2 = \min_C E(y_{n+1} - C^T Y)^2, \quad (6.8)$$

где минимум берется только по таким векторам-столбцам коэффициентов $C = (c^{(1)}, c^{(2)}, \dots, c^{(n)})^T$, которые дают *несмещенные* оценки для y_{n+1} , то есть они должны удовлетворять условию:

$$E(C^T Y) = E y_{n+1}. \quad (6.9)$$

Обозначим

$$\hat{y}_C(X_{n+1}) = C^T Y = C^T (X\Theta + \epsilon) \quad (6.10)$$

линейный прогноз для $y_{n+1} = y(X_{n+1})$, удовлетворяющий условию несмещенности (6.9).

Чтобы получить решение поставленной задачи, выпишем и проанализируем средний квадрат ошибки прогноза (то есть левую часть соотношения (6.8)). Сама ошибка прогноза будет иметь вид:

$$\begin{aligned} \hat{y}_C(X_{n+1}) - y_{n+1} &= C^T (X\Theta + \epsilon) - X_{n+1}\Theta - \epsilon_{n+1} = \\ &= (C^T X - X_{n+1}^T)\Theta + (C^T \epsilon - \epsilon_{n+1}) = \\ &= C^T \epsilon - \epsilon_{n+1}, \end{aligned} \quad (6.11)$$

так как из условия несмещенности (6.9) следует:

$$E(C^T Y - y_{n+1}) = (C^T X - X_{n+1}^T)\Theta + E(C^T \epsilon - \epsilon_{n+1}) = (C^T X - X_{n+1}^T)\Theta = 0,$$

то есть вектор C должен удовлетворять условию

$$C^T X - X_{n+1}^T = 0. \quad (6.9')$$

Возвращаясь к (6.11), имеем:

$$\begin{aligned} E(\hat{y}_C(X_{n+1}) - y_{n+1})^2 &= E(C^T \epsilon - \epsilon_{n+1})^2 = \\ &= E[(C^T \epsilon - \epsilon_{n+1})(C^T \epsilon - \epsilon_{n+1})^T] = \\ &= E[C^T \epsilon \epsilon^T C + \epsilon_{n+1}^2 - 2C^T \epsilon \epsilon_{n+1}] = \\ &= C^T \Sigma_\epsilon C + E(\epsilon_{n+1}^2) - 2C^T \sigma_\epsilon^{(n+1)}. \end{aligned} \quad (6.12)$$

Далее минимизируем (6.12) при условии (6.9') по C . С этой целью выписываем соответствующую функцию Лагранжа $\varphi(\Lambda)$, дифференцируем ее по C и по Λ и приравниваем эти производные к нулю:

$$\varphi(\Lambda) = C^T \Sigma_\epsilon C + \Delta^2 - 2C^T \sigma_\epsilon^{(n+1)} - (C^T X - X_{n+1}^T)\Lambda, \quad (6.13)$$

где $\Lambda = (\lambda^{(0)}, \lambda^{(1)}, \dots, \lambda^{(p)})^T$.

Система уравнений

$$\begin{cases} \frac{\partial \varphi(\Lambda)}{\partial \Lambda} = 0_{p+1}, \\ \frac{\partial \varphi(\Lambda)}{\partial C} = 0_n \end{cases}$$

имеет вид

$$\begin{pmatrix} \mathbf{S}_\varepsilon & \mathbf{X} \\ \mathbf{X}^\top & \mathbf{0}_{(p+1) \times (p+1)} \end{pmatrix} \begin{pmatrix} C \\ -\Lambda \end{pmatrix} = \begin{pmatrix} \sigma_\varepsilon^{(n+1)} \\ \tilde{X}_{n+1} \end{pmatrix} \quad (6.14)$$

(здесь $\mathbf{0}_{(p+1) \times (p+1)}$, как и прежде, матрица размерности $(p+1) \times (p+1)$, состоящая из нулей).

Решение системы (6.14) относительно компонент вектора C и дает нам по построению коэффициенты C_0 для оптимального прогноза:

$$C_0 = \mathbf{S}_\varepsilon^{-1} [\mathbf{I}_n - \mathbf{X}(\mathbf{X}^\top \mathbf{S}_\varepsilon^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{S}_\varepsilon^{-1}] \sigma_\varepsilon^{(n+1)} + \mathbf{S}_\varepsilon^{-1} \mathbf{X}(\mathbf{X}^\top \mathbf{S}_\varepsilon^{-1} \mathbf{X})^{-1} \tilde{X}_{n+1}. \quad (6.15)$$

Таким образом, наилучшим линейным несмещенным прогнозом значения $y_{n+1} = y(X_{n+1})$ будет

$$\begin{aligned} \hat{y}_{C_0}(X_{n+1}) &= C_0^\top Y = X_{n+1}^\top (\mathbf{X}^\top \mathbf{S}_\varepsilon^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{S}_\varepsilon^{-1} Y + \\ &\quad + \sigma_\varepsilon^{(n+1)\top} \mathbf{S}_\varepsilon^{-1} Y - \sigma_\varepsilon^{(n+1)\top} \mathbf{S}_\varepsilon^{-1} \mathbf{X}(\mathbf{X}^\top \mathbf{S}_\varepsilon^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{S}_\varepsilon^{-1} Y = \\ &= X_{n+1}^\top \hat{\Theta}_{\text{ОМНК}} + (\sigma_\varepsilon^{(n+1)})^\top \mathbf{S}_\varepsilon^{-1} (Y - \mathbf{X} \hat{\Theta}_{\text{ОМНК}}) = \\ &= X_{n+1}^\top \hat{\Theta}_{\text{ОМНК}} + (\sigma_\varepsilon^{(n+1)})^\top \mathbf{S}_\varepsilon^{-1} \hat{\varepsilon}. \end{aligned} \quad (6.16)$$

В преобразованиях, приведших к соотношению (6.16), мы воспользовались правилом транспонирования правой части выражения (6.15), формулой (5.6) для $\hat{\Theta}_{\text{ОМНК}}$ и выражением вектора «невязок», $\hat{\varepsilon}$ в виде $Y - \mathbf{X} \hat{\Theta}_{\text{ОМНК}}$.

Анализируя формулу (6.16), мы видим, что наилучший прогноз значения $y(X_{n+1})$ определяется оцененным с помощью ОМНК значением функции регрессии в точке X_{n+1} (то есть величиной $X_{n+1}^\top \hat{\Theta}_{\text{ОМНК}}$) только в случае отсутствия корреляции между ε_{n+1} и остальными регрессионными остатками (из рассмотренных выше моделей этому условию удовлетворяют классическая ЛММР и модель с взаимно некоррелированными, но гетероскедастичными регрессионными остатками). В этом случае второе слагаемое в правой части соотношения (6.16) тождественно равно нулю, так как $\sigma_\varepsilon^{(n+1)} = (0, 0, \dots, 0)^\top$.

Итак, для КЛММР и для ОЛММР с гетероскедастичными (но взаимно некоррелированными) остатками наилучший (в смысле минимума среднего квадрата ошибки) прогноз значения $y(X_{n+1})$ определяется значением оцененной функции регрессии в этой точке, то есть

$$\hat{y}(X_{n+1}) = \hat{\Theta}^\top \cdot X_{n+1}, \quad (6.16')$$

где $\hat{\Theta} = \hat{\Theta}_{\text{МНК}}$ для КЛММР и $\hat{\Theta} = \hat{\Theta}_{\text{ОМНК}}$ для ЛММР с гетероскедастичными остатками.

Посмотрим, какой результат дает общая формула (6.16) в *рамках модели с автокоррелированными остатками*.

Модель с автокоррелированными остатками предполагает, что мы располагаем наблюдениями (X, Y) за n «тактов» времени, так что заданные значения объясняющих переменных $X_{n+1} = (1, x_{n+1}^{(1)}, \dots, x_{n+1}^{(p)})^T$ интерпретируются как прогнозные (или планируемые) значения предикторов в будущем такте времени « $n + 1$ ».

По условиям модели (5.36) регрессионные остатки ϵ подчиняются авторегрессионной схеме первого порядка, следовательно

$$\sigma_{\epsilon}^{(n+1)} = \sigma^2 \begin{pmatrix} \rho^n \\ \rho^{n-1} \\ \vdots \\ \rho \end{pmatrix} = \rho \sigma^2 \begin{pmatrix} \rho^{n-1} \\ \rho^{n-2} \\ \vdots \\ 1 \end{pmatrix},$$

так что вектор $\sigma_{\epsilon}^{(n+1)}$ может быть получен умножением на ρ последнего столбца матрицы Σ_{ϵ} , определенной формулой (5.4'). Но поскольку $S_{\epsilon} S_{\epsilon}^{-1} = I_n$, то произведение $(\sigma_{\epsilon}^{(n+1)})^T S_{\epsilon}^{-1}$ дает умноженную на ρ последнюю строку матрицы I_n и, следовательно:

$$(\sigma_{\epsilon}^{(n+1)})^T \Sigma_{\epsilon}^{-1} \hat{\epsilon} = \rho \hat{\epsilon}_n. \quad (6.17)$$

Возвращаясь к формуле (6.16) и принимая во внимание (6.17), имеем:

$$\hat{y}(X_{n+1}) = X_{n+1}^T \hat{\Theta}_{\text{ОМНК}} + \rho \hat{\epsilon}_n. \quad (6.18)$$

Строго говоря, использование прогноза (6.18) предполагает априорное знание величины ρ (напомним, что и оценка $\hat{\Theta}_{\text{ОМНК}}$ вычисляется с использованием этого значения). Однако она может применяться и при неизвестном значении ρ , если заменить в ней величину ρ ее значением, полученным на последней итерации *практически реализуемого ОМНК* (см. п. 5.4.3 и 5.5).

Построение наилучшего (в смысле минимума среднего квадрата ошибки) точечного прогноза для значения функции регрессии $f(X_{n+1}) = \Theta^T \cdot X_{n+1}$ при заданных значениях X_{n+1} объясняющих переменных опирается на теорему Гаусса–Маркова для КЛММП (см. п. 4.2.3) и теорему Эйткена для ОЛММП (см. п. 5.2), в соответствии с которыми наилучшей (среди всех линейных относительно $y_1, y_2, \dots, y_n$ и несмещенных оценок) оценкой для любой линейной комбинации неизвестных параметров $\theta_0, \theta_1, \dots, \theta_p$ является соответствующая линейная комбинация МНК- (в случае КЛММП) или ОМНК- (в случае ОЛММП) оценок этих параметров. Но функция регрессии y по X при $X = X_{n+1}$ это и есть линейная комбинация параметров:

$$f(X_{n+1}) = E(y|X = X_{n+1}) = \Theta^T \cdot X_{n+1},$$

так что наилучшая точечная оценка $\hat{f}(X_{n+1})$ функции $f(X_{n+1})$ задается соотношением:

$$\hat{f}(X_{n+1}) = \begin{cases} \hat{\Theta}_{\text{МНК}}^T \cdot X_{n+1} & \text{в КЛММР,} \\ \hat{\Theta}_{\text{ОМНК}}^T \cdot X_{n+1} & \text{в ОЛММР.} \end{cases} \quad (6.19)$$

6.3 Интервальный прогноз $y(X)$ и $f(X) = E(y|X)$, основанный на ОЛММР

Теперь перед нами поставлена следующая задача:

пользуясь результатами оценивания модели (6.1), построить (при заданной величине доверительной вероятности P_0) основанный на наилучшем линейном несмещенном точечном прогнозе $\hat{y}(X_{n+1})$ или $\hat{f}(X_{n+1})$ интервальный прогноз для значения результирующего показателя $y(X_{n+1})$ (соответственно, для значения функции регрессии $f(X_{n+1}) = E(y|X = X_{n+1})$), то есть указать такой интервал значений $[\hat{y}_{\min}^{(P_0)}(X_{n+1}), \hat{y}_{\max}^{(P_0)}(X_{n+1})]$ (соответственно, $[\hat{f}_{\min}^{(P_0)}(X_{n+1}), \hat{f}_{\max}^{(P_0)}(X_{n+1})]$), который с вероятностью, не меньшей P_0 , покроет неизвестное значение $y(X_{n+1})$ (соответственно, $f(X_{n+1})$).

Решение этой задачи для $y(X_{n+1})$ в общем случае основано на следующей схеме. По существу речь идет об оценке точности точечного прогноза, а точность определяется средним квадратом (или дисперсией) ошибки. Средний квадрат ошибки оптимального прогноза $\sigma_{\text{прогн.}}^2(X_{n+1}) = E(\hat{y}_{C_0}(X_{n+1}) - y(X_{n+1}))^2$ получаем, подставив в правую часть соотношения (6.12) оптимальные значения коэффициентов C_0 , определенные формулой (6.15) (из-за громоздкости получающегося при этом выражения не будем здесь его выписывать).

Поскольку точечный прогноз $\hat{y}_{C_0}(X_{n+1})$ по построению является линейной функцией от $y_1, y_2, \dots, y_n$, а последние в силу нормальности распределения остатков сами подчиняются нормальному закону, то и разность $\hat{y}_{C_0}(X_{n+1}) - y(X_{n+1})$ при весьма широких условиях является нормально распределенной случайной величиной со средним значением, равным нулю (в силу (6.9), то есть из-за свойства несмещенности прогноза), и с дисперсией $\sigma_{\text{прогн.}}^2(X_{n+1})$, определяемой соотношениями (6.12) и (6.15). Подставив в полученное таким образом выражение для $\sigma_{\text{прогн.}}^2(X_{n+1})$ оценки σ^2 и $\hat{\Sigma}_0$ для σ^2 и Σ_0 и воспользовавшись статистической взаимной независимостью оценок $\hat{\Theta}$ и $(\hat{\sigma}^2, \hat{\Sigma}_0)$, можно опереться далее на факт $t(n-p-1)$ -распределенности статистики

$$\frac{\hat{y}_{C_0}(X_{n+1}) - y(X_{n+1})}{\hat{\sigma}_{\text{прогн.}}(X_{n+1})} \sim t(n-p-1). \quad (6.20)$$

Это дает основание определить концы доверительного интервала для

$y(X_{n+1})$ в форме:

$$y_{\min}^{(P_0)}(X_{n+1}) = \hat{y}_{C_0}(X_{n+1}) \pm t_{\frac{1-P_0}{2}}(n-p-1) \cdot \hat{\sigma}_{\text{прогн.}}(X_{n+1}), \quad (6.21)$$

где $\hat{y}_{C_0}(X_{n+1})$ вычисляется по формуле (6.16).

Другими словами, можно гарантировать (с заданной вероятностью P), что величина результирующего показателя y при заданных значениях объясняющих переменных X_{n+1} будет заключена в пределах интервала $[\hat{y}_{\min}^{(P)}(\tilde{X}_{n+1}), \hat{y}_{\max}^{(P)}(\tilde{X}_{n+1})]$, где концы интервала определяются по формуле (6.30), а участвующая в них величина $\hat{y}_{C_0}(X_{n+1})$ — по формуле (6.16).

Случай классической линейной модели регрессии. Полученный результат весьма нагляден и прост в применении к классической ЛММР. Действительно, в соответствии с (4.1') в данном случае имеем:

$$\begin{aligned} \mathbf{S}_\varepsilon &= \sigma^2 \mathbf{I}_n, \\ \Delta^2 &= \mathbf{D}\varepsilon_{n+1} = \sigma^2, \\ \sigma_\varepsilon^{(n+1)} &= 0_n. \end{aligned}$$

Соответственно, с учетом формул (6.15), (6.16) и (6.12), получаем:

$$\begin{aligned} C_0 &= \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1} X_{n+1}, \\ \hat{y}_{C_0}(X_{n+1}) &= X_{n+1}^\top \hat{\Theta}_{\text{мнк}}, \\ \sigma_{\text{прогн.}}^2(X_{n+1}) &= \sigma^2 [X_{n+1}^\top (\mathbf{X}^\top \mathbf{X})^{-1} X_{n+1} + 1]. \end{aligned} \quad (6.22)$$

Применим полученный результат к построению интервального прогноза в классической модели *парной* регрессии (то есть при $p = 1$). Ранее (в п. 4.2.1) мы выписывали конкретный вид матриц $\mathbf{X}^\top \mathbf{X}$ и $\mathbf{X}^\top \mathbf{Y}$ в данной схеме. В частности:

$$\mathbf{X}^\top \mathbf{X} = \begin{pmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^n x_i^2 \end{pmatrix}, \quad \text{где } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Соответственно:

$$\begin{aligned} (\mathbf{X}^\top \mathbf{X})^{-1} &= \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n x_i^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix}, \\ X_{n+1}^\top (\mathbf{X}^\top \mathbf{X})^{-1} X_{n+1} &= (1 \quad x_{n+1}) \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n x_i^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix} \begin{pmatrix} 1 \\ x_{n+1} \end{pmatrix} \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}. \end{aligned}$$

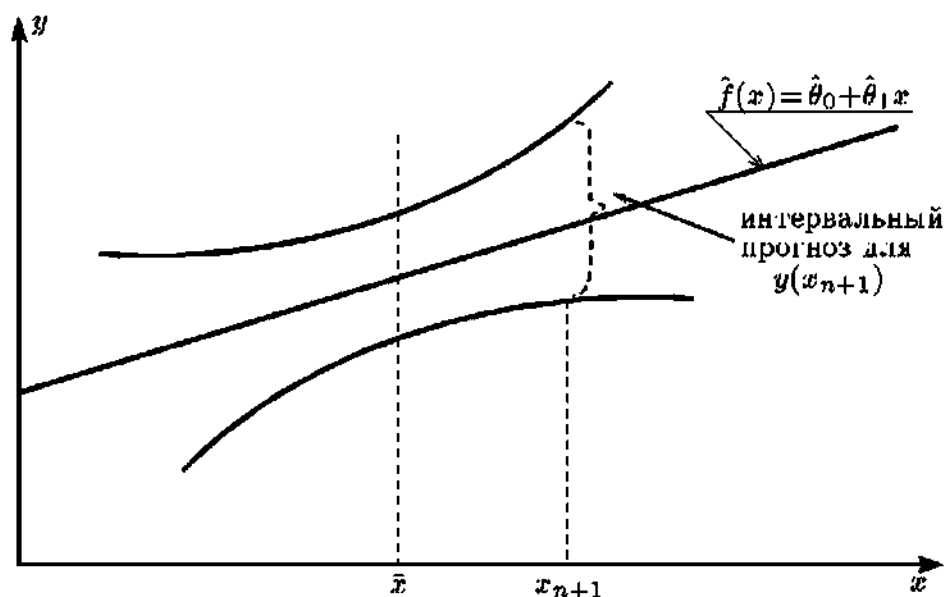


Рис. 6.1. Интервальный прогноз для значения зависимой переменной $y(x_{n+1})$ при заданной величине объясняющей переменной x_{n+1}

Подставляя полученное выражение в (6.22) и используя формулу (6.21), с учетом свойств классической модели имеем: прогнозируемое значение $y(X_{n+1})$ заключено (с вероятностью P_0) между величинами

$$\hat{y}_{\min}^{(P)}(\tilde{X}_{n+1}) = \hat{\theta}_{0. \text{мнк}} + \hat{\theta}_{1. \text{мнк}} x_{n+1} \mp t_{\frac{1-P_0}{2}}(n-2) \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}},$$

где $\hat{\theta}_{0. \text{мнк}}$, $\hat{\theta}_{1. \text{мнк}}$ — оценки коэффициентов парной регрессии обычным методом наименьших квадратов (см. (4.15)), x_{n+1} — заданное значение объясняющей переменной, а $\hat{\sigma}^2$ — оценка $\sigma^2 = D\varepsilon$ — дисперсии регрессионных остатков в классической модели линейной регрессии.

Рис. 6.1 дает представление о зависимости (в рассмотренной модели) ширины интервального прогноза от удаленности заданного значения объясняющей переменной x_{n+1} от ее средней величины $\bar{x}$.

Случай линейной модели с гетероскедастичными остатками. В этом случае (см. модель (5.1)~(5.2)) $E\varepsilon_{n+1}^2 = \frac{\sigma^2}{\lambda_{n+1}}$, остатки взаимнонекоррелированы (так что $\sigma_\varepsilon^{(n+1)} = 0_n$), поэтому в соответствии с (6.15) $C_0 = \Sigma_\varepsilon^{-1} X(X^\top \Sigma_\varepsilon^{-1} X)^{-1} X_{n+1}$, что после постановки этого выражения в правую часть (6.12) дает (с учетом того, что $\Sigma_\varepsilon = \sigma^2 \cdot \Sigma_0$):

$$\begin{aligned} \sigma_{\text{прогн.}}^2(X_{n+1}) &= E\left(\hat{y}_{C_0}(X_{n+1}) - y(X_{n+1})\right)^2 = \\ &= \sigma^2 \left[X_{n+1}^\top (X^\top \Sigma_0^{-1} X)^{-1} X_{n+1} + \frac{1}{\lambda_{n+1}} \right]. \end{aligned}$$

Подставляя сюда вместо σ^2 и Σ_0 их оценки $\hat{\sigma}^2$ и $\hat{\Sigma}_0$ (последняя получена, например, в ходе реализации критерия Бреуша-Пагана, см. п. 5.3.3)

и используя полученное выражение для $\hat{\sigma}_{\text{прогн.}}^2(X_{n+1})$ в (6.20), имеем:

$$\frac{\hat{\Theta}_{\text{ОМНК}}^T \cdot X_{n+1} - y(X_{n+1})}{\hat{\sigma} \cdot \sqrt{X_{n+1}^T (\mathbf{X}^T \hat{\Sigma}_0^{-1} \mathbf{X})^{-1} X_{n+1} + \lambda_{n+1}^{-1}}} \sim t(n-p-1),$$

откуда получаем интервальную оценку для $y(X_{n+1})$ вида:

$$y(X_{n+1}) \in \left[\hat{\Theta}_{\text{ОМНК}}^T \cdot X_{n+1} - t_{\alpha}(n-p-1) \cdot \hat{\sigma} \cdot \sqrt{X_{n+1}^T \mathbf{A} X_{n+1} + \lambda^{-1}}; \right. \\ \left. \hat{\Theta}_{\text{ОМНК}}^T \cdot X_{n+1} + t_{\alpha}(n-p-1) \cdot \hat{\sigma} \cdot \sqrt{X_{n+1}^T \mathbf{A} X_{n+1} + \lambda^{-1}} \right],$$

справедливую с вероятностью $P_0 = 1 - 2\alpha$ (здесь $\mathbf{A} = (\mathbf{X}^T \hat{\Sigma}_0^{-1} \mathbf{X})^{-1}$).

Случай линейной модели с автокоррелированными остатками (см. (5.36)). Это третий частный случай ОЛММР, подробно анализируемый в рамках основных курсов по эконометрике. Общая схема действий, приводящих к построению интервальной оценки для $y(X_{n+1})$, так же как и в предыдущих двух случаях, основана на соотношениях (6.12), (6.15), (6.16) и (6.18), а именно:

1) Оцениваются неизвестные параметры ρ , σ^2 , Σ_0 и $\mathbf{0}$ модели (см. § 5.4).

2) По формуле (6.15) вычисляются вектор коэффициентов C_0 оптимального линейного несмещенного прогноза (с подстановкой в нее $\Sigma_{\varepsilon} = \hat{\sigma}^2 \cdot \hat{\Sigma}_0$).

3) Формула (6.18) позволяет выписать разность $\hat{y}_{C_0}(X_{n+1}) - y(X_{n+1})$.

4) С использованием формулы (6.12) (и с учетом уже подсчитанных C_0 , а также заданных $E\varepsilon_{n+1}^2$ и $\sigma_{\varepsilon}^{(n+1)}$) вычисляется оценка $\hat{\sigma}_{\text{прогн.}}^2(X_{n+1})$ для $\mathbf{E}(\hat{y}_{C_0}(X_{n+1}) - y(X_{n+1}))^2$.

5) Основываясь на $t(n-p-1)$ -распределенности статистики $(\hat{y}_{C_0}(X_{n+1}) - y(X_{n+1}))/\hat{\sigma}_{\text{прогн.}}(X_{n+1})$ и, соответственно, на неравенстве

$$\frac{|\hat{y}_{C_0}(X_{n+1}) - y(X_{n+1})|}{\hat{\sigma}_{\text{прогн.}}(X_{n+1})} < t_{\alpha}(n-p-1) \text{ с вероятностью } P_0 = 1 - 2\alpha,$$

определяются концы доверительного интервала для $y(X_{n+1})$ в форме (6.21).

Интервальные прогнозные оценки значений функции регрессии в заданной точке. В некоторых ситуациях исследователя может интересовать прогноз не «индивидуального», значения результирующего признака y при заданных величинах объясняющих переменных $X_{n+1} = (1, x_{n+1}^{(1)}, \dots, x_{n+1}^{(p)})^T$, а *условного среднего значения* $f(\tilde{X}_{n+1}) = \mathbf{E}(y | X_{n+1})$. Так, в примере 2.1 (см. пп. 2.1, 5.1 и 5.3.2) вовсе не надуманным выглядит вопрос об оценке (точечной и интервальной) *средних* денежных сбережений семей с заданной величиной среднедушевого дохода x_{n+1} , а это и есть задача оценки значения функции регрессии при

заданном значении аргумента. И следует отличать ее от предыдущей задачи оценки (или прогноза) *индивидуального* значения результирующей переменной! В данном примере решение последней задачи потребовало бы построения точечного или интервального прогноза для значения денежных сбережений *одной отдельно взятой семьи* с уровнем душевого дохода, равным x_{n+1} .

Как уже было отмечено, ответ на вопрос о наилучшей *точечной* оценке значения линейной функции регрессии при заданном значении аргумента дается теоремой Гаусса–Маркова для классической модели (см. п. 4.3.4) и ее расширением на случай обобщенной модели (см. теорему Эйткена в п. 5.2): наилучшая точечная оценка (наилучший точечный прогноз) определяется значением эмпирической функции регрессии $\hat{f}(X_{n+1}) = X_{n+1}^\top \hat{\Theta}$ при заданном значении X_{n+1} объясняющей переменной (в качестве оценок $\hat{\Theta}$ используются МНК-оценки $\hat{\Theta}_{\text{МНК}}$ в условиях классической ЛММР и ОМНК-оценки $\hat{\Theta}_{\text{ОМНК}}$ в условиях обобщенной ЛММР).

Выведем формулу для интервального прогноза неизвестного значения линейной функции регрессии $f(X) = X^\top \Theta$ при заданном значении $X = X_{n+1}$ в условиях обобщенной ЛММР. Для этого вычислим сначала средний квадрат ошибки наилучшего точечного прогноза:

$$\begin{aligned} \sigma_f^2(X_{n+1}) &= \mathbf{E}(X_{n+1}^\top \hat{\Theta}_{\text{ОМНК}} - X_{n+1}^\top \Theta)^2 = \\ &= \mathbf{E}[X_{n+1}^\top (\hat{\Theta}_{\text{ОМНК}} - \Theta)(\hat{\Theta}_{\text{ОМНК}} - \Theta)^\top X_{n+1}] = \\ &= X_{n+1}^\top \mathbf{E}[(\hat{\Theta}_{\text{ОМНК}} - \Theta)(\hat{\Theta}_{\text{ОМНК}} - \Theta)^\top] X_{n+1} = \\ &= X_{n+1}^\top \Sigma_{\hat{\Theta}_{\text{ОМНК}}} X_{n+1} = \sigma^2 X_{n+1}^\top (\mathbf{X}^\top \mathbf{S}_0^{-1} \mathbf{X})^{-1} X_{n+1}. \end{aligned} \quad (6.23)$$

При выводе соотношения (6.23) мы воспользовались возможностью представить $[X_{n+1}^\top (\hat{\Theta}_{\text{ОМНК}} - \Theta)]^2$ в виде $[X_{n+1}^\top (\hat{\Theta}_{\text{ОМНК}} - \Theta)][X_{n+1}^\top (\hat{\Theta}_{\text{ОМНК}} - \Theta)]^\top$, а также видом ковариационной матрицы оценок $\hat{\Theta}_{\text{ОМНК}}$ (см. формулу (5.15')).

Итак, наилучшая линейная несмещенная точечная оценка $X_{n+1}^\top \hat{\Theta}_{\text{ОМНК}}$ $= X_{n+1}^\top (\mathbf{X}^\top \mathbf{S}_0^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{S}_0^{-1} \mathbf{Y}$ неизвестного значения функции регрессии $X_{n+1}^\top \Theta$ — это нормально распределенная случайная величина (поскольку она является линейной комбинацией нормально распределенных случайных величин $\mathbf{Y}$) со средним значением, равным $X_{n+1}^\top \Theta$, и с дисперсией $\sigma_f^2(X_{n+1})$, определяемой по формуле (6.23).

Поэтому, заменяя вычисленный по формуле (6.23) средний квадрат ошибки $\sigma_f^2(X_{n+1})$ ее оценкой $\hat{\sigma}_f^2(X_{n+1})$ и учитывая, что отношение $(\hat{f}(X_{n+1}) - f(X_{n+1})) / \hat{\sigma}_f(X_{n+1})$ должно подчиняться $t(n-p-1)$ -распределению, получаем следующие выражения для концов $\hat{f}_{\min}^{(P_0)}(X_{n+1})$ довери-

тельного интервала для $f(X_{n+1}) = \mathbf{E}(y|X = X_{n+1})$:

$$\begin{aligned} \hat{f}_{\min}^{(P_0)}(X_{n+1}) &= \\ &= \hat{\Theta}_{\text{омнк}}^\top \cdot X_{n+1} \pm t_\alpha(n-p-1) \cdot \hat{\sigma} \cdot \sqrt{X_{n+1}^\top \cdot (\mathbf{X}^\top \hat{\Sigma}_0^{-1} \mathbf{X}) X_{n+1}}. \end{aligned} \quad (6.24)$$

Таким образом, можно гарантировать, что с доверительной вероятностью $P_0 = 1 - 2\alpha$ прогнозируемое значение функции регрессии $f(X_{n+1}) = \mathbf{E}(y|X = X_{n+1})$ будет принадлежать интервалу $[\hat{f}_{\min}^{(p_0)}(X_{n+1}), \hat{f}_{\max}^{(P_0)} \times \times (X_{n+1})]$, концы которого определены формулой (6.24).

Рекомендации для классической ЛММР получаются из (6.23)-(6.24) в качестве частного случая $\mathbf{S}_0 = \mathbf{I}_n$. В частности, для классической модели парной регрессии (то есть при $p = 1$) нами было подсчитано (см. выше) выражение $X_{n+1}^\top (\mathbf{X}^\top \mathbf{X})^{-1} X_{n+1}$, необходимое для вычисления $\sigma_f^2(\tilde{X}_{n+1})$. Используя его, с помощью формулы (6.23) получаем выражение для $\hat{\sigma}_f^2(X_{n+1})$ в данном простом случае:

$$\hat{\sigma}_f^2(X_{n+1}) = \hat{\sigma}^2 \left[\frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right].$$

Подставляя это выражение в (6.24), получаем интервальный прогноз неизвестного значения функции регрессии в заданной точке x_{n+1} в условиях классической линейной модели парной регрессии.

З а м е ч а н и е. Следует отметить один достаточно общий факт. Точность прогноза значения функции регрессии $f(X_{n+1})$ мы можем сколь угодно повышать (то есть сколь угодно сжимать ширину соответствующего доверительного интервала при любом заданном значении доверительной вероятности P_0) за счет роста числа n имеющихся в нашем распоряжении наблюдений, в то время как при прогнозировании индивидуального значения $y(X_{n+1})$ это сжатие ширины доверительного интервала (при $n \rightarrow \infty$) возможно лишь до определенного уровня, обусловленного величиной дисперсии $D\varepsilon(X_{n+1})$. Этот факт наглядно демонстрируется на примере парной классической регрессии (см. выше). Ширина доверительного интервала для $f(x_{n+1})$ пропорциональна значению

$$\Delta_n(f) = \sqrt{\frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}},$$

которое при достаточно общих условиях стремится к нулю при $n \rightarrow \infty$. В то же время ширина доверительного интервала для $y(x_{n+1})$ пропорциональна значению $\Delta_n(y) = \sqrt{\Delta_n^2(f) + 1}$, которое уже не стремится к нулю при $n \rightarrow \infty$.

6.4 Анализ точности регрессионной модели и прогнозирование в условиях реалистической ситуации

Приведенные в п. 6.1–6.3 правила и рекомендации по анализу точности построенной регрессионной модели основаны на ряде базовых допущений, среди которых, в первую очередь, следует назвать:

- (А) *Общий функциональный вид модели (ее структура) угадан точно (в нашем случае постулировался линейный относительно $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ вид).*
- (Б) *Число и состав объясняющих переменных модели (в рамках угаданного точно общего вида модели) определены без ошибок.*
- (В) *Заданные значения $X_{n+1} = (1, x_{n+1}^{(1)}, \dots, x_{n+1}^{(p)})^T$ объясняющих переменных, при которых строится прогноз для $y(X_{n+1})$ и/или для $f(X_{n+1}) = E(y|X = X_{n+1})$, также определены без ошибок.*

К сожалению, одновременное выполнение на практике всех этих постулатов следует отнести, скорее, к счастливым исключениям, чем к обычному положению вещей. Поэтому когда мы говорим об условиях реалистической ситуации, мы имеем в виду нарушение хотя бы одного из постулатов (А)–(В).

Наиболее неприятные последствия влечет нарушение постулата (А), связанного со спецификацией регрессионной модели, а именно, — *неправильное определение параметрического семейства функций*, к которому должна принадлежать искомая функция регрессии. В подобных ситуациях приходится говорить лишь об аппроксимационных вариантах регрессионных моделей. Применительно к построению и исследованию линейных регрессионных моделей (классической и обобщенной) это означает, что функцию $f_a(X; \Theta) = \theta_0 + \theta_1 x^{(1)} + \dots + \theta_p x^{(p)}$ мы должны рассматривать лишь как некоторую аппроксимацию неизвестной нам функции $f(X) = E(y | X)$. В этих случаях могут оказаться полезными следующие полуэвристические рекомендации при построении прогнозов и исследовании точности статистических выводов в регрессионном анализе:

1) При анализе точности аппроксимационных вариантов регрессионных моделей не следует претендовать на построение *точных* доверительных интервалов ни для неизвестных значений параметров Θ (они, как правило, не имеют в данной ситуации самостоятельной содержательной интерпретации), ни для функции регрессии $f(X)$ или результирующего показателя $y(X)$ (поскольку, пользуясь аппроксимацией $\hat{f}_a(X)$, отличающейся по структуре от истинной функции регрессии $f(X)$, мы

не можем иметь достоверной априорной информации о вероятностной природе остатков $\hat{\varepsilon}_i = y_i - \hat{f}_a(X_i)$.

2) Имеющуюся выборку наблюдений $\tilde{B}_n = \{(X_1, y_1), \dots, (X_n, y_n)\}$ целесообразно разбить (одним или несколькими различными способами) на две непересекающиеся подвыборки объемов n_1 и n_2 ($n_1 + n_2 = n$): обучающую $\tilde{B}_{n_1}^{об}$, на основании наблюдений которой строятся оценки $\hat{\Theta}^{(n_1)}$ неизвестных параметров аппроксимационной функции регрессии $f_a(X; \Theta)$, и экзаменующую (или контрольную) $\tilde{B}_{n_2}^{экз}$, по наблюдениям которой оцениваются основные характеристики точности анализируемой модели (в первую очередь регрессионные остатки $\hat{\varepsilon}_i = y_i - \hat{f}_a(X_i; \hat{\Theta}^{(n_1)})$). Пусть мы произвели такое разбиение исходной выборки $\tilde{B}_n = (X, Y)$ на две подвыборки — обучающую $\tilde{B}_{n_{1j}} = (X_j(1), Y_j(1))$ и экзаменующую $\tilde{B}_{n_{2j}} = (X_j(2), Y_j(2))$ k различными способами (то есть $j = 1, 2, \dots, k$).

3) Одной из основных характеристик точности аппроксимационного варианта линейной регрессионной модели является оценка $\hat{\sigma}$ среднеквадратической ошибки аппроксимации σ , вычисляемая по формуле

$$\hat{\sigma}^2 = \frac{1}{\sum_{j=1}^k (n_{2j} - p - 1)} \sum_{j=1}^k (Y_j(2) - X_j(2) \hat{\Theta}^{(n_{1j})})^T \times \times (Y_j(2) - X_j(2) \hat{\Theta}^{(n_{1j})}). \quad (6.25)$$

В соотношении (6.25) оценки $\hat{\Theta}^{(n_{1j})}$ вычислены в соответствии с формулами соответственно (4.19) и (5.6) обычного или обобщенного методов наименьших квадратов по наблюдениям *только обучающей* выборки (соответствующей j -му способу разбиения выборки $\tilde{B}_n$ на обучающую и экзаменующую). Знание $\hat{\sigma}$ позволяет оценить максимально возможную погрешность аппроксимации неизвестной функции регрессии $f(X)$ (в пределах обследованного диапазона значений X) приблизительно величиной порядка $\pm 2\hat{\sigma}/\sqrt{n}$, а результирующего показателя $y(X)$ — величиной порядка $\pm 2\hat{\sigma}$;

Поясним подробнее реализацию положений 2) и 3) на примере использования широко применяемого метода *скользящего экзамена*.¹ Определим n разбиений исходной выборки $\tilde{B}_n = \{(X_1, y_1), \dots, (X_n, y_n)\}$ на обучающую ($B_{1j}^{об}$) и экзаменующую ($B_{n_{2j}}^{экз}$) следующим образом:

$$\begin{cases} \tilde{B}_{n_{1j}}^{об} = \{(X_1, y_1), \dots, (X_{j-1}, y_{j-1}), (X_{j+1}, y_{j+1}), \dots, (X_n, y_n)\}; \\ B_{n_{2j}}^{экз} = \{X_j, y_j\}, \quad j = 1, 2, \dots, n. \end{cases}$$

¹ Этот метод (в зарубежной литературе он называется «методом складного ножа», или «jackknife method») является одним из вариантов реализации общего подхода, впервые предложенного, по-видимому, в связи с задачей устранения смещения в статистических оценках (см. [Кендалл, Стьюарт (1973)]).

Таким образом: $n_{1j} = n - 1$ и $n_{2j} = 1$ для всех $j = 1, 2, \dots, n$; j -й вариант обучающей выборки содержит все наблюдения исходной выборки $\tilde{B}_n$ кроме одного — (X_j, y_j) ; соответственно j -й вариант экзаменующей выборки содержит единственное наблюдение — (X_j, y_j) . Применение к такой последовательности обучающих и экзаменующих выборок формулы (6.25) и составляет метод скользящего экзамена оценки точности статистического вывода.

Величина среднеквадратической погрешности $\hat{\sigma}$, подсчитанная с помощью метода скользящего экзамена, в аппроксимационных схемах регрессии оказывается, как правило, существенно больше аналогичной характеристики, вычисленной с помощью обычной формулы (4.31).

З а м е ч а н и е (по поводу вычислительной реализации метода скользящего экзамена). На первый взгляд реализация метода скользящего экзамена связана с многократным повторением громоздких вычислений на ЭВМ. Действительно, процедура предусматривает n -кратное вычисление оценок $\hat{\Theta}^{(n_{11})}, \hat{\Theta}^{(n_{12})}, \dots, \hat{\Theta}^{(n_{1n})}$, n^2 -кратное вычисление выборочных функций регрессии $f_a(X_i; \hat{\Theta}^{(n_{1j})}) (i, j = 1, 2, \dots, n)$ и т. д. Однако непосредственный анализ основных формул метода наименьших квадратов в случае линейного вида аппроксимирующих функций $f_a(X; \Theta) = \tilde{X}^\top \Theta$ позволяет установить полезные соотношения между интересующими нас характеристиками, подсчитанными по всей выборке $\tilde{B}_n$, и теми же характеристиками, подсчитанными по выборке, в которой нет наблюдения (X_i, y_i) :

$$\hat{\Theta}^{(n_{1i})} = \hat{\Theta} + \frac{y_i - \tilde{X}_i^\top \hat{\Theta}}{1 - q_i} \cdot (\mathbf{X}^\top \mathbf{X})^{-1} \cdot \tilde{X}_i,$$

где $q_i = \tilde{X}_i^\top \cdot (\mathbf{X}^\top \mathbf{X})^{-1} \cdot \tilde{X}_i$;

$$\tilde{X}_i^\top \hat{\Theta}^{(n_{1i})} = \tilde{X}_i^\top \hat{\Theta} + q_i \frac{y_i - \tilde{X}_i^\top \hat{\Theta}}{1 - q_i};$$

$$\hat{\varepsilon}_i(\hat{\Theta}^{(n_{1i})}) = \frac{\hat{\varepsilon}_i(\hat{\Theta})}{1 - q_i}, \quad \text{где} \quad \hat{\varepsilon}_i(\hat{\Theta}) = y_i - f_a(X_i; \hat{\Theta});$$

$$\sum_{i=1}^n \hat{\varepsilon}_i^2(\hat{\Theta}^{(n_{1i})}) = \sum_{i=1}^n \frac{1}{(1 - q_i)^2} \hat{\varepsilon}_i^2(\hat{\Theta}).$$

Эти соотношения позволяют избежать многократной вычислительной «прогонки» процедур метода наименьших квадратов на различных вариантах обучающей выборки за счет пересчета значений $\hat{\Theta}^{(n_{1i})}$,

² При самых естественных ограничениях на структуру матрицы $\mathbf{X}$ величина q_i , несмотря на зависимость от X_i , ведет себя при $n \rightarrow \infty$ (при фиксированной размерности $p + 1$ оцениваемого параметра) как n^{-1} ; если же и $p \rightarrow \infty$, то $q_i \sim (p + 1)/n$.

$f(X_i; \hat{\Theta}^{(n_{1i})}) = \tilde{X}_i^T \hat{\Theta}^{(n_{1i})}$ и т. д. по соответствующим характеристикам, подсчитанным по наблюдениям *всей* выборки $\tilde{B}_n$.

4) Если наблюдения (X_t, y_t) , $-t = 1, 2, \dots, n$, — *упорядочены во времени* (то есть номер наблюдения t определяет такт времени регистрации значений X и y), то весьма распространенной является так называемая **ретроспективная оценка относительной погрешности $\Delta_{\text{отн.}}(\tau)$ прогноза величины y на τ тактов времени вперед**. Величина $\Delta_{\text{отн.}}(\tau)$ определяется по формуле

$$\Delta_{\text{отн.}}(\tau) = \frac{1}{n - n_1 - \tau + 1} \sum_{t=n_1}^{n-\tau} \left| \frac{\hat{y}_{t+\tau}(t) - y_{t+\tau}}{y_{t+\tau}} \right| \cdot 100\%, \quad (6.26)$$

в которой n — общее число имеющихся наблюдений (и одновременно, номер самого позднего такта времени регистрации наблюдений), τ -горизонт прогнозирования, $\hat{y}_{t+\tau}(t)$ — прогнозное (модельное) значение результирующего показателя в момент времени $t + \tau$, подсчитанное по модели, параметры которой *оценены только по первым t наблюдениям*, а n_1 — объем минимальной обучающей выборки. При этом число n_1 выбирается обычно так, чтобы, с одной стороны, оно, по возможности, существенно (хотя бы в 2–3 раза) превосходило число оцениваемых параметров анализируемой модели, а с другой — чтобы число слагаемых $n - n_1 - \tau + 1$ в сумме правой части (6.26) оказывалось не меньше 8–10. Как мы видим, этот способ анализа точности прогноза также является одной из специальных форм реализации идей разбиения имеющейся выборки на «обучающую» и «экзаменующую» с использованием приема *скользящего экзамена*.

5) Следует проявлять известную сдержанность и осторожность при использовании аппроксимационных вариантов регрессионных моделей для решения задач интерполяции и (особенно) экстраполяции, то есть при восстановлении неизвестного значения функции регрессии $f(X)$ или результирующего показателя $\eta(X)$ по значению предиктора X , *лежащему вне статистически обследованной области значений объясняющих переменных*.

Характер возможных искажений в оценивании анализируемой линейной модели регрессии, **возникающих при нарушении постулата (Б)** (то есть при потерях или излишествах в наборе объясняющих переменных), был обсужден в п. 4.4.3.

В заключение рассмотрим влияние нарушения постулата (В) на точность оцененной КЛММР и основанного на ней прогноза $y(X_{n+1})$. пусть значения объясняющих переменных X_{n+1} заданы с некоторой ошибкой δ , то есть мы располагаем значениями

$$\tilde{X}_{n+1} = X_{n+1} + \delta, \quad (6.27)$$

где случайные ошибки $\delta = (\delta^{(0)}, \delta^{(1)}, \dots, \delta^{(p)})^T$ нормальны и статистически взаимно независимы с регрессионными остатками $(\varepsilon(X_1), \varepsilon(X_2), \dots, \varepsilon(X_n), \varepsilon(X_{n+1}))$ (а значит, и с оценками $\hat{\Theta}$), причем, $E\delta = 0_{p+1}$ и $\Sigma_\delta = \sigma_0^2 \cdot I_{p+1}$.

Тогда в качестве прогноза мы будем иметь

$$\hat{y}(X_{n+1}) = \hat{\Theta}^T \tilde{X}_{n+1}. \quad (6.28)$$

Нетрудно показать, что

$$E(\hat{y}(X_{n+1}) - y(X_{n+1})) = 0,$$

то есть что оценка (6.28) является несмещенной, и что

$$E(\hat{y}(X_{n+1}) - y(X_{n+1}))^2 = \sigma^2 [X_{n+1}^T (X^T X)^{-1} X_{n+1} + 1] + \sigma^2 \cdot \sigma_0^2 \cdot \text{tr}((X^T X)^{-1}) + \sigma_0^2 \cdot \Theta^T \cdot \Theta \quad (6.30)$$

(оставляем доказательство этих фактов в качестве самостоятельного упражнения читателю).

Сравнение (6.30) с (6.22) показывает, что наличие ошибок в задании объясняющих переменных увеличивает погрешность прогноза, добавляя к среднему квадрату ошибки прогноза (6.22) два положительных слагаемых, пропорциональных дисперсии $\sigma_0^2 = D\delta^{(j)}$.

Отметим, к тому же, что при добавлении ошибок в задание значений X_{n+1} прогнозное значение $\hat{y}(X_{n+1})$ представляется уже не как нормальная случайная величина $\hat{\Theta}^T X_{n+1}$, а как скалярное произведение двух независимых нормальных векторов (см. (6.28)), что лишает нас возможности *аналитически* вывести доверительный интервал для $y(X_{n+1})$.

Выводы

1. Восстановление (прогноз) неизвестного значения зависимой переменной y по заданным значениям X_{n+1} объясняющих переменных является главной целью, с которой строится и анализируется модель регрессии y по X . При этом, прогноз значения $y(X_{n+1})$ может быть *точечным* (определенным в виде числа $\hat{y}(X_{n+1})$) и *интервальным* (определенным в виде интервала значений $[\hat{y}_{\min}^{(P_0)}(X_{n+1}), \hat{y}_{\max}^{(P_0)}(X_{n+1})]$), попадание в который прогнозируемого значения $y(X_{n+1})$ гарантируется с наперед заданной доверительной вероятностью P_0). Та же задача может быть поставлена и в отношении неизвестного *регрессионного* значения $f(X_{n+1}) = E(y|X = X_{n+1})$.

2. Построение точечного и интервального прогноза для $y(X_{n+1})$ и $f(X_{n+1}) = E(y|X = X_{n+1})$ опирается на результаты анализа точности анализируемой ЛММР и, в частности, на свойства (моменты, законы

распределения вероятностей) оценок параметров этой модели и ее регрессионных остатков. Эти свойства (см. (6.2)–(6.4)) позволяют строить интервальные оценки и доверительные области для самих параметров анализируемой модели (см. (6.5)–(6.7)).

3. Наилучший (в смысле среднего квадрата ошибки) линейный (относительно $y_1, y_2, \dots, y_n$) несмещенный точечный прогноз $\hat{y}(X_{n+1})$ неизвестного значения $y(X_{n+1})$ по заданным значениям объясняющих переменных $X_{n+1} = (1, x_{n+1}^{(1)}, \dots, x_{n+1}^{(p)})^T$ в КЛММР и в линейной модели множественной регрессии с гетероскедастичными, взаимнонекоррелированными остатками определяется значениями, соответственно, $\hat{\Theta}_{\text{мнк}}^T \cdot X_{n+1}$ и $\hat{\Theta}_{\text{омнк}}^T \cdot X_{n+1}$ оцененной функции регрессии в точке X_{n+1} . Однако для ЛММР с автокоррелированными остатками прогнозное значение $\hat{y}(X_{n+1})$ определяется другой формулой, а именно:

$$\hat{y}(X_{n+1}) = \hat{\Theta}_{\text{омнк}}^T \cdot X_{n+1} + \hat{\rho} \cdot (y_n - \hat{\Theta}_{\text{омнк}}^T \cdot X_n).$$

4. Наилучший точечный прогноз $\hat{f}(X_{n+1})$ значения функции регрессии $f(X_{n+1}) = E(y|X = X_{n+1})$ в точке X_{n+1} определяется значением оцененной функции регрессии в этой точке, то есть $\hat{f}(X_{n+1}) = \hat{\Theta}^T \cdot X_{n+1}$. При этом, в случае КЛММР в качестве $\hat{\Theta}$ используется $\hat{\Theta}_{\text{мнк}}$, а в моделях с гетероскедастичными и автокоррелированными остатками — $\hat{\Theta}_{\text{омнк}}$.

5. Вывод общей формулы (6.21) для интервального прогноза значения $y(X_{n+1})$ зависимой переменной в точке X_{n+1} основан на $t(n-p-1)$ -распределенности отношения

$$\frac{\hat{y}(X_{n+1}) - y(X_{n+1})}{\hat{\sigma}_{\text{прогн.}}(X_{n+1})},$$

где $\hat{y}(X_{n+1})$ — наилучший линейный точечный прогноз значения $y(X_{n+1})$ (см. п. 3 «Выводов»), а $\hat{\sigma}_{\text{прогн.}}^2(X_{n+1})$ — оценка среднего квадрата $E(\hat{y}(X_{n+1}) - y(X_{n+1}))^2$ ошибки точечного прогноза $\hat{y}(X_{n+1})$. Умение вычислить наилучший точечный прогноз $\hat{y}(X_{n+1})$, а также $E(\hat{y}(X_{n+1}) - y(X_{n+1}))^2$ и ее оценку (см. (6.16), (6.12)) позволяет выписать в аналитической форме интервальный прогноз значения $y(X_{n+1})$ для всех трех рассматриваемых ЛММР-классической, с гетероскедастичными остатками и с автокоррелированными остатками.

6. Вывод интервальных прогнозов для неизвестных значений $f(X_{n+1}) = E(y|X = X_{n+1})$ функций регрессии y по X основан на той же идее, а именно, на $t(n-p-1)$ -распределенности отношения

$$\frac{\hat{f}(X_{n+1}) - f(X_{n+1})}{\hat{\sigma}_f(X_{n+1})},$$

где $\hat{f}(X_{n+1})$ — полученный ранее (см. п. 2 «Выводов») наилучший точечный линейный прогноз значения $f(X_{n+1})$, а $\hat{\sigma}_f^2(X_{n+1})$ — оценка среднего

квадрата ошибки $E(\hat{f}(X_{n+1}) - f(X_{n+1}))^2$. Реализация этой идеи позволяет получить относительно простые аналитические выражения интервальных прогнозов для $f(X_{n+1})$ в рамках всех трех упомянутых выше линейных моделей множественной регрессии (см. (6.24)).

7. Строгие математические выводы по анализу точности построенной модели регрессии и прогнозированию значений зависимой переменной $y(X)$ и функции регрессии $f(X) = E(y|X)$ приходится дополнять полуэвристическими рекомендациями в случае так называемых *реалистических ситуаций*, характеризуемых неизбежными на практике нарушениями ряда постулатов, на которых базируются рассматриваемые ЛММР. Среди этих нарушений: (А) отклонение общего функционального вида анализируемой модели от постулируемого (в нашем случае – от линейного); (Б) нехватка или избыток в рассматриваемой модели объясняющих переменных; (В) неточность задания значений объясняющих переменных, при которых требуется построить прогноз для y или f .

8. Среди рекомендаций, которые могут оказаться полезными в подобных реалистических ситуациях, выделим идею деления имеющихся исходных данных на две подвыборки: *обучающую*, по которой оцениваются параметры модели, и *экзаменующую*, которая используется для анализа точности построенной модели и прогнозирования. Идею, дополненную приемом «скользящего экзамена» (см. метод «Jackknife», а также «метод ретроспективной оценки относительной погрешности прогноза» (6.26)).

Глава 7

Линейные модели регрессии со стохастическими объясняющими переменными

До сих пор мы изучали линейную модель множественной регрессии (ЛММР) в предположении, что объясняющие переменные $X = (1, x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$ и, соответственно, матрица наблюдаемых значений объясняющих переменных X не являются случайными, стохастическими по своей природе. Другими словами, значения $x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)}$ ($i = 1, 2, \dots, n$) играли роль неслучайных параметров, от которых зависит закон распределения вероятностей случайной величины y , связанной с $x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)}$ соотношениями типа (4.2). В этом соотношении функция регрессии $E(y | X)$, строго говоря, должна пониматься не как *условное математическое ожидание* случайной величины y (при условии, что значения стохастических объясняющих переменных зафиксированы на уровнях X), а как *безусловное математическое ожидание случайной величины y при значениях параметров (от которых зависит ее закон распределения вероятностей), равных X .*

В ряде ситуаций «неслучайность» объясняющих переменных очевидна и обусловлена, например, способом получения исходных статистических данных. Так, в примере 2.1 анализируемые семьи были предварительно разбиты на 4 группы, однородные по величине объясняющей переменной x : низкодоходные ($x = x_1^0 = 80$ ден.ед.), с доходом ниже среднего уровня ($x = x_2^0 = 120$ ден.ед.), среднедоходные ($x = x_3^0 = 160$ ден.ед.) и высокодоходные ($x = x_4^0 = 200$ ден.ед.). Конечно, в данном примере объясняющая переменная играет роль неслучайного параметра x , от значения которого зависит закон распределения интересующей нас случайной величины — суммы удельных сбережений в семье $y(x)$. Однако видоизменив способ получения исходных статистических данных,

а именно, извлекая случайным образом семью из анализируемой генеральной совокупности семей и фиксируя в каждой такой (i -й) извлеченной семье одновременно значения *двух* признаков (x_i, y_i) , мы уже будем иметь дело с наблюдениями *двумерной случайной величины* (x, y) , так что *и объясняющая переменная x должна интерпретироваться при такой схеме наблюдений как случайная величина*. Соответственно и функция регрессии $f(x) = E(y | x)$ должна пониматься уже как условное математическое ожидание величины удельных денежных сбережений в семье (при условии, что значение среднедушевого семейного дохода зафиксировано на уровне x).

Надо признать, что в практике статистических исследований и эконометрического моделирования очень часто довольно трудно определить, можно ли считать объясняющие переменные неслучайными. К тому же существует весьма широкий спектр реальных задач регрессионного анализа, в которых *объясняющие переменные имеют явно стохастическую природу*. Помимо схем получения исходных статистических данных, аналогичных описанной выше видоизмененной схеме примера 2.1, к таким задачам относятся все ситуации, в которых *значения объясняющих переменных могут измеряться только со случайной ошибкой*, а также весьма распространенные в эконометрических исследованиях задачи, в которых в качестве объясняющих переменных используются значения случайных по своей природе *результатирующих* показателей в *предыдущие* моменты времени. Поэтому хотелось бы уметь проводить статистический анализ моделей регрессии, в которых объясняющие переменные стохастичны по своей природе. Нам будет удобно разбить описание регрессионного анализа линейных моделей со *стохастическими* объясняющими переменными на три случая:

- *объясняющие переменные X статистически независимы от регрессионных остатков ϵ и их распределение не зависит от оцениваемых параметров модели Θ и σ^2 ;*
- *объясняющие переменные X коррелированы с регрессионными остатками ϵ ;*
- *объясняющие переменные могут быть измерены только со случайными ошибками.*

При этом, как и прежде, рассматривается ЛММР вида (обозначения определены выше соотношениями (4.1а), (4.3)–(4.10))

$$Y = X\Theta + \epsilon, \quad (7.1)$$

где

$$\text{ранг } \mathbf{X} = p + 1 \text{ (с вероятностью единица)}^1. \quad (7.2)$$

Что касается спецификации условий, определяющих природу регрессионных остатков $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^\top$, то пока мы их сформулируем в достаточно общем виде:

$$\mathbf{E}\varepsilon = \mathbf{O}_n, \quad (7.3)$$

$$\Sigma_\varepsilon = \mathbf{E}(\varepsilon\varepsilon^\top) = \sigma^2 \Sigma_0, \quad (7.4)$$

где усреднение (операция вычисления математического ожидания $\mathbf{E}$) производится в *безусловном* смысле, то есть по *совместному* з.р.в. случайных величин ε и $\mathbf{X}$. При этом, вообще говоря, $(n \times n)$ -матрица Σ_0 может зависеть от неизвестных параметров, в том числе и от параметров Θ .

7.1 Случайные остатки ε не зависят от предикторов $\mathbf{X}$ и оцениваемых коэффициентов регрессии Θ

Из взаимной независимости ε и $\mathbf{X}$ следует взаимная независимость ε и $\mathbf{X}$. В этом случае *безусловные* соотношения (7.3) и (7.4) влекут за собой справедливость соответствующих соотношений для *условных* математических ожиданий

$$\mathbf{E}(\varepsilon | \mathbf{X}) = \mathbf{O}_n, \quad (7.3')$$

$$\Sigma_\varepsilon(\mathbf{X}) = \mathbf{E}(\varepsilon\varepsilon^\top | \mathbf{X}) = \sigma^2 \Sigma_0, \quad (7.4')$$

где матрица Σ_0 не зависит ни от $\mathbf{X}$, ни от Θ .

Заметим, что условия (7.3') и (7.4') эквивалентны условиям

$$\mathbf{E}(Y | \mathbf{X}) = \mathbf{X}\Theta, \quad (7.3'')$$

$$\Sigma_Y(\mathbf{X}) = \mathbf{E}\left\{(Y - \mathbf{X}\Theta)(Y - \mathbf{X}\Theta)^\top | \mathbf{X}\right\} = \sigma^2 \Sigma_0 \quad (7.4'')$$

и что из общих свойств многомерного нормального распределения следует, в частности, выполнение этих условий в ситуациях, когда вектор

¹Условие (7.2) сопровождается требованием его выполнения *с вероятностью единица*, поскольку матрица $\mathbf{X}$ в рамках данной постановки задачи формируется из элементов случайной выборки $x_i^{(j)}$ ($i = 1, 2, \dots, n$; $j = 1, 2, \dots, p$).

анализируемых стохастических переменных $(x^{(1)}, x^{(2)}, \dots, x^{(p)}, y)$ подчинен $(p+1)$ -мерному нормальному закону (см. Приложение 3.1).

Покажем, что к модели, специфицированной соотношениями (7.1)–(7.2)–(7.3')–(7.4'), применимы все методы и результаты, полученные нами в рамках КЛММР (при $\Sigma_0 = I_n$) и ОЛММР (при произвольной симметричной и положительно определенной матрице Σ_0). Правда, интерпретация этих результатов при стохастических регрессорах X базируется на понятии *условного* распределения регрессионных остатков $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)^\top$, в котором в качестве условия используется *заданность* (фиксированность) матрицы X наблюдаемых значений объясняющих переменных.

Проведем доказательство возможности распространения формул МНК и свойств МНК-оценок на случай некоррелированных с регрессионными остатками *стохастических* объясняющих переменных в рамках условий *классической* ЛММР. Переход к *обобщенной* ЛММР аналогичен тому, как это было сделано при неслучайных предикторах.

В рамках допущений классической ЛММР матрица Σ_0 в (7.4') полагается *единичной*, то есть

$$\Sigma_\epsilon(X) = \sigma^2 I_n. \quad (7.4a)$$

Мы будем предполагать также, что существуют следующие пределы (*по вероятности*):

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n} X^\top X \right) = \Sigma, \quad (7.5)$$

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n} X^\top \epsilon \right) = O_{p+1}, \quad (7.6)$$

где $(p+1) \times (p+1)$ -матрица Σ положительно определена (а значит существует матрица Σ^{-1}).

Рассмотрим МНК-оценку $\hat{\Theta} = (X^\top X)^{-1} X^\top Y$ неизвестных параметров Θ модели (7.1)–(7.2)–(7.3')–(7.4^a) и проанализируем ее статистические свойства. Ее существование при любых выборочных данных X, Y обеспечивается условием (7.2).

Состоятельность МНК-оценки $\hat{\Theta}$ следует из ее представления в виде

$$\begin{aligned} \hat{\Theta} &= (X^\top X)^{-1} X^\top (X\Theta + \epsilon) = \\ &= (X^\top X)^{-1} (X^\top X) \Theta + (X^\top X)^{-1} X^\top \epsilon = \\ &= \Theta + \left(\frac{1}{n} X^\top X \right)^{-1} \left(\frac{1}{n} X^\top \epsilon \right). \end{aligned} \quad (7.7)$$

В соответствии с условиями (7.5)–(7.6) второе слагаемое правой части выражения (7.7) стремится по вероятности к нулю при $n \rightarrow \infty$, а следовательно, $\hat{\Theta} \rightarrow \Theta$ (по вероятности) при $n \rightarrow \infty$.

Анализ остальных статистических свойств оценок $\hat{\Theta}$ и $\hat{\sigma}^2$ в рамках моментов соответствующих условных распределений (то есть при зафиксированных значениях элементов матрицы $\mathbf{X}$), проведенный точно так же, как это было сделано в п. 4.2, дает:

$$\mathbf{E}(\hat{\Theta} | \mathbf{X}) = \Theta \quad (\text{см. (4.25)-(4.26)});$$

$$\Sigma_{\hat{\Theta}}(\mathbf{X}) = \mathbf{E}\{[(\hat{\Theta} - \Theta)(\hat{\Theta} - \Theta)^T] | \mathbf{X}\} = \sigma^2(\mathbf{X}^T \mathbf{X})^{-1} \quad (\text{см. (4.33)});$$

$$\hat{\varepsilon}(\mathbf{X}) = Y - \mathbf{X}\hat{\Theta} = \mathbf{Z}\varepsilon \quad (\text{см. (4.27)});$$

$$\mathbf{E}(\hat{\varepsilon} | \mathbf{X}) = \mathbf{O}_n;$$

$$\Sigma_{\hat{\varepsilon}}(\mathbf{X}) = \mathbf{E}(\hat{\varepsilon}\hat{\varepsilon}^T | \mathbf{X}) = \mathbf{E}(\mathbf{Z}\varepsilon\varepsilon^T\mathbf{Z}^T | \mathbf{X}) = \sigma^2\mathbf{Z};$$

$$\mathbf{E}(\hat{\varepsilon}^T \hat{\varepsilon} | \mathbf{X}) = \sigma^2 \text{tr } \mathbf{Z} = \sigma^2(n - p - 1) \quad (\text{см. (4.29)});$$

$$\mathbf{E}(\hat{\sigma}^2 | \mathbf{X}) = \sigma^2 \quad (\text{см. (4.30)});$$

$$\mathbf{E}[\hat{\sigma}^2(\mathbf{X}^T \mathbf{X})^{-1} | \mathbf{X}] = \sigma^2(\mathbf{X}^T \mathbf{X})^{-1}.$$

Вывод этих соотношений опирается на тот факт, что сомножитель, зависящий от «условия» (то есть от матрицы $\mathbf{X}$), можно выносить из-под знака *условного* математического ожидания.

С помощью известного метода последовательного (повторного) вычисления математического ожидания вида²

$$\mathbf{E}_{(\xi, \eta)}\varphi(\xi, \eta) = \mathbf{E}_X\{\mathbf{E}_\eta[\varphi(\xi, \eta) | \xi = X]\}$$

легко доказывается и безусловная несмещенность оценок $\hat{\Theta}$ и $\hat{\sigma}^2$:

$$\mathbf{E}\hat{\Theta} = \mathbf{E}_X[\mathbf{E}_Y(\hat{\Theta} | \mathbf{X})] = \mathbf{E}_X(\Theta) = \Theta,$$

$$\mathbf{E}\hat{\sigma}^2 = \mathbf{E}_X[\mathbf{E}_Y(\hat{\sigma}^2 | \mathbf{X})] = \mathbf{E}_X(\sigma^2) = \sigma^2.$$

²Нижний индекс у знака математического ожидания $\mathbf{E}$ определяет, по значениям какой именно случайной величины производится усреднение. Поясним смысл (выведем) упомянутое правило повторного вычисления математического ожидания. Пусть $\varphi(\xi, \eta)$ — некоторая функция случайных (вообще говоря, векторных) переменных (ξ, η) , совместное распределение которых задается функцией плотности $p_{(\xi, \eta)}(X, Y)$. Предположим, что существует $\mathbf{E}_{(\xi, \eta)}\varphi(\xi, \eta)$. Тогда, учитывая, что функция $p_{(\xi, \eta)}(X, Y)$ может быть представлена в виде $p_{(\xi, \eta)}(X, Y) = p_\eta(Y | \xi = X)p_\xi(X)$, имеем:

$$\begin{aligned} \mathbf{E}_{(\xi, \eta)}(\varphi(\xi, \eta)) &= \int_X \int_Y \varphi(X, Y) p_\eta(Y | \xi = X) p_\xi(X) dX dY = \\ &= \int_X \left[\int_Y \varphi(X, Y) p_\eta(Y | X) dY \right] p_\xi(X) dX = \\ &= \int_X \mathbf{E}_\eta(\varphi(X; \eta) | X) p_\xi(X) dX = \mathbf{E}_X\{\mathbf{E}_\eta[\varphi(X; \eta) | X]\}, \end{aligned}$$

что и требовалось доказать.

Наконец, точно так же, как при неслучайных предикторах, доказывается *условная оптимальность* МНК-оценок $\hat{\Theta}$, а именно тот факт, что *среди всех линейных (относительно Y) условно несмещенных оценок вектора Θ его МНК-оценка обладает «наименьшей ковариационной матрицей» в смысле соотношений (4.36)–(4.37)*. При этом участвующие в этих соотношениях моменты должны быть заменены на соответствующие *условные* моменты (при условии, что матрица наблюдаемых значений объясняющих переменных равна X).

7.2 Общий случай: стохастические предикторы X коррелированы с регрессионными остатками ϵ . Метод инструментальных переменных

Из (7.7) видно, что если хотя бы одна из объясняющих переменных X асимптотически (по $n \rightarrow \infty$) коррелирует с регрессионными остатками ϵ , то $X^T \epsilon / n$ не стремится (по вероятности) к нулевому вектору (то есть не выполняется условие (7.6)) и, следовательно, МНК-оценки $\hat{\Theta}$ уже не будут состоятельными. При этом *всего один* ненулевой элемент вектора $\lim_{n \rightarrow \infty} (X^T \epsilon / n)$ может приводить к несостоятельности *всех* компонент МНК-оценки $\hat{\Theta}$. Действительно, возвращаясь к (7.7), имеем³:

$$\lim_{n \rightarrow \infty} \hat{\Theta} = \Theta + \lim_{n \rightarrow \infty} \left(\frac{1}{n} X^T X \right)^{-1} \lim_{n \rightarrow \infty} \left(\frac{1}{n} X^T \epsilon \right).$$

Предположим, что в вектор-столбце $\lim_{n \rightarrow \infty} (X^T \epsilon / n)$ на l -м месте стоит постоянная $r_l = \text{cov}(x^{(l)}, \epsilon) \neq 0$, а на всех остальных местах — нули, и пусть $s^l = (s^{1l}, s^{2l}, \dots, s^{p+1l})^T$ — l -й столбец матрицы Σ^{-1} , где матрица Σ определена соотношением (7.5). Тогда

$$\lim_{n \rightarrow \infty} \hat{\Theta} = \Theta + r_l s^l$$

и, следовательно, все компоненты вектора $\hat{\Theta}$ окажутся *смещенными и несостоятельными* оценками соответствующих компонент вектора Θ (за исключением того редкого частного случая, когда все исходные объясняющие переменные ортогональны, так как в этом случае все s^{jl} , кроме одного — s^{ll} , равны нулю).

Мы видим, что корреляция между объясняющими переменными и регрессионными остатками является серьезным препятствием к использованию обыкновенного метода наименьших квадратов. Поэтому для

³Все участвующие в данном пункте пределы (по $n \rightarrow \infty$) понимаются в смысле *сходимости по вероятности*, поскольку их обоснование опирается, в конечном счете, на закон больших чисел.

анализа моделей такого типа используется альтернативный метод, известный в литературе как **метод инструментальных переменных**.

В качестве основного *инструмента* метода используются вспомогательные (*сопутствующие*) переменные $U = (1, u^{(1)}, \dots, u^{(p)})^T$, которые требуется подобрать таким образом, чтобы:

(i) имелась бы принципиальная возможность измерить их значения на тех же объектах (или в те же «моменты времени»), на которых (или в которые) мы располагаем значениями исходных объясняющих переменных $X = (1, x^{(1)}, \dots, x^{(p)})$; следовательно, наряду с $n \times (p+1)$ -матрицей значений X мы будем располагать и $n \times (p+1)$ -матрицей значений U ;

(ii) переменные U должны быть асимптотически (по $n \rightarrow \infty$) некоррелированы с регрессионными остатками ε , то есть

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n} U^T \varepsilon \right) = O_{p+1}; \quad (7.8)$$

(iii) перекрестные (смешанные) вторые моменты переменных U и X в пределе (по $n \rightarrow \infty$) не все равны нулю и образуют, в асимптотике по $n \rightarrow \infty$, невырожденную (положительно определенную) $(p+1) \times (p+1)$ -матрицу Σ_{UX} , то есть

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n} U^T X \right) = \Sigma_{UX}; \quad (7.9)$$

(iv) должна существовать (в асимптотике по $n \rightarrow \infty$) положительно определенная матрица Σ_{UU} такая, что

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n} U^T U \right) = \Sigma_{UU}. \quad (7.10)$$

Отметим, что если некоторые из исходных объясняющих переменных X можно считать некоррелированными с регрессионными остатками ε , то их можно использовать для формирования столбцов матрицы U и находить дополнительно вспомогательные переменные только для оставшихся столбцов.

Найденные в соответствии с условиями (i)–(iv) переменные U называют **инструментальными**, а оценки неизвестных параметров Θ в модели (7.1) определяются по формуле

$$\hat{\Theta}_U = (U^T X)^{-1} U^T Y. \quad (7.11)$$

В состоятельности оценки $\hat{\Theta}_U$ можно убедиться, подставив Y , выраженное соотношением (7.1), в (7.11), и воспользовавшись свойствами (i)–(iv) инструментальных переменных U :

$$\begin{aligned} \hat{\Theta}_U &= (U^T X)^{-1} U^T (X\Theta + \varepsilon) = (U^T X)^{-1} (U^T X)\Theta + (U^T X)^{-1} U^T \varepsilon = \\ &= \Theta + (U^T X)^{-1} U^T \varepsilon. \end{aligned}$$

Следовательно:

1)

$$\begin{aligned}\lim_{n \rightarrow \infty} \hat{\Theta}_U &= \Theta + \lim_{n \rightarrow \infty} \left(\frac{1}{n} \mathbf{U}^\top \mathbf{X} \right)^{-1} \lim_{n \rightarrow \infty} \left(\frac{1}{n} \mathbf{U}^\top \boldsymbol{\varepsilon} \right) = \\ &= \Theta + \Sigma_{UX}^{-1} \mathbf{O}_{p+1} = \Theta;\end{aligned}$$

2)

$$\begin{aligned}\mathbf{E}(\hat{\Theta}_U | \mathbf{X}, \mathbf{U}) &= \Theta \quad \text{и} \\ \mathbf{E}\hat{\Theta}_U &= \mathbf{E}_{\mathbf{X}, \mathbf{U}}\{\mathbf{E}(\hat{\Theta}_U | \mathbf{X}, \mathbf{U})\} = \mathbf{E}_{\mathbf{X}, \mathbf{U}}\Theta = \Theta,\end{aligned}$$

то есть оценка $\hat{\Theta}_U$ является состоятельной, а также условно и безусловно несмещенной.

Говоря о других статистических свойствах оценки $\hat{\Theta}_U$, следует отметить, что она, вообще говоря, уже *не является оптимальной*. Оценка для ее условной ковариационной матрицы $\Sigma_{\hat{\Theta}_U}(\mathbf{X}, \mathbf{U}) = \mathbf{E}\{(\hat{\Theta}_U - \Theta)(\hat{\Theta}_U - \Theta)^\top | \mathbf{X}, \mathbf{U}\}$ может быть вычислена по формуле

$$\hat{\Sigma}_{\hat{\Theta}_U} = \hat{\sigma}^2 (\mathbf{U}^\top \mathbf{X})^{-1} (\mathbf{U}^\top \mathbf{U}) (\mathbf{X}^\top \mathbf{U})^{-1}, \quad (7.12)$$

где

$$\hat{\sigma}^2 = \frac{1}{n - p - 1} (Y - \mathbf{X}\hat{\Theta}_U)^\top (Y - \mathbf{X}\hat{\Theta}_U). \quad (7.13)$$

Анализ правой части (7.12) приводит к выводу, что *чем выше корреляция между U и X , тем точнее будут оценки метода инструментальных переменных $\hat{\Theta}_U$* . И наоборот: при слабой коррелированности инструментальных переменных U с исходными объясняющими переменными X среднеквадратические ошибки оценок оказываются чрезмерно большими, устремляясь в бесконечность при приближении всех ковариаций между $x^{(j)}$ и $u^{(k)}$ к нулевым значениям. Так, например, в случае единственной объясняющей переменной ($p = 1$) «инструментальная» оценка коэффициента θ_1 наклона линии регрессии будет равна (в соответствии с (7.11)):

$$\hat{\theta}_{1U} = \frac{\sum_{i=1}^n (u_i - \bar{u})(y_i - \bar{y})}{\sum_{i=1}^n (u_i - \bar{u})(x_i - \bar{x})}, \quad (7.11')$$

а ее выборочная дисперсия (в соответствии с (7.12)–(7.13))

$$s_{\hat{\theta}_{1U}}^2 = \hat{\sigma}^2 \frac{\sum_{i=1}^n (u_i - \bar{u})^2}{\left[\sum_{i=1}^n (u_i - \bar{u})(x_i - \bar{x}) \right]^2}. \quad (7.12')$$

Так что, если u и x слабо взаимно коррелированы, то знаменатель в выражении для $s_{\hat{\theta}_{1U}}^2$ будет близок к нулю и, следовательно, выборочная дисперсия оценки $\hat{\theta}_{1U}$ окажется чрезмерно большой.

Отсюда следует, что *при подборе инструментальных переменных надо стараться определить такие переменные U , которые практически не коррелируют с регрессионными остатками ϵ и одновременно достаточно сильно коррелированы с исходными объясняющими переменными X* . В этом и состоит главная трудность реализации метода инструментальных переменных (ведь истинное вероятностное распределение X , U и ϵ ненаблюдаемо и поэтому трудно быть уверенными в том, что подобранные инструментальные переменные действительно практически некоррелированы с ϵ и одновременно сильно коррелированы с X !).

Проиллюстрируем технику метода инструментальных переменных на простом примере.

Пример 7.1. С целью оценки эластичностей потребления определенного вида товаров (y) по доходу (x) на основании результатов выборочных обследований домашних хозяйств анализируется линейная регрессия y по x

$$y = \theta_0 + \Theta_1 x + \epsilon. \quad (7.14)$$

В ходе выборочных обследований семей регистрировались значения следующих показателей:

y_i — *удельные (то есть приходящиеся на одного члена семьи в «единицу времени») расходы на рассматриваемый вид товаров или услуг в i -й обследованной семье;*

x_i^* — *общие удельные расходы в той же семье (включая статью «денежные сбережения»);*

x_i^0 — **объявленный среднедушевой доход** в i -й семье ($i = 1, 2, \dots, n$).

Специальный статистический и профессиональный (экономический) анализ позволил принять в качестве рабочих гипотез следующие допущения:

(а) ненаблюдаемый истинный среднедушевой доход i -й семьи x_i связан с общими удельными расходами x_i^* соотношением

$$x_i^* = x_i + \delta_i,$$

где случайная ошибка δ_i имеет нулевое среднее значение ($E\delta_i = 0$), некоррелирована с x_i (то есть $\text{cov}(x, \delta) = E[(x - Ex)\delta] = 0$) и с ϵ_i (то есть $\text{cov}(\delta, \epsilon) = E(\delta\epsilon) = 0$);

(б) **объявленный** среднедушевой доход i -й семьи x_i^0 дает заниженную оценку для x_i , некоррелирован ни с регрессионными остатками ϵ_i , ни со

случайными ошибками δ_i (то есть $\text{cov}(x^0, \varepsilon) = \text{cov}(x^0, \delta) = 0$), но весьма сильно коррелирован с истинным доходом x_i и с общими расходами x_i^* ;

(в) общие удельные расходы x_i^* практически некоррелированы с регрессионными остатками ε_i , то есть $\text{cov}(x^*, \varepsilon) = \mathbf{E}((x^* - \mathbf{E}x^*)\varepsilon) = 0$.

Проведем статистический анализ данной модели. Поскольку в соответствии с (а) общие расходы x^* дают в среднем правильное представление о величине истинного дохода x , то вставим в (7.14) значения x_i , выраженные через *наблюдаемые* величины x_i^* , и будем анализировать модель (7.14) на основании данных (x_i^*, y_i) , $i = 1, 2, \dots, n$:

$$y_i = \theta_0 + \theta_1(x_i^* - \delta_i) + \varepsilon_i = \theta_0 + \theta_1 x_i^* + \varepsilon_i^*, \quad (7.14')$$

где $\varepsilon_i^* = \varepsilon_i - \theta_1 \delta_i$. Перепишем уравнение (7.14') в терминах *центрированных* переменных:

$$y_i - \bar{y} = \theta_1(x_i^* - \bar{x}^*) + \varepsilon_i^*. \quad (7.14'')$$

Далее можно было бы попытаться воспользоваться МНК-оценками параметров θ_0 и θ_1 по данным (x_i^*, y_i) , в частности:

$$\hat{\theta}_{1. \text{мнк}} = \frac{\sum_{i=1}^n (x_i^* - \bar{x}^*)(y_i - \bar{y})}{\sum_{i=1}^n (x_i^* - \bar{x}^*)^2}. \quad (7.15)$$

Однако легко убедиться в том, что объясняющая переменная x^* коррелирована с регрессионными остатками ε^* в модели (7.14') и, следовательно, в соответствии с (7.7) МНК-оценки будут несостоятельными и обладать асимптотически неустрашимым смещением.

Действительно:

$$\begin{aligned} \text{cov}(x^*, \varepsilon^*) &= \mathbf{E}((x^* - \mathbf{E}x^*)\varepsilon^*) = \mathbf{E}((x^* - \mathbf{E}x^*)(\varepsilon - \theta_1 \delta)) = \\ &= \text{cov}(x^*, \varepsilon) - \theta_1 \text{cov}(x^*, \delta) = -\theta_1 \text{cov}(x + \delta, \delta) = -\theta_1 \text{cov}(\delta, \delta) = \\ &= -\theta_1 \cdot \mathbf{D}\delta \neq 0 \quad \text{при} \quad \theta_1 \neq 0. \end{aligned}$$

То есть x^* коррелирована с ε^* при $\theta_1 \neq 0$.

Вычислим смещение оценки (7.15):

$$\begin{aligned} \hat{\theta}_{1. \text{мнк}} &= \frac{\sum_{i=1}^n (x_i^* - \bar{x}^*)[\theta_1(x_i^* - \bar{x}^*) + (\varepsilon_i - \theta_1 \delta_i)]}{\sum_{i=1}^n (x_i^* - \bar{x}^*)^2} = \\ &= \theta_1 + \frac{1}{\frac{1}{n} \sum_{i=1}^n (x_i^* - \bar{x}^*)^2} \left[\widehat{\text{cov}}(x^*, \varepsilon) - \theta_1 \cdot \frac{1}{n} \sum_{i=1}^n (x_i + \delta_i - \bar{x} - \bar{\delta})\delta_i \right] = \\ &= \theta_1 + \frac{1}{s_{x^*}^2} [\widehat{\text{cov}}(x^*, \varepsilon) - \theta_1 \widehat{\text{cov}}(x, \delta) - \theta_1 s_\delta^2]. \end{aligned}$$

В выполненных тождественных преобразованиях мы использовали следующие обозначения:

$$s_{x^*}^2 = \frac{1}{n} \sum_{i=1}^n (x_i^* - \bar{x}^*)^2 - \text{выборочная дисперсия переменной } x^*;$$

$$\widehat{\text{cov}}(\xi, \eta) = \frac{1}{n} \sum_{i=1}^n (\xi_i - \bar{\xi})(\eta_i - \bar{\eta}) - \text{выборочная ковариация переменных } \xi \text{ и } \eta;$$

$$s_{\delta}^2 = \widehat{\text{cov}}(\delta, \delta) = \frac{1}{n} \sum_{i=1}^n (\delta_i - \bar{\delta})^2 - \text{выборочная дисперсия ошибки } \delta.$$

По закону больших чисел выборочные ковариации и дисперсии стремятся (по вероятности) при $n \rightarrow \infty$ к соответствующим теоретическим значениям, так что с учетом рабочих гипотез (а)–(б)–(в) имеем:

$$\lim_{n \rightarrow \infty} \hat{\theta}_{1.\text{мнк}} = \theta_1 \left(1 - \frac{D\delta}{Dx^*} \right) = \theta_1 \left(1 - \frac{D\delta}{Dx + D\delta} \right) = \theta_1 \frac{1}{1 + \frac{D\delta}{Dx}}. \quad (7.16)$$

А поскольку ни дисперсии $D\delta$ и Dx , ни их отношение нам не известны (и, как правило, их не удастся хорошо оценить по исходным статистическим данным), то придется обратиться к *методу инструментальных переменных* для оценивания параметра θ_1 . Очевидно, в нашем примере на роль инструментальной переменной вполне подходит переменная x^0 — величина *объявленного* среднедушевого дохода семьи: во-первых, она некоррелирована со случайными остатками $\varepsilon^* = \varepsilon - \theta_1 \delta$ (то есть x^0 не коррелирована ни с ε , ни с δ , см. рабочую гипотезу (б)), а во-вторых, достаточно сильно коррелирована с истинным среднедушевым доходом x и общими расходами x^* .

Поэтому в качестве состоятельной (и весьма эффективной) оценки параметра θ_1 предлагается использовать оценку метода инструментальных переменных

$$\hat{\theta}_{1.u} = \frac{\sum_{i=1}^n (x_i^0 - \bar{x}^0)(y_i - \bar{y})}{\sum_{i=1}^n (x_i^* - \bar{x}^*)(x_i^0 - \bar{x}^0)}.$$

7.3 Случайные ошибки в измерении значений объясняющих переменных

Именно ситуации, в которых объясняющие переменные могут измеряться только с некоторыми случайными ошибками, чаще других приводят к необходимости рассматривать модели регрессии со *стохастическими* объясняющими переменными и обращаться соответственно к *методу инструментальных переменных* (заметим, что и в рассмотренном выше примере 7.1 природа стохастичности объясняющей переменной и

ее коррелированности с регрессионными остатками как раз и была обусловлена тем, что о значениях объясняющей переменной — истинного дохода x мы могли судить лишь по соответствующим значениям общих семейных расходов x^* , представляющим собой измеренные с некоторой случайной ошибкой величины (x).

Прежде всего отметим, что случайные ошибки в измерении *результатирующего показателя* y (при условии их нулевых средних значений и некоррелированности с объясняющими переменными X и регрессионными остатками ϵ) приводят лишь к увеличению остаточной дисперсии (или дисперсии остатка), но не влияют ни на несмещенность, ни на состоятельность МНК-оценок. Действительно, пусть Δ_Y — вектор-столбец (высоты n) случайных ошибок измерения анализируемого результирующего показателя Y , так что в качестве его наблюдаемых значений мы располагаем вектор-столбцом

$$Y^* = Y + \Delta_Y. \quad (7.17)$$

Подставляя выраженные из (7.1) значения Y в модель (7.17), получаем:

$$Y^* = X\Theta + (\epsilon + \Delta_Y). \quad (7.1a)$$

Анализ (по стандартной схеме, см. п. 4.2) полученных по наблюдениям (X, Y^*) МНК-оценок

$$\hat{\Theta}_{\text{МНК}} = (X^T X)^{-1} X^T Y^* = (X^T X)^{-1} X^T (X\Theta + \epsilon + \Delta_Y), \quad (7.18)$$

проведенный с учетом упомянутых выше свойств вектора случайных ошибок Δ_Y ($E\Delta_Y = O_n$, $E(\Delta_Y \epsilon^T) = O_{n,n}$, $E(\Delta_Y X^T) = O_{n,p}$), подтверждает, что МНК-оценки (7.18) остаются несмещенными, состоятельными и оптимальными (в смысле (4.36)–(4.37)).

Теперь проанализируем, как влияют случайные ошибки в измерении значений объясняющих переменных на МНК-оценки параметров классической ЛММР (7.1). При этом мы будем предполагать:

$$X^* = X + \Delta_X; \quad (7.19)$$

$$E\Delta_X = O_{p+1}; \quad (7.20)$$

$$E(\Delta_X \epsilon^T) = O_{p+1,n}; \quad (7.21)$$

$$E(\Delta_X X^T) = O_{p+1,p+1}; \quad (7.22)$$

$$E(X^* \epsilon^T) = O_{p+1,n}; \quad (7.23)$$

$$\Sigma_{\Delta_X} = E(\Delta_X \Delta_X^T). \quad (7.24)$$

В соотношениях (7.19)–(7.24) $\Delta_X = (\Delta_X^{(0)}, \Delta_X^{(1)}, \dots, \Delta_X^{(p)})^T$ — вектор случайных ошибок в измерении объясняющих переменных $X =$

$= (1, x^{(1)}, \dots, x^{(p)})^\top$; $\Delta_X - n \times (p+1)$ -матрица значений ошибок $\Delta_i^{(j)}$ в измерениях $x_i^{(j)}$; соответственно $X^* - n \times (p+1)$ -матрица искаженных наблюдаемых значений объясняющих переменных, а $X^* = (1, x^{*(1)}, \dots, x^{*(p)})^\top$ — вектор наблюдаемых (искаженных) значений объясняющих переменных; соотношения (7.21)–(7.22) свидетельствуют о некоррелированности ошибок измерения с регрессионными остатками и объясняющими переменными (откуда следует некоррелированность искаженных объясняющих переменных с регрессионными остатками, см. (7.23)); и, наконец, соотношением (7.24) определена ковариационная матрица вектора ошибок измерения Δ_X .

Итак, мы рассматриваем классическую ЛММР со стохастическими объясняющими переменными (7.1)–(7.3)–(7.4а)–(7.5)–(7.6) в условиях случайных ошибок в измерении предикторов (7.19)–(7.24). Подставим в модель (7.1) значения X , выраженные из (7.19) в терминах *наблюдаемых* (но искаженных) значений объясняющих переменных:

$$Y = X^* \Theta + (\varepsilon - \Delta_X \Theta). \quad (7.16)$$

МНК-оценка для Θ из (7.16) есть

$$\begin{aligned} \hat{\Theta}_{\text{МНК}} &= (X^{*\top} X^*)^{-1} X^{*\top} Y = (X^{*\top} X^*)^{-1} X^{*\top} [X^* \Theta + \varepsilon - \Delta_X \Theta] = \\ &= \Theta + \left(\frac{1}{n} X^{*\top} X^* \right)^{-1} \left[\frac{1}{n} X^{*\top} \varepsilon - \left(\frac{1}{n} X^{*\top} \Delta_X \right) \Theta \right]. \end{aligned} \quad (7.25)$$

Проанализируем поведение правой части (7.25) при $n \rightarrow \infty$.

1)

$$\begin{aligned} \lim_{n \rightarrow \infty} \left(\frac{1}{n} X^{*\top} X^* \right) &= \lim_{n \rightarrow \infty} \left(\frac{1}{n} (X + \Delta_X)^\top (X + \Delta_X) \right) = \\ &= \lim_{n \rightarrow \infty} \left(\frac{1}{n} X^\top X \right) + \lim_{n \rightarrow \infty} \left(\frac{1}{n} \Delta_X^\top X \right) + \lim_{n \rightarrow \infty} \left(\frac{1}{n} X^\top \Delta_X \right) + \\ &+ \lim_{n \rightarrow \infty} \left(\frac{1}{n} \Delta_X^\top \Delta_X \right) = \Sigma + \Sigma_{\Delta_X} \end{aligned} \quad (7.26)$$

(полученный результат опирается на закон больших чисел (З.Б.Ч) и соотношения (7.5), (7.22) и (7.24)).

2)

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n} X^{*\top} \varepsilon \right) = O_{p+1} \quad (7.27)$$

(полученный результат опирается на З.Б.Ч. и соотношение (7.23)).

3)

$$\begin{aligned} \lim_{n \rightarrow \infty} \left(\frac{1}{n} \mathbf{X}^{*\top} \Delta_X \right) &= \lim_{n \rightarrow \infty} \left(\frac{1}{n} (\mathbf{X} + \Delta_X)^\top \Delta_X \right) = \\ &= \lim_{n \rightarrow \infty} \left(\frac{1}{n} \mathbf{X}^\top \Delta_X \right) + \lim_{n \rightarrow \infty} \left(\frac{1}{n} \Delta_X^\top \Delta_X \right) = \Sigma_{\Delta_X} \end{aligned} \quad (7.28)$$

(полученный результат опирается на З.Б.Ч. и соотношения (7.22) и (7.24)).

Переходя к пределу (по вероятности) при $n \rightarrow \infty$ в (7.25) и учитывая (7.26)–(7.28), имеем:

$$\begin{aligned} \lim_{n \rightarrow \infty} \hat{\Theta}_{\text{МНК}} &= \Theta - (\Sigma + \Sigma_{\Delta_X})^{-1} \Sigma_{\Delta_X} \Theta = \\ &= [\mathbf{I}_{p+1} - (\Sigma + \Sigma_{\Delta_X})^{-1} \Sigma_{\Delta_X}] \Theta. \end{aligned} \quad (7.29)$$

Как видим из (7.29), оценка $\hat{\Theta}_{\text{МНК}}$ в данном случае не является ни состоятельной, ни несмещенной (ее асимптотически-неустранимое смещение оказывается равным, в соответствии с (7.29), величине $-(\Sigma + \Sigma_{\Delta_X})^{-1} \Sigma_{\Delta_X} \Theta$). Степень смещения оценки определяется истинным значением параметра, структурой матрицы наблюдаемых значений объясняющих переменных и ковариационной матрицей ошибок измерения. Заметим, что в простейшем случае парной регрессии (то есть при $p = 1$) формула (7.29), естественно, повторяет ранее выведенную нами (при анализе примера 7.1) формулу (7.16).

Как же поступают при анализе подобных моделей? Это зависит от априорной информации, которой располагает исследователь. Если из предыстории или с помощью специальных экспертных опросов мы можем оценить матрицы Σ и Σ_{Δ_X} , то на основании (7.29) можно построить «подправленную МНК-оценку»

$$\hat{\Theta}_{\text{МНК}}^* = [\mathbf{I}_{p+1} - (\Sigma + \Sigma_{\Delta_X})^{-1} \Sigma_{\Delta_X}]^{-1} \hat{\Theta}_{\text{МНК}}. \quad (7.30)$$

В остальных случаях, как правило, обращаются к методу инструментальных переменных.

Следующий пример иллюстрирует возможность привлечения в качестве инструментальных переменных не реальных показателей, а построенных некоторым специальным способом искусственных признаков.

Вернемся к примеру 7.1 (см. модель (7.14) – (7.14') – (7.14'')). Предположим, что сведения о значениях объявленного дохода x_i^0 утеряны и, следовательно, мы не можем использовать этот признак в качестве инструментальной переменной. Пусть для определенности число наблюдений n четно. Итак, для модели (7.14''), записанной в терминах центрированных переменных

$$\mathbf{X}_n^{*\top} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1^* - \bar{x}^* & x_2^* - \bar{x}^* & \dots & x_n^* - \bar{x}^* \end{pmatrix}$$

и

$$Y_{\text{ц}} = (y_1 - \bar{y}, y_2 - \bar{y}, \dots, y_n - \bar{y})^T,$$

определим матрицу значений инструментальных переменных U следующим образом:

$$U^T = \begin{pmatrix} 1 & 1 & \dots & 1 \\ -1 & +1 & \dots & -1 \end{pmatrix},$$

где i -й элемент второй строки матрицы U^T равен единице с плюсом или с минусом в зависимости от того, будет ли соответствующее значение $x_i^* - \bar{x}^*$ в матрице $X_{\text{ц}}^{*T}$ выше или ниже медианы выборки $(x_1^* - \bar{x}^*, x_2^* - \bar{x}^*, \dots, x_n^* - \bar{x}^*)$. В соответствии с (7.11) оценка по методу инструментальной переменной окажется равной в данном случае

$$\hat{\Theta}_U = \begin{pmatrix} n & 0 \\ 0 & \frac{n}{2}(\bar{x}_{\text{ц}2}^* - \bar{x}_{\text{ц}1}^*) \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \frac{n}{2}(\bar{y}_{\text{ц}2} - \bar{y}_{\text{ц}1}) \end{pmatrix}, \quad (7.31)$$

где $\bar{x}_{\text{ц}2}^*$ и $\bar{x}_{\text{ц}1}^*$ обозначают средние из отклонений значений x_i^* соответственно вверх и вниз от медианы, а $\bar{y}_{\text{ц}2}$ и $\bar{y}_{\text{ц}1}$ — средние из соответствующих этому разбиению значений $y_i - \bar{y}$. Из (7.31) получаем:

$$\hat{\theta}_{0U} = 0; \quad \hat{\theta}_{1U} = \frac{\bar{y}_{\text{ц}2} - \bar{y}_{\text{ц}1}}{\bar{x}_{\text{ц}2}^* - \bar{x}_{\text{ц}1}^*}. \quad (7.32)$$

Соотношения (7.31)–(7.32) определяют способ оценивания, впервые предложенный Вальдом.⁴ Когда значение n нечетно, прежде чем приступить к вычислениям, нужно исключить среднее значение. При весьма общих предположениях оценка (7.32) состоятельна, однако ее выборочная дисперсия может оказаться большой. Бартлетт показал, что эффективность оценки можно повысить, разбив упорядоченные значения объясняющей переменной на три группы одинакового объема, из которых первая содержит наименьшие значения x , а третья — наибольшие.⁵ Исключив центральные $n/3$ наблюдений, получим оценку для коэффициента наклона линии регрессии

$$\hat{\theta}'_{1U} = \frac{\bar{y}_3 - \bar{y}_1}{\bar{x}_3 - \bar{x}_1},$$

где $\bar{x}_i, \bar{y}_i$ ($i = 1, 3$) обозначают средние для наблюдений, попавших в две крайние группы.

Распространение метода Вальда и Бартлетта на случай двух и более объясняющих переменных весьма трудоемко (см., например, статью Hooper J. W., Theil H. The Extension of Wald's Method of Fitting Straight

⁴ Wald A. The Fitting of Straight Lines if Both Variables are Subject to Error. Ann. Math. Stat., 1940, vol. 11, pp. 284–300.

⁵ Bartlett M. S. Fitting a Straight Line when Both Variables are Subject to Error. Biometrics, 1949, vol. 5, pp. 207–212.

Lines to Multiple Regression. Rev. Intern. Statist. Inst., 1958, vol. 26, pp.37–47).

Влияние ошибок измерения объясняющих переменных на точность прогноза (в рамках КЛММР). Вернемся к задаче построения точечного прогноза для неизвестного значения результирующего показателя $y(X_{n+1})$ по заданному значению объясняющей переменной $X_{n+1} = (1, x_{n+1}^{(1)}, \dots, x_{n+1}^{(p)})^T$ (см. п. 6.2). Пусть значение X_{n+1} задается со случайной ошибкой (что чаще всего и бывает в действительности, так как в качестве X_{n+1} берутся обычно *прогнозные* или *планируемые* значения предикторов), то есть мы располагаем значениями

$$X_{n+1}^* = X_{n+1} + \Delta_X,$$

где Δ_X — вектор-столбец ошибок измерений, не зависящий от $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n, \varepsilon_{n+1}$, причем,

$$E\Delta_X = O_{p+1} \quad \text{и} \quad \Sigma_{\Delta_X} = E(\Delta_X \Delta_X^T) = \sigma_{\Delta}^2 I_{p+1}.$$

Тогда можно показать, что если мы проводим регрессионный анализ в рамках КЛММР со стохастическими переменными (то есть в рамках модели (7.1)–(7.3')–(7.4a)), то наилучший линейный несмещенный прогноз $\hat{y}(X_{n+1})$ неизвестного значения $y(X_{n+1})$ задается формулой

$$\hat{y}(X_{n+1}) = X^{*T} \hat{\Theta}_{\text{мнк}}, \quad (7.33)$$

а его средний квадрат ошибки определяется соотношением:

$$E(\hat{y}(X_{n+1}) - y(X_{n+1}))^2 = \sigma^2 [1 + X_{n+1}^T (X^T X)^{-1} X_{n+1} + \sigma_{\Delta}^2 \text{tr}(X^T X)^{-1}] + \sigma_{\Delta}^2 \Theta^T \Theta. \quad (7.34)$$

В соотношении (7.33) оценки $\hat{\Theta}_{\text{мнк}}$ строятся по *наблюдаемой* матрице X^* (то есть $\hat{\Theta}_{\text{мнк}} = (X^{*T} X^*)^{-1} X^{*T} Y$). При статистической оценке величины среднего квадрата ошибки прогноза значения $\sigma^2, \sigma_{\Delta}^2$ и Θ заменяются их статистическими оценками, а *ненаблюдаемые* вектор X_{n+1} и матрица X — соответственно наблюдаемыми X_{n+1}^* и X^* .

Сравнение правой части (7.34) с правой частью ранее выведенной формулы (6.22) показывает, что ошибки в измерении значений объясняющих переменных, как и следовало ожидать, увеличивают погрешность прогноза (добавляются два новых положительных слагаемых, пропорциональных дисперсии ошибки измерений σ_{Δ}^2).

Выводы

1. Существует широкий спектр задач регрессионного анализа, в которых объясняющие переменные модели имеют *стохастическую* природу. Генезис возникновения моделей подобного рода включает в себя ситуации, когда значения объясняющих переменных могут измеряться *только со случайными ошибками*, когда в качестве предикторов используются *лагированные* (то есть измеренные в предыдущие моменты времени) значения самой зависимой переменной и др.

2. С точки зрения возможности распространения описанных в главах 4–6 методов анализа линейных *моделей регрессии с неслучайными объясняющими переменными* на модели со *стохастическими* предикторами следует выделить две принципиально различные схемы:

(а) *стохастические объясняющие переменные* $(x^{(1)}, x^{(2)}, \dots, x^{(p)})$ статистически не коррелированы с регрессионными остатками ϵ модели;

(б) *хотя бы одна из объясняющих переменных модели коррелирована с регрессионными остатками*.

В случае (а) методы построения МНК- и ОМНК-оценок и их свойства остаются в силе (при соблюдении достаточно широких условий) и для линейных моделей регрессии со стохастическими предикторами. При этом, если свойства оценок формулируются в терминах их моментов (средних, ковариационных матриц), то в схемах со стохастическими предикторами они понимаются *условно* (при условии заданности матрицы X).

В случае (б) МНК- и ОМНК-оценки оказываются *смещенными и несостоятельными*.

3. Для получения состоятельных оценок в случае (б) используется так называемый *метод инструментальных переменных* (МИП). Реализация этого метода требует от исследователя возможности определения не менее p (где p — число объясняющих переменных в исходной модели) вспомогательных (инструментальных) переменных таких, что:

- *значения этих переменных могут быть определены для каждого из n статистически обследованных объектов (для каждого из n измерений);*
- *ИП должны быть не коррелированы с регрессионными остатками модели и, по возможности, как можно более тесно коррелированы с исходными объясняющими переменными*

При этом в число ИП могут быть включены те из исходных объясняющих переменных, которые не коррелированы с регрессионными остатками модели.

4. Являющиеся состоятельными МИП-оценки модели и их характеристики задаются формулами (7.11)–(7.13).

5. Возможно использование в качестве ИП не реальных показателей, а построенных некоторым специальным способом *искусственных признаков*. Примерами такого способа реализации МИП являются *метод Вальда* и *метод Бартлетта*. Правда, эти методы являются относительно просто конструируемыми *только в рамках модели парной линейной регрессии*.

Глава 8

Линейные регрессионные модели с переменной структурой

Линейные регрессионные модели с переменной структурой рассматриваются в ситуациях, когда в ходе сбора исходных статистических данных $\tilde{B}_n = \{(X_1, y_1), (X_2, y_2), \dots, (X_n, y_n)\}$ имеет место косвенное воздействие (во времени и/или пространстве) некоторых качественных факторов (*сопутствующих переменных*), в результате которого происходят скачкообразные сдвиги в структуре анализируемых линейных связей (то есть в значениях коэффициентов регрессии $\theta_0, \theta_1, \dots, \theta_p$). Попробуем формализовать сказанное.

8.1 Проблема неоднородных (в регрессионном смысле) данных

Пусть функция плотности вероятности анализируемого результирующего показателя y зависит не только от значений объясняющих переменных (регрессоров) $(x^{(1)}, x^{(2)}, \dots, x^{(p)})$, но и от того, на каких конкретных уровнях (градациях) зафиксированы так называемые *сопутствующие качественные переменные* $Z = (z^{(1)}, z^{(2)}, \dots, z^{(k)})^\top$, определяющие условия, в рамках которых проводится сбор исходных статистических данных $\tilde{B}_n = \{(X_1, y_1), (X_2, y_2), \dots, (X_n, y_n)\}$. Тогда эта функция плотности $p_y(\tilde{y} \mid X; Z)$ интерпретируется как *условная плотность вероятности* случайной величины y в точке $\tilde{y}$ при значениях объясняющих и сопутствующих переменных, зафиксированных соответственно на уровнях X и Z , — если объясняющие и сопутствующие переменные случайны по

своей природе, — или как *безусловная* плотность вероятности случайной величины y в точке $\tilde{y}$ при значениях неслучайных параметров, равных соответственно X и Z , если векторные переменные X и Z выполняют роль неслучайных параметров, от которых зависит закон распределения вероятностей результирующей переменной y .

Если исходить из допущения, что при любых условиях (то есть при любых «значениях» сопутствующих переменных Z)¹ функция регрессии y по X остается линейной, то естественно предположить, что коэффициенты этой линейной зависимости (то есть параметры $\Theta = (\theta_0, \theta_1, \dots, \theta_p)^\top$), вообще говоря, зависят от Z , то есть

$$f(X | Z) = \mathbf{E}(y | X, Z) = \theta_0(Z) + \theta_1(Z)x^{(1)} + \dots + \theta_p(Z)x^{(p)}. \quad (8.1)$$

Исходные статистические данные

$$\tilde{B}_n = \{(X_1, y_1, Z_1), (X_2, y_2, Z_2), \dots, (X_n, y_n, Z_n)\}.$$

называются **неоднородными в регрессионном смысле** (или просто — *регрессионно неоднородными*), если в ходе процесса их регистрации найдется хотя бы одна пара наблюдений (с номерами i_1 и i_2) и хотя бы один коэффициент регрессии θ_{j_0} такие, что $Z_{i_1} \neq Z_{i_2}$, причем $\theta_{j_0}(Z_{i_1}) \neq \theta_{j_0}(Z_{i_2})$. Тогда анализируемую модель (8.1) называют **моделью с переменной структурой**.

Проблема, которой посвящена данная глава, состоит в следующем: как наилучшим (в определенном смысле) образом использовать данные $\tilde{B}_n = \{(X_1, y_1), (X_2, y_2), \dots, (X_n, y_n)\}$ для статистического анализа моделей типа (8.1)? Казалось бы, ответ тривиален: разбить имеющиеся в нашем распоряжении исходные статистические данные $\tilde{B}_n$ на заведомо однородные порции (то есть на такие подвыборки, внутри каждой из которых «значения» сопутствующих переменных Z не меняются при переходе от одного наблюдения к другому), а затем оценивать значения коэффициентов $\theta_0(Z), \theta_1(Z), \dots, \theta_p(Z)$ одним из описанных выше методов (выбранным с учетом природы регрессионных остатков модели) *по каждой из таких выборок отдельно*. Тогда для каждого фиксированного сочетания градаций Z^* сопутствующих переменных Z будет определена своя *однородная* подвыборка $\tilde{B}_{n(Z^*)}^*$ объема $n(Z^*)$ ($n(Z^*) < n$) и каждой такой подвыборке соответствует своя оценка искомой функции регрессии (8.1):

$$\hat{f}(X | Z^*) = \hat{\theta}_0(Z^*) + \hat{\theta}_1(Z^*)x^{(1)} + \dots + \hat{\theta}_p(Z^*)x^{(p)}. \quad (8.2)$$

При этом оценки $\hat{f}(X | Z^*)$ и $\hat{f}(X | Z^{**})$, полученные по двум различным однородным подвыборкам, *могут, вообще говоря, статистически*

¹Поскольку сопутствующая переменная является по своей природе нечисловой (качественной), то под ее «значением» подразумевается закодированная градация одного из ее возможных состояний.

значимо отличаться друг от друга, то есть могут представлять собой эмпирические версии двух различных теоретических функций регрессии.

Пример 8.1. Вернемся к условиям примера 7.1, в котором ставилась задача исследования зависимости удельного потребления определенного вида товаров или услуг (y) от величины располагаемого дохода (x) на основании результатов выборочных обследований бюджетов домашних хозяйств. Роль сопутствующих переменных Z в данном примере могут играть как социо-демографо-экономические характеристики, определяющие принадлежность семьи к той или иной социально-экономической страте (или к тому или иному типу потребительского поведения), так и индикатор сезонности, определяющий к какому именно времени года «привязаны» собранные исходные статистические данные. Действительно, предположим, что речь идет о потреблении продуктов питания. Тогда количество потребляемого продукта y , определяемого «склонностью к потреблению» (коэффициентом θ_1) и свободным членом (θ_0) в уравнении (7.14), будет существенно различным для беднейших и состоятельных социально-экономических страт общества: для первых величина θ_1 может достигать значений 0,6–0,7, в то время как для вторых она колеблется в пределах 0,2–0,4. Если же речь идет о потреблении, скажем, пива и прохладительных напитков, то даже в рамках одной и той же страты населения и одинаковых «склопностей к потреблению» θ_1 количество потребляемого продукта будет подвержено существенному сезонному варьированию: очевидно, в модели регрессии, построенной по летним исходным данным, оценка $\hat{\theta}_0$ будет статистически значимо превышать оценку $\hat{\theta}_0$, построенную, например, по зимним исходным данным. В дальнейшем мы еще вернемся к рассмотрению примера 8.1.

К сожалению, предложенный выше рецепт преодоления проблемы регрессионной неоднородности исходных статистических данных, заключающийся в необходимости предварительного разбиения имеющейся выборки $\tilde{B}_n$ на однородные подвыборки $\tilde{B}_{n(Z^*)}^*$, далеко не всегда дает наилучшее решение, а в определенных ситуациях вообще не может привести к сколько-нибудь удовлетворительному решению. Поясним этот тезис.

а) Сопутствующие переменные Z ненаблюдаемы либо их «значения» не были своевременно зарегистрированы при сборе исходных статистических данных. Обращаясь к примеру 8.1, мы можем пояснить эту ситуацию следующим образом: при выборочном статистическом обследовании бюджетов домашних хозяйств не регистрировались ни социо-демографо-экономические характеристики семей, ни времена года, к которым относились данные $\tilde{B}_n = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$. В этих случаях *прямое* (то есть по «значениям» Z) разбиение выборки $\tilde{B}_n$ на регрессионно однородные подвыборки невозможно и с этой целью приходится привлекать подходящие методы кластер-анализа (в

том числе методы расщепления смесей вероятностных распределений) в объединенном $(p+1)$ -мерном пространстве $(x^{(1)}, x^{(2)}, \dots, x^{(p)}, y)$. Описание этих методов см. в Приложении П3.3.

б) Сопутствующие переменные Z наблюдаемы (то есть известны значения $Z_i, i = 1, 2, \dots, n$), но прямое (то есть по «значениям» Z) разбиение выборки $\tilde{B}_n$ на регрессионно однородные подвыборки приводит к слишком малым подвыборкам $\tilde{B}_{n(Z^*)}^*$, то есть к подвыборкам таких объемов, которых оказывается недостаточно для статистически надежной оценки искомых функций регрессии (8.1). В этих случаях для оценки искомых функций регрессии привлекают подходы, связанные с введением так называемых *фиктивных переменных*², или с использованием аппарата дисперсионного или *ковариационного анализа* (описание этих методов ковариационного анализа читатель найдет, например, в книге [Айвазян, Енюков, Мешалкин (1985), п. 13.5]).

8.2 Введение «манекенов» (фиктивных переменных) в линейную модель регрессии

Прием введения в анализируемую линейную модель регрессии так называемых фиктивных переменных (или «переменных манекенов»), отражающих влияние на исследуемый результирующий показатель y *сопутствующих качественных переменных*, используется обычно при работе с *неоднородными* (в регрессионном смысле) исходными статистическими данными в ситуациях, когда изменение значений сопутствующих переменных влияет на изменение значений *только части* коэффициентов регрессии y по X . Использование этого приема в подобных ситуациях оказывается удобным и выгодным по меньшей мере в двух отношениях:

- статистическая надежность (точность) получаемых при этом оценок искомых коэффициентов регрессии будет выше той, которую мы бы имели, оценивая эти коэффициенты *отдельно по каждой однородной выборке*;
- в ходе построения регрессионной модели с фиктивными переменными мы получаем возможность одновременно проверять гипоте-

²В англоязычной литературе по эконометрике эти переменные называются «dummy variables». В соответствующей русскоязычной литературе исторически «привился» неудачный, на наш взгляд, перевод — «фиктивные переменные». Гораздо точнее смысловую нагрузку этого термина передает перевод «*переменные-манекены*» или «*манекенные переменные*», так как из самой постановки задачи следует, что каждая из переменных $z^{(j)}$ *вполне реальна*, а слово «dummy» определяет лишь способ ее кодирования в искомом уравнении.

зы о наличии или отсутствии статистически значимого влияния сопутствующих переменных на структуру анализируемой модели.

Учет влияния сопутствующих переменных на структуру модели осуществляют, как правило, с помощью аддитивно-линейного введения в правую часть регрессионного уравнения определенного числа фиктивных, по своей природе — *дихотомических* (бинарных) переменных, то есть таких переменных, которые могут принимать одно из двух возможных значений.

Обычно рекомендуется использовать в качестве фиктивных переменных *булевы переменные*, то есть переменные, принимающие значения 0 или 1. И делается это в соответствии со следующими правилами.

П р а в и л о 1. *Каждая сопутствующая качественная переменная порождает столько фиктивных переменных, сколько различных «значений» она может принимать, минус единица. Другими словами, если $z^{(j)}$ может принимать m_j возможных «значений», то в правую часть уравнения (8.1) сводятся переменные*

$$z_i^{(j,l)} = \begin{cases} 1, & \text{если в } i\text{-м наблюдении переменная } z^{(j)} \text{ была} \\ & \text{зарегистрирована на уровне своей } (l+1)\text{-й градации;} \\ 0 & \text{в противном случае,} \end{cases}$$

$$l = 1, 2, \dots, m_j - 1.$$

П р а в и л о 2. *Если постулируется зависимость от $z^{(j)}$ свободного члена уравнения (8.1), то в его правую часть добавляются члены $\theta_{0j,l} \cdot z_i^{(j,l)}$, $l = 1, 2, \dots, m_j - 1$. Если же постулируется зависимость от $z^{(j)}$ коэффициента θ_k при $x^{(k)}$, то в правую часть (8.1) добавляются члены вида $\theta_{kj,l} \cdot (z_i^{(j,l)} \cdot x_i^{(k)})$, $l = 1, 2, \dots, m_j - 1$.*

Для пояснения практической реализации этих правил вернемся к нашему примеру 8.1.

П р и м е р 8.1 (продолжение). Предположим, что $y_{\text{руб}}$ — это удельное (то есть в расчете на одного члена семьи и один такт времени) потребление пива и прохладительных напитков. Тогда векторная сопутствующая качественная переменная Z состоит из двух компонент, то есть $Z = (z^{(1)}, z^{(2)})^T$, где $z^{(1)}$ — номер социально-экономической страты (типа потребительского поведения), к которой принадлежит статистически обследованное домашнее хозяйство, а $z^{(2)}$ — номер квартала (сезона). Будем предполагать, что обследуемая совокупность семей подразделяется по уровню дохода на 3 страты: 1 — «низкодоходные», 2 — «среднедоходные» и 3 — «высокодоходные» (то есть число градаций m_1 качественной переменной $z^{(1)}$ равно трем). Из определения $z^{(2)}$ следует, что число ее градаций (m_2) равно четырем (1 — зима, 2 — весна, 3 — лето, 4 — осень).

Чтобы учесть влияние сопутствующей переменной $z^{(1)}$ на структуру модели (7.14), введем $m_1 - 1 = 2$ фиктивные переменные:

$$z_i^{(1.1)} = \begin{cases} 1, & \text{если } i\text{-е наблюдение относилось к домашнему хозяйству,} \\ & \text{принадлежащему страте 2 (то есть к } \textit{среднедоходным}); \\ 0 & \text{в противном случае;} \end{cases}$$

$$z_i^{(1.2)} = \begin{cases} 1, & \text{если } i\text{-е наблюдение относилось к домашнему хозяйству,} \\ & \text{принадлежащему страте 3 (то есть к } \textit{высокодоходным}); \\ 0 & \text{в противном случае.} \end{cases}$$

Учет сезонности (то есть влияния сопутствующей переменной $z^{(2)}$) на структуру модели (7.14) осуществляется введением $m_2 - 1 = 3$ фиктивных переменных:

$$z_i^{(2.1)} = \begin{cases} 1, & \text{если } i\text{-е наблюдение производилось } \textit{весной}; \\ 0 & \text{в противном случае;} \end{cases}$$

$$z_i^{(2.2)} = \begin{cases} 1, & \text{если } i\text{-е наблюдение производилось } \textit{летом}; \\ 0 & \text{в противном случае;} \end{cases}$$

$$z_i^{(2.3)} = \begin{cases} 1, & \text{если } i\text{-е наблюдение производилось } \textit{осенью}; \\ 0 & \text{в противном случае.} \end{cases}$$

Конкретная форма введения этих фиктивных переменных в уравнение (7.14) зависит, как выше сказано, от допущений о характере влияния $z^{(1)}$ и $z^{(2)}$ на структуру модели. Рассмотрим два варианта таких допущений.

Вариант 1. Факторы $z^{(1)}$ и $z^{(2)}$ *влияют на количество потребляемого продукта, но не влияют на склонность к потреблению* (то есть на коэффициент θ_1).

Тогда анализируемое уравнение регрессии следует записать в виде:

$$y_i = \theta_0 + \theta_1 x_i + \theta_{01} z_i^{(1.1)} + \theta_{02} z_i^{(1.2)} + \theta_{03} z_i^{(2.1)} + \theta_{04} z_i^{(2.2)} + \theta_{05} z_i^{(2.3)} + \varepsilon_i. \quad (8.3)$$

Далее, в зависимости от допущений о природе регрессионных остатков ε_i используем обычный или обобщенный метод наименьших квадратов для получения оценок $\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_{01}, \hat{\theta}_{02}, \hat{\theta}_{03}, \hat{\theta}_{04}$ и $\hat{\theta}_{05}$. Располагая значениями этих оценок и их среднеквадратических ошибок, мы можем выписать вид искомой модели для любых сочетаний значений сопутствующих переменных. А именно:

$$y_i = \hat{\theta}_0 + \hat{\theta}_1 x_i + \varepsilon_i \quad \text{— для семей 1-й страты в зимний период;}$$

$$y_i = (\hat{\theta}_0 + \hat{\theta}_{03}) + \hat{\theta}_1 x_i + \varepsilon_i \quad \text{— для семей 1-й страты в весенний период;}$$

$y_i = (\hat{\theta}_0 + \hat{\theta}_{01}) + \hat{\theta}_1 x_i + \varepsilon_i$ — для семей 1-й страты в летний период;
 $y_i = (\hat{\theta}_0 + \hat{\theta}_{05}) + \hat{\theta}_1 x_i + \varepsilon_i$ — для семей 1-й страты осенний период;
 $y_i = (\hat{\theta}_0 + \hat{\theta}_{01}) + \hat{\theta}_1 x_i + \varepsilon_i$ — для семей 2-й страты в зимний период;
 $y_i = (\hat{\theta}_0 + \hat{\theta}_{01} + \hat{\theta}_{03}) + \hat{\theta}_1 x_i + \varepsilon_i$ — для семей 2-й страты в весенний период;

и т. д.

$y_i = (\hat{\theta}_0 + \hat{\theta}_{02} + \hat{\theta}_{05}) + \hat{\theta}_1 x_i + \varepsilon_i$ — для семей 3-й страты в осенний период.

Мы видим, что в рамках допущений варианта 1 структурные изменения модели при переходе из одной страты в другую и из одного сезона в другой выражаются *лишь в изменении значений свободного члена регрессионного уравнения* и не касаются значения параметра $\hat{\theta}_1$, определяющего «склонность к потреблению».

Сделаем ряд важных замечаний.

З а м е ч а н и е 1 (*о статистической надежности метода оценивания, использующего фиктивные переменные*). Надежность статистических выводов при построении модели, как мы знаем, существенно зависит от соотношения объема исходных статистических данных (n) и числа оцениваемых параметров модели ($p+1$). Это, в частности, видно из формул (3.34) и (3.35) для вычисления оценки основной характеристики прогностической силы регрессионной модели (коэффициента детерминации $\hat{R}^2$) и ее среднеквадратической ошибки ($\sqrt{D\hat{R}^2}$): чем больше отношение $n/(p+1)$, тем точнее соответствующие оценки. Попробуем сравнить величины отношений $n/(p+1)$ в двух конкурирующих способах оценки модели (7.14): в способе, использующем аппарат фиктивных переменных, и в обычном методе, примененном *отдельно* к каждой из *однородных* подвыборок (предварительно полученных путем разбиения исходных неоднородных данных $\tilde{B}_n$ на однородные по «значениям» сопутствующих переменных Z порции). Предположим для определенности, что в примере 8.1 статистическому обследованию подверглись 12 семей (по 4 семьи от каждой страты) в течение 12 тактов времени (за такт времени был выбран месяц). Тогда вся совокупность исходных статистических данных будет состоять из 144 наблюдений ($n = 144$), которые распадаются на 12 однородных (по «стратам-сезонам») подвыборок, каждая из которых будет содержать 12 наблюдений (то есть $n_1 = n_2 = \dots = n_{12} = 12$). Поэтому при *изолированной* оценке параметров θ_0 и θ_1 модели (7.14) *отдельно по каждой однородной выборке* отношение $n/(p+1)$ составит величину $12/2 = 6$, в то время как при оценке *семи* параметров модели (8.3) по всем 144 наблюдениям это отношение оказывается равным $144/7 = 20,57$, то есть почти в 3,5 раза больше!

З а м е ч а н и е 2 (*о выявлении наличия/отсутствия статистически значимого влияния сопутствующих переменных на структуру модели*). Предположим, метод наименьших квадратов дал следующий

результат статистического оценивания модели (8.3) (в скобках под значениями оценок коэффициентов регрессии приведены значения среднеквадратических ошибок этих оценок):

$$y_i = \underset{(2,1)}{6,8} + \underset{(0,006)}{0,025} x + \underset{(1,2)}{3,1} z_i^{(1,1)} + \underset{(2,1)}{5,2} z_i^{(1,2)} + \underset{(1,20)}{0,80} z_i^{(2,1)} + \underset{(1,1)}{4,7} z_i^{(2,2)} - \underset{(11,8)}{1,2} z_i^{(2,3)} + \varepsilon_i.$$

Отсюда непосредственно следует, что фиктивные переменные $z^{(2,1)}$ и $z^{(2,3)}$ статистически незначимо влияют на y , то есть удельное потребление пива и прохладительных напитков в зимние месяцы в среднем не отличается от удельного потребления этих продуктов в весенние и осенние месяцы для семей, принадлежащих одним и тем же социально-экономическим стратам. Этот пример демонстрирует возможности регрессионной модели с фиктивными переменными в задаче выявления наличия или отсутствия статистически значимого влияния сопутствующих переменных на структуру регрессионной связи. *В частности, если окажется, что ни одна из фиктивных переменных $z^{(j,1)}, \dots, z^{(j,k_j-1)}$, введенных в модель с целью учета влияния сопутствующей переменной $z^{(j)}$, не влияет на y , то, следовательно, изменение значений этой сопутствующей переменной не влечет за собой неоднородности соответствующих исходных статистических данных.*

З а м е ч а н и е 3 (о необходимости избегать «ловушек», связанных с введением фиктивных переменных). Поясним смысл приведенной выше рекомендации, в соответствии с которой для учета влияния сопутствующей переменной $z^{(j)}$, имеющей m_j градаций, следует вводить $m_j - 1$ фиктивных переменных. Предположим, что в примере 8.1 мы с целью учета влияния фактора принадлежности семьи к той или иной социально-экономической страте ввели не 2, а 3 фиктивные переменные, дополнив определенные выше переменные $z^{(1,1)}$ и $z^{(1,2)}$ переменной

$$z_i^{(1,3)} = \begin{cases} 1, & \text{если } i\text{-е наблюдение относилось к домашнему хозяйству,} \\ & \text{принадлежащему низкодходной страте;} \\ 0 & \text{в противном случае.} \end{cases}$$

Тогда, если сгруппировать имеющиеся n наблюдений таким образом, что первыми пронумерованы все n_1 наблюдений, относящихся к семьям из среднедоходной страты, затем n_2 наблюдений, относящихся к высокодоходной страте, и наконец, n_3 наблюдений, относящихся к низкодходной страте, то столбцы матрицы наблюдений X , задающие наблюдения по переменным $z^{(1,1)}$, $z^{(1,2)}$ и $z^{(1,3)}$, будут иметь вид, представленный в табл. 8.1.

Просуммировав эти три столбца матрицы X , мы получим столбец, состоящий из одних единиц, то есть столбец, повторяющий 1-й столбец

общей матрицы X , соответствующий свободному члену анализируемого уравнения (8.3). А это значит, что столбцы матрицы X *линейно зависимы*, следовательно, ее *ранг меньше числа оцениваемых параметров*. Таким образом, мы оказываемся в условиях *строгой мультиколлинеарности*: матрица $X^T X$ оказывается вырожденной и метод наименьших квадратов «не работает» (см. п. 4.4.1). Легко видеть, что к такому же нежелательному результату мы будем приходить всякий раз, когда число введенных в модель фиктивных переменных, отражающих влияние любой из сопутствующих переменных $z^{(j)}$, будет совпадать с числом градаций этой переменной.

Таблица 8.1. Значения фиктивных переменных, порожденных сопутствующей переменной $z^{(1)}$

i (номер наблюдения)	$z_i^{(1.1)}$	$z_i^{(1.2)}$	$z_i^{(1.3)}$
1	1	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$
n_1	1	0	0
$n_1 + 1$	0	1	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$n_1 + n_2$	0	1	0
$n_1 + n_2 + 1$	0	0	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$n = n_1 + n_2 + n_3$	0	0	1

Вариант 2. Предположим теперь, что фактор принадлежности домашнего хозяйства к той или иной социально-экономической страте ($z^{(1)}$) влияет как раз на характеристику склонности к потреблению θ_1 (фактор сезонности $z^{(2)}$, как и прежде, влияет лишь на количество потребляемого блага). Тогда введенные выше фиктивные переменные следует представить в анализируемом уравнении регрессии в следующей форме:

$$y_i = \theta_0 + \theta_1 x_i + \theta_{11}(z_i^{(1.1)} x_i) + \theta_{12}(z_i^{(1.2)} x_i) + \theta_{01} z_i^{(2.1)} + \theta_{02} z_i^{(2.2)} + \theta_{03} z_i^{(2.3)} + \varepsilon_i. \quad (8.3')$$

При этом очевидным образом меняются способ и результат комплектования общей ($n \times 7$)-матрицы наблюдений X и соответственно результаты метода наименьших квадратов³. Легко видеть, что в данном случае в качестве оценок «склонности к потреблению» мы получим:

³Очевидно, 3-й и 4-й столбцы матрицы X (соответствующие наблюдениям переменных $\tilde{x}^{(2)} = z^{(1.1)}x$ и $\tilde{x}^{(3)} = z^{(1.2)}x$) будут состоять теперь не из нулей и единиц, как это было в варианте 1, а из нулей и значений x_i , причем, $\tilde{x}_i^{(2)} = x_i$ для тех значений i , которые соответствуют наблюдениям над *среднедоходными* семьями (для остальных i : $\tilde{x}_i^{(2)} = 0$), а $\tilde{x}_i^{(3)} = x_i$ для тех значений i , которые соответствуют наблюдениям над *высокодоходными* семьями (для остальных i : $\tilde{x}_i^{(3)} = 0$).

для страты 1: $\hat{\theta}_1$;

для страты 2: $\hat{\theta}_1 + \hat{\theta}_{11}$;

для страты 3: $\hat{\theta}_1 + \hat{\theta}_{12}$.

Сделанные выше замечания 1,2 и 3 в равной мере относятся и к результатам анализа варианта 2.

Учет эффекта взаимодействия сопутствующих переменных

До сих пор введение фиктивных переменных отражало *взаимонезависимое* влияние каждой из сопутствующих переменных на структуру модели. Однако в ряде ситуаций может оказаться существенным влияние, порожденное *взаимодействием* различных сопутствующих переменных. Это взаимодействие также можно отразить в уравнении регрессии введением в него дополнительных (соответствующим образом построенных) фиктивных переменных. А именно: эффект взаимодействия сопутствующей переменной $z^{(j)}$ (представленной в уравнении фиктивными переменными $z^{(j,1)}, z^{(j,2)}, \dots, z^{(j,m_j-1)}$) и сопутствующей переменной $z^{(l)}$ (представленной в уравнении фиктивными переменными $z^{(l,1)}, z^{(l,2)}, \dots, z^{(l,m_l-1)}$) может быть учтен введением в анализируемое уравнение дополнительно $N = (m_j - 1)(m_l - 1)$ фиктивных переменных $\tilde{z}^{(1)}, \tilde{z}^{(2)}, \dots, \tilde{z}^{(N)}$, образуемых всевозможными попарными произведениями вида $\tilde{z} = z^{(j,q)}z^{(l,s)}$, где $q = 1, 2, \dots, m_j - 1$ и $s = 1, 2, \dots, m_l - 1$ (напомним, что m_j и m_l — это числа градаций сопутствующих переменных соответственно $z^{(j)}$ и $z^{(l)}$). Конечно, в ходе статистического анализа полученной таким образом линейной модели множественной регрессии оценки коэффициентов регрессии при многих (или даже всех) переменных — «взаимодействиях» $\tilde{z}^{(t)}$ могут оказаться статистически незначимо отличающимися от нуля, что будет означать отсутствие влияния соответствующих взаимодействий на структуру модели. Поясним это на примере.

Пример 8.2. Пусть y (руб.) — заработная плата работника; $X = (1, x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$ — набор количественных признаков, от которых может зависеть величина y (трудовой стаж; оценки, выставленные работнику по ряду профессиональных тестов, и т. п.);⁴ $z^{(1)}$ — сопутствующая переменная, определяющая уровень образования работника ($m_1 = 3$: начальное, среднее и высшее) и $z^{(2)}$ — сопутствующая переменная, определяющая пол работника ($m_2 = 2$: мужской и женский). В соответствии с приведенными выше рекомендациями вводим фиктивные

⁴ В действительности y и $x^{(j)}$ — это *логарифмы* соответствующих характеристик, так как связь между заработной платой и определяющими ее признаками имеет мультипликативный (степенной) характер. Логарифмирование степенной зависимости позволяет перейти к *линейной аддитивной* модели.

переменные:

$$z_i^{(1.1)} = \begin{cases} 1, & \text{если } i\text{-е наблюдение относится к работнику} \\ & \text{со средним образованием;} \\ 0 & \text{в противном случае;} \end{cases}$$

$$z_i^{(1.2)} = \begin{cases} 1, & \text{если } i\text{-е наблюдение относится к работнику} \\ & \text{с высшим образованием;} \\ 0 & \text{в противном случае;} \end{cases}$$

$$z_i^{(2.1)} = \begin{cases} 1, & \text{если } i\text{-е наблюдение относится к женщине;} \\ 0 & \text{в противном случае;} \end{cases}$$

$$\tilde{z}_i^{(1)} = z_i^{(1.1)} z_i^{(2.1)};$$

$$\tilde{z}_i^{(2)} = z_i^{(1.2)} z_i^{(2.1)}.$$

Таблица 8.2 отражает правила формирования элементов матрицы наблюдений X в той ее части, которая относится к «значениям» фиктивных переменных.

Таблица 8.2. Значения фиктивных переменных при учете «эффекта взаимодействия»

Уровень образования	Пол	Фиктивные переменные				
		$z_i^{(1.1)}$	$z_i^{(1.2)}$	$z_i^{(2.1)}$	$\tilde{z}_i^{(1)}$	$\tilde{z}_i^{(2)}$
Начальное	Мужской	0	0	0	0	0
Начальное	Женский	0	0	1	0	0
Среднее	Мужской	1	0	0	0	0
Среднее	Женский	1	0	1	1	0
Высшее	Мужской	0	1	0	0	0
Высшее	Женский	0	1	1	0	1

Анализируемая линейная модель регрессии в данном случае имеет вид:

$$y_i = \theta_0 + \theta_1 x_i^{(1)} + \dots + \theta_p x_i^{(p)} + \theta_{01} z_i^{(1.1)} + \theta_{02} z_i^{(1.2)} + \theta_{03} z_i^{(2.1)} + \theta_{04} \tilde{z}_i^{(1)} + \theta_{05} \tilde{z}_i^{(2)} + \varepsilon_i. \quad (8.4)$$

Подобный способ представления фиктивных переменных в уравнении регрессии означает, что влияние рассматриваемых сопутствующих переменных (уровня образования и пола работника) может отразиться *лишь на величине свободного члена уравнения*. Это выглядит вполне реалистичным с учетом того факта, что на самом деле мы рассматриваем линейную аддитивную модель между *логарифмами* анализируемых переменных, так что после возвращения к исходным переменным (то

есть после потенцирования уравнения (8.4)) свободный член будет играть роль некоторого поправочного *множителя* в степенной функции от $x^{(1)}, x^{(2)}, \dots, x^{(p)}$.

Применение МНК с целью получения оценок коэффициентов регрессии уравнения (8.4) позволяет выписать *специфицированные* (то есть относящиеся только к однородным по $z^{(1)}$ и $z^{(2)}$ ситуациям) модели зависимости заработной платы от рассматриваемых объясняющих переменных, см. табл. 8.3.

Таблица 8.3. Модели зависимости заработной платы для разных сочетаний градаций сопутствующих переменных

№ п/п	Сочетание значений сопутствующих переменных (образование — пол)	Соответствующая модель регрессии
1	Начальное — мужской	$y_i = \hat{\theta}_0 + \sum_{k=1}^p \hat{\theta}_k x_i^{(k)} + \varepsilon_i$
2	Начальное — женский	$y_i = (\hat{\theta}_0 + \hat{\theta}_{03}) + \sum_{k=1}^p \hat{\theta}_k x_i^{(k)} + \varepsilon_i$
3	Среднее — мужской	$y_i = (\hat{\theta}_0 + \hat{\theta}_{01}) + \sum_{k=1}^p \hat{\theta}_k x_i^{(k)} + \varepsilon_i$
4	Среднее — женский	$y_i = (\hat{\theta}_0 + \hat{\theta}_{01} + \hat{\theta}_{03} + \hat{\theta}_{04}) + \sum_{k=1}^p \hat{\theta}_k x_i^{(k)} + \varepsilon_i$
5	Высшее — мужской	$y_i = (\hat{\theta}_0 + \hat{\theta}_{02}) + \sum_{k=1}^p \hat{\theta}_k x_i^{(k)} + \varepsilon_i$
6	Высшее — женский	$y_i = (\hat{\theta}_0 + \hat{\theta}_{02} + \hat{\theta}_{03} + \hat{\theta}_{05}) + \sum_{k=1}^p \hat{\theta}_k x_i^{(k)} + \varepsilon_i$

Знание оценок $\hat{\theta}_k$ ($k = 0, 1, \dots, p$), $\hat{\theta}_{0j}$ ($j = 1, 2, \dots, 5$) и соответственно их среднеквадратических ошибок $(D\hat{\theta}_k)^{1/2}$ и $(D\hat{\theta}_{0j})^{1/2}$ позволит, в частности, сделать ряд важных в прикладном плане выводов относительно характера анализируемой зависимости. Так, например, если величина $\hat{\theta}_{03}$ окажется статистически значимо отличающейся от нуля *отрицательной* величиной, то придется признать, что существующая система оплаты труда характеризуется определенным дискриминационным (по отношению к женскому труду) свойством. Статистически незначимое отличие от нуля величин $\hat{\theta}_{04}$ и $\hat{\theta}_{05}$ говорило бы о том, что взаимодействие двух рассматриваемых сопутствующих факторов (уровень образования — пол) никак не влияет на структуру анализируемой модели, и т. д.

8.3 Проверка регрессионной однородности двух групп наблюдений (критерий Г. Чоу)

Предположим, сопутствующая переменная z принимала в процессе сбора регрессионных наблюдений

$$\tilde{B}_n = \{(X_1, y_1), (X_2, y_2), \dots, (X_n, y_n)\} \quad (8.5)$$

всего два значения: z_1 и z_2 . Тогда общая выборка $\tilde{B}_n$ может быть поделена на две подвыборки

$$B_{n_1}^{(1)} = \{(X_{i_1}, y_{i_1}), (X_{i_2}, y_{i_2}), \dots, (X_{i_{n_1}}, y_{i_{n_1}}) \mid z = z_1\}, \quad (8.5a)$$

$$B_{n_2}^{(2)} = \{(X_{j_1}, y_{j_1}), (X_{j_2}, y_{j_2}), \dots, (X_{j_{n_2}}, y_{j_{n_2}}) \mid z = z_2\} \quad (8.5b)$$

таким образом, что все наблюдения внутри каждой из подвыборок были произведены *при одном и том же* значении сопутствующей переменной z : 1-я подвыборка — при $z = z_1$ и 2-я подвыборка — при $z = z_2$.

Мы хотели бы получить ответ на вопрос: действительно ли подвыборки $B_{n_1}^{(1)}$ и $B_{n_2}^{(2)}$ неоднородны в регрессионном смысле или переход от градации z_1 к градации z_2 в условиях сбора исходных статистических данных никак не влияет на структуру линейной модели регрессии y по X , и, следовательно, подвыборки $B_{n_1}^{(1)}$ и $B_{n_2}^{(2)}$ можно объединить и строить искомую функцию регрессии по объединенной (общей) выборке $\tilde{B}_n$? Иными словами, мы хотели бы статистически проверить гипотезу

$$H_0: \Theta^{(1)} = \Theta^{(2)}, \mathbf{D}\varepsilon^{(1)} = \mathbf{D}\varepsilon^{(2)} = \sigma^2,$$

где $\Theta^{(j)} = (\theta_0^{(j)}, \theta_1^{(j)}, \dots, \theta_p^{(j)})^\top$ и $\varepsilon^{(j)}$ соответственно коэффициенты регрессии и случайные регрессионные остатки в КЛМНР, генерирующей наблюдения выборки $B_{n_j}^{(j)}$ ($j = 1, 2$).

Если $n_1 > p + 1$ и $n_2 > p + 1$, то можно предложить много способов проверки гипотезы H_0 (например, для принятия гипотезы H_0 достаточно проверить тот факт, что точечные оценки $\hat{\Theta}^{(1)} = (\hat{\theta}_0^{(1)}, \hat{\theta}_1^{(1)}, \dots, \hat{\theta}_p^{(1)})^\top$, полученные по 1-й подвыборке, попадают внутрь интервальных оценок коэффициентов регрессии, построенных по 2-й подвыборке).

Если же объем одной из подвыборок (например, второй) не позволяет провести сколько-нибудь надежную оценку неизвестных коэффициентов (это всегда так при $n_2 \leq p + 1$), то требуется более изощренная процедура проверки гипотезы H_0 . Кстати, именно в такой ситуации достаточно часто оказывается статистик-эконометрист, если к уже имеющейся у него основной выборке $B_{n_1}^{(1)}$ он хочет присоединить полученную позже (или в другом месте) небольшую дополнительную порцию данных вида $B_{n_2}^{(2)}$,

но не знает, можно ли считать выборки $B_{n_1}^{(1)}$ и $B_{n_2}^{(2)}$ регрессионно однородными.

Г. Чоу (G. Chow, Econometrica, vol. 28, 1960, pp. 591–605) предложил следующую процедуру для статистической проверки гипотезы H_0 в данном случае.

1) По основной выборке $B_{n_1}^{(1)}$ строим МНК-оценки $\hat{\Theta}^{(1)}$ параметров модели (4.1) и вычисляем вектор невязок $\hat{\varepsilon}^{(1)} = Y^{(1)} - X^{(1)}\hat{\Theta}^{(1)}$, а затем и сумму квадратов этих невязок $\hat{\varepsilon}^{(1)\top}\hat{\varepsilon}^{(1)}$ (здесь $Y^{(1)}$ и $X^{(1)}$ соответственно $(n_1 \times 1)$ -вектор и $n_1 \times (p+1)$ -матрица наблюдаемых значений анализируемых переменных, соответствующие выборке $B_{n_1}^{(1)}$).

2) По объединенной (общей) выборке $\tilde{B}_n$, содержащей в себе все наблюдения обеих выборок $B_{n_1}^{(1)}$ и $B_{n_2}^{(2)}$, строим МНК-оценки $\hat{\Theta}$ параметров модели (4.1) и вычисляем вектор невязок $\hat{\varepsilon} = Y - X\hat{\Theta}$, а затем и сумму квадратов этих невязок $\hat{\varepsilon}^\top\hat{\varepsilon}$ (здесь Y и X — соответственно $(n \times 1)$ -вектор и $n \times (p+1)$ -матрица наблюдаемых значений анализируемых переменных, соответствующие объединенной выборке $\tilde{B}_n$, $n = n_1 + n_2$).

3) Критическая статистика γ вычисляется по формуле

$$\gamma_{n,n_1} = \frac{(\hat{\varepsilon}^\top\hat{\varepsilon} - \hat{\varepsilon}^{(1)\top}\hat{\varepsilon}^{(1)})/n_2}{\hat{\varepsilon}^{(1)\top}\hat{\varepsilon}^{(1)}/(n_1 - p - 1)}.$$

В предположении справедливости гипотезы H_0 статистика γ_{n,n_1} должна «вести себя» как случайная величина, распределенная по закону $F(n_2, n_1 - p - 1)$. Поэтому, если $\gamma_{n,n_1} > F_\alpha(n_2, n_1 - p - 1)$, то гипотезу H_0 о регрессионной однородности выборок $B_{n_1}^{(1)}$ и $B_{n_2}^{(2)}$ следует отвергнуть (и принять в противном случае). Здесь $F_\alpha(n_2, n_1 - p - 1)$ — $100\alpha\%$ -ная точка F -распределения с числами степеней свободы числителя и знаменателя, равными соответственно n_2 и $n_1 - p - 1$, а α — заданный уровень значимости критерия.

З а м е ч а н и е. Если объем «дополнительной порции» наблюдений n_2 достаточно велик для того, чтобы можно было получить статистически надежные оценки $\hat{\Theta}^{(2)}$ неизвестных параметров регрессии и соответствующие невязки $\hat{\varepsilon}^{(2)} = Y^{(2)} - X^{(2)}\hat{\Theta}^{(2)}$ только по наблюдениям 2-й подвыборки, то для проверки гипотезы H_0 более предпочтительно использовать критическую статистику

$$\gamma_{n_1,n_2} = \frac{(\hat{\varepsilon}^\top\hat{\varepsilon} - \hat{\varepsilon}^{(1)\top}\hat{\varepsilon}^{(1)} - \hat{\varepsilon}^{(2)\top}\hat{\varepsilon}^{(2)})/(p+1)}{(\hat{\varepsilon}^{(1)\top}\hat{\varepsilon}^{(1)} + \hat{\varepsilon}^{(2)\top}\hat{\varepsilon}^{(2)})/(n_1 + n_2 - 2p - 2)}.$$

В предположении справедливости гипотезы H_0 статистика γ_{n_1,n_2} должна «вести себя» как случайная величина, распределенная по закону $F(p+1, n_1 + n_2 - 2p - 2)$. Так что, если $\gamma_{n_1,n_2} > F_\alpha(p+1, n_1 + n_2 - 2p - 2)$, то гипотезу H_0 отвергают с уровнем значимости α (и принимают в противном случае).

8.4 Построение КЛММР по неоднородным данным в условиях, когда значения сопутствующих переменных неизвестны

Если воздействия сопутствующих качественных факторов на структуру КЛММР скрыты, то есть их «значения» не поддаются учету и контролю или не были своевременно (в процессе сбора исходных статистических данных) зарегистрированы, то мы можем оказаться в ситуации, когда собранные исходные статистические данные $\tilde{B}_n$ в действительности представляют собой *смесь нескольких регрессионно однородных подвыборок*, однако выявить эти подвыборки по значениям сопутствующих переменных мы не можем. Игнорирование этого обстоятельства, выражающееся в построении *единой* регрессионной зависимости на основании *всех* имеющихся в распоряжении эконометриста исходных данных, является причиной многих недоразумений и неудач в прикладных эконометрических исследованиях. Для подтверждения этого тезиса рассмотрим следующий пример.

Пример 8.3. В целях исследования зависимости интенсивности региональных миграционных процессов ($y\%$) от уровня (общей продолжительности) полученного образования (x лет) были собраны исходные статистические данные вида (8.5) за определенный промежуток времени в анализируемом регионе. Так что элемент (x_i, y_i) выборки (8.5) интерпретируется в данном случае следующим образом: x_i (лет) — общая продолжительность процесса обучения взрослого (то есть в возрасте не менее 25 лет) жителя региона, y_i — процент уехавших из региона за рассматриваемый промежуток времени взрослых жителей среди всех взрослых жителей с уровнем образования x_i (лет). Несмотря на гипотетическую уверенность специалистов в существовании достаточно тесной статистической связи между x и y , регрессионный анализ данных (8.5) свидетельствовал об отсутствии какой бы то ни было зависимости между этими переменными. Обратившись к визуальному анализу исходных статистических данных (их геометрическое изображение представлено на рис. 8.1), естественно предположить, что наша выборка регрессионно неоднородна и состоит из двух пересекающихся «крестом» подвыборок, каждая из которых имеет вид линейно вытянутого эллиптического облака точек-наблюдений. Более внимательный содержательный анализ каждой из этих подвыборок позволил обнаружить скрытый (не учтенный в ходе сбора статистических данных) сопутствующий признак z . Им оказался *тип полученного образования* с двумя градациями: 1 — гуманитарное, 2 — естественно-научно-техническое. Разделение всех имеющихся данных (8.5) на подвыборки (8.5а) и (8.5б) по этой сопутствующей переменной и построение искомой регрессионной зависимости *отдельно*

для каждой из этих подвыборок дают две различные модели достаточно высокой прогностической силы и противоположной направленности: если для жителей с гуманитарным образованием (им соответствуют «крестики» на рис. 8.1) интенсивность эмиграции *падает* (по линейному закону) с ростом уровня образования, то для жителей с естественно-научно-техническим образованием (им соответствуют точки на рис. 8.1) мы наблюдаем обратную картину.

Возвращаясь к общей проблеме построения линейной регрессионной модели по неоднородным данным в условиях, когда значения сопутствующих переменных неизвестны, отметим, что уже при трех объясняющих переменных (то есть при $p = 3$) визуальный анализ геометрической структуры исходной выборки (8.5) *принципиально невозможен*. Поэтому с целью предварительного анализа геометрической структуры данных (8.5) и выделения в них регрессионно однородных подвыборок необходимо использовать методы кластер-анализа, в том числе методы расщепления смеси многомерных распределений вероятностей (см. Приложение П3.3). Подчеркнем, что разбиение исходных данных (8.5) на регрессионно однородные подвыборки (кластеры) производится в *объединенном* $(p + 1)$ -мерном пространстве $(x^{(1)}, x^{(2)}, \dots, x^{(p)}, y)$, а не в пространстве только объясняющих переменных (как это подчас делается).

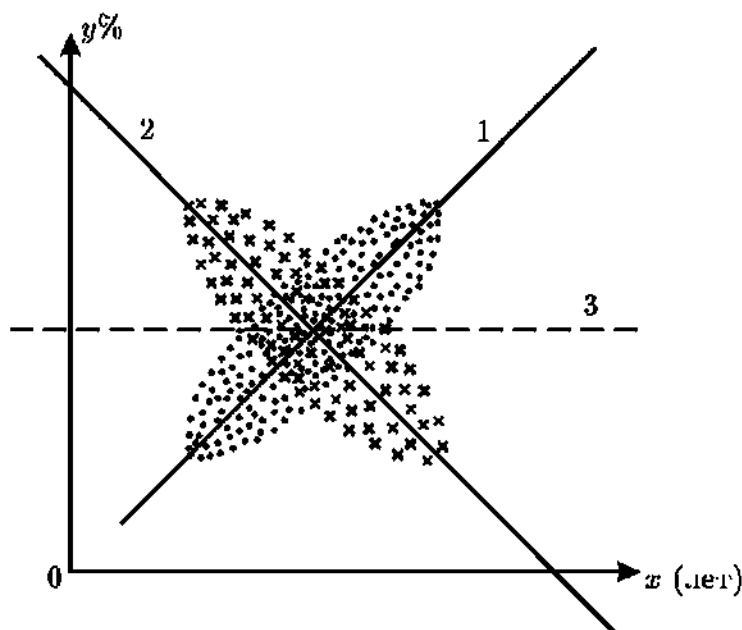


Рис. 8.1. Прямые 1, 2 и 3 — графики аппроксимирующих функций регрессии, построенных соответственно по наблюдениям подвыборок: 1 (точки), 2 (крестики) и по объединенной выборке, состоящей из тех и других наблюдений

Общая процедура построения регрессионной модели по неоднородным данным (8.5) в условиях отсутствия измерений по качественным сопутствующим переменным Z («процедура типологической регрессии») имеет *итерационный* характер. При этом,

каждая итерация состоит из двух шагов. На первом шаге (шаг «Типология») осуществляется разбиение данных (8.5) в соответствующем $(p+1)$ -мерном пространстве $(x^{(1)}, x^{(2)}, \dots, x^{(p)}; y)$ на некоторое число k (вообще говоря, — неизвестное) регрессионно однородных подвыборок $B_{n_j}^{(j)}$ ($j = 1, 2, \dots, k$), а на втором шаге (шаг «Регрессия») строится функция $f_j(X)$ регрессии y по X *отдельно по данным каждой из подвыборок* $B_{n_j}^{(j)}$. Затем построенные на 2-м шаге функции регрессии $f_j(X)$ используются в описании j -го эталона, который является базовым понятием в осуществлении следующей итерации шага «Типология» (то есть происходит уточнение полученного на предыдущей итерации разбиения данных (8.5) на регрессионно однородные подвыборки). После этого по обновленным составам подвыборок $B_{n_j}^{(j)}$ снова строятся регрессии $f_j(X)$ и т. д.

Прокомментируем каждый из двух шагов этой итерационной процедуры.

Шаг 1 («Типология»). На старте процедуры (то есть на самой начальной итерации) мы не располагаем результатами регрессионного анализа отдельно по каждой (j -й) регрессионно однородной подвыборке $B_{n_j}^{(j)}$ и, соответственно, в задаче разбиения данных (8.5) на подвыборки $B_{n_j}^{(j)}$ мы не можем использовать эти результаты для описания эталона j -го класса. Поэтому искомое разбиение на начальной итерации производится либо с помощью метода k -средних, либо (если есть основания постулировать $(p+1)$ -мерную нормальность распределения наблюдений $(x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)}; y_i)$ *внутри каждой регрессионно однородной подсовкупности*) с помощью метода расщепления смеси нормальных распределений (см. краткое описание этих методов в п. ПЗ.3 Приложения 3). Но, начиная со следующей итерации, типологизация наблюдений (8.5) осуществляется с помощью *одного из общих алгоритмов эталонного типа*, когда каждое из наблюдений $(x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)}; y_i)$ относится к классу с тем номером j_0 , расстояние $\rho_i(j_0)$ до эталона которого оказывается наименьшим, то есть $\rho_i(j_0) = \min_{1 \leq j \leq k} \rho_i(j)$. Неоднозначным остается ответ на вопрос о выборе метрики $\rho_i(j)$. Однако вполне естественным представляется задание расстояния $\rho_i(j)$ в моделях **линейных** регрессий в форме

$$\rho_i^2(j) = \rho^2((X_i, y_i); \Pi T(j)) + \rho^2((X_i, y_i); f_j),$$

где $\rho^2((X_i, y_i); \Pi T(j))$ — это квадрат евклидова расстояния от точки (X_i, y_i) до центра тяжести j -й подвыборки, определенной на предыдущей итерации ($\Pi T(j) = \frac{1}{n_j} \sum_{(X_i, y_i) \in B_{n_j}^{(j)}} (X_i, y_i)$), а $\rho^2((X_i, y_i); f_j)$ — это квадрат

длины перпендикуляра, опущенного из точки (X_i, y_i) на гиперплоскость регрессии $y = f_j(X)$.

Примечание. В случае наличия *нескольких* (m) минимумов по j у расстояния $\rho_i(j)$ принимается *рандомизированное* (с вероятностями $\frac{1}{m}$) решение об отнесении точки (X_i, y_i) к одному из этих m классов.

На рис. 8.2а–8.2в схематично представлены варианты возможных расположений подвыборок $B_{n_j}^{(j)}$ для случая парной регрессии ($p = 1$) и двух регрессионно однородных подвыборок ($k = 2$), а также — геометрический смысл предложенного выше расстояния $\rho_i(j)$.

Шаг 2 («Регрессия») осуществляется с помощью стандартных методов статистического оценивания линейной модели множественной регрессии (см. главы 4 и 5) отдельно по данным каждой из полученных на шаге 1 подвыборок $B_{n_j}^{(j)}$ ($j = 1, 2, \dots, k$).

Сходимость подобных итерационных процедур анализируется (и доказывается при некоторых достаточно общих условиях), например, в [Диде и др. (1985)].

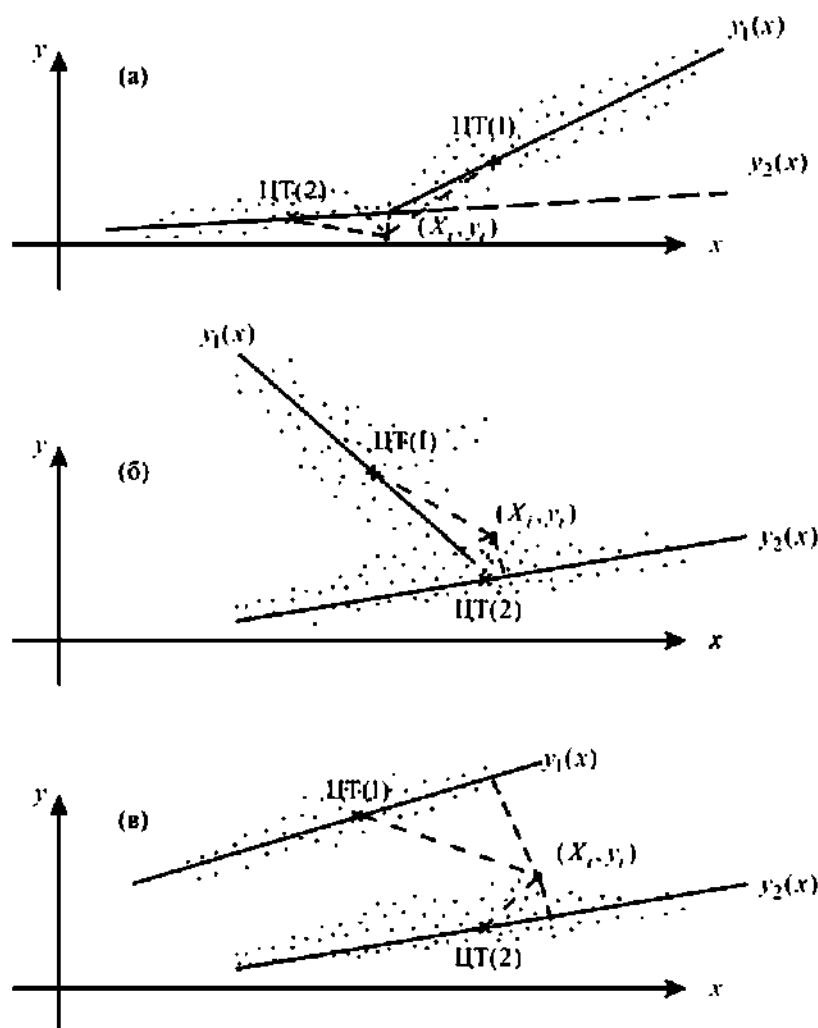


Рис. 8.2. Прямые линии регрессии $y_i(x)$, центры тяжести $\text{ЦТ}(j)$ и расстояния от точки (X_i, y_i) до j -го эталона

Выводы

1. Иногда в ходе сбора исходных статистических данных имеет место косвенное воздействие (во времени и/или в пространстве) некоторых качественных факторов, в результате которого происходят скачкообразные сдвиги в структуре анализируемых линейных связей (то есть в значениях регрессионных коэффициентов $\theta_0, \theta_1, \dots, \theta_p$). В этих случаях говорят об исследовании *регрессионных моделей с переменной структурой* или *о построении ЛММР по неоднородным данным*, а само исследование в зависимости от условий проводится по одной из трех нижеследующих схем.

2. Воздействующие качественные факторы *наблюдаемы* в ходе сбора исходных статистических данных, а объем последних (n) позволяет разбить всю имеющуюся выборку $\{(x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)}; y_i)\}_{i=1,2,\dots,n}$ на регрессионно однородные подвыборки таких объемов $(n_1, n_2, \dots)$, которые обеспечивают возможность статистически надежного регрессионного анализа *отдельно по каждой подвыборке*; при этом воздействие качественных факторов может приводить к скачкообразному изменению *практически всех* регрессионных коэффициентов. В этом случае анализируется столько регрессионных зависимостей, сколько имеется регрессионно однородных подвыборок, причем **построение и анализ модели регрессии производятся отдельно по каждой такой регрессионно однородной подвыборке**.

3. Этот случай отличается от предыдущего тем, что *воздействие качественных факторов может приводить к скачкообразному изменению лишь части регрессионных коэффициентов модели*. Тогда анализ регрессионной модели производится на базе объединенной (регрессионно неоднородной) выборки с помощью введения в модель так называемых фиктивных переменных или в рамках модели ковариационного анализа. Этот подход особенно актуален в условиях дефицита исходных статистических данных.

4. Качественные переменные, под воздействием которых происходят скачкообразные изменения в структуре модели, *ненаблюдаемы или их значения не были своевременно зарегистрированы* при сборе исходных статистических данных. Это значит, что мы не имеем принципиальной возможности разбить имеющиеся исходные статистические данные на регрессионно однородные «порции», ориентируясь при этом на значения этих качественных переменных. В этих случаях рекомендуется использовать **итерационную процедуру «типологической регрессии»**, каждая итерация которой состоит из двух шагов. **На первом шаге** с помощью специальных методов кластер-анализа осуществляется разбиение исходных данных на регрессионно однородные подвыборки, а **на втором** строится регрессия y по X *отдельно по данным*

каждой такой подвыборки. Затем построенные на втором шаге функции регрессии используются для уточнения разбиения исходных данных на регрессионно однородные подвыборки, осуществляемого на первом шаге и т.д.

Глава 9

Модели с дискретными и дискретно-непрерывными зависимыми переменными

В главах 4–8 изучались модели, описывающие зависимость *непрерывного количественного* результирующего показателя y от набора объясняющих переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$, среди которых могли быть и неколичественные показатели (например бинарные фиктивные переменные, см. главу 8). Однако многие практические задачи эконометрического анализа приводят к необходимости строить и статистически оценивать модели зависимости **дискретного** или **дискретно-непрерывного** результирующего показателя y от набора объясняющих переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$. При этом под *дискретным показателем* мы понимаем переменную, множество возможных значений которой *дискретно*. Дискретные переменные могут быть *порядковыми* (ординальными) и *номинальными*.

Дискретная порядковая (ординальная) переменная может принимать только дискретные значения, поддающиеся естественному упорядочению. П р и м е р ы: *уровень благосостояния домашнего хозяйства* (сверхвысокий, высокий, выше среднего, средний, ниже среднего, низкий, ниже черты бедности); *эффективность работы компании* (высокая, средняя, низкая); *уровень образования индивидуума* (высшее, среднее специальное, среднее, незаконченное среднее, начальное); *число просроченных платежей компании за год* и т.п. Заметим, что последний пример относится к *количественным* переменным, однако для нас решающее значение имеет дискретная природа признака.

Дискретная номинальная переменная может принимать только дискретные «значения» (метки), *не поддающиеся упорядочению*. П р и м е р ы: *форма собственности* (государственная, смешанная,

частная); *тип места проживания* (мегаполис, город, поселок городского типа, сельская местность), *профессия опрашиваемого*; *мотив миграции*; *реакция опрашиваемого на предложение заполнить анкету* (согласие, отказ); *решение потенциального покупателя о приобретении определенного товара* (купить, не покупать); *выбор одного из кандидатов при голосовании* и т.п.

Дискретно-непрерывная переменная имеет непрерывную область возможных значений, но при этом может принимать с положительной вероятностью одно или несколько дискретных возможных значений. Часто это может быть, например, тогда, когда зависимая переменная y равна нулю для определенной доли генеральной совокупности, но положительна (с непрерывной областью возможных исходов) для остальной ее части. Например, удельные расходы домашнего хозяйства на товары длительного пользования; официальная сумма проработанного индивидуумом в месяц времени; величина прямых инвестиций за рубежом (за год) и т.п.

Отметим, что рассмотрение моделей, описывающих зависимость дискретно-непрерывной зависимой переменной y от ряда объясняющих переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$, тесно связано с *проблемой цензурированных выборок* и с более общей проблемой смещения в выводах из-за ограничений на процесс формирования выборки («*sample selection problem*»).

Очевидно, рассмотренные в главах 4–8 линейные модели множественной регрессии, позволяющие описывать зависимость *непрерывной количественной* переменной y от набора объясняющих переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ при дискретной и дискретно-непрерывной природе зависимой переменной нуждаются в модификации: ведь правая часть моделей типа (4.1), (5.1), (7.1)–(7.4), всегда определяет величину, способную *непрерывно* меняться в *неограниченном* диапазоне значений числовой оси, в то время как дискретная зависимая переменная y может принимать лишь отдельные значения, а в случае *номинальной* природы — лишь «значения-метки», не имеющие числового смысла.

Для конструирования и анализа зависимостей *дискретных* результирующих показателей от набора объясняющих переменных в эконометрике разработан специальный класс моделей, называемых **моделями бинарного и множественного (упорядоченного или нет) выбора**. Описанию этих моделей посвящены п. 9.1–9.3. Для анализа таких зависимостей в *случае дискретно-непрерывной природы* результирующего показателя (включая ситуации, характеризуемые ограничениями на процесс формирования выборки) привлекаются так называемые «тобит-модели». Их описанию посвящен п. 9.4 данной главы.

9.1 Модели бинарного выбора

Представим себе, что результирующий показатель y , «поведение» которого существенно зависит от количественных объясняющих переменных $X = (1, x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$, является *качественной* переменной, определяющей *одно из двух* возможных состояний характеризуемого ею объекта. Так, например, объясняющие переменные X_i могут описывать различные социально-демографо-экономические характеристики i -го индивида (его доход, возраст, стаж работы и т. п.), а результирующему показателю y_i приписывается значение единица, если i -й индивидум оказался безработным в обследуемом периоде времени, и нуль — в противном случае.

В подобных ситуациях вектор $Y = (y_1, y_2, \dots, y_n)^T$ исходных статистических данных зависимой переменной будет содержать только *дихотомические (бинарные) признаки*, то есть его компоненты y_i могут принимать только два значения: «0» или «1». Как построить регрессионную модель, описывающую зависимость y от X в данном случае?

Для исследования такой статистической связи между y и X строят некоторую специальную регрессионную модель зависимости *вероятности* $P\{y = 1 | X\}$ от линейной формы $\Theta^T X$. Но описывать вероятность *непосредственно линейной функцией* $\Theta^T X$, по-видимому, нецелесообразно, так как в этом случае значения *предсказанной* (вычисленной по модели) вероятности могут быть как отрицательными, так и превосходящими единицу. Вместо этого для моделирования значений $P\{y = 1 | X\}$ подбирают функции, область значений которых определяется отрезком $[0, 1]$, а линейная форма $\Theta^T X$ играет роль аргумента этой функции, то есть

$$P\{y = 1 | X\} = F(\Theta^T X), \quad (9.1)$$

причем функции $F(z)$ должны удовлетворять следующим требованиям:

$$\left. \begin{array}{l} \text{(а)} \quad F(z) \text{ монотонно возрастает по } z; \\ \text{(б)} \quad 0 \leq F(z) \leq 1; \\ \text{(в)} \quad F(z) \rightarrow 1 \text{ при } z \rightarrow \infty; \\ \text{(г)} \quad F(z) \rightarrow 0 \text{ при } z \rightarrow -\infty. \end{array} \right\} \quad (9.2)$$

График одной из функций данного вида приведен на рис. 9.1.

Модели типа (9.1) при ограничениях (9.2) называют обычно *моделями бинарного выбора*. Наиболее распространенными моделями бинарного выбора являются так называемые *логит- и пробит-модели*.

Логит-модель. Модель типа (9.1)–(9.2) называется логит-моделью, если в качестве $F(z)$ рассматривается *логистическая* функция, то есть

$$P\{y_i = 1 | X_i\} = \frac{e^{\Theta^T X_i}}{1 + e^{\Theta^T X_i}}. \quad (9.1a)$$

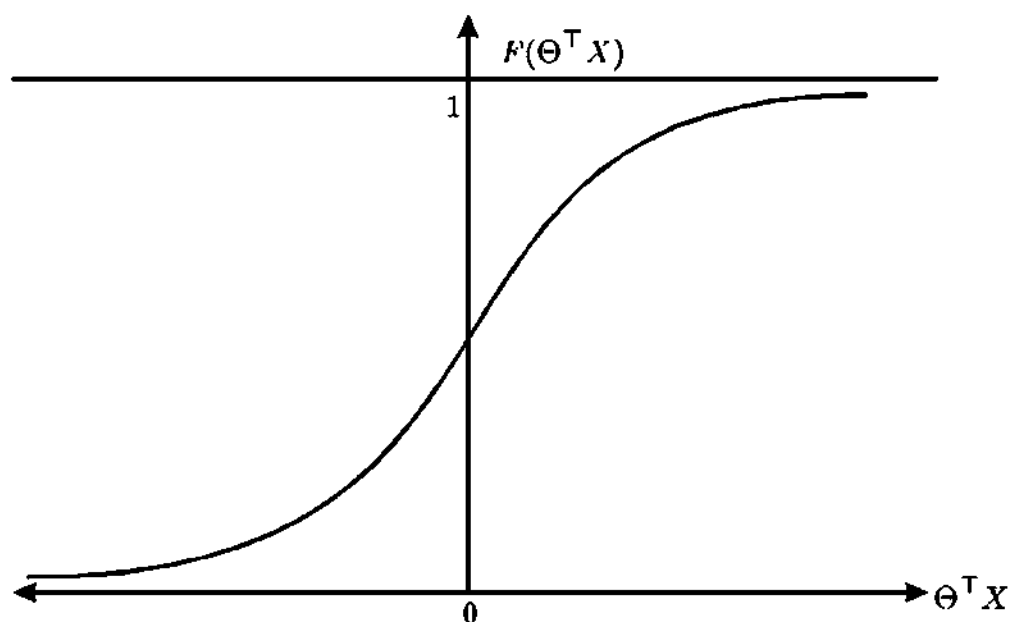


Рис. 9.1. График функции $F(\Theta^T X)$, используемой в модели бинарного выбора

Нетрудно проверить, что эта функция удовлетворяет условиям (9.2). Непосредственной проверкой можно убедиться также, что эта функция симметрична относительно точки $\Theta^T X = 0$, то есть $\Lambda(-z) = 1 - \Lambda(z)$, где

$$\Lambda(z) = \frac{e^z}{1 + e^z}. \quad (9.3)$$

Пробит-модель. Модель типа (9.1)–(9.2) называется пробит-моделью, если в качестве $F(z)$ рассматривается *стандартная нормальная вероятностная функция распределения* $\Phi(z)$, то есть

$$P\{y_i = 1 \mid X_i\} = \Phi(\Theta^T X_i),$$

где $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{x^2}{2}} dx.$ (9.16)

Так же, как и функция $\Lambda(z)$, нормальная вероятностная функция распределения $\Phi(z)$ удовлетворяет всем условиям (9.2) и является симметричной относительно $z = 0$ (то есть $\Phi(-z) = 1 - \Phi(z)$).

Интерпретация модели бинарного выбора с помощью латентного показателя полезности (качества). Словосочетание «*бинарный выбор*» в названии модели оправдано тем, что в широком классе приложений кодирование значений «0» и «1» зависимой переменной y_i соответствует *выбору* лица, принимающего решение (ЛПР), *между двумя альтернативами или состояниями* i -го объекта, характеризуемого значениями $x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)}$ объясняющих переменных. В подобной ситуации естественно предположить, что этот выбор ЛПР производит, руководствуясь величиной некоторой **латентной** (то есть не поддающейся

непосредственному измерению) **характеристики** y_i^* **полезности или качества**, которую, с точностью до традиционной для ЛММР случайной остаточной компоненты ε_i , можно считать линейной функцией от X_i , то есть

$$y_i^* = \Theta^\top \cdot X_i + \varepsilon_i, \quad (9.4)$$

где, как и прежде, $\Theta = (\theta_0, \theta_1, \dots, \theta_p)^\top$, $X_i = (1, x_i^{(1)}, \dots, x_i^{(p)})^\top$, а $E\varepsilon_i = 0$ (предположения о ковариационной матрице вектора остатков $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^\top$ определяются видом распределения вектора остатков ε).

Тогда, без ограничения общности (за счет подбора «подходящего» значения свободного члена θ_0), можно предположить, что ЛПР будет руководствоваться следующим правилом: он выберет решение, закодированное как событие $\{y_i = 1\}$, если полезность (качество) $y_i^* > 0$; в противном случае, то есть при $y_i^* \leq 0$, он выберет решение, закодированное как событие $\{y_i = 0\}$.

Итак:

$$\begin{aligned} P\{y_i = 1 | X_i\} &= P\{y_i^* > 0\} = P\{\Theta^\top X_i + \varepsilon_i > 0\} = \\ &= P\{\varepsilon_i > -\Theta^\top X_i\} = F(\Theta^\top X_i), \end{aligned} \quad (9.5)$$

где $F(z) = P\{\varepsilon < z\}$ — функция распределения случайного остатка ε (справедливость (9.5) следует из гипотетической симметричности распределения $F(z)$ относительно нуля, то есть из $P\{\varepsilon_i > -z\} = P\{\varepsilon_i < z\} = F(z)$). Таким образом, мы получили модель бинарного выбора, вид которой определяется предполагаемой формой распределения ε . Обычно масштаб в измерении полезности (или качества) не играет роли, так что дисперсия распределения ε может быть выбрана по нашему усмотрению. Полагая распределение $F(z)$ стандартным нормальным, мы получаем *пробит-модель* бинарного выбора, а при логистическом распределении $F(z)$ (см. (9.3)) мы имеем *логит-модель*.

Оценивание модели. Для оценивания параметров Θ модели бинарного выбора (9.1) обычно используют *метод максимального правдоподобия*, то есть оценка $\hat{\Theta}_{\text{МП}}$ определяется из условия

$$\hat{\Theta}_{\text{МП}} = \arg \max_{\Theta} \ln L(y_1, y_2, \dots, y_n | \Theta; X_1, X_2, \dots, X_n),$$

где $L(y_1, y_2, \dots, y_n | \Theta; X_1, X_2, \dots, X_n)$ — функция правдоподобия наблюдений $(y_1, y_2, \dots, y_n)$, определенная при значении оцениваемого параметра, равном Θ , и значениях объясняющих переменных, зафиксированных на уровнях $X_1, X_2, \dots, X_n$. Опираясь на взаимную статистическую независимость наблюдений $y_1, y_2, \dots, y_n$ и на тот факт (см. (9.1)), что $P\{y_1 = 1 | X_i\} = F(\Theta^\top X_i)$ и, соответственно, $P\{y_i = 0 | X_i\} = 1 - F(\Theta^\top X_i)$,

имеем:

$$\begin{aligned} L(y_1, \dots, y_n | \Theta; X_1, \dots, X_n) &= \prod_{i: y_i=1} F(\Theta^\top X_i) \times \prod_{i: y_i=0} (1 - F(\Theta^\top X_i)) = \\ &= \prod_{i=1}^n (F(\Theta^\top X_i))^{y_i} (1 - F(\Theta^\top X_i))^{1-y_i}. \end{aligned}$$

Переходя к *логарифмической* функции правдоподобия $l = \ln L$ и дифференцируя ее по Θ , получаем систему уравнений (условий 1-го порядка существования экстремума у функции l) для определения оценок $\hat{\Theta}_{\text{ММП}} = (\hat{\theta}_{0,\text{ММП}}, \hat{\theta}_{1,\text{ММП}}, \dots, \hat{\theta}_{p,\text{ММП}})^\top$:

$$\frac{\partial l}{\partial \Theta} = \sum_{i=1}^n \left[\frac{y_i - F(\Theta^\top X_i)}{F(\Theta^\top X_i)(1 - F(\Theta^\top X_i))} f(\Theta^\top X_i) \right] X_i = \mathbf{0}_{p+1}, \quad (9.6)$$

где $f(z) = F'(z)$ — функция плотности распределения, задаваемого функцией распределения $F(z)$.

В общем случае система (9.6) является системой *нелинейных* относительно Θ уравнений, не поддающейся аналитическому решению (то есть требующей применения специальных численных методов, которые и реализованы в эконометрических и статистических пакетах E-views, STATA, SPSS и др.). Правда, для логит-модели (то есть при $F(z)$ вида (9.3)), с учетом непосредственно проверяемого тождества $\Lambda'(z) = \Lambda(z)(1 - \Lambda(z))$, система (9.6) сводится к более простому виду

$$\sum_{i=1}^n \left(y_i - \frac{e^{\Theta^\top X_i}}{1 + e^{\Theta^\top X_i}} \right) X_i = \mathbf{0}_{p+1}. \quad (9.6')$$

Напомним, что условия 1-го порядка (из которых выведена система (9.6)) являются лишь *необходимыми* условиями существования экстремума у функции $l = \ln L$. Однако анализ условий 2-го порядка (см., например, [Greene (2000)]) свидетельствует о том, что матрица производных функций l по Θ второго порядка является *отрицательно определенной* (если объясняющие переменные в векторе X неколлинеарны), а значит логарифмическая функция правдоподобия l является *вогнутой* по Θ функцией. А это уже гарантирует, что решения системы (9.6) действительно определяют точку локального максимума по Θ функции l .

Проверка гипотез о значениях коэффициентов Θ и влияние ошибок спецификации. Проверка гипотез о значениях коэффициентов Θ в логит- и пробит-моделях и, в частности, гипотез о статистически незначимом отклонении от нуля одного или нескольких из коэффициентов $\theta_0, \theta_1, \dots, \theta_p$, осуществляется в рамках критериев F, W, LR и LM , описанных выше (см. п. 4.6). Соответствующие процедуры реализованы в пакетах E-views, STATA и др.

Ошибки спецификации могут возникать из-за неугадывания общего вида модели $F(z)$, из-за неправильного определения состава объясняющих переменных X , наконец, из-за отклонения закона распределения остатков ε в модели (9.4) от постулируемого.

Существуют статистические критерии выявления подобных ошибок спецификации в логит- и пробит-моделях и методы избавления от них (см., например, [Вербик (2008), п. 7.1.7], [Greene (2000), п. 19.4]). Отметим лишь, что в отличие от КЛММР, в моделях бинарного выбора гетероскедастичность остатков ε_i в (9.4), так же как и избыточность набора объясняющих переменных в векторе X , являются причинами несостоятельности получаемых оценок $\hat{\Theta}$ параметров Θ .

Пример 9.1. Построение и оценка функции отказов респондента от обследования. Данный пример заимствован из [Айвазян, Колеников (2001)]. Речь в этой работе шла об оценке функции плотности распределения домашних хозяйств России по величине среднедушевых расходов и основанных на ней характеристик бедности и дифференциации на базе данных выборочных бюджетных обследований домашних хозяйств (ВБОДХ), ежеквартально проводимых органами государственной статистики. Дело в том, что план ВБОДХ, обеспечивающий репрезентативность выборки, на практике не удается реализовать из-за отказов ряда попавших в план домашних хозяйств от обследования. В итоге выборка оказывается нерепрезентативной, а основанные на ней выводы — смещенными (один из случаев так называемой *проблемы выборочной селективности*, или *the Sample Selection Problem*, см., например, [Вербик (2008), пп. 7.5 и 10.7]). Существуют различные процедуры «взвешивания» («калибровки») имеющихся данных ВБОДХ, нацеленные на устранение (а точнее, — на смягчение) получаемых при этом смещений в статистических выводах. Предложенная в [Айвазян, Колеников (2001)] процедура основана именно на «функции отказов от обследования» $p(Z)$, задающей вероятность того, что домашнее хозяйство, характеризующееся значениями переменных $Z = (1, z^{(1)}, z^{(2)}, z^{(3)})^T$, откажется от участия в ВБОДХ.

В качестве переменных, от которых зависит «вероятность уклонения» p , в данной работе рассматривались характеристики:

$z^{(1)} = \ln \xi$ — логарифм (натуральный) душевых расходов домашнего хозяйства (ДХ);

$z^{(2)}$ — характеристика места проживания ДХ (с градациями «город» — метка «U», «столичные регионы» — метка «M», «сельская местность» — метка «R» и «поселки городского типа» — метка «P»);

$z^{(3)}$ — уровень образования главы семьи (с градациями «ниже среднего» — метка «L», «среднее» — метка «S», «профессионально-технические училища» — метка «P», «техникумы» — метка «T»

и «высшее» – метка «Н»).

Зависимость «вероятности уклонения» p от $Z = (1, z^{(1)}, z^{(2)}, z^{(3)})^T$ анализировалась в рамках логит-модели вида

$$p(Z) = P\{\eta_i = 0|Z\} = \frac{e^{\beta^T Z}}{1 + e^{\beta^T Z}}, \quad (9.7)$$

где $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)^T$ – вектор-столбец искомых (подлежащих статистической оценке) параметров модели (9.7)¹, а

$$\eta_i = \begin{cases} 0, & \text{если соответствующее ДХ уклонилось от обследования,} \\ 1, & \text{если не уклонилось.} \end{cases}$$

При этом, характеристики $z^{(2)}$ и $z^{(3)}$ вводятся в модель в форме *фиктивных переменных* (при гипотезе независимости значения параметра β_1 от разных сочетаний переменных $z^{(2)}$ и $z^{(3)}$), так что переменные $z^{(2)}$ и $z^{(3)}$ можно рассматривать как «сопутствующие», а полученная в результате эконометрического анализа модели (9.7) функция $p((Z) = (1, z^{(1)}, z^{(2)}, z^{(3)})$ в действительности дает нам *целый набор моделей*, описывающих связь между «вероятностью уклонения от обследования» p и логарифмом совокупных душевых расходов $z = z^{(1)}$ при различных сочетаниях градаций переменных $z^{(2)}$ и $z^{(3)}$. Нам будет удобно обозначить элементы этого набора функций с помощью

$$p_{kl}(z) = p(z|z_k^{(2)}, z_l^{(3)}) = P\{\eta_i = 0|z^{(1)} = z, z^{(2)} = z_k^{(2)}, z_l^{(3)}\}, \\ k = 1, 2, 3, 4; \quad l = 1, 2, 3, 4, 5 \quad (9.7')$$

(очевидно, общее число таких функций составит 20).

В таблице 9.1 приведены результаты оценивания параметров β функции (9.7) по данным RLMS, раунды 5–8 (см. [RLMS (1996)] и [Mroz et al. (1997)]). Расчеты подтвердили статистически значимую монотонно возрастающую зависимость вероятности p от z при любых сочетаниях градаций сопутствующих переменных $z^{(2)}$ и $z^{(3)}$.

¹В исходной постановке задачи априорный набор объясняющих переменных $Z = (1, z^{(1)}, z^{(2)}, \dots, z^{(p)})^T$ в логит-модели (9.7) был существенно шире: помимо переменных $z^{(1)}$ (логарифма совокупных душевых расходов ДХ), $z^{(2)}$ (характеристики места проживания семьи) и $z^{(3)}$ (уровня образования главы семьи), включались также 6 переменных, характеризующих социально-демографический состав ДХ (количества детей различного возраста, мужчин-пенсионеров, женщин-пенсионеров, отдельно мужчин и женщин трудоспособного возраста), а также демографическая характеристика главы семьи. Однако последующий статистический анализ коэффициентов β на их статистически значимое отличие от нуля, дополненный анализом существенности различия получаемых функций (9.7') при разных наборах объясняющих переменных, привел, в конечном счете, именно к такому варианту модели.

Там же приведены результаты эконометрического анализа упрощенного (парного) варианта модели (9.7), в котором исследуется зависимость вероятности p только от $z = z^{(1)} = \ln \xi$:

$$p(z) = P\{\eta_i = 0 | z^{(1)} = z\} = \frac{e^{\beta_0 + \beta_1 z}}{1 + e^{\beta_0 + \beta_1 z}}. \quad (9.7'')$$

Таблица 9.1 Результаты анализа многофакторной модели для вероятности отказа от участия в обследовании

	(1)	(2)	(3)	(4)
Медиана расходов	0,396 (0,084)**	0,355 (0,075)**	—	—
Средние расходы	—	—	0,429 (0,089)**	0,399 (0,079)**
Столичные регионы (M)	—	1,052 (0,206)**	—	1,043 (0,203)**
Сельская местность (R)	—	-1,583 (0,292)**	—	-1,576 (0,291)**
ПГТ (P)	—	-0,876 (0,310)**	—	-0,878 (0,308)**
Среднее образование (S)	—	-0,862 (0,156)**	—	-0,868 (0,156)**
ПТУ (P)	—	-1,826 (0,184)**	—	-1,825 (0,182)**
Техникум (T)	—	-1,268 (0,212)**	—	-1,277 (0,213)**
Высшее (H)	—	-0,857 (0,142)**	—	-0,880 (0,142)**
Константа	-4,532 (0,653)**	-3,140 (0,588)**	-4,788 (0,691)**	-3,464 (0,632)**
Кол-во наблюдений	4239	4239	4239	4239
Тест Вальда	Wald (1)= =22,05	Wald(8)= =317,86	Wald(1)= =23,39	Wald(8)= =334,78
Эмпирический уровень значимости	0,00	0,00	0,00	0,00

В скобках указаны среднеквадратические ошибки оценок коэффициентов.

* – значимость на уровне 5%; ** – значимость на уровне 1%.

Природа и структура исходных статистических данных, использованных в качестве информационной базы в примере 9.1.

В каждом раунде RLMS второй волны (то есть начиная с 1994 г.) интервьюеры обходят все домохозяйства, входящие в первичную выборку (4718 домохозяйств) и записывают, был ли проведен опрос в данном домохозяйстве или нет, и если нет, то по какой причине. Наблюдавшаяся частота отказов от обследования приведена в таблице 9.2.

Таблица 9.2. Частота отказов от участия в обследовании

	Раунд 5	Раунд 6	Раунд 7	Раунд 8
Опрос не проведен	743	963	1118	1254
В том числе из-за отказов	410	539	489	701
Отказ из-за нежелания предоставлять информацию о благосостоянии	нет сведений	нет сведений	17	19
Опрос проведен	3973	3781	3750	3831

Конечной целью анализа является ответ на вопрос: «Зависит ли вероятность уклонения от социологического обследования от благосостояния семьи?». На основе вышеупомянутых данных об отказах в сочетании с данными о доходных и расходных характеристиках домохозяйств, приводимых в основных файлах RLMS, были сформированы эконометрические модели с *бинарной зависимой* переменной (отказ/участие) и расходами домохозяйства и рядом других его характеристик в качестве объясняющих переменных.

Очевидно, если домохозяйство отказалось от участия в каком-либо раунде опроса, то данные о его расходах не могут быть получены. Однако, поскольку данные RLMS являются *панельными*, — то есть в разных раундах участвуют одни и те же домохозяйства (по крайней мере потенциально), и другие домохозяйства попасть в выборку не могут, — информацию об уровне благосостояния семьи можно получить из других раундов, считая, что благосостояние семьи в разные раунды примерно постоянно. Отметим, что доходы ДХ относительно более вариабельны, чем его расходы (известный феномен "consumption smoothing").

В расчетах использовалось среднее значение дефлированных расходов за все периоды наблюдения домохозяйства. Эта величина интерпретируется как *постоянные расходы* (постоянное потребление), рассматриваемые в контексте гипотезы Фридмана о доходах в течение жизненного цикла [Deaton (1992)].

Базовыми переменными, которые использовались в анализе вероятности уклонения от обследования в зависимости от дохода, служили потребительские расходы домохозяйств, дефлированные к единому начальному периоду (в базе данных RLMS используется дефлятор «Обзора российской экономики», конструируемый Российско-Европейским центром экономической политики), а также уровень урбанизации местности, в которой проживает домохозяйство ($z^{(2)}$), и уровень образования его главы, точнее, члена домохозяйства с наибольшим доходом ($z^{(3)}$). Используемой мерой благосостояния домохозяйства служит среднее значение логарифмов дефлированных расходов за доступные периоды (максимум — четыре периода, раунды 5–8). Использование среднего за несколько периодов позволяет приблизиться к значению «постоянного потребления» за указанные периоды. В качестве зависимой переменной η фигу-

рирует наблюдаемый факт, отказывалось ли домохозяйство от участия хотя бы в одном из четырех раундов RLMS.

Переменные, характеризующие уровень благосостояния домохозяйства — медиана или среднее за четыре раунда логарифмов реальных (дефлированных к 1992 г.) расходов. В качестве базовой категории региональных переменных при введении соответствующих фиктивных переменных используется градация «город» (U; буквы в скобках относятся к графику, приведенному ниже). Образовательные категории построены по накопительной, а не индикаторной системе, следующим образом: базовая категория — образование ниже среднего (L); далее идет среднее образование (коэффициент при этой категории показывает отличие от категории образование «ниже среднего»); далее ПТУ, техникум, высшее — дополнительно к среднему (то есть измеряют отличия указанных категорий от лиц со средним образованием; среднее образование не исключает наличия технического или высшего, поэтому индикаторы S не исключают индикаторов P, T или H).

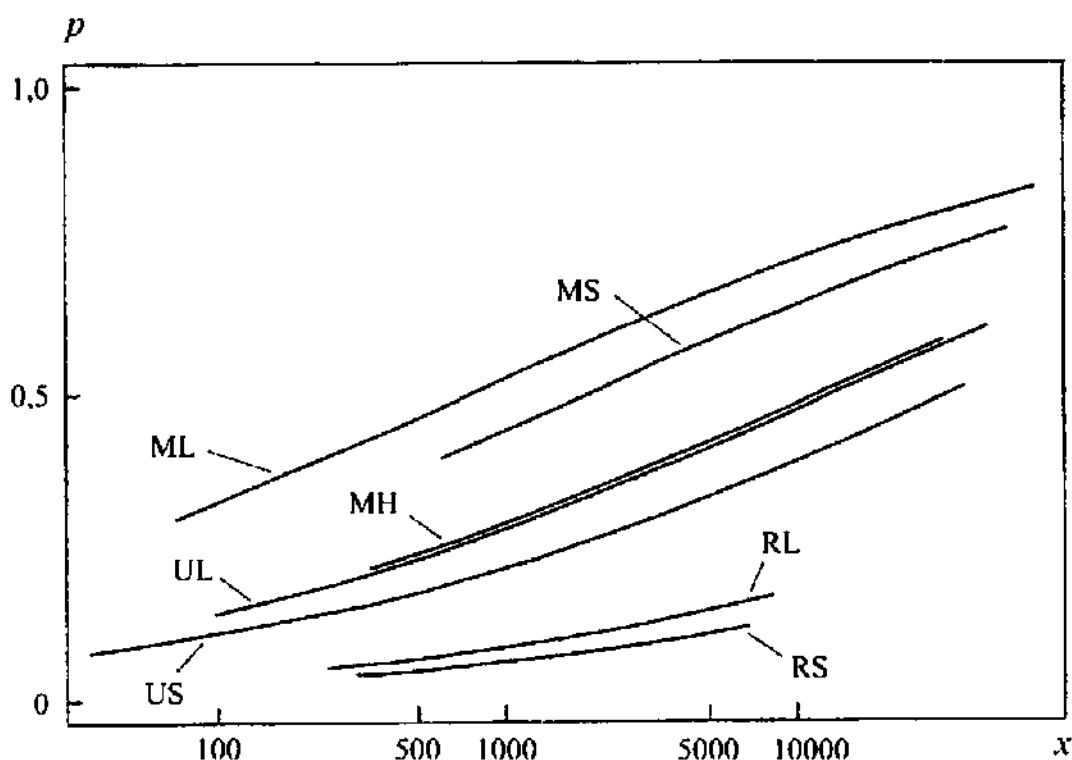


Рис. 9.2. Семейство кривых, описывающих зависимость вероятности p от совокупных душевых расходов x

На рис. 9.2 приводятся графики прогнозных значений вероятности отказа от участия в обследовании для нескольких категорий домохозяйств (шкала совокупных душевых расходов дана в логарифмическом масштабе). Поскольку модель насчитывает четыре географических и

пять образовательных категорий, общее количество частных логистических кривых на графике должно составлять 20. На рис. 9.2 показаны некоторые наиболее «населенные» и представительные из этих кривых.

9.2 Модели множественного выбора

В начале этой главы приводились примеры, когда число рассматриваемых альтернатив, из которых производится выбор, оказывается *больше двух* (выбор одного из нескольких кандидатов при голосовании, уровень благосостояния домашнего хозяйства и т.д.). *Модели множественного выбора описывают зависимость вероятности каждого из возможных исходов для анализируемого дискретного результирующего признака y от значений объясняющих переменных X , то есть*

$$P\{y = j|X\} = \varphi_j(X), \quad j = 1, 2, \dots, k.$$

Подходы к построению моделей такого типа зависят от того, являются ли рассматриваемые альтернативы *упорядочиваемыми* (тогда говорят о *моделях упорядочиваемого множественного выбора*) или они не поддаются упорядочению (в этом случае речь идет о *мультиномиальных моделях множественного выбора* или о *моделях неупорядочиваемого множественного выбора*).

9.2.1 Модели упорядочиваемого множественного выбора

Начнем с примера. Исходные статистические данные при анализе одной из важнейших синтетических категорий качества жизни населения (КЖН) n регионов России — *уровня материального благосостояния (УМБ)* имели вид

$$(x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)}; y_i), \quad i = 1, 2, \dots, n, \quad (9.8)$$

где официальные статистические показатели $x_i^{(j)}$ характеризовали различные аспекты УМБ (среднюю величину и уровень дифференциации душевного дохода, долю бедных, обеспеченность жильем, личными автомобилями, мощностями инфраструктуры и т.п.), а y_i — полученные от экспертов оценки категории УМБ, построенные следующим образом:

$$y_i = \begin{cases} 1, & \text{если } i\text{-й регион отнесен к категории «аутсайдеров»;} \\ 2, & \text{если } i\text{-й регион отнесен к категории «середняков»;} \\ 3, & \text{если } i\text{-й регион отнесен к категории «лидеров»}. \end{cases}$$

Синтетическая категория УМБ относится к категориям *латентным*, то есть непосредственно не поддающимся измерению. **Целью нашего исследования** в данном случае было построить такую модель, которая

позволила бы по значениям объясняющих переменных $x_0^{(1)}, x_0^{(2)}, \dots, x_0^{(p)}$, характеризующим регион, не попавший в число оцененных экспертами, оценить вероятности $P\{y = j|X\}$. В подобных ситуациях зависимая переменная y является порядковой (ординальной), и естественно строить искомую модель на предположении, что существует некоторый количественный интегральный измеритель y^* анализируемой латентной категории, связанный с исходными данными (9.8) множественным соответствием вида:

$$\begin{cases} y_i^* = \Theta^\top X_i + \varepsilon_i, \\ y_i = j, \text{ если } c_{j-1} < y_i^* \leq c_j, \quad j = 1, 2, \dots, k, \end{cases} \quad (9.9)$$

где коэффициенты $\Theta = (\theta_0, \theta_1, \dots, \theta_p)^\top$ и $(c_0, c_1, \dots, c_k)$ подлежат статистическому оцениванию по исходным данным (9.8) (очевидно, в нашем примере $k = 3$). Нетрудно убедиться, что данная схема является простым обобщением модели (9.4)-(9.5) бинарного выбора, *использующей понятие латентного показателя полезности (качества)*.

Условия нормировки (необходимые для идентифицируемости модели (9.9), см. [Вербик (2008), п. 7.2.2]) позволяют положить априори $c_0 = -\infty, c_k = \infty$ и $D\varepsilon_i = 1$, так что из (9.9) следует

$$P\{y_i = j|X_i\} = F(c_j - \Theta^\top X_i) - F(c_{j-1} - \Theta^\top X_i), \quad j = 1, \dots, k, \quad (9.10)$$

где $F(z)$, как и прежде, функция распределения случайного остаточного компонента ε_i из (9.9). Выбирая $F(z)$ в форме (9.3), мы получаем *логит-модель множественного упорядоченного выбора*, а выбирая $F(z)$ в форме (9.16), мы имеем соответствующую *пробит-модель*.

Оценивание параметров Θ и $C = (c_1, c_2, \dots, c_{k-1})$ модели (9.9)-(9.10) производится с помощью *метода максимального правдоподобия*. С учетом взаимной статистической независимости остатков $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ можно выразить их функцию правдоподобия в форме:

$$\begin{aligned} L(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n | \Theta; C; X_1, \dots, X_n) = \\ = \prod_{j=1}^k \prod_{i: y_i=j} (F(c_j - \Theta^\top X_i) - F(c_{j-1} - \Theta^\top X_i)), \end{aligned} \quad (9.11)$$

так что оценки $\hat{\Theta}_{\text{ммп}}$ и $\hat{C}_{\text{ммп}}$ получаются как решение задачи

$$(\hat{\Theta}_{\text{ммп}}; \hat{C}_{\text{ммп}}) = \arg \max_{\Theta; C} L(\varepsilon_1, \dots, \varepsilon_n | \Theta; C; X_1, \dots, X_n),$$

где функция L определена соотношением (9.11).

9.2.2 Модели неупорядочиваемого множественного выбора

Если результирующая (зависимая) переменная y является по своей природе *номинальной* и число возможных «значений» (состояний) этой переменной больше двух, то нам приходится иметь дело с *моделями неупорядочиваемого множественного выбора*. Примеры подобных ситуаций приводились в начале главы (форма собственности; отрасль, в которой занят работник; профессия опрашиваемого респондента и т.п.).

Построение моделей в данном случае основано на весьма сильных (и часто – *малореалистичных*) модельных допущениях и усложнено проблемами идентифицируемости (подробное изложение этой части инструментария, предназначенного для построения и анализа моделей множественного выбора, дано, например, в [Greene (2000), п. 19.7]). Мы приведем здесь без вывода лишь один частный вид подобных моделей, являющийся обобщением логит-модели *бинарного* выбора на случай *нескольких* ($k > 2$) альтернатив, а потому называющийся **логит-моделью неупорядочиваемого множественного выбора**. Построение данного типа моделей основано на предположении, что для индивидуума (ЛПР) i альтернатива j имеет полезность $u_{ij} = c_{ij} + \varepsilon_{ij}$, где c_{ij} – некоторая нестохастическая функция наблюдаемых переменных X_i и какого-то числа неизвестных параметров, которые, вообще говоря, могут зависеть от номера альтернативы j , а случайные остатки ε_{ij} имеют функцию распределения $F(z) = \exp\{-e^{-z}\}$. Тогда можно показать, что

$$P\{y_i = j | X_i\} = \frac{e^{\Theta(j) \cdot X_i}}{\sum_{j=1}^k e^{\Theta(j) \cdot X_i}}, \quad j = 1, 2, \dots, k, \quad (9.12)$$

где $\Theta(j) = (\theta_0(j), \theta_1(j), \dots, \theta_p(j))^T$, причем $\Theta(1) = \mathbf{0}_{p+1}$ (условие нормировки, обеспечивающее идентифицируемость модели (9.12)).

Параметры $\Theta(2), \Theta(3), \dots, \Theta(k)$ модели (9.12) оцениваются с помощью метода максимального правдоподобия, то есть

$$\begin{aligned} &(\hat{\Theta}(2), \dots, \hat{\Theta}(k)) = \\ &= \arg \max_{\Theta(2), \dots, \Theta(k)} L(y_1, y_2, \dots, y_n | \Theta(2), \dots, \Theta(k); X_1, \dots, X_n), \end{aligned}$$

где функция правдоподобия $L(y_1, \dots, y_n | \Theta(2), \dots, \Theta(k); X_1, \dots, X_n)$ в предположении независимости наблюдений имеет вид:

$$L(y_1, \dots, y_n | \Theta(2), \dots, \Theta(k); X_1, \dots, X_n) = \prod_{j=1}^k \prod_{i: y_i=j} \frac{e^{\Theta(j) \cdot X_i}}{\sum_{j=1}^k e^{\Theta(j) \cdot X_i}}$$

(напомним, что $\Theta(1) = \mathbf{0}_{p+1}$).

Кратко напомним постановку задачи *дискриминантного анализа* (см. также п. П3.2 Приложения 3). Дискретная зависимая переменная y_i , порядковая или номинальная, определяет номер класса j ($j = 1, 2, \dots, k$), к которому относится объект i , характеризуемый значениями объясняющих переменных $X_i = (x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)})$. Исследователь располагает данными вида (9.8), которые в дискриминантном анализе чаще представляются в форме:

где X_{jl} ($l = 1, 2, \dots, n_j$) — это значения объясняющих переменных тех n_j наблюдений, у которых значение y было зафиксировано на уровне $y = j$ (очевидно, $n_1 + n_2 + \dots + n_k = n$). Подвыборки данных, представленные в (9.13), называются в дискриминантном анализе «обучающими выборками». Задача дискриминантного анализа: используя данные (9.13), построить модель, позволяющую наилучшим (в определенном смысле) образом прогнозировать неизвестное значение (номер класса) y по заданным значениям объясняющих переменных $(x^{(1)}, x^{(2)}, \dots, x^{(p)})$.

Таким образом, мы видим, что и модели бинарного и множественного выбора, и модели дискриминантного анализа нацелены на решение одной и той же задачи построения и анализа регрессии порядковой или номинальной переменной y по объясняющим переменным $(x^{(1)}, x^{(2)}, \dots, x^{(p)})$.

При этом, правда, в моделях бинарного и множественного выбора центральным объектом исследования являются *апостериорные вероятности* $p_j(X) = P\{y = j|X\}$, в то время как оптимальное (в смысле максимизации вероятности правильной классификации объектов) правило классификации объекта $(x^{(1)}, x^{(2)}, \dots, x^{(p)})$ в дискриминантном анализе основано на *отношениях правдоподобия* $\gamma_{jl}(X) = \pi_j f_j(X) / \pi_l \cdot f_l(X)$, где π_j и $f_j(X)$ — это, соответственно, априорная вероятность принадлежности объекта классу j и функция плотности вероятности распределения вектора $X = (x^{(1)}, \dots, x^{(p)})$ внутри j -го класса.

Можно показать, однако, что эти характеристики (то есть $p_j(X)$ и $\gamma_{ij}(X)$) связаны чисто функциональными соотношениями, а способы параметризации моделей, используемых в задачах бинарного и множественного выбора, могут находить свое обоснование (или опровержение) в постулатах дискриминантного анализа.

Продemonстрируем это на примере $k = 2$ (то есть на моделях *бинарного* выбора и дискриминантного анализа при *двух* классах). С этой

целью докажем два утверждения.

Утверждение 1. Апостериорная вероятность $p(X) = P\{j = 1|X\}$ в модели бинарного выбора и отношение правдоподобия $\gamma(X) = \frac{\pi_1 f_1(X)}{\pi_2 f_2(X)}$ в модели дискриминантного анализа при двух классах связаны соотношением

$$p(X) = \frac{\gamma(X)}{1 + \gamma(X)}. \quad (9.14)$$

Доказательство. Формула (9.14) немедленно следует из формулы Байеса $P(A_j|B) = \frac{P(A_j) \cdot P(B|A_j)}{\sum_{j=1}^k P(B|A_j) \cdot P(A_j)}$ (где $A_1, A_2, \dots, A_k$ — полная система событий). В нашем случае $k = 2$, $A_j = \{y = j\}$ ($j = 1, 2$), $P(A_j) = P\{y = j\} = \pi_j$, $B = \{X\}$ и $P(B|A_j) = f(X|y = j) = f_j(X)$. Так что

$$P\{y = 1|X\} = \frac{\pi_1 f_1(X)}{\pi_1 f_1(X) + \pi_2 f_2(X)}$$

и после деления числителя и знаменателя правой части этого выражения на $\pi_2 \cdot f_2(X)$ получаем формулу (9.14) (предполагается, что вероятностные меры $f_1(X)$ и $f_2(X)$ абсолютно непрерывны относительно друг друга).

Утверждение 2. Если вектор объясняющих переменных X распределен нормально внутри каждого из двух анализируемых классов и распределения X внутри классов различаются только векторами средних значений (то есть $(X|j) \in N_p(a(j); \Sigma)$), то апостериорная вероятность $p(X) = P\{y = 1|X\}$ может быть представлена в виде

$$p(X) = \frac{e^{\beta^T X}}{1 + e^{\beta^T X}}, \quad (9.15)$$

где $X = (1, x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$, $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$ — вектор некоторых коэффициентов, выражающихся в терминах $a(1), a(2)$, Σ , π_1 и π_2 .

Следствие. Утверждение 2 дает строгое обоснование использованию именно логит-модели (9.3) в случае $(X|j) \in N_p(a(j); \Sigma)$, $-j = 1, 2$, — в качестве модели бинарного выбора.

Доказательство. Воспользовавшись видом функции плотности многомерного нормального распределения, проанализируем отношение правдоподобия $\gamma(X)$:

$$\begin{aligned} \gamma(X) &= \frac{\pi_1 f_1(X)}{\pi_2 f_2(X)} = \frac{\pi_1}{\pi_2} e^{\frac{1}{2}[(X-a(1))^T \Sigma^{-1}(X-a(1)) - (X-a(2))^T \Sigma^{-1}(X-a(2))]} = \\ &= e^{\ln \frac{\pi_1}{\pi_2} + D(X)}, \end{aligned} \quad (9.16)$$

где

$$D(X) = -\frac{1}{2} [((X - a(1))^T \Sigma^{-1} (X - a(1)) - (X - a(2))^T \Sigma^{-1} (X - a(2)))]. \quad (9.17)$$

Раскрытие скобок и выполнение требуемых действий в правой части соотношения (9.17) показывает, что $D(X)$ — **линейная** функция от X , поэтому подстановка полученного в (9.16) выражения $\gamma(X)$ в (9.14) (с учетом линейности $D(X)$) завершает доказательство справедливости формулы (9.15).

Помимо явной связи, существующей между задачами и инструментарием моделей бинарного и множественного выбора, с одной стороны, и дискриминантного анализа — с другой, приведенные выше результаты свидетельствуют о том, что возможны ситуации, когда *подход, основанный на методах дискриминантного анализа, может оказаться более гибким и эффективным*. Действительно, если например, распределение вектора X внутри классов не удовлетворяет условиям утверждения 2 (не является нормальным или, при нормальности, не соблюдается равенство ковариационных матриц в разных классах), то дискриминантный анализ оставляет возможности использования **нелинейной** дискриминантной функции $D(X)$ (при неодинаковых ковариационных матрицах) или применения методов непараметрического оценивания отношения правдоподобия $\gamma(X)$. Поэтому традиционное пренебрежение подобными методами многомерного статистического анализа, к сожалению, присущее даже лучшим мировым образцам учебной эконометрической литературы, мы расцениваем как сильно затянувшееся во времени недоразумение.

9.4 Модель с дискретно-непрерывной зависимой переменной (тобит-модель)

В параграфах 9.1 и 9.2 нам пришлось модифицировать стандартную линейную модель множественной регрессии (ЛММР) с учетом дискретной (в том числе — неколичественной) природы зависимой переменной y . Как правило, в основе таких модификаций лежало введение некоей *латентной* (ненаблюдаемой) переменной y^* , которую, с одной стороны, мы могли анализировать в рамках ЛММР по рассматриваемым объясняющим переменным X , а с другой, — поставить ее в такое соответствие с *наблюдаемой* зависимой переменной y , при котором закон распределения последней определяется наблюдаемыми значениями объясняющих переменных $X = (1, x^{(1)}, \dots, x^{(p)})^T$.

Теперь мы рассмотрим ситуацию, которая также требует модификации привычной ЛММР по *причине специфической природы зависимой переменной y* . Однако в данном случае эта специфическая природа зависимой переменной выражается в том, что, имея некоторую *непрерыв-*

ную область своих возможных значений, эта переменная, тем не менее, может принимать, одно или несколько *дискретных* значений (каждое из них – с некоторой положительной вероятностью). Классические примеры подобного рода – это исследование зависимости расходов (y) на табак или предметы длительного пользования от доходов и других социально-демографо-экономических характеристик X . Очевидно, что, во-первых, $y \geq 0$, а во-вторых, может существовать определенная ненулевая доля генеральной совокупности потребителей, которая имеет нулевые расходы на этот товар, то есть $P\{y = 0\} = q > 0$. В этом случае, как мы увидим чуть позднее, если строить с помощью МНК обычную линейную регрессию y по X , то полученные оценки коэффициентов Θ окажутся несостоятельными и смещенными. Выйти из положения нам и здесь помогает введение некоторой латентной переменной y^* . В частности, рассматривают модель в форме:

$$y_i^* = \Theta^\top \cdot X_i + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (9.18)$$

$$y_i = \begin{cases} y_i^*, & \text{если } y_i^* > 0, \\ 0, & \text{если } y_i^* \leq 0, \end{cases} \quad (9.19)$$

где (9.18) – это стандартная *нормальная* КЛММП (см. (4.1')). Модель (9.18)–(9.19) носит название **тобит-модели** по сочетанию двух причин: впервые подобная схема была рассмотрена Джеймсом Тобином (см. *Tobin J. Relationships for Limited Dependent Variables. – Econometrica*, vol. 26 (1958)), pp. 24–36), обнаруживая, при этом, идейное сходство с логит- и пробит-моделями.

Теперь, опираясь на (9.18)–(9.19) и на некоторые известные факты, касающиеся усеченного нормального распределения, попробуем вывести функцию регрессии y по X :

$$\mathbf{E}(y_i|X_i) = \mathbf{E}(y_i|X_i; y_i^* \leq 0) \cdot P(y_i^* \leq 0) + \mathbf{E}(y_i|X_i; y_i^* > 0) \cdot P(y_i^* > 0). \quad (9.20)$$

Рассмотрим каждое из слагаемых правой части этого выражения:

$$\mathbf{E}(y_i|X_i; y_i^* \leq 0) = 0 \quad (9.20a)$$

по построению (см. (9.19));

$$\begin{aligned} P(y_i^* > 0) &= P(\Theta^\top X_i + \varepsilon_i > 0) = P(\varepsilon_i > -\Theta^\top X_i) = \\ &= 1 - \Phi\left(-\frac{\Theta^\top X_i}{\sigma}\right) = \Phi\left(\frac{\Theta^\top X_i}{\sigma}\right); \end{aligned} \quad (9.20б)$$

$$\begin{aligned} \mathbf{E}(y_i|X_i; y_i^* > 0) &= \mathbf{E}(y_i^*|X_i; y_i^* > 0) = \mathbf{E}(\Theta^\top X_i + \varepsilon_i|X_i; \Theta^\top X_i + \varepsilon_i > 0) \\ &= \Theta^\top X_i + \mathbf{E}(\varepsilon_i|X_i; \varepsilon_i > -\Theta^\top X_i). \end{aligned} \quad (9.20в)$$

Но последнее слагаемое в правой части (9.20в) это среднее значение $(0; \sigma^2)$ -нормального распределения, усеченного в точке $-\Theta^\top X_i$. Для его вычисления воспользуемся известным фактом (см., например, *Johnson N. and S. Kotz. Distributions in Statistics. Continuous Univariate Distributions. – John Wiley and Sons, 1970*):

если $\xi \in N(a; \sigma^2)$, то

$$E(\xi | \xi > c) = a + \sigma \frac{\varphi\left(\frac{c-a}{\sigma}\right)}{1 - \Phi\left(\frac{c-a}{\sigma}\right)}, \quad (9.20г)$$

где $\varphi(z)$ и $\Phi(z)$ — значения функций, соответственно, плотности и распределения в точке z стандартного нормального закона.

Поэтому, возвращаясь к (9.20) и учитывая (9.20а)–(9.20г), имеем:

$$E(y_i | X_i) = \Theta^\top X_i \Phi\left(\frac{\Theta^\top X_i}{\sigma}\right) + \sigma \varphi\left(\frac{\Theta^\top X_i}{\sigma}\right). \quad (9.20')$$

Мы видим, что спецификация функции регрессии y по X существенно отличается от спецификации функции регрессии y^* по X , определенной соотношением (9.18), и, соответственно, МНК-оценки коэффициентов Θ в уравнении

$$y_i = \Theta^\top X_i + \tilde{\varepsilon}_i,$$

вообще говоря, будут несостоятельными и смещенными.

Оценивание тобит-модели. Состоятельные и асимптотически несмещенные оценки параметров Θ и σ^2 тобит-модели (9.18)–(9.19) получают с помощью метода максимального правдоподобия. Учитывая смешанный характер наблюдений (9.19), получаем функцию их правдоподобия:

$$\begin{aligned} L(y_1, \dots, y_n | \Theta; \sigma^2; X_1, \dots, X_n) &= \\ &= \prod_{i: y_i=0} \left(1 - \Phi\left(\frac{\Theta^\top X_i}{\sigma}\right)\right) \times \prod_{i: y_i>0} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i - \Theta^\top X_i)^2}. \end{aligned}$$

Сомножители под первым знаком произведения $\left(\prod_{i: y_i=0}\right)$ соответствуют тем наблюдениям, для которых значения y_i оказались нулевыми, а сомножители под вторым знаком произведения $\left(\prod_{i: y_i>0}\right)$ соответствуют всем остальным наблюдениям.

Переходя к логарифмической функции правдоподобия $l = \ln L$, получаем оценки $\hat{\Theta}_{\text{МП}}$ и $\hat{\sigma}_{\text{МП}}^2$ как решение следующей оптимизационной

задачи:

$$(\hat{\Theta}_{\text{ММП}}, \hat{\sigma}_{\text{ММП}}^2) = \arg \max_{\Theta, \sigma^2} \left[\sum_{i: y_i=0} \ln \left(1 - \Phi \left(\frac{\Theta^\top X_i}{\sigma} \right) \right) + \right. \\ \left. + \sum_{i: y_i>0} \ln \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i - \Theta^\top X_i)^2} \right) \right]. \quad (9.21)$$

Вычисление оценок $\hat{\Theta}_{\text{ММП}}$ и $\hat{\sigma}_{\text{ММП}}^2$ программно реализовано в ряде эконометрических и статистических пакетов (E-views, STATA и др.).

Интерпретация коэффициентов тобит-модели. Во-первых, оценив модель (9.18)–(9.19)–(9.20'), мы можем оценить долю нулевых значений наблюдаемой зависимой переменной y для наблюдений, характеризуемых значениями X_i анализируемых объясняющих переменных, а именно:

$$\hat{P}(y=0|X=X_i) = 1 - \Phi \left(\frac{\hat{\Theta}^\top X_i}{\hat{\sigma}} \right).$$

Кроме того, мы можем оценить предельный эффект влияния переменной $x^{(j)}$ на ожидаемый исход y_i :

$$\frac{\partial \mathbf{E}(y_i|X_i)}{\partial x_i^{(j)}} = \hat{\theta}_j \Phi \left(\frac{\hat{\Theta}^\top \cdot X_i}{\hat{\sigma}} \right). \quad (9.22)$$

Мы видим, что, в отличие от *линейных* моделей, измеритель этого эффекта является величиной *переменной*, зависящей от значений предикторов.

Ошибки в спецификации модели (9.18)–(9.19) и, в частности, нарушение нормальности и/или гомоскедастичности остатков ε_i в общем случае может приводить к несостоятельности оценок (9.21). Статистические тесты на выявление этих нарушений и некоторые приемы, позволяющие избежать несостоятельности оценок параметров тобит-модели, описаны, например, в книгах [Вербик (2008)], [Greene (2000)], [Johnston, Di Nardo (1997)].

Обобщения тобит-модели и проблема выборочной селективности. Рассмотренная выше тобит-модель устроена таким образом, что и доля нулевых значений y , и зависимость положительных значений y от объясняющих переменных определяются *одними и теми же предикторами* X . Это существенно «сковывает» модель, делая ее иногда противоречащей некоторым соображениям реальности. Кроме того, бывают ситуации, когда ограничения на процесс формирования выборки (*не заданные априори и регулируемые, а случайные, спонтанные*) не позволяют получать репрезентативную (случайную) выборку из анализируемой генеральной совокупности, например, как бы искусственно

(«принудительно») заставляя принимать элементы определенной ее доли некоторое дискретное значение. Это создает дополнительную проблему в обеспечении состоятельных и несмещенных статистических выводов, которую принято называть **«проблемой выборочной селективности»** (“Sample Selection Problem”). Широкий класс обобщений тобит-модели мотивирован именно этими двумя обстоятельствами (см., например, [Вербик (2008), пп. 7.4 и 7.5], [Greene (2000), пп. 20.3 и 20.4], [Heckman (1990)], [Manski (1989, 1995)], [Amemiya (1984)]). Все это — важные и интересные вопросы, однако в рамках данного издания их анализ не предусмотрен.

Выводы

1. Широкий диапазон практических задач эконометрического анализа приводит к необходимости строить и статистически оценивать модели зависимости *дискретной* (в том числе — *номинальной*) и *дискретно-непрерывной результирующей переменной* y от набора объясняющих переменных $X = (1, x^{(1)}, x^{(2)}, \dots, x^{(p)})$. Это мотивировало разработку специального класса регрессионных моделей — *моделей бинарного и множественного выбора* (в том числе, — логит- и пробит-моделей), *моделей с дискретно-непрерывной зависимой переменной* (тобит-моделей) и их обобщений, учитывающих проблему, возникающую с ограничениями на процесс формирования выборки (проблему «селективности выборки»).

2. **Модели бинарного выбора** (логит- и пробит-модели) позволяют строить и статистически оценивать зависимость вероятности данного (одного из двух) состояния анализируемого объекта от набора объясняющих переменных X в форме

$$P(y_i = 1|X_i) = F(\Theta^T X_i)$$

и

$$P(y_i = 2|X_i) = 1 - F(\Theta^T X_i),$$

где $F(z)$ — некоторая функция распределения вероятностей непрерывной случайной величины (в *логит-модели* в качестве $F(z)$ используется логистическая функция распределения, см. (9.1а), а в *пробит-модели* — функция распределения стандартного нормального закона, см. (9.1б)). Оценка параметров $\Theta = (\theta_0, \theta_1, \dots, \theta_p)^T$ моделей производится с помощью метода максимального правдоподобия.

3. **Модели множественного выбора** обобщают модели бинарного выбора на случай *нескольких* (более чем двух) возможных значений дискретной зависимой переменной y и, соответственно, позволяют строить и статистически оценивать зависимости вида

$$P(y_i = j|X_i) = \varphi_j(\Theta^T(j) \cdot X_i).$$

При построении моделей бинарного и множественного выбора используется *идея введения полезности* каждого из анализируемых состояний зависимой переменной и соответствующей *латентной* переменной y^* , связанной с *наблюдаемой* переменной y множественными соотношениями типа (9.4)–(9.5) и (9.9).

4. Существует тесная связь между задачами и моделями бинарного и множественного выбора с одной стороны и дискриминантного анализа, — с другой (см. Утверждения 1 и 2 в п. 9.3). При этом подход, основанный на методах дискриминантного анализа, может оказаться, при решении практических задач определенного класса, более гибким и дееспособным.

5. Специфическая природа *дискретно-непрерывной* зависимой переменной y , которая, имея непрерывную область своих возможных значений, может, тем не менее, принимать одно или несколько дискретных значений (*каждое из них — с некоторой положительной вероятностью*), обуславливает необходимость построения специальных регрессионных моделей зависимости y от X , приспособленных для подобных случаев. Одна из наиболее распространенных в эконометрической практике ситуаций подобного рода: $P(y = 0) = q > 0$, а остальная доля $(1 - q)$ генеральной совокупности анализируемой зависимой переменной y распределена непрерывно на положительной числовой оси или на ограниченном ее отрезке. В этом случае используют так называемую *тобит-модель*, тоже основанную на введении латентной переменной y^* (см. (9.18)–(9.19)).

6. В эконометрической практике бывают ситуации, когда дискретно-непрерывная природа анализируемой зависимой переменной y образуется, в том числе, под воздействием некоторых ограничений на процесс формирования выборки (то есть при некоторой *выборочной селективности*). Это обстоятельство явилось существенной мотивацией для разработки ряда обобщений *тобит-модели* (тобит-2 и др.).

Глава 10

Анализ одномерных временных рядов (модели и прогнозирование)

Всякий эконометрический анализ основывается на *исходных статистических данных*. Их общая форма в широком классе случаев может быть представлена в виде последовательности $n \times p$ -матриц

$$X(t_k) = \begin{pmatrix} x_1^{(1)}(t_k) & x_1^{(2)}(t_k) & \dots & x_1^{(p)}(t_k) \\ x_2^{(1)}(t_k) & x_2^{(2)}(t_k) & \dots & x_2^{(p)}(t_k) \\ \vdots & \vdots & \dots & \vdots \\ x_n^{(1)}(t_k) & x_n^{(2)}(t_k) & \dots & x_n^{(p)}(t_k) \end{pmatrix}, \quad k = 1, 2, \dots, N,$$

где $x_i^{(j)}(t_k)$ — значение j -го анализируемого показателя, характеризующего состояние i -го статистически обследованного объекта в момент времени t_k ($j = 1, 2, \dots, p$; $i = 1, 2, \dots, n$; $k = 1, 2, \dots, N$). Если зафиксировать номер переменной j и номер статистически обследуемого объекта i , то расположенную в хронологическом порядке последовательность значений

$$x_i^{(j)}(t_1), x_i^{(j)}(t_2), \dots, x_i^{(j)}(t_k) \quad (10.1)$$

называют *одномерным временным рядом*. Если же одновременно рассматривать p одномерных временных рядов вида (10.1), то есть исследовать закономерности во *взаимосвязанном* поведении временных рядов (10.1) для $j = 1, 2, \dots, p$, характеризующих динамику p переменных, *измеренных на каком-то одном (i -м) объекте*, то тогда говорят о *статистическом анализе многомерного временного ряда* $X(t) = (x^{(1)}(t_k), x^{(2)}(t_k), \dots, x^{(p)}(t_k))^T$ ($k = 1, 2, \dots, N$). По существу, все задачи, связанные с анализом экономической динамики, предусматривают использо-

вание в качестве своей статистической базы временных рядов тех или иных показателей.

Данная глава практически полностью посвящена *методам построения, идентификации (то есть — статистического оценивания параметров) и верификации (то есть — статистической проверки адекватности) моделей одномерных временных рядов*. При этом с точки зрения прикладного назначения этих моделей нас будет интересовать прежде всего *проблема прогнозирования* экономических показателей. Это значит, что вне рамок данной главы остаются такие важные (в первую очередь, в области систем управления технологическими процессами) прикладные проблемы, как анализ и построение так называемой *передаточной функции* систем или *проектирование регулирующих динамических систем с прямой и обратной связями* (см., например, [Бокс, Дженкинс (1974)]). Добавим к этому, что в данной главе будут рассматриваться лишь *дискретные (по времени наблюдения)* одномерные временные ряды для *равноотстоящих моментов наблюдения*, то есть $t_2 - t_1 = t_3 - t_2 = \dots = t_N - t_{N-1} = \Delta$, где Δ — заданный временной такт (минута, час, сутки, неделя, месяц, квартал, год и т. п.). Поэтому в дальнейшем исследуемый временной ряд нам будет удобнее представлять в виде

$$x(1), x(2), \dots, x(N), \quad (10.1')$$

где $x(t)$ — значение анализируемого показателя, зарегистрированное в t -м такте времени ($t = 1, 2, \dots, N$).

Говоря о проблеме прогнозирования, мы имеем в виду *кратко- и среднесрочный прогноз*, поскольку построение *долгосрочного* прогноза подразумевает обязательное использование методов организации и статистического анализа *специальных экспертных оценок*. Тем не менее, использование доступных к моменту времени $t = N$ наблюдений временного ряда (10.1') для прогнозирования значения $x(t)$ на один или несколько временных тактов вперед (то есть — для прогнозной оценки значений $\hat{x}(N + l)$, $l = 1, 2, \dots, \tau$) может явиться основой для:

- планирования в экономике, производстве, торговле;
- управления и оптимизации протекающих в обществе социально-экономических процессов;
- частичного управления важными параметрами демографических процессов и экологической ниши общества;
- принятия оптимальных решений в бизнесе.

Отметим, что анализ и моделирование данных, представленных временными рядами (10.1'), потребует определенной ревизии принятого до

сих пор в данном учебнике понимания как *механизмов генерации исходных статистических данных*, так и того, что мы называем *выборкой из генеральной совокупности*.

10.1 Временной ряд: определения, примеры, формулировка основных задач

Выше, говоря о *расположенной в хронологическом порядке последовательности наблюдаемых значений* (10.1) или (10.1') *какого-либо признака*, мы, по существу, уже дали определение временного ряда. Однако в дальнейшем нас будет интересовать лишь некоторый подкласс подобных последовательностей, а именно тот, который связан с наблюдениями *стохастических* по своей природе признаков. Другими словами, мы исключаем из рассмотрения *детерминированные* схемы динамических наблюдений, при которых элементы последовательности (10.1') могут быть в точности вычислены как значения некоторой *нестатистической функции* $f(t)$, то есть $x(t) = f(t)$. Поэтому уточним понятие временного ряда, принятое в данной главе.

О п р е д е л е н и е 10.1. *Ряд наблюдений $x(t_1), x(t_2), \dots, x(t_N)$ анализируемой случайной величины $\xi(t)$, произведенных в последовательные моменты времени $t_1, t_2, \dots, t_N$, называется временным рядом.*

Как уже было отмечено, предметом нашего анализа в данной главе будут *временные ряды с равноотстоящими моментами наблюдений*. А это позволяет представлять их в форме (10.1').

Определение 10.1 опирается на понятие случайной величины $\xi(t)$, зависящей от параметра t , интерпретируемого как время. То есть, по существу, речь идет об *однопараметрическом семействе случайных величин* $\{\xi(t)\}$. Это значит, что закон распределения вероятностей (з.р.в.) этих случайных величин, и в частности, их первые и вторые моменты, также, вообще говоря, могут зависеть от времени t .

В чем же состоят принципиальные отличия временного ряда (10.1') от последовательности наблюдений $x_1, x_2, \dots, x_n$, образующих случайную выборку? Этих отличий два.

(а) Во-первых, в отличие от элементов случайной выборки *члены временного ряда не являются статистически независимыми*.

(б) Во-вторых, члены временного ряда *не являются одинаково распределенными*, то есть $P\{x(t_1) < x\} \neq P\{x(t_2) < x\}$ при $t_1 \neq t_2$.

Это значит, что мы не можем распространять свойства и правила статистического анализа случайной выборки на временные ряды. С другой стороны, взаимозависимость членов временного ряда создает свою специфическую базу для построения прогнозных значений анализируемого показателя (то есть для построения оценок $\hat{x}(N + k)$ для неизвестных

значений $x(N + k)$) по наблюдаемым значениям $x(1), x(2), \dots, x(N)$.

Генезис наблюдений, образующих временной ряд. Речь идет о структуре и классификации основных факторов, под воздействием которых формируются значения элементов временного ряда. Целесообразно выделить следующие 4 типа таких факторов.

(А) *Долговременные*, формирующие общую (в длительной перспективе) тенденцию в изменении анализируемого признака $x(t)$. Обычно эта тенденция описывается с помощью той или иной неслучайной функции $f_{\text{тр}}(t)$, как правило, монотонной. Эту функцию называют *функцией тренда* или просто — *трендом*.

(Б) *Сезонные*, формирующие периодически повторяющиеся в определенное время года колебания анализируемого признака. Условимся обозначать результат действия сезонных факторов с помощью неслучайной функции $\varphi(t)$. Поскольку эта функция должна быть *периодической* (с периодами, кратными «сезонам»), в ее аналитическом выражении участвуют гармоники (тригонометрические функции), периодичность которых, как правило, обусловлена содержательной сущностью задачи.

(В) *Циклические (конъюнктурные)*, формирующие изменения анализируемого признака, обусловленные действием долговременных циклов экономической, демографической или астрофизической природы (волны Кондратьева, демографические «ямы», циклы солнечной активности и т. п.). Результат действия циклических факторов будем обозначать с помощью неслучайной функции $\psi(t)$.

(Г) *Случайные (нерегулярные)*, не поддающиеся учету и регистрации. Их воздействие на формирование значений временного ряда как раз и обуславливает *стохастическую природу* элементов $x(t)$, а следовательно, и необходимость интерпретации $x(1), x(2), \dots, x(N)$ как наблюдений, произведенных над случайными величинами, соответственно, $\xi(1), \xi(2), \dots, \xi(N)$. Будем обозначать результат воздействия случайных факторов с помощью случайных величин («остатков», «ошибок») $\varepsilon(t)$ ¹.

Конечно, вовсе не обязательно, чтобы в процессе формирования значений всякого временного ряда участвовали одновременно факторы *всех* четырех типов. Из приведенных ниже примеров мы увидим, что в одних случаях значения временного ряда формируются под воздействием фак-

¹Случайные факторы, в свою очередь, могут быть двойкой природы: *внезапными* («разладочными»), приводящими к скачкообразным структурным изменениям в механизме формирования значений $x(t)$ (что выражается, например, в радикальных скачкообразных изменениях основных структурных характеристик функций $f_{\text{тр}}(t)$, $\varphi(t)$ и $\psi(t)$ анализируемого временного ряда в случайный момент времени), и *эволюционными остаточными*, обуславливающими относительно небольшие случайные отклонения значений $x(t)$ от тех, которые должны были бы получиться только под воздействием факторов (А), (Б) и (В). Однако в данном учебнике будут рассмотрены схемы формирования временных рядов, включающие в себя действие только эволюционных остаточных случайных факторов.

торов (А), (В) и (Г) (см. пример 10.1), в других — под воздействием факторов (А), (В) и (Г) (см. пример 10.2) и, наконец, — исключительно под воздействием одних только случайных факторов (Г) (см. пример 10.3). Однако во всех случаях *предполагается непрерывное участие случайных (эволюционных) факторов (Г)*. Кроме того, в дальнейшем везде (если специально не оговорено противное) мы будем следовать *аддитивной структурной схеме* влияния факторов А, Б, В и Г на формирование значений $x(t)$, которая означает правомерность представления значений членов временного ряда в виде разложения:

$$x(t) = \chi(A) f_{\text{тр}}(t) + \chi(B) \varphi(t) + \chi(V) \psi(t) + \varepsilon(t), \quad t = 1, 2, \dots, N, \quad (10.2)$$

где

$$\chi(C) = \begin{cases} 1, & \text{если факторы типа } C \text{ участвуют в формировании} \\ & \text{значений } x(t); \\ 0 & \text{в противном случае;} \end{cases}$$

$C = A, B \text{ или } V.$

Выводы о том, участвуют или нет факторы данного типа в формировании значений $x(t)$, могут базироваться как на анализе содержательной сущности задачи (то есть быть *априорно-экспертными по своей природе*), так и на специальном *статистическом анализе исследуемого временного ряда*.

Примеры временных рядов. Рассмотрим несколько конкретных временных рядов. Они иллюстрируют различные варианты вышеописанной схемы генезиса исходных статистических данных, имеющих динамическую природу. Одновременно эти данные понадобятся нам для иллюстрации смысла и работоспособности описанных далее методов и моделей анализа временных рядов.

Пример 10.1. В таблице 10.1 и на рис. 10.1 приведены данные о суммарных месячных расстояниях $x(t)$ (в тысячах миль), пройденных британскими авиалайнерами за 96 месяцев с января 1963 г. по декабрь 1970 г. (то есть $t = 1, 2, \dots, 96$; временной такт Δ равен одному месяцу). Данные заимствованы из [Кендэл (1981), с. 18].

Таблица 10.1. Расстояния, пройденные британскими виалайнерами за месяц (тыс. миль)

	1963	1964	1965	1966	1967	1968	1969	1970
Январь	6827	7269	8350	8186	8334	8639	9491	10840
Февраль	6178	6775	7829	7444	7899	8772	8919	10436
Март	7084	7819	8829	8484	9994	10894	11607	13589
Апрель	8162	8371	9948	9864	10078	10455	8852	13402
Май	8462	9069	10638	10252	10801	11179	12537	13103
Июнь	9644	10248	11253	12282	12950	10588	14759	14933
Июль	10466	11030	11424	11637	12222	10794	13667	14147
Август	10748	10882	11391	11577	12246	12770	13731	14057
Сентябрь	9963	10333	10665	12417	13281	13812	15110	16234
Октябрь	8194	9109	9396	9637	10366	10857	12185	12389
Ноябрь	6848	7685	7775	8094	8730	9290	10645	11595
Декабрь	7027	7602	7933	9280	9614	10925	12161	12772

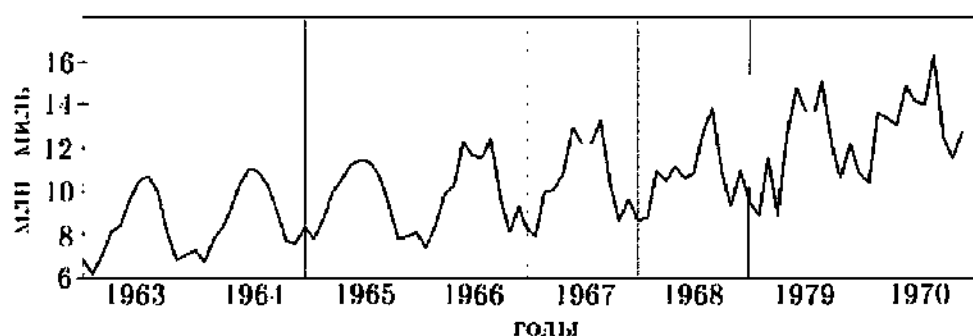


Рис. 10.1. График данных из таблицы 10.1 (расстояния, пройденные авиалайнерами Соединенного Королевства за месяц)

Пример 10.2. В таблице 10.2 и на рис. 10.2 представлены данные Financial Times (FT) о квартальной динамике среднего индекса курса акций ведущих компаний на Лондонской бирже за 1960–1971 гг. (соответственно: временной такт Δ равен одному кварталу, а общее число членов временного ряда $N = 48$). Данные заимствованы из [Кендэл (1981), с. 20].

Таблица 10.2. Индекс FT курса обыкновенных акций ведущих компаний: квартальные средние за 1960–1971 гг.

Год и квартал	Индекс	Год и квартал	Индекс	Год и квартал	Индекс
1960 1	323,8	1964 1	335,1	1968 1	409,1
2	314,1	2	344,4	2	461,1
3	321,0	3	360,9	3	491,4
4	312,9	4	346,5	4	490,5
1961 1	323,7	1965 1	340,6	1969 1	491,0
2	349,3	2	340,3	2	433,0
3	310,4	3	323,3	3	378,0
4	295,8	4	345,6	4	382,6
1962 1	301,2	1966 1	349,3	1970 1	403,4
2	285,8	2	359,7	2	354,7
3	271,7	3	320,0	3	343,0
4	283,6	4	299,9	4	345,4
1963 1	295,7	1967 1	318,5	1971 1	330,4
2	309,3	2	343,1	2	372,8
3	295,7	3	360,8	3	409,2
4	342,0	4	397,8	4	427,6



Рис. 10.2. График временного ряда, представленного данными таблицы 10.2 (индекс FT, квартальные средние)

Пример 10.3. В таблице 10.3 и на рис. 10.3 представлены данные об урожае ячменя в Англии и Уэльсе за 56 лет (с 1884 по 1939 гг.). Так что временной такт Δ в данном случае равен одному году, а длина временного ряда $N = 56$.

Таблица 10.3. Урожайность ячменя в Англии и Уэльсе с 1884 по 1939 гг. (в англ. центнерах на акр², данные из «Agricultural Statistics»)

Год	Урожайность	Год	Урожайность	Год	Урожайность
1884	15,2	1903	15,1	1922	14,0
1885	16,9	1904	14,6	1923	14,5
1886	15,3	1905	16,0	1924	15,4
1887	14,9	1906	16,8	1925	15,3
1888	15,7	1907	16,8	1926	16,0
1889	15,1	1908	15,5	1927	16,4
1890	16,7	1909	17,3	1928	17,2
1891	16,3	1910	15,5	1929	17,8
1892	16,5	1911	15,5	1930	14,4
1893	13,3	1912	14,2	1931	15,0
1894	16,5	1913	15,8	1932	16,0
1895	15,0	1914	15,7	1933	16,8
1896	15,9	1915	14,1	1934	16,9
1897	15,5	1916	14,8	1935	16,6
1898	16,9	1917	14,4	1936	16,2
1899	16,4	1918	15,6	1937	14,0
1900	14,9	1919	13,9	1938	18,1
1901	14,5	1920	14,7	1939	17,5
1902	16,6	1921	14,3		

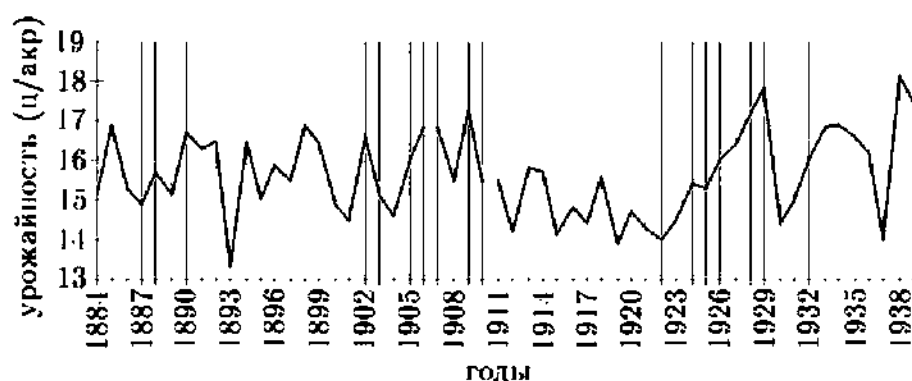


Рис. 10.3. График временного ряда, представленного данными таблицы 10.3 (урожайность ячменя, ц/акр)

Попробуем проанализировать приведенные примеры временных рядов, опираясь на описанную выше общую схему их генезиса и вытекающее из нее разложение вида (10.2).

² Английский центнер равен 50,8 кг, один акр равен 0,405 га.

Данные об авиапассажирах (пример 10.1) представляют собой типичный образец *сезонных колебаний*, наслаивающихся на монотонно *растущий тренд*. Сезонный эффект легко расшифровывается. Мы наблюдаем в течение года три «всплеска» активности пассажирских авиаперевозок, и все они объясняются перелетами в праздничный и отпускной периоды: один из них приходится на Пасху, второй — на лето и третий — на Рождественские праздники. Правда, перелеты на Пасху приходятся, как и сама Пасха, то на одни, то на другие дни года, и картина колебаний меняется из года в год частично из-за возрастающего парка авиапассажиров, а частично из-за увеличения в настоящее время периода праздников.

Иногда требуется особая внимательность, чтобы не спутать *сезонные* циклические компоненты с такими колебаниями, как *экономические* циклы или циклы проявления солнечной активности (последние правильнее отнести к колебаниям псевдоциклического типа). Так, в примере 10.2 улавливается *четырёхлетний экономический цикл*, который, как считают некоторые специалисты, следует объяснять не долговременной внутренней логикой экономического развития, а лишь приблизительно таким же циклом проведения общих парламентских выборов в Великобритании.

Наконец, анализ графика временного ряда, представленного в примере 10.3 (см. рис. 10.3), приводит к гипотезе, что в генерировании этих данных не участвовали ни долговременные (поскольку тренд, скорее всего, отсутствует), ни сезонные, ни циклические факторы (поскольку осциллирование значений $x(t)$ около некоторого постоянного уровня носит, скорее, случайный, а не строго периодический характер).

Основные задачи анализа временных рядов. Отправляясь от приведенного выше *аддитивного разложения* (10.2) временного ряда $x(t)$, можно дать общую формулировку *базисной цели* его статистического анализа³:

- по имеющейся траектории (10.1') анализируемого временного ряда $x(t)$ требуется:
 - 1) определить, какие из неслучайных функций $f_{\text{тр}}(t)$, $\varphi(t)$ и $\psi(t)$ присутствуют в разложении (10.2), то есть определить значения индикаторов $\chi(C)$ ($C = A, B, V$) в разложении (10.2);
 - 2) построить «хорошие» оценки для тех неслучайных функций, которые присутствуют в разложении (10.2);

³Мы отличаем базисную цель анализа от конечных прикладных целей исследования. Без достижения базисной цели невозможна реализация конечных прикладных целей исследования. Другими словами, базисная цель формулируется *внутри* исследования и является необходимым этапом для достижения конечных прикладных целей, результаты реализации которых формулируются «на выходе» исследования.

- 3) подобрать модель, адекватно описывающую поведение «случайных остатков» $\varepsilon(t)$, и статистически оценить параметры этой модели.

Успешное решение задач (1)–(3), обусловленных базисной целью статистического анализа временного ряда, является основой для достижения конечных прикладных целей исследования и, в первую очередь, для решения задачи кратко- и среднесрочного прогноза значений временного ряда.

10.2 Стационарные временные ряды и их основные характеристики

Поиск модели, адекватно описывающей поведение случайных остатков $\varepsilon(t)$ анализируемого временного ряда $x(t)$ (см. выше формулировку задачи (3)), производят, как правило, в рамках некоторого специального класса случайных временных последовательностей — *класса стационарных временных рядов*. На интуитивном уровне стационарность временного ряда мы связываем с требованием, чтобы он имел *постоянное* среднее значение и колебался вокруг этого среднего с *постоянной* дисперсией. В некоторых случаях временные последовательности этого класса могут воспроизводить и поведение *самого анализируемого временного ряда* $x(t)$ (из приведенных выше примеров внешние признаки стационарного временного ряда демонстрирует график динамики урожайности ячменя, см. пример 10.3).

О п р е д е л е н и е 10.2. Ряд $x(t)$ называется **строго стационарным** (или **стационарным в узком смысле**), если совместное распределение вероятностей m наблюдений $x(t_1), x(t_2), \dots, x(t_m)$ такое же, как и для m наблюдений $x(t_1 + \tau), x(t_2 + \tau), \dots, x(t_m + \tau)$, при любых $m, t_1, t_2, \dots, t_m$ и τ .

Другими словами, свойства строго стационарного временного ряда не меняются при изменении начала отсчета времени. В частности, при $m = 1$ из предположения о строгой стационарности временного ряда $x(t)$ следует, что закон распределения вероятностей (з.р.в.) случайной величины $x(t)$ не зависит от t , а значит, не зависят от t и все его основные числовые характеристики, в том числе:

$$\text{среднее значение } E x(t) = a \quad (10.3)$$

и

$$\text{дисперсия } D x(t) = E(x(t) - a)^2 = \sigma^2. \quad (10.4)$$

Очевидно, значение a определяет постоянный уровень, относительно которого флуктуирует анализируемый временной ряд $x(t)$, а постоянная величина σ характеризует размах этой флуктуации. Поскольку з.р.в. случайной величины $x(t)$ *одинаков при всех t* , то он сам и его основные числовые характеристики могут быть оценены по наблюдениям $x(1), x(2), \dots, x(N)$ ⁴. В частности:

$$\hat{a} = \frac{1}{N} \sum_{t=1}^N x(t) - \text{оценка среднего значения}, \quad (10.3')$$

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{t=1}^n (x(t) - \hat{a})^2 - \text{оценка дисперсии}. \quad (10.4')$$

Автоковариационная функция $\gamma(\tau)$. Из предположения о строгой стационарности временного ряда $x(t)$ при $m = 2$ следует, что совместные двумерные распределения для пар случайных величин $(x(t_1), x(t_2))$, $(x(0), x(t_2 - t_1))$, $(x(\tau), x(t_2 - t_1 + \tau))$ совпадают при любых t_1, t_2 и τ и зависят только от разности $t_2 - t_1$. Соответственно, ковариация между значениями $x(t)$ и $x(t \pm \tau)$ будет зависеть только от величины «сдвига по времени» τ (и не будет зависеть от t). Эта ковариация называется *автоковариацией* (поскольку измеряет ковариацию для различных значений *одного и того же* временного ряда $x(t)$) и определяется соотношением:

$$\gamma(\tau) = \text{cov}(x(t), x(t + \tau)) = \mathbf{E}[(x(t) - a)(x(t + \tau) - a)]. \quad (10.5)$$

При анализе изменения величины $\gamma(\tau)$ в зависимости от значения τ принято говорить об *автоковариационной функции $\gamma(\tau)$* . Значения автоковариационной функции могут быть *статистически оценены* по имеющимся наблюдениям временного ряда (10.1') по формуле

$$\hat{\gamma}(\tau) = \frac{1}{N - \tau} \sum_{t=1}^{N-\tau} (x(t) - \hat{a})(x(t + \tau) - \hat{a}), \quad (10.5')$$

где $\tau = 1, 2, \dots, N - 1$, а $\hat{a}$ вычислено по формуле (10.3').

Очевидно, значение автоковариационной функции при $\tau = 0$ есть не что иное, как дисперсия временного ряда, то есть

$$\gamma(0) = \sigma^2 = \mathbf{E}(x(t) - a)^2$$

⁴В общем случае (то есть при отсутствии свойства стационарности) для оценки з.р.в. случайной величины $x(t)$ и ее основных числовых характеристик мы должны были бы иметь *по несколько* наблюдений временного ряда для каждого фиксированного «момента времени» t . Однако неизменность з.р.в. $x(t)$ во времени t позволяет (при некоторых дополнительных условиях *эргодичности*, см. теорему Биркгофа-Хинчина в книге [Кендалл, Стьюарт, (1976)]), оценивать среднее значение, дисперсию и ковариации стационарного временного ряда по его *единственной реализации* (10.1').

и, соответственно,

$$\hat{\gamma}(0) = \hat{\sigma}^2 = \frac{1}{N} \sum_{t=1}^N (x(t) - \hat{a})^2. \quad (10.4'')$$

Автокорреляционная функция $r(\tau)$. Одно из главных отличий последовательности наблюдений, образующих временной ряд, от *случайной выборки* заключается, как мы видели (см. свойство (а) в начале п. 10.1), в том, что члены временного ряда не являются, вообще говоря, статистически взаимозависимыми. Степень тесноты статистической связи между двумя случайными величинами может быть измерена *парным коэффициентом корреляции* (см. главу 3). Так что степень тесноты статистической связи между наблюдениями стационарного временного ряда, разнесенными (по времени) на τ единиц, определится величиной коэффициента корреляции

$$r(\tau) = \frac{\mathbf{E}[(x(t) - a)(x(t + \tau) - a)]}{[\mathbf{E}(x(t) - a)^2 \cdot \mathbf{E}(x(t + \tau) - a)^2]^{\frac{1}{2}}} = \frac{\gamma(\tau)}{\gamma(0)}, \quad (10.6)$$

поскольку $\mathbf{E}(x(t) - a)^2 = \mathbf{E}(x(t + \tau) - a)^2 = \gamma(0)$.

Коэффициент $r(\tau)$ измеряет корреляцию, существующую между членами *одного и того же* временного ряда, поэтому его принято называть коэффициентом *автокорреляции*. При анализе изменения величины $r(\tau)$ в зависимости от значения τ принято говорить об *автокорреляционной функции* $r(\tau)$. График автокорреляционной функции иногда называют *коррелограммой*. Заметим, что автокорреляционная функция (в отличие от автоковариационной) *безразмерна*, то есть не зависит от масштаба измерения анализируемого временного ряда. Ее значения, по определению, могут колебаться от -1 до $+1$. Кроме того, из стационарности следует, что $r(\tau) = r(-\tau)$, так что при анализе поведения автокорреляционных функций ограничиваются рассмотрением *только положительных значений τ* .

Выборочный аналог автокорреляционной функции (ее статистическая оценка $\hat{r}(\tau)$) определяется формулой

$$\hat{r}(\tau) = \frac{\frac{1}{N-\tau} \sum_{t=1}^{N-\tau} (x(t) - \hat{a})(x(t + \tau) - \hat{a})}{\frac{1}{N} \sum_{t=1}^N (x(t) - \hat{a})^2} = \frac{\hat{\gamma}(\tau)}{\hat{\gamma}(0)}, \quad (10.6')$$

$$\tau = 1, 2, \dots, N - 1.$$

Существуют ли какие-либо общие характерные особенности, отличающие поведение автокорреляционной функции стационарного временного ряда? Другими словами, можно ли описать в общих чертах схематичный вид коррелограммы стационарного временного ряда? Ответ —

утвердительный, и это обусловлено следующим общим соображением: очевидно, чем больше разнесены во времени члены временного ряда $x(t)$ и $x(t + \tau)$ (то есть чем больше величина сдвига τ), тем слабее взаимосвязь этих членов и, соответственно, тем меньше должно быть по абсолютной величине значение $r(\tau)$. При этом в ряде случаев существует такое пороговое значение τ_0 , начиная с которого все значения $r(\tau)$ будут тождественно равны нулю (то есть $r(\tau) \equiv 0$ для всех $\tau \geq \tau_0$).

В левых частях рис. 10.4, 10.5 и 10.6 приведены графики автокорреляционных функций для ряда моделей временных рядов, рассмотренных ниже (см. пп. 10.4–10.7). Обращаем внимание читателя на тот факт, что в отличие от теоретических автокорреляционных функций их выборочные (эмпирические) аналоги (10.6') часто допускают нарушения свойства монотонного убывания (по абсолютной величине) при возрастании аргумента τ .

Частная автокорреляционная функция $r_{\text{част}}(\tau)$. С помощью этой функции реализуется идея измерения автокорреляции, существующей между разделенными τ тактами времени членами временного ряда $x(t)$ и $x(t + \tau)$, *при устраненном опосредованном влиянии на эту взаимозависимость всех промежуточных* (то есть расположенных между $x(t)$ и $x(t + \tau)$) *членов этого временного ряда*. Идея «очищенной» от опосредованного влияния, то есть *частной*, корреляции уже обсуждалась в этой книге (см. п. 3.2.4). Поэтому нам остается лишь применить понятия, результаты и формулы этого раздела корреляционного анализа к последовательности членов временного ряда. Так, частная автокорреляция 1-го порядка может быть подсчитана с использованием соотношения (3.26) из главы 3:

$$\begin{aligned} r_{\text{част}}(2) &= r(x(t), x(t+2) \mid x(t+1) = a) \\ &= \frac{r(x(t), x(t+2)) - r(x(t), x(t+1))r(x(t+2), x(t+1))}{\sqrt{[1 - r^2(x(t), x(t+1))][1 - r^2(x(t+2), x(t+1))]} \quad (10.7) \\ &= \frac{r(2) - r^2(1)}{1 - r^2(1)}, \end{aligned}$$

где a — среднее значение анализируемого стационарного процесса.

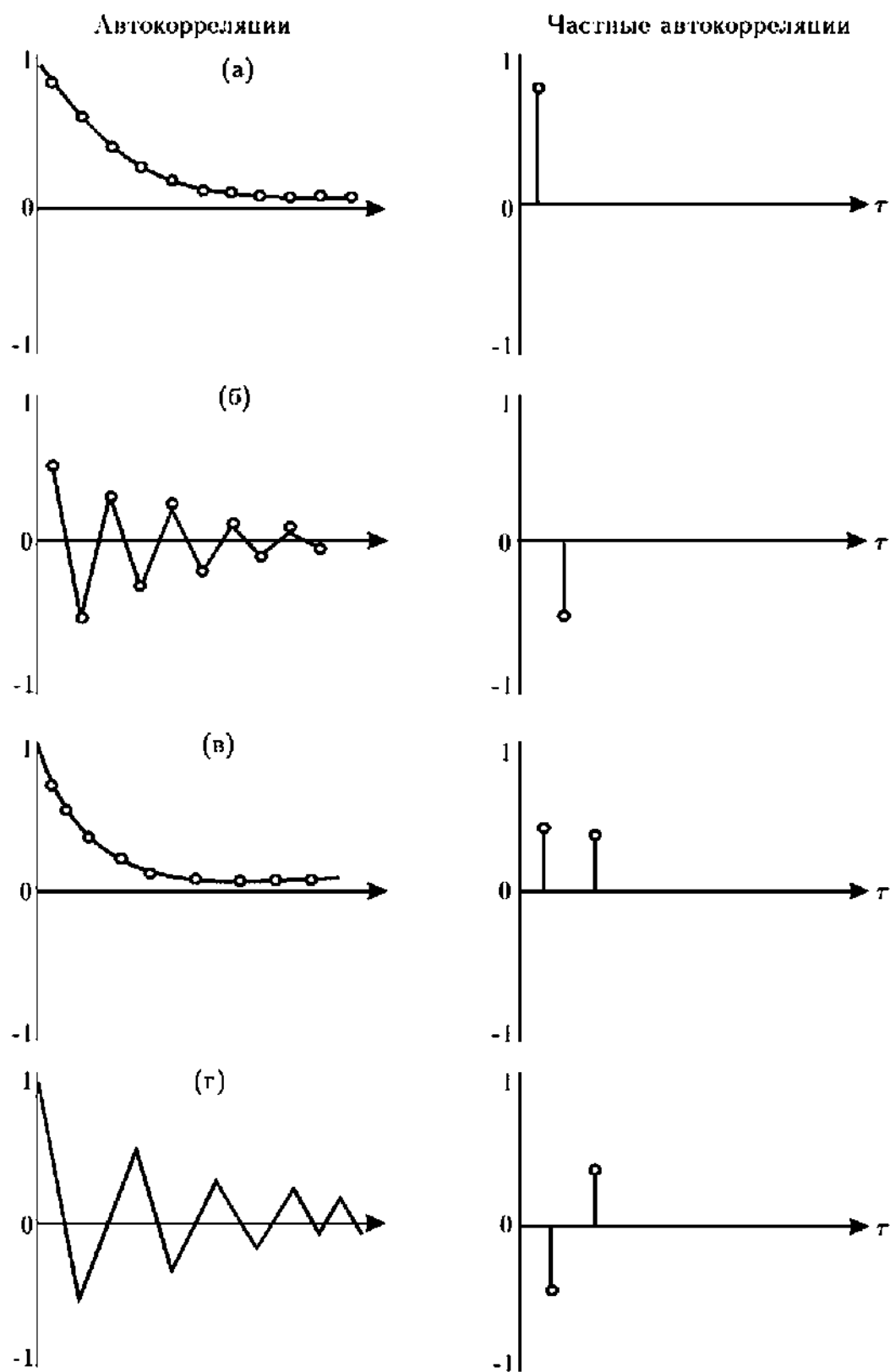


Рис. 10.4. Схематичные графики автокорреляционной и частной автокорреляционной функций для моделей авторегрессии 1-го ((а) и (б)) и 2-го ((в) и (г)) порядков

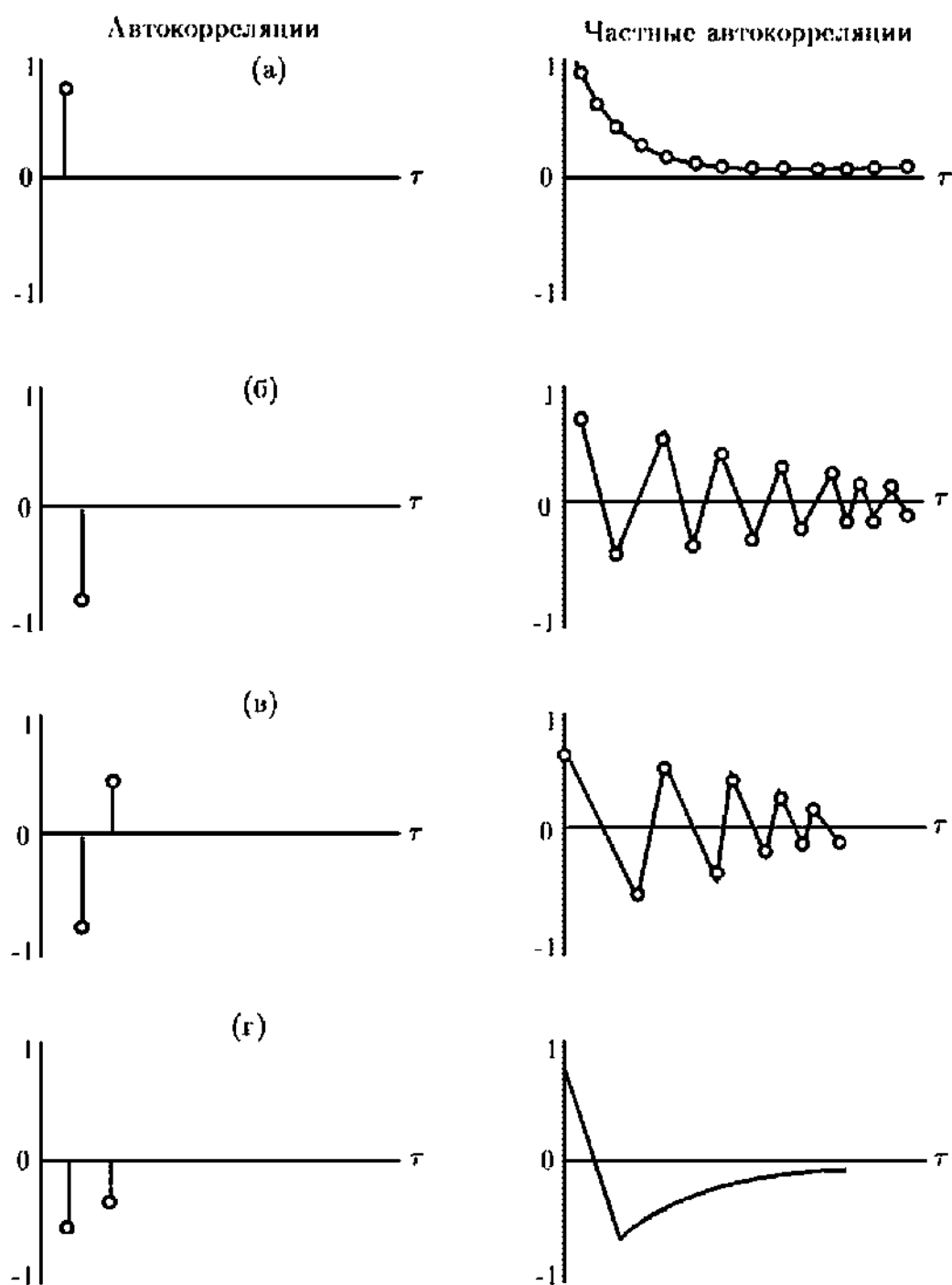


Рис. 10.5. Схематичные графики автокорреляционной и частной автокорреляционной функции для моделей скользящего среднего 1-го ((а) и (б)) и 2-го ((в) и (г)) порядков

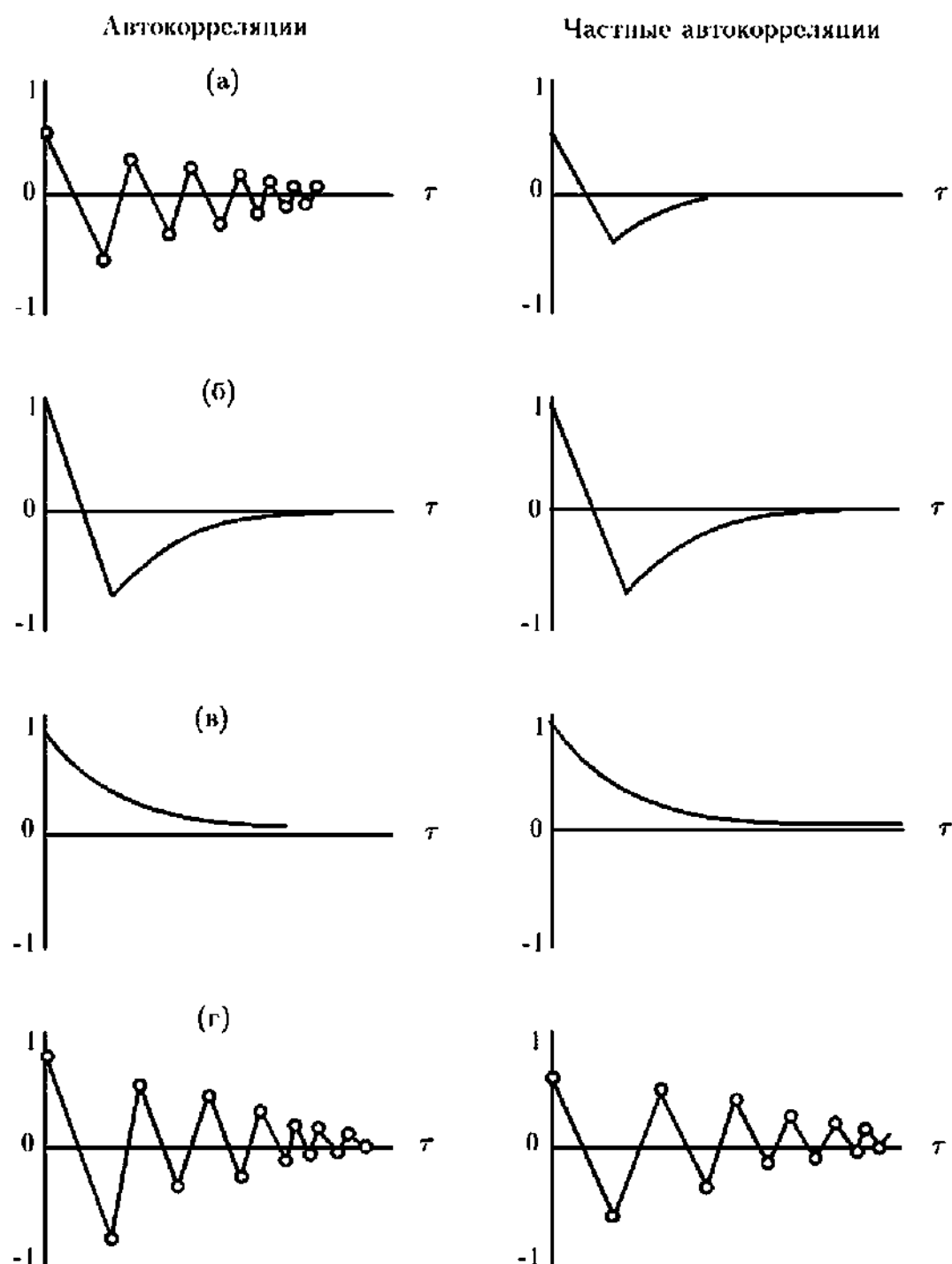


Рис. 10.6. Схематичные графики автокорреляционной и частной автокорреляционной функций для моделей вида $x(t) = \alpha x(t-1) + \delta(t) - \theta \delta(t-1)$ при различных сочетаниях знаков коэффициентов α и θ

Частные автокорреляции более высоких порядков могут быть подсчитаны аналогичным образом с использованием либо рекуррентной формулы (3.26'), либо с помощью формулы (3.25) по элементам общей корреляционной $(N \times N)$ — матрицы $\mathbf{R}$, в которой $r_{ij} = r(x(i), x(j)) = r(|i-j|)$, где $i, j = 1, 2, \dots, N$ (и, очевидно, $r(0) = 1$). Так, например, частная

автокорреляция *2-го порядка* (то есть корреляция между членами временного ряда, разделенными тремя тактами времени, подсчитанная при условии, что значения *двух* промежуточных членов временного ряда зафиксированы на среднем уровне) может быть определена по формуле (см. (3.26')):

$$r_{\text{част}}(3) = r(x(t), x(t+3) \mid x(t+1) = x(t+2) = a) = \frac{r_{03.1} - r_{02.1}r_{32.1}}{\sqrt{(1 - r_{02.1}^2)(1 - r_{32.1}^2)}}, \quad (10.8)$$

где

$$\begin{aligned} r_{03.1} &= r(x(t), x(t+3) \mid x(t+1) = a), \\ r_{02.1} &= r(x(t), x(t+2) \mid x(t+1) = a), \\ r_{32.1} &= r(x(t+3), x(t+2) \mid x(t+1) = a) \end{aligned}$$

и в соответствии с формулой (3.26), имеем:

$$\begin{aligned} r_{03.1} &= r(x(t), x(t+3) \mid x(t+1) = a) = \frac{r(3) - r(1)r(2)}{\sqrt{(1 - r^2(1))(1 - r^2(2))}}, \\ r_{02.1} &= r(x(t), x(t+2) \mid x(t+1) = a) = \frac{r(2) - r^2(1)}{1 - r^2(1)}, \\ r_{32.1} &= r(x(t+3), x(t+2) \mid x(t+1) = a) = \frac{r(1) - r(2)r(1)}{\sqrt{(1 - r^2(2))(1 - r^2(1))}}. \end{aligned}$$

Эмпирические (выборочные) версии частных автокорреляционных функций получаются с помощью тех же соотношений (10.7), (10.8) при замене участвующих в них теоретических значений автокорреляций $r(\tau)$ их статистическими оценками $\hat{r}(\tau)$ (см. формулу (10.6')).

Полученные таким образом частные автокорреляции $\hat{r}_{\text{част}}(1)$, $\hat{r}_{\text{част}}(2)$, $\hat{r}_{\text{част}}(3)$, ... можно нанести на график, в котором роль абсциссы выполняет величина сдвига τ .

Знание автокорреляционных функций $\hat{r}(\tau)$ и $\hat{r}_{\text{част}}(\tau)$ оказывает существенную помощь в решении задачи *подбора и идентификации модели анализируемого временного ряда*. В правых частях рис. 10.4, 10.5 и 10.6 приведены схематические графики частных автокорреляционных функций для рассмотренных ниже (см. пп. 10.4–10.7) конкретных моделей временных рядов. Заметим, что для тех временных рядов, которые укладываются в рамки той или иной конкретной модели (например, в рамки *модели авторегрессии*, см. п. 10.4), можно избежать прямого вычисления частных корреляционных функций по формулам (10.7)–(10.8), воспользовавшись более простой процедурой.

Спектральная плотность $p(\omega)$. Определим спектральную плотность стационарного временного ряда $x(t)$ через его автокорреляцион-

ную функцию $r(\tau)$ соотношением вида

$$p(\omega) = \sum_{\tau=-\infty}^{\infty} r(\tau) e^{i\tau\omega}, \quad (10.9)$$

где $i = \sqrt{-1}$ (мнимая единица). Благодаря тому, что $r(\tau) = r(-\tau)$, спектральная плотность $p(\omega)$ может быть записана в виде

$$p(\omega) = 1 + 2 \sum_{\tau=1}^{\infty} r(\tau) \cos(\tau\omega). \quad (10.9')$$

Мы видим, что функция (10.9') является *гармонической* и имеет период $2\pi^5$. Поскольку $\cos[(2\pi - \omega)\tau] = \cos(\omega\tau)$, то график спектральной плотности $p(\omega)$, называемый *спектром*, симметричен относительно $\omega = \pi$. Поэтому при анализе поведения функции $p(\omega)$ ограничиваются значениями ω , лежащими между 0 и π . Ниже мы убедимся в том, что спектральная плотность $p(\omega)$ может принимать *только неотрицательные значения*.

Использование свойств этой функции в прикладном анализе временных рядов определяется как «*спектральный анализ временных рядов*». Весьма полное описание этого подхода читатель может найти, например, в книге: Г. Дженикинс, Д. Ватс. Спектральный анализ и его приложения. М.: Мир, вып. 1 (1971); вып. 2 (1972). Применительно к статистическому анализу *экономических* рядов динамики этот подход не получил широкого распространения, так как *эмпирический* (выборочный) анализ спектральной плотности требует в качестве своей информационной базы либо *достаточно длинных стационарных* временных рядов, либо по *несколько* траекторий анализируемого временного ряда (и та, и другая ситуация весьма редки в практике статистического анализа рядов экономической динамики). Поэтому мы ограничимся здесь обсуждением лишь следующих двух важных свойств спектральной плотности.

1) *Выражение автокорреляций через спектральную плотность $p(\omega)$* . Из общей теории преобразований Фурье непосредственно следует

$$r(k) = \frac{1}{\pi} \int_0^{\pi} p(\omega) e^{-ik\omega} d\omega = \frac{1}{\pi} \int_0^{\pi} \cos(k\omega) p(\omega) d\omega. \quad (10.10)$$

Этот результат получается, в частности, из соотношения (10.9') после его домножения на $\cos(k\omega)$ и почленного интегрирования получившегося

⁵Аргумент ω измеряется в радианах на единицу времени и называется *угловой частотой* (так как значения $\cos(\omega t)$ повторяются с периодом $2\pi/\omega$, то число циклов этой функции в единице времени равно $\omega/2\pi$). Соответственно, период $2\pi/\omega$ имеет размерность времени t и называется *длиной волны*. Мы увидим ниже, что наличие или отсутствие гармонических членов в разложении (10.2) анализируемого временного ряда $x(t)$ по-своему отражается на поведении спектральной плотности $p(\omega)$.

выражения по ω от 0 до π , с учетом того, что

$$\int_0^{\pi} \cos(\tau\omega) \cos(k\omega) d\omega = \begin{cases} 0 & \text{при } \tau \neq k, \\ \pi/2 & \text{при } \tau = k. \end{cases}$$

2) *Спектральная плотность $p(\omega)$ как индикатор наличия гармонических составляющих в представлении анализируемого временного ряда $x(t)$.* Рассмотрим взаимосвязь наблюдаемого ряда (10.1') (измеренного относительно его среднего, то есть мы полагаем, что $E x(t) = 0$) с гармоническим членом, имеющим период $2\pi/\omega$. Пусть

$$a(\omega) = \frac{1}{\sqrt{\pi N}} \sum_{t=1}^N x(t) \cos(\omega t),$$

$$b(\omega) = \frac{1}{\sqrt{\pi N}} \sum_{t=1}^N x(t) \sin(\omega t).$$

Рассмотрим функцию $I(\omega) = a^2(\omega) + b^2(\omega)$, называемую *интенсивностью*. Заметим, что поскольку $a(\omega)$ и $b(\omega)$ прямопропорциональны величинам коэффициентов корреляции наблюдаемого временного ряда $x(t)$ с гармониками, соответственно, $\cos(\omega t)$ и $\sin(\omega t)$, то интенсивность $I(\omega)$ можно рассматривать как характеристику степени тесноты связи между $x(t)$ и гармоническим членом, имеющим период $2\pi/\omega$. В то же время функция $I(\omega)$ может быть представлена в виде:

$$\begin{aligned} I(\omega) &= \frac{1}{\pi N} \left[\left(\sum_{t=1}^N x(t) \cos(\omega t) \right)^2 + \left(\sum_{t=1}^N x(t) \sin(\omega t) \right)^2 \right] = \\ &= \frac{1}{\pi N} \left[\sum_{t=1}^N x^2(t) + 2 \sum_{k=1}^{N-1} \sum_{t=1}^{N-k} (\cos(\omega t) \cos[\omega(t+k)] + \right. \\ &\quad \left. + \sin(\omega t) \sin[\omega(t+k)]) x(t) x(t+k) \right] = \\ &= \frac{1}{\pi N} \left[\sum_{t=1}^N x^2(t) + 2 \sum_{k=1}^{N-1} \sum_{t=1}^{N-k} x(t) x(t+k) \cos(k\omega) \right] = \\ &= \frac{\hat{\sigma}^2}{\pi} \left[1 + 2 \sum_{k=1}^{N-1} \hat{r}'(k) \cos(k\omega) \right], \end{aligned}$$

где $\hat{\sigma}^2 = \sum_{t=1}^N x^2(t)/N$ — оцененная дисперсия анализируемого ряда, а $\hat{r}'(k) = \hat{r}(k)(N-k)/N$ — несколько «испорченная» оценка автокорреля-

ции $r(k)$. Однако, устремив $N \rightarrow \infty$ и переходя к пределу, мы получим:

$$\mathbf{E}(I(\omega)) = \frac{\sigma^2}{\pi} \left[1 + 2 \sum_{k=1}^{\infty} r(k) \cos(k\omega) \right],$$

что, как это следует из (10.9'), эквивалентно

$$\mathbf{E}(I(\omega)) = \frac{\sigma^2}{\pi} p(\omega) \quad (10.11)$$

(откуда, в частности, следует неотрицательность всех значений спектральной плотности, поскольку усредняемая в левой части величина $I(\omega)$ неотрицательна по определению).

Таким образом, с учетом упомянутой выше интерпретации интенсивности $I(\omega)$, мы получаем, что *величина спектральной плотности при значении аргумента, равном ω , характеризует силу взаимосвязи, существующей между анализируемым временным рядом $x(t)$ и гармоникой с периодом $2\pi/\omega$* . Это позволяет использовать спектр как средство улавливания периодичностей в анализируемом временном ряду: при движении вдоль спектра в заданном диапазоне частот значения ординат должны оставаться относительно малыми до тех пор, пока не достигнута частота гармонического компонента анализируемого ряда; при этой частоте на спектре обнаружится высокий пик. Соответственно, совокупность пиков спектра и будет определять набор гармонических компонентов в разложении (10.2). Отметим, что если в ряде содержится скрытая гармоника частоты ω (дающая в спектре пик в точке ω), то, конечно, в нем присутствуют также периодические члены с частотами $\omega/2$, $\omega/3$ и т. д. То есть, скажем, месячная периодичность порождает эффект появления пиков, соответствующих двухмесячным, трехмесячным и т. д. периодам. Это так называемое «эхо», повторяемое спектром на низких частотах.

Пример 10.4. На рис. 10.7 представлен пример графика спектральной плотности (то есть *спектра*), в котором присутствует упомянутый эффект «эха». Пример заимствован из статьи: *Granger C. W. J. The effect of varying month-length in the analysis of economic time-series. L'Industria, 1 (1963), 3, Milano*. Анализировался ряд ежемесячных безналичных расчетов между банками США за 1875–1958 гг. после исключения из него экспоненциальной неслучайной составляющей. Как видно из графика, спектральная функция этого ряда обнаруживает эффект появления пиков в точках, соответствующих интервалам в 2 месяца, 3 месяца и т. д.

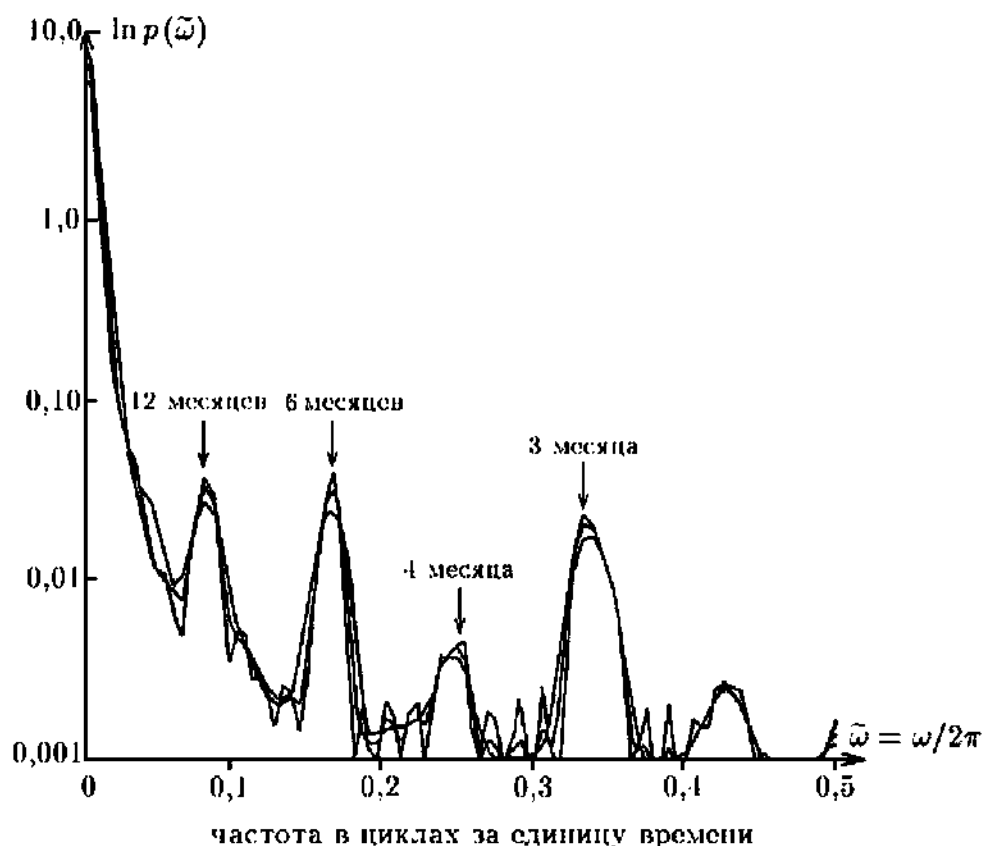


Рис. 10.7. Спектральная функция очищенных от тренда данных о банковском клиринге в США

В заключение разговора о спектральном анализе стационарного временного ряда подчеркнем еще раз тот факт, что *статистические оценки (выборочные аналоги)* введенных выше теоретических спектральных характеристик, как правило, сильно флуктуируют и обладают весьма посредственной точностью.

И, наконец, можно несколько расширить класс моделей стационарных временных рядов, используемых при анализе конкретных рядов экономической динамики. С этой целью введем понятие *стационарного в широком смысле* временного ряда.

О п р е д е л е н и е 10.3. Ряд $x(t)$ называется **слабо стационарным** (или **стационарным в широком смысле**), если его среднее значение, дисперсия и ковариации не зависят от t , то есть если выполняются соотношения (10.3)–(10.4)–(10.5).

Очевидно, все строго стационарные (или стационарные в узком смысле, см. определение 10.2) временные ряды являются одновременно и стационарными в широком смысле, но не наоборот.

10.3 Неслучайная составляющая временного ряда и методы его сглаживания

В п. 10.1 были описаны основные задачи (1)–(3) статистического анализа временного ряда (10.1'). Существенную роль в решении задач выявления и оценивания трендовой ($f_{\text{тр}}(t)$), сезонной ($\varphi(t)$) и циклической ($\psi(t)$) составляющих в разложении (10.2) играет начальный этап анализа, на котором:

- *выявляется сам факт наличия/отсутствия неслучайной (и зависящей от времени t) составляющей в разложении (10.2); по существу, речь идет о статистической проверке гипотезы*

$$H_0: \mathbf{E}x(t) = a = \text{const} \quad (10.12)$$

(включая утверждение о взаимной статистической независимости членов анализируемого временного ряда (10.1')) при различных вариантах конкретизации альтернативных гипотез типа

$$H_1: \mathbf{E}x(t) \neq \text{const}; \quad (10.13)$$

- *строится оценка (аппроксимация) для неизвестной интегральной неслучайной составляющей $f(t) = \chi(A)f_{\text{тр}}(t) + \chi(B)\varphi(t) + \chi(C)\psi(t)$, то есть решается задача сглаживания (элиминирования случайных остатков $\varepsilon(t)$) анализируемого временного ряда $x(t)$ (функции χ определены в (10.2)).*

Данный пункт посвящен именно этому начальному этапу статистического анализа временного ряда.

10.3.1 Проверка гипотезы о неизменности среднего значения временного ряда

Критерий серий, основанный на медиане. Расположим члены анализируемого временного ряда (10.1') в порядке возрастания, то есть образуем по наблюдениям (10.1') *вариационный ряд*

$$x_{(1)}, x_{(2)}, \dots, x_{(N)}.$$

Определим *выборочную медиану* $x_{\text{med}}^{(N)}$ по формуле

$$x_{\text{med}}^{(N)} = \begin{cases} x_{(\frac{N+1}{2})}, & \text{если } N \text{ нечетно,} \\ \frac{1}{2} \left(x_{(\frac{N}{2})} + x_{(\frac{N}{2}+1)} \right), & \text{если } N \text{ четно.} \end{cases}$$

После этого мы образуем «серии» из плюсов и минусов, на статистическом анализе которых основана процедура проверки гипотезы (10.12). В частности, возвращаясь к исходному временному ряду (10.1'), мы будем вместо каждого его члена $x(t)$, $t = 1, 2, \dots, N$, ставить плюс, если $x(t) > x_{\text{med}}^{(N)}$, и минус, если $x(t) < x_{\text{med}}^{(N)}$ (члены временного ряда, равные $x_{\text{med}}^{(N)}$, в полученной таким образом последовательности плюсов и минусов не учитываются).

Образованная последовательность плюсов и минусов характеризуется общим числом серий $\nu(N)$ и протяженностью самой длинной серии $\tau(N)$. При этом под «серией» понимается последовательность подряд идущих плюсов и подряд идущих минусов (в частном случае серия может состоять только из одного плюса или только из одного минуса, и тогда ее протяженность равна единице). Очевидно, что если анализируемая последовательность (10.1') состоит из статистически независимых наблюдений, случайно варьирующих около некоторого *постоянного* уровня a (то есть если справедлива гипотеза (10.12)), то чередование плюсов и минусов в построенной указанным выше способом последовательности должно быть более или менее случайным, то есть эта последовательность не должна содержать слишком длинных серий подряд идущих плюсов или подряд идущих минусов, и, соответственно, общее число серий $\nu(N)$ не должно быть слишком малым. Так что в данном критерии целесообразно рассматривать *одновременно пару критических статистик* $(\nu(N); \tau(N))$.

Тогда, следуя общей логической схеме построения статистического критерия, мы должны были бы вывести и затабулировать двумерный закон распределения случайной величины $(\nu(N); \tau(N))$, справедливый при условии, что проверяемая гипотеза (10.12) верна. Мы ограничимся здесь изложением *приближенного* критерия. Для его построения воспользуемся следующими утверждениями, справедливыми в условиях верности гипотезы (10.12):

- $(\frac{N+2}{2}, \frac{N-1}{4})$ — нормальным приближением одномерного (частного) распределения случайной величины $\nu(N)$;
- приближенной оценкой для пятипроцентной точки $\tau_{0.05}(N)$ частного распределения случайной величины $\tau(n)$:

$$\tau_{0.05}(n) \approx [1,43 \ln(n+1)]$$

(здесь и далее в этом пункте $[x]$ означает «целая часть числа x »);

- оценками сверху и снизу для вероятности

$$P\{\nu(N) < \nu_{0.95}(N); \tau(N) > \tau_{0.05}(N)\},$$

основанными на знании вероятностей $P\{\nu(N) < \nu_{0,95}(N)\}$ и $P\{\tau(N) > \tau_{0,05}(N)\}$ (здесь $\nu_{0,95}(N)$ — 95%-ная точка частного распределения $\nu(N)$).

Это позволяет сформулировать следующее приближенное правило проверки гипотезы (10.12):

если хотя бы одно из неравенств

$$\begin{aligned}\nu(N) &> \left\lfloor \frac{1}{2}(N + 2 - 1,96\sqrt{N - 1}) \right\rfloor, \\ \tau(N) &< \lceil 1,43 \ln(N + 1) \rceil\end{aligned}\quad (10.14)$$

окажется нарушенным, то гипотеза (10.12) отвергается с вероятностью ошибки α , заключенной между 0,05 и 0,0975 (и, следовательно, подтверждается наличие зависящей от времени неслучайной составляющей в разложении (10.2) анализируемого временного ряда).

Пример 10.5. Имеются результаты испытаний на долговечность 58 образцов, отобранных в хронологическом порядке из текущей продукции: 38, 33, 29, 16, 44, 21, 16, 17, 19, 1, 22, 28, 22, 14, 7, 13, 21, 15, 34, 23, 15, 19, 32, 24, 14, 13, 22, 8, 30, 11, 15, 24, 26, 14, 11, 25, 17, 10, 19, 5, 6, 16, 7, 10, 1, 5, 2, 8, 14, 14, 15, 16, 13, 11, 9, 11, 19, 21. (Подчеркнуты те выборочные данные, на месте которых в соответствующей последовательности знаков стояли бы плюсы.)

Ряд факторов, от которых существенно зависит качество образцов (сырье, квалификация персонала, сменность и т.п.), подвержен неизбежным колебаниям с течением времени, характер которых может быть как случайным, так и систематическим. Нас будет интересовать, было ли это должным образом учтено при назначении способа отбора образцов, то есть производился ли отбор так, чтобы результаты наблюдений были бы стохастически независимыми, образовывали бы случайную выборку? Так, характер изменения выборочных данных *во времени* (порядок отбора образцов из текущей продукции во времени определяется в нашем примере движением по строкам слева направо) наводит на мысль, что имела место некоторая систематическая тенденция к монотонному снижению долговечности. Ответить на вопрос, являются ли наши сомнения достаточно обоснованными, нам поможет только что описанный критерий серий.

Необходимые подсчеты дают: $\hat{x}_{\text{med}}(N) = 15,5$; $\tau(N) = 9$; $\nu(N) = 25$.

Так что из двух неравенств (10.14) лишь одно (первое) оказалось выполненным. Поэтому приходится признать, что результаты наблюдений, представленные выше, не являются стохастически независимыми и обнаруживают временную тенденцию к снижению долговечности.

Критерий «восходящих» и «нисходящих» серий. Этот критерий «улавливает» постепенное смещение (по ходу выборочного обследо-

вания) среднего значения в исследуемом распределении не только монотонного, но и более общего, например периодического, характера.

Так же, как и в предыдущем критерии, исследуется последовательность знаков — плюсов и минусов, однако правило образования этой последовательности в данном критерии иное. Исходным пунктом, как обычно, является анализируемая последовательность результатов наблюдения, то есть ряд $x(1), x(2), \dots, x(N)$; на i -м месте вспомогательной последовательности ставится плюс, если $x(i+1) - x(i) > 0$, и минус, если $x(i+1) - x(i) < 0$ (если два или несколько следующих друг за другом наблюдений равны между собой, то принимается во внимание только одно из них). Очевидно, последовательность подряд идущих плюсов будет соответствовать тогда возрастанию результатов наблюдения (восходящая серия), а последовательность минусов — их убыванию (нисходящая серия). Критерий основан на том же соображении, что и предыдущий: если выборка случайна (наблюдения независимы и одинаково распределены), то в образованной нами последовательности знаков общее число серий не может быть слишком малым, а их протяженность (измеренная в количестве подряд идущих плюсов или минусов) — слишком большой.

В частности, при уровне значимости $0,050 < \alpha < 0,0975$ количественное выражение этого правила имеет вид:

$$\nu(N) > \left[\frac{1}{3}(2N - 1) - 1,96\sqrt{\frac{16N - 29}{90}} \right], \quad (10.15)$$

$$\tau(N) < \tau_0(N),$$

где под $\nu(N)$ и $\tau(N)$, как и прежде, понимается, соответственно, общее число серий и количество подряд идущих плюсов или минусов в самой длинной серии, а величина $\tau_0(N)$ в зависимости от N определяется следующим образом:

N	$N \leq 26$	$26 < N \leq 153$	$153 < N \leq 1170$
$\tau_0(N)$	$\tau_0 = 5$	$\tau_0 = 6$	$\tau_0 = 7$

Если хотя бы одно из неравенств (10.15) окажется нарушенным, то гипотезу (10.12) следует отвергнуть (и, соответственно, признать, что в разложении (10.2) анализируемого временного ряда присутствует неслучайная, зависящая от времени t компонента) при вероятности ошибки, заключенной в пределах от 0,05 до 0,0975.

Пример 10.6. Вернемся к данным примера 10.3 (см. выше таблицу 10.3 и рис. 10.3) и воспользуемся критерием восходящих и нисходящих серий для проверки гипотезы о неизменности средней урожайности ячменя в Англии и Уэльсе на отрезке времени с 1884 по 1939 гг.

Общее число наблюдений (длина временного ряда) N в данном примере равно 56, а образованная описанным выше способом последовательность плюсов и минусов будет содержать 53 элемента (так как анализируемый временной ряд содержит две пары совпадающих соседних

наблюдений: за 1906–1907 гг. и за 1910–1911 гг.). Эта последовательность из плюсов и минусов, в частности, имеет вид:

№ п/п	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Член последо- ватель- ности	+	–	–	+	–	+	–	+	–	+	–	+	–	+	–	–	–	+

№ п/п	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
Член последо- ватель- ности	–	–	+	+	–	+	–	–	+	–	–	+	–	+	–	+	–	–

№ п/п	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53
Член последо- ватель- ности	+	+	–	+	+	+	+	–	+	+	+	+	–	–	–	+	–

Анализ полученной последовательности плюсов и минусов дает:

$$\nu(N) = 36 \quad \text{и} \quad \tau(N) = 4.$$

Воспользовавшись приведенной выше таблицей значений $\tau_0(N)$ и выражением для правой части первого из неравенств (10.15), определяем:

$$\left[\frac{1}{3}(2 \cdot 56 - 1) - 1,96 \sqrt{\frac{16 \cdot 56 - 29}{90}} \right] = [30, 92] = 30,$$

$$\tau_0(56) = 6.$$

Поскольку оба неравенства (10.15) в нашем случае оказываются выполненными, делаем вывод о непротиворечивости гипотезы о неизменности средней урожайности ячменя в Англии и Уэльсе в период времени от 1884 до 1939 г.

Критерий квадратов последовательных разностей (критерий Аббе). Если есть основания полагать, что случайный разброс наблюдений $x(1), x(2), \dots, x(N)$ относительно своих средних значений в условиях

справедливости гипотезы (10.12) подчиняется нормальному закону распределения вероятностей, то для выяснения вопроса о возможном систематическом смещении среднего в ходе выборочного обследования целесообразнее воспользоваться критерием квадратов последовательных разностей⁶.

Для проверки гипотезы (10.12) с помощью данного критерия подсчитывают величину

$$\gamma^{(N)} = \frac{q^2(N)}{s'^2(N)},$$

где $q^2(N) = \frac{1}{2(N-1)} \sum_{i=1}^{N-1} (x(i+1) - x(i))^2$;

$$s'^2(N) = \frac{1}{N-1} \sum_{i=1}^N (x(i) - \bar{x})^2,$$

$$\bar{x} = \bar{x}(N) = \frac{1}{N} \sum_{i=1}^N x(i).$$

Если окажется, что

$$\gamma(N) \leq \gamma_{\alpha}^{\min}(N), \quad (10.16)$$

то гипотеза (10.12) отвергается. При этом величина $\gamma_{\alpha}^{\min}(N)$ для $N > 60$ подсчитывается по формуле

$$\gamma_{\alpha}^{\min}(N) = 1 + \frac{u_{\alpha}}{\sqrt{N + 0,5(1 + u_{\alpha}^2)}},$$

где u_{α} — α -квантиль стандартного нормального распределения.

Величины $\gamma_{\alpha}^{\min}(N)$ при $N \leq 60$ для трех наиболее употребительных значений уровня значимости α даны в таблице 4.9 книги [Большев, Смирнов (1965)].

10.3.2 Методы сглаживания временного ряда (выделение неслучайной составляющей)

Методы выделения неслучайной составляющей в траектории, отражающей поведение анализируемого временного ряда (10.1'), можно условно разделить на два типа.

⁶В этом случае данный критерий оказывается более мощным, чем предыдущий. Это означает, в частности, что если мы воспользуемся обоими этими критериями при данном объеме выборки N и заданном уровне значимости α (то есть при заданной вероятности ошибочного отвержения гипотезы (10.12)), то вероятность ошибиться в другую сторону (то есть принять гипотезу (10.12), в то время как на самом деле она является ошибочной) окажется меньшей в случае критерия квадратов последовательных разностей.

Методы первого типа (аналитические) основаны на допущении, что нам известен общий вид неслучайной составляющей в разложении (10.2)

$$f(t) = \chi(A)f_{\text{тр}}(t) + \chi(B)\varphi(t) + \chi(C)\psi(t). \quad (10.17)$$

Например, рассматривая график временного ряда из примера 10.2 (см. рис. 10.2), исследователь может предположить, что неслучайная составляющая динамики курса анализируемых акций описывается *линейной функцией времени t* , то есть

$$f(t) = \theta_0 + \theta_1 t,$$

где θ_0 и θ_1 — некоторые неизвестные (подлежащие статистическому оцениванию) параметры модели. Тогда задача выделения неслучайной составляющей (или — *задача элиминирования случайных остатков в разложении* (10.2), или, что то же, — *задача сглаживания временного ряда* (10.1')) сведется к задаче построения «хороших» оценок $\hat{\theta}_0$ и $\hat{\theta}_1$ для параметров θ_0 и θ_1 .

Соответствующие методы мы назвали *аналитическими*, поскольку «на выходе» задачи мы будем иметь явное аналитическое задание (оценку, приближение) $\hat{f}(t)$ для искомой неслучайной составляющей $f(t)$. Другими словами, $\hat{f}(t)$ будет представлена в виде формулы $f(t; \hat{\Theta})$ *функции известного вида*, в которой неизвестные параметры $\Theta = (\theta_0, \theta_1, \dots)$ заменены их статистическими оценками $\hat{\Theta}$.

Методы второго типа (алгоритмические) не связаны ограничительным допущением о том, что общий аналитический вид искомой функции (10.17) известен исследователю. В этом смысле они являются более гибкими, более привлекательными. Однако «на выходе» задачи они доставляют исследователю *лишь алгоритм расчета* оценки $\hat{f}(t)$ для искомой функции $f(t)$ в любой наперед заданной точке t и не претендуют на аналитическое (то есть заданное в виде явной формулы) представление функции (10.17).

Аналитические методы выделения (оценки) неслучайной составляющей временного ряда. Эти методы реализуются в рамках моделей регрессии, в которых в роли зависимой (объясняемой) переменной выступает переменная $x(t)$, генерирующая анализируемый временной ряд (10.1'), а в роли единственной объясняющей переменной — время t .

Таким образом, рассматривается модель регрессии вида

$$x(t) = f(t; \Theta) + \varepsilon(t), \quad t = 1, 2, \dots, N,$$

в которой общий вид функции $f(t; \Theta)$ известен, но неизвестны значения параметров $\Theta = (\theta_0, \theta_1, \dots, \theta_m)^\top$. Оценки $\hat{\Theta}$ параметров Θ строятся по наблюдениям $\{t; x(t)\}_{t=1,2,\dots,N}$ с помощью соответствующих методов.

Выбор метода оценивания зависит от гипотетического вида функции $f(t; \Theta)$ и стохастической природы случайных регрессионных остатков $\varepsilon(t)$. Так, например, если функция $f(t; \Theta)$ имеет вид алгебраического полинома степени p , то есть

$$f(t; \Theta) = \theta_0 + \theta_1 t + \dots + \theta_p t^p, \quad (10.18)$$

и при этом длина временного ряда n существенно превышает степень этого полинома p^7 , а регрессионные остатки $\varepsilon(1), \varepsilon(2), \dots, \varepsilon(N)$ взаимно некоррелированы, то оценки $\hat{\Theta}$ параметров Θ могут быть получены с помощью обычного метода наименьших квадратов (см. п. 4.2.1, формулу (4.22)):

$$\hat{\Theta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}. \quad (10.19)$$

В этой формуле $n \times (p + 1)$ — матрица $\mathbf{X}$ наблюдаемых значений объясняющих переменных имеет в данном случае вид:

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 1^2 & \dots & 1^p \\ 1 & 2 & 2^2 & \dots & 2^p \\ \dots & \dots & \dots & \dots & \dots \\ 1 & N & N^2 & \dots & N^p \end{pmatrix},$$

а вектор-столбец наблюдаемых значений зависимой переменной $\mathbf{Y} = (x(1), x(2), \dots, x(N))^T$.

Отказ от взаимной некоррелированности регрессионных остатков в данной модели повлечет за собой необходимость использования *практически реализуемого обобщенного МНК* (см. п. 5.5); исследование моделей (10.18), в которых неслучайная компонента $f(t; \Theta)$ *нелинейна* относительно оцениваемых параметров Θ , потребует техники статистического анализа нелинейных моделей регрессии и т. д.

Алгоритмические методы выделения неслучайной составляющей временного ряда (методы скользящего среднего). В основе этих методов элиминирования случайных флуктуаций в поведении анализируемого временного ряда лежит простая идея: если «индивидуальный» разброс значений члена временного ряда $x(t)$ около своего среднего (сглаженного) значения a характеризуется дисперсией σ^2 , то разброс среднего из K членов временного ряда $(x(1) + x(2) + \dots + x(K))/K$ около того же значения a будет характеризоваться гораздо меньшей величиной дисперсии, а именно дисперсией, равной σ^2/K . А уменьшение меры случайного разброса (дисперсии) и означает как раз *сглаживание* соответствующей траектории. Поэтому выбирают некоторую нечетную «длину усреднения» $K = 2m + 1$, измеренную в числе подряд идущих членов анализируемого временного ряда (конкретный выбор числа m

⁷Обычно удовлетворительная надежность статистических оценок $\hat{\Theta}$ параметров Θ достигается уже при $N \geq 4p$, но, как минимум, требуется, чтобы $N > p + 1$.

зависит от специфики исходных данных, но, как правило, m выбирают таким образом, чтобы оно удовлетворяло неравенству $m < n/3$; обычно m не превышает трех; более подробно о выборе m см. ниже). А затем сглаженное значение $\hat{f}(t)$ временного ряда $x(t)$ вычисляют по значениям $x(t-m), x(t-m+1), \dots, x(t), x(t+1), \dots, x(t+m)$ по формуле

$$\hat{f}(t) = \sum_{k=-m}^m w_k x(t+k), \quad t = m+1, m+2, \dots, n-m, \quad (10.20)$$

где w_k ($k = -m, -m+1, \dots, m$) — некоторые «весовые» коэффициенты (или просто «веса»), в сумме равные единице, то есть $\sum_{k=-m}^m w_k = 1$.

Поскольку, изменяя t от $m+1$ до $n-m$, мы как бы «скользим» по оси времени (так что при переходе от t к $t+1$ в составе слагаемых правой части (10.20) происходит замена только одного слагаемого $x(t-m)$ слагаемым $x(t-m+1)$), то и методы, основанные на формуле (10.20), принято называть *методами скользящей средней (МСС)*.

Очевидно, один МСС отличается от другого выбором параметров m и w_k ($k = -m, -m+1, \dots, m$). Остановимся на этом подробнее.

Определение параметров w_k основано на следующей процедуре. В соответствии с известной теоремой Вейерштрасса любая гладкая функция $f(x)$ при самых общих допущениях может быть *локально* (то есть в ограниченном интервале изменения ее аргумента t) представлена алгебраическим полиномом подходящей степени p . Поэтому берем первые $2m+1$ членов временного ряда $x(1), x(2), \dots, x(2m+1)$, строим с помощью МНК полином $\hat{x}_1(t)$ степени p , аппроксимирующий поведение этой начальной части траектории временного ряда, и используем этот полином для определения оценки $\hat{f}(t)$ сглаженного значения $f(x)$ временного ряда в *средней* (то есть $(m+1)$ -й) точке этого отрезка ряда, то есть полагаем $\hat{f}(m+1) = \hat{x}_1(m+1)$. Затем «скользим» по оси времени на один такт и таким же способом подбираем полином $\hat{x}_2(t)$ той же степени p к отрезку временного ряда $x(2), x(3), \dots, x(2m+2)$ и определяем оценку сглаженного значения временного ряда в средней точке сдвинутого на единицу отрезка временного ряда, то есть $\hat{f}(m+2) = \hat{x}_2(m+2)$, и т. д.

В результате мы найдем оценки для сглаженных значений $\hat{f}(t)$ анализируемого временного ряда при всех t , кроме $t = 1, 2, \dots, m$ и $t = n, n-2, \dots, n-m+1$.

Покажем, что подбор наилучшего (в смысле критерия МНК) аппроксимирующего полинома к траектории анализируемого временного ряда (на заданном временном интервале) действительно приводит к формуле вида (3.20), причем результат (то есть значения коэффициентов w_k , $k = -m, -m+1, \dots, m$) не зависит от того, для какого именно из «скользящих» временных интервалов был осуществлен этот подбор.

Начнем с простого примера. Будем полагать, что локальное поведе-

ние сглаженной функции $f(t)$ описывается алгебраическим полиномом 1-й степени ($p = 1$), то есть что на выбранном временном интервале (протяженностью $2m + 1$ временных тактов) неслучайная составляющая $f(x)$ может быть аппроксимирована *линейной функцией времени*:

$$f(t) = \theta_0 + \theta_1 t, \quad t = 1, 2, \dots, 2m + 1.$$

Без ограничения общности можно переобозначить моменты времени таким образом, чтобы рассматривать нашу модель на временном отрезке $t' = -m, -m + 1, \dots, -1, 0, 1, \dots, m - 1, m$ (тогда средняя точка будет соответствовать $t' = 0$). Очевидно, «новое» (условное) время t' связано со «старым» временем t соотношением $t' = t - (m + 1)$.

Итак, мы хотим подобрать коэффициенты θ_0 и θ_1 таким образом, чтобы минимизировать критерий МНК, то есть

$$\sum_{t'=-m}^m (x(t') - \theta_0 - \theta_1 t')^2 \rightarrow \min_{\theta_0, \theta_1}.$$

Дифференцирование левой части по θ_0 и θ_1 и приравнивание полученных частных производных к нулю дает систему уравнений для определения МНК-оценок $\hat{\theta}_0$ и $\hat{\theta}_1$:

$$\begin{cases} (2m + 1)\hat{\theta}_0 + \hat{\theta}_1 \sum_{t'=-m}^m t' = \sum_{t'=-m}^m x(t'), \\ \hat{\theta}_0 \sum_{t'=-m}^m t' + \hat{\theta}_1 \sum_{t'=-m}^m (t')^2 = \sum_{t'=-m}^m t' x(t'). \end{cases}$$

Поскольку $\sum_{t'=-m}^m t' = 0$, а оценка сглаженного значения временного ряда определяется в *средней точке* рассматриваемого временного интервала соответствующим значением функции $\hat{\theta}_0 + \hat{\theta}_1 t'$ при $t' = 0$, то:

$$\begin{aligned} \hat{f}(m + 1) = \hat{\theta}_0 + \hat{\theta}_1 t']_{t'=0} = \hat{\theta}_0 &= \frac{1}{2m + 1} \sum_{t'=-m}^m x(t') = \\ &= \frac{1}{2m + 1} \sum_{t=1}^{2m+1} x(t). \end{aligned} \quad (10.21)$$

Проделав тот же самый анализ для любого другого временного интервала $t = k + 1, k + 2, \dots, k + 2m + 1$ ($k = 1, 2, \dots, n - 3m - 1$), мы получим аналогичный результат. Таким образом, в общем случае имеем:

$$\hat{f}(t) = \frac{1}{2m + 1} \sum_{k=-m}^m x(t + k), \quad (10.21')$$

то есть при линейном характере локальной аппроксимации траектории временного ряда в качестве его сглаженного значения в точке t следует брать среднее арифметическое из окаймляющих его $2m + 1$ соседних значений $x(t - m), x(t - m + 1), \dots, x(t), \dots, x(t + m)$, или, в терминах весовых коэффициентов: $w_{-m} = w_{-m+1} = \dots = w_m = 1/(2m + 1)$.

Нетрудно показать, что если локальное поведение сглаженной функции $f(t)$ описывается алгебраическим полиномом нулевой степени ($p = 0$), то есть, другими словами, если функция $f(x)$ локально ведет себя как постоянная величина ($f(t) = \theta_0 = \text{const}$), то оценка сглаженного значения временного ряда в точке t определяется той же самой формулой (10.21).

Рассмотрим теперь случай $p = 2, m = 2$. Это означает, что в «скользящий» временной отрезок, по которому будет производиться усреднение значений временного ряда, мы будем включать 5 точек ($K = 2m + 1 = 2 \cdot 2 + 1 = 5$) и что локальное поведение сглаженного временного ряда внутри каждого такого отрезка мы будем аппроксимировать параболой 2-го порядка, то есть

$$f(t') = \theta_0 + \theta_1 t' + \theta_2 (t')^2, \quad t' = -2, -1, 0, 1, 2$$

(здесь t' — уже «переобозначенное» время, то есть сдвинутое по отношению к реальному времени таким образом, чтобы точка $t' = 0$ была бы средней на каждом временном отрезке усреднения).

Определяя оценки $\hat{\theta}_0, \hat{\theta}_1$ и $\hat{\theta}_2$, соответственно, коэффициентов θ_0, θ_1 и θ_2 по методу наименьших квадратов, имеем:

- критерий МНК: $Q(\Theta) = \sum_{t'=-2}^2 (x(t') - \theta_0 - \theta_1 t' - \theta_2 t'^2)^2$;
- систему уравнений МНК:

$$\left\{ \begin{aligned} \frac{\partial Q(\Theta)}{\partial \theta_0} &= -2 \left[\sum_{t'=-2}^2 x(t') - 5\theta_0 - \theta_1 \sum_{t'=-2}^2 t' - \theta_2 \sum_{t'=-2}^2 (t')^2 \right] = 0, \\ \frac{\partial Q(\Theta)}{\partial \theta_1} &= -2 \left[\sum_{t'=-2}^2 t' x(t') - \theta_0 \sum_{t'=-2}^2 t' - \theta_1 \sum_{t'=-2}^2 (t')^2 - \theta_2 \sum_{t'=-2}^2 (t')^3 \right] = 0, \\ \frac{\partial Q(\Theta)}{\partial \theta_2} &= -2 \left[\sum_{t'=-2}^2 (t')^2 x(t') - \theta_0 \sum_{t'=-2}^2 (t')^2 - \theta_1 \sum_{t'=-2}^2 (t')^3 - \theta_2 \sum_{t'=-2}^2 (t')^4 \right] = 0; \end{aligned} \right. \quad (10.22)$$

• решение системы уравнений МНК (с учетом того, что $\sum_{t'=-2}^2 t' = 0$):

$$\begin{aligned}\hat{\theta}_0 &= \frac{\sum_{t'=-2}^2 (t')^4 \cdot \sum_{t'=-2}^2 x(t') - \sum_{t'=-2}^2 (t')^2 \sum_{t'=-2}^2 (t')^2 x(t')}{5 \sum_{t'=-2}^2 (t')^4 - \left(\sum_{t'=-2}^2 (t')^2 \right)^2} = \\ &= \frac{34 \sum_{t'=-2}^2 x(t') - 10 \sum_{t'=-2}^2 (t')^2 x(t')}{170 - 100} = \\ &= \frac{1}{35} [-3x(-2) + 12x(-1) + 17x(0) + 12x(1) - 3x(2)], \\ \hat{\theta}_1 &= \frac{\sum_{t'=-2}^2 t' x(t')}{\sum_{t'=-2}^2 (t')^2} = \frac{1}{10} [-2x(-2) - x(-1) + x(1) + 2x(2)], \\ \hat{\theta}_2 &= \frac{5 \sum_{t'=-2}^2 (t')^2 x(t') - \sum_{t'=-2}^2 (t')^2 \sum_{t'=-2}^2 x(t')}{5 \sum_{t'=-2}^2 (t')^4 - \left(\sum_{t'=-2}^2 (t')^2 \right)^2} = \\ &= \frac{1}{14} [2x(-2) - x(-1) - 2x(0) - x(1) + 2x(2)].\end{aligned}$$

Соответственно, оценка сглаженного значения $f(t)$ анализируемого временного ряда в точке t определится значением оцененной параболы $\hat{\theta}_0 + \hat{\theta}_1 t' + \hat{\theta}_2 (t')^2$ при $t' = 0$, то есть:

$$\begin{aligned}\hat{f}(t) &= \hat{\theta}_0 + \hat{\theta}_1 t' + \hat{\theta}_2 (t')^2 \Big|_{t'=0} = \hat{\theta}_0 = \\ &= \frac{1}{35} [-3x(t-2) + 12x(t-1) + 17x(t) + 12x(t+1) - 3x(t+2)].\end{aligned}$$

Таким образом, мы снова получили формулу вида (10.20), в которой $w_{-2} = w_2 = -3/35$, $w_{-1} = w_1 = 12/35$ и $w_0 = 17/35$.

Продемонстрированная на двух простых примерах процедура определения весов в формуле (10.20) имеет общий характер. Если по $2m+1$ точкам подбирают с помощью МНК алгебраический полином степени p , то необходимо минимизировать критерий вида

$$Q(\Theta) = \sum_{t'=-m}^m (x(t') - \theta_0 - \theta_1 t' - \dots - \theta_p (t')^p)^2. \quad (10.23)$$

Минимизация $Q(\Theta)$ по Θ приводит к уравнениям относительно Θ , аналогичным (10.22), которые разбиваются на две более простые подсистемы из-за того, что $\sum_{t'=-m}^m (t')^k = 0$ при всех нечетных значениях k . Решения

этих систем зависят от численных значений сумм вида $\sum_{t'=-m}^m (t')^k$ (где k

четно) и линейных функций от $x(t')$ вида $\sum_{t'=-m}^m (t')^k x(t')$ (это следует из

самой формулы МНК-оценок (10.19) и вида вектора Y в данной задаче).

Оценкой $\hat{f}(t)$ неслучайной составляющей $f(t)$ анализируемого временно-го ряда $x(t)$ в точке $t = t_0$ ($t_0 = m + 1, m + 2, \dots, 2n - m$) будет величина $\hat{\theta}_0$, представляющая собой взвешенное среднее значений от $x(t_0 - m)$ до $x(t_0 + m)$. В таблице 10.4 приводятся подсчитанные указанным выше способом значения весовых коэффициентов w_k ($k = -m, -m + 1, \dots, -1, 0$) для ряда значений m и p .

Значения w_k для положительных k не приводятся, так как в силу имеющего место свойства симметрии этих коэффициентов $w_k = w_{-k}$.

Таблица 10.4. Значения весовых коэффициентов в формуле метода скользящего среднего (10.20) для различных длин отрезков усреднения m и порядков аппроксимирующих полиномов p

m	p	w_{-m}	w_{-m+1}	$\dots$	w_0
m_0	0 или 1	$\frac{1}{2m_0+1}$	$\frac{1}{2m_0+1}$	$\dots$	$\frac{1}{2m_0+1}$
2	2 или 3	$-\frac{3}{35}$	$\frac{12}{35}$		$\frac{17}{35}$
3	2 или 3	$-\frac{2}{21}$	$\frac{3}{21}$	$\frac{6}{21}$	$\frac{7}{21}$
4	2 или 3	$-\frac{21}{231}$	$\frac{14}{231}$	$\frac{39}{231}$ $\frac{54}{231}$	$\frac{59}{231}$
3	4 или 5	$\frac{5}{231}$	$-\frac{30}{231}$	$\frac{75}{231}$	$\frac{131}{231}$
4	4 или 5	$\frac{15}{429}$	$-\frac{55}{429}$	$\frac{30}{429}$ $\frac{135}{429}$	$\frac{179}{429}$

Из самого хода процедуры вычисления оценок параметров Θ с помощью минимизации критерия $Q(\Theta)$ (см. (10.23)) и вытекающего из него вида θ_0 (определяющего значения весовых коэффициентов w_k в формуле (10.20)) можно вывести следующие важные свойства этих коэффициентов:

- веса w_k симметричны относительно серединного значения w_0 , то есть $w_k = w_{-k}$, $k = 1, 2, \dots, m$ (это следует из того факта, что они получены как функции сумм $\sum_{t'=-m}^m (t')^j x(t')$, которые сами симметричны);

- сумма всех весов w_k равна единице, то есть $\sum_{k=-m}^m w_k = 1$ (то, что это должно быть так, можно проверить на таком частном случае: пусть все значения членов временного ряда на анализируемом временном отрезке равны одной и той же константе c ; тогда скользящее среднее $\hat{f}(t) = \sum_{k=-m}^m w_k x(t+k) = c \sum_{k=-m}^m w_k$ тоже должно быть равно этой константе c , а это может быть только в случае $\sum_{k=-m}^m w_k = 1$);
- при одной и той же длине $N = 2m + 1$ временного отрезка, по которому производится усреднение, веса w_k в формуле (10.20) для полиномов четной степени (то есть для $p = 2l, l = 0, 1, 2, \dots$) будут теми же самыми, что и для полиномов степени, на единицу большей (то есть для $p = 2l + 1$); этот результат виден из того, что при составлении систем уравнений, аналогичных (10.22), для $p = 2l$ и для $p = 2l + 1$ мы получим две системы, каждая из которых распадается на две подсистемы: одна из этих подсистем будет содержать параметр θ_0 в качестве неизвестного задачи, а вторая — нет; так вот, первые две подсистемы уравнений (то есть подсистемы, содержащие θ_0) в системах, соответствующих $p = 2l$ и $p = 2l + 1$, будут одними и теми же, а следовательно, и решения для $\hat{\theta}_0$ будут одинаковыми (а ведь именно значениями $\hat{\theta}_0$ определяются веса w_k).

З а м е ч а н и е 1: определение сглаженных значений в крайних точках.

В качестве сглаженных значений $f(t)$ в m первых точках и в m последних точках всего анализируемого временного интервала используются соответствующие значения локально аппроксимирующих полиномов, построенных, соответственно, по $2m + 1$ первым и по $2m + 1$ последним точкам анализируемого временного ряда, а именно:

$$\hat{f}(t) = \begin{cases} \hat{x}_1(t) = \hat{\theta}_0^{(1)} + \hat{\theta}_1^{(1)}t + \dots + \hat{\theta}_p^{(1)}t^p, & t = 1, 2, \dots, m; \\ \hat{x}_{N-2m}(t) = \hat{\theta}_0^{(N-2m)} + \hat{\theta}_1^{(N-2m)}t + \dots + \hat{\theta}_p^{(N-2m)}t^p, & t = N - m + 1, \dots, N. \end{cases} \quad (10.24)$$

В (10.24) $\hat{x}_1(t)$ — сглаживающий полином, построенный по 1-му интервалу сглаживания, то есть — по точкам $(1, x(1)), (2, x(2)), \dots, (2m+1, x(2m+1))$ (соответственно, $\hat{\theta}_j^{(1)}, j = 0, 1, \dots, p$, — МНК-оценки параметров этого полинома, вычисленные по точкам 1-го интервала сглаживания), а $\hat{x}_{N-2m}(t)$ — сглаживающий полином, построенный по $(N - 2m)$ -му интервалу сглаживания, то есть — по точкам $(N - 2m, x(N - 2m)),$

$(N - 2m + 1, x(N - 2m + 1)), \dots, (N, x(N))$ (соответственно, $\hat{\theta}_j^{(N-2m)}$, $j = 0, 1, \dots, p$, — МНК-оценки параметров этого полинома, вычисленные по точкам $(N - 2m)$ -го интервала сглаживания).

З а м е ч а н и е 2: усреднение по четному числу точек.

До сих пор мы использовали *нечетное* число $(2m + 1)$ точек в качестве базы для расчета скользящего среднего. Однако специфика конкретной задачи зачастую обуславливает необходимость использования интервала скользящего усреднения, содержащего четное число точек. С подобными ситуациями мы столкнемся при вычислении среднесуточных часовых данных (24 часа в сутках), среднемесячных недельных данных (4 недели в месяце), среднегодовых квартальных или месячных данных (4 квартала и 12 месяцев в году) и т. п.

В этих случаях, как и прежде, сглаженное значение $\hat{f}(t)$ временного ряда $x(t)$ вычисляется в средней точке скользящего интервала усреднения. Очевидно, таковой (при длине интервала усреднения, равной $2m$, и при начальной точке k этого интервала) будет точка

$$t^* = \frac{k + (k + 2m) - 1}{2} = k + m - \frac{1}{2}, \quad k = 1, 2, \dots, N - 2m + 1. \quad (10.25)$$

Другими словами, мы будем получать при этом сглаженные значения анализируемого временного ряда не в точках его наблюдения $t = 1, 2, \dots, n$, а для моментов времени t^* (см. (10.25)), лежащих посередине между точками наблюдения. Чтобы получить сглаженное значение временного ряда $x(t)$ именно в одной из точек его наблюдения (например, в точке $t = k + m$), необходимо вычислить его сглаженные значения для двух окаймляющих эту точку промежуточных моментов времени (то есть для $t_1^* = k + m - \frac{1}{2}$ и $t_2^* = t_1^* + 1 = k + m + \frac{1}{2}$) и взять их среднее арифметическое, то есть:

$$\hat{f}(k + m) = \frac{1}{2} \left[\hat{f} \left(k + m - \frac{1}{2} \right) + \hat{f} \left(k + m + \frac{1}{2} \right) \right].$$

Предположим, например, что имеются помесечные наблюдения с января по декабрь, «привязанные» к последнему дню каждого месяца. Простое скользящее среднее (то есть МСС с $p = 0$ или 1) по двенадцати точкам (то есть $K = 2m = 2 \cdot 6 = 12$) дает оценку значения неслучайной составляющей для *середины июля* (то есть для точки $t^* = (6+7)/2 = 6,5$). Поэтому, чтобы получить сглаженное значение *на конец июля*, мы возьмем арифметическое среднее сглаженных значений, вычисленных для середины июля и середины августа. Нетрудно убедиться, что это эквивалентно вычислению 13-месячного среднего с весами $w_1 = w_{13} = 1/24$, $w_2 = w_3 = \dots = w_{12} = 1/12$. То есть при расчете сглаженного значения берутся два январских наблюдения, но каждое входит с весом, в два раза меньшим, чем наблюдения за другие месяцы.

З а м е ч а н и е 3: влияние скользящего усреднения на остаточную случайную компоненту в разложении (10.2).

При выделении неслучайной составляющей $f(t)$ анализируемого временного ряда $x(t)$ с помощью методов скользящего среднего мы должны отдавать себе отчет в том, что получаемая при этом оценка $\hat{f}(t)$ лишь *приблизженно* описывает поведение функции (10.17) и сама продолжает содержать в себе некоторую, хотя и менее ярко выраженную (в смысле уменьшения дисперсии случайных колебаний) случайную составляющую $\tilde{\varepsilon}(t)$.

Рассмотрим теперь как влияет выделение неслучайной составляющей с помощью МСС на исходные случайные остатки $\varepsilon(t)$ анализируемого временного ряда $x(t)$ (см. (10.2)) в ситуации, когда эти исходные случайные остатки взаимно некоррелированы и гомоскедастичны, то есть

$$\mathbf{E} \varepsilon(t) \equiv 0,$$

$$\mathbf{E} (\varepsilon(t)\varepsilon(t+\tau)) = \begin{cases} \sigma^2 & \text{при } \tau = 0, \\ 0 & \text{при } \tau \neq 0. \end{cases} \quad (10.26)$$

Пусть $w_1, w_2, \dots, w_{2m+1}$ будут весами, используемыми при вычислении среднего по $2m+1$ последовательным членам временного ряда $x(t+1), \dots, x(t+2m), x(t+2m+1)$. С теми же весами будут суммироваться и остаточные компоненты $\varepsilon(t+1), \varepsilon(t+2), \dots, \varepsilon(t+2m+1)$. Поэтому случайный остаток $\tilde{\varepsilon}(t+m+1)$, характеризующий оценку сглаженного значения $\hat{f}(t+m+1)$ в средней точке $t+m+1$, будет иметь вид

$$\tilde{\varepsilon}(t+m+1) = \sum_{k=1}^{2m+1} w_k \varepsilon(t+k).$$

Легко вычисляются:

$$\mathbf{E} \tilde{\varepsilon}(t) = \mathbf{E} \tilde{\varepsilon}(t+m+1) = 0,$$

$$\mathbf{D} \tilde{\varepsilon}(t) = \mathbf{D} \tilde{\varepsilon}(t+m+1) = \mathbf{D} \varepsilon(t) \sum_{k=1}^{2m+1} w_k^2 = \sigma^2 \sum_{k=1}^{2m+1} w_k^2, \quad (10.27)$$

$$\begin{aligned} \text{cov} (\tilde{\varepsilon}(t+m+1), \tilde{\varepsilon}(t+m+1+\tau)) &= \\ &= \mathbf{E} [w_1 \varepsilon(t+1) + w_2 \varepsilon(t+2) + \dots + w_{2m+1} \varepsilon(t+2m+1)] \times \\ &\times [w_1 \varepsilon(t+1+\tau) + \dots + w_{2m+1} \varepsilon(t+2m+1+\tau)]. \end{aligned} \quad (10.28)$$

Учитывая взаимную некоррелированность исходных случайных остатков $\varepsilon(1), \varepsilon(2), \dots, \varepsilon(N)$ (см. (10.26)), соотношение (10.28) может быть преобразовано к виду

$$\text{cov} (\tilde{\varepsilon}(t+m+1), \tilde{\varepsilon}(t+m+1+\tau)) = \sigma^2 \sum_{j=1}^{2m+1-\tau} w_j w_{j+\tau}, \quad \tau = 1, 2, \dots, 2m. \quad (10.28')$$

Из (10.28') и (10.27), в частности, следует, что автокорреляция между преобразованными случайными остатками $\tilde{\varepsilon}(t)$, разнесенными друг от друга на τ единиц, равна

$$r_{\tilde{\varepsilon}} = \frac{\sum_{j=1}^{2m+1-\tau} w_j w_{j+\tau}}{\sum_{j=1}^{2m+1} w_j^2}, \quad \tau = 1, 2, \dots, 2m. \quad (10.29)$$

Отсюда следует, что последовательность сглаженных значений остатков $\tilde{\varepsilon}(t)$ имеет ненулевые автокорреляции, определяемые формулой (10.29), вплоть до порядка $2m$ (то, что автокорреляции более высокого порядка равны нулю, следует непосредственно из (10.28)). Более того, для скользящих средних, наиболее часто применяемых на практике, величина $r_{\tilde{\varepsilon}}(1) = r_{\hat{f}}(1)$ будет положительной и может принимать достаточно высокие (близкие к единице) значения. Таким образом, ряд сглаженных значений $\hat{f}(t)$ будет *более гладким*, чем исходный временной ряд $x(t)$ (так как дисперсия $\hat{f}(t)$, равная дисперсии $\tilde{\varepsilon}(t)$, будет меньше исходной дисперсии σ^2 , что следует из (10.27)), однако в нем могут появиться систематические колебания, обусловленные автокоррелированностью его последовательных значений (эффект Слущкого-Юла).

Метод экспоненциально взвешенного скользящего среднего (метод Брауна⁸). До сих пор все методы скользящего среднего построения оценок $\hat{f}(t)$ для неслучайных составляющих $f(t)$ анализируемого временного ряда $x(t)$ основывались на критерии (10.23) *классического* МНК, в соответствии с которым все исходные статистические данные $(t, x(t)), t = 1, 2, \dots, N$, имеют равный вес. Однако в задачах прогноза, в которых сглаженная функция $\hat{f}(t)$ используется обычно для *экстраполяции в будущее*, кажется более естественным сравнительно «недавним» исходным данным придавать, в определенном смысле, больший вес, чем наблюдениям, относящимся к далекому прошлому (в таких случаях говорят о *дисконтировании* наблюдений). Обратимся к одному из наиболее распространенных методов выделения «более свежих» наблюдений — *методу экспоненциально взвешенного скользящего среднего* (МЭВСС).

В соответствии с этим методом оценка сглаженного значения $\hat{f}(t)$ в точке t определяется как решение оптимизационной задачи вида

$$Q(f) = \sum_{k=0}^{t-1} \lambda^k (x(t-k) - f)^2 \rightarrow \min_f, \quad (10.30)$$

где λ — некоторое положительное число, меньшее единицы (то есть $0 < \lambda < 1$).

⁸См. Brown R. G. Smoothing, Forecasting and Prediction. Prentice-Hall, Englewood Cliffs, N.Y., 1963.

Мы видим, что веса λ^k при «невязках» в критерии $Q(f)$ обобщенного («взвешенного») МНК *уменьшаются экспоненциально* по мере удаления наблюдений $x(t - k)$ в прошлое (то есть по мере роста k), — отсюда и название метода.

Решение оптимизационной задачи (10.30), то есть дифференцирование $Q(f)$ по f , приравнивание полученной производной к нулю и решение образовавшегося таким образом уравнения относительно f , дает:

$$\hat{f}(t) = \frac{1 - \lambda}{1 - \lambda^t} \sum_{k=0}^{t-1} \lambda^k x(t - k). \quad (10.31)$$

Другими словами, речь вновь идет о скользящем усреднении исходных значений $x(t)$ анализируемого временного ряда. Однако в отличие от обычного МСС в данном случае, во-первых, скользит только правый конец интервала усреднения (левый его конец закреплён в точке $t = 1$) и, во-вторых, веса при $x(t - k)$ экспоненциально уменьшаются по мере «удаления в прошлое» (то есть по мере роста k). Кроме того, формула (10.31) даёт оценку сглаженного значения временного ряда не в средней, а в правой конечной точке интервала усреднения.

Посмотрим, как преобразовались взаимно некоррелированные случайные остатки $\varepsilon(t)$ из разложения (10.2) при переходе от исходных значений временного ряда $x(t)$ к сглаженным $\hat{f}(t)$ по формуле (10.31).

Поскольку преобразованные остатки $\tilde{\varepsilon}(t)$ связаны с исходными остатками $\varepsilon(t)$ тем же самым соотношением, которым связаны значения $\hat{f}(t)$ и $x(t)$ (см. (10.31)), то очевидно:

$$\begin{aligned} \mathbf{E} \tilde{\varepsilon}(t) &= 0, \\ \mathbf{D} \tilde{\varepsilon}(t) &= \sigma^2 \frac{(1 - \lambda)^2}{(1 - \lambda^t)^2} \sum_{k=0}^{t-1} \lambda^{2k} = \sigma^2 \frac{(1 - \lambda)(1 + \lambda^t)}{(1 + \lambda)(1 - \lambda^t)}. \end{aligned} \quad (10.32)$$

Из (10.32) видно, что при значениях λ , не слишком близких к единице, и для достаточно удалённых от прошлого значений t случайные остатки $\tilde{\varepsilon}(t)$ (а следовательно, и сглаженные значения $\hat{f}(t)$) подвержены существенно меньшему случайному разбросу, чем $x(t)$, то есть ведут себя *более гладко*.

З а м е ч а н и е 4: случай «бесконечно удалённого» прошлого.

Если анализируемый временной ряд достаточно длинный (то есть N достаточно велико), то для представляющих главный интерес в прогнозе «свежих» значений временного ряда $x(t)$ (то есть при t близких к n) можно приближенно допускать, что их прошлое как бы «уходит в бесконечность». Тогда соотношения (10.30), (10.31) и (10.32) могут быть

заменены, соответственно, соотношениями

$$Q(f) = \sum_{k=0}^{\infty} \lambda^k (x(t-k) - f)^2 \rightarrow \min_f, \quad (10.30')$$

$$\hat{f}(t) = (1 - \lambda) \sum_{k=0}^{\infty} \lambda^k x(t-k), \quad (10.31')$$

$$D\tilde{\varepsilon}(t) = \sigma^2 \frac{1 - \lambda}{1 + \lambda}. \quad (10.32')$$

Кроме того, в этом случае легко непосредственно проверить справедливость следующего полезного рекуррентного соотношения

$$\hat{f}(t) = \lambda \hat{f}(t-1) + (1 - \lambda)x(t). \quad (10.33)$$

10.3.3 Подбор порядка аппроксимирующего полинома с помощью метода последовательных разностей

Реализация *алгоритмических* методов выделения неслучайной составляющей временного ряда, и в частности методов скользящего среднего, связана с необходимостью подбора порядка p локально-аппроксимирующего полинома. Эта же задача возникает и при реализации *аналитических* методов выделения неслучайной составляющей (см. (10.18)). При решении этой задачи широко используется так называемый *метод последовательных разностей* членов анализируемого временного ряда, который основан на следующем математическом факте: *если анализируемый временной ряд $x(t)$ содержит в качестве своей неслучайной составляющей алгебраический полином $f(t) = \theta_0 + \theta_1 t + \dots + \theta_p t^p$ порядка (степени) p , то переход к последовательным разностям⁹ $x(1), x(2), \dots, x(N)$, повторенный $p + 1$ раз (то есть переход к последовательным разностям порядка $p + 1$), исключает неслучайную составляющую (включая константу θ_0), оставляя элементы, выражающиеся только через остаточную случайную компоненту $\varepsilon(t)$.*

Продemonстрируем справедливость этого утверждения на простых примерах.

⁹Если имеется ряд чисел $x(1), x(2), \dots, x(N)$, то последовательные разности этого ряда обозначаются с помощью $\Delta x(t) = x(t) - x(t-1)$, $t = 2, \dots, N$. Последовательные разности 2-го порядка — это разности от последовательных разностей, то есть $\Delta^2 x(t) = \Delta(\Delta x(t)) = \Delta x(t) - \Delta x(t-1) = (x(t) - x(t-1)) - (x(t-1) - x(t-2)) = x(t) - 2x(t-1) + x(t-2)$. Аналогично определяется последовательная разность любого (k -го, $k \geq 3$) порядка: $\Delta^k x(t) = \Delta(\Delta^{k-1} x(t))$. Нетрудно показать (например, методом индукции), что

$$\Delta^k x(t) = x(t) - C_k^1 x(t-1) + C_k^2 x(t-2) - \dots + (-1)^k x(t-k), \quad (10.34)$$

где C_k^j — это число сочетаний из k элементов по j .

1) Пусть $p = 1$, то есть

$$x(t) = \theta_0 + \theta_1 t + \varepsilon(t), \quad t = 1, 2, \dots, N.$$

Тогда

$$\Delta x(t) = x(t) - x(t-1) = \theta_1 + (\varepsilon(t) - \varepsilon(t-1)),$$

$$\Delta^2 x(t) = \Delta x(t) - \Delta x(t-1) = \varepsilon(t) - 2\varepsilon(t-1) + \varepsilon(t-2).$$

Мы видим, что если неслучайная составляющая $f(t)$ временного ряда выражается *линейной* функцией времени (то есть является алгебраическим полиномом времени порядка $p = 1$), то последовательность *вторых* разностей $\Delta^2 x(t)$ анализируемого временного ряда уже не будет содержать неслучайной составляющей, а может быть представлена только случайными остатками $\tilde{\varepsilon}(t)$, связанными с исходными остатками $\varepsilon(t)$ соотношением

$$\tilde{\varepsilon}(t) = \varepsilon(t) - 2\varepsilon(t-1) + \varepsilon(t-2), \quad t = 3, 4, \dots, N.$$

2) Пусть $p = 2$, то есть

$$x(t) = \theta_0 + \theta_1 t + \theta_2 t^2 + \varepsilon(t), \quad t = 1, 2, \dots, N.$$

Тогда

$$\begin{aligned} \Delta x(t) &= \theta_0 + \theta_1 + \theta_2 t^2 - \theta_0 - \theta_1(t-1) - \theta_2(t-1)^2 + (\varepsilon(t) - \varepsilon(t-1)) \\ &= \theta_1 - \theta_2 + 2\theta_2 t + (\varepsilon(t) - \varepsilon(t-1)), \end{aligned}$$

$$\Delta^2 x(t) = \Delta x(t) - \Delta x(t-1) = 2\theta_2 + (\varepsilon(t) - 2\varepsilon(t-1) + \varepsilon(t-2)),$$

$$\begin{aligned} \Delta^3 x(t) &= \Delta^2 x(t) - \Delta^2 x(t-1) = \varepsilon(t) - 3\varepsilon(t-1) + 3\varepsilon(t-2) - \varepsilon(t-3), \\ t &= 4, 5, \dots, N. \end{aligned}$$

Мы вновь получили подтверждение сформулированного выше математического факта: при неслучайной составляющей временного ряда, выраженной в виде алгебраического полинома времени t порядка $p = 2$, последовательность третьих разностей $\Delta^3 x(t)$ уже не будет содержать неслучайной составляющей, а будет представлена только в терминах случайных остатков $\varepsilon(t)$ разложения (10.2).

Теперь рассмотрим общий случай. Пусть

$$x(t) = \theta_0 + \theta_1 t + \dots + \theta_p t^p + \varepsilon(t), \quad t = 1, 2, \dots, N, \quad (10.35)$$

где случайные остатки $\varepsilon(t)$ удовлетворяют требованиям (10.26).

Последовательно переходя к разностям порядка $p+1$, с учетом (10.34) получаем:

$$\begin{aligned} \Delta^{p+1} x(t) &= \varepsilon(t) - C_{p+1}^1 \varepsilon(t-1) + \\ &+ C_{p+1}^2 \varepsilon(t-2) - \dots + (-1)^{p+1} \varepsilon(t-p-1), \\ t &= p+2, p+3, \dots, N. \end{aligned} \quad (10.36)$$

Вычислим среднее значение и дисперсию членов последовательности $\Delta^{p+1}x(t)$:

$$\begin{aligned} \mathbf{E}(\Delta^{p+1}x(t)) &= 0, \\ \mathbf{D}(\Delta^{p+1}x(t)) &= \mathbf{D}x(t)[1 + (C_{p+1}^1)^2 + (C_{p+1}^2)^2 + \dots + (C_{p+1}^p)^2 + 1] = \\ &= \sigma^2 C_{2(p+1)}^{p+1}. \end{aligned} \quad (10.37)$$

При получении выражения для $\mathbf{D}(\Delta^{p+1}x(t))$ мы воспользовались взаимной некоррелированностью исходных остатков $\varepsilon(1), \varepsilon(2), \dots, \varepsilon(n)$, а также тождеством

$$1 + (C_{p+1}^1)^2 + \dots + (C_{p+1}^p)^2 + 1 = C_{2(p+1)}^{p+1}.$$

Для обоснования последнего тождества заметим, что из разложения

$$\begin{aligned} (1+z)^{p+1}(1+z)^{p+1} &= (1 + C_{p+1}^1 z + \dots + C_{p+1}^p z^p + C_{p+1}^{p+1} z^{p+1}) \times \\ &\times (1 + C_{p+1}^1 z + \dots + C_{p+1}^p z^p + z^{p+1}) \end{aligned}$$

следует, что коэффициент при z^{p+1} подсчитывается по формуле

$$1 \cdot C_{p+1}^{p+1} + C_{p+1}^1 C_{p+1}^p + C_{p+1}^2 C_{p+1}^{p-1} + \dots + C_{p+1}^{p+1} \cdot 1 = 1 + (C_{p+1}^1)^2 + \dots + (C_{p+1}^p)^2 + 1;$$

но тот же самый коэффициент, подсчитанный с использованием разложения

$$(1+z)^{2(p+1)} = 1 + C_{2(p+1)}^1 z + \dots + C_{2(p+1)}^{2(p+1)-1} z^{2(p+1)-1} + z^{2(p+1)},$$

оказывается равным $C_{2(p+1)}^{p+1}$.

Теперь можно обсудить способ подбора порядка p полинома, представляющего собой неслучайную составляющую $f(t)$ в разложении анализируемого временного ряда $x(t)$. Заметим, прежде всего, что если мы знаем, что среднее значение $\mathbf{E}\xi$ наблюдаемой случайной величины ξ равно нулю, то выборочным аналогом ее дисперсии является величина $\hat{\mu}^2(N) = \sum_{i=1}^N \xi_i^2 / N$, где $\xi_1, \xi_2, \dots, \xi_N$ — наблюдаемые значения этой случайной величины. Если же $\mathbf{E}\xi \neq 0$, то выборочным аналогом дисперсии будет статистика $\sum_{i=1}^N \xi_i^2 / N - (\sum_{i=1}^N \xi_i / N)^2$, так что величина $\hat{\mu}^2(N)$ будет давать в этом случае существенно завышенные оценки для $\mathbf{D}\xi$. Возвращаясь к последовательному переходу к разностям $\Delta^k x(t)$, $k = 1, 2, \dots, p+1$, — в общем случае (10.35), отметим, что при всех $k < p+1$ средние значения этих разностей будут отличны от нуля, так как будут выражаться не только через остатки $\varepsilon(t)$, но и через коэффициенты $\theta_0, \theta_1, \dots, \theta_p$ и степени t . И только начиная с $k = p+1$

(и для всех $k > p + 1$) оказывается справедливой формула (10.37), а следовательно, при $k \geq p + 1$ можно утверждать, что:

$$E(\Delta^k x(t)) = 0 \quad \text{и} \quad \sigma^2 = D(\Delta^k x(t)) / C_{2(p+1)}^{p+1}. \quad (10.37')$$

С учетом этих замечаний можно сформулировать следующее правило подбора порядка сглаживающего полинома p , называемое *методом последовательных разностей*.

Последовательно для $k = 1, 2, \dots$ вычисляем разности $\Delta^k x(t)$ ($t = k + 1, k + 2, \dots, N$), а также величины

$$\hat{\sigma}^2(k) = \frac{\frac{1}{n-k} \sum_{t=k+1}^N (\Delta^k x(t))^2}{C_{2k}^k}. \quad (10.38)$$

Анализируем поведение величины $\hat{\sigma}^2(k)$ в зависимости от k . Учитывая сделанные выше замечания, величина $\hat{\sigma}^2(k)$ как функция k будет демонстрировать явную тенденцию к убыванию до тех пор, пока k не достигнет величины $p + 1$. Начиная с момента $k = p + 1$ величина (10.38) стабилизируется, оставаясь (при дальнейшем увеличении k) приблизительно на одном уровне. Поэтому значение $k = k_0$, начиная с которого величина $\hat{\sigma}^2(k)$ стабилизируется, и будет давать завышенный на единицу искомый порядок сглаживающего полинома, то есть $p = k_0 - 1$.

Для упрощения использования описанного метода в таблице 10.5 приведены значения вспомогательных коэффициентов C_{2k}^k для $k = 1, 2, \dots, 10$.

Таблица 10.5. Значения вспомогательных коэффициентов C_{2k}^k

k	1	2	3	4	5	6	7	8	9	10
C_{2k}^k	2	6	20	70	252	924	3432	12870	48260	184756

Этот метод привлекателен своей простотой, но его практическое применение требует определенной осторожности. Последовательные значения $\hat{\sigma}^2(k)$ не являются независимыми, и часто обнаруживается тенденция их медленного убывания (а иногда возрастания) без видимой сходимости к постоянному значению. Кроме того процесс перехода к разностям имеет тенденцию уменьшать относительное значение любого систематического движения, кроме сезонных эффектов с периодом, близким к временному интервалу, так что сходимость отношения $\hat{\sigma}^2(k)$ не *доказывает*, что ряд первоначально состоял из полинома плюс случайный остаток, а только то, что он может быть приближенно представлен таким образом. Однако для нас этот метод ценен лишь тем, что он дает *верхний предел* порядка полинома p , который целесообразно использовать для элиминирования неслучайной составляющей.

10.4 Модели стационарных временных рядов и их идентификация

В п. 10.2 был введен в рассмотрение *класс стационарных временных рядов*, в рамках которого подбирается, в частности, модель, пригодная для описания поведения случайных остатков $\varepsilon(t)$ анализируемого временного ряда (10.2). В данном пункте мы займемся изучением определенного *набора линейных параметрических моделей* из этого класса, в том числе *методами идентификации этих моделей*, то есть методами статистического оценивания участвующих в описании модели параметров (структурных и неструктурных) по имеющейся в нашем распоряжении траектории (10.1') анализируемого временного ряда.

Таким образом, речь пока идет о моделировании не самих анализируемых временных рядов, а лишь их случайных остатков $\varepsilon(t)$, то есть того, что мы получаем после элиминирования из исходного временного ряда $x(t)$ его неслучайной составляющей (10.17) (см. п. 10.3.2). И наши усилия направлены на построение такой модели случайного остатка, которая позволила бы прогнозировать значения этого остатка по его же значениям в предыдущие моменты времени.

Так что в отличие от прогноза, основанного, например, на классической регрессионной модели (когда в качестве прогнозного значения мы используем значение функции регрессии при заданных значениях объясняющих переменных, *игнорируя значения случайных остатков*, см. п. 6.2, формулу (6.16')), в прогнозе временных рядов существенно используются *взаимозависимость и прогноз самих случайных остатков*.

Условимся об обозначениях, принятых в данном пункте. Поскольку представленные ниже модели временных рядов предназначены для описания поведения *случайных остатков*, то анализируемый (моделируемый) временной ряд мы будем обозначать с помощью $\varepsilon(t)$ и, соответственно, полагать, что его среднее значение (математическое ожидание) тождественно, то есть при всех значениях t , равно нулю, то есть $E\varepsilon(t) \equiv 0$. Временные последовательности, образующие так называемый «белый шум», будем обозначать с помощью $\delta(t)$. Это значит, что

$$E\delta(t) \equiv 0, \quad (10.39)$$

$$E(\delta(t)\delta(t \pm \tau)) = \begin{cases} \sigma_0^2 & \text{при } \tau = 0 \\ 0 & \text{при } \tau \neq 0, \end{cases} \quad (10.40)$$

причем величина дисперсии σ_0^2 не зависит от t .

Описание и анализ рассмотренных ниже моделей могут формулироваться в терминах *общего линейного процесса*, представимого в ви-

де взвешенной суммы настоящего и прошлых значений белого шума, а именно:

$$\varepsilon(t) = \delta(t) + \beta_1 \delta(t-1) + \beta_2 \delta(t-2) + \dots = \sum_{k=0}^{\infty} \beta_k \delta(t-k), \quad (10.41)$$

$$\text{где } \beta_0 = 1 \text{ и } \sum_{k=0}^{\infty} \beta_k^2 < \infty.$$

Таким образом, белый шум можно рассматривать как серию импульсов, в широком классе реальных ситуаций генерирующих случайные остатки анализируемого временного ряда.

Существует эквивалентная соотношению (10.41) запись общего линейного процесса, при которой анализируемый временной ряд $\varepsilon(t)$ получается в виде классической линейной модели множественной регрессии, когда в роли объясняющих переменных выступают значения самого временного ряда во все прошлые моменты времени (включая и бесконечно удаленные), то есть

$$\varepsilon(t) = \pi_1 \varepsilon(t-1) + \pi_2 \varepsilon(t-2) + \dots + \delta(t) = \sum_{k=1}^{\infty} \pi_k \varepsilon(t-k) + \delta(t). \quad (10.41')$$

При этом весовые коэффициенты $\pi_1, \pi_2, \dots$ связаны определенными условиями, обеспечивающими стационарность ряда $\varepsilon(t)$. Переход от формы (10.41') к форме (10.41) осуществляется с помощью последовательной (и повторенной бесконечное число раз!) подстановки в правую часть (10.41') вместо $\varepsilon(t-1), \varepsilon(t-2), \dots$ их выражений, вычисленных в соответствии с (10.41') для моментов времени $t-1, t-2$ и т. д.

Введем в рассмотрение, наряду с общим линейным процессом вида (10.41) или (10.41'), процесс *смешанного* типа, в представлении которого присутствуют как авторегрессионные члены самого процесса, так и скользящее суммирование элементов белого шума:

$$\varepsilon(t) = \sum_{k=1}^p \pi_k \varepsilon(t-k) + \delta(t) + \sum_{j=1}^q \beta_j \delta(t-j). \quad (10.41'')$$

Мы будем подразумевать, что p и q могут принимать и бесконечные значения, а также то, что в частных случаях некоторые (или даже все) коэффициенты π или β равны нулю. Нам понадобится в дальнейшем следующая специальная форма представления спектральной плотности $p(\tilde{\omega})$ ($\tilde{\omega} = \omega/2\pi$, так что $0 \leq \tilde{\omega} \leq 1/2$), справедливая для процессов типа (10.41'') (обоснование такого представления можно найти, например, в книге [Бокс, Дженкинс, (1974), с. 67, 98–99]:

$$p(\tilde{\omega}) = 2\sigma_0^2 \frac{|B(e^{-i2\pi\tilde{\omega}})|^2}{|A(e^{-i2\pi\tilde{\omega}})|^2}, \quad 0 \leq \tilde{\omega} \leq \frac{1}{2}, \quad (10.9a)$$

где многочлены $A(z)$ и $B(z)$ определяются коэффициентами, соответственно, π_k ($k = 1, 2, \dots, p$) и β_j ($j = 1, 2, \dots, q$):

$$A(z) = 1 - \sum_{k=1}^p \pi_k z^k, \quad B(z) = 1 + \sum_{j=1}^q \beta_j z^j,$$

$\sigma_0^2 = D\delta(t)$, а $i = \sqrt{-1}$ — мнимая единица.

Идентификация моделей. В общем смысле под идентификацией модели понимается и определение ее структурных параметров p и q в соотношении (10.41''), и статистическая оценка (при заданных целочисленных значениях p и q) ее неструктурных параметров π_k ($k = 1, 2, \dots, p$) и β_j ($j = 1, 2, \dots, q$). Рекомендации по подбору значений структурных параметров анализируемых моделей и общие сведения о современных методах статистического оценивания ее неструктурных параметров будут приведены в п. 10.4.4. В контексте же рассмотрения *каждой конкретной модели* (AR(1), AR(2), CC(1), CC(2) и ARCC(1;1)) мы обсудим лишь один из возможных методов оценивания неструктурных параметров — *метод моментов*, который, как это следует из теории статистического оценивания, в общем случае уступает в эффективности методам полного и условного максимального правдоподобия. Однако, мы руководствовались, при этом, двумя мотивами: в о - п е р в ы х, методы максимального правдоподобия требуют постулирования *нормальности* элементов белого шума $\beta(t)$, а в о - в т о р ы х, их численная реализация из-за сложности «спрятана» в соответствующих модулях эконометрического программного обеспечения пакетов E-views, STATA, SPSS и др., в то время как метод моментов достаточно прозрачен и прост в своей численной реализации.

Теперь перейдем к рассмотрению *конкретных* линейных моделей стационарного временного ряда.

10.4.1 Модели авторегрессии порядка p (AR(p)-модели)¹⁰

Начнем описание этого параметрического семейства моделей с анализа простейших частных случаев.

Модель авторегрессии 1-го порядка — AR(1) (марковский процесс). Эта модель представляет собой простейший вариант линейного авторегрессионного процесса типа (10.41'), когда все коэффициенты π_j кроме первого равны нулю. Соответственно, она может быть определена выражением

$$\varepsilon(t) = \alpha \varepsilon(t-1) + \delta(t), \quad (10.42)$$

¹⁰В англоязычной специальной литературе эти модели известны как *AutoRegressive processes of order p* и кратко обозначаются *AR(p)-models*.

где α — некоторый числовой коэффициент, не превосходящий по абсолютной величине единицу ($|\alpha| < 1$), а $\delta(t)$ — последовательность случайных величин, образующая белый шум (то есть последовательность случайных величин, удовлетворяющая соотношениям (10.39)–(10.40)).

Внимательный читатель вспомнит, что мы уже имели дело с этой моделью. Ведь именно соотношением (10.42) были связаны между собой регрессионные остатки в *обобщенной модели множественной регрессии с автокоррелированными остатками* (см. п. 5.4, соотношения (5.31)–(5.32), в которых роль параметра α выполнял параметр ρ). В специальной литературе последовательности $\varepsilon(t)$, удовлетворяющие соотношению (10.42), называют также *марковским процессом*. Мы не будем повторять здесь выкладки (5.33)–(5.35), из которых следовало:

$$\mathbf{E}\varepsilon(t) \equiv 0, \quad (10.43a)$$

$$r(\varepsilon(t), \varepsilon(t \pm k)) = \alpha^k, \quad (10.43б)$$

$$\mathbf{D}\varepsilon(t) = \frac{\sigma_0^2}{1 - \alpha^2}, \quad (10.43в)$$

$$\text{cov}(\varepsilon(t), \varepsilon(t \pm k)) = \alpha^k \mathbf{D}\varepsilon(t). \quad (10.43г)$$

Проанализируем некоторые свойства временной последовательности, представимой в виде (10.42). Из представления $\varepsilon(t)$ в виде (5.33) следует, что $\varepsilon(t)$ зависит от $\delta(t)$ и всех предшествующих δ , но не зависит от будущих значений δ , то есть от $\delta(t+1)$, $\delta(t+2)$ и т. д.

Заметим, что формула (5.33) представляет $\varepsilon(t)$ как скользящее среднее с бесконечной длиной отрезка усреднения, но с весами, дающими в сумме не единицу, а $1/(1 - \alpha)$. Может показаться, что случайная величина $\varepsilon(t)$, представляемая в (5.33) как взвешенная сумма большого числа случайных слагаемых, должна быть распределена (в соответствии с центральной предельной теоремой), *нормально*. Однако одно из условий ц.п.т. нарушено: веса в этом среднем — величины различного порядка.

Одно важное следствие (10.43в) состоит в том, что если величина $|\alpha|$ близка к единице, то дисперсия $\varepsilon(t)$ будет намного больше дисперсии $\delta(t)$. А это значит (с учетом того, что в соответствии с (10.43б) $\alpha = r(\varepsilon(t), \varepsilon(t \pm 1)) = r(1)$), что если соседние значения ряда $\varepsilon(t)$ сильно коррелированы, то ряд довольно слабых возмущений $\delta(t)$ будет порождать размашистые колебания остатков $\varepsilon(t)$.

На графике марковский процесс представляется осцилляциями более или менее регулярного типа (см. рис. 10.8). Может быть вычислено

среднее расстояние $d_{\text{ср}}$ между «пиками»¹¹ такой осциллирующей ломаной (см. [Кендэл (1981), с. 78]):

$$d_{\text{ср}}(\alpha) = \frac{2\pi}{\arccos \left[-\frac{1}{2}(1 - \alpha) \right]}.$$

В частности, для последовательности независимых и одинаково (0; 1)-нормально распределенных случайных величин (то есть при $\alpha = 0$) это среднее расстояние оказывается равным 3,0 (так как $\arccos(-\frac{1}{2}) = \frac{2}{3}\pi$); для примеров, представленных на рис. 10.8, эти средние значения между пиками будут, соответственно, равны $2\pi/\arccos(-0,1) = \frac{360^\circ}{96^\circ} = 3,75$ (для $\alpha = 0,8$) и $2\pi/\arccos(-0,9) = 360^\circ/154^\circ = 2,34$ (для $\alpha = -0,8$).

Исследуем теперь, опираясь на соотношения (10.43), основные характеристики процесса авторегрессии 1-го порядка и попробуем оценить параметры этой модели (α и σ_0^2) по имеющейся в нашем распоряжении реализации (10.1') анализируемого временного ряда $x(t)$.

Условие стационарности ряда (10.42) определяется требованием к коэффициенту α :

$$|\alpha| < 1, \quad (10.44)$$

или, что то же, корень z_0 уравнения $1 - \alpha z = 0$ должен быть по абсолютной величине больше единицы¹².

Автокорреляционная функция марковского процесса определяется соотношением (10.436).

Отсюда, в частности, следует простая вероятностная интерпретация параметра α :

$$\alpha = r(1) = r(\varepsilon(t), \varepsilon(t \pm 1)),$$

то есть значение α определяет величину коэффициента парной корреляции между двумя соседними членами ряда $\varepsilon(t)$.

¹¹Диагностика модели по числу пиков в анализируемом временном ряду или по среднему расстоянию между пиками основана на том же подходе, который был использован в п. 10.3.1 при построениях критериев для проверки гипотезы о том, что исследуемый временной ряд представляет собой последовательность независимых и одинаково распределенных случайных величин (см. (10.12)). Обращаем внимание читателя на то, что наличие «пика» определяется значением временного ряда, большим, чем два соседних его значения, и что при таком понимании пика среднее расстояние между пиками будет всегда несколько меньше периода (если таковой существует) или псевдопериода анализируемой временной последовательности.

¹²Уравнение $1 - \alpha z = 0$ представляет собой частный случай характеристического уравнения общего линейного процесса авторегрессии. Поскольку в общем случае у характеристического уравнения допускается наличие комплексных корней, то вместо требования к абсолютным величинам корней условие стационарности чаще формулируют так: «все корни соответствующего характеристического уравнения должны лежать вне единичного круга» (см. ниже описание общей модели авторегрессии p -го порядка).

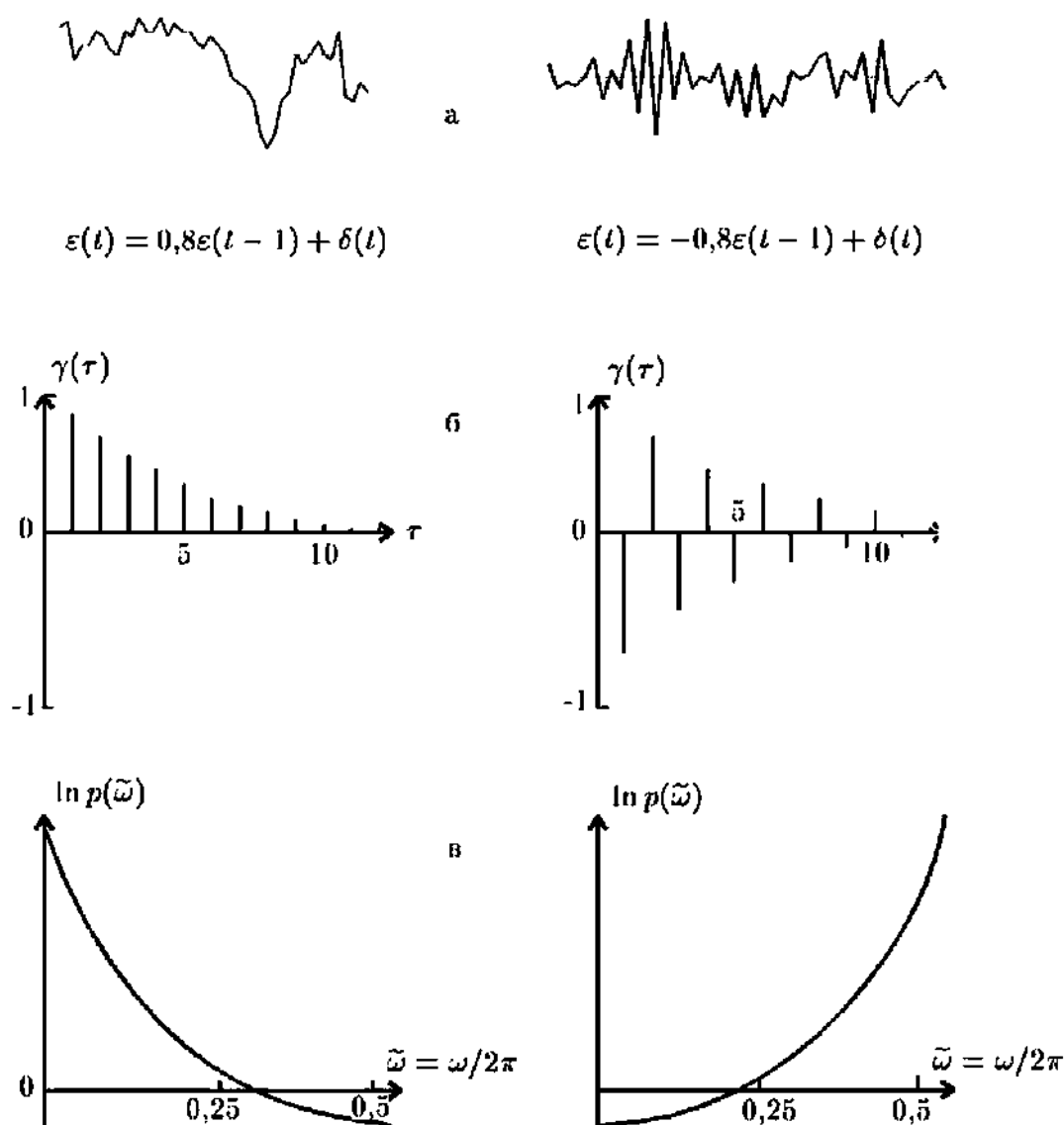


Рис. 10.8. Реализация процессов авторегрессии первого порядка (а), соответствующие им автокорреляционные функции (б) и логарифмы плотности (в)

Из (10.43б) видно, что степень тесноты корреляционной связи между членами последовательности (10.42) экспоненциально убывает по мере их взаимного удаления друг от друга во времени.

Частная автокорреляционная функция $r_{\text{част}}(\tau) = r(\varepsilon(t), \varepsilon(t+\tau) | \varepsilon(t+1) = \varepsilon(t+2) = \dots = \varepsilon(t+\tau-1) = 0)$ может быть подсчитана с помощью формул (10.7)–(10.8). Непосредственное вычисление по этим формулам дает следующий простой результат: значения частной корреляционной функции $r_{\text{част}}(\tau)$ равны нулю для всех $\tau = 2, 3, \dots$. Это свойство частной автокорреляции может быть использовано при подборе модели: если вычисленные по оцененным «невязкам» $\varepsilon(t) = x(t) - \hat{f}(t)$ выборочные частные корреляции $\hat{r}_{\text{част}}(\tau)$ статистически незначимо отличаются от нуля при $\tau = 2, 3, \dots$, то использование модели авторегрессии 1-го порядка для описания поведения случайных остатков анализируемого

временного ряда считается не противоречащим исходным статистическим данным.

Спектральная плотность $p(\tilde{\omega})$ марковского процесса (10.42) может быть подсчитана либо по формуле (10.9а), либо по формуле (10.9) с учетом известного вида автокорреляционной функции $r(\tau) = r(\varepsilon(t), \varepsilon(t \pm \tau)) = \alpha^\tau$:

$$p(\tilde{\omega}) = \frac{2\sigma_0^2}{1 + \alpha^2 - 2\alpha \cos(2\pi\tilde{\omega})},$$

$$0 \leq \tilde{\omega} \leq \frac{1}{2} \quad (\tilde{\omega} = \omega/2\pi).$$

На рис. 10.8 видно, что в случае большого положительного значения параметра α (при $\alpha = 0,8$) соседние значения ряда $\varepsilon(t)$ близки друг к другу по величине, автокорреляционная функция убывает (оставаясь положительной) по экспоненциальному закону, а в спектре преобладают низкие частоты (а значит, среднее расстояние между пиками осциллирующего временного ряда $\varepsilon(t)$ достаточно велико). При большом по абсолютной величине, но отрицательном значении параметра α (при $\alpha = -0,8$) ряд быстро осциллирует (в спектре преобладают высокие частоты), а график автокорреляционной функции экспоненциально спадает до нуля с попеременным изменением знака.

Идентификация модели, то есть статистическое оценивание ее параметров α и σ_0^2 по имеющейся реализации (10.1') исходного анализируемого временного ряда $x(t)$ (а не его остатков, которые являются ненаблюдаемыми), основана на соотношениях (10.43) (при $k = 0$ и $k = 1$) и может быть осуществлена, например, с помощью метода моментов. Для этого следует предварительно решить задачу выделения неслучайной составляющей $\hat{f}(t)$ (см. п. 10.3), что позволит оперировать в дальнейшем остатками («невязками»)

$$\hat{\varepsilon}(t) = x(t) - \hat{f}(t).$$

Затем подсчитывается выборочная дисперсия $\hat{\gamma}(0)$ остатков по формуле

$$\hat{\gamma}(0) = \frac{1}{N} \sum_{t=1}^N (\hat{\varepsilon}(t) - \bar{\varepsilon})^2, \quad (10.45)$$

где $\bar{\varepsilon} = (\sum_{t=1}^N \hat{\varepsilon}(t))/N$.

Оценку $\hat{\alpha}$ параметра α получаем с помощью формулы (10.43б), подставляя в нее вместо коэффициента корреляции $r(\varepsilon(t), \varepsilon(t \pm 1))$ его выборочное значение $\hat{r}(\varepsilon(t), \varepsilon(t + 1))$, то есть

$$\hat{\alpha} = \frac{\frac{1}{N-1} \sum_{t=1}^{N-1} (\hat{\varepsilon}(t) - \bar{\varepsilon})(\hat{\varepsilon}(t+1) - \bar{\varepsilon})}{\hat{\gamma}(0)}. \quad (10.46)$$

Наконец, оценка $\hat{\sigma}_0^2$ параметра σ_0^2 основана на соотношении (10.43в), в котором величины $D\varepsilon(t)$ и α заменяются их оценками, соответственно, $\hat{\gamma}(0)$ и $\hat{\alpha}$:

$$\hat{\sigma}_0^2 = (1 - \hat{\alpha}^2)\hat{\gamma}(0).$$

Заметим, что в большинстве методов сглаживания из самой техники построения $\hat{f}(t)$ автоматически следует равенство нулю среднего значения «невязок» (то есть $\bar{\varepsilon} = (\sum_{t=1}^N \varepsilon(t))/N = 0$). В этих случаях в формулах (10.45) и (10.46) значение $\bar{\varepsilon}$ следует положить равным нулю.

Обращаем внимание читателя на следующий факт: *прямая* оценка значений автокорреляционной функции (то есть оценка ее значений с помощью формулы (10.6')) при *реалистичных* соотношениях N и τ дает, как правило, неудовлетворительный результат (хотя и является состоятельной). В частности, в то время, как теоретическая коррелограмма анализируемого ряда затухает, построенная с помощью (10.6') выборочная коррелограмма может сохранять колебательный (*незатухающий*) эффект. Поэтому, в частности, при анализе марковского процесса элементы его автокорреляционной функции $r(\tau)$ лучше оценивать по формуле $\hat{r}(\tau) = \hat{\alpha}^\tau$.

Модели авторегрессии 2-го порядка — АР(2) (процессы Юла). Эта модель, как и АР(1), представляет собой частный случай авторегрессионного процесса типа (10.41'), когда все коэффициенты π_j в правой части (10.41'), *кроме первых двух*, равны нулю. Соответственно, она может быть определена выражением

$$\varepsilon(t) = \alpha_1 \varepsilon(t-1) + \alpha_2 \varepsilon(t-2) + \delta(t), \quad (10.47)$$

где последовательность случайных величин $\delta(1), \delta(2), \dots$ образует белый шум (см. выше).

Последовательно подставляя (бесконечное число раз) в правую часть (10.47) вместо $\varepsilon(t-1), \varepsilon(t-2), \dots$ их выражения, вычисленные по формуле (10.47), мы убедимся в том, что, как и в АР(1)-модели, $\varepsilon(t)$ зависит от $\delta(t), \delta(t-1), \delta(t-2), \dots$, но *не зависит от будущих значений δ* , то есть от $\delta(t+1), \delta(t+2), \dots$.

Умножая обе части (10.47) по очереди на $\varepsilon(t-1)$ и $\varepsilon(t-2)$ и беря математические ожидания от двух полученных таким образом соотношений, получаем систему уравнений, связывающих между собой параметры модели α_1 и α_2 с дисперсией $\gamma(0) = D\varepsilon(t)$ и первыми двумя ковариациями $\gamma(1) = E[\varepsilon(t)\varepsilon(t-1)]$ и $\gamma(2) = E[\varepsilon(t)\varepsilon(t-2)]$ анализируемого процесса $\varepsilon(t)$:

$$\begin{cases} \gamma(1) - \alpha_1 \gamma(0) - \alpha_2 \gamma(1) = 0, \\ \gamma(2) - \alpha_1 \gamma(1) - \alpha_2 \gamma(0) = 0, \end{cases}$$

или, после деления обоих уравнений на $\gamma(0)$:

$$\begin{cases} r(1) - \alpha_1 - \alpha_2 r(1) = 0, \\ r(2) - \alpha_1 r(1) - \alpha_2 = 0. \end{cases} \quad (10.48)$$

Разрешая эту систему относительно α_1 и α_2 , имеем:

$$\begin{aligned} \alpha_1 &= \frac{r(1)(1 - r(2))}{1 - r^2(1)}, \\ \alpha_2 &= \frac{r(2) - r^2(1)}{1 - r^2(1)}. \end{aligned} \quad (10.49)$$

Уравнения (10.48) могут быть решены относительно $r(1)$ и $r(2)$:

$$\begin{aligned} r(1) &= \frac{\alpha_1}{1 - \alpha_2}, \\ r(2) &= \alpha_2 + \frac{\alpha_1^2}{1 - \alpha_2}. \end{aligned} \quad (10.50)$$

Обобщая использованный прием, домножим обе части (10.47) на $\varepsilon(t - k)$ (k — любое целое число, большее нуля) и возьмем математическое ожидание от всех членов получившегося соотношения:

$$\gamma(k) = \alpha_1 \gamma(k - 1) + \alpha_2 \gamma(k - 2),$$

или (после деления на $\gamma(0)$):

$$r(k) = \alpha_1 r(k - 1) + \alpha_2 r(k - 2). \quad (10.51)$$

Рекуррентная формула (10.51) используется для вычисления любого значения автокорреляционной функции $r(\tau)$ лишь по двум ее первым значениям $r(1)$ и $r(2)$ (в модели авторегрессии 1-го порядка аналогичную роль играет формула $r(\tau) = r^{\tau}(1)$).

Наконец, для того, чтобы получить соотношение, связывающее между собой параметры $\gamma(0) = \mathbf{D}\varepsilon(t)$ и $\sigma_0^2 = \mathbf{D}\delta(t)$, домножим обе части (10.47) на $\varepsilon(t)$:

$$\varepsilon^2(t) = \alpha_1 \varepsilon(t - 1) \varepsilon(t) + \alpha_2 \varepsilon(t - 2) \varepsilon(t) + \delta(t) [\alpha_1 \varepsilon(t - 1) + \alpha_2 \varepsilon(t - 2) + \delta(t)].$$

После операции усреднения (с учетом того, что $\delta(t)$ и $\varepsilon(t - k)$ взаимно некоррелированы при $k \geq 1$) получаем:

$$\gamma(0) = \alpha_1 \gamma(1) + \alpha_2 \gamma(2) + \sigma_0^2,$$

откуда

$$\sigma_0^2 = \gamma(0) - \alpha_1 \gamma(1) - \alpha_2 \gamma(2), \quad (10.52)$$

или в терминах $\gamma(0) = \mathbf{D}\varepsilon(t)$, α_1 , α_2 , $r(1)$ и $r(2)$:

$$\sigma_0^2 = \gamma(0)(1 - \alpha_1 r(1) - \alpha_2 r(2)) = \gamma(0) \frac{1 + \alpha_2}{1 - \alpha_2} [(1 - \alpha_2)^2 - \alpha_1^2]. \quad (10.52')$$

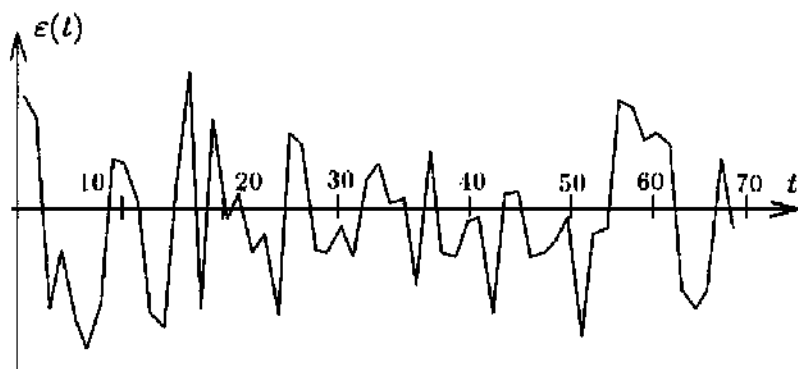


Рис. 10.9. Временной ряд, генерируемый моделью авторегрессии второго порядка

На рис. 10.9 изображен временной ряд, генерируемый моделью авторегрессии 2-го порядка при $\alpha_1 = 0,75$ и $\alpha_2 = -0,50$. Чисто внешне его трудно отличить от рядов, генерируемых, например, моделью авторегрессии 1-го порядка: мы наблюдаем приблизительно такие же осцилляции так называемого «псевдопериодического» характера. Можно было бы попытаться отличать реализации АР(2)-модели по среднему расстоянию $d_{\text{ср}}(\alpha_1, \alpha_2)$ между пиками ряда (наличие «пика», напомним, определяется значением $\varepsilon(t)$, большим, чем два соседних значения), которое подсчитывается по значениям α_1 и α_2 по формуле (см. [Кендэл (1981), с. 81]):

$$d_{\text{ср}}(\alpha_1, \alpha_2) = \frac{2\pi}{\arccos \left[-\frac{1}{2}(1 - \alpha_1 + \alpha_2) \right]}.$$

Однако диагностика модели, основанная на среднем расстоянии между пиками, не очень чувствительна. Так, для белого шума это среднее расстояние, как мы видели, оказывается равным трем, для марковских процессов оно варьирует в диапазоне от 3 до 4, а для процессов Юла — между 4 и 6. В частности, для ряда, изображенного на рис. 10.9, среднее расстояние между пиками

$$d(\alpha_1; \alpha_2) = d(0,75; -0,50) = \frac{2\pi}{\arccos(0,125)} = 4,34.$$

Более обстоятельные и обоснованные выводы о типе модели, генерирующей анализируемый временной ряд, делаются, конечно, на основании комплексного анализа его основных характеристик — автокорреляционной функции, спектра и т. п. Условия стационарности ряда (10.47) могут быть получены из естественных требований к абсолютным величинам автокорреляций $r(1)$ и $r(2)$, выраженным в терминах α_1 и α_2 (см. соотношения (10.50)):

$$\begin{cases} -1 < \frac{\alpha_1}{1 - \alpha_2} < 1 \\ -1 < \alpha_2 + \frac{\alpha_1^2}{1 - \alpha_2} < 1. \end{cases} \quad (10.53)$$

Отсюда получаем следующие необходимые и достаточные условия стационарности ряда (10.47):

$$\begin{cases} |\alpha_1| < 2, \\ \alpha_2 < 1 - |\alpha_1|. \end{cases} \quad (10.53')$$

В рамках *общей теории* моделей авторегрессии те же самые условия стационарности получаются из требования, чтобы *все корни соответствующего характеристического уравнения лежали бы вне единичного круга*. Характеристическое уравнение для модели авторегрессии 2-го порядка имеет вид:

$$1 - \alpha_1 z - \alpha_2 z^2 = 0.$$

Кстати, из условий стационарности (10.53') и выражений (10.50) для $r(1)$ и $r(2)$ следует, что далеко не всякая пара значений $r(1)$ и $r(2)$, по модулю меньших единицы, годится для описания поведения первых двух членов автокорреляционной функции процесса Юла. Используя упомянутые соотношения, можно показать, что допустимые значения $r(1)$ и $r(2)$ для стационарного процесса АР(2) должны лежать в области

$$\begin{cases} -1 < r(1) < 1, \\ -1 < r(2) < 1, \\ r(2) > 2r^2(1) - 1. \end{cases}$$

Автокорреляционная функция процесса Юла подсчитывается следующим образом. Два первых значения $r(1)$ и $r(2)$ определены соотношениями (10.50), а значения $r(\tau)$ для $\tau = 3, 4, \dots$ вычисляются с помощью рекуррентного соотношения (10.51).

На рис. 10.10 представлен график автокорреляционной функции для рассмотренного выше примера модели авторегрессии 2-го порядка (см. также рис. 10.9).

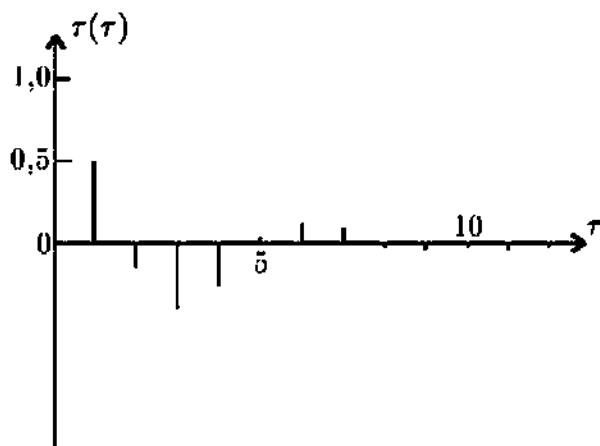


Рис. 10.10. Автокорреляционная функция процесса авторегрессии второго порядка $\varepsilon(t) = 0,75\varepsilon(t-1) - 0,50\varepsilon(t-2) + \delta(t)$

Частная автокорреляционная функция $r_{\text{част}}(\tau) = r(\varepsilon(t), \varepsilon(t + \tau) \mid \varepsilon(t + 1) = \varepsilon(t + 2) = \dots = \varepsilon(t + \tau - 1) = 0)$ временного ряда, сгенерированного моделью авторегрессии 2-го порядка, обладает следующим отличительным свойством:

$$r_{\text{част}}(\tau) = 0 \quad \text{при всех} \quad \tau = 3, 4, \dots$$

Этот результат доказывается прямым вычислением частных коэффициентов корреляции по формулам (10.7)–(10.8) для $\tau = 2, 3$ и с использованием рекуррентных формул (3.26') для $\tau > 3$. Действительно, подставляя в формулу (10.7) значения $r(1)$ и $r(2)$, выраженные через α_1 и α_2 с помощью (10.50), получаем

$$r_{\text{част}}(2) = \frac{r(2) - r^2(1)}{1 - r^2(1)} = \alpha_2.$$

Далее, числитель в формуле (10.8) для $r_{\text{част}}(3)$ приводится в нашем случае к выражению, в свою очередь, числитель которого равен

$$(1 - r^2(1))(r(3) - r(2)r(1)) - (r(1) - r(2)r(1))(r(2) - r^2(1)).$$

Равенство нулю последнего выражения устанавливается на базе соотношений (10.50) и (10.51).

Спектральная плотность $p(\tilde{\omega})$ процесса Юла может быть вычислена с помощью общей формулы (10.9а):

$$\begin{aligned} p(\tilde{\omega}) &= \frac{2\sigma_0^2}{|1 - \alpha_1 e^{-i2\pi\tilde{\omega}} - \alpha_2 e^{-i4\pi\tilde{\omega}}|^2} = \\ &= \frac{2\sigma_0^2}{1 + \alpha_1^2 + \alpha_2^2 - 2\alpha_1(1 - \alpha_2) \cos(2\pi\tilde{\omega}) - 2\alpha_2 \cos(4\pi\tilde{\omega})}, \quad (10.54) \\ &0 \leq \tilde{\omega} \leq \frac{1}{2}. \end{aligned}$$

На рис. 10.11 представлен график спектральной плотности рассмотренного ранее (см. рис. 10.9) примера модели авторегрессии 2-го порядка. Из рисунка видно, что максимум спектра приходится приблизительно на значение частоты $\tilde{\omega}_0 = 0,14 \sim 0,15$. А это значит, что средний кажущийся («псевдо-») период T_0 нашего временного ряда близок к 6 (так как $T_0 = 1/\tilde{\omega}_0$). Кстати, обращаем внимание читателя на несовпадение (как и следовало ожидать!) вычисленного ранее среднего расстояния между пиками этого ряда ($d_{\text{ср}}(0,75; -0,50) = 4,34$) и его псевдопериода T_0 .

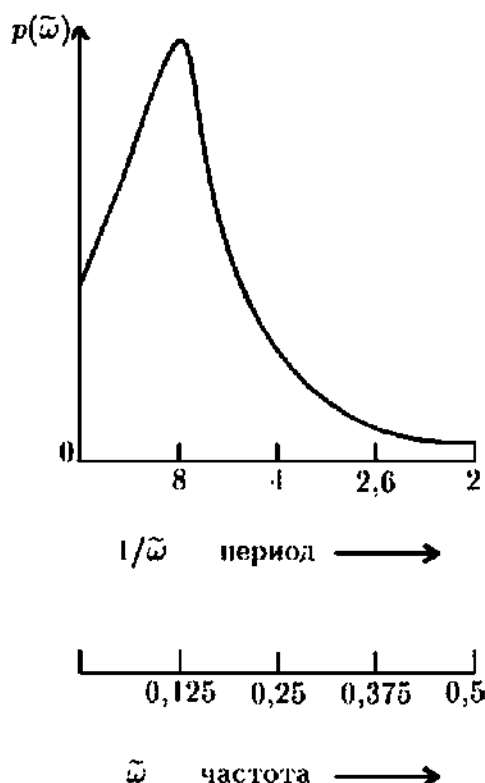


Рис. 10.11. Спектральная плотность процесса авторегрессии второго порядка

Идентификация модели авторегрессии 2-го порядка основана на соотношениях (10.49)–(10.52'), связывающих между собой неизвестные параметры модели α_1, α_2 и σ_0^2 со значениями различных моментов «наблюдаемого» временного ряда $\varepsilon(t)$ (с его дисперсией, автокорреляциями и т.п.). Слово «наблюдаемого» взято в кавычки, поскольку в действительности наблюдается анализируемый временной ряд $x(t)$, а «наблюдаемые» значения остатков $\hat{\varepsilon}(t)$ получаются после выделения неслучайной составляющей $\hat{f}(t)$ и образования разностей $\hat{\varepsilon}(t) = x(t) - \hat{f}(t)$.

Итак, по значениям $\hat{\varepsilon}(t)$ прежде всего вычисляются оценки $\hat{\gamma}(0), \hat{r}(1)$ и $\hat{r}(2)$, соответственно, дисперсии $D\varepsilon(t)$ и автокорреляций $r(1)$ и $r(2)$. Это делается с помощью соотношений (10.4'') и (10.6'):

$$\hat{\gamma}(0) = \frac{1}{N} \sum_{t=1}^N (\hat{\varepsilon}(t) - \bar{\varepsilon})^2,$$

$$\hat{r}(k) = \frac{\frac{1}{N-k} \sum_{t=1}^{N-k} (\hat{\varepsilon}(t) - \bar{\varepsilon})(\hat{\varepsilon}(t+k) - \bar{\varepsilon})}{\hat{\gamma}(0)}, \quad k = 1, 2$$

(если из техники построения неслучайной составляющей $\hat{f}(t)$ автоматически следует равенство $\sum_{t=1}^N \hat{\varepsilon}(t) = 0$, то значение $\bar{\varepsilon}$ в этих формулах следует положить равным нулю).

После этого можно получить оценки $\hat{\alpha}_1$ и $\hat{\alpha}_2$ для параметров α_1 и α_2 , подставив в правые части формул (10.49) значения $\hat{r}(1)$ и $\hat{r}(2)$ вместо, соответственно, величин $r(1)$ и $r(2)$. Наконец, оценку параметра σ_0^2 получаем с помощью формулы (10.52'), заменив в ней теоретическую ковариацию $\gamma(0)$ и параметры α_1, α_2 ранее полученными оценками, соответственно, $\hat{\gamma}(0), \hat{\alpha}_1$ и $\hat{\alpha}_2$.

Заметим, что при построении выборочной автокорреляционной функции $\hat{r}(\tau)$ ее значения для $\tau \geq 3$ предпочтительнее вычислять с помощью рекуррентного соотношения (10.51) на базе начальных значений $\hat{r}(1)$ и $\hat{r}(2)$.

Модели авторегрессии p -го порядка — $AP(p)$ ($p \geq 3$). Эти модели, образуя подмножество в классе *общих линейных моделей* (10.41'), сами составляют достаточно широкий класс моделей авторегрессии, включающий в себя в качестве частных случаев ранее рассмотренные $AP(1)$ -и $AP(2)$ -модели. Если в общей линейной модели (10.41') полагать все параметры π_j , *кроме первых p коэффициентов*, равными нулю, то мы приходим к определению $AP(p)$ -модели. Итак, авторегрессионной моделью порядка p называется модель вида:

$$\varepsilon(t) = \sum_{j=1}^p \alpha_j \varepsilon(t-j) + \delta(t), \quad (10.55)$$

где последовательность случайных величин $\delta(1), \delta(2), \dots$ образует белый шум, то есть удовлетворяет условиям (10.39)–(10.40).

Условия стационарности процесса, генерируемого моделью (10.55), также формулируются в терминах корней его характеристического уравнения

$$1 - \alpha_1 z - \alpha_2 z^2 - \dots - \alpha_p z^p = 0. \quad (10.56)$$

Для стационарности процесса (10.55) необходимо и достаточно, чтобы все корни уравнения (10.56) лежали бы вне единичного круга, то есть превосходили бы по модулю единицу.

Автокорреляционная функция процесса (10.55) так же, как и для $AP(2)$ -модели, может быть вычислена с помощью рекуррентного соотношения по первым p ее значениям $r(1), r(2), \dots, r(p)$. Это рекуррентное соотношение выводится следующим образом. Умножим все члены соотношения (10.55) на $\varepsilon(t-\tau)$ ($\tau > p$). Получим $\varepsilon(t-\tau)\varepsilon(t) = \alpha_1 \varepsilon(t-\tau) \times \varepsilon(t-1) + \alpha_2 \varepsilon(t-\tau)\varepsilon(t-2) + \dots + \alpha_p \varepsilon(t-\tau)\varepsilon(t-p) + \varepsilon(t-\tau)\delta(t)$. Переходя к математическим ожиданиям величин, участвующих в этом соотношении, получаем:

$$\gamma(\tau) = \alpha_1 \gamma(\tau-1) + \alpha_2 \gamma(\tau-2) + \dots + \alpha_p \gamma(\tau-p). \quad (10.57)$$

Отметим, что $E(\varepsilon(t-k)\delta(t)) = 0$ при $k > 0$, так как $\varepsilon(t-k)$, выраженное через δ , может включать лишь импульсы $\delta(j)$ для $j \leq t-k$ (то есть $\varepsilon(t)$ не

(см. (10.436) и (10.48)). Оценки $\hat{\alpha}_k$ для параметров α_k получим, заменив теоретические значения автокорреляций $r(k)$ их оценками $\hat{r}(k)$ ($k = 1, 2, \dots, p$). Для того, чтобы выписать решение в явном виде, перейдем к матричным обозначениям:

$$\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{pmatrix}, \quad \mathbf{r} = \begin{pmatrix} r(1) \\ r(2) \\ \vdots \\ r(p) \end{pmatrix},$$

$$\mathbf{R} = \begin{pmatrix} 1 & r(1) & r(2) & \dots & r(p-1) \\ r(1) & 1 & r(1) & \dots & r(p-2) \\ \dots & \dots & \dots & \dots & \dots \\ r(p-1) & r(p-2) & r(p-3) & \dots & 1 \end{pmatrix}.$$

Тогда система (10.60) может быть представлена в форме

$$\mathbf{R}\alpha = \mathbf{r}, \quad (10.60')$$

а ее решение, соответственно, будет иметь вид:

$$\alpha = \mathbf{R}^{-1}\mathbf{r}. \quad (10.61)$$

Для вывода соотношения, необходимого при построении оценки $\hat{\sigma}_0^2$ параметра $\sigma_0^2 = \mathbf{D}\delta(t)$, домножим обе части (10.55) на $\varepsilon(t)$:

$$\begin{aligned} \varepsilon^2(t) &= \alpha_1 \varepsilon(t-1)\varepsilon(t) + \alpha_2 \varepsilon(t-2)\varepsilon(t) + \dots + \alpha_p \varepsilon(t-p)\varepsilon(t) + \delta(t) \times \\ &\times (\alpha_1 \varepsilon(t-1) + \alpha_2 \varepsilon(t-2) + \dots + \alpha_p \varepsilon(t-p) + \delta(t)). \end{aligned}$$

Переходя к математическим ожиданиям участвующих в этом соотношении величин и учитывая взаимную некоррелированность $\delta(t)$ и $\varepsilon(t-k)$ при $k = 1, 2, \dots, p$, получаем:

$$\gamma(0) = \alpha_1 \gamma(1) + \alpha_2 \gamma(2) + \dots + \alpha_p \gamma(p) + \sigma_0^2.$$

Отсюда после деления всех членов на $\gamma(0)$ имеем:

$$\sigma_0^2 = \gamma(0)(1 - \alpha_1 r(1) - \alpha_2 r(2) - \dots - \alpha_p r(p)). \quad (10.62)$$

Оценка $\hat{\sigma}_0^2$ для параметра σ_0^2 получается с помощью (10.62) заменой всех участвующих в правой части величин их оценками $(\hat{\gamma}(0), \hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_p)$.

10.4.2 Модели скользящего среднего порядка q (СС(q)-модели)

Рассмотрим частный случай общего линейного процесса (10.41), когда только первые q из весовых коэффициентов β_j ненулевые. В этом случае процесс имеет вид

$$\varepsilon(t) = \delta(t) - \theta_1 \delta(t-1) - \theta_2 \delta(t-2) - \dots - \theta_q \delta(t-q), \quad (10.63)$$

где символы $-\theta_1, -\theta_2, \dots, -\theta_q$ используются для обозначения *конечного* набора параметров β , участвующих в (10.41). Процесс (10.63) называется *моделью скользящего среднего порядка q* и сокращенно обозначается СС(q)¹⁴. Особенно распространены в практике статистического моделирования СС-модели первого ($q = 1$) и второго ($q = 2$) порядка.

$$\text{СС (1): } \varepsilon(t) = \delta(t) - \theta \delta(t-1), \quad (10.64)$$

$$\text{СС (2): } \varepsilon(t) = \delta(t) - \theta_1 \delta(t-1) - \theta_2 \delta(t-2). \quad (10.65)$$

Двойственность в представлении АР- и СС-моделей и понятие обратимости СС-моделей. В начале п. 10.4 были приведены две эквивалентные формы записи *общего линейного процесса*. Из (10.41) и (10.41') видно, что один и тот же общий линейный процесс может быть представлен либо в виде АР-модели бесконечного порядка, либо в виде СС-модели бесконечного порядка. Именно это мы понимаем под «двойственностью» в представлении анализируемого временного ряда. Проанализируем более подробно эту проблему в рамках *конечных* АР(p)- и СС(q)-моделей, представляющих собой в совокупности некоторый подкласс в классе общих линейных моделей.

Начнем с простейших представителей этого подкласса, а именно с моделей АР(1) и СС(1). Мы уже видели (см. (5.33)), где АР(1)-модель $\varepsilon(t) = \alpha \varepsilon(t-1) + \delta(t)$ может быть представлена также в виде

$$\varepsilon(t) = \delta(t) + \alpha \delta(t-1) + \alpha^2 \delta(t-2) + \dots,$$

то есть в виде модели скользящего среднего бесконечного порядка. При этом оказалось, что для обеспечения *стационарности* $\varepsilon(t)$ необходимо и достаточно потребовать, чтобы параметр α по абсолютной величине был меньше единицы (или, что то же, чтобы корень характеристического уравнения $1 - \alpha z = 0$ лежал бы вне единичного круга).

¹⁴Термин «скользящее среднее» здесь употреблен в несколько ином смысле, чем в п. 10.3.2, где речь шла об алгоритмических методах сглаживания временного ряда. В частности, в моделях СС(q) сумма весовых коэффициентов не равна единице. В англоязычной специальной литературе СС(q)-модели называются «Moving Average models» и сокращенно обозначаются MA(q)-models.

Рассмотрим теперь $CC(1)$ -модель и попробуем выразить $\delta(t)$ через текущее и все прошлые значения $\varepsilon(t)$. Уединяя $\delta(t)$ в соотношении (10.64), имеем:

$$\begin{aligned}\delta(t) &= \varepsilon(t) + \theta\delta(t-1) = \varepsilon(t) + \theta[\varepsilon(t-1) + \theta\delta(t-2)] = \\ &= \varepsilon(t) + \theta\varepsilon(t-1) + \theta^2[\varepsilon(t-2) + \theta\delta(t-3)] = \\ &= \varepsilon(t) + \theta\varepsilon(t-1) + \theta^2\varepsilon(t-2) + \theta^3[\varepsilon(t-3) + \theta\delta(t-4)] = \\ &\dots\dots\dots \\ &= \varepsilon(t) + \theta\varepsilon(t-1) + \theta^2\varepsilon(t-2) + \theta^3\varepsilon(t-3) + \dots \quad (10.66)\end{aligned}$$

Соотношение (10.66) можно переписать в виде

$$\varepsilon(t) = - \sum_{k=1}^{\infty} \theta^k \varepsilon(t-k) + \delta(t). \quad (10.66')$$

Это означает, что ряд $\varepsilon(t)$, сгенерированный $CC(1)$ -моделью (10.64), может быть представлен также в *виде модели авторегрессии бесконечного порядка*. Заметим, что, в отличие от $AR(p)$ -моделей, в $CC(q)$ -моделях не требуется накладывать никаких ограничений на параметры $\theta_1, \theta_2, \dots, \theta_q$ для обеспечения их стационарности: *ряды вида (10.63) стационарны при любых вещественных значениях параметров θ_j ($j = 1, 2, \dots, q$)*. Однако, если, например, в модели $CC(1)$ параметр θ по абсолютной величине больше или равен единице ($|\theta| \geq 1$), то текущее значение анализируемого ряда $\varepsilon(t)$ в соответствии с (10.66') будет зависеть от своих прошлых значений $\varepsilon(t-1), \varepsilon(t-2), \dots$, *берущихся с весами, бесконечно растущими по мере удаления в прошлое*. С целью избежать этой неестественной (с прикладной точки зрения) ситуации потребуем, чтобы веса в обратном разложении (10.66') образовывали бы сходящийся ряд, то есть чтобы $|\theta| < 1$. Это требование равносильно условию, сформулированному в терминах *характеристического уравнения модели* (10.64) $1 - \theta z = 0$ и заключающемуся в том, чтобы корень этого характеристического уравнения был бы по модулю больше единицы (лежал бы вне единичного круга). Итак, *$CC(1)$ -модель называется обратимой, если в обратном разложении (10.66') бесконечный ряд весов при $\varepsilon(t-k)$ ($k = 1, 2, \dots$) сходится* (или, что то же, если $|\theta| < 1$). Еще раз заметим, что условия обратимости CC -моделей не имеют никаких связей с условиями стационарности временного ряда.

Обобщим этот результат на случай $CC(q)$ -модели при *произвольном конечном* значении q . Уединяя $\delta(t)$ в соотношении (10.63), имеем:

$$\delta(t) = \varepsilon(t) + \theta_1\delta(t-1) + \theta_2\delta(t-2) + \dots + \theta_q\delta(t-q). \quad (10.63')$$

Действуя аналогично тому, как мы это делали при выводе соотношения (10.66), то есть заменяя бесконечное число раз $\delta(t-1), \delta(t-2), \dots$,

$\delta(t - q)$ в правой части (10.63') их выражениями, вычисленными по формуле (10.63'), получим разложение вида

$$\delta(t) = \varepsilon(t) - \pi_1 \varepsilon(t - 1) - \pi_2 \varepsilon(t - 2) - \dots, \quad (10.67)$$

в котором коэффициенты π_j ($j = 1, 2, \dots$) определенным образом выражаются через параметры $\theta_1, \theta_2, \dots, \theta_p$. Соотношение (10.67) может быть записано в виде модели авторегрессии бесконечного порядка (то есть в виде *обращенного разложения*)

$$\varepsilon(t) = \sum_{j=1}^{\infty} \pi_j \varepsilon(t - j) + \delta(t). \quad (10.67')$$

Можно показать (см., например, [Бокс, Дженкинс (1974), с. 84]), что *условие обратимости* СС(q)-модели (то есть условие сходимости ряда $\sum_{j=1}^{\infty} \pi_j$) формулируется в терминах характеристического уравнения модели (10.63) следующим образом:

$$\left. \begin{aligned} &\bullet \text{ все корни } z_1, z_2, \dots, z_q \text{ характеристического} \\ &\text{уравнения } 1 - \theta_1 z - \theta_2 z^2 - \dots - \theta_q z^q = 0 \\ &\text{должны лежать вне единичного круга, то есть} \\ &|z_j| > 1 \text{ для всех } j = 1, 2, \dots, q. \end{aligned} \right\} \quad (10.68)$$

Основные характеристики процесса СС(q). Получим вспомогательные соотношения, связывающие между собой дисперсию и ковариации процесса скользящего среднего порядка q с параметрами $\theta_1, \theta_2, \dots, \theta_q$. Для этого выразим $\varepsilon(t - k)$ в соответствии с (10.63)

$$\varepsilon(t - k) = \delta(t - k) - \theta_1 \delta(t - k - 1) - \dots - \theta_q \delta(t - k - q), \quad (10.63)$$

перемножим почленно выражения (10.63) и (10.63а) и вычислим математические ожидания всех членов получившегося соотношения (учитывая, конечно, взаимную некоррелированность элементов белого шума $\delta(t_1)$ и $\delta(t_2)$ при $t_1 \neq t_2$). В результате получим следующее выражение для ковариаций $\gamma(\tau) = E(\varepsilon(t)\varepsilon(t - \tau))$:

$$\gamma(\tau) = \begin{cases} \sigma_0^2(1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) & \text{при } \tau = 0; \\ \sigma_0^2(-\theta_\tau + \theta_1\theta_{\tau+1} + \theta_2\theta_{\tau+2} + \dots \\ \dots + \theta_{q-\tau}\theta_q) & \text{при } \tau = 1, 2, \dots, q; \\ 0 & \text{при } \tau > q \end{cases} \quad (10.69)$$

(при этом, естественно, полагается, что $\theta_j = 0$ при $j > q$).

Автокорреляционная функция процесса $CC(q)$ получается непосредственно из (10.69):

$$r(\tau) = \begin{cases} \frac{-\theta_\tau + \theta_1\theta_{\tau+1} + \theta_2\theta_{\tau+2} + \dots + \theta_{q-\tau}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} & \text{при } \tau = 1, 2, \dots, q; \\ 0 & \text{при } \tau > q. \end{cases} \quad (10.70)$$

Мы видим, что автокорреляционная функция $r(\tau)$ процесса $CC(q)$ равна нулю для всех значений τ , больших порядка процесса q . Это важное свойство используется при подборе порядка $CC(q)$ -модели по экспериментальным данным.

Спектральная плотность $p(\tilde{\omega})$ процесса $CC(q)$ может быть вычислена с помощью соотношения (10.9а) с учетом того, что участвующий в нем оператор $B(z)$ для модели (10.63) имеет вид

$$B(z) = 1 - \theta_1 z - \theta_2 z^2 - \dots - \theta_q z^q.$$

Соответственно имеем:

$$p(\tilde{\omega}) = 2\sigma_0^2 |1 - \theta_1 e^{-i2\pi\tilde{\omega}} - \theta_2 e^{-i4\pi\tilde{\omega}} - \dots - \theta_q e^{-i2\pi q\tilde{\omega}}|^2, \quad 0 \leq \tilde{\omega} \leq \frac{1}{2}. \quad (10.71)$$

Идентификация модели $CC(q)$ производится на базе соотношений (10.70), а именно: 1) по значениям $\hat{\varepsilon}(t) = x(t) - \hat{f}(t)$ с помощью формулы

$$\hat{r}(\tau) = \frac{\frac{1}{N-\tau} \sum_{t=1}^{N-\tau} (\hat{\varepsilon}(t) - \bar{\varepsilon})(\hat{\varepsilon}(t+\tau) - \bar{\varepsilon})}{\frac{1}{N} \sum_{t=1}^N (\varepsilon(t) - \bar{\varepsilon})^2}, \quad \tau = 1, 2, \dots, q, \quad (10.6'')$$

подсчитываются значения $\hat{r}(1), \hat{r}(2), \dots, \hat{r}(q)$; 2) в соотношения (10.70) последовательно подставляются значения $\tau = 1, 2, \dots, q$ с заменой в левой их части величин $r(\tau)$ полученными ранее оценками $\hat{r}(\tau)$; 3) полученная таким образом система из q уравнений разрешается относительно неизвестных значений $\theta_1, \theta_2, \dots, \theta_q$; решения этой системы $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q$ и дадут нам оценки неизвестных параметров модели, соответственно, $\theta_1, \theta_2, \dots, \theta_q$; 4) оценка параметра σ_0^2 может быть получена с помощью первого из соотношений (10.69) подстановкой в него вместо $\gamma(0), \theta_1, \theta_2, \dots, \theta_q$ их оценок, соответственно, $\hat{\gamma}(0), \hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q$.

Заметим, однако, что в отличие от системы линейных уравнений Юла-Уокера (10.60), выведенных в связи с задачей оценивания параметров процесса авторегрессии, уравнения для определения оценок параметров $CC(q)$ -модели, полученные на базе соотношений (10.70), *нелинейны*. Поэтому, за исключением простого случая $q = 1$, который будет

рассмотрен ниже, эти уравнения приходится решать с помощью итерационных процедур (одна из таких процедур описана, например, в книге [Бокс, Дженкинс (1974), с. 225]).

Рассмотрим теперь специально два наиболее важных частных случая $CC(q)$ -модели.

Модель скользящего среднего 1-го порядка ($CC(1)$ -модель). Эта модель описывается соотношением (10.64) и, как уже было показано, *стационарна* при любом значении параметра θ и *обратима* при условии $|\theta| < 1$.

Автокорреляционная функция $CC(1)$ -модели определяется соотношениями (10.70) при $q = 1$, то есть

$$r(\tau) = \begin{cases} \frac{-\theta}{1 + \theta^2} & \text{при } \tau = 1, \\ 0 & \text{при } \tau \geq 2. \end{cases} \quad (10.72)$$

Частная корреляционная функция $r_{\text{част}}(\tau)$ процесса скользящего среднего 1-го порядка, определяющая степень тесноты корреляционной связи между $\varepsilon(t)$ и $\varepsilon(t \pm \tau)$ при фиксированных значениях всех промежуточных элементов этого ряда, задается выражением

$$r_{\text{част}}(\tau) = -\theta^\tau \frac{1 - \theta^2}{1 - \theta^{2(\tau+1)}} \quad (10.73)$$

(вывод этого соотношения см., например, в книге [Бокс, Дженкинс (1974), с. 87]). Мы видим, что (с учетом $|\theta| < 1$) $|r_{\text{част}}(\tau)| < \theta^\tau$ и что затухающая экспонента определяет поведение частной автокорреляционной функции. Если $\tau(1)$ положительно (и, следовательно, θ отрицательно), $r_{\text{част}}(\tau)$ осциллирует с переменами знака. Если же $r(1)$ отрицательно, все значения $r_{\text{част}}(\tau)$ отрицательны.

Спектральную плотность $CC(1)$ -модели получаем из (10.71) при $q = 1$:

$$\begin{aligned} p(\tilde{\omega}) &= 2\sigma_0^2 |1 - \theta e^{-i2\pi\tilde{\omega}}|^2 = \\ &= 2\sigma_0^2 (1 + \theta^2 - 2\theta \cos(2\pi\tilde{\omega})), \quad 0 \leq \tilde{\omega} \leq \frac{1}{2}. \end{aligned} \quad (10.74)$$

Из (10.72) и (10.74) видно, что когда θ отрицательно, то $r(1)$ положительно и в спектре доминируют низкие частоты (то есть соответствующие временные ряды имеют относительно длинные псевдопериоды). Напротив, при положительном θ величина $r(1)$ отрицательна, а в спектре доминируют высокие частоты.

Идентификация $CC(1)$ -модели сводится, в соответствии с описанным выше общим процессом оценивания параметров $CC(q)$ -модели, к решению единственного уравнения

$$\theta^2 + \frac{1}{\hat{r}(1)} \theta + 1 = 0. \quad (10.75)$$

Из двух решений этого квадратного уравнения выбирается то, которое удовлетворяет условию обратимости СС(1)-модели (то есть условию $|\theta| < 1$). То, что такое решение существует и единственно, следует из известного свойства корней приведенного квадратного уравнения: их произведение равно свободному члену. В нашем случае это означает: если $\theta^{(1)}$ — корень уравнения (10.75), то величина $1/\theta^{(1)}$ является вторым корнем этого уравнения. Поэтому, если один из корней по абсолютной величине меньше единицы, то другой будет больше единицы.

Модель скользящего среднего 2-го порядка (СС(2)-модель). Эта модель описывается соотношением (10.65) и *стационарна* при любых значениях параметров θ_1 и θ_2 . Однако, как это следует из (10.68), СС(2)-процесс *обратим* лишь при условии, что корни его характеристического уравнения

$$1 - \theta_1 z - \theta_2 z^2 = 0$$

лежат вне единичного круга, то есть

$$\begin{cases} |\theta_1| < 2, \\ \theta_2 < 1 - |\theta_1|. \end{cases} \quad (10.68a)$$

Заметим, что эти условия аналогичны условиям (10.53') *стационарности* процесса АР(2).

Автокорреляционная функция процесса СС(2) определяется соотношениями (10.70) при $q = 2$, то есть

$$\begin{aligned} r(1) &= \frac{-\theta_1(1 - \theta_2)}{1 + \theta_1^2 + \theta_2^2}, \\ r(2) &= \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2}, \\ r(\tau) &= 0 \quad \text{при всех } \tau \geq 3. \end{aligned} \quad (10.76)$$

Из (10.68a) и (10.76) следует, что первые две автокорреляции обратимого процесса СС(2) должны лежать на плоскости $(r(1), r(2))$ внутри площади, ограниченной отрезками линий

$$\begin{cases} r(2) + r(1) = -0,5, \\ r(2) - r(1) = -0,5, \\ r^2(1) = 4r(2)(1 - 2r(2)), \end{cases} \quad (10.77)$$

см. рис. 10.12. Рисунок позволяет оценить, согласуется ли вычисленная по экспериментальным данным с помощью формулы (10.6'') пара значений $\hat{r}(1), \hat{r}(2)$ с гипотезой, что анализируемый временной ряд $\varepsilon(t)$ может быть описан моделью обратимого процесса СС(2).

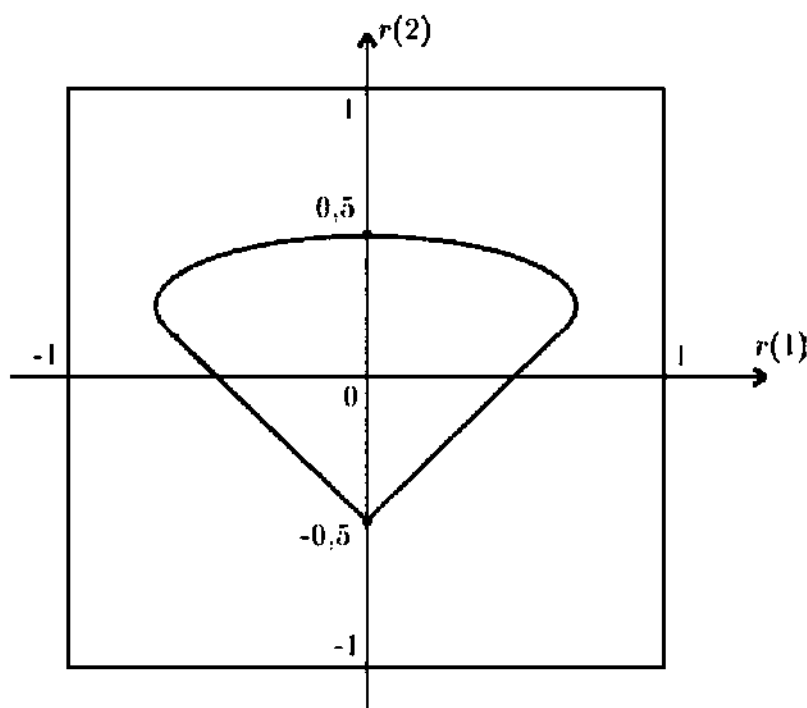


Рис. 10.12. Допустимые области значений $r(1)$ и $r(2)$ для обратимого процесса $CC(2)$

Частная автокорреляционная функция имеет весьма сложную форму аналитического представления. Однако можно показать, что главную роль в этом представлении играет либо сумма двух экспоненциально убывающих (с ростом τ) членов (если корни характеристического уравнения действительны), либо затухающая синусоида (если корни характеристического уравнения комплексны). То есть эта функция ведет себя так же, как автокорреляционная функция процесса $AR(2)$.

Спектральную плотность $p(\tilde{\omega})$ процесса $CC(2)$ получаем из (10.71) при $q = 2$:

$$\begin{aligned} p(\tilde{\omega}) &= 2\sigma_0^2 |1 - \theta_1 e^{-i2\pi\tilde{\omega}} - \theta_2 e^{-i4\pi\tilde{\omega}}|^2 = \\ &= 2\sigma_0^2 (1 + \theta_1^2 + \theta_2^2 - 2\theta_1(1 - \theta_2) \cos(2\pi\tilde{\omega}) - 2\theta_2 \cos(4\pi\tilde{\omega})), \\ &0 \leq \tilde{\omega} \leq \frac{1}{2}. \end{aligned}$$

Отметим, что этот спектр с точностью до постоянного множителя $2\sigma_0^2$ обратен спектру (10.54) процесса авторегрессии 2-го порядка.

Идентификация $CC(2)$ -модели сводится, в соответствии с описанным выше общим процессом оценивания параметров $CC(q)$ -модели, к решению системы из двух нелинейных (относительно неизвестных θ_1 и θ_2) уравнений:

$$\begin{cases} \hat{r}(1) = \frac{-\theta_1(1 - \theta_2)}{1 + \theta_1^2 + \theta_2^2}, \\ \hat{r}(2) = \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2}. \end{cases} \quad (10.78)$$

Взаимосвязь процессов $AP(p)$ и $CC(q)$. Результаты, изложенные в пунктах 10.4.1 и 10.4.2, позволяют сделать ряд заключений о взаимосвязях, существующих между процессами авторегрессии и скользящего среднего.

1) Для конечного процесса авторегрессии $\varepsilon(t)$ порядка p элемент белого шума $\delta(t)$ может быть представлен как *конечная* взвешенная сумма предшествующих ε , или $\varepsilon(t)$ может быть представлен как бесконечная сумма предшествующих δ . В то же время, если $\varepsilon(t)$ — конечный процесс скользящего среднего порядка q , то $\delta(t)$ может быть представлен как бесконечная взвешенная сумма предшествующих ε .

2) Конечный процесс CC имеет автокорреляционную функцию, обращающуюся в нуль после некоторой точки, но так как он эквивалентен бесконечному процессу AP , его частная автокорреляционная функция бесконечно протяженная. Главную роль в ней играют затухающие экспоненты и (или) затухающие синусоиды. И наоборот: процесс AP имеет частную автокорреляционную функцию, обращающуюся в нуль после некоторой точки, но его автокорреляционная функция имеет бесконечную протяженность и состоит из совокупности затухающих экспонент и (или) затухающих синусоид.

3) Параметры процесса авторегрессии конечного порядка p не должны удовлетворять каким-нибудь условиям, для того чтобы этот процесс был обратимым. Однако, чтобы процесс был стационарен, корни его характеристического уравнения должны лежать вне единичного круга. В то же время параметры процесса CC не должны удовлетворять каким-нибудь условиям для того, чтобы процесс был стационарным. Однако для того чтобы процесс CC был обратимым, корни его характеристического уравнения должны лежать вне единичного круга.

4) Спектр процесса скользящего среднего обратен спектру соответствующего процесса авторегрессии.

10.4.3 Авторегрессионные модели со скользящими средними в остатках ($ARCC(p, q)$ -модели)

В п. 10.4.2 обсуждалась *двойственность* в представлении AP - и CC -моделей. В частности, мы видели, что *конечный* процесс скользящего среднего, например,

$$\varepsilon(t) = \delta(t) - \theta\delta(t-1),$$

может быть записан как *бесконечный* процесс авторегрессии

$$\varepsilon(t) = - \sum_{k=1}^{\infty} \theta^k \varepsilon(t-k) + \delta(t).$$

И обратно: простейший процесс авторегрессии может быть представлен как бесконечный процесс скользящего среднего (см. (5.33)).

Но если мы анализируем процесс действительно типа $CC(1)$, то его представление в виде процесса авторегрессии неэкономично с точки зрения формы его параметризации. Аналогично процесс $AP(1)$ не может быть экономично представлен с помощью модели скользящего среднего. Поэтому на практике для получения экономичной параметризации анализируемого процесса иногда бывает необходимо включить в модель как члены, описывающие авторегрессию, так и члены, моделирующие остаток в виде скользящего среднего. Такой линейный процесс имеет вид

$$\varepsilon(t) = \alpha_1 \varepsilon(t-1) + \dots + \alpha_p \varepsilon(t-p) + \delta(t) - \theta_1 \delta(t-1) - \dots - \theta_q \delta(t-q) \quad (10.79)$$

и называется *процессом авторегрессии — скользящего среднего порядка (p, q)* . Примем для него сокращенное обозначение $APCC(p, q)$ ¹⁵.

Отметим, что, последовательно выражая бесконечное число раз в правой части (10.79) величины $\varepsilon(t-1), \varepsilon(t-2), \dots, \varepsilon(t-p), \dots$ по формуле (10.79), мы убеждаемся в том, что $\varepsilon(t)$ не зависит от будущих значений δ , то есть от $\delta(t+1), \delta(t+2), \dots$.

Стационарность и обратимость $APCC(p, q)$ -процессов. Записывая процесс (10.79) в виде

$$\varepsilon(t) = \sum_{j=1}^p \alpha_j \varepsilon(t-j) + \bar{\delta}_q(t), \quad (10.79')$$

где $\bar{\delta}_q(t) = \delta(t) - \theta_1 \delta(t-1) - \dots - \theta_q \delta(t-q)$, мы можем провести анализ стационарности (10.79') точно по той же схеме, что и для $AP(p)$ -процессов. При этом различие «остатков» $\bar{\delta}_q(t)$ и $\delta(t)$ никак не повлияет на выводы, определяющие условия стационарности процесса авторегрессии. Поэтому процесс (10.79) является стационарным тогда и только тогда, когда все корни характеристического уравнения $AP(p)$ -процесса

$$1 - \alpha_1 z - \alpha_2 z^2 - \dots - \alpha_p z^p = 0$$

лежат вне единичного круга.

Аналогично, обозначив $\bar{\varepsilon}_p(t) = \varepsilon(t) - \sum_{j=1}^p \alpha_j \varepsilon(t-j)$ и рассматривая процесс (10.79) в виде

$$\bar{\varepsilon}_p(t) = \delta(t) - \theta_1 \delta(t-1) - \dots - \theta_q \delta(t-q), \quad (10.79'')$$

¹⁵ Модель (10.79) может интерпретироваться как линейная модель множественной регрессии, в которой в качестве объясняющих переменных выступают прошлые значения самой зависимой переменной, а в качестве регрессионного остатка — скользящее среднее из элементов белого шума. Поэтому название этой модели, вынесенное в заголовок данного пункта, лучше отражает ее сущность. Однако для краткости чаще используется название «модель авторегрессии — скользящего среднего». В англоязычной литературе эти модели называют **AutoRegressive — Moving Average Models**, или сокращенно **ARMA-models**.

получаем те же выводы относительно условий обратимости этого процесса, что и для процесса СС (q), а именно: для обратимости АРСС (p, q)-процесса необходимо и достаточно, чтобы все корни характеристического уравнения СС (q)-процесса

$$1 - \theta_1 z - \theta_2 z^2 - \dots - \theta_q z^q = 0$$

лежали бы вне единичного круга.

Автокорреляционная функция анализируется методами, аналогичными тем, что мы использовали при выводе автокорреляционных функций для АР- и СС-процессов. Умножим все члены (10.79) на $\varepsilon(t - \tau)$ и перейдем к математическим ожиданиям получившегося выражения:

$$\gamma(\tau) = \alpha_1 \gamma(\tau - 1) + \dots + \alpha_p \gamma(\tau - p) + \gamma_{\varepsilon\delta}(\tau) - \theta_1 \gamma_{\varepsilon\delta}(\tau - 1) - \dots - \theta_q \gamma_{\varepsilon\delta}(\tau - q), \quad (10.80)$$

где $\gamma_{\varepsilon\delta}(k) = E(\varepsilon(t - k)\delta(t))$ — «перекрестная» ковариационная функция случайных последовательностей $\varepsilon(t)$ и $\delta(t)$. Так как $\varepsilon(t - k)$ не зависит от будущих (по отношению к моменту $t - k$) значений δ , то $\gamma_{\varepsilon\delta}(k) = 0$ при всех $k > 0$ и $\gamma_{\varepsilon\delta}(k) \neq 0$ при всех $k \leq 0$. Из (10.80) следует:

$$\gamma(\tau) = \alpha_1 \gamma(\tau - 1) + \alpha_2 \gamma(\tau - 2) + \dots + \alpha_p \gamma(\tau - p) \quad \text{при} \quad \tau \geq q + 1$$

и, соответственно (после деления всех членов на $\gamma(0)$),

$$r(\tau) = \alpha_1 r(\tau - 1) + \alpha_2 r(\tau - 2) + \dots + \alpha_p r(\tau - p) \quad \text{при} \quad \tau \geq q + 1. \quad (10.81)$$

Анализ соотношений (10.80) и (10.81) позволяет сделать следующие выводы.

1) Из соотношений (10.80) для $\tau = 0, 1, 2, \dots, q$ следует, что ковариации $\gamma(0), \gamma(1), \dots, \gamma(q)$ и, соответственно, автокорреляции $r(1), r(2), \dots, r(q)$ связаны определенной системой зависимостей с q параметрами скользящего среднего $\theta_1, \theta_2, \dots, \theta_q$ и p параметрами авторегрессии $\alpha_1, \alpha_2, \dots, \alpha_p$. При этом имеется в виду, что перекрестные ковариации $\gamma_{\varepsilon\delta}(\tau), \gamma_{\varepsilon\delta}(\tau - 1), \dots, \gamma_{\varepsilon\delta}(\tau - q)$ при положительных значениях сдвига по времени, как уже было подмечено, равны нулю, а при отрицательных — тоже могут быть выражены в терминах параметров $\alpha_1, \dots, \alpha_p, \theta_1, \dots, \theta_q$ с помощью следующего приема: пусть $k > 0$; тогда $\gamma_{\varepsilon\delta}(-k) = E(\varepsilon(t + k)\delta(t))$; в произведении $\varepsilon(t + k)\delta(t)$ с помощью $(k + 1)$ -кратной последовательной подстановки первого сомножителя по формуле (10.79) он заменяется линейной комбинацией $\varepsilon(t - 1)$, элементов белого шума δ и параметров модели, что после применения к получившемуся произведению операции усреднения E дает выражение, зависящее только от параметров модели (поскольку $E(\varepsilon(t - 1)\delta(t)) = 0$).

2) Значения автокорреляционной функции $r(\tau)$ для $\tau \geq q + 1$ вычисляются по рекуррентному соотношению (10.81), которое в точности повторяет аналогичное рекуррентное соотношение (10.58) для ав-

токорреляционной функции процесса $AP(p)$. А это значит, что, начиная с $\tau = q + 1$, автокорреляционная функция процесса $APCC(p, q)$ ведет себя так же, как и автокорреляционная функция процесса $AP(p)$, то есть она будет состоять из совокупности затухающих экспонент и (или) затухающих синусоид, и ее свойства определяются коэффициентами $\alpha_1, \alpha_2, \dots, \alpha_p$ и начальными значениями $r(1), r(2), \dots, r(p)$.

Частная автокорреляционная функция $r(\tau)$ процесса $APCC(p, q)$ при больших τ ведет себя как частная автокорреляционная функция $CC(q)$ -процесса. Это значит, что в ней преобладают члены типа затухающих экспонент и (или) затухающих синусоид (соотношение между теми и другими зависит от порядка скользящего среднего q и значений параметров процесса).

Спектральная плотность $p(\tilde{\omega})$ процесса $APCC(p, q)$ может быть вычислена с помощью соотношения (10.9а):

$$p(\tilde{\omega}) = 2\sigma_0^2 \frac{|1 - \theta_1 e^{i2\pi\tilde{\omega}} - \theta_2 e^{-i4\pi\tilde{\omega}} - \dots - \theta_q e^{-i2\pi q\tilde{\omega}}|^2}{|1 - \alpha_1 e^{-i2\pi\tilde{\omega}} - \alpha_2 e^{-i4\pi\tilde{\omega}} - \dots - \alpha_p e^{-i2\pi p\tilde{\omega}}|^2}, \quad (10.82)$$

$$0 \leq \tilde{\omega} \leq \frac{1}{2}.$$

Идентификация процесса $APCC(p, q)$ может базироваться (так же как и в AP - и CC -моделях) на статистическом оценивании параметров модели с помощью метода моментов. Соотношения, связывающие ковариации $\gamma(\tau)$ с неизвестными значениями параметров, могут быть получены различными способами. Но большинство подходов к идентификации модели основано на домножении соотношения (10.79) (или вытекающих из этого соотношения равенств) на $\varepsilon(k)$ ($k = t, t-1, t-2, \dots$), переходе к математическим ожиданиям от получившихся выражений и использовании образовавшихся связей между автоковариациями процесса $\varepsilon(t)$ и параметрами $APCC$ -модели для определения неизвестных значений этих параметров по необходимому набору подсчитанных по исходным статистическим данным выборочных автоковариаций $\hat{\gamma}(\tau)$ ($\tau = 0, 1, 2, \dots$).

Наметим один из вариантов таких подходов. Разобьем процедуру получения оценок параметров α_k ($k = 1, 2, \dots, p$), θ_j ($j = 1, 2, \dots, q$) и σ_0^2 на два этапа. На 1-м этапе мы получим оценки параметров α_k , на 2-м — оценки параметров θ_j и σ_0^2 .

1 - й э т а п. Перепишем соотношение (10.79) в виде

$$\varepsilon(t) - \sum_{k=1}^p \alpha_k \varepsilon(t-k) = \delta(t) - \sum_{j=1}^q \theta_j \delta(t-j), \quad (10.79a)$$

домножим поочередно обе части этого соотношения на $\varepsilon(t-q-1), \varepsilon(t-q-2), \dots, \varepsilon(t-q-p)$ и перейдем к математическим ожиданиям в каждом из p получившихся таким образом соотношений. Учитывая взаимную

формулой

$$\varepsilon(t) = \alpha\varepsilon(t-1) + \delta(t) - \theta\delta(t-1), \quad (10.84)$$

или, что то же,

$$\varepsilon(t) - \alpha\varepsilon(t-1) = \delta(t) - \theta\delta(t-1). \quad (10.84')$$

Стационарность и обратимость. В соответствии с общей теорией АРСС-процессов процесс АРСС(1,1) стационарен, если корень характеристического уравнения АР(1)-модели, то есть уравнения $1 - \alpha z = 0$, по модулю больше единицы. Это значит, что стационарность гарантируется условием $|\alpha| < 1$. Обратимость АРСС(1,1)-процесса обеспечивается требованием, чтобы корень характеристического уравнения СС(1)-процесса, то есть уравнения $1 - \theta z = 0$, был бы по модулю больше единицы. Это означает, что обратимость АРСС(1,1)-процесса гарантируется условием $|\theta| < 1$.

Автокорреляционная функция может быть построена с помощью соотношений (10.80), выписанных для $\tau = 0$, $\tau = 1$ и $\tau \geq 2$:

$$\begin{cases} \gamma(0) = \alpha\gamma(1) + \sigma_0^2 - \theta\gamma_{\varepsilon\delta}(-1), \\ \gamma(1) = \alpha\gamma(0) - \theta\sigma_0^2, \\ \gamma(\tau) = \alpha\gamma(\tau-1) \quad \text{при } \tau \geq 2. \end{cases}$$

Чтобы выразить $\gamma_{\varepsilon\delta}(-1)$ через параметры модели, умножим все члены (10.84') на $\delta(t-1)$ и перейдем к математическим ожиданиям от получившихся выражений:

$$\gamma_{\varepsilon\delta}(-1) = (\alpha - \theta)\sigma_0^2.$$

Теперь мы можем описать автоковариационную и автокорреляционную функции АРСС(1,1)-процесса в терминах его параметров:

$$\begin{aligned} \gamma(0) &= \frac{1 + \theta^2 - 2\alpha\theta}{1 - \alpha^2} \sigma_0^2, \\ \gamma(1) &= \frac{(1 - \alpha\theta)(\alpha - \theta)}{1 - \alpha^2} \sigma_0^2, \\ \gamma(\tau) &= \alpha\gamma(\tau-1) \quad \text{при } \tau \geq 2 \end{aligned}$$

и, следовательно,

$$\begin{aligned} r(1) &= \frac{(1 - \alpha\theta)(\alpha - \theta)}{1 + \theta^2 - 2\alpha\theta}, \\ r(\tau) &= \alpha r(\tau-1) = \alpha^{\tau-1} r(1) \quad \text{для } \tau \geq 2. \end{aligned} \quad (10.85)$$

Мы видим, что автокорреляционная функция экспоненциально убывает от начального значения $r(1)$, причем это убывание монотонно, если

α положительно, и колебательно (знакопеременно), если α отрицательно.

Из (10.85) и условий стационарности и обратимости следует, что $r(1)$ и $r(2)$ должны удовлетворять соотношениям:

$$\begin{aligned} |r(2)| &< |r(1)|, \\ r(2) &> r(1)(2r(1) + 1) \quad \text{при} \quad r(1) < 0, \\ r(2) &> r(1)(2r(1) - 1) \quad \text{при} \quad r(1) > 0. \end{aligned}$$

Эти условия бывают полезными при проверке гипотезы о том, что анализируемый процесс может быть описан АРСС(1,1)-моделью (по выборочным значениям $\hat{r}(1)$ и $\hat{r}(2)$ коэффициентов автокорреляции).

Частная автокорреляционная функция $r(\tau)$ АРСС(1,1)-процесса определяется единственным начальным значением $r_{\text{част}}(1)$, а затем экспоненциально убывает. При этом если θ положительно, то она монотонно убывает от значения $r_{\text{част}}(1)$, знак которого совпадает со знаком величины $(\alpha - \theta)$. При отрицательном θ величина $r_{\text{част}}(\tau)$ убывает экспоненциально-знакопеременно.

Спектральная плотность $p(\tilde{\omega})$ определится в соответствии с общим соотношением (10.82), в котором следует положить $p = q = 1$:

$$p(\tilde{\omega}) = 2\sigma_0^2 \frac{1 + \theta^2 - 2\theta \cos(2\pi\tilde{\omega})}{1 + \alpha^2 - 2\alpha \cos(2\pi\tilde{\omega})}, \quad 0 \leq \tilde{\omega} \leq \frac{1}{2}. \quad (10.86)$$

Идентификация АРСС(1,1)-модели может быть проведена в соответствии с общей двухэтапной процедурой, описанной выше (см. соотношения (10.79а), (10.83) и т.д.). Система (10.83) в нашем случае состоит из единственного уравнения

$$r(2) - \alpha r(1) = 0,$$

так что на 1-м этапе мы получаем оценку параметра α в виде

$$\hat{\alpha} = \frac{\hat{r}(2)}{\hat{r}(1)}. \quad (10.87)$$

На 2-м этапе выписываем соотношения

$$\varepsilon(t) - \hat{\alpha}\varepsilon(t-1) = \delta(t) - \theta\delta(t-1), \quad (10.88)$$

$$\varepsilon(t+1) - \hat{\alpha}\varepsilon(t) = \delta(t+1) - \theta\delta(t). \quad (10.88')$$

Возводя в квадрат (10.88), перемножая почленно соотношения (10.88) и (10.88'), переходя к математическим ожиданиям полученных выражений и заменяя $\gamma(k)$ их выборочными значениями $\hat{\gamma}(k)$, имеем следующую систему из двух уравнений относительно неизвестных θ и σ_0^2 :

$$\begin{cases} \hat{\gamma}(0)(1 + \hat{\alpha}^2) - 2\hat{\alpha}\hat{\gamma}(1) = \sigma_0^2(1 + \theta^2), \\ \hat{\gamma}(1)(1 + \hat{\alpha}^2) - \hat{\alpha}(\hat{\gamma}(0) + \hat{\gamma}(2)) = -\theta\sigma_0^2. \end{cases}$$

Решение этой системы не представляет принципиальных трудностей. Поделив 1-е уравнение на 2-е, получаем квадратное уравнение относительно θ :

$$\frac{\hat{\gamma}(0)(1 + \hat{\alpha}^2) - 2\hat{\alpha}\hat{\gamma}(1)}{\hat{\gamma}(1)(1 + \hat{\alpha}^2) - \hat{\alpha}(\hat{\gamma}(0) + \hat{\gamma}(2))} = -\frac{1 + \theta^2}{\theta}.$$

Решив это уравнение относительно θ и выбрав из двух корней тот, который удовлетворяет условию обратимости $|\theta| < 1$, возвращаемся к любому из уравнений системы и определяем оценку параметра σ_0^2 .

10.4.4 Общие замечания, связанные с моделями АРСС

Операторы F_+ и F_- сдвига по времени и действия с ними. Удобным понятием при анализе АРСС-моделей является понятие операторов сдвига по времени анализируемого процесса на один такт времени, соответственно, вперед (оператор F_+) и назад (оператор F_-). Операторы F_+ и F_- определяются с помощью следующих соотношений:

$$F_+\varepsilon(t) = \varepsilon(t + 1);$$

$$F_-\varepsilon(t) = \varepsilon(t - 1).$$

Эта запись означает, что оператор F_+ (или оператор F_-), примененный к значению $\varepsilon(t)$ временного ряда в точке t , преобразует это значение в $\varepsilon(t + 1)$ (или, соответственно, в $\varepsilon(t - 1)$), то есть как бы «сдвигает» временной ряд на один такт времени вперед (соответственно, назад).

Из данного определения операторов F_+ и F_- непосредственно вытекают следующие простые правила действий с ними:

(1) $F_+ = F_-^{-1}$.

Действительно, $F_+(F_-\varepsilon(t)) = F_+F_-\varepsilon(t) = F_+\varepsilon(t - 1) = \varepsilon(t)$, то есть $F_+F_- = 1$.

(2) $F_+^k\varepsilon(t) = \varepsilon(t + k)$ и $F_-^k\varepsilon(t) = \varepsilon(t - k)$.

Действительно, например, $F_-^2\varepsilon(t) = F_-(F_-\varepsilon(t)) = F_-\varepsilon(t - 1) = \varepsilon(t - 2)$ и т. д.

(3) $(c_0 + c_1F_- + \dots + c_mF_-^m)\varepsilon(t) = c_0\varepsilon(t) + c_1\varepsilon(t - 1) + \dots + c_m\varepsilon(t - m)$ и аналогично

$$(c_0 + c_1F_+ + \dots + c_mF_+^m)\varepsilon(t) = c_0\varepsilon(t) + c_1\varepsilon(t + 1) + \dots + c_m\varepsilon(t + m)$$

(в этих соотношениях $c_0, c_1, \dots, c_m$ — некоторые постоянные числа).

В частности, АРСС(p, q)-модель может быть записана в виде

$$A_p(F_-, \alpha)\varepsilon(t) = B_q(F_-\theta)\delta(t), \quad (10.796)$$

где

и

$$B_q(F_-, \theta) = 1 - \theta_1 F_- - \theta_2 F_-^2 - \dots - \theta_q F_-^q. \quad (10.89')$$

Приведем здесь еще одно полезное представление АРСС(p, q)-процессов. Пусть $z_1(\alpha), z_2(\alpha), \dots, z_p(\alpha)$ — корни характеристического уравнения АР(p)-модели

$$A_p(z, \alpha) = 0, \quad (10.90)$$

а $\tilde{z}_1(\theta), \tilde{z}_2(\theta), \dots, \tilde{z}_q(\theta)$ — корни характеристического уравнения СС(q)-модели

$$B_q(z, \theta) = 0, \quad (10.90')$$

где полиномы $A_p(z, \alpha)$ и $B_q(z, \theta)$ определяются соотношениями, соответственно, (10.89) и (10.89'). И пусть все эти корни удовлетворяют условиям стационарности процесса АР(p) и обратимости процесса СС(q). Тогда с помощью несложных рассуждений, используя представимость полиномов (10.89) и (10.89') в виде

$$\begin{aligned} A_p(z, \alpha) &= (z - z_1(\alpha))(z - z_2(\alpha)) \cdots (z - z_p(\alpha)), \\ B_q(z, \theta) &= (z - \tilde{z}_1(\theta))(z - \tilde{z}_2(\theta)) \cdots (z - \tilde{z}_q(\theta)), \end{aligned}$$

переходим от (10.79б) к соотношению

$$\prod_{i=1}^p \left(1 - \frac{1}{z_i(\alpha)} F_-\right) \varepsilon(t) = \prod_{j=1}^q \left(1 - \frac{1}{\tilde{z}_j(\theta)} F_-\right) \delta(t), \quad (10.79в)$$

представляющему собой еще одну форму записи АРСС(p, q)-процесса.

Наконец отметим связь, существующую между оператором F_- и оператором введенной в п. 10.3.3 *последовательной разности* Δ . Из определения Δ (напомним: $\Delta\varepsilon(t) = \varepsilon(t) - \varepsilon(t-1)$) и F_- непосредственно следует:

$$\Delta = 1 - F_-, \quad (10.91)$$

так как $\Delta\varepsilon(t) = (1 - F_-)\varepsilon(t) = \varepsilon(t) - \varepsilon(t-1)$.

Неоднозначность в определении параметров АРСС-моделей и «прогноз назад». Как известно, поведение всякой последовательности случайных величин $\{\varepsilon(t)\}$ полностью определяется совместным законом распределения вероятностей для многомерных случайных величин вида $(\varepsilon(t_1), \varepsilon(t_2), \dots, \varepsilon(t_N))$ для любого наперед заданного N и любого набора значений $t_1, t_2, \dots, t_N$. В данном пункте мы анализируем определенный подкласс случайных последовательностей, а именно *стационарные в широком смысле временные ряды* $\varepsilon(t)$, $t = 1, 2, \dots$, и каждый из них определяется своей автоковариационной функцией $\gamma(\tau)$. О моделях (временных рядах, процессах), имеющих одну и ту же ковариационную функцию, будем говорить как о моделях с *идентичной ковариационной структурой*. В ходе анализа АР(p)-, СС(q)- и АРСС(p, q)-моделей, идя

от известной автоковариационной функции исследуемого процесса $\gamma(\tau)$ к определению неизвестных значений параметров модели, мы видели, что одной и той же автоковариационной функции могут соответствовать *различные* его представления в виде линейных параметрических моделей (см., например, выше уравнение (10.75) для определения параметра θ в СС(1)-модели). Выбор *единственного* решения осуществлялся с помощью введения соответствующих ограничений на значения параметров α_k ($k = 1, 2, \dots, p$) и θ_j ($j = 1, 2, \dots, q$), обеспечивающих стационарность и обратимость АРСС-модели. Эти ограничения в общем случае формируются в терминах корней $z_1(\alpha), \dots, z_p(\alpha)$ и $\tilde{z}_1(\theta), \dots, \tilde{z}_q(\theta)$ характеристических уравнений, соответственно, (10.90) и (10.90'). Напомним, что *стационарность* процессов АР(p) обеспечивается требованием, чтобы все корни $z_k(\alpha)$ ($k = 1, 2, \dots, p$) лежали бы вне единичного круга, *обратимость* процессов СС(q) — аналогичным требованием к корням $\tilde{z}_j(\theta)$, $j = 1, 2, \dots, q$ (СС — процессы стационарны при любых значениях θ_j), а для стационарности и обратимости АРСС(p, q)-процессов необходимо и достаточно одновременного выполнения обоих этих требований.

Теперь мы можем сформулировать три существенных факта, относящихся к проблеме неоднозначности в определении параметров АРСС-модели (их обоснование можно найти, например, в книге [Бокс, Дженкинс, (1974), с. 218–221]).

1) Все линейные параметрические модели вида

$$\prod_{i=1}^p \left(1 - \frac{1}{z_i(\alpha)} F_{+}^{-}\right) \varepsilon(t) = \prod_{j=1}^q \left(1 - \frac{1}{\tilde{z}_j(\theta)} F_{+}^{-}\right) \delta(t) \quad (10.92)$$

имеют идентичную ковариационную структуру. При этом допускается любая комбинация знаков «+» и «−» в качестве нижних индексов оператора F в левых и правых частях (10.92).

2) Существует *единственное* представление (10.79в) (или равносильные ему представления (10.79), (10.79а), (10.79б)), связывающее в рамках модели АРСС(p, q) значение $\varepsilon(t)$ *только с прошлым этого ряда.*

3) Беря в левой и правой частях соотношения (10.92) оператор сдвига F со знаком «+», приходим к соотношению

$$\prod_{i=1}^p \left(1 - \frac{1}{z_i(\alpha)} F_{+}\right) \varepsilon(t) = \prod_{j=1}^q \left(1 - \frac{1}{\tilde{z}_j(\theta)} F_{+}\right) \delta(t), \quad (10.92')$$

или, что то же,

$$\varepsilon(t) = \alpha_1 \varepsilon(t+1) + \dots + \alpha_p \varepsilon(t+p) + \delta(t) - \theta_1 \delta(t+1) - \dots - \theta_q \delta(t+q). \quad (10.92'')$$

Другими словами, существует стационарное обратимое представление, в котором $\varepsilon(t)$ выражено целиком через будущие значения ε и настоящее и будущие значения $\delta(t)$. На первый взгляд, с прикладной точки

зрения, представление (10.92'') (которое часто называют *возвратным*) кажется искусственным, странным, однако оно оказывается полезным в ситуациях, когда требуется восстанавливать неизвестные (потерянные, «стертые», незарегистрированные) прошлые значения анализируемого процесса, то есть *строить «прогноз назад»*.

Проблема «перепараметризации» при подборе модели. Если отправляться в представлении АРСС(p, q)-модели от (10.79б)–(10.79в), то очевидно, что модель (10.79б) идентична модели

$$(1 - \lambda F_-)A_p(F_-, \alpha) = (1 - \lambda F_-)B_q(F_-, \theta)\delta(t),$$

в которой λ — некоторое число, а операторы авторегрессии и скользящего среднего умножены на один и тот же множитель $(1 - \lambda F_-)$. Поэтому при подборе модели следует позаботиться о том, чтобы в ней не появлялись *избыточные* (с точки зрения экономной параметризации) или «почти избыточные» множители подобного типа.

Например, общий множитель в модели АРСС(2,1)

$$(1 - 1,3F_- + 0,4F_-^2)\varepsilon(t) = (1 - 0,5F_-)\delta(t)$$

можно увидеть только после разложения левой части на множители в соответствии с (10.79в):

$$(1 - 0,5F_-)(1 - 0,8F_-)\varepsilon(t) = (1 - 0,5F_-)\delta(t).$$

Так что на самом деле мы имеем не АРСС(2,1)-модель, а АР(1)-модель вида

$$(1 - 0,8F_-)\varepsilon(t) = \delta(t),$$

или, что то же,

$$\varepsilon(t) = 0,8\varepsilon(t-1) + \delta(t).$$

На практике трудности вызывает не столько наличие *одинаковых* множителей в представлении (10.79в), сколько ситуация, когда имеются *почти одинаковые* множители. Например, пусть истинная модель имеет вид

$$(1 - 0,4F_-)(1 - 0,8F_-)\varepsilon(t) = (1 - 0,5F_-)\delta(t). \quad (10.93)$$

Если делается попытка подогнать эту модель по экспериментальным данным, то можно ожидать крайнюю нестабильность в оценках параметров¹⁶ из-за близости множителей $(1 - 0,4F_-)$ и $(1 - 0,5F_-)$ в разных частях (10.93). В подобных ситуациях следует идти на упрощение модели. Это

¹⁶Под *нестабильностью в оценках* подразумевается их относительно большой разброс, например, статистически незначимое расхождение в оценках параметра 0,4 в левой части и параметра 0,5 в правой части представления анализируемого процесса. Однако вопросы точности оценивания параметров моделей временных рядов остались за рамками нашего учебника (за исключением некоторых частных случаев).

достигается различными способами. Например, для близких значений корней $z_i(\alpha)$ и $\tilde{z}_j(\theta)$ полагают $z_i(\alpha) \approx \tilde{z}_j(\theta)$ и сокращают соответствующие сомножители в левой и правой частях представления (10.79в). В нашем примере (10.93) это привело бы к рассмотрению АР(1)-модели

$$(1 - 0,8F_-)\varepsilon(t) = \delta(t)$$

вместо АРСС(2,1)-модели (10.93), поскольку эти две модели на практике оказались *статистически неразличимы*.

Можно подойти к этому вопросу с несколько иной точки зрения, «загоняя» множители вида $(1 - F_-/\tilde{z}_j(\theta))$ правой части (3.79в) в знаменатель левой части этого соотношения и формально разлагая получившийся таким образом оператор левой части

$$\frac{\prod_{i=1}^p \left(1 - \frac{1}{z_i(\alpha)} F_-\right)}{\prod_{j=1}^q \left(1 - \frac{1}{\tilde{z}_j(\theta)} F_-\right)}$$

по степеням $(F_-/\tilde{z}_j(\theta))$ (при этом полагается, что $|F_-/\tilde{z}_j(\theta)| < 1$, используется формула суммы бесконечно убывающей геометрической прогрессии и в получившемся бесконечном разложении ограничиваются двумя первыми членами). Так, например, реализация этой процедуры применительно к (10.93) дает

$$\begin{aligned} \frac{(1 - 0,4F_-)(1 - 0,8F_-)}{1 - 0,5F_-} &= (1 - 0,4F_-)(1 - 0,8F_-) \times \\ &\times (1 + 0,5F_- + 0,25F_-^2 + 0,125F_-^3 + \dots) = \\ &= 1 - 0,700F_- - 0,030F_-^2 - 0,015F_-^3 - \\ &- 0,008F_-^4 - \dots \end{aligned}$$

Ограничиваясь первыми двумя членами этого разложения и возвращаясь к (10.93), получаем АР(1)-модель

$$(1 - 0,7F_-)\varepsilon(t) = \delta(t),$$

также статистически неразличимую с истинной моделью (10.93).

Подбор структурных параметров модели. Из сказанного следует, что окончательный выбор *структурных параметров* модели (то есть параметров p и q) надо производить *лишь после идентификации первого (априорного) гипотетического вида* АРСС(p, q)-модели. Если в результате оценивания параметров α_k и θ_j и вычисления корней характеристических уравнений (10.90) и (10.90') (в которых вместо значений параметров α_k и θ_j подставлены их оценки, соответственно, $\hat{\alpha}_k$ и $\hat{\theta}_j$) окажутся

близкие по величине пары $z_i(\alpha)$ и $\bar{z}_j(\theta)$, то в представлении (10.79в) соответствующие сомножители в его левой и правой частях должны быть сокращены. После этого значения p и q соответствующим образом подправляются (в сторону их уменьшения, конечно) и весь процесс анализа и идентификации модели повторяется заново. При этом, конечно, следует ориентироваться на характерные свойства моделей, определяемые поведением их автокорреляционных и частных автокорреляционных функций (см. таблицу 10.6).

На заключительном этапе построения модели обычно проводится проверка ее адекватности (*диагностика модели*). Она основана на проверке статистически значимого отличия от нуля неструктурных коэффициентов модели и на анализе оцененных остатков $\hat{\delta}(t)$, которые должны удовлетворять требованиям белого шума. Например, выбрав модель $\text{APCC}(p; q)$, оценив и проверив на статистически значимое отличие от нуля ее коэффициенты $\hat{\alpha}_k (k = 1, \dots, p)$ и $\hat{\theta}_j (j = 1, \dots, q)$, можно оценить «соседние» модели $\text{APCC}(p+1; q)$ и $\text{APCC}(p, q+1)$ и протестировать (на значимое отличие от нуля) возникших при этом дополнительных коэффициентов $\hat{\alpha}_{p+1}$ и $\hat{\theta}_{q+1}$. Проверка, основанная на поведении оцененных остатков $\hat{\delta}(t)$, требует предварительно выразить и оценить эти остатки по модели (общий прием использует представление модели (10.41'') в форме (10.41'), см. выше), а затем — построения по этим остаткам $\hat{\delta}(t)$ соответствующей критической статистики. В качестве последней обычно используется *статистика Льюнга–Бокса* (см. *Ljung G.M., Box G.E.P. On a Measure of Lack of Fit in Time Series Models. — Biometrika, 65 (1978), pp. 297–303*):

$$\gamma = N(N+2) \sum_{\tau=1}^K \frac{1}{N-\tau} \hat{r}(\hat{\delta}_t, \hat{\delta}_{t-\tau}),$$

где

$$\hat{r}(\hat{\delta}_t, \hat{\delta}_{t-\tau}) = \frac{\sum_{t=\tau+1}^N \hat{\delta}_t \hat{\delta}_{t-\tau}}{\sum_{t=1}^N \hat{\delta}_t^2}$$

— это коэффициент автокорреляции между $\hat{\delta}_t$ и $\hat{\delta}_{t-\tau}$, N , как и прежде, общее число наблюдений (длина имеющейся реализации анализируемого ряда), а значение $K > p + q$ выбирается исследователем (обычно значения критической статистики γ вычисляются для разных значений $K = p + q + 1, p + q + 2, \dots$ с учетом ограничений по N).

Таблица 10.6. Характерные свойства моделей, основанные на поведении их автокорреляционных и частных автокорреляционных функций (а.к.ф. и ч.а.к.ф.)

Модель	А.к.ф. $r(k)$	Ч.а.к.ф. $r_{\text{частн.}}(k)$	Характерное свойство
$AR(1):$ $x(t) = \alpha x(t-1) + \delta(t)$	$r(k) = \alpha^k$ $k = 1, 2, \dots$	$r_{\text{частн.}}(k) = 0$ при $k = 2, 3, \dots$	$r_{\text{частн.}}(k) = 0$ при $k = 2, 3, \dots$
$AR(2):$ $x(t) = \alpha x(t-1) + \alpha_2 x(t-2) + \delta(t)$	$r(1) = \frac{\alpha_1}{1-\alpha_2}$ $r(2) = \alpha_2 + \frac{\alpha_1^2}{1-\alpha_2}$ $r(k) = \alpha_1 r(k-1) + \alpha_2 r(k-2)$ при $k = 3, 4, \dots$	$r_{\text{частн.}}(2) = \frac{r(2)-r^2(1)}{1-r^2(1)}$ $r_{\text{частн.}}(k) = 0$ при $k = 3, 4, \dots$	$r_{\text{частн.}}(k) = 0$ при $k = 3, 4, \dots$
$AR(p):$ $x(t) = \alpha_1 x(t-1) + \alpha_2 x(t-2) + \dots + \alpha_p x(t-p) + \delta(t)$	$r(k)$ убывает по k как затухающая экспонента или затухающая синусоида	$r_{\text{частн.}}(k) = 0$ при $k = p+1, p+2, \dots$	$r_{\text{частн.}}(k) = 0$ при $k = p+1, p+2, \dots$
$MA(1):$ $x(t) = \delta(t) - \theta \delta(t-1)$	$r(1) = \frac{\theta}{1+\theta^2}$ $r(k) = 0$ при $k = 2, 3, \dots$	$r_{\text{частн.}}(k)$ экспоненциально по k	$r(k) = 0$ при $k = 2, 3, \dots$
$MA(2):$ $x(t) = \delta(t) - \theta_1 \delta(t-1) - \theta_2 \delta(t-2)$	$r(1) = -\frac{\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2}$ $r(2) = -\frac{\theta_2}{1+\theta_1^2+\theta_2^2}$ $r(k) = 0$ при $k = 3, 4, \dots$	$r_{\text{частн.}}(k)$ убывает по k как затухающая экспонента и/или затухающая синусоида	$r(k) = 0$ при $k = 3, 4, \dots$

Продолжение таблицы 10.6.

Модель	А.к.ф. $r(k)$	Ч.к.ф. $r_{\text{частн.}}(k)$	Характерное свойство
$MA(q):$ $x(t) = \delta(t) - \theta_1\delta(t-1) - \dots - \theta_q\delta(t-q)$	$r(k) = 0$ при $k = q + 1, q + 2, \dots$	$r_{\text{частн.}}(k)$ убывает по k как затухающая экспонента и/или затухающая синусоида	$r(k) = 0$ при $k = q + 1, q + 2, \dots$
$ARMA(1; 1):$ $x(t) = \alpha x(t-1) + \delta(t) - \theta\delta(t-1)$	$r(k)$ экспоненциально (при $\alpha > 0$) или осциллирующе (при $\alpha < 0$ убывает, начиная с $k = 1$; знак $r(1)$ совпадает со знаком $\alpha - \theta$)	$r_{\text{частн.}}(k)$ осциллирующе (при $\alpha > 0$) или экспоненциально (при $\alpha < 0$ убывает, начиная с $k = 1$; $r_{\text{частн.}}(1) = r(1)$)	См. поведение $r(k)$ и $r_{\text{частн.}}(k)$ в предыдущих двух столбцах
$ARMA(p; q):$ $x(t) = \alpha_1 x(t-1) + \dots + \alpha_p x(t-p) + \delta(t) - \theta_1\delta(t-1) - \dots - \theta_q \cdot \delta(t-q)$	Осциллирующее или прямое убывание $r(k)$, начиная с $k = q$	$r_{\text{частн.}}(k)$ осциллирующе или, прямо убывает, начиная с $k = p$	См. поведение $r(k)$ и $r_{\text{частн.}}(k)$ в предыдущих двух столбцах

В условиях справедливости проверяемой гипотезы (об адекватности выбранной модели) статистика γ должна подчиняться (асимптотически по $n \rightarrow \infty$) $\chi^2(K - p - q)$ -распределению. Так что, неравенство $\gamma > \chi^2_\alpha(K - p - q)$ сигнализирует о неудачном выборе модели с вероятностью ошибки, равной α ($\chi^2_\alpha(m)$, как и прежде, 100%-ная точка распределения $\chi^2(m)$).

В качестве косвенных критериев правильности выбора модели, учитывающих упомянутый выше эффект ее *перепараметризации*, используются также различные *информационные критерии*, построенные на идее штрафа, налагаемого за увеличение числа параметров (которое, само по себе, снижает в то же время средний квадрат ошибки модели $\hat{\sigma}^2$). Среди наиболее распространенных (и представленных в соответствующих пакетах программ) критериев такого рода

- *информационный критерий Akaike (AIC-критерий)*

$$AIC = \ln \hat{\sigma}^2 + \frac{(p + q) \cdot 2}{N}$$

(см. Akaike H. Information Theory and an Extension of the Maximum Likelihood Principle. — In: Second International Symposium on Information Theory. Budapest, 1973, pp. 267–281);

- *байесовский информационный критерий Шварца (BIC-критерий)*

$$BIC = \ln \hat{\sigma}^2 + \frac{(p + q) \cdot \ln N}{N}$$

(см. Schwarz G. Estimating the Dimension of a Model. — Annals of Statistics, 6(1978), pp. 461–464).

Оба критерия построены на компромиссе между качеством подгонки модели, характеризуемом значением соответствующей функции правдоподобия или величиной среднего квадрата ошибки

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{t=1}^N \hat{\delta}_t^2,$$

и штрафом за «перематризацию» (который, как мы видим, в критерии Шварца выше, что часто отвечает более адекватному выбору). Обычно выбирают модель среди вариантов с наименьшими значениями AIC и BIC.

Статистическое оценивание неструктурных параметров. Современные, реализованные в эконометрических и статистических пакетах программ, способы оценивания параметров α_k ($k = 1, \dots, p$) и

$\theta_j (j = 1, \dots, q)$ моделей $APCC(p; q)$ используют *метод наименьших квадратов* и различные версии *методы максимального правдоподобия*. Реализация МНК в $APCC(p; 0)$ — (то есть в $AR(p)$) моделях не представляет трудностей, так как в уравнении

$$\varepsilon(t) = \alpha_1 \varepsilon(t-1) + \dots + \alpha_p \varepsilon(t-p) + \delta(t)$$

лагированные объясняющие переменные $\varepsilon(t-\tau)$ в правой части по построению *не коррелированы* с $\delta(t)$ и, следовательно, обычные МНК-оценки параметров $\alpha_1, \dots, \alpha_p$ состоятельны.

Для реализации МНК в моделях $APCC(p; q)$ при $q \geq 1$ требуется предварительно преобразовать модель к виду (10.41').

Правомерность использования *метода максимального правдоподобия* (и, соответственно, ММП-оценок) основана на гипотезе нормальности остатков $\delta(t)$ и на максимизации так называемой *условной функции правдоподобия* (в качестве условий фиксируются начальные значения ε и δ). Подробнее с процедурой вычисления ММП-оценок параметров читатель может познакомиться, например, по книгам [Вербик (2008)], [Greene (2000)].

В эконометрической литературе и приложениях обсуждаются и используются также *многомерные* (или *векторные*) модели $APCC$ и, как их частные случаи, векторные AR - и CC -модели. Эта тема относится к проблематике анализа *многомерных* временных рядов и систем одновременных уравнений, которая выходит за рамки данного издания.

10.4.5 Простая и обобщенная модели авторегрессионных условно гетероскедастичных остатков

В ряде прикладных эконометрических работ, в частности, при анализе и моделировании макроэкономических данных, характеризующих процессы инфляции и внешней торговли, механизмы формирования нормы процента и т. п.¹⁷, была выявлена некоторая общая закономерность в поведении случайных остатков (ошибок прогноза) исследуемых моделей: их малые и большие значения *группировались целыми кластерами*, или *сериями*. Причем это не приводило к нарушению их стационарности и, в частности, их гомоскедастичности для относительно больших временных интервалов, то есть гипотеза $D\varepsilon(t) = \gamma(0) = \text{const}$ не противоречила имеющимся экспериментальным данным. Однако в рамках

¹⁷См. Engle R. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. «Econometrica», 50 (1982), pp. 987–1008; Engle R. Estimates of the Variance of U.S. Inflation Based on the ARCH Model. «Journal of Money, Credit, and Banking», 15 (1983), pp. 286–301; Cragg J. More Efficient Estimation in the Presence of Heteroscedasticity of Unknown Form. «Econometrica», 51 (1983), pp. 751–763.

моделей АРСС удовлетворительно объяснить этот феномен не удалось. Требовалась определенная модификация известных моделей.

Такая модификация была предложена впервые Р. Энглom в 1982 г. (см. работу 1982 г. в сноске¹⁷). Он рассматривал остатки $\varepsilon(t)$ как *условно гетероскедастичные*, связанные друг с другом простейшей авторегрессионной зависимостью, а именно:

$$\begin{cases} [\varepsilon(t) | \varepsilon(t-1)] \in N(0; \sigma_t^2), \\ \text{где } \sigma_t^2 = D(\varepsilon(t) | \varepsilon(t-1)) = \theta_0 + \theta_1 \varepsilon^2(t-1), \end{cases} \quad (10.94)$$

или, что то же,

$$\varepsilon(t) = \delta(t) [\theta_0 + \theta_1 \varepsilon^2(t-1)]^{\frac{1}{2}}, \quad (10.94')$$

где последовательность $\delta(t)$, $t = 1, 2, \dots$, — образует *стандартизованный нормальный* белый шум (то есть $\delta(t_1)$ и $\delta(t_2)$ независимы при $t_1 \neq t_2$ и $\delta(t) \in N(0; 1)$), а параметры θ_0 и θ_1 должны удовлетворять ограничениям, обеспечивающим *безусловную* гомоскедастичность $\varepsilon(t)$ (такими ограничениями являются требования $\theta_0 > 0, |\theta_1| < 1$). При этом под $[\varepsilon(t) | \varepsilon(t-1)]$ мы подразумеваем, что речь идет о случайной величине, рассматриваемой в предположении, что ее значение в предшествующий момент времени *зафиксировано* (задано). Соответственно, ее поведение будет описываться *условным* законом распределения вероятностей.

Следуя установившейся терминологии, будем называть модель (10.94) *авторегрессионной условно гетероскедастичной* (сокращенно АРУГ). В англоязычной литературе такие модели называют **AutoRegressive Conditional Heteroscedasticity** (сокращенно ARCH-model).

Использование такой модели для описания поведения остатков моделей регрессии и временных рядов в упомянутых выше типовых ситуациях оказывается более адекватным действительности и позволяет строить более эффективные оценки параметров рассматриваемых моделей, чем обычные или даже обобщенные МНК-оценки (см., например, описание алгоритма построения *нелинейных* оценок максимального правдоподобия для параметров линейной модели множественной регрессии с авторегрессионными и условно гетероскедастичными остатками в книге [Greene W. H. (2000), pp. 439–440]; получающиеся оценки оказываются более эффективными, чем даже наиболее эффективные, в *классе линейных оценок*, МНК-оценки).

Естественное обобщение моделей типа (10.94) было предложено Р. Энглom и Д. Крафтом в 1983 г. (см. *Engle R., Kraft D.* in «Applied Time Series Analysis of Economic Data», Washington D. C.: Bureau of the Census, 1983):

$$\begin{cases} [\varepsilon(t) | \varepsilon(t-1)] \in N(0; \sigma_t^2), \\ \text{где } \sigma_t^2 = \theta_0 + \theta_1 \varepsilon^2(t-1) + \dots + \theta_q \varepsilon^2(t-q), \end{cases} \quad (10.94a)$$

а параметры $\theta_0, \theta_1, \dots, \theta_q$ связаны некоторыми ограничениями, обеспечивающими *безусловную* гомоскедастичность остатков $\varepsilon(t)$.

Модели (10.94а) называются моделями АРУГ порядка q (сокращенно АРУГ(q)). Очевидно, модель (10.94) является АРУГ(1)-моделью и соответствует частному случаю (10.94а) при $q = 1$. Содержательно переход к $q > 1$ в моделях (10.94а) означает, что процесс формирования значений остатков $\varepsilon(t)$ имеет «более длинную память» о величинах предшествующих остатков $\varepsilon(t-1)$, $\varepsilon(t-2)$, ... Кстати, АРУГ(q)-модель (10.94а) может рассматриваться как *некая специальная форма СС(q)-модели*, что и используется при ее анализе.

Дальнейшее обобщение моделей этого типа было сделано в 1986 г. Т. Боллерслевом (см. *Bollerslev T. Generalized Autoregressive Conditional Heteroscedasticity. «Journal of Econometrics», 31 (1986), pp. 307–327*). Он предложил описывать поведение остатков $\varepsilon(t)$ с помощью *обобщенной авторегрессионной условно гетероскедастичной модели* (ОАРУГ-модели, или, в англоязычном варианте, GARCH-model), которая записывается в виде

$$\begin{cases} [\varepsilon(t) | \psi(t)] \in N(0; \sigma_t^2), \\ \text{где условная дисперсия } \sigma_t^2 = \mathbf{D}(\varepsilon(t) | \psi(t)) \text{ имеет вид} \\ \sigma_t^2 = \alpha_1 \sigma_{t-1}^2 + \alpha_2 \sigma_{t-2}^2 + \dots + \alpha_p \sigma_{t-p}^2 + \theta_0 + \theta_1 \varepsilon(t-1) + \dots + \theta_q \varepsilon(t-q). \end{cases} \quad (10.95)$$

В соотношениях (10.95) под $\psi(t)$ подразумевается *вся информация о процессе $\varepsilon(t)$, которой мы располагаем к моменту времени t* (то есть — все значения $\varepsilon(\tau)$ и σ_τ^2 для $\tau < t$), а параметры α_k и θ_j ($k = 1, 2, \dots, p$; $j = 0, 1, \dots, q$) связаны ограничениями, обеспечивающими *безусловную* гомоскедастичность остатков $\varepsilon(t)$. Модель ОАРУГ(p, q), задаваемая соотношениям (10.95), может интерпретироваться как специальная форма АРСС(p, q)-модели. На ряде примеров показано, что использование ОАРУГ(p, q)-модели позволяет добиваться *более экономной параметризации* в описании поведения остатков $\varepsilon(t)$, чем в рамках АРУГ(q)-моделей (то есть модели ОАРУГ(p, q) при малых значениях p и q оказываются более точными, чем АРУГ(q)-модели при больших значениях q).

Достаточно полный критический обзор, посвященный описанию АРУГ- и ОАРУГ-моделей и их приложениям в экономике и финансах, читатель найдет, например, в журнале «Обозрение прикладной и промышленной математики», серия «Финансовая и страховая математика», т. 3 (1996), вып. 6.

10.5 Модели нестационарных временных рядов и их идентификация

10.5.1 Нестационарность и тренды: детерминированный и стохастический

До сих пор в этой главе мы занимались анализом и моделированием случайных остатков $\varepsilon(t)$, с которыми приходится иметь дело как в обобщенных линейных моделях множественной регрессии (5.1), так и в моделях динамических показателей (10.2).

Но если в моделях регрессии типа (5.1) с нестохастическими объясняющими переменными удовлетворительные (в прикладном смысле) результаты в широком классе случаев можно получить, полагая стохастические остатки $\varepsilon(t)$ стационарными, то при анализе реальных динамических показателей $x(t)$ (см. (10.2)), встречающихся в экономике, как правило, приходится считаться с их нестационарностью. При этом, к относительно *«легкому случаю»* нестационарности можно отнести нестационарность $x(t)$, выражающуюся только в наличии **детерминированного тренда**, то есть неслучайной составляющей $f(t)$, объединяющей в себе, в общем случае, три первых слагаемых правой части (10.2) (соответственно, *ряд стохастических остатков $\varepsilon(t)$ тогда будет стационарным*). В этих случаях анализируемые ряды называют **«тренд-стационарными»** или **«ТС-рядами»**. Наряду с этим нестационарность ряда $x(t)$ может выражаться также в наличии так называемого **«стохастического тренда»**, определяемого зависимостью от времени t его моментов более высокого (чем первый) порядка. Среди подобных нестационарных рядов существуют такие, которые становятся стационарными после их «дифференцирования», то есть *после перехода к их последовательным разностям $\Delta x(t), \Delta^2 x(t), \dots$ определенного порядка* (см. п. 10.3.3). Такие ряды принято называть **«дифференциально-стационарными»** или **«ДС-рядами»**.

Заметим, что если ТС-ряд $x(t)$ имеет детерминированную составляющую в форме *алгебраического от времени t* , то есть

$$x(t) = \beta_0 + \beta_1 t + \dots + \beta_k t^k + \varepsilon(t),$$

то k -кратное дифференцирование такого ряда (то есть переход к последовательной разности $x_k(t) = \Delta^k x(t)$) также приведет к его «остационариванию» (см. выше, п. 10.3.3). Однако, как мы увидим, два ряда — ТС- и ДС — «остационаренные» с помощью дифференцирования, имеют свои специфические различия, обуславливающие разницу в их интерпретации и методах дальнейшего прикладного анализа.

З а м е ч а н и е. Поскольку до сих пор в этой главе мы рассматривали модели *остатков* $\varepsilon(t)$, то их средние значения мы полагали рав-

ными нулю. Соответственно, в частности, в правой части общей формулы (10.79) модели АРСС отсутствуют члены, определяющие отличное от нуля среднее значение анализируемого ряда. Теперь же, переходя к анализу *динамических показателей* $x(t)$, мы будем допускать существование их ненулевых средних значений $\mu = \mathbf{E}x(t)$. Тогда общая форма моделей АРСС($p; q$) принимает вид

$$x(t) - \mu = \sum_{k=1}^p \alpha_k (x(t-k) - \mu) + \delta(t) - \sum_{j=1}^q \theta_j \delta(t-j) \quad (10.79''')$$

или, что то же:

$$x(t) = a + \sum_{k=1}^p \alpha_k x(t-k) + \delta(t) - \sum_{j=1}^q \theta_j \delta(t-j), \quad (10.79'''a)$$

где

$$a = \mu \left(1 - \sum_{k=1}^p \alpha_k \right).$$

Рассмотрим несколько примеров ТС- и ДС-рядов.

Пример 10.7. Ряд, содержащий детерминированный тренд.

Рассмотрим ряд $x(t)$ вида

$$x(t) = a + bt + \delta(t), \quad t = 1, 2, \dots, N, \quad (10.96)$$

где остатки $\delta(t)$ удовлетворяют условиям белого шума, то есть $\mathbf{E}\delta(t) = 0$, $\mathbf{D}\delta(t) = \sigma_0^2$ и $\text{cov}(\delta(t_1), \delta(t_2)) = 0$ при $t_1 \neq t_2$. Поскольку $\mathbf{E}x(t) = a + bt$ — функция, зависящая от времени, то ряд (10.96) — нестационарный. Однако уже первая его разность $\Delta x(t) = x(t) - x(t-1) = b + \delta(t) - \delta(t-1)$ будет удовлетворять условиям стационарности (см. выше Определение 3). Мы видим, при этом, что ряд $\Delta x(t)$ относится к классу *необратимых* СС(1)-моделей (так как $\theta_1 = 1$), а это означает (см. п. 10.4.2 и формулу (10.66')), что в соответствующем авторегрессионном представлении он будет зависеть от своих прошлых значений с весами, не убывающими по мере удаления в прошлое (свойство, мало пригодное для практического использования такого ряда). Тем не менее, ряд (10.96) относится к категории ТС-рядов.

Пример 10.8. Ряд, содержащий стохастический тренд (процесс случайного блуждания).

Проанализируем теперь ряд

$$x(t) = x(t-1) + \delta(t), \quad t = 0, 1, \dots, N, \quad (10.97)$$

который формально может быть отнесен к классу АР(1)-моделей (правда, с коэффициентом $\alpha = 1$, не удовлетворяющим условиям стационарности). Чтобы разобраться в природе нестационарности этого ряда, выразим первое слагаемое правой части (10.97) последовательно $t-1$ раз

по формуле (10.97), то есть

$$\begin{aligned} x(t) &= (x(t-2) + \delta(t-1)) + \delta(t) = (x(t-3) + \delta(t-2)) + \\ &+ \delta(t-1) + \delta(t) = \dots = x(0) + \sum_{k=1}^t \delta(k). \end{aligned} \quad (10.97')$$

Используя (10.97') и свойства белого шума $\{\delta(1), \delta(2), \dots, \delta(t)\}$, вычислим условные (при условии заданности значения $x(0)$) и безусловные 2-е моменты ряда $x(t)$.

$$\mathbf{D}(x(t) | x(0)) = \mathbf{D} \left(\sum_{k=1}^t \delta(k) \right) = \sum_{k=1}^t \mathbf{D} \delta(k) = t\sigma_0^2.$$

Поскольку правая часть не зависит от значения $x(0)$, то и безусловная дисперсия $\mathbf{D}x(t) = t\sigma_0^2$.

$$\begin{aligned} \text{Cov}(x(t), x(t-1) | x(0)) &= \mathbf{E} [(x(t) - x(0))(x(t-1) - x(0)) | x(0)] = \\ &= \mathbf{E} \left[\left(\sum_{k=1}^t \delta(k) \right) \left(\sum_{k=1}^{t-1} \delta(k) \right) \middle| x(0) \right] = (t-1)\sigma_0^2. \end{aligned}$$

И здесь правая часть не зависит от $x(0)$, так что и безусловная ковариация $\text{cov}(x(t), x(t-1)) = (t-1)\sigma_0^2$. Отсюда следует, в частности, что коэффициент корреляции

$$r(x(t), x(t-1)) = \frac{\text{cov}(x(t), x(t-1))}{\sqrt{\mathbf{D}x(t)\mathbf{D}x(t-1)}} = \frac{(t-1)\sigma_0^2}{\sqrt{t\sigma_0^2(t-1)\sigma_0^2}} = \sqrt{1 - \frac{1}{t}}.$$

Мы видим, что вторые моменты ряда $x(t)$ зависят от времени (присутствует стохастический тренд) и что соседние значения анализируемого ряда $x(t)$ все более тесно (по мере роста t) положительно коррелированы. А это значит, что попав после первых нескольких шагов в определенное состояние (выше или ниже начального уровня $x(0)$), траектория $x(t)$ может в течение относительно длительного времени оставаться, соответственно, выше или ниже этого уровня, не пересекая его. При этом варьирование $x(t)$ относительно своего начального значения $x(0)$ будет все более «раскачиваться» с течением времени t , т. к. дисперсия $\mathbf{D}x(t) = t \cdot \sigma_0^2$ растет пропорционально $t!$ (см. ниже рис. 10.13).

Обратим внимание читателя на тот факт, что и в данном случае «дифференцирование» анализируемого ряда $x(t)$ (то есть переход к его первой разности $x_1(t) = \Delta x(t) = x(t) - x(t-1)$), как это непосредственно следует из (10.97), приводит к его «остационариванию» (то есть $\Delta x(t) = \delta(t)$). Однако в отличие от «остационаренного» ТС-ряда, рас-

смотренного в примере 10.7, «остационаренный» ДС-ряд (10.97), как и сам ряд (10.97), широко используется в экономических приложениях¹⁸

Пример 10.9. **Ряд, содержащий и детерминированный, и стохастический тренды (процесс случайного блуждания с линейным сносом).**

Теперь добавим в правую часть соотношения (10.97) константу a и рассмотрим ряд вида

$$x(t) = a + x(t-1) + \delta(t), \quad t = 0, 1, \dots, N. \quad (10.98)$$

Последовательное $(t-1)$ -кратное выражение второго слагаемого правой части (10.98) по формуле (10.98) дает:

$$\begin{aligned} x(t) &= a(a + x(t-2) + \delta(t-1)) + \delta(t) = 2a + (a + x(t-3) + \delta(t-2)) + \\ &+ \delta(t-1) + \delta(t) = \dots = ta + x(0) + \sum_{k=1}^t \delta(k). \end{aligned} \quad (10.98')$$

Используя правую часть (10.98') для вычисления условных (при условии заданности начального значения $x(0)$) среднего значения и дисперсии ряда $x(t)$, имеем:

$$\mathbf{E}(x(t) | x(0)) = ta + x(0); \quad (10.99)$$

$$\mathbf{D}(x(t) | x(0)) = t\sigma_0^2. \quad (10.100)$$

Мы видим, что и среднее значение, и дисперсия ряда $x(t)$ (при заданном значении $x(0)$) ведут себя как линейные функции времени, а это значит, что ряд содержит оба тренда: *детерминированный* (10.99) и *стохастический* (10.100), — и соответственно, является нестационарным. Из (10.98) непосредственно следует, что дифференцирование ряда (переход к его 1-й разности) дает ряд $x_1(t) = \Delta x(t) = a + \delta(t)$, который удовлетворяет всем требованиям стационарности. Таким образом, *ряд (10.98) является ДС-рядом.*

Очевидно траектория ряда (10.98) будет виться около прямой (10.99), обнаруживая при этом все свойства случайного блуждания (то есть все больше «раскачиваясь» вокруг этой прямой, относительно редко ее пересекая). Поэтому его часто называют «*процессом случайного блуждания с линейным сносом*».

Сравнивая ряды (10.96) (ТС-ряд с детерминированным линейным трендом) и (10.98') (ДС-ряд с линейным детерминированным и стохастическим трендами), еще раз подчеркнем разницу, существующую между

¹⁸Это не означает, что сами ТС-ряды не используются в экономических приложениях. Используются, и весьма широко! Однако процедуру выделения (элиминирования) неслучайной составляющей такого ряда можно строить и методами, отличными от «дифференцирования» (см. выше п. 10.3).

ТС- и ДС-рядами: если исключение детерминированного тренда (например, с помощью методов, описанных в п. 10.3.2) из ТС-ряда делает его стационарным, то эта же операция, примененная к ДС-ряду, оставляет его нестационарным (он по-прежнему будет содержать стохастический тренд!).

10.5.2 Проблема выявления нестационарности ряда

Проблема выявления нестационарности ряда тесно связана с проблемой отнесения анализируемого экономического ряда к одному из двух типов — ТС-рядам или ДС-рядам. В общем случае эти проблемы сложны и запутанны (см., например, [Enders (1995)], [Hamilton (1994)], [Hatanaka (1996)]). В указанных работах можно найти описание многих процедур, нацеленных на решение упомянутых выше двух проблем, причем, подчас применение разных процедур к одним и тем же временным рядам приводит к противоположным выводам. Тем не менее, построение модели конкретного экономического ряда, достаточно дееспособной в задачах описания динамики и прогноза значений рассматриваемого показателя, а также — в построении адекватных моделей связей этого показателя с другими динамическими показателями (то есть моделей *многомерных* временных рядов), практически невозможно без выяснения природы анализируемого временного ряда и рядов, с ним связываемых.

В данном учебнике мы ограничимся рассмотрением этих проблем *лишь в классе моделей АРСС*. Прежде всего посмотрим, как ведут себя траектории простейшей АРСС-модели авторегрессии первого порядка в зависимости от значения корня ее характеристического уравнения.

Сравнение стационарных и нестационарных АРСС-моделей. С этой целью рассмотрим позаимствованные из [Носко (2002)] смоделированные реализации шести вариантов процесса АР(1):

$$x(t) = \alpha x(t-1) + \delta(t), \quad (10.101)$$

где $\delta(t)$ — элементы нормального белого шума с $E\delta(t) = 0$ и $D\delta(t) = 1$, а $\alpha = 0; 0,5; 0,7; 0,9; 1,0; 1,05$ (последние два варианта представляют *нестационарные* версии АР(1)). На рис. 10.13 ((a)–(e)) представлены траектории соответствующих вариантов АР(1)-процесса.

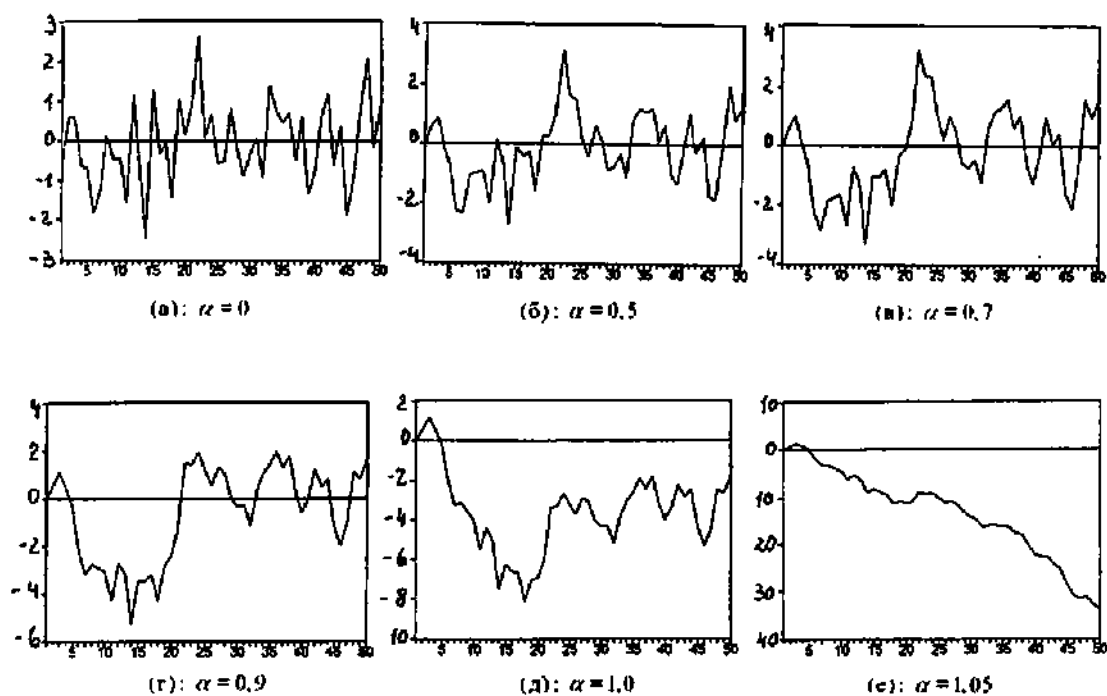


Рис. 10.13. Реализации временного ряда $AR(1)$ (см. (10.101)) при различных значениях параметра α

В действительности с помощью метода Монте-Карло моделировались траектории ряда $AR(1)$ длины $N = 500$. На рис. 10.13 представлены лишь первые 50 наблюдений этих траекторий. Начальное значение $x(0)$ полагалось равным нулю (соответственно, $Ex(t) = 0$).

Мы видим, что поведение смоделированных траекторий при разных значениях параметра α оказалось существенно различным. Эти различия выражаются, в первую очередь, в числе пересечений траекториями начального (в данном случае — нулевого) уровня, в интенсивности однонаправленного отклонения от этого уровня и, соответственно, в значениях выборочного среднего по выборке ограниченного объема (в нашем случае по выборке объема $N = 50$), см. таблицу 10.7.

Таблица 10.7. Характеристики поведения траекторий ряда $AP(1)$ при различных значениях параметра α

№№ пп	Параметр α	Корень характери- стического уравнения z	Число пересечений нулевого уровня	Среднее значение ряда по выборке объема $N = 50$
1	0,00	∞	25	-0,046
2	0,50	2,00	14	-0,097
3	0,70	1,43	8	-0,191
4	0,90	1,11	8	-0,649
5	1,00	1,00	1	-3,582
6	1,05	0,95	1	-13,511

Отметим, что из шести рассмотренных вариантов модели $AP(1)$ первые пять нам хорошо знакомы, в том числе первая (с $\alpha = 0$), определяющая в качестве частного случая $AP(1)$ модель белого шума, и пятая (с $\alpha = 1$), являющаяся моделью нестационарного процесса случайного блуждания, рассмотренного в начале этого параграфа.

Обратим внимание на последнюю модель с $\alpha = 1,05$. Поведение соответствующей траектории в англоязычной литературе называют «*взрывным*» (*explosive*), имея в виду как бы ускоряющееся ее удаление от начального уровня в течение достаточно протяженного отрезка времени. В экономических приложениях такого рода ряды практически не встречаются, в отличие от другого представителя нестационарных рядов — модели случайного блуждания, встречающейся достаточно часто и характеризующейся *единичным значением корня характеристического уравнения*.

Именно предположение о том, что в реально встречающихся в эконометрической практике временных рядах, описываемых моделями $APCC$, либо все корни характеристического уравнения (10.90) находятся вне единичного круга (тогда соответствующий ряд стационарен), либо хотя бы один из корней попадает на границу этого круга (**но нет корней внутри круга**, как это случилось в нашем искусственно смоделированном варианте ряда $AP(1)$ под номером 6, см. таблицу 10.6), лежит в основе предложенных Дики и Фуллером серии процедур под общим названием «**проверка гипотезы единичного корня**» ([Dickey, Fuller (1979)]).

Критерии Дики–Фуллера (простой и расширенный). Эти критерии предназначены для проверки гипотезы о нестационарности $APCC$ -рядов. Точнее, с помощью этих критериев проверяется гипотеза о том, что анализируемый AP -ряд принадлежит классу $ДС$ -рядов при альтернативе, что ряд стационарен или принадлежит классу $ТС$ -рядов (в последнем случае его «остационаривают» оценкой и исключением де-

терминированного тренда). То, что в основе построений лежит *авторегрессионная* модель, а применимость процедур распространяется и на АРСС-модели, объясняется тем, что любая обратимая АРСС-модель имеет представление в форме АР.

В рамках простого и расширенного критериев Дики–Фуллера рассматриваются три варианта модели (в зависимости от наличия/отсутствия в ней свободного члена или детерминированного линейного тренда).

Простой критерий Дики–Фуллера (в англоязычной литературе — **DF-тест**: *Dickey-Fuller Test*) базируется во всех трех упомянутых вариантах на простейшей модели авторегрессии АР(1).

В а р и а н т 1 :

$$x(t) = \alpha x(t-1) + \delta(t)$$

или, что то же:

$$\Delta x(t) = (\alpha - 1)x(t-1) + \delta(t). \quad (10.101')$$

Проверяется гипотеза:

$$H_0: \alpha = 1 \text{ (или, что то же: } \beta = \alpha - 1 = 0), \quad (10.102)$$

$$H_1: |\alpha| < 1 \text{ (или, что то же: } \beta < 0), \quad (10.102')$$

что означает проверку гипотезы о том, что корень характеристического уравнения модели (10.101) равен единице. К сожалению, как показали Дики и Фуллер, стандартная статистика

$$\gamma(N) = \frac{\hat{\beta}}{s_{\hat{\beta}}}, \quad (10.103)$$

с помощью которой проверялась бы гипотеза о нулевом значении коэффициента регрессии β в классической модели парной регрессии $\Delta x(t)$ по $x(t-1)$, см. (10.101') (где $\hat{\beta}$ — МНК-оценка коэффициента $\beta = 1 - \alpha$, а $s_{\hat{\beta}}$ — среднеквадратическая ошибка этой оценки), в данной схеме не подчиняется $t(N-1)$ -распределению (в условиях справедливости гипотезы H_0 и нормальности $\beta(t)$). Дики и Фуллер исследовали распределение статистики (10.103) в условиях справедливости гипотезы H_0 и вычислили соответствующие критические значения этой статистики для определенных наборов значений уровня значимости критерия ε и длин траекторий анализируемого ряда N . Выяснилось, в частности, что вычисленные ими критические значения $\gamma_{\varepsilon}(N)$ статистики $\gamma(N)$ (то есть такие значения, что при $\gamma(N) < \gamma_{\varepsilon}(N)$ гипотеза H_0 о нестационарности ряда отвергается) оказались существенно меньшими соответствующих классических процентных точек $t(N-1)$ -распределения. А это значит, что если бы мы пользовались стандартными процедурами, принятыми в парной регрессии, то мы бы часто необоснованно принимали решение

о стационарности нашего ряда. В таблице 10.8 приведены критические значения статистики критерия Дики–Фуллера для уровней значимости критерия $\varepsilon = 0,05$ и $0,01$ и для ряда значений длин траектории N анализируемого временного ряда.

Таблица 10.8. Критические значения $\gamma_\varepsilon(N)$ для статистики $\gamma(N)$ критерия Дики–Фуллера*)

Длина траектории ряда N	Модель (10.101') (без константы и тренда)		Модель (10.104') (с константой, без тренда)		Модель (10.105') (с константой и трендом)	
	$\varepsilon = 0,05$	$\varepsilon = 0,01$	$\varepsilon = 0,05$	$\varepsilon = 0,01$	$\varepsilon = 0,05$	$\varepsilon = 0,01$
$N = 25$	-1,95	-2,66	-3,00	-3,75	-3,60	-4,38
$N = 50$	-1,95	-2,62	-2,93	-3,58	-3,50	-4,15
$N = 100$	-1,95	-2,60	-2,89	-3,51	-3,45	-4,04
$N = 250$	-1,95	-2,58	-2,88	-3,46	-3,43	-3,99
$N = 500$	-1,95	-2,58	-2,87	-3,44	-3,42	-3,98
$N = \infty$	-1,95	-2,58	-2,86	-3,43	-3,41	-3,96

*) Из книги *Fuller W.A. (1976). Introduction to Statistical Time Series.* — John Wiley and Sons, New York, p. 373.

В а р и а н т 2 :

$$x(t) = a + \alpha x(t-1) + \delta(t) \quad (10.104)$$

или, что то же:

$$\Delta x(t) = a + (\alpha - 1)x(t-1) + \delta(t). \quad (10.104')$$

При таком варианте $AR(1)$ -модели допускается наличие константы a в правой части $AR(1)$ -модели, и это оказывает влияние на распределение критической статистики (10.103) в задаче проверки гипотезы (10.102) при альтернативе (10.102'). Поэтому критические значения $\gamma_\varepsilon(N)$ для этого варианта модели, как мы видим из таблицы 10.8, отличаются от соответствующих критических значений для модели (10.101'). Заметим, что из нестационарности анализируемого $AR(1)$ -ряда (то есть при $\alpha = 1$) автоматически следует и $a = 0$ (так как $a = \mu(1 - \alpha)$, см. (10.79'''–(10.79'''a)).

В а р и а н т 3 :

$$x(t) = a + bt + \alpha x(t-1) + \delta(t) \quad (10.105)$$

или, что то же:

$$\Delta x(t) = a + bt + (\alpha - 1)x(t-1) + \delta(t). \quad (10.105')$$

Здесь мы имеем ситуацию, когда наряду с возможным *стохастическим* трендом (наличие которого и проверяется статистически) присутствует тренд *детерминированный*. Это также вносит свою специфику в

закон распределения критической статистики (10.103), что и отражено в последних двух столбцах таблицы 10.8.

Итак, критерий Дики–Фуллера (DF-тест) предназначен для проверки нестационарности рядов, представимых на базе АР(1)-моделей в форме (10.101'), (10.104') или (10.105'), и реализуется с помощью построения МНК-оценок $\hat{\beta}$ параметра $\beta = \alpha - 1$ в упомянутых моделях и проверки этих оценок на статистически значимое отличие от нуля с помощью критических значений, приведенных в таблице 10.8.

Расширенный критерий Дики–Фуллера (или ADF-тест: Augmented Dickey–Fuller Test) предназначен для проверки нестационарности временного ряда $x(t)$, представленного соотношениями (10.101'), (10.104') или (10.105'), в которых модель АР(1) заменена более общей моделью АР(p), $p \geq 1$. В этом случае (10.101') может быть представлена в форме

$$\Delta x(t) = (\alpha_1 + \alpha_2 + \dots + \alpha_p - 1)x(t-1) + c_1 \Delta x(t-1) + \dots + c_{p-1} \Delta x(t-p+1) + \delta(t), \quad (10.106)$$

и проверяется гипотеза

$$H_0: \alpha_1 + \alpha_2 + \dots + \alpha_p - 1 = 0. \quad (10.107)$$

Действительно, тождество (10.107) означает нестационарность ряда АР(p), так как его характеристическое уравнение имеет вид

$$1 - \alpha_1 z - \alpha_2 z^2 - \dots - \alpha_p z^p = 0, \quad (10.108)$$

и если хоть один из корней z_j этого уравнения равен единице, то с одной стороны, это сигнализирует о нестационарности ряда, а с другой — подстановка единичного корня в (10.108) и дает нам тождество (10.107).

Соответственно, ряды (10.104') и (10.105') модифицируются с учетом (10.106), и процедура проверки нестационарности полученных таким образом рядов реализуется следующим образом. В каждом из них строится МНК-оценка $\hat{\beta}(p)$ параметра $\beta(p) = \alpha_1 + \alpha_2 + \dots + \alpha_p - 1$, а затем с помощью критической статистики

$$\gamma(N; p) = \frac{\hat{\beta}(p)}{s_{\hat{\beta}(p)}} \quad (10.103')$$

проверяется статистически незначимое отличие от нуля оценки $\hat{\beta}(p)$ с использованием тех же самых критических значений из таблицы 10.8.

У п р а ж н е н и е . Доказать справедливость представления процесса АР(2) в форме (10.106).

10.5.3 Модель авторегрессии-проинтегрированного скользящего среднего (АРПСС(p, k, q)-модель)

Эта модель предложена Дж. Боксом и Г. Дженкинсом (см. [Бокс, Дженкинс (1974), с. 102–132]) и поэтому в специальной литературе известна также как «модель Бокса–Дженкинса» (в англоязычном варианте — **AutoRegressive Integrated Moving Average Model**, или сокращенно **ARIMA-model**). Она предназначена для описания нестационарных временных рядов $x(t)$, $t = 1, 2, \dots, N$, обладающих следующим свойством:

существует такое целое положительное число k (минимальное из возможных), что ряд $x_k(t)$, $t = 1, 2, \dots, N - k$, получившийся из $x(t)$ после применения к нему k -кратной процедуры метода последовательных разностей (см. п. 10.3.3), может быть описан моделью АРСС(p, q).

Это означает, что АРПСС(p, k, q)-модель анализируемого процесса $x(t)$, $t = 1, 2, \dots, N$, может быть записана в виде

$$x_k(t) = \alpha_1 x_k(t-1) + \alpha_2 x_k(t-2) + \dots + \alpha_p x_k(t-p) + \delta(t) - \theta_1 \delta(t-1) - \dots - \theta_q \delta(t-q), \quad (10.109)$$

где

$$x_k(t) = \Delta^k x(t) = x(t) - C_k^1 x(t-1) + C_k^2 x(t-2) - \dots + (-1)^k x(t-k), \\ t = k+1, k+2, \dots, N.$$

С учетом представления АРСС(p, q)-модели в виде (10.79б) и связи между операторами Δ и F_- (см. (10.91)) модель АРПСС(p, k, q) может быть записана в форме

$$A_p(F_-, \alpha)(1 - F_-)^k x(t) = B_q(F_-, \theta)\delta(t), \quad (10.109')$$

или, что то же,

$$A_p(F_-, \alpha)\Delta^k x(t) = B_q(F_-, \theta)\delta(t), \quad (10.109'')$$

где полиномы $A_p(F_-, \alpha)$ и $B_q(F_-, \theta)$ определены соотношениями, соответственно, (10.89) и (10.89'), а под $\Delta^k x(t)$ понимается k -я последовательная разность анализируемого процесса $x(t)$. Обозначив произведение операторов $A_p(F_-, \alpha)$ и $(1 - F_-)^k$ с помощью $\varphi(F_-, \alpha)$, то есть полагая

$$\varphi(F_-, \alpha) = A_p(F_-, \alpha)(1 - F_-)^k, \quad (10.110)$$

представление АРПСС(p, k, q)-модели можно записать в виде

$$\varphi(F_-, \alpha)x(t) = B_q(F_-, \theta)\delta(t). \quad (10.109a)$$

Оператор $\varphi(F_-, \alpha)$ принято называть *обобщенным оператором авторегрессии*, в то время как $A_p(F_-, \alpha)$ — просто *оператор авторегрессии*, $B_q(F_-, \theta)$ — *оператор скользящего среднего*.

Из определения (10.110) следует, что $\varphi(F_-, \alpha)$ есть полином от F_- степени $p + k$ со свободным членом, равным единице, то есть

$$\begin{aligned}\varphi(F_-, \alpha) &= (1 - \alpha_1 F_- - \alpha_2 F_-^2 - \dots - \alpha_p F_-^p)(1 - F_-)^k = \\ &= 1 - \varphi_1 F_- - \varphi_2 F_-^2 - \dots - \varphi_{p+k} F_-^{p+k},\end{aligned}\quad (10.110')$$

где коэффициенты $\varphi_1, \varphi_2, \dots, \varphi_{p+k}$ очевидным образом выражаются через $\alpha_1, \alpha_2, \dots, \alpha_p$. Например, для модели АРПСС(1,1,1) имеем в соответствии с (10.89) и (10.110)

$$\varphi(F_-, \alpha) = (1 - \alpha F_-)(1 - F_-) = 1 - (1 + \alpha)F_- + \alpha F_-^2,$$

так что форма (10.109а) для модели АРПСС(1,1,1) будет иметь вид

$$[1 - (1 + \alpha)F_- + \alpha F_-^2]x(t) = (1 - \theta F_-)\delta(t),\quad (10.111)$$

или, что то же,

$$x(t) = (1 + \alpha)x(t-1) - \alpha x(t-2) + \delta(t) - \theta \delta(t-1).\quad (10.111')$$

Для многих целей, и в частности для вычисления прогнозов (см. п. 10.6), уравнения типа (10.111') являются наиболее удобной формой описания АРПСС-моделей.

Поясним теперь слово «проинтегрированного» в названии АРПСС-моделей. С этой целью введем в рассмотрение *бесконечный оператор суммирования* S , определенный как

$$Sx(t) = x(t) + x(t-1) + x(t-2) + \dots = \sum_{\tau=-\infty}^t x(\tau).$$

Непосредственно из определения операторов F_- , Δ и S следует:

$$S = \Delta^{-1} = (1 - F_-)^{-1} = 1 + F_- + F_-^2 + \dots \quad (10.112)$$

Соответственно определяются степени оператора S :

$$\begin{aligned}S^2 x(t) &= S(Sx(t)) = Sx(t) + Sx(t-1) + \dots = \sum_{l=-\infty}^t \sum_{\tau=-\infty}^l x(\tau), \\ S^3 x(t) &= S(S^2 x(t)) = \sum_{\nu=-\infty}^t \sum_{l=-\infty}^{\nu} \sum_{\tau=-\infty}^l x(\tau) \quad \text{и т.д.}\end{aligned}$$

Из (10.112) следует (и непосредственной проверкой можно в этом убедиться), что процесс $x(t)$ можно получить из процесса $x_k(t) = \Delta^k x(t)$ суммированием (интегрированием) последнего k раз, то есть

$$x(t) = S^k(\Delta^k x(t)).$$

Применяя эту операцию к обеим частям АРПСС-процесса, записанного в форме (10.109''), имеем

$$A_p(F_-, \alpha)x(t) = S^k(B_q(F_-, \theta)\delta(t)), \quad (10.1096)$$

то есть имеем именно *процесс авторегрессии (левая часть) — проинтегрированного скользящего среднего (правая часть)*. Правда, правильнее было бы сказать «*просуммированного*» вместо «*проинтегрированного*».

Заметим, что при описании несезонных временных рядов (сезонные модели будут рассмотрены в п. 10.5.4) редко встречаются с ситуацией, в которой порядки p, q или k были бы больше 2. Наиболее распространенные в прикладных эконометрических исследованиях АРПСС-модели имеют порядки

- 1) $p = 0, q = 1, k = 1$;
- 2) $p = 0, q = 2, k = 2$;
- 3) $p = 1, q = 1, k = 1$ (см. (10.111'));
- 4) $p = 1, q = 0, k = 1$;
- 5) $p = 2, q = 0, k = 1$.

Идентификация АРПСС-моделей. В первую очередь, следует подобрать *порядок k модели*. При этом руководствуются соображениями двух типов. Первый тип основан на эвристическом критерии, описанном в п. 10.3.3. Он состоит в отслеживании поведения величины $\hat{\sigma}^2(k)$ (см. (10.38)) в зависимости от k : в качестве верхней оценки для порядка k определяется то значение k_0 , начиная с которого тенденция к убыванию $\hat{\sigma}^2(k)$ гасится и само значение $\hat{\sigma}^2(k)$ относительно стабилизируется. Второй тип соображений, на которых основан подбор порядка k модели АРПСС(p, k, q), относится к анализу поведения автокорреляционных функций процессов $\Delta x(t), \Delta^2 x(t), \dots$. Последовательные преобразования анализируемого процесса $x(t)$ с помощью операторов Δ, Δ^2 и т.д. нацелены на устранение его нестационарности. Пока мы не «доберемся» до нужного порядка k (то есть при всех $l < k$) процессы $\Delta^l x(t)$ будут оставаться нестационарными, что, в частности, будет выражаться в *отсутствии быстрого спада в поведении их выборочной автокорреляционной функции*. Поэтому предполагается, что необходимая для получения стационарности степень k разности Δ достигнута, если автокорреляционная функция ряда $x_k(t) = \Delta^k x(t)$ быстро затухает. На практике k обычно равно 0, 1 или 2. Поэтому достаточно бывает вычислить по несколько первых значений автокорреляций исходного ряда

$x(t)$, его первых и вторых разностей (то есть рядов $x_1(t) = \Delta x(t)$ и $x_2(t) = \Delta^2 x(t)$).

После подбора порядка k мы практически анализируем уже не сам ряд $x(t)$, а его k -е разности, то есть ряд $x_k(t) = \Delta^k x(t)$. А идентификация этого ряда сводится к идентификации АРСС(p, q)-моделей. Процедуры же идентификации моделей АРСС(p, q) описаны в пп. 10.4.3.–10.4.4.

10.5.4 Модели рядов, содержащих сезонную компоненту

В разделе п. 10.1, посвященном генезису наблюдений, образующих временной ряд $x(t)$, мы подразделяли факторы, под воздействием которых формируются значения $x(t)$, на *долговременные, сезонные, циклические и случайные*. В примере 10.1 и на рис. 10.1 был описан типичный представитель временных рядов, в формировании значений которых существенную роль играют сезонные факторы. В дальнейшем под *временными рядами, содержащими сезонную компоненту*, мы будем понимать процессы, при формировании значений которых обязательно присутствовали сезонные и/или циклические факторы.

Напомним, что построение модели временного ряда — не самоцель, а лишь средство его анализа и *прогнозирования*. Один из распространенных подходов к прогнозированию состоит в следующем: решают задачу разложения ряда на долговременную, сезонную (объединяющую в себе собственно сезонную и циклическую компоненты) и случайную составляющие; затем долговременную составляющую стараются «подогнать» полиномом, сезонную — рядом Фурье, после чего прогноз производится экстраполяцией этих «подогнанных» значений в будущее. Однако такой подход может оказаться неэффективным, более того, приводить к серьезным ошибкам по двум причинам. Во-первых, *короткие участки в действительности стационарного ряда* (а в экономических приложениях мы редко имеем достаточно длинные временные ряды) могут выглядеть приблизительно как фрагменты полиномиальных или гармонических функций, что повлечет за собой их неправомерную аппроксимацию и представление в качестве долговременных или сезонных *неслучайных* составляющих (см. ниже пример 10.14 в п. 10.6.2). Во-вторых, даже в случаях, когда анализируемый ряд действительно включает в себя неслучайные полиномиальные или гармонические компоненты, их формальная аппроксимация может потребовать привлечения слишком большого числа параметров, то есть *получающаяся параметризация модели оказывается неэкономичной*. Так, например, потребуется много гармонических компонент, чтобы описать в динамике данные о сбыте товаров, на который влияют всевозможные праздники и сезонные распродажи.

Принципиально другой подход основан на модификации описанных выше конструкций АРСС-моделей с помощью так называемых «упро-

щающих операторов». Мы уже, по существу, пользовались этим приемом, когда для элиминирования нестационарности из анализируемого временного ряда $x(t)$ применяли к нему оператор $\Delta^k = (1 - F_-)^k$. Оценивая затем параметры в *упрощенной* таким образом модели стационарного временного ряда $x_k(t) = (1 - F_-)^k x(t)$ и возвращаясь с помощью обратного оператора $S^k = (1 - F_-)^{-k}$ к исходному ряду $x(t)$, мы получали в итоге модель именно для ряда $x(t)$. В данном случае в роли упрощающего оператора выступал оператор $1 - F_-$. В других задачах может оказаться полезным применение других типов упрощающих операторов. В частности, для временного исключения (в целях упрощения) из анализируемого ряда $x(t)$ сезонной составляющей, имеющей период T , в левую часть АРПСС(p, k, q)-модели, представленной уравнением (10.109а), мультипликативно вводится упрощающий оператор $\nabla_T = 1 - F_-^T$ в «подходящей» степени K , а в правую часть — множители (упрощающие операторы) вида $1 - \theta^* F_-^T$ (такие модели называются *мультипликативными*) либо в оператор-полином $B_q(F_-, \theta)$ добавляются члены вида $-\theta_T^* F_-^T$ и/или $-\theta_{T+1}^* F_-^{T+1}$.

Ниже описываются четыре относительно простые конкретные модели такого типа (1-я и 2-я — мультипликативные, 3-я и 4-я — не мультипликативные), которые используются при эконометрическом моделировании в достаточно широком спектре приложений.

Модель сезонных рядов №1. Уравнение модели имеет вид

$$\Delta^k \nabla_T^K x(t) = (1 - \theta F_-)(1 - \theta^* F_-^T) \delta(t),$$

или, что то же,

$$\Delta^k \nabla_T^K x(t) = \delta(t) - \theta \delta(t-1) - \theta^* \delta(t-T) + \theta \theta^* \delta(t-T-1), \\ T \geq 3.$$

При *идентификации* модели подбор параметров k, K, T производится в процессе сочетания статистического и качественного анализа с использованием приведенных выше рекомендаций. Оценка параметров основана на следующих соотношениях, связывающих автоковариации $\gamma(\tau)$ процесса $x_{k,K}^{(T)}(t) = \Delta^k \nabla_T^K x(t)$ с неизвестными параметрами $\sigma_0^2 = D\delta(t)$ и θ^* :

$$\begin{aligned} \gamma(0) &= \sigma_0^2(1 + \theta^2)(1 + \theta^{*2}); \\ \gamma(1) &= -\sigma_0^2\theta(1 + \theta^{*2}); \\ \gamma(T-1) &= \gamma(T+1) = \sigma_0^2\theta\theta^*; \\ \gamma(T) &= -\sigma_0^2\theta^*(1 + \theta^2). \end{aligned}$$

Все остальные ковариации равны нулю.

Модель сезонных рядов №2. Уравнение модели имеет вид:

$$(1 - a F_-^T) \Delta^k \nabla_T^K x(t) = (1 - \theta F_-)(1 - \theta^* F_-^T) \delta(t),$$

или, что то же,

$$\begin{aligned}\Delta^k \nabla_T^K x(t) - a \Delta^k \nabla_T^K x(t-T) &= \\ &= \delta(t) - \theta \delta(t-1) - \theta^* \delta(t-T) + \theta \theta^* \delta(t-T-1), \\ T &\geq 3, \quad |a| < 1.\end{aligned}$$

Идентификация модели включает в себя подбор параметров k , K и T (см. выше), а также оценивание остальных параметров методом моментов с использованием оценок $\hat{\gamma}(\tau)$ автоковариаций $\gamma(\tau)$ процесса $x_{k,K}^{(T)}(t) = \Delta^k \nabla_T^K x(t)$ и связей, существующих между этими ковариациями и неизвестными параметрами $\sigma_0^2 = \mathbf{D}\delta(t)$, a , θ и θ^* :

$$\begin{aligned}\gamma(0) &= \sigma_0^2(1 + \theta^2) \left(1 + \frac{(\theta^* - a)^2}{1 - a^2} \right); \\ \gamma(1) &= -\sigma_0^2 \theta \left(1 + \frac{(\theta^* - a)^2}{1 - a^2} \right); \\ \gamma(T-1) &= \gamma(T+1) = \sigma_0^2 \theta \left(\theta^* - a - \frac{a(\theta^* - a)^2}{1 - a^2} \right); \\ \gamma(T) &= -\sigma_0^2(1 + \theta^2) \left(\theta^* - a - \frac{a(\theta^* - a)^2}{1 - a^2} \right); \\ \gamma(\tau) &= a\gamma(\tau-T) \quad \text{при} \quad \tau \geq T+2;\end{aligned}$$

Модель сезонных рядов №3. Уравнение модели имеет вид:

$$\Delta^k \nabla_T^K x(t) = (1 - \theta F_- - \theta^* F_-^T) \delta(t),$$

или, что то же,

$$\begin{aligned}\Delta^k \nabla_T^K x(t) &= \delta(t) - \theta \delta(t-1) - \theta^* \delta(t-T), \\ T &\geq 3.\end{aligned}$$

Идентификация модели включает в себя подбор параметров k , K и T (см. выше), а также оценивание остальных параметров методом моментов с использованием оценок $\hat{\gamma}(\tau)$ ковариаций $\gamma(\tau)$ процесса $x_{k,K}^{(T)}(t) = \Delta^k \nabla_T^K x(t)$ и связей, существующих между этими ковариациями и неизвестными параметрами $\sigma_0^2 = \mathbf{D}\delta(t)$, θ и θ^* :

$$\begin{aligned}\gamma(0) &= \sigma_0^2(1 + \theta^2 + \theta^{*2}); \\ \gamma(1) &= -\sigma_0^2 \theta; \\ \gamma(T-1) &= \sigma_0^2 \theta \theta^*; \\ \gamma(T) &= -\sigma_0^2 \theta^*;\end{aligned}$$

все остальные ковариации равны нулю.

Модель сезонных рядов №4. Уравнение модели имеет вид:

$$(1 - aF_-^T)\Delta^k\nabla_T^K x(t) = (1 - \theta F_- - \theta^* F_-^T)\delta(t),$$

или, что то же,

$$\Delta^k\nabla_T^K x(t) - a\Delta^k\nabla_T^K x(t-T) = \delta(t) - \theta\delta(t-1) - \theta^*\delta(t-T),$$

$$T \geq 3, \quad |a| < 1.$$

Оценивание параметров a , $\sigma_0^2 = \mathbf{D}\delta(t)$, θ и θ^* производится методом моментов с использованием оценок $\hat{\gamma}(\tau)$ ковариаций $\gamma(\tau)$ процесса $x_{k,K}^{(T)}(t) = \Delta^k\nabla_T^K x(t)$ и связей, существующих между этими ковариациями и неизвестными параметрами:

$$\begin{aligned}\gamma(0) &= \sigma_0^2 \left(1 + \frac{\theta^2 + (\theta^* - a)^2}{1 - a^2} \right); \\ \gamma(1) &= -\sigma_0^2 \theta \left(1 - a \frac{(\theta^* - a)^2}{1 - a^2} \right); \\ \gamma(T-1) &= \sigma_0^2 \frac{\theta(\theta^* - a)^2}{1 - a^2}; \\ \gamma(T) &= \sigma_0^2 \frac{a\theta^2 - (\theta^* - a)(1 - a\theta^*)}{1 - a^2}; \\ \gamma(\tau) &= a\gamma(\tau - T) \quad \text{при } \tau \geq T+1.\end{aligned}$$

Схематично процедура построения сезонных моделей, основанных на АРССС-конструкциях, модифицированных с помощью упрощающих операторов $\nabla_T = 1 - F_-^T$, может быть описана следующим образом:

- 1) применяем к наблюдаемому ряду $x(t)$ операторы Δ и ∇_T для достижения стационарности;
- 2) по виду автокорреляционной функции преобразованного ряда $x_{k,K}^{(T)}(t)$ подбираем пробную модель в классе АРССС- или модифицированных (в правой части) АРССС-моделей;
- 3) по значениям соответствующих автоковариаций ряда $x_{k,K}^{(T)}(t)$ получаем (методом моментов) оценки параметров пробной модели;
- 4) диагностическая проверка полученной модели (анализ остатков в описании реального ряда $x(t)$ с помощью построенной модели) может либо подтвердить правильность модели, либо указать пути ее улучшения, что приводит к новой подгонке и повторению всей процедуры.

Более детальное описание этих процедур можно найти в книге [Бокс, Дженкинс (1974)].

10.6 Прогнозирование экономических показателей, основанное на использовании моделей временных рядов

Познакомившись в п. 10.3 с различными способами сглаживания временного ряда, а в пп. 10.4 и 10.5 — с разнообразными моделями, соответственно, стационарных и нестационарных временных рядов, рассмотрим теперь, как все эти методы и модели используются при прогнозировании экономических показателей. Причем наше внимание будет сконцентрировано на методах *автопрогноза*, в которых имеющийся в наличии ряд экстраполируется вперед, а другие ряды, которые, возможно, тоже несут определенную информацию о поведении нашего ряда, остаются без внимания.

Отметим прежде всего, что *не существует универсально предпочтительных методов прогнозирования на все случаи жизни*. Выбор метода прогнозирования и его эффективность зависят от многих условий, и в частности от:

- (а) требуемого горизонта l прогнозирования, то есть от того, на сколько временных тактов (l) вперед мы собираемся строить наш прогноз; при $l \leq 3$ прогноз обычно называется *краткосрочным*, при $3 < l \leq 6$ — *среднесрочным*, а при $l > 6$ — *долгосрочным* (приведенная здесь классификация, конечно, условна; она может зависеть от конкретного смысла одного такта времени — подразумевается ли под ним день, неделя, месяц, квартал, год и т. д.);
- (б) длины N анализируемого временного ряда (условно говоря, при $N \leq 50$ ряд считается *коротким*, а при $N > 50$ — *длинным*);
- (в) наличия или отсутствия в анализируемом ряду *сезонной составляющей* или каких-либо «нестандартностей» (скачкообразных изменений в поведении тренда, слишком большой величины дисперсии σ_0^2 случайных остатков и т. п.).

Поэтому выбор метода прогнозирования следует производить с учетом всех специфических особенностей как целей прогноза, так и анализируемого временного ряда.

Кстати, с точки зрения целевых установок мы ограничим наше рассмотрение *задачами кратко- и среднесрочного прогноза*, которые могут решаться (и весьма эффективно) с привлечением *только статистически-экстраполяционных методов*. Серьезное решение задач долгосрочного прогноза требует использования *комплексных подходов* и, в первую очередь, привлечения различных технологий сбора и анализа *экспертных оценок*.

10.6.1 Прогнозирование на базе АРПСС-моделей

Из результатов пп. 10.4–10.5 следует, что АРПСС-модели охватывают достаточно широкий спектр как стандартных, так и нестандартных временных рядов, а небольшие модификации этих моделей позволяют весьма точно описывать и временные ряды с сезонностью (см. п. 10.5.4). Поэтому начнем обсуждение проблемы прогнозирования временных рядов с методов, основанных на использовании АРПСС-моделей. Мы говорим об АРПСС-моделях, имея в виду, что сюда входят как частные случаи АР-, СС- и АРСС-модели. Кроме того, мы будем исходить из того, что уже осуществлен подбор подходящей АРПСС-модели для анализируемого временного ряда, включая идентификацию этой модели. Поэтому в дальнейших наших рассуждениях предполагается, что все параметры модели уже оценены, то есть *известны*. Для упрощения обозначений мы будем использовать при записи моделей в качестве ее параметров символы $p, q, k; \alpha_1, \dots, \alpha_p; \theta_1, \dots, \theta_q$, подразумевая при этом их оценки, соответственно, $\hat{p}, \hat{q}, \hat{k}; \hat{\alpha}_1, \dots, \hat{\alpha}_p; \hat{\theta}_1, \dots, \hat{\theta}_q$ (значениями которых мы располагаем).

Итак, займемся прогнозированием неизвестного значения $x(t+l)$, $l \geq 1$, полагая, что $x(t)$ — это последнее по времени наблюдение анализируемого временного ряда, которое имеется в нашем распоряжении (будем говорить, что t — это *текущий момент времени*). Другими словами, мы делаем прогноз в момент времени t на l тактов времени вперед (или с упреждением l). Обозначим такой прогноз $\hat{x}(t; l)$.

Заметим, что хотя $\hat{x}(t; l)$ и $\hat{x}(t-1; l+1)$ обозначают прогноз одного и того же неизвестного значения $x(t+l)$, но вычисляются они по-разному, т.к. являются решениями разных задач. Поэтому мы не используем, казалось бы, естественное обозначение $\hat{x}(t+l)$ для прогноза величины $x(t+l)$.

Известно (см. (10.109)–(10.109')), что анализируемый в рамках АРПСС(p, k, q)-модели ряд $x(\tau)$ представим (при любом $\tau \geq k+1$) уравнением вида

$$(1 - \alpha_1 F_- - \dots - \alpha_p F_-^p)(x(\tau) - C_k^1 x(\tau-1) + \dots + (-1)^k x(\tau-k)) = \delta(\tau) - \theta_1 \delta(\tau-1) - \dots - \theta_q \delta(\tau-q), \quad (10.109в)$$

где F_- , как и прежде, оператор сдвига функции времени на один временной такт назад (соответственно, оператор F_-^j сдвигает функцию, к которой он применяется, на j тактов времени назад).

Перепишем соотношение (10.109в) последовательно для $\tau = t-q, t-q+1, \dots, t+1, t+2, \dots, t+l$, уединив при этом $x(\tau)$ в левой его части:

$$\begin{aligned}
 x(t-q) = & \left(\sum_{j=1}^p \alpha_j F_-^j \right) x(t-q) + \left(1 - \sum_{j=1}^p \alpha_j F_-^j \right) \times \\
 & \times (C_k^1 x(t-q-1) - \dots + (-1)^k x(t-q-k)) + \\
 & + \delta(t-q) - \theta_1 \delta(t-q-1) - \dots - \theta_q \delta(t-2q), \quad (10.113)
 \end{aligned}$$

$$\begin{aligned}
 x(t-q+1) = & \left(\sum_{j=1}^p \alpha_j F_-^j \right) x(t-q+1) + \left(1 - \sum_{j=1}^p \alpha_j F_-^j \right) \times \\
 & \times (C_k^1 x(t-q) - \dots + (-1)^k x(t-q+1-k)) + \delta(t-q+1) - \\
 & - \theta_1 \delta(t-q) - \dots - \theta_q \delta(t-2q+1), \quad (10.113')
 \end{aligned}$$

$$\begin{aligned}
 x(t+1) = & \left(\sum_{j=1}^p \alpha_j F_-^j \right) x(t+1) - \left(1 - \sum_{j=1}^p \alpha_j F_-^j \right) \times \\
 & \times (C_k^1 x(t) - \dots + (-1)^k x(t+1-k)) + \\
 & + \delta(t+1) - \theta_1 \delta(t) - \dots - \theta_q \delta(t+1-q), \quad (10.113'')
 \end{aligned}$$

$$\begin{aligned}
 x(t+l) = & \left(\sum_{j=1}^p \alpha_j F_-^j \right) x(t+l) - \left(1 - \sum_{j=1}^p \alpha_j F_-^j \right) \times \\
 & \times (C_k^1 x(t+l-1) - \dots + (-1)^k x(t+l-k)) + \delta(t+l) - \\
 & - \theta_1 \delta(t+l-1) - \dots - \theta_q \delta(t+l-q). \quad (10.113''')
 \end{aligned}$$

Обращаем внимание на тот факт, что правые части этих соотношений представляют собой линейные комбинации $p+k$ предшествующих (по отношению к левой части) значений анализируемого процесса $x(\tau)$, дополненные линейными комбинациями *текущего и q предшествующих* значений случайных остатков $\delta(\tau)$. Причем коэффициенты, с помощью которых эти линейные комбинации подсчитываются, *известны*, так как выражаются в терминах уже оцененных нами параметров модели $\alpha_1, \alpha_2, \dots, \alpha_p$ и $\theta_1, \theta_2, \dots, \theta_q$.

Этот факт и дает нам возможность использовать соотношения (10.113)–(10.113''') для построения прогнозных значений анализируемого временного ряда на l тактов времени вперед. Теоретическую базу такого подхода к прогнозированию обеспечивает известный результат, в соответствии с которым *наилучшим (в смысле среднеквадратической ошибки) линейным прогнозом в момент времени t с упреждением l является условное математическое ожидание случайной величины $x(t+l)$, вычисленное при условии, что все значения $x(\tau)$ до момента времени t (включительно) известны* (этот факт является частным случаем более

общих результатов теории прогнозирования, развитой Волдом, Колмогоровым, Винером¹⁹).

Условное математическое ожидание $E(x(t+l) | x(1), x(2), \dots, x(t))$ получается применением операции условного усреднения к обеим частям (10.113''') с учетом следующих соотношений:

$$E(x(t-j) | x(1), \dots, x(t)) = x(t-j) \quad \text{при всех } j = 0, 1, 2, \dots, t-1; \quad (10.114a)$$

$$E(x(t+j) | x(1), \dots, x(t)) = \hat{x}(t; j) \quad \text{при всех } j = 1, 2, \dots; \quad (10.114б)$$

$$E(\delta(t+j) | x(1), \dots, x(t)) = 0 \quad \text{при всех } j = 1, 2, \dots; \quad (10.114в)$$

$$E(\delta(t-j) | x(1), \dots, x(t)) = x(t-j) - \hat{x}(t-j-1; 1) \quad \text{при всех } j = 0, 1, 2, \dots; \quad (10.114г)$$

Таким образом, определяется следующая процедура построения прогноза $\hat{x}(t; l)$ для значения $x(t+l)$ анализируемого временного ряда по известной до момента t (включительно) его траектории $x(1), x(2), \dots, x(t)$:

1) на 1-м этапе используются формулы (10.113), (10.113'), ... для вычисления *ретроспективных* прогнозов $\hat{x}(t-q-1; 1)$, $\hat{x}(t-q; 1)$ и т. д. до $\hat{x}(t-1; 1)$ по предыдущим значениям временного ряда; при этом при вычислении *начальных* прогнозных значений $\hat{x}(t-q+m-1)$ для $x(t-q+m)$ ($m = 0, 1, \dots$) по формулам (10.113)–(10.113')–... вместо *условных* средних $E(\delta(t-q+m-j) | x(1), \dots, x(t-q+m))$ (которые в общем случае следовало бы вычислять по формулам (10.114г), однако на начальном этапе значения $\hat{x}(t-j-1; 1)$ нам не известны) подставляются их *безусловные* значения, равные, как известно, нулю; ретроспективный прогноз нужен для того, чтобы при построении обычного прогноза (начиная с $\hat{x}(t; 1)$ для $x(t+1)$ по формуле (10.113'')) мы могли бы подсчитать в правой части (10.113'') значения $E(\delta(t-j) | x(1), \dots, x(t))$ ($j = 0, 1, \dots, q-1$) по формулам (10.114г);

2) на 2-ом этапе используются формулы (10.113''), ..., (10.113''') и правила (10.114а)–(10.114г) подсчета условных математических ожиданий участвующих в правых частях (10.113'')–(10.113''') выражений для вычисления прогнозных значений $\hat{x}(t; 1)$, $\hat{x}(t; 2)$, ..., $\hat{x}(t; l)$.

З а м е ч а н и е. Возможен и *упрощенный вариант*, при котором сразу приступают к реализации 2-го этапа, начиная с формул (10.113''); при этом те значения $E(\delta(t-j) | x(1), \dots, x(t))$, которые должны были бы, но еще не могут, быть вычислены по формуле (10.114г), аппроксимируются нулями.

¹⁹См.: 1) Wold H. O. A Study in the Analysis of Stationary Time Series. Almqvist and Wicksell, Uppsala, 1932; 2) Kolmogoroff A. Sur l'interpolation et l'extrapolation des suites stationnaires. Compt. Rend., 208 (1939), p. 2043; 3) Wiener N. Extrapolation, Interpolation and Smoothing of Stationary Time Series. John Wiley, New York, 1949.

Описанная процедура выглядит достаточно сложной. Мы сейчас покажем на примерах, что при *реалистичных* значениях параметров p, q и k (как правило, не превышающих двух) эта процедура в действительности оказывается весьма простой. Тем не менее, в качестве альтернативного способа построения прогноза, *казалось бы*, могла бы рассматриваться следующая схема.

Из (10.109в) непосредственно следует, что описываемый АРПСС(p, k, q)-моделью временной ряд $x(\tau)$ может быть представлен в форме модели авторегрессии вида

$$x(\tau) = \sum_{j=1}^{p+k} \pi_j x(\tau - j) + \varepsilon(\tau), \quad (10.115)$$

где коэффициенты π_j очевидным образом выражаются в виде линейных комбинаций параметров α и θ , а случайные остатки $\varepsilon(\tau)$ — в виде линейных комбинаций элементов белого шума δ . Почему бы, не проводя предварительную процедуру статистического оценивания параметров α и θ , не оценить параметры π_j с помощью обычного или обобщенного метода наименьших квадратов по наблюдениям $(x(\tau), x(\tau+1), \dots, x(\tau+p+k-1); x(\tau+p+k))$, $\tau = 1, 2, \dots, t-p-k$? А получив оценки $\hat{\pi}_j$ параметров π_j и подставив их в (10.115) при $\tau = t+1$, можно было бы вычислить прогнозное значение $\hat{x}(t+1)$ для $x(t+1)$, по нему — прогнозное значение $\hat{x}(t+2)$ для $x(t+2)$ и таким рекуррентным способом добраться до $\hat{x}(t+l)$.

К сожалению, эта схема в общем случае не позволяет получить **состоятельные** оценки параметров π_j по той причине, что регрессионные остатки $\varepsilon(\tau)$ в модели (10.115) оказываются существенно коррелированными с объясняющими переменными $x(\tau - j)$ (см. п. 7.2).

Рассмотрим ряд примеров.

Пример 10.10. Прогнозирование ряда, описываемого моделью $АРСС(1,1,0)$. Анализ и идентификация имеющегося временного ряда $x(\tau)$ привели нас к $АРСС(1,1,0)$ -модели вида

$$(1 - 0,8F_-)(1 - F_-)x(\tau) = \delta(\tau),$$

или, что то же,

$$(1 - 1,8F_- + 0,8F_-^2) x(\tau) = \delta(\tau).$$

Воспользуемся упрощенным вариантом описанной выше процедуры прогнозирования. Уравнения (10.113'')–(10.113''') в данном случае имеют вид

$$\begin{aligned} x(t+1) &= 1,8x(t) - 0,8x(t-1) + \delta(t+1), \\ &\dots\dots\dots(10.116) \\ x(t+l) &= 1,8x(t+l-1) - 0,8x(t+l-2) + \delta(t+l). \end{aligned}$$

Мы знаем, что прогнозные значения $\hat{x}(t; m)$ для $x(t+m)$ ($m = 1, 2, \dots, l$) получаются вычислением условных математических ожиданий $E(x(t+m) | x(1), x(2), \dots, x(t))$. Для этого в правых частях соотношений (10.116) фиксируются значения $x(\tau)$, известные к моменту времени t (то есть $x(1), \dots, x(t)$), а будущие значения $x(t+m)$ заменяются их прогнозами $\hat{x}(t; m)$; вместо будущих значений δ и тех прошлых значений, которые нельзя подсчитать по формулам (10.114г) из-за нехватки прогнозных значений $\hat{x}(t-j-1; 1)$, вставляются нули, а остальные прошлые значения δ подсчитываются по формулам (10.114г). В результате от (10.116) приходим к следующим выражениям для прогнозов:

$$\begin{aligned}\hat{x}(t; 1) &= 1,8x(t) - 0,8x(t-1); \\ \hat{x}(t; 2) &= 1,8\hat{x}(t; 1) - 0,8x(t); \\ \hat{x}(t; l) &= 1,8\hat{x}(t; l-1) - 0,8\hat{x}(t; l-2), \quad l \geq 3.\end{aligned}$$

Пример 10.11. *Прогнозирование ряда, описываемого моделью АРПСС(0,1,1)²⁰.* Анализ ежедневных данных о биржевых ценах акций IBM в период с 17 мая по 2 ноября 1962 г. (общая длина ряда $N = 369$) показал, что исследуемый временной ряд $x(\tau)$ хорошо может быть описан АРПСС(0,1,1)-моделью с параметром $\theta = 0,1$. Это значит, что он может быть представлен в форме $\Delta x(\tau) = \delta(\tau) - \theta\delta(\tau-1)$, где $\Delta x(\tau) = x(\tau) - x(\tau-1)$. Поэтому уравнения (10.113'')-(10.113''') в данном случае имеют вид

$$\begin{aligned}x(t+1) &= x(t) + \delta(t+1) - 0,1\delta(t), \\ &\dots\dots\dots \\ x(t+l) &= x(t+l-1) + \delta(t+l) - 0,1\delta(t).\end{aligned}$$

Отсюда в соответствии с описанной выше процедурой получаем

$$\begin{aligned}\hat{x}(t; 1) &= x(t) - 0,1\delta(t), \\ \hat{x}(t; l) &= \hat{x}(t; l-1) \quad \text{при } l \geq 2.\end{aligned}$$

Мы видим, что для всех упреждений (в момент t) прогнозы будут следовать прямой линии, параллельной оси времени. Учитывая то, что $\delta(t) = x(t) - \hat{x}(t-1; 1)$, прогноз $\hat{x}(t; l)$ можно записать также в виде

$$\hat{x}(t; l) = (1 - \theta)x(t) + \theta\hat{x}(t-1; l),$$

что в нашем конкретном случае дает

$$\hat{x}(t; l) = 0,9x(t) + 0,1\hat{x}(t-1; l).$$

То есть прогноз в момент t с упреждением l есть взвешенное среднее текущего значения $x(t)$ и аналогичного прогноза в предыдущий момент

²⁰Пример заимствован из [Бокс, Дженкинс (1974), с. 188–190.]

времени. Причем если θ близко к нулю (в нашем случае $\theta = 0,1$), то прошлый прогноз $\hat{x}(t-1; l)$, представляющий собой взвешенное среднее данных «о прошлом» (то есть — до момента $t-1$), практически игнорируется и прогноз «на все будущие времена» определяется текущим значением $x(t)$. Мы еще вернемся к этому примеру в п. 10.6.2 в связи с сопоставлением описываемого метода прогнозирования с методом экспоненциально взвешенного скользящего среднего Брауна.

Пример 10.12. *Прогнозирование ряда, описываемого моделью АРПСС(0,2,2).* Анализируемый временной ряд $x(\tau)$ представим в форме $\Delta^2 x(\tau) = \delta(\tau) - \theta_1 \delta(\tau-1) - \theta_2 \delta(\tau-2)$, где $\Delta^2 x(\tau) = x(\tau) - 2x(\tau-1) + x(\tau-2)$. Поэтому уравнение (10.113''') в данном конкретном случае имеет вид

$$x(t+l) = 2x(t+l-1) - x(t+l-2) + \delta(t+l) - \theta_1 \delta(t+l-1) - \theta_2 \delta(t+l-2).$$

Беря условные математические ожидания (при известном до момента t прошлом) от обеих частей этого соотношения, получаем для $l=1$, $l=2$ и в общем случае

$$\begin{aligned}\hat{x}(t; 1) &= 2x(t) - x(t-1) - \theta_1 \delta(t) - \theta_2 \delta(t-1), \\ \hat{x}(t; 2) &= 2\hat{x}(t; 1) - x(t) - \theta_2 \delta(t), \\ \hat{x}(t; l) &= 2\hat{x}(t; l-1) - \hat{x}(t; l-2) \quad \text{при } l \geq 3.\end{aligned}$$

Напомним, что в этих выражениях $\delta(t) = x(t) - \hat{x}(t-1; 1)$, $\delta(t-1) = x(t-1) - \hat{x}(t-2; 1)$, и процесс прогнозирования можно начинать подстановкой вместо неизвестных значений δ их безусловных математических ожиданий, равных нулю.

Пример 10.13. *Прогнозирование ряда, описываемого моделью АРПСС(1,1,1).* Подчиняющийся закономерностям этой модели временной ряд представим, как известно, в форме

$$(1 - \alpha F_-) \Delta x(\tau) = \delta(\tau) - \theta \delta(\tau-1),$$

где $\Delta x(\tau) = x(\tau) - x(\tau-1)$. Применяя к $\Delta x(\tau)$ оператор $(1 - \alpha F_-)$ и уединяя $x(\tau)$ в левой части, можно выписать уравнение (10.113'''):

$$x(t+l) = (1 + \alpha)x(t+l-1) - \alpha x(t+l-2) + \delta(t+l) - \theta \delta(t+l-1).$$

Беря условные математические ожидания от обеих частей этого соотношения (при условии, что все прошлое этого процесса до момента t известно), получаем для $l=1$ и в общем случае

$$\begin{aligned}\hat{x}(t; 1) &= (1 + \alpha)x(t) - \alpha x(t-1) - \theta \delta(t), \\ \hat{x}(t; l) &= (1 + \alpha)\hat{x}(t; l-1) - \alpha \hat{x}(t; l-2) \quad \text{при } l \geq 2.\end{aligned} \tag{10.117}$$

Значение $\delta(t)$ можно вычислить по формуле $\delta(t) = x(t) - \hat{x}(t-1; 1)$, построив предварительно ретроспективный прогноз $\hat{x}(t-1; 1)$ так, как это

было описано выше. В упрощенном варианте значение $\delta(t)$ полагается равным нулю.

Можно показать, что объединенной формой соотношений (10.117) является эквивалентное им представление прогноза $\hat{x}(t; l)$ в виде

$$\hat{x}(t; l) = \left[1 - \frac{\theta - \alpha}{1 - \alpha} (1 - \alpha^l) \right] x(t) + \frac{\theta - \alpha}{1 - \alpha} (1 - \alpha^l) \bar{x}_\theta(t - 1), \quad (10.117')$$

где $\bar{x}_\theta(t - 1)$ — экспоненциально взвешенное скользящее среднее (с параметром θ) — см. выше п. 10.3.2 — анализируемого «длинного» ряда, просуммированного от бесконечно удаленного прошлого до момента времени $t - 1$, то есть

$$\bar{x}_\theta(t - 1) = (1 - \theta) \sum_{j=1}^{\infty} \theta^{j-1} x(t - j). \quad (10.118)$$

Из (10.117') следует, что прогноз $\hat{x}(t; l)$ получается в результате линейной интерполяции величин $x(t)$ и $\bar{x}_\theta(t - 1)$ с весами $1 - \lambda(l)$ и $\lambda(l)$, где $\lambda(l) = (\theta - \alpha)(1 - \alpha^l)/(1 - \alpha)$, то есть $\hat{x}(t; l) = (1 - \lambda(l))x(t) + \lambda(l)\bar{x}_\theta(t - 1)$ (напомним, что $|\alpha| < 1$ и $|\theta| < 1$, так как это обеспечивает условия, соответственно, стационарности и обратимости анализируемого ряда, см. п. 10.4.3).

Мы не касались здесь важных вопросов *оценки точности получающихся прогнозов*. Теоретические аспекты этой проблемы рассмотрены, например, в [Бокс, Дженкинс (1974)].

10.6.2 Адаптивные методы прогнозирования

Считается, что характерной чертой адаптивных методов прогнозирования является их способность непрерывно учитывать эволюцию динамических характеристик изучаемых процессов, «подстраиваться» под эту эволюцию, придавая, в частности, тем больший вес и тем более высокую информационную ценность имеющимся наблюдениям, чем ближе они к текущему моменту прогнозирования. Однако деление методов и моделей на «адаптивные» и «неадаптивные» достаточно условно. В известном смысле каждый из рассмотренных в данном учебнике методов прогнозирования адаптивный, так как практически все они учитывают вновь поступающие наблюдения, в том числе сделанные с момента последнего прогноза. Общее значение термина заключается, по-видимому, в том, что «адаптивное» прогнозирование позволяет обновлять прогнозы с минимальной задержкой и с помощью относительно несложных математических процедур. Но, как мы увидим, это вовсе не значит, что в любой ситуации адаптивные методы оказываются эффективнее тех, которые традиционно не относятся к таковым.

Методы экспоненциального сглаживания (Брауна²¹). Простейший вариант этого метода был описан в п. 10.3.2 в связи с задачей выделения неслучайной составляющей из анализируемого временного ряда. Постановка задачи прогнозирования с использованием простейшего варианта метода экспоненциального сглаживания формулируется следующим образом.

Пусть анализируемый временной ряд $x(\tau)$ ($\tau = 1, 2, \dots, t$) представлен в виде

$$x(\tau) = a_0 + \varepsilon(\tau), \quad (10.119)$$

где a_0 — неизвестный параметр, не зависящий от времени, а $\varepsilon(\tau)$ — случайный остаток со средним значением, равным нулю, и конечной дисперсией. Как мы знаем (см. п. 10.3.2), экспоненциально взвешенная скользящая средняя (ЭВСС) ряда $x(\tau)$ в точке t $\bar{x}_\lambda(t)$ с *параметром сглаживания* (или — *параметром адаптации*) λ ($0 < \lambda < 1$) определяется формулой

$$\bar{x}_\lambda(t) = \frac{1 - \lambda}{1 - \lambda^t} \sum_{j=0}^{t-1} \lambda^j x(t - j), \quad (10.120)$$

которая дает решение задачи на минимум экспоненциально взвешенного критерия метода наименьших квадратов, а именно:

$$\bar{x}_\lambda(t) = \arg \min_a \sum_{j=0}^{t-1} \lambda^j [x(t - j) - a]^2.$$

Коэффициент сглаживания λ можно интерпретировать и как *коэффициент дисконтирования*, характеризующий меру обесценивания наблюдения за единицу времени.

Для рядов с «бесконечным прошлым» формула (10.120) сводится к виду

$$\bar{x}_\lambda(t) = (1 - \lambda) \sum_{j=0}^{\infty} \lambda^j x(t - j). \quad (10.120')$$

В соответствии с простейшим вариантом метода экспоненциального сглаживания прогноз $\hat{x}(t; 1)$ для неизвестного значения $x(t+1)$ по известной до момента времени t траектории ряда $x(\tau)$ строится по формуле

$$\hat{x}(t; 1) = \bar{x}_\lambda(t), \quad (10.121)$$

где значение $\bar{x}_\lambda(t)$ определено формулой (10.120) (для короткого временного ряда) или формулой (10.120') (для длинного временного ряда).

²¹См.: Brown R.G. Smoothing, Forecasting and Prediction of Discrete Time-Series. Prentice-Hall, New Jersey, 1962.

Формула (10.121) удобна, в частности, тем, что при появлении следующего $(t+1)$ -го наблюдения $x(t+1)$ пересчет прогнозирующей функции $\hat{x}(t+1; 1) = \bar{x}_\lambda(t+1)$ производится с помощью простого соотношения

$$\bar{x}_\lambda(t+1) = \lambda \bar{x}_\lambda(t) + (1-\lambda)x(t+1). \quad (10.122)$$

Для вывода этого соотношения выпишем (10.120') для моментов времени $t+1$ и t :

$$\begin{aligned} \bar{x}_\lambda(t+1) &= (1-\lambda)[x(t+1) + \lambda x(t) + \lambda^2 x(t-1) + \dots], \\ \bar{x}_\lambda(t) &= (1-\lambda)[x(t) + \lambda x(t-1) + \lambda^2 x(t-2) + \dots]. \end{aligned}$$

Вычтя из первого соотношения второе, умноженное на λ , получим $\bar{x}_\lambda(t+1) - \lambda \bar{x}_\lambda(t) = (1-\lambda)x(t+1)$, что и доказывает справедливость (10.122). Для *конечных* временных рядов соотношение (10.122) имеет *приближенный* смысл, поскольку в аналогичных выкладках мы должны будем пренебречь разницей в знаменателях правых частей (10.120) (легко видеть, что приближение это высокой точности, так как разность

$$\frac{1}{1-\lambda^{t+1}} - \frac{1}{1-\lambda^t}$$

достаточно быстро стремится к нулю при $t \rightarrow \infty$).

Метод экспоненциального сглаживания можно обобщить на случаи полиномиальной неслучайной составляющей анализируемого временного ряда, то есть на ситуации, когда вместо (10.119) постулируется

$$x(t+\tau) = a_0 + a_1\tau + \dots + a_k\tau^k + \varepsilon(\tau), \quad (10.119')$$

где $k \geq 1$. В соотношении (10.119') начальная точка отсчета времени сдвинута в текущий момент времени t , что облегчает дальнейшие выкладки и вычисления. Соответственно, в схеме простейшего варианта метода прогноз $\hat{x}(t; 1)$ значения $x(t+1)$ будет определяться соотношениями (10.119') при $\tau = 1$ и (10.121):

$$\hat{x}(t; 1) = \hat{x}(t+1) = \hat{a}_0^{(k)}(t; \lambda) + \hat{a}_1^{(k)}(t; \lambda) + \dots + \hat{a}_k^{(k)}(t; \lambda), \quad (10.122')$$

где оценки $\hat{a}_0(t; \lambda)$, $\hat{a}_1(t; \lambda)$, ..., $\hat{a}_k(t; \lambda)$ получаются как решение экспоненциально взвешенной оптимизационной задачи метода наименьших квадратов, то есть

$$\sum_{j=0}^{\infty} \lambda^j [x(t-j) - a_0 - a_1j - \dots - a_kj^k]^2 \rightarrow \min_{a_0, a_1, \dots, a_k}. \quad (10.123)$$

Решение задачи (10.123) не представляет принципиальных трудностей. Мы выпишем здесь решения задачи (10.123) для случаев $k = 1$ и $k = 2$ и для рядов с «бесконечным прошлым».

1) *Случай линейного тренда, то есть* $\mathbf{E}x(t + \tau) = a_0 + a_1\tau$.

$$\begin{aligned}\hat{a}_0^{(1)}(t; \lambda) &= \bar{x}_\lambda(t) = (1 - \lambda) \sum_{j=0}^{\infty} \lambda^j x(t - j), \\ \hat{a}_1^{(1)}(t; \lambda) &= (1 - \lambda)\bar{x}_\lambda(t) + (1 - \lambda)^2 \sum_{j=0}^{\infty} (j + 1)\lambda^j x(t - j).\end{aligned}\quad (10.124)$$

Так что прогноз $\hat{x}(t; 1)$ будущего значения $x(t + 1)$ при линейном тренде ряда будет вычисляться в соответствии с (10.122')

$$\hat{x}(t; 1) = \hat{a}_0^{(1)}(t; \lambda) + \hat{a}_1^{(1)}(t; \lambda),$$

где коэффициенты $\hat{a}_0^{(1)}(t; \lambda)$ и $\hat{a}_1^{(1)}(t; \lambda)$ определяются по формулам (10.124).

Пересчет коэффициентов $\hat{a}_0^{(1)}$ и $\hat{a}_1^{(1)}$ при появлении следующего $(t + 1)$ -го члена ряда производится по формулам

$$\begin{aligned}\hat{a}_0^{(1)}(t + 1; \lambda) &= (1 - \lambda^2)x(t + 1) + \lambda^2\hat{a}_0^{(1)}(t) + \lambda^2\hat{a}_1^{(1)}(t), \\ \hat{a}_1^{(1)}(t + 1; \lambda) &= (1 - \lambda)^2x(t + 1) - (1 - \lambda)^2\hat{a}_1^{(1)}(t) - (1 - \lambda)^2\hat{a}_0^{(1)}(t) + \hat{a}_1^{(1)}(t).\end{aligned}$$

2) *Случай квадратичного тренда, то есть* $\mathbf{E}x(t + \tau) = a_0 + a_1\tau + a_2\tau^2$.

Мы выпишем для этого случая лишь так называемые «формулы обновления», позволяющие пересчитывать оценки $\hat{a}_0^{(2)}(s; \lambda)$, $\hat{a}_1^{(2)}(s; \lambda)$ и $\hat{a}_2^{(2)}(s; \lambda)$, полученные как решения оптимизационной задачи (10.123) при $t = s$, когда появляется очередное $(s + 1)$ -е наблюдение временного ряда $x(s + 1)$:

$$\begin{aligned}\hat{a}_0^{(2)}(s + 1; \lambda) &= \hat{a}_0^{(2)}(s; \lambda) + \hat{a}_1^{(2)}(s; \lambda) + \hat{a}_2^{(2)}(s; \lambda) + (1 - \lambda^3)\hat{\varepsilon}(s + 1), \\ \hat{a}_1^{(2)}(s + 1; \lambda) &= \hat{a}_1^{(2)}(s; \lambda) + 2\hat{a}_2^{(2)}(s; \lambda) + \frac{3}{2}(1 - \lambda)(1 - \lambda^2)\hat{\varepsilon}(s + 1), \\ \hat{a}_2^{(2)}(s + 1; \lambda) &= \hat{a}_2^{(2)}(s; \lambda) + \frac{1}{2}(1 - \lambda)^3\hat{\varepsilon}(s + 1),\end{aligned}$$

где $\hat{\varepsilon}(s + 1) = x(s + 1) - \hat{a}_0^{(2)}(s; \lambda) - \hat{a}_1^{(2)}(s; \lambda) - \hat{a}_2^{(2)}(s; \lambda)$ — ошибка прогноза на последнем шаге. Решив для сравнительно небольшого $t = s$ оптимизационную задачу (10.123), затем последовательно пересчитывают по этим формулам значения прогнозирующих коэффициентов $\hat{a}_j^{(2)}(s + m; \lambda)$ ($j = 0, 1, 2; m = 1, 2, \dots$), пока не «доберутся» до текущего момента времени t .

Пример 10.14. Мы возвращаемся к примеру 10.11 (см. п. 10.6.1), чтобы сравнить эффективность применения квадратичного варианта метода Брауна и метода, основанного АРПСС(0,1,1)-модели, к решению одной и той же задачи — прогнозу биржевых цен на акции IBM на три

дня вперед. Описание исходных данных задачи и результатов построения АРПСС-модели было дано выше. На рис. 10.14 наглядно представлено сравнение двух методов, а в таблице 10.9 приведены результаты расчетов среднеквадратических ошибок прогнозов на базе ряда цен акций за период с 11 июля по 10 февраля 1961 г. (150 наблюдений). Для этого участка ряда оценка параметра θ в модели АРПСС (0,1,1) оказалась равной $\hat{\theta} = 0,1$, а это значит, что прогнозы по АРПСС-модели на 3 дня вперед, грубо говоря, эквивалентны использованию сегодняшней цены (так как $\hat{x}(t; l) = 0,9x(t) + 0,1\hat{x}(t-1; l)$), что тем не менее оказывается значительно точнее, чем прогнозы, полученные с помощью существенно более сложной методики Брауна.

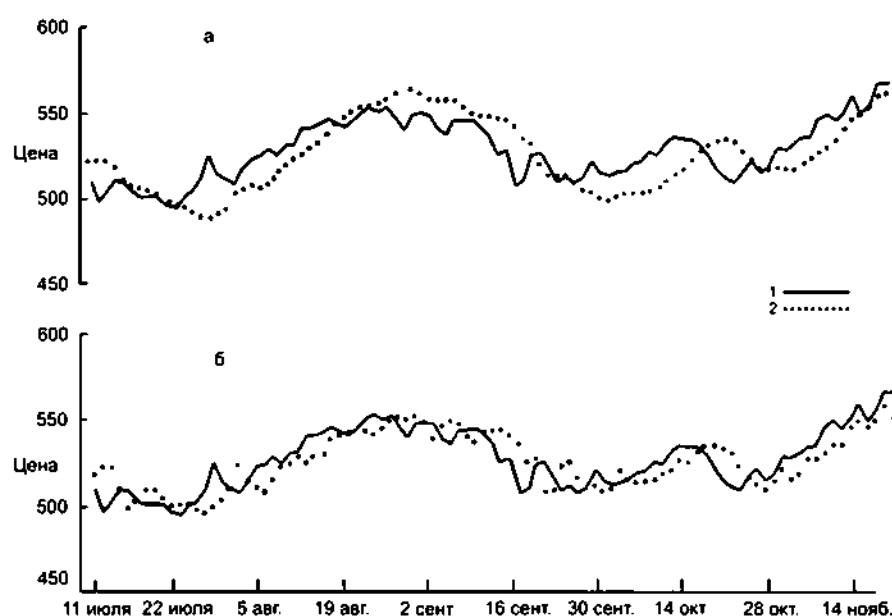


Рис. 10.14. Ряд биржевых цен акций IBM; сравнение с прогнозом на три шага вперед, основанным на лучшей модели АРПСС (0,1,1), и квадратичным прогнозом Брауна для периода, начинающегося 11 июля 1960 г. а — квадратичный прогноз Брауна, б — прогноз с минимальной среднеквадратичной ошибкой.
1 — данные; 2 — прогнозы на три шага вперед

Таблица 10.9. Сравнение среднеквадратичных ошибок прогнозов, полученных при различных упреждениях, по наилучшему из процессов АРПСС(0,1,1) и по методике Брауна

Упреждение l	1	2	3	4	5	6	7	8	9	10
СКО (Браун)	102	158	218	256	363	452	554	669	799	944
СКО АРПСС ($\theta = 0,1$)	42	91	136	180	282	266	317	371	427	483

Попробуем понять причины неудач применения квадратичного метода экспоненциального сглаживания в данном случае.

1) Выбор вида прогнозирующей функции, основанный на подборе операторов авторегрессии $(1 - \sum \alpha_j F_-^j)$ и конечных разностей $\Delta^k = 1 - F_-^k$, нам представляется более гибким и обоснованным, чем формальная аппроксимация полиномом траектории анализируемого ряда. Посмотрим, например, на траекторию биржевых цен акций IBM на рис. 10.14. Видно, что квадратичная функция может быть использована для аппроксимации лишь *коротких участков* ряда, причем эта аппроксимация носит чисто формальный характер и не имеет отношения к описанию генезиса анализируемых наблюдений.

2) Узким местом всех адаптивных методов, и методов экспоненциального сглаживания в частности, является *подбор подходящего к данной конкретной задаче параметра сглаживания (адаптации) λ* .

3) Наконец, в моделях экспоненциального сглаживания вся специфика анализируемого ряда должна быть отражена в *единственном параметре λ* . Это, конечно, сильно ограничивает класс допустимых в рамках этого метода моделей.

Кратко описанные ниже методы также используют «идеологию» экспоненциального сглаживания, но развивают методы Брауна в различных направлениях.

Метод Хольта. По-видимому, Хольт первым попытался ослабить ограничения метода Брауна, связанные с его *однопараметричностью*, за счет введения *двух параметров сглаживания λ_1 и λ_2* ($0 < \lambda_1 < 1, 0 < \lambda_2 < 1$)²². В его модели прогноз $\hat{x}(t; l)$ на l тактов времени в текущий момент t также определяется линейным трендом вида

$$\hat{x}(t; l) = \hat{a}_0(t; \lambda_1, \lambda_2) + l\hat{a}_1(t; \lambda_1, \lambda_2),$$

где обновление прогнозирующих коэффициентов проводится по формулам:

$$\begin{aligned}\hat{a}_0(t+1; \lambda_1, \lambda_2) &= \lambda_1 x(t) + (1 - \lambda_1)(\hat{a}_0(t; \lambda_1, \lambda_2) + \hat{a}_1(t; \lambda_1, \lambda_2)), \\ \hat{a}_1(t+1; \lambda_1, \lambda_2) &= \lambda_2(\hat{a}_0(t+1; \lambda_1, \lambda_2) - \hat{a}_1(t; \lambda_1, \lambda_2)) + \\ &\quad + (1 - \lambda_2)\hat{a}_1(t; \lambda_1, \lambda_2).\end{aligned}$$

Таким образом, прогноз по данному методу является функцией прошлых и текущих данных, параметров λ_1 и λ_2 , а также начальных значений $\hat{a}_0(0; \lambda_1, \lambda_2)$ и $\hat{a}_1(0; \lambda_1, \lambda_2)$.

Метод Хольта–Уинтерса. Метод Хольта был развит Уинтерсом (см.: *Winters P.R. Forecasting Sales by Exponentially Weighted Moving*

²²См.: *Holt C.C. Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages. Carnegie Inst. Tech. Res. Mem. №52, 1957.*

Averages. Mgmt. Sci., 6, p. 324) так, чтобы он охватывал, помимо линейного тренда, еще и сезонные эффекты. Прогноз, сделанный в момент t на l тактов времени вперед, равен

$$\hat{x}(t; l) = [\hat{a}_0(t) + l\hat{a}_1(t)]\omega(t + l - T),$$

где $\omega(\tau)$ — так называемый коэффициент сезонности, а T — число временных тактов, содержащихся в полном сезонном цикле, то есть в одном году (например, при месячных данных $T = 12$). Мы видим, что сезонность в этой формуле представлена *мультипликативно*. Метод использует *три параметра сглаживания (адаптации)* $\lambda_1, \lambda_2, \lambda_3$ ($0 < \lambda_j < 1, j = 1, 2, 3$), а его формулы обновления (пересчета прогнозирующих коэффициентов) имеют вид:

$$\hat{a}_0(t + 1) = \lambda_1 \frac{x(t + 1)}{\omega(t + 1 - T)} + (1 - \lambda_1)[\hat{a}_0(t) + \hat{a}_1(t)],$$

$$\omega(t + 1) = \lambda_2 \frac{x(t + 1)}{\hat{a}_0(t + 1)} + (1 - \lambda_2)\omega(t + 1 - T),$$

$$\hat{a}_1(t + 1) = \lambda_3 [\hat{a}_0(t + 1) - \hat{a}_0(t)] + (1 - \lambda_3)\hat{a}_1(t).$$

Как и в предыдущем методе, прогноз здесь строится на основании прошлых и текущих значений временного ряда, параметров адаптации λ_1, λ_2 и λ_3 , а также начальных значений $\hat{a}_0(0)$, $\hat{a}_1(0)$ и $\omega(0)$.

Аддитивная модель сезонности Тейла–Вейджа. Надо признать, что в экономической практике чаще встречаются экспоненциальные тенденции с мультипликативно наложенной сезонностью. Поэтому перед использованием *аддитивной* модели члены анализируемого временного ряда обычно заменяют их логарифмами и тем самым преобразуют экспоненциальную тенденцию в линейную и одновременно — мультипликативную сезонность в аддитивную. Преимущество же аддитивной модели — в относительной простоте ее вычислительной реализации. Будем считать, что логарифмическое преобразование исходных данных уже выполнено, и рассмотрим аддитивную модель вида

$$\begin{aligned} x(\tau) &= a_0(\tau) + \omega(\tau) + \delta(\tau), \\ a_0(\tau) &= a_0(\tau - 1) + a_1(\tau), \end{aligned}$$

где $a_0(\tau)$ — уровень процесса после элиминирования сезонных колебаний, $a_1(\tau)$ — аддитивный коэффициент роста, $\omega(\tau)$ — аддитивный коэффициент сезонности и $\delta(\tau)$ — белый шум.

Прогноз, сделанный в момент t на l временных тактов вперед, подсчитывается по формуле

$$\hat{x}(t; l) = \hat{a}_0(t) + l\hat{a}_1(t) + \hat{\omega}(t - T + l),$$

где коэффициенты $\hat{a}_0$, $\hat{a}_1$ и $\hat{\omega}$ вычисляются рекуррентным способом с помощью следующих формул обновления:

$$\begin{aligned}\hat{a}_0(\tau) &= \hat{a}_0(\tau - 1) + \hat{a}_1(\tau - 1) + \lambda_1[x(\tau) - \hat{x}(\tau - 1; 1)]; \\ \hat{a}_1(\tau) &= \hat{a}_1(\tau_1) + \lambda_1\lambda_2[x(\tau) - \hat{x}(\tau - 1; 1)]; \\ \hat{\omega}(\tau) &= \hat{\omega}(\tau - T) + (1 - \lambda_1)\lambda_3[x(\tau) - \hat{x}(\tau - 1; 1)].\end{aligned}$$

В этих соотношениях, как и прежде, T — это число временных тактов, содержащихся в полном сезонном цикле (как правило, в году), а λ_1 , λ_2 и λ_3 — параметры адаптации.

Мы уже упоминали при рассмотрении примера 10.14, что подбор подходящих значений для параметров адаптации является узким местом всех методов, основанных на экспоненциальном сглаживании. Обычно вся сложность решения этой проблемы перекладывается на компьютер: запасаются наборами значений — «кандидатов» в параметры сглаживания, для каждого набора значений подсчитывают серию прогнозов по данному методу и на основе сравнения получающихся при этом среднеквадратических ошибок прогнозов выбирают оптимальные значения для λ_j .

Выводы

1. Большинство математико-статистических методов, описанных в предыдущих главах, имеет дело с моделями, в которых наблюдения предполагаются *независимыми* и *одинаково распределенными*. При этом зависимость между наблюдениями чаще всего рассматривалась как помеха в эффективном применении этих методов. Однако разнообразные данные в экономике, социологии, финансах, коммерции и других сферах человеческой деятельности поступают в форме временных рядов, в которых наблюдения зависимы, и характер этой зависимости как раз и представляет главный интерес для исследователя. Совокупность существующих методов и моделей исследования таких рядов зависимых наблюдений называется **анализом временных рядов**.

2. Главная цель эконометрического анализа временных рядов состоит в построении по возможности простых и *экономично параметризованных* моделей, адекватно описывающих имеющиеся ряды наблюдений и составляющих базу для решения, в первую очередь, следующих задач:

- вскрытие *механизма генезиса* наблюдений, составляющих анализируемый временной ряд;
- построение *оптимального (то есть наиболее точного) прогноза* для будущих значений временного ряда;

- выработка *стратегии управления и оптимизации* анализируемых процессов.

3. Говоря о *генезисе образующих временной ряд наблюдений*, следует иметь в виду (и по возможности модельно описать) четыре типа факторов, под воздействием которых могут формироваться эти наблюдения: *долговременные, сезонные, циклические (или конъюнктурные) и случайные*. При этом не обязательно в процессе формирования значений конкретного временного ряда должны одновременно участвовать факторы всех четырех типов. Успешное решение задач выявления и моделирования действия этих факторов является основой, базисным отправным пунктом для достижения конечных прикладных целей исследования, главные из которых упомянуты в предыдущем пункте.

4. Приступая к анализу дискретного ряда наблюдений, расположенных в хронологическом порядке, следует, в первую очередь, убедиться, *действительно ли в формировании значений этого ряда участвовали какие-либо факторы, помимо чисто случайных*. При этом под «ч и с т о с л у ч а й н ы м и» понимаются лишь те случайные факторы, под воздействием которых генерируются последовательности взаимно не коррелированных и одинаково распределенных случайных величин, обладающих постоянными (не зависящими от времени) средними значениями и дисперсиями. Ответ на поставленный вопрос получают, проводя статистическую проверку соответствующей гипотезы с помощью одного из «критериев серий» или критерия Аббе (см. п. 10.3.1).

5. Если в результате проверки такой статистической гипотезы выяснилось, что имеющиеся наблюдения взаимно зависимы и неодинаково распределены (то есть образуют действительно *временной ряд*), то приступают к подбору подходящей модели для этого ряда. Множество моделей, в рамках которого ведется этот подбор, ограничивается обычно двумя классами линейных моделей: классом *стационарных временных рядов* (которые часто используются для описания поведения «случайных остатков») и классом *нестационарных временных рядов, которые могут содержать детерминированные и/или стохастические тренды*.

6. Стационарные (в широком смысле) временные ряды $x(t)$ характеризуются тем, что их средние значения $E x(t)$, дисперсии $D x(t)$ и ковариации $\gamma(\tau) = E[(x(t) - E x(t))(x(t + \tau) - E x(t + \tau))]$ не зависят от t , то есть остаются постоянными при изменении момента времени t , для которого они подсчитываются (см. п. 10.2). Взаимозависимости, существующие между членами стационарного временного ряда в широком классе случаев могут быть адекватно описаны в рамках моделей *авторегрессии порядка p* (АР(p)-моделей), моделей *скользящего среднего порядка q* (СС(q)-моделей) или моделей *авторегрессии со скользящими средними в остатках порядка p и q* (АРСС(p, q)-моделей), см. п. 10.4.

7. Среди нестационарных временных рядов выделим те, нестационарность которых определяется *только наличием детерминированной составляющей* в разложении (10.2) (это так называемые «**тренд-стационарные**» или просто — **ТС-ряды**), а также те, которые содержат *стохастический тренд*, но которые становятся стационарными при переходе к их последовательной разности соответствующего порядка (это так называемые «**дифференциально-стационарные**» или просто — **ДС-ряды**).

8. Проблема выявления нестационарности ряда тесно связана с проблемой отнесения анализируемого экономического временного ряда к одному из двух типов — *ТС- или ДС-рядам*. Можно констатировать, что в общем случае решение этих проблем еще далеко не завершено, однако в классе моделей АРСС предложен ряд статистических критериев для определения нестационарности/стационарности рассматриваемого ряда. Наиболее распространенными в эконометрической практике среди таких критериев можно признать **простой и расширенный тесты Дики-Фуллера (DF-тест и ADF-тест)**.

9. Поведение нестационарных временных рядов, в том числе рядов, содержащих сезонную компоненту, в эконометрических прикладных задачах в широком классе случаев достаточно успешно описывают с помощью *моделей авторегрессии — проинтегрированного скользящего среднего порядка p, q и k* (АРПСС(p, k, q)-моделей) или некоторых их модификаций (см. пп. 10.5.3–10.5.4).

10. Подобрать модель для конкретного временного ряда $\{x(t)\}$, $t = 1, 2, \dots, N$ — это значит определить подходящее параметрическое семейство моделей в качестве допустимого множества решений, а затем статистически оценить параметры модели на основании имеющихся наблюдений $x(1), x(2), \dots, x(N)$. Весь этот процесс принято называть *процессом идентификации модели*, или просто *идентификацией*.

11. Важную роль в системах поддержки принятия экономических решений играет прогнозирование экономических показателей. В главе 6 мы рассматривали методы прогнозирования, основанные на моделях регрессии. В данной главе внимание было сконцентрировано на методах *автопрогноза*, в которых имеющийся в наличии ряд экстраполируется вперед только на основании информации, содержащейся в нем самом. *Такого рода прогноз может оказаться эффективным лишь в краткосрочном, максимум, в среднесрочной перспективе*. Серьезное решение задач *долгосрочного* прогнозирования требует использования комплексных подходов и, в первую очередь, привлечения различных (в том числе статистических) технологий сбора и анализа экспертных оценок.

12. Наиболее эффективный подход к решению задач кратко- и среднесрочного автопрогноза — это прогнозирование, основанное на использовании «*подогнанных*» (идентифицированных) моделей типа АРПСС(p, k, q), включая, в качестве частных случаев, и модели АР-

СС- и АРСС.

13. Весьма широко распространены в решении прикладных задач кратко- и среднесрочного автопрогноза и так называемые адаптивные методы, позволяющие по мере поступления новых данных обновлять ранее сделанные прогнозы с минимальной задержкой и с помощью относительно несложных математических процедур.

ПРИЛОЖЕНИЕ 1. ТАБЛИЦЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Таблица П1.1. Значения функции плотности $\phi(x) = \phi(x; 0; 1)$ стандартного нормального закона распределения.

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

x	$\phi(x)$	x	$\phi(x)$	x	$\phi(x)$	x	$\phi(x)$	x	$\phi(x)$	x	$\phi(x)$
0,00	0,3989	0,50	0,3521	1,00	0,2420	1,50	0,1295	2,00	0,0540	2,50	0,0175
0,05	0,3984	0,55	0,3429	1,05	0,2299	1,55	0,1200	2,05	0,0488	2,55	0,0154
0,10	0,3970	0,60	0,3332	1,10	0,2179	1,60	0,1109	2,10	0,0440	2,60	0,0136
0,15	0,3945	0,65	0,3230	1,15	0,2059	1,65	0,1023	2,15	0,0396	2,65	0,0119
0,20	0,3910	0,70	0,3123	1,20	0,1942	1,70	0,0940	2,20	0,0355	2,70	0,0104
0,25	0,3867	0,75	0,3011	1,25	0,1826	1,75	0,0863	2,25	0,0317	2,75	0,0091
0,30	0,3814	0,80	0,2897	1,30	0,1714	1,80	0,0790	2,30	0,0283	2,80	0,0079
0,35	0,3752	0,85	0,2780	1,35	0,1604	1,85	0,0721	2,35	0,0252	2,85	0,0069
0,40	0,3683	0,90	0,2661	1,40	0,1497	1,90	0,0656	2,40	0,0224	2,90	0,0060
0,45	0,3605	0,95	0,2541	1,45	0,1394	1,95	0,0596	2,45	0,0198	2,95	0,0051
										3,00	0,0044

З а м е ч а н и е 1. Значения функции плотности $\phi(x_0; a; \sigma^2)$ нормального закона со средним a и дисперсией σ^2 подсчитывается по формуле

$$\phi(x_0; a; \sigma^2) = \frac{1}{\sigma} \phi\left(\frac{x_0 - a}{\sigma}; 0; 1\right) = \frac{1}{\sigma} \phi\left(\frac{x_0 - a}{\sigma}\right) \quad (\text{П1.1})$$

(если величина аргумента $(x_0 - a)/\sigma$ попадает *между* табличными значениями x , то для определения $\phi(\frac{x_0 - a}{\sigma})$ пользуются линейной интерполяцией функции $\phi(x)$).

З а м е ч а н и е 2. При определении значений функции $\phi(x)$ для отрицательных величин аргумента x следует использовать тождество (выражающее свойство *четности* функции $\phi(x)$)

$$\phi(-x) \equiv \phi(x). \quad (\text{П1.2})$$

П р и м е р П1.1. Требуется определить значение $\phi(x_0; a; \sigma^2)$ при $x_0 = 3,36$, $a = 1$ и $\sigma^2 = 4$.

Решение: $x = \frac{x_0 - a}{\sigma} = \frac{3,36 - 1}{2} = 1,18$. Два окаймляющих x соседних табличных значения аргумента — это $x_1 = 1,15$ и $x_2 = 1,20$, поэтому, используя линейную интерполяцию, получаем:

$$\phi(1,18) = \phi(1,15) - \frac{1,18 - 1,15}{1,20 - 1,15} [\phi(1,15) - \phi(1,20)] = 0,199.$$

Значение $\phi(3,36; 1; 4)$ получаем по формуле (П1.1):

$$\phi(x_0 = 3,36; 1; 4) = \frac{1}{2} \phi(1,18) = \frac{1}{2} 0,199 = 0,0995.$$

Таблица П1.2. Значения функции $\Phi(x) = \Phi(x; 0; 1)$ стандартного нормального распределения.

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,00	0,500000	1,00	0,841345	2,00	0,977250
0,05	0,519939	1,05	0,853141	2,05	0,979818
0,10	0,539828	1,10	0,864334	2,10	0,982136
0,15	0,559618	1,15	0,874928	2,15	0,984222
0,20	0,579260	1,20	0,884930	2,20	0,986097
0,25	0,589706	1,25	0,894350	2,25	0,987776
0,30	0,617911	1,30	0,903200	2,30	0,989276
0,35	0,636831	1,35	0,911492	2,35	0,990613
0,40	0,655422	1,40	0,919243	2,40	0,991802
0,45	0,673645	1,45	0,926471	2,45	0,992857
0,50	0,691463	1,50	0,933193	2,50	0,993790
0,55	0,708840	1,55	0,939429	2,55	0,994614
0,60	0,725747	1,60	0,945201	2,60	0,995339
0,65	0,742154	1,65	0,950528	2,65	0,995975
0,70	0,758036	1,70	0,955434	2,70	0,996533
0,75	0,773373	1,75	0,959941	2,75	0,997020
0,80	0,788145	1,80	0,964070	2,80	0,997445
0,85	0,802338	1,85	0,967843	2,85	0,997814
0,90	0,815940	1,90	0,971283	2,90	0,998134
0,95	0,828944	1,95	0,974412	2,95	0,998411
				3,00	0,998650

З а м е ч а н и е 1. Значение функции распределения $\Phi(x_0; a; \sigma^2)$ нормального закона при среднем значении a и дисперсии σ^2 в заданной точке x_0 подсчитывается по значениям функции $\Phi(x) = \Phi(x; 0; 1)$ с помощью формулы

$$\Phi(x_0; a; \sigma^2) = \Phi\left(\frac{x_0 - a}{\sigma}; 0; 1\right) = \Phi\left(\frac{x_0 - a}{\sigma}\right) \quad (\text{П1.3})$$

(если величина аргумента $(x_0 - a)/\sigma$ попадает *между* табличными значениями x , то для определения $\Phi((x_0 - a)/\sigma)$ пользуются линейной интерполяцией функции $\Phi(x)$).

З а м е ч а н и е 2. При определении значений функции $\Phi(x)$ для *отрицательных* величин аргумента x следует использовать тождество

$$\Phi(x) \equiv 1 - \Phi(-x). \quad (\text{П1.4})$$

П р и м е р П1.2 В условиях примера П1.1 имеем:

$$\begin{aligned} \Phi(x_0 = 3,36; a = 1; \sigma^2 = 4) &= \Phi(1,18) = \\ &= \Phi(1,15) + \frac{1,18 - 1,15}{1,20 - 1,15} [\Phi(1,20) - \Phi(1,15)] = 0,881. \end{aligned}$$

Таблица П1.3. Значения q -квантилей u_q стандартного нормального распределения

q	u_q	q	u_q	q	u_q	q	u_q
0,50	0,000000	0,70	0,524401	0,90	1,281552	0,983	120072
51	025069	71	553385	91	340755	984	144411
52	050154	72	582842	92	405072	0,985	2,170090
53	075270	73	612813	93	475791	986	197286
54	100434	74	643345	94	554774	987	226212
0,55	0,125661	0,75	0,674490	0,95	1,644854	988	257129
56	150969	76	706303	96	750686	989	290368
57	176374	77	738847	97	880794	0,990	2,326348
58	201893	78	772193	971	895698	991	365618
59	227545	79	806421	972	911036	992	408916
0,60	0,253347	0,80	0,841621	973	926837	993	457263
61	279319	81	877896	974	943134	994	512144
62	305481	82	915365	0,975	1,959964	0,995	2,575829
63	331853	83	954165	976	977368	996	652070
64	358459	84	994458	977	995393	997	747781
0,65	0,385320	0,85	1,036433	978	2,014091	998	878162
66	412463	86	080319	979	033520	999	3,090232
67	439913	87	126391	0,980	2,053749		
68	467699	88	174987	981	074855		
69	495850	89	226528	982	096927		

Пример П1.3. Найти 0,9-квантиль $u_{0,9}$. Величину $u_{0,9}$ находим из таблицы в графе, расположенной справа от соответствующего значения $q = 0,9$, то есть $u_{0,9} = 1,281552$.

Замечание 1. Если заданная величина q попадает между двумя соседними табличными значениями q_1 и q_2 ($q_1 < q_2$; это может случиться при графической проверке нормальности распределения), то следует воспользоваться линейной интерполяцией, а именно формулой

$$u_q = u_{q_1} + \frac{q - q_1}{q_2 - q_1} (u_{q_2} - u_{q_1}).$$

Замечание 2. При нахождении q -квантилей для значений $q < 0,5$ следует воспользоваться соотношением $u_q = -u_{1-q}$. Например, $u_{0,4} = -u_{1-0,4} = -u_{0,6} = -0,25335$.

Замечание 3. При отыскании 100Q%-ных точек w_Q следует воспользоваться соотношением $w_q = u_{1-q}$. Например, $w_{0,05} = u_{0,95} = 1,64485$.

Замечание 4. Из тождества (П1.3) следует, что q -квантиль $u_q(a; \sigma^2)$ нормального распределения, имеющего среднее значение a и дисперсию σ^2 , определяется по формуле

$$u_q(a; \sigma^2) = a + \sigma u_q. \quad (\text{П1.5})$$

Таблица П1.4. Значения $100Q\%$ -ных точек $\chi^2_Q(\nu)$ χ^2 -распределения с ν степенями свободы

ν	Q									
	0,995	0,990	0,975	0,950	0,900	0,100	0,050	0,025	0,010	0,005
1	$392704 \cdot 10^{-10}$	$157088 \cdot 10^{-9}$	$982069 \cdot 10^{-9}$	$393214 \cdot 10^{-8}$	0,0157908	2,70554	3,84146	5,02389	6,63490	7,87944
2	0,0100251	0,0201007	0,0506356	0,102587	0,210720	4,60517	5,99147	7,37776	9,21034	10,5966
3	0,0717212	0,114832	0,215795	0,351846	0,584375	6,25139	7,81473	9,34840	11,3449	12,8381
4	0,206990	0,297110	0,484419	0,710721	1,063623	7,77944	9,48773	11,1433	13,2767	14,8602
5	0,411740	0,554300	0,831211	1,145476	1,61031	9,23635	11,0705	12,8325	15,0863	16,7496
6	0,675727	0,872085	1,237347	1,63539	2,20413	10,6446	12,5916	14,4494	16,8119	18,5476
7	0,989265	1,239043	1,68987	2,16735	2,83311	12,0170	14,0671	16,0128	18,4753	20,2777
8	1,344419	1,646482	2,17973	2,73264	3,48954	13,3616	15,5073	17,5346	20,0902	21,9550
9	1,734926	2,087912	2,70039	3,32511	4,16816	14,6837	16,9190	19,0228	21,6660	23,5893
10	2,15585	2,55821	3,24697	3,94030	4,86518	15,9871	18,3070	20,4831	23,2093	25,1882
11	2,60321	3,05347	3,81575	4,57481	5,57779	17,2750	19,6751	21,9200	24,7250	26,7569
12	3,07382	3,57056	4,40379	5,22603	6,30380	18,5494	21,0261	23,3367	26,2170	28,2995
13	3,56503	4,10691	5,00874	5,89186	7,04150	19,8119	22,3621	24,7356	27,6883	29,8194
14	4,07468	4,66043	5,62872	6,57063	7,78953	21,0642	23,6848	26,1190	29,1413	31,3193
15	4,60094	5,22935	6,26214	7,26094	8,54675	22,3072	24,9958	27,4884	30,5779	32,8013
16	5,14224	5,81221	6,90766	7,96164	9,31223	23,5418	26,2962	28,8454	31,9999	34,2672
17	5,69724	6,40776	7,56418	8,67176	10,0852	24,7690	27,5871	30,1910	33,4087	35,7185
18	6,26481	7,01491	8,23075	9,39046	10,8649	25,9894	28,8693	31,5264	34,8053	37,1564
19	6,84398	7,63273	8,90655	10,1170	11,6509	27,2036	30,1435	32,8523	36,1908	38,5822

20	7,43386	8,26040	9,59083	10,8508	12,4426	28,4120	31,4104	34,1696	37,5662	39,9968
21	8,03366	8,89720	10,28293	11,5913	13,2396	29,6151	32,6705	35,4789	38,9321	41,4010
22	8,64272	9,54249	10,9823	12,3380	14,0415	30,8133	33,9244	36,7807	40,2894	42,7956
23	9,26042	10,19567	11,6885	13,0905	14,8479	32,0069	35,1725	38,0757	41,6384	44,1813
24	9,88623	10,8564	12,4011	13,8484	15,6587	33,1963	36,4151	39,3641	42,9798	45,5585
25	10,5197	11,5240	13,1197	14,6114	16,4734	34,3816	37,6525	40,6465	44,3141	46,9278
26	11,1603	12,1981	13,8439	15,3791	17,2919	35,5631	38,8852	41,9232	45,6417	48,2899
27	11,8076	12,8786	14,5733	16,1513	18,1138	36,7412	40,1133	43,1944	46,9630	49,6449
28	12,4613	13,5648	15,3079	16,9279	18,9392	37,9159	41,3372	44,4607	48,2782	50,9933
29	13,1211	14,2565	16,0471	17,7083	19,7677	39,0875	42,5569	45,7222	49,5879	52,3356
30	13,7867	14,9535	16,7908	18,4926	20,5992	40,2560	43,7729	46,9792	50,8922	53,6720
40	20,7065	22,1643	24,4331	26,5093	29,0505	51,8050	55,7585	59,3417	63,6907	66,7659
50	27,9907	29,7067	32,3574	34,7642	37,6886	63,1671	67,5048	71,4202	76,1539	79,4900
60	35,5346	37,4848	40,4817	43,1879	46,4589	74,3970	79,0819	83,2976	88,3794	91,9517
70	43,2752	45,4418	48,7576	51,7393	55,3290	85,5271	90,5312	95,0231	100,425	104,215
80	51,1720	53,5400	57,1532	60,3915	64,2778	96,5782	101,879	106,629	112,329	116,321
90	59,1963	61,7541	65,6466	69,1260	73,2912	107,565	113,145	118,136	124,116	128,299
100	67,3276	70,0648	74,2219	77,9295	82,3581	118,498	124,342	129,561	135,807	140,169

Таблица П1.5. Значения 100Q%-ных точек $v_Q^2(\nu_1, \nu_2)$ F-распределения с числом степеней свободы числителя ν_1 и знаменателя ν_2

ν_2	ν_1																			
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞	
$Q=0.1$																				
1	39,86	49,50	53,59	55,83	57,24	58,20	58,91	59,44	59,86	60,19	60,71	61,22	61,74	62,00	62,26	62,53	62,79	63,06	63,33	
2	8,53	9,00	9,16	9,24	9,29	9,33	9,35	9,37	9,38	9,39	9,41	9,42	9,44	9,45	9,46	9,47	9,47	9,48	9,49	
3	5,54	5,46	5,39	5,34	5,31	5,28	5,27	5,25	5,24	5,23	5,22	5,20	5,18	5,18	5,17	5,16	5,15	5,14	5,13	
4	4,54	4,32	4,19	4,11	4,05	4,01	3,98	3,95	3,94	3,92	3,90	3,87	3,84	3,83	3,82	3,80	3,79	3,78	3,76	
5	4,06	3,78	3,62	3,52	3,45	3,40	3,37	3,34	3,32	3,30	3,27	3,24	3,21	3,19	3,17	3,16	3,14	3,12	3,10	
6	3,78	3,46	3,29	3,18	3,11	3,05	3,01	2,98	2,96	2,94	2,90	2,87	2,84	2,82	2,80	2,78	2,76	2,74	2,72	
7	3,59	3,26	3,07	2,96	2,88	2,83	2,78	2,75	2,72	2,70	2,67	2,63	2,59	2,58	2,56	2,54	2,51	2,49	2,47	
8	3,46	3,11	2,92	2,81	2,73	2,67	2,62	2,59	2,56	2,54	2,50	2,46	2,42	2,40	2,38	2,36	2,34	2,32	2,29	
9	3,36	3,01	2,81	2,69	2,61	2,55	2,51	2,47	2,44	2,42	2,38	2,34	2,30	2,28	2,25	2,23	2,21	2,18	2,16	
10	3,29	2,92	2,73	2,61	2,52	2,46	2,41	2,38	2,35	2,32	2,28	2,24	2,20	2,18	2,16	2,13	2,11	2,08	2,06	
11	3,23	2,86	2,66	2,54	2,45	2,39	2,34	2,30	2,27	2,25	2,21	2,17	2,12	2,10	2,08	2,05	2,03	2,00	1,97	
12	3,18	2,81	2,61	2,48	2,39	2,33	2,28	2,24	2,21	2,19	2,15	2,10	2,06	2,04	2,01	1,99	1,96	1,93	1,90	
13	3,14	2,76	2,56	2,43	2,35	2,28	2,23	2,20	2,16	2,14	2,10	2,05	2,01	1,98	1,96	1,93	1,90	1,88	1,85	
14	3,10	2,73	2,52	2,39	2,31	2,24	2,19	2,15	2,12	2,10	2,05	2,01	1,96	1,94	1,91	1,89	1,86	1,83	1,80	
15	3,07	2,70	2,49	2,36	2,27	2,21	2,16	2,12	2,09	2,06	2,02	1,97	1,92	1,90	1,87	1,85	1,82	1,79	1,76	
16	3,05	2,67	2,46	2,33	2,24	2,18	2,13	2,09	2,06	2,03	1,99	1,94	1,89	1,87	1,84	1,81	1,78	1,75	1,72	
17	3,03	2,64	2,44	2,31	2,22	2,15	2,10	2,06	2,03	2,00	1,96	1,91	1,86	1,84	1,81	1,78	1,75	1,72	1,69	
18	3,01	2,62	2,42	2,29	2,20	2,13	2,08	2,04	2,00	1,98	1,93	1,89	1,84	1,81	1,78	1,75	1,72	1,69	1,66	
19	2,99	2,61	2,40	2,27	2,18	2,11	2,06	2,02	1,98	1,96	1,91	1,86	1,81	1,79	1,76	1,73	1,70	1,67	1,63	
20	2,97	2,59	2,38	2,25	2,16	2,09	2,04	2,00	1,96	1,94	1,89	1,84	1,79	1,77	1,74	1,71	1,68	1,64	1,61	
21	2,96	2,57	2,36	2,23	2,14	2,08	2,02	1,98	1,95	1,92	1,87	1,83	1,78	1,75	1,72	1,69	1,66	1,62	1,59	
22	2,95	2,56	2,35	2,22	2,13	2,06	2,01	1,97	1,93	1,90	1,86	1,81	1,76	1,73	1,70	1,67	1,64	1,60	1,57	
23	2,94	2,55	2,34	2,21	2,11	2,05	1,99	1,95	1,92	1,89	1,84	1,80	1,74	1,72	1,69	1,66	1,62	1,59	1,55	
24	2,93	2,54	2,33	2,19	2,10	2,04	1,98	1,94	1,91	1,88	1,83	1,78	1,73	1,70	1,67	1,64	1,61	1,57	1,53	

25	2,92	2,53	2,32	2,18	2,09	2,02	1,97	1,93	1,89	1,87	1,82	1,77	1,72	1,69	1,66	1,63	1,59	1,56	1,52
26	2,91	2,52	2,31	2,17	2,08	2,01	1,96	1,92	1,88	1,86	1,81	1,76	1,71	1,68	1,65	1,61	1,58	1,54	1,50
27	2,90	2,51	2,30	2,17	2,07	2,00	1,95	1,91	1,87	1,85	1,80	1,75	1,70	1,67	1,64	1,60	1,57	1,53	1,49
28	2,89	2,50	2,29	2,16	2,06	2,00	1,94	1,90	1,87	1,84	1,79	1,74	1,69	1,66	1,63	1,59	1,56	1,52	1,48
29	2,89	2,50	2,28	2,15	2,06	1,99	1,93	1,89	1,86	1,83	1,78	1,73	1,68	1,65	1,62	1,58	1,55	1,51	1,47
30	2,88	2,49	2,28	2,14	2,05	1,98	1,93	1,88	1,85	1,82	1,77	1,72	1,67	1,64	1,61	1,57	1,54	1,50	1,46
40	2,84	2,44	2,23	2,09	2,00	1,93	1,87	1,83	1,79	1,76	1,71	1,66	1,61	1,57	1,54	1,51	1,47	1,42	1,38
60	2,79	2,39	2,18	2,04	1,95	1,87	1,82	1,77	1,74	1,71	1,66	1,60	1,54	1,51	1,48	1,44	1,40	1,35	1,29
120	2,75	2,35	2,13	1,99	1,90	1,82	1,77	1,72	1,68	1,65	1,60	1,55	1,48	1,45	1,41	1,37	1,32	1,26	1,19
∞	2,71	2,30	2,08	1,94	1,85	1,77	1,72	1,67	1,63	1,60	1,55	1,49	1,42	1,38	1,34	1,30	1,24	1,17	1,00

$Q=0,05$

1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	243,9	245,9	248,0	249,1	250,1	251,1	252,2	253,3	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41	19,43	19,45	19,45	19,46	19,47	19,48	19,49	19,50
3	10,13	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	5,06	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

Продолжение таблицы П1.5

v_1	v_2																			∞
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120		
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	
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	2.49	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	1.66	1.61	1.55	1.50	1.43	1.35	1.25	
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00	

$Q=0.01$

1	4052	399.5	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

 $Q = 0,01$

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
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

Примечание. При вычислении 100Q%-ых точек для $Q \geq 0,9$ следует воспользоваться тождеством $u_Q^2(v_1, v_2) = (u_{1-Q}^2(v_1, v_2))^{-1}$.

Таблица П1.6. Значения 100Q%-ных точек $t_Q(\nu)$ распределения Стьюдента с ν степенями свободы

ν	$Q=0.4$ $2Q=0.8$	0.25 0.5	0.1 0.2	0.05 0.1	0.025 0.05	0.01 0.02	0.005 0.01	0.0025 0.005
1	0,325	1,000	3,078	6,314	12,706	31,821	63,657	127,32
2	289	0,816	1,886	2,920	4,303	6,965	9,925	14,089
3	277	765	1,638	2,353	3,182	4,541	5,841	7,453
4	271	741	1,533	2,132	2,776	3,747	4,604	5,598
5	0,267	0,727	1,476	2,015	2,571	3,365	4,032	4,773
6	265	718	1,440	1,943	2,447	3,143	3,707	4,317
7	263	711	1,415	1,895	2,365	2,998	3,499	4,029
8	262	706	1,397	1,860	2,306	2,896	3,355	3,833
9	261	703	1,383	1,833	2,262	2,821	3,250	3,690
10	0,260	0,700	1,372	1,812	2,228	2,764	3,169	3,581
11	260	697	1,363	1,796	2,201	2,718	3,106	3,497
12	259	695	1,356	1,782	2,179	2,681	3,055	3,428
13	259	694	1,350	1,771	2,160	2,650	3,012	3,372
14	258	692	1,345	1,761	2,145	2,624	2,977	3,326
15	0,258	0,691	1,341	1,753	2,131	2,602	2,947	3,286
16	258	690	1,337	1,746	2,120	2,583	2,921	3,252
17	257	689	1,333	1,740	2,110	2,567	2,898	3,222
18	257	688	1,330	1,734	2,101	2,552	2,878	3,197
19	257	688	1,328	1,729	2,093	2,539	2,861	3,174
20	0,257	0,687	1,325	1,725	2,086	2,528	2,845	3,153
21	257	686	1,323	1,721	2,080	2,518	2,831	3,135
22	256	686	1,321	1,717	2,074	2,508	2,819	3,119
23	256	685	1,319	1,714	2,069	2,500	2,807	3,104
24	256	685	1,318	1,711	2,064	2,492	2,797	3,091
25	0,256	0,684	1,316	1,708	2,060	2,485	2,787	3,078
26	256	684	1,315	1,706	2,056	2,479	2,779	3,067
27	256	684	1,314	1,703	2,052	2,473	2,771	3,057
28	256	683	1,313	1,701	2,048	2,467	2,763	3,047
29	256	683	1,311	1,699	2,045	2,462	2,756	3,038
30	0,256	0,683	1,310	1,697	2,042	2,457	2,750	3,030
40	255	681	1,303	1,684	2,021	2,423	2,704	2,971
60	254	679	1,296	1,671	2,000	2,390	2,660	2,915
120	254	677	1,289	1,658	1,980	2,358	2,617	2,860
∞	253	674	1,282	1,645	1,960	2,326	2,576	2,807

Таблица П1.7. Преобразование Фишера (z -преобразование) выборочного коэффициента корреляции $\hat{r}$ ($z = \operatorname{arctanh} \hat{r}$)

$\hat{r}$	,000	,002	,004	,006	,008	$\hat{r}$	,000	,002	,004	,006	,008
0,00	0000	0020	0040	0060	0080	0,50	5493	5520	5547	5573	5600
1	0100	0120	0140	0160	0180	1	5627	5654	5682	5709	5736
2	0200	0220	0240	0260	0280	2	5763	5791	5818	5846	5874
3	0300	0320	0340	0360	0380	3	5901	5929	5957	5985	6013
4	0400	0420	0440	0460	0480	4	6042	6070	6098	6127	6155
0,05	0500	0520	0541	0561	0581	0,55	6184	6213	6241	6270	6299
6	0601	0621	0641	0661	0681	6	6328	6358	6387	6416	6446
7	0701	0721	0741	0761	0782	7	6475	6505	6535	6565	6595
8	0802	0822	0842	0862	0882	8	6625	6655	6685	6716	6746
9	0902	0923	0943	0963	0983	9	6777	6807	6838	6869	6900
0,10	1003	1024	1044	1064	1084	0,60	6931	6963	6994	7026	7057
1	1104	1125	1145	1165	1186	1	7089	7121	7153	7185	7218
2	1206	1226	1246	1267	1287	2	7250	7283	7315	7348	7381
3	1307	1328	1348	1368	1389	3	7414	7447	7481	7514	7548
4	1409	1430	1450	1471	1491	4	7582	7616	7650	7684	7718
0,15	1511	1532	1552	1573	1593	0,65	7753	7788	7823	7858	7893
6	1614	1634	1655	1676	1696	6	7928	7964	7999	8035	8071
7	1717	1737	1758	1779	1799	7	8107	8144	8180	8217	8254
8	1820	1841	1861	1882	1903	8	8291	8328	8366	8404	8441
9	1923	1944	1965	1986	2007	9	8480	8518	8556	8595	8634
0,20	2027	2048	2069	2090	2111	0,70	8673	8712	8752	8792	8832
1	2132	2153	2174	2195	2216	1	8872	8912	8953	8994	9035
2	2237	2258	2279	2300	2321	2	9076	9118	9160	9202	9245
3	2342	2363	2384	2405	2427	3	9287	9330	9373	9417	9461
4	2448	2469	2490	2512	2533	4	9505	9549	9594	9639	9684
0,25	2554	2575	2597	2618	2640	0,75	0,973	0,978	0,982	0,987	0,991
6	2661	2683	2704	2726	2747	6	0,996	1,001	1,006	1,011	1,015
7	2769	2790	2812	2833	2855	7	1,020	1,025	1,030	1,035	1,040
8	2877	2899	2920	2942	2964	8	1,045	1,050	1,056	1,061	1,066
9	2986	3008	3029	3051	3073	9	1,071	1,077	1,082	1,088	1,093
0,30	3095	3117	3139	3161	3183	0,80	1,099	1,104	1,110	1,116	1,121
1	3205	3228	3250	3272	3294	1	1,127	1,133	1,139	1,145	1,151
2	3316	3339	3361	3383	3406	2	1,157	1,163	1,169	1,175	1,182
3	3428	3451	3473	3496	3518	3	1,188	1,195	1,201	1,208	1,214
4	3541	3564	3586	3609	3632	4	1,221	1,228	1,235	1,242	1,249
0,35	3654	3677	3700	3723	3746	0,85	1,256	1,263	1,271	1,278	1,286
6	3769	3792	3815	3838	3861	6	1,293	1,301	1,309	1,317	1,325
7	3884	3907	3931	3954	3977	7	1,333	1,341	1,350	1,358	1,367
8	4001	4024	4047	4071	4094	8	1,376	1,385	1,394	1,403	1,412
9	4118	4142	4165	4189	4213	9	1,422	1,432	1,442	1,452	1,462

Продолжение таблицы П1.7

$\hat{r}$	,000	,002	,004	,006	,008	$\hat{r}$	,000	,002	,004	,006	,008
0,40	4236	4260	4284	4308	4332	0,90	1,472	1,483	1,494	1,505	1,516
1	4356	4380	4404	4428	4453	1	1,528	1,539	1,551	1,564	1,576
2	4477	4501	4526	4550	4574	2	1,589	1,602	1,616	1,630	1,644
3	4599	4624	4648	4673	4698	3	1,658	1,673	1,689	1,705	1,721
4	4722	4747	4772	4797	4822	4	1,738	1,756	1,774	1,792	1,812
0,45	4847	4872	4897	4922	4948	0,95	1,832	1,853	1,874	1,897	1,921
6	4973	4999	5024	5049	5075	6	1,946	1,972	2,000	2,029	2,060
7	5101	5126	5152	5178	5204	7	2,092	2,127	2,165	2,205	2,249
8	5230	5256	5282	5308	5334	8	2,298	2,351	2,410	2,477	2,555
9	5361	5387	5413	5440	5466	9	2,647	2,759	2,903	3,106	3,453
$\hat{r}$	,000	,002	,004	,006	,008	$\hat{r}$	,000	,002	,004	,006	,008

Примеры.

1. Дано $\hat{r} = 0,206$. Определить $z = \operatorname{arcth} 0,206$.

Находим (в левом столбце таблицы) строку, соответствующую $\hat{r} = 0,20$. Чтобы получить заданное значение $\hat{r}$, к 0,20 надо прибавить 0,006, а потому искомое число находится (в этой строке) в столбце, расположенном под 0,006. Итак, $z = \operatorname{arcth} 0,206 = 0,2090$.

2. Дано $\hat{r} = -0,515$. Определить $z = \operatorname{arcth} (-0,515)$.

Находим (в левом столбце таблицы) строку, соответствующую $\hat{r} = 0,51$. Чтобы получить значение $\hat{r} = 0,515$, к 0,51 надо прибавить 0,005, а потому $\operatorname{arcth} 0,515$ находится как среднее арифметическое двух чисел данной строки, расположенных в столбцах, соответствующих верхним индексам 0,006 и 0,004, то есть

$$\operatorname{arcth} 0,515 = \frac{0,5709 + 0,5682}{2} = 0,56955.$$

Соответственно, $z = \operatorname{arcth} (-0,515) = -\operatorname{arcth} 0,515 = -0,56955$.

3. Дано $z = 0,8752$. Определить $\hat{r} = \operatorname{th} z$.

Находим в таблице число, равное 0,8752, и определяем, какому значению $\hat{r}$ оно соответствует. В нашем случае $\hat{r} = 0,704$.

Примечание. В тех случаях, когда в таблице не найдется в точности заданного числа, берут два приближенных (ближайших к нему) значения — с недостатком и с избытком. Искомое значение $\hat{r}$ будет лежать между двумя значениями $\hat{r}_1$ и $\hat{r}_2$, соответствующими этим приближенным величинам z .

Таблица П1.8. Верхняя граница доверительного интервала для истинного значения коэффициента корреляции при условии отсутствия линейной корреляционной связи (при доверительной вероятности $P = 1 - 2Q$)

n - 2	Q					
	0,05	0,025	0,01	0,005	0,0025	0,0005
1	0,9877	0,94692	0,94507	0,93877	0,94692	0,94877
2	9000	9500	9800	92000	92500	92000
3	805	878	9343	9587	9740	9114
4	729	811	882	9172	9417	9741
5	669	754	833	875	9056	9509
6	0,621	0,707	0,789	0,834	0,870	0,9249
7	582	666	750	798	836	898
8	549	632	715	765	805	872
9	521	602	685	735	776	847
10	497	576	658	708	750	823
11	0,476	0,553	0,634	0,684	0,726	0,801
12	457	532	612	661	703	780
13	441	514	592	641	683	760
14	426	497	574	623	664	742
15	412	482	558	606	647	725
16	0,400	0,468	0,543	0,590	0,631	0,708
17	389	456	529	575	616	693
18	378	444	516	561	602	679
19	369	433	503	549	589	665
20	360	423	492	537	576	652
25	0,323	0,381	0,445	0,487	0,524	0,597
30	296	349	409	449	484	554
35	275	325	381	418	452	519
40	257	304	358	393	425	490
45	243	288	338	372	403	465
50	0,231	0,273	0,322	0,354	0,384	0,443
60	211	250	295	325	352	408
70	195	232	274	302	327	380
80	183	217	257	283	307	357
90	173	205	242	267	290	338
100	164	195	230	254	276	321

П р и м е ч а н и е. Верхний индекс (2,3 и т. д.) над цифрой 9 означает, что эта цифра занимает первые 2,3 и т. д. разряда десятичной дроби. Например, $0,9^{4692} = 0,9999692$.

П р и м е р. Если мы оцениваем корреляционную связь по $n = 20$ наблюдениям, то при доверительной вероятности $P = 0,95$ (т. е. при $2Q = 0,05$) значение коэффициента корреляции, не превосходящее по абсолютной величине 0,444, еще не говорит о статистической значимости этой корреляционной связи (т. е. о том, что истинное значение коэффициента корреляции ρ отлично от нуля).

Таблица П1.9.

Проверка статистической значимости корреляционной связи
с помощью рангового коэффициента корреляции Спирмена $\hat{\tau}(S)$

$n=4$		$n=5$		$n=6$		$n=7$		$n=8$		$n=9$		$n=10$	
S_C	Q	S_C	Q	S_C	Q	S_C	Q	S_C	Q	S_C	Q	S_C	Q
12	0,458	22	0,475	50	0,210	74	0,249	108	0,250	156	0,218	208	0,235
14	375	24	392	52	178	78	198	114	195	164	168	218	184
16	208	26	342	54	149	82	151	120	150	172	125	228	139
18	167	28	258	56	121	86	118	126	108	180	089	238	102
20	042	30	225	58	088	90	083	132	076	188	060	248	072
		32	0,175	60	0,068	94	0,055	138	0,048	196	0,038	258	0,048
		34	117	62	051	98	033	144	029	204	022	268	030
		36	067	64	029	102	017	150	014	212	011	278	017
		38	042	66	017	106	0062	156	0054	220	0041	288	0087
		40	0083	68	0083	110	0014	162	0011	228	0010	298	0036
				70	0,0014							308	0,001
20		40		70		112		168		240		330	

Таблица П1.10.

Проверка статистической значимости корреляционной связи
с помощью рангового коэффициента корреляции Кендалла $\hat{\tau}(K)$

S_K	n				S_K	n		
	4	5	8	9		6	7	10
0	0,625	0,592	0,548	0,540	1	0,500	0,500	0,500
2	375	408	452	460	3	360	386	431
4	167	242	360	381	5	235	281	364
6	042	117	274	306	7	136	191	300
8		042	199	238	9	068	119	242
10		0,0083	0,138	0,179	11	0,028	0,068	0,190
12			089	130	13	0083	035	146
14			054	090	15	0014	015	108
16			031	060	17		0054	078
18			016	038	19		0014	054
20			0,0071	0,022	21		0,0002	0,036
22			0028	012	23			023
24			0009	0063	25			014
26			0002	0029	27			0083
28				0012	29			0046
30				0,0004	31			0,0023
					33			0011

Продолжение Таблицы П1.11^а

Вероятность того, что данное значение S будет достигнуто или превзойдено, для $n=4$, $m=3$ и $m=5$

S	$m=3$	$m=5$	S	$m=5$
1	1,000	1,000	61	0,055
3	0,958	0,975	65	0,044
5	0,910	0,944	67	0,034
9	0,727	0,857	69	0,031
11	0,608	0,771	73	0,023
13	0,524	0,709	75	0,020
17	0,446	0,652	77	0,017
19	0,342	0,561	81	0,012
21	0,300	0,521	83	0,0087
25	0,207	0,445	85	0,0067
27	0,175	0,408	89	0,0055
29	0,148	0,372	91	0,0031
33	0,075	0,298	93	0,0023
35	0,054	0,260	97	0,0018
37	0,033	0,226	99	0,0016
41	0,017	0,210	101	0,0014
43	0,0017	0,162	105	0,00064
45	0,0017	0,141	107	0,00033
49		0,123	109	0,00021
51		0,107	113	0,00014
53		0,093	117	0,000048
57		0,075	125	0,000030
59		0,067		

Вероятность того, что данное значение S будет достигнуто или превзойдено, для $n=4$ и $m=2$, $m=4$ и $m=6$

S	$m=2$	$m=4$	$m=6$	S	$m=6$
0	1,000	1,000	1,000	82	0,035
2	0,958	0,992	0,996	84	0,032
4	0,833	0,928	0,957	86	0,029
6	0,792	0,900	0,940	88	0,023
8	0,625	0,800	0,874	90	0,022
10	0,542	0,754	0,844	94	0,017
12	0,458	0,677	0,789	96	0,014
14	0,375	0,649	0,772	98	0,013
16	0,208	0,524	0,679	100	0,010
18	0,167	0,508	0,668	102	0,0096
20	0,042	0,432	0,609	104	0,0085
22		0,389	0,574	106	0,0073
24		0,355	0,541	108	0,0061

Продолжение Таблицы П1.11*

S	$m=2$	$m=4$	$m=6$	S	$m=6$
26		0,324	0,512	110	0,0057
30		0,242	0,431	114	0,0040
32		0,200	0,386	116	0,0033
34		0,190	0,375	118	0,0028
36		0,158	0,338	120	0,0023
38		0,141	0,317	122	0,0020
40		0,105	0,270	126	0,0015
42		0,094	0,256	128	0,0010
44		0,077	0,230	130	0,00087
46		0,068	0,218	132	0,00073
48		0,054	0,197	134	0,00065
50		0,052	0,194	136	0,00040
52		0,036	0,163	138	0,00036
54		0,033	0,155	140	0,00028
56		0,019	0,127	144	0,00024
58		0,014	0,114	146	0,00022
62		0,012	0,108	148	0,00012
64		0,0069	0,089	150	0,000095
66		0,0062	0,088	152	0,000062
68		0,0027	0,073	154	0,000046
70		0,0027	0,066	158	0,000024
72		0,0016	0,060	160	0,000016
74		0,00094	0,056	162	0,000012
76		0,00094	0,043	164	0,0000080
78		0,00094	0,041	170	0,0000024
80		0,00072	0,037	180	0,0000013

Вероятность того, что данное значение S будет достигнуто или превзойдено, для $n=5$ и $m=3$

S	$m=3$	S	$m=3$	S	$m=3$	S	$m=3$
0	1,000	22	0,649	44	0,236	66	0,038
2	1,000	24	0,595	46	0,213	68	0,028
4	0,988	26	0,559	48	0,172	70	0,026
6	0,972	28	0,493	50	0,163	72	0,017
8	0,941	30	0,475	52	0,127	74	0,015
10	0,914	32	0,432	54	0,117	76	0,0078
12	0,845	34	0,406	56	0,096	78	0,0053
14	0,831	36	0,347	58	0,080	80	0,0040
16	0,768	38	0,326	60	0,063	82	0,0028
18	0,720	40	0,291	62	0,056	86	0,00090
20	0,682	42	0,253	64	0,045	90	0,000069

Таблица П1.11⁶. Критические значения статистики S при уровне значимости $\alpha = 0,05$ для проверки статистической значимости выборочного значения коэффициента конкордации $\widehat{W}(m)$

m	n					Дополнительные значения для $n = 3$	
	3	4	5	6	7	m	S
3			64,4	103,9	157,3	9	54,0
4		49,5	88,4	143,3	217,0	12	71,9
5		62,6	112,3	182,4	276,2	14	83,8
6		75,7	136,1	221,4	335,2	16	95,8
8	48,1	101,7	183,7	299,0	453,1	18	107,7
10	60,0	127,8	231,2	376,7	571,0		
15	89,8	192,9	349,8	570,5	864,9		
20	119,7	258,0	468,5	764,4	1158,7		

Таблица П1.12^а. Значения статистик d_L и d_U критерия

 Дарбина-Уотсона при уровне значимости $\alpha = 0,05$

 (n — число наблюдений, p — число объясняющих переменных)

n	$p = 1$		$p = 2$		$p = 3$		$p = 4$		$p = 5$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
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

Таблица П1.12⁶. Значения статистик d_L и d_U критерияДарбина–Уотсона при уровне значимости $\alpha = 0,01$ $(n$ — число наблюдений, p — число объясняющих переменных)

n	$p = 1$		$p = 2$		$p = 3$		$p = 4$		$p = 5$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	0,81	1,07	0,70	1,25	0,59	1,46	0,49	1,70	0,39	1,96
16	0,84	1,09	0,74	1,25	0,63	1,44	0,534	1,66	0,44	1,90
17	0,87	1,10	0,77	1,25	0,67	1,43	0,57	1,63	0,48	1,85
18	0,90	1,12	0,80	1,26	0,71	1,42	0,61	1,60	0,52	1,80
19	0,93	1,13	0,83	1,26	0,74	1,41	0,65	1,58	0,56	1,77
20	0,95	1,15	0,86	1,27	0,77	1,41	0,68	1,57	0,60	1,74
21	0,97	1,16	0,89	1,27	0,80	1,41	0,72	1,55	0,63	1,71
22	1,00	1,17	0,91	1,28	0,83	1,40	0,75	1,54	0,66	1,69
23	1,02	1,19	0,94	1,29	0,86	1,40	0,77	1,53	0,70	1,67
24	1,04	1,20	0,96	1,30	0,88	1,41	0,80	1,53	0,72	1,66
25	1,05	1,21	0,98	1,30	0,90	1,41	0,83	1,52	0,75	1,65
26	1,07	1,22	1,00	1,31	0,93	1,41	0,85	1,52	0,78	1,64
27	1,09	1,23	1,02	1,32	0,95	1,41	0,88	1,51	0,81	1,63
28	1,10	1,24	1,04	1,32	0,97	1,41	0,90	1,51	0,83	1,62
29	1,12	1,25	1,05	1,33	0,99	1,42	0,92	1,51	0,85	1,61
30	1,13	1,26	1,07	1,34	1,01	1,42	0,94	1,51	0,88	1,61
31	1,15	1,27	1,08	1,34	1,02	1,42	0,96	1,51	0,90	1,60
32	1,16	1,28	1,10	1,35	1,04	1,43	0,98	1,51	0,92	1,60
33	1,17	1,29	1,11	1,36	1,05	1,43	1,00	1,51	0,94	1,59
34	1,18	1,30	1,13	1,36	1,07	1,43	1,01	1,51	0,95	1,59
35	1,19	1,31	1,14	1,37	1,08	1,44	1,03	1,51	0,97	1,59
36	1,21	1,32	1,15	1,38	1,10	1,44	1,04	1,51	0,99	1,59
37	1,22	1,32	1,16	1,38	1,11	1,45	1,06	1,51	1,00	1,59
38	1,23	1,33	1,18	1,39	1,12	1,45	1,07	1,52	1,02	1,58
39	1,24	1,34	1,19	1,39	1,14	1,45	1,09	1,52	1,03	1,58
40	1,25	1,34	1,20	1,40	1,15	1,46	1,10	1,52	1,05	1,58
45	1,29	1,38	1,24	1,42	1,20	1,48	1,16	1,53	1,11	1,58
50	1,32	1,40	1,28	1,45	1,24	1,49	1,20	1,54	1,16	1,59
55	1,36	1,43	1,32	1,47	1,28	1,51	1,25	1,55	1,21	1,59
60	1,38	1,45	1,35	1,48	1,32	1,52	1,28	1,56	1,25	1,60
65	1,41	1,47	1,38	1,50	1,35	1,53	1,31	1,57	1,28	1,61
70	1,43	1,49	1,40	1,52	1,37	1,55	1,34	1,58	1,31	1,61
75	1,45	1,50	1,42	1,53	1,39	1,56	1,37	1,59	1,34	1,62
80	1,47	1,52	1,44	1,54	1,42	1,57	1,39	1,60	1,36	1,62
85	1,48	1,53	1,46	1,55	1,43	1,58	1,41	1,60	1,39	1,63
90	1,50	1,54	1,47	1,56	1,45	1,59	1,43	1,61	1,41	1,64
95	1,51	1,55	1,49	1,57	1,47	1,60	1,45	1,62	1,42	1,64
100	1,52	1,56	1,50	1,58	1,48	1,60	1,46	1,63	1,44	1,65

ПРИЛОЖЕНИЕ 2. НЕОБХОДИМЫЕ СВЕДЕНИЯ ИЗ МАТРИЧНОЙ АЛГЕБРЫ

Все модели и методы эконометрики, эксплуатирующие понятия и приемы *многомерного статистического анализа*, — а именно таковыми являются модели и методы, рассматриваемые в данном издании, — *требуют для своего компактного и эффективного описания и анализа использования аппарата матричной алгебры*. И если при описании этих моделей мы еще подчас можем обойтись без такого аппарата (правда, за счет утомительных и чрезвычайно громоздких выкладок и выражений, приводящих к *покомпонентной* записи требуемых соотношений), то при анализе их свойств и выводе связанных с этим математических результатов без использования методов линейной алгебры мы вынуждены были бы в ряде ситуаций просто отступить, не добившись желаемого.

Поэтому специалисту (или студенту, готовящемуся стать таковым), предполагающему всерьез работать с эконометрическими методами и моделями, необходимо внутренне настроиться на неизбежность расходования некоторого времени и интеллектуальных усилий на овладение определенным минимумом сведений из матричной алгебры.

При описании этих сведений мы будем следовать обозначениям, принятым в данном учебнике, а именно: буквы, набранные *жирным* шрифтом, будут обозначать *матрицы*, а буквы *прописные* будут использоваться для обозначения *векторов* (представляющих собой, правда, важный частный случай тех же матриц). Доказательства некоторых из приводимых здесь фактов из линейной алгебры будут опускаться.

П 2.1. Виды матриц и действия с ними

Матрицей называется прямоугольная таблица чисел (элементов) вида

$$\mathbf{A} = (a_{ij})_{\substack{i=1,2,\dots,n; \\ j=1,2,\dots,m}} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix}. \quad (\text{П2.1})$$

В общем обозначении элемента a_{ij} матрицы $\mathbf{A}$ первый нижний индекс (i) указывает номер той строки, а второй нижний индекс (j) — номер того столбца, на пересечении которых находится этот элемент. Если матрица $\mathbf{A}$ содержит nm элементов, расположенных на n строках и m столбцах, то говорят, что мы имеем $n \times m$ -матрицу $\mathbf{A}$ или матрицу $\mathbf{A}$ размерности $n \times m$.

Квадратной матрицей называется матрица, у которой число строк равно числу столбцов, то есть $n \times n$ -матрица (при любом це-

лом $n \geq 1$). Элементы $a_{11}, a_{22}, \dots, a_{nn}$ квадратной матрицы образуют ее *главную диагональ* и называются *диагональными*.

Среди квадратных матриц выделим:

- *диагональную матрицу* $\mathbf{D}$, у которой все элементы, кроме элементов, стоящих на главной диагонали, равны нулю, то есть

$$\mathbf{D} = \begin{pmatrix} d_{11} & 0 & 0 & \dots & 0 & 0 \\ 0 & d_{22} & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & d_{nn} \end{pmatrix} = \text{diag}(d_{11}, d_{22}, \dots, d_{nn});$$

заметим, что ковариационная матрица $\Sigma = (\sigma_{ij})$ любой последовательности $\xi^{(1)}, \xi^{(2)}, \dots, \xi^{(n)}$ *взаимнонекоррелированных* случайных величин является диагональной, так как $\sigma_{ij} = \mathbf{E}[(\xi^{(i)} - \mathbf{E}\xi^{(i)})(\xi^{(j)} - \mathbf{E}\xi^{(j)})] = 0$ при любых $i \neq j$.

- *единичную матрицу* $\mathbf{I}_n$, которая является частным случаем $n \times n$ диагональной матрицы, поскольку все ее диагональные элементы равны единице:

$$\mathbf{I}_n = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix};$$

нижний индекс n определяет размерность матрицы, и в тех случаях, когда эта размерность очевидна из контекста, он может опускаться.

- *симметрическую (симметричную)* матрицу, а именно такую матрицу, у которой элементы, симметрично расположенные относительно главной диагонали, равны между собой, то есть $a_{ij} = a_{ji}$ для всех $i, j = 1, 2, \dots, n$; заметим, что всякая ковариационная матрица $\Sigma = (\sigma_{ij})$ является симметрической по определению, так как $\sigma_{ij} = \mathbf{E}[(\xi^{(i)} - \mathbf{E}\xi^{(i)})(\xi^{(j)} - \mathbf{E}\xi^{(j)})] = \mathbf{E}[(\xi^{(j)} - \mathbf{E}\xi^{(j)})(\xi^{(i)} - \mathbf{E}\xi^{(i)})] = \sigma_{ji}$.

Равенство матриц. Две матрицы $\mathbf{A} = (a_{ij})$ и $\mathbf{B} = (b_{ij})$ называются равными, если они имеют одну и ту же размерность и если $a_{ij} = b_{ij}$ для всех i и j . Это означает, что равные матрицы *совпадают поэлементно*.

Вектор-строка — это матрица размерности $1 \times m$ ($m > 1$), то есть матрица, состоящая из единственной строки длины m . Например, результаты наблюдения значений p анализируемых переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ на одном объекте, зарегистрированные в определенный момент времени t , образуют вектор-строку

$$X_t = (x_t^{(1)}, x_t^{(2)}, \dots, x_t^{(p)}). \quad (\text{П2.2})$$

Вектор-столбец — это матрица размерности $n \times 1$ ($n > 1$), то есть матрица, состоящая из единственного столбца длины n . Например, результаты наблюдения одной какой-либо переменной $x^{(j)}$ на n статистически обследованных объектах (или на одном объекте, но зарегистрированные в n последовательных моментов времени) можно представить в виде вектора-столбца

$$X^{(j)} = \begin{pmatrix} x_1^{(j)} \\ x_2^{(j)} \\ \vdots \\ x_n^{(j)} \end{pmatrix}. \quad (\text{П2.3})$$

Чтобы выделить эти специальные частные случаи матриц, мы будем обозначать векторы-строки и векторы-столбцы *прописными* буквами алфавита, а их компоненты — *теми же самыми, но строчными* буквами.

Условимся обозначать в дальнейшем вектор-строку длины m , состоящую из одних нулей, с помощью $0_{1,m}$ (или просто 0_m , если из контекста ясно, что речь идет о строках). Аналогично вектор-столбец из нулей длины n будем обозначать $0_{n,1}$ (или просто 0_n).

Перейдем к описанию *действий над матрицами*.

Транспонирование матрицы A определяется как действие, в результате которого из A получается новая матрица A^T , строками которой служат столбцы матрицы A , а столбцами — строки матрицы A при сохранении их порядка. Таким образом, первая строка матрицы A становится первым столбцом матрицы A^T , вторая строка матрицы A — вторым столбцом матрицы A^T и т. д., так что (i, j) -й элемент a_{ij} матрицы A становится (j, i) -м элементом матрицы A^T . Так, при построении регрессионных моделей мы оперировали как с $n \times (p + 1)$ -матрицей наблюдений объясняющих переменных

$$X = \begin{pmatrix} 1 & x_1^{(1)} & x_1^{(2)} & \dots & x_1^{(p)} \\ 1 & x_2^{(1)} & x_2^{(2)} & \dots & x_2^{(p)} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_n^{(1)} & x_n^{(2)} & \dots & x_n^{(p)} \end{pmatrix}, \quad (\text{П2.4})$$

так и с *транспонированной* $(p + 1) \times n$ -матрицей

$$X^T = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1^{(1)} & x_2^{(1)} & \dots & x_n^{(1)} \\ \dots & \dots & \dots & \dots \\ x_1^{(p)} & x_2^{(p)} & \dots & x_n^{(p)} \end{pmatrix}. \quad (\text{П2.5})$$

Заметим, что из определений операции транспонирования и свойства симметричности матрицы следует непосредственно, что *транспонирование не меняет симметрическую матрицу*, то есть если A — симмет-

рична, то $A^T = A$ и обратно: если $A^T = A$, то матрица A квадратна и симметрична.

Очевидно также, что последовательное двукратное применение к матрице A операции транспонирования приводит к *исходной* матрице A , то есть $(A^T)^T = A$.

Сложение двух матриц. Если A и B — матрицы одной размерности, то можно определить новую матрицу $C = A + B$, которая будет иметь ту же размерность, что и A и B , а ее элементы c_{ij} определяются для всех i и j соотношениями $c_{ij} = a_{ij} + b_{ij}$. Заметим, что из определений операций транспонирования и сложения непосредственно следует, что

$$(A + B)^T = A^T + B^T. \quad (\text{П2.6})$$

Умножение матрицы на число. Произведение матрицы A на число λ определяется как

$$\lambda A = (\lambda a_{ij})_{\substack{i=1,2,\dots,m; \\ j=1,2,\dots,n}},$$

то есть каждый элемент матрицы A умножается на это число.

Произведение матриц. Если число столбцов $n \times m$ -матрицы A равно числу строк $m \times k$ -матрицы B , то может быть определена операция умножения AB матрицы A на матрицу B (при этом говорят, что «матрица A слева умножается на матрицу B »). Элементы c_{ij} такого произведения $C = AB$ определяются для всех $i = 1, 2, \dots, n$ и $j = 1, 2, \dots, k$ соотношениями

$$c_{ij} = \sum_{l=1}^m a_{il}b_{lj}. \quad (\text{П2.7})$$

Таким образом, (i, j) -й элемент произведения C вычисляется как *скалярное произведение* i -й строки матрицы A и j -го столбца матрицы B , то есть как сумма попарных произведений элементов i -й строки первой матрицы на соответствующие (то есть стоящие на тех же по порядку местах) элементы j -го столбца второй матрицы. При этом автоматически определяется размерность матрицы-произведения C : она будет иметь столько же строк (n), сколько имел первый сомножитель, и столько же столбцов (k), сколько их имел второй сомножитель.

Заметим, что при перестановке сомножителей произведение BA может просто не существовать, но даже если оно существует (как это и бывает в случае, если матрицы A и B — квадратные и одной размерности или если сомножители имеют размерности $n \times m$ и $m \times n$), то, *вообще говоря, коммутативный закон для умножения матриц не имеет места, то есть*

$$AB \neq BA. \quad (\text{П2.8})$$

Отметим еще несколько полезных для эконометрических приложений свойств произведения матриц.

Ассоциативный закон:

$$ABC = (AB)C = A(BC). \quad (\text{П2.9})$$

Дистрибутивный закон:

$$A(B + C) = AB + AC, \quad (B + C)A = BA + CA. \quad (\text{П2.10})$$

Умножение на единичную матрицу: для любой $n \times m$ -матрицы A имеют место тождества

$$AI_m = I_n A = A. \quad (\text{П2.11})$$

Транспонирование произведения матриц:

$$(A_1 A_2 \dots A_k)^T = A_k^T A_{k-1}^T \dots A_2^T A_1^T \quad (\text{П2.12})$$

(доказывается индукцией: непосредственным сравнением элементов матриц $(A_1 A_2)^T$ и $A_2^T A_1^T$ доказывается справедливость (П2.12) для случая $k = 2$ и т. д.).

Произведения вида AA^T и $A^T A$ играют заметную роль в эконометрических построениях. Так, произведение $X^T X$, в котором матрица X определена соотношением (П2.4), является непрременным участником всех основных формул классического метода наименьших квадратов. Заметим, что если A — матрица размерности $n \times m$, то произведение AA^T будет иметь размерность $n \times n$, в то время как произведение $A^T A$ — это матрица размерности $m \times m$. Но в любом случае *произведения вида AA^T и $A^T A$ всегда являются квадратными симметрическими матрицами.*

Обратим внимание на специальный случай, когда $X = (x_1, x_2, \dots, x_n)^T$ — это вектор-столбец, состоящий из n элементов (или *вектор-столбец длины n*). Тогда

$$X^T X = \sum_{i=1}^n x_i^2 \quad (\text{П2.13})$$

— это 1×1 -матрица, то есть *число*, а

$$XX^T = \begin{pmatrix} x_1^2 & x_1 x_2 & \dots & x_1 x_n \\ x_2 x_1 & x_2^2 & \dots & x_2 x_n \\ \dots & \dots & \dots & \dots \\ x_n x_1 & x_n x_2 & \dots & x_n^2 \end{pmatrix} \quad (\text{П2.14})$$

— это *матрица* размерности $n \times n$.

Матрицы типа (П2.13) и (П2.14) играют заметную роль в регрессионном анализе и в системах одновременных уравнений. Действительно, если в качестве компонент x_i вектора X рассмотреть регрессионные «невязки» $y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i^{(1)} - \dots - \hat{\theta}_p x_i^{(p)}$ ($i = 1, 2, \dots, n$), то произведение

(П2.13) даст нам сумму квадратов «невязок», которая играет важную роль в анализе точности регрессионной модели. Если же в качестве компонент x_i вектора X рассмотреть отклонения i -й объясняющей переменной $x^{(i)}$ от своего среднего значения $a^{(i)}$ ($i = 1, 2, \dots, p$), то произведение (П2.14) после применения к нему операции усреднения (математического ожидания E) даст $p \times p$ -ковариационную матрицу объясняющих переменных Σ_X .

Прямое (или кронекерово) произведение $n \times m$ -матрицы A и $k \times l$ -матрицы B обозначается $A \otimes B$, имеет размерность $nk \times ml$ и подсчитывается по формуле

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1m}B \\ a_{21}B & a_{22}B & \dots & a_{2m}B \\ \dots & \dots & \dots & \dots \\ a_{n1}B & a_{n2}B & \dots & a_{nm}B \end{pmatrix}, \quad (\text{П2.15})$$

где правая часть представляет собой матрицу, *составленную из блоков вида $a_{ij}B$* . Каждый такой блок сам является матрицей и, как легко видеть, имеет размерность $k \times l$. Повторенные n раз по строкам, они дают общее количество строк, равное nk , а повторенные m раз по столбцам, они дают общее количество столбцов, равное ml (подробнее о *матрицах блочного типа*, или о *составленных матрицах*, так же, как и свойствах прямого произведения, см. п. П2.8).

Отметим, что некоторые эконометрические построения (в первую очередь, связанные с вычислительными проблемами *систем регрессионных уравнений*) значительно упрощаются, если использовать понятие прямого (кронекерова) произведения.

Матричное дифференцирование. Достаточно полное изложение аппарата векторного и матричного дифференциального исчисления читатель найдет в книге [Магнус, Нейдеккер (2002)]. Мы же остановимся лишь на тех определениях и результатах, которые непосредственно использованы в нашем учебнике.

Пусть $\Theta = (\theta_1, \theta_2, \dots, \theta_m)^\top$ — $m \times 1$ -матрица (то есть вектор-столбец длины m), компоненты которой играют роль неизвестных параметров эконометрической модели, а $A(\Theta) = (a_1(\Theta), a_2(\Theta), \dots, a_n(\Theta))^\top$ — $n \times 1$ -матрица (то есть вектор-столбец длины n), компоненты которой интерпретируются как некоторые характеристики этой модели, зависящие от Θ .

Производной $n \times 1$ -векторной функции $A(\Theta)$ по $m \times 1$ -векторному аргументу Θ называется $n \times m$ -матрица

$$\frac{\partial A(\Theta)}{\partial \Theta} = \begin{pmatrix} \frac{\partial a_1(\Theta)}{\partial \theta_1} & \frac{\partial a_1(\Theta)}{\partial \theta_2} & \dots & \frac{\partial a_1(\Theta)}{\partial \theta_m} \\ \frac{\partial a_2(\Theta)}{\partial \theta_1} & \frac{\partial a_2(\Theta)}{\partial \theta_2} & \dots & \frac{\partial a_2(\Theta)}{\partial \theta_m} \\ \dots & \dots & \dots & \dots \\ \frac{\partial a_n(\Theta)}{\partial \theta_1} & \frac{\partial a_n(\Theta)}{\partial \theta_2} & \dots & \frac{\partial a_n(\Theta)}{\partial \theta_m} \end{pmatrix}. \quad (\text{П2.16})$$

Матрицы вида (П2.16) используются, в частности, в качестве матриц преобразований («якобианов») в теории преобразований случайных величин (см., например, п. 4.4 в [Айвазян, Мхитарян (2001)]).

Рассмотрим важные для эконометрических приложений частные случаи функции $A(\Theta)$:

$$1^0. \quad A(\Theta) = \Theta^T \mathbf{B} \Theta, \text{ где } \mathbf{B} \text{ — квадратная матрица размерности } m \times m. \\ \text{Тогда } \frac{\partial A(\Theta)}{\partial \Theta} = \frac{\partial (\Theta^T \mathbf{B} \Theta)}{\partial \Theta} = (\mathbf{B} + \mathbf{B}^T) \Theta.$$

$$2^0. \quad A(\Theta) = \Theta^T Z, \text{ где } Z \text{ — вектор-столбец длины } m. \\ \text{Тогда } \frac{\partial A(\Theta)}{\partial \Theta} = \frac{\partial (\Theta^T Z)}{\partial \Theta} = Z.$$

П2.2. Основные числовые характеристики квадратной матрицы: определитель, след

Любой квадратной матрице $\mathbf{A}$ можно сопоставить некоторый набор ее числовых характеристик. В данном пункте остановимся на двух из них. Одна из таких характеристик называется *определителем* (детерминантом) матрицы $\mathbf{A}$ и обозначается $\det \mathbf{A}$ или $|\mathbf{A}|$, а другая определяется как *след матрицы* $\mathbf{A}$ и обозначается $\text{tr}(\mathbf{A})$.

Определитель (детерминант) $n \times n$ -матрицы $\mathbf{A}$ вычисляется по формуле

$$\det \mathbf{A} = \sum_{j_1=1}^n \sum_{j_2=1}^n \dots \sum_{j_n=1}^n (-1)^{\nu(j_1, j_2, \dots, j_n)} a_{1j_1} a_{2j_2} \dots a_{nj_n}, \quad (\text{П2.17})$$

где суммирование ведется по всем возможным комбинациям *различных* столбцов (то есть по всем возможным перестановкам вторых индексов), а $\nu(j_1, j_2, \dots, j_n)$ — это минимальное число инверсий (то есть парных обменов местами), которое надо совершить с элементами *исходной* перестановки $(1, 2, \dots, n)$, чтобы получить перестановку $(j_1, j_2, \dots, j_n)$. Очевидно, общее число слагаемых в правой части (П2.17) составит при таком определении n -факториал $(n!)$.

Для малых размерностей матриц это определение приводит, в частности, к следующим результатам:

$$\text{а) } n = 2: \quad \det \mathbf{A} = \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = a_{11}a_{22} - a_{12}a_{21};$$

$$\text{б) } n = 3: \quad \det \mathbf{A} = \det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \\ = a_{11}a_{22}a_{33} + a_{13}a_{21}a_{32} + a_{12}a_{23}a_{31} - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33}.$$

Приведем здесь важнейшие свойства определителя:

$$1) \det(\mathbf{A} \mathbf{B}) = \det(\mathbf{B} \mathbf{A}) = \det \mathbf{A} \cdot \det \mathbf{B};$$

- 2) $\det(\lambda A) = \lambda^n \det A$ (λ — число, n — размерность матрицы A);
- 3) $\det[\operatorname{diag}(a_{11}, a_{22}, \dots, a_{nn})] = a_{11}a_{22} \cdot \dots \cdot a_{nn}$;
- 4) $\det I_n = 1$;
- 5) $\det A^T = \det A$;
- 6) $\det A = 0$, если в матрице A есть две одинаковые строки (или два одинаковых столбца);
- 7) инверсия (обмен местами) двух строк (столбцов) матрицы A приводит к изменению знака ее определителя;
- 8) значение определителя матрицы A не изменится, если к любой его строке (столбцу) добавить линейную комбинацию других строк (столбцов).

Последнее свойство нуждается в пояснении. Обозначим $A_{i.} = (a_{i1}, a_{i2}, \dots, a_{in})$ — i -ю строку $n \times n$ матрицы A . Линейной комбинацией строк с номерами $i_1, i_2, \dots, i_k$ называется строка, определяемая в соответствии с правилами действий с матрицами по формуле

$$\lambda_1 A_{i_1.} + \lambda_2 A_{i_2.} + \dots + \lambda_k A_{i_k.},$$

где $\lambda_1, \lambda_2, \dots, \lambda_k$ — некоторые числа. Точно так же определяется линейная комбинация столбцов $A_{.j} = (a_{1j}, a_{2j}, \dots, a_{nj})^T$, $j = j_1, j_2, \dots, j_l$;

- 9) разложение определителя по элементам строки (или столбца).

$$\det A = \sum_{j=1}^n a_{ij} (-1)^{i+j} \det A_{ij},$$

где A_{ij} — $(n-1) \times (n-1)$ -матрица, получающаяся из матрицы A вычеркиванием из нее i -й строки и j -го столбца. Величина $\det A_{ij}$ называется *минором* матрицы A , а величина $(-1)^{i+j} \det A_{ij}$ — *алгебраическим дополнением элемента a_{ij}* в матрице A . Кстати, понятие алгебраического дополнения используется в учебнике, например, при вычислении коэффициента детерминации R^2 и частных коэффициентов корреляции по корреляционной матрице $\mathbf{R}$ исследуемого многомерного признака (см. п. 3.2).

Отметим, что если квадратная матрица имеет отличный от нуля определитель, то она называется **невырожденной**.

След квадратной $n \times n$ -матрицы A (обозначается $\operatorname{tr} A$, от английского слова *trace*) определяется как сумма ее диагональных элементов, то есть

$$\operatorname{tr} A = a_{11} + a_{22} + \dots + a_{nn}. \quad (\text{П2.18})$$

Основные свойства следа матрицы:

- 1) $\text{tr}(\mathbf{A} \mathbf{B}) = \text{tr}(\mathbf{B} \mathbf{A})$;
- 2) $\text{tr} \mathbf{I}_n = n$;
- 3) $\text{tr}(\lambda \mathbf{A}) = \lambda \text{tr} \mathbf{A}$, где λ — число;
- 4) $\text{tr} \mathbf{A}^\top = \text{tr} \mathbf{A}$;
- 5) $\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr} \mathbf{A} + \text{tr} \mathbf{B}$;
- 6) в качестве частного случая свойства 1) особо отметим ситуации, в которых роль $\mathbf{A}$ играет вектор-столбец $\mathbf{X} = (x_1, x_2, \dots, x_n)^\top$, а роль $\mathbf{B} = \mathbf{X}^\top$; и хотя ни для $\mathbf{X}$, ни для $\mathbf{X}^\top$ след не определен, их произведения $\mathbf{X} \mathbf{X}^\top$ и $\mathbf{X}^\top \mathbf{X}$ являются *квадратными*, соответственно, $n \times n$ - и 1×1 -матрицами и для них действует правило 1), то есть

$$\text{tr}(\mathbf{X} \mathbf{X}^\top) = \text{tr}(\mathbf{X}^\top \mathbf{X}),$$

где, обращаем внимание читателя, матрица $\mathbf{X}^\top \mathbf{X} = x_1^2 + x_2^2 + \dots + x_n^2$ — это *число*.

Заметим, что и определитель, и след матрицы имеют четкую вероятностную интерпретацию, например, когда в качестве $\mathbf{A}$ рассматривается *ковариационная матрица* Σ_ξ многомерной случайной величины $\xi = (\xi^{(1)}, \xi^{(2)}, \dots, \xi^{(p)})^\top$. В этом случае и $\text{tr} \Sigma_\xi$ и $\det \Sigma_\xi$ характеризуют *степень многомерного рассеяния* значений этой случайной величины, а $\det \Sigma_\xi$ называется *обобщенной дисперсией* ξ .

П2.3. Обратная матрица и ее свойства

В операциях с *числами* для любого *отличного от нуля* числа a существует число $a^{-1} = 1/a$, которое мы называем *обратным* и которое обладает тем характеристическим свойством, что $aa^{-1} = a^{-1}a = 1$. В матричной алгебре роль единицы, как мы видели, выполняет *единичная матрица* $\mathbf{I}_n$, поскольку при умножении на $\mathbf{I}_n$ любой квадратной матрицы размерности $n \times n$ справа и слева эта матрица не меняется (см. (П2.11)).

Поэтому по аналогии с алгеброй чисел определим:

- пусть $\mathbf{A}$ — *квадратная невырожденная* (то есть $\det \mathbf{A} \neq 0$) матрица; тогда матрица $\mathbf{A}^{-1}$ называется *обратной*, если $\mathbf{A} \mathbf{A}^{-1} = \mathbf{A}^{-1} \mathbf{A} = \mathbf{I}$.

Можно показать, что такое определение обратной матрицы $\mathbf{A}^{-1} = (a_{ij}^{\text{обр}})$ приводит к следующей формуле для вычисления ее элементов $a_{ij}^{\text{обр}}$:

где A_{ji} — как и прежде, матрица, получающаяся из матрицы A вычеркиванием из нее j -й строки и i -го столбца (то есть числитель правой части (П2.19) является алгебраическим дополнением элемента a_{ji} в исходной матрице A , или, что то же, — алгебраическим дополнением (i, j) -го элемента в транспонированной матрице A^T).

Пример П2.1. *Двухпродуктовая версия статической модели «затраты-выпуск» Леонтьева.* Пусть a_{ij} — затраты продукта i на выпуск единицы продукта j ($i, j = 1, 2$). И пусть x_i — общий выпуск продукта i и c_i — конечный спрос на этот продукт. Тогда уравнения, связывающие между собой введенные выше величины, будут иметь вид (в предположении, что нет ни потерь, ни излишков продуктов):

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + c_1 = x_1, \\ a_{21}x_1 + a_{22}x_2 + c_2 = x_2, \end{cases}$$

или, после приведения подобных членов,

$$\begin{cases} (1 - a_{11})x_1 - a_{12}x_2 = c_1, \\ -a_{21}x_1 + (1 - a_{22})x_2 = c_2. \end{cases}$$

Запишем эту систему, используя матричные обозначения:

$$AX = C, \quad (\text{П2.20})$$

где $X = (x_1, x_2)^T$, $C = (c_1, c_2)^T$, а

$$A = \begin{pmatrix} 1 - a_{11} & -a_{12} \\ -a_{21} & 1 - a_{22} \end{pmatrix}.$$

Мы хотим разрешить систему (П2.20) относительно x_1 и x_2 , то есть определить, какое общее количество первого и второго продукта должно быть произведено, чтобы обеспечить производственное потребление и конечный спрос. Чтобы уединить X в уравнении (П2.20), домножим обе части этого уравнения на матрицу A^{-1} слева:

$$X = A^{-1}C,$$

где в соответствии с (П2.19)

$$A^{-1} = \frac{1}{\det A} \begin{pmatrix} 1 - a_{22} & a_{12} \\ a_{21} & 1 - a_{11} \end{pmatrix}.$$

С учетом того, что $\det A = (1 - a_{11})(1 - a_{22}) - a_{12}a_{21}$, имеем:

$$\begin{aligned} x_1 &= \frac{1}{(1 - a_{11})(1 - a_{22}) - a_{12}a_{21}} [(1 - a_{22})c_1 + a_{12}c_2], \\ x_2 &= \frac{1}{(1 - a_{11})(1 - a_{22}) - a_{12}a_{21}} [a_{21}c_1 + (1 - a_{11})c_2]. \end{aligned}$$

Основные свойства обратной матрицы:

- 1) матрица A^{-1} для любой невырожденной матрицы A — единственна;
- 2) $\det A^{-1} = (\det A)^{-1}$;
- 3) $(A^T)^{-1} = (A^{-1})^T$;
- 4) $(A^{-1})^{-1} = A$;
- 5) $(AB)^{-1} = B^{-1}A^{-1}$,
 $(ABC)^{-1} = C^{-1}B^{-1}A^{-1}$ и т. д.
 (напомним, что все матрицы, участвовавшие в формулировке свойств обратной матрицы, квадратные и невырожденные).

Теперь мы можем дополнить перечень основных типов матриц, приведенный в п. П2.1. Введем в рассмотрение класс *ортогональных матриц*, определив:

- *квадратная невырожденная матрица A называется ортогональной, если $A^T = A^{-1}$.*

Из определения ортогональной матрицы непосредственно следует, в частности, что $A^T A = A A^T = I$. Нетрудно также вывести, что *определитель ортогональной матрицы всегда равен по абсолютной величине единице*, то есть $|\det A| = 1$. Действительно, поскольку $\det A = \det A^T$ (свойство 5 из п. П2.2), а $\det (AB) = \det A \cdot \det B$ (свойство 1 из п. П2.2), то для ортогональной матрицы $\det (A A^T) = \det A \det A^T = (\det A)^2 = \det I = 1$. Отсюда получаем, что $|\det A| = 1$, если A ортогональна.

Мы еще будем обращаться к ортогональным матрицам в пп. П2.5 и П2.6.

П2.4. Ранг матрицы и линейная зависимость ее строк (столбцов)

Наряду с рассмотренными в п. П2.2 двумя основными числовыми характеристиками *квадратной* матрицы — ее *определителем* и *следом*, в общем случае, включающем и *прямоугольные* матрицы, существует еще одна очень важная числовая характеристика матрицы — ее *ранг*.

Перед тем, как сформулировать строгое определение этого понятия, рассмотрим понятие *линейной зависимости строк* (столбцов) анализируемой $n \times m$ -матрицы A .

Строки $A_i = (a_{i1}, a_{i2}, \dots, a_{im})$, $i = 1, 2, \dots, n$ матрицы A называются линейно зависимыми, если существуют действительные числа $\lambda_1, \lambda_2, \dots, \lambda_n$, не все равные нулю и такие, что

$$\lambda_1 A_1 + \lambda_2 A_2 + \dots + \lambda_n A_n = 0_{1..m} \quad (\text{П2.21})$$

(здесь $0_{1,m}$, в соответствии с принятыми выше обозначениями — это строка длины m , состоящая из нулей).

Аналогично: столбцы $A_j = (a_{1j}, a_{2j}, \dots, a_{nj})^T$, $j = 1, 2, \dots, m$ матрицы A называются **линейно зависимыми**, если существуют действительные числа $\mu_1, \mu_2, \dots, \mu_m$, не все равные нулю и такие, что

$$\mu_1 A_{.1} + \mu_2 A_{.2} + \dots + \mu_m A_{.m} = 0_{n,1}. \quad (\text{П2.21}')$$

В противном случае строки (столбцы) называются **линейно независимыми**.

Можно доказать, что максимальное число линейно независимых строк $n \times m$ -матрицы A совпадает с максимальным числом ее линейно независимых столбцов и, одновременно, — с максимальным порядком ее не равного нулю минора (напомним, что минором порядка k матрицы A называется определитель $k \times k$ -матрицы, получающейся из матрицы A вычеркиванием из нее $n - k$ строк и $m - k$ столбцов).

- **Ранг** $n \times m$ -матрицы A (будем обозначать его **ранг** A) определяется (при $n \geq m$) как максимальное число ее линейно независимых столбцов.

Очевидно, что в силу приведенного выше свойства, ранг матрицы A может быть определен и как максимальное число ее линейно независимых строк, и как максимальный порядок ее отличного от нуля минора. Кстати, последнее определение часто бывает наиболее удобным с точки зрения возможности *практического вычисления* ранга конкретной матрицы. При этом, определяя ранг матрицы как максимальный порядок ее отличного от нуля минора, подразумевают, что достаточно того, чтобы нашелся *хотя бы один* ненулевой минор порядка k , в то время как все миноры порядка $k + 1$ уже будут равны нулю. Так, например, в 4×3 -матрице

$$A = \begin{pmatrix} 4 & 8 & 2 \\ 2 & 4 & 4 \\ 2 & 4 & -2 \\ 3 & 6 & -3 \end{pmatrix}$$

все четыре минора 3-го порядка равны нулю. Следовательно, **ранг** $A < 3$. И хотя большинство миноров 2-го порядка тоже равно нулю, все-таки существуют миноры этого порядка, отличные от нуля. А это значит, что **ранг** $A = 2$.

Заметим, что применительно к матрице X наблюдаемых значений объясняющих переменных (см. (П2.4)) в регрессионном анализе *линейная зависимость столбцов означает линейную зависимость объясняющих переменных*.

Из определения ранга матрицы более или менее непосредственно вытекают следующие его основные свойства:

- 1) $\text{ранг } A \leq \min(n, m)$; если $\text{ранг } A = \min(n, m)$, то говорят, что матрица A — это матрица *полного ранга*;
- 2) $\text{ранг } I_n = n$;
- 3) $\text{ранг}(\text{diag}(a_{11}, a_{22}, \dots, a_{nn})) = k$, где k — число ненулевых элементов в ряду $a_{11}, a_{22}, \dots, a_{nn}$;
- 4) $\text{ранг}(AB) \leq \min\{\text{ранг } A, \text{ранг } B\}$;
- 5) $\text{ранг } B = n$, если B — квадратная *невырожденная* матрица размерности $n \times n$;
- 6) $\text{ранг}(AB) = \text{ранг } A$, если A — произвольная матрица размерности $n \times m$, а B — *невырожденная* $m \times m$ -матрица;
- 7) $\text{ранг}(BA) = \text{ранг } A$, если A — произвольная матрица размерности $n \times m$, а B — *невырожденная* $n \times n$ -матрица;
- 8) $\text{ранг}(B_1 A B_2) = \text{ранг } A$, где A — произвольная матрица порядка $n \times m$, а B_1 и B_2 — невырожденные квадратные матрицы размерностей, соответственно, $(n \times n)$ и $(m \times m)$;
- 9) $\text{ранг } A = \text{ранг } A^T$.

Отметим также один важный для эконометрических приложений (особенно для проблемы идентифицируемости модели) факт, связанный с понятием ранга матрицы. Пусть $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$ — набор переменных (среди которых, в целях большей общности, мы допускаем присутствие и переменной, тождественно равной единице), а A — матрица размерности $n \times p$. Система уравнений относительно X может быть записана в виде

$$AX = 0_{n,1}. \quad (\text{П2.22})$$

При решении системы (П2.22) важным моментом является соотношение между числом неизвестных и числом линейно независимых уравнений, содержащихся в системе. Так вот, оказывается, что *число линейно независимых уравнений в системе (П2.22) равно рангу матрицы A* .

П2.5. Характеристические (собственные) числа квадратной матрицы и соответствующие им собственные векторы

Широкий круг задач многомерного статистического анализа и эконометрики сводится (в вычислительном плане) к необходимости анализа и решения параметрического семейства систем уравнений типа

$$(A - \lambda I_n)X = 0_{n,1}, \quad (\text{П2.23})$$

где $\mathbf{A}$ — некоторая $n \times n$ -матрица, $X = (x_1, x_2, \dots, x_n)^T$ — вектор-столбец неизвестных, а λ — некоторый числовой параметр. Для того, чтобы существовало нетривиальное (отличное от нулевого) решение, необходимо, чтобы матрица $\mathbf{A} - \lambda \mathbf{I}_n$ была вырожденной, то есть необходимо потребовать, чтобы

$$\det(\mathbf{A} - \lambda \mathbf{I}_n) = 0. \quad (\text{П2.24})$$

Из правил вычисления определителя матрицы (см. (П2.17)) следует, что левая часть (П2.24) представляет собой алгебраический полином от λ степени n , так что соотношение (П2.24) — это алгебраическое уравнение степени n относительно λ . Само это уравнение, а следовательно, и его корни $\lambda_1, \lambda_2, \dots, \lambda_n$ по построению полностью определяются элементами матрицы $\mathbf{A}$. Приходим к определению:

- **Характеристическими (собственными) числами $n \times n$ -матрицы $\mathbf{A}$ называются корни характеристического уравнения (П2.24).**

Беря любое из собственных чисел λ_i и подставляя его в исходное соотношение (П2.23), мы получаем уже конкретную систему уравнений (относительно X) вида

$$(\mathbf{A} - \lambda_i \mathbf{I}_n) X = \mathbf{0}_{n,1}. \quad (\text{П2.24}')$$

Существование нетривиального решения $X(i)$ этого уравнения обеспечивается равенством нулю определителя $\det(\mathbf{A} - \lambda_i \mathbf{I}_n)$. Соответственно, приходим к еще одному определению:

- **Вектор-столбец $X(i) = (x_1^{(i)}, x_2^{(i)}, \dots, x_n^{(i)})^T$, являющийся решением уравнения (П2.24'), называется характеристическим (собственным) вектором матрицы $\mathbf{A}$, соответствующим характеристическому (собственному) числу λ_i .**

А поскольку алгебраическое уравнение степени n имеет n корней (среди которых, вообще говоря, могут быть совпадающие и комплексные), то всякая квадратная $n \times n$ -матрица $\mathbf{A}$ имеет n собственных чисел (не обязательно различных) и n соответствующих им собственных векторов.

В дальнейшем, отправляясь от интересов прикладной статистики и эконометрики, мы сосредоточим свое внимание на свойствах характеристических чисел и характеристических векторов только действительных симметрических положительно (или неотрицательно) определенных матриц $\mathbf{A}$ (будем для краткости их называть далее, соответственно, спо- и сно-матрицами)¹.

¹ Симметрическая $n \times n$ -матрица $\mathbf{A}$ называется положительно (неотрицательно)

1) Всякая спо- (сно-) $n \times n$ -матрица имеет n положительных (неотрицательных) действительных характеристических чисел; и наоборот, матрица, все характеристические числа которой положительны (неотрицательны) является спо- (сно-) матрицей.

2) Собственные векторы $X(i)$ и $X(j)$, соответствующие разным собственным числам, всегда *взаимно ортогональны*, то есть

$$X^T(i)X(j) = 0 \quad \text{при} \quad \lambda_i \neq \lambda_j.$$

Поскольку собственный вектор определяется с точностью до коэффициента пропорциональности, то можно нормировать собственные векторы $X(1), X(2), \dots, X(n)$ так, что все они будут образовывать *ортонормированную систему*, то есть

$$X^T(i)X(j) = \begin{cases} 1 & \text{при } i = j, \\ 0 & \text{при } i \neq j \end{cases} \quad \text{для всех } i, j = 1, 2, \dots, n.$$

Заметим, что $n \times n$ -матрица A , составленная из столбцов $X(i)$ ($i = 1, 2, \dots, n$), то есть матрица

$$X = (X(1) : X(2) : \dots : X(n))$$

будет *ортогональной*, то есть $X^T X = I_n$.

3) Всякая спо- и сно- матрица A может быть приведена с помощью ортогонального преобразования X к диагональному виду, в котором на диагонали будут стоять характеристические числа, то есть

$$X^T A X = \text{diag} (\lambda_1, \lambda_2, \dots, \lambda_n).$$

П2.6. Некоторые свойства симметричных положительно (неотрицательно) определенных матриц

На протяжении всего данного пункта мы будем полагать (если специально не оговорено другое), что A — *симметрическая положительно определенная матрица размерности n* , то есть для любого ненулевого вектора $X = (x_1, x_2, \dots, x_n)^T$ выполняется неравенство

$$X^T A X > 0. \quad (\text{П2.25})$$

1) При любой $m \times n$ -матрице B матрица $B^T B$ будет всегда симметрична и неотрицательно определена. Действительно, для любой $(n \times 1)$ -матрицы X (то есть для вектора-столбца длины n)

но) *определенной, если для любого ненулевого вектора $X = (x_1, x_2, \dots, x_n)^T$ с действительными значениями x_i выполняется неравенство $X^T A X > 0$ ($X^T A X \geq 0$).* Более подробно о симметрических положительно и неотрицательно определенных матрицах см. в п. П2.6.

$$X^T (B^T B) X = (B X)^T B X = Y^T Y = \sum_{i=1}^n y_i^2 \geq 0,$$

так как произведение вектора-строки $Y^T = (B X)^T$ на вектор-столбец $Y = B X$ приводит, очевидно, к неотрицательному числу.

2) Если $n \geq m$ и $n \times m$ -матрица B — матрица полного ранга (то есть $\text{rang } B = m$), то матрица $B^T A B$ — положительно определенная; в частном случае $n = m$ B является невырожденной квадратной матрицей размерности $n \times n$.

3) Матрица A^{-1} , обратная к A , также является симметрической и положительно определенной.

4) Единичная матрица I_n является симметрической положительно определенной (так как для любого вектора-столбца X длины n $X^T I_n X = X^T X = \sum_{i=1}^n x_i^2 > 0$).

5) $\det A > 0$, то есть определитель любой спо-матрицы положителен; отсюда, в частности, следует, что и все *главные миноры* матрицы A положительны (минор матрицы называется *главным*, если он является определителем подматрицы, образованной из исходной матрицы вычеркиванием из нее строк и столбцов, имеющих *одинаковые* номера).

6) След матрицы A может быть вычислен как сумма ее собственных значений, то есть

$$\text{tr} A = \lambda_1 + \lambda_2 + \dots + \lambda_n.$$

где, напомним, A — спо-матрица размерности $n \times n$, а $\lambda_1, \lambda_2, \dots, \lambda_n$ — ее собственные числа.

7) Всякая спо $(m \times m)$ -матрица A может быть представлена в виде

$$A = C C^T, \quad (\text{П2.25a})$$

где C некоторая невырожденная $(m \times m)$ -матрица.

Действительно, известно (см., например, п. 13.2.3 в [Айвазян, Мхитарян (2001)]), что любая спо-матрица A может быть приведена к диагональному виду при помощи некоторого ортогонального преобразования L , то есть

$$L A L^T = \Lambda, \quad (\text{П2.25б})$$

где

$$\Lambda = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_m \end{pmatrix},$$

а $\lambda_1, \lambda_2, \dots, \lambda_m$ — собственные значения (*действительные и положительные*) матрицы A . Домножая обе части соотношения (П2.25б) слева

на L^T и справа на L , получаем

$$A = L^T \Lambda L. \quad (\text{П2.25в})$$

Представляя матрицу Λ в виде $\Lambda = \Lambda^{1/2} \Lambda^{1/2}$, где

$$\Lambda^{1/2} = \begin{pmatrix} \sqrt{\lambda_1} & & & 0 \\ & \sqrt{\lambda_2} & & \\ & & \ddots & \\ 0 & & & \sqrt{\lambda_m} \end{pmatrix},$$

и обозначая $C = L^T \Lambda^{1/2}$, получаем из (П2.25в) доказательство возможности представления матрицы A в виде (П2.25а).

Пример П2.2. Продемонстрируем использование свойств симметрических положительно определенных матриц при анализе многомерных распределений и их статистик. Пусть $X = (x_1, x_2, \dots, x_n)^T$ — последовательность одинаково $(0, \sigma^2)$ -нормально распределенных случайных величин, то есть

$$\begin{aligned} E x_i &= 0, \quad i = 1, 2, \dots, n; \\ D x_i &= E x_i^2 = \sigma^2, \quad i = 1, 2, \dots, n; \\ \text{cov}(x_i, x_j) &= E(x_i x_j) = 0 \quad \text{при } i \neq j, \end{aligned}$$

или, в компактной записи, $X \in N_n(0_{n,1}; \sigma^2 I_n)$. И пусть в процессе анализа модели нам пришлось заниматься исследованием квадратичной формы от анализируемых случайных величин вида $X^T A X$, где A , как и ранее, споматрица.

В соответствии со свойством 3) споматриц (см. п. П2.5) существует такая ортогональная матрица C , с помощью которой матрица A может быть приведена к диагональному виду, а именно:

$$C^T A C = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n),$$

где $\lambda_1, \lambda_2, \dots, \lambda_n$ — собственные числа матрицы A , а элементы матрицы C полностью определяются собственными векторами той же матрицы A . Введем вспомогательные переменные

$$Y = C^T X \quad (\text{соответственно, } X = C Y) \quad (\text{П2.26})$$

и посмотрим, как будет выражаться анализируемая квадратичная форма $X^T A X$ в терминах этих новых переменных:

$$\begin{aligned} X^T A X &= (C Y)^T A C Y = Y^T (C^T A C) Y = \\ &= Y^T \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n) Y = \sum_{i=1}^n \lambda_i y_i^2. \end{aligned} \quad (\text{П2.27})$$

Следовательно, для того чтобы понять, как «ведет себя» квадратичная форма $X^T A X$, необходимо выяснить, какому закону распределения вероятностей подчиняются переменные $y_1, y_2, \dots, y_n$.

Из (П2.26) следует, что Y , будучи линейной функцией от X , также подчиняется нормальному закону со средним значением

$$EY = E(C^T X) = C^T EX = 0_{n,1}$$

и с ковариационной матрицей

$$\begin{aligned} \Sigma_Y &= E(YY^T) = E(C^T X X^T C) = C^T E(X X^T) C = \\ &= C^T \Sigma_X C = C^T (\sigma^2 I_n) C = \sigma^2 C^T C = \sigma^2 I_n \end{aligned}$$

(в этой выкладке мы воспользовались тем, что ковариационная матрица Σ_X вектора X равна по условию $\sigma^2 I_n$, а также — ортогональностью матрицы C , из чего следует $C^T C = I_n$).

Таким образом, мы убедились в том, что вспомогательные переменные $y_1, y_2, \dots, y_n$ так же, как и исходные случайные величины $x_1, x_2, \dots, x_n$, $(0, \sigma^2)$ -нормальны и взаимно независимы. А это позволяет описать распределение интересующей нас квадратичной формы (П2.27) законом суммы масштабированных $\chi^2(1)$ — распределенных случайных величин.

П2.7. Идемпотентные матрицы

Особенно важны в прикладной статистике и эконометрике соотношения типа (П2.27), в которых матрица A относится к классу так называемых *идемпотентных матриц*.

- Идемпотентной называется симметрическая² матрица A , обладающая тем свойством, что она совпадает со своим квадратом, то есть

$$A = A^2. \quad (\text{П2.28})$$

Перечислим основные свойства идемпотентных матриц.

1) Любая целая положительная степень k идемпотентной матрицы A дает снова исходную матрицу A , то есть $A^k = A$ (следует непосредственно из определения).

2) Собственные числа идемпотентной матрицы могут принимать только одно из двух значений: 0 или 1; а следовательно, в соответствии со свойством 1 из п. П2.5 все идемпотентные матрицы неотрицательно определены.

Докажем справедливость этого утверждения. В соответствии с характеристическим уравнением (П2.23) имеем:

$$A X = \lambda X.$$

²Иногда требование симметричности не включают в определение идемпотентной матрицы.

После домножения обеих частей этого характеристического уравнения слева на матрицу A получаем:

$$A^2 X = \lambda (AX) = \lambda (\lambda X) = \lambda^2 X. \quad (\text{П2.29})$$

Но одновременно с этим, используя идемпотентность матрицы A , мы можем записать:

$$A^2 X = A X = \lambda X. \quad (\text{П2.29}')$$

Сравнивая правые части (П2.29) и (П2.29'), получаем

$$(\lambda^2 - \lambda)X = 0_{n,1},$$

а так как $X \neq 0_{n,1}$, то

$$\lambda^2 - \lambda = 0,$$

то есть либо $\lambda = 0$, либо $\lambda = 1$, что и доказывает свойство 2).

3) Ранг идемпотентной матрицы равен числу ее ненулевых собственных чисел, что совпадает со следом исходной матрицы.

4) Если ранг $n \times n$ -идемпотентной матрицы A равен $n - k$, то использование ее собственных векторов для преобразований типа (П2.26) будет приводить квадратичную форму $X^T A X$ относительно n независимых $(0, \sigma^2)$ -нормально распределенных случайных величин к виду

$$X^T A X = y_1^2 + y_2^2 + \dots + y_{n-k}^2, \quad (\text{П2.30})$$

где элементы $y_1, y_2, \dots, y_{n-k}$ также взаимно независимы и $(0, \sigma^2)$ -нормально распределены. А это значит, что квадратичная форма $X^T A X$ подчиняется $\sigma \chi^2(n - k)$ -распределению.

П2.8. Матрицы блочной структуры

Содержание анализируемой задачи зачастую обуславливает целесообразность разбиения матриц, которыми мы оперируем, на отдельные блоки. Так, например, при анализе модели многомерной регрессии или систем одновременных уравнений в рамки нашего анализа попадают m зависимых (эндогенных) переменных $Y = (y^{(1)}, y^{(2)}, \dots, y^{(m)})^T$ и p объясняющих (экзогенных, предопределенных) переменных $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$. Важным объектом анализа является при этом *ковариационная матрица Σ_Z всего вектора переменных $Z = \begin{pmatrix} Y \\ X \end{pmatrix}$* , которая будет состоять, соответственно, из четырех блоков Σ_{YY} , Σ_{YX} , Σ_{XY} и Σ_{XX} , а именно:

$$\Sigma_Z = \begin{pmatrix} \Sigma_{YY} & \vdots & \Sigma_{YX} \\ \dots & \dots & \dots \\ \Sigma_{XY} & \vdots & \Sigma_{XX} \end{pmatrix}. \quad (\text{П2.31})$$

Очевидно, матрица Σ_Z будет иметь размерность $(m + p) \times (m + p)$, в то время как блок Σ_{YY} (матрица ковариаций между компонентами вектора Y) имеет размерность $m \times m$, блок Σ_{YX} (матрица «перекрестных»

ковариаций эндогенных переменных с предопределенными) имеет размерность $m \times p$, блок Σ_{xy} (матрица «перекрестных» ковариаций предопределенных переменных с эндогенными) имеет размерность $p \times m$, и, наконец, блок Σ_{xx} (матрица ковариаций между компонентами вектора X) имеет размерность $p \times p$.

Заметим, что для того, чтобы пользоваться некоторыми общими правилами оперирования с блочными матрицами, следует соблюдать *необходимые* требования при их разбиении, а именно:

- блоки, стоящие друг *над* (или *под*) другом по *вертикали*, должны иметь *одинаковое количество столбцов*; а блоки, стоящие *рядом* (слева или справа друг от друга) по *горизонтали*, должны иметь *одинаковое количество строк*.

Сложение блочных матриц. Если две матрицы A и B одинаковой размерности разбиты на блоки, соответственно, A_{ij} и B_{ij} ($i = 1, 2, \dots, k_1$; $j = 1, 2, \dots, k_2$) *одинаковым образом* (то есть размерности блоков A_{ij} и B_{ij} совпадают), то их сложение производится по правилу

$$A + B = \begin{pmatrix} A_{11} + B_{11} & \vdots & A_{12} + B_{12} & \vdots & \dots & \vdots & A_{1k_2} + B_{1k_2} \\ A_{21} + B_{21} & \vdots & A_{22} + B_{22} & \vdots & \dots & \vdots & A_{2k_2} + B_{2k_2} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ A_{k_11} + B_{k_11} & \vdots & A_{k_12} + B_{k_12} & \vdots & \dots & \vdots & A_{k_1k_2} + B_{k_1k_2} \end{pmatrix}.$$

Умножение блочных матриц формально также производится по правилам умножения *обычных* матриц, однако разбиение A и B на блоки A_{ij} и B_{ij} должно быть произведено таким образом, чтобы количество столбцов в первом сомножителе совпадало бы с количеством строк во втором сомножителе. Например:

$$\begin{aligned} AB &= \begin{pmatrix} A_{11} & \vdots & A_{12} \\ \dots & \dots & \dots \\ A_{21} & \vdots & A_{22} \end{pmatrix} \begin{pmatrix} B_{11} & \vdots & B_{12} \\ \dots & \dots & \dots \\ B_{21} & \vdots & B_{22} \end{pmatrix} = \\ &= \begin{pmatrix} A_{11}B_{11} + A_{12}B_{21} & \vdots & A_{11}B_{12} + A_{12}B_{22} \\ \dots & \dots & \dots \\ A_{21}B_{11} + A_{22}B_{21} & \vdots & A_{21}B_{12} + A_{22}B_{22} \end{pmatrix}. \end{aligned}$$

Определитель блочной матрицы. Пусть матрица A разбита на 4 блока

$$A = \begin{pmatrix} A_{11} & \vdots & A_{12} \\ \dots & \dots & \dots \\ A_{21} & \vdots & A_{22} \end{pmatrix} \quad (\text{П2.32})$$

таким образом, что блоки A_{11} и A_{22} являются *квадратными* матрицами (как это было в нашем примере (П2.31)). Тогда можно воспользоваться следующей общей формулой подсчета определителя матрицы A , во многих случаях существенно упрощающей вычисления:

$$\det \begin{pmatrix} A_{11} & \vdots & A_{12} \\ \dots & \dots & \dots \\ A_{21} & \vdots & A_{22} \end{pmatrix} = \det A_{11} \cdot \det (A_{22} - A_{21} A_{11}^{-1} A_{12}) = \\ = \det A_{22} \cdot \det (A_{11} - A_{12} A_{22}^{-1} A_{21}).$$

В частном случае *блочной-диагональной* матрицы A (то есть когда матрицы A_{12} и A_{21} состоят из одних нулей) данная формула принимает вид

$$\det \begin{pmatrix} A_{11} & \vdots & 0 \\ \dots & \dots & \dots \\ 0 & \vdots & A_{22} \end{pmatrix} = \det A_{11} \cdot \det A_{22}.$$

Обращение блочной матрицы (П2.32) производится по формуле

$$\begin{pmatrix} A_{11} & \vdots & A_{12} \\ \dots & \dots & \dots \\ A_{21} & \vdots & A_{22} \end{pmatrix}^{-1} = \begin{pmatrix} A^{11} & \vdots & -A_{11}^{-1} A_{12} A^{22} \\ \dots & \dots & \dots \\ -A^{22} A_{21} A_{11}^{-1} & \vdots & A^{22} \end{pmatrix},$$

где

$$A^{22} = (A_{22} - A_{21} A_{11}^{-1} A_{12})^{-1} \quad \text{и} \quad A^{11} = (A_{11} - A_{12} A_{22}^{-1} A_{21})^{-1}.$$

Прямое (кронекерово) произведение $n \times m$ -матрицы A и $k \times l$ -матрицы B было введено в п. П2.1 (см. соотношение (П2.15)). Поскольку результат такого перемножения матриц A и B представляется *матрицей блочной структуры*, рассмотрим несколько полезных свойств прямого произведения:

1) если матрицы A и B квадратны и обратимы, то

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1};$$

2) если A — $n \times n$ -матрица, а B — $k \times k$ -матрица, то

$$\det (A \otimes B) = (\det A)^n (\det B)^k;$$

$$3) (A \otimes B)^T = (A^T \otimes B^T);$$

$$4) \operatorname{tr} (A \otimes B) = \operatorname{tr} A \cdot \operatorname{tr} B.$$

ПРИЛОЖЕНИЕ 3. МНОГОМЕРНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ

Из самого определения эконометрики (см. п. 1.3) следует, что *математико-статистический инструментарий* является одной из трех ее базисных составных частей (наряду с *экономической теорией* и *экономическими измерениями*). И если такие разделы этого математического инструментария как элементы теории вероятностей и *одномерной* математической статистики, дополненные стандартными сведениями из линейного регрессионного анализа, традиционно включаются в учебную эконометрическую литературу (чаще всего, в форме приложений), то ряд актуальнейших для современной эконометрики **методов многомерного статистического анализа** (МСА) остается, как правило, вне поля зрения авторов этих изданий. А мы видим (см. задачи 1–4 в п. 1.2, а также главы 6, 9), что без хорошего понимания *многомерного нормального закона и связанных с ним распределений* (в том числе — многомерных), без знания постановок задач и подходов к их решению в *дискриминантном, кластерном анализе, в методе главных компонент* добиться высокого профессионального уровня владения современными эконометрическими методами практически невозможно! Именно упомянутым разделам МСА и посвящено данное приложение. В то же время, мы решили обойтись в этом учебнике без приложений, посвященных элементам теории вероятностей и одномерной математической статистики, полагая нашего читателя уже подготовленным в данной области (или имеющим «под рукой» соответствующие руководства).

П3.1. Многомерный нормальный закон распределения вероятностей (з.р.в.) и его свойства

1. **Функция плотности** $\varphi_{\xi}(X; \mathbf{a}; \Sigma)$ распределения p -мерной нормальной случайной величины $\xi = (\xi^{(1)}, \dots, \xi^{(p)})^T$ в точке $X = (x^{(1)}, \dots, x^{(p)})^T$ задается соотношением

$$\varphi_{\xi}(X; \mathbf{a}, \Sigma) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(X-\mathbf{a})^T \Sigma^{-1}(X-\mathbf{a})}, \quad (\text{П3.1})$$

где $\mathbf{a} = (a^{(1)}, \dots, a^{(p)})^T$ — вектор средних значений (то есть $a^{(j)} = \mathbf{E}\xi^{(j)}$), $\Sigma = (\sigma_{ij})_{i,j=\overline{1,p}}$ — ковариационная матрица анализируемой случайной величины (то есть $\sigma_{ij} = \text{cov}(\xi^{(i)}, \xi^{(j)}) = \mathbf{E}[(\xi^{(i)} - a^{(i)})(\xi^{(j)} - a^{(j)})]$), а $|\Sigma|$ — определитель матрицы Σ .

Всякая ковариационная матрица *симметрична* (очевидно!) и *неотрицательно определена*. Чтобы убедиться в последнем, возьмем вектор $\lambda = (\lambda^{(1)}, \dots, \lambda^{(p)})^T$ с вещественнозначными компонентами и рассмотрим $\mathbf{E}(\lambda^T(\xi - \mathbf{a}))^2 = \mathbf{E}(\lambda^T(\xi - \mathbf{a})(\xi - \mathbf{a})^T \lambda) = \lambda^T \mathbf{E}[(\xi - \mathbf{a})(\xi - \mathbf{a})^T] \cdot \lambda = \lambda^T \Sigma \lambda \geq 0$, причем знак равенства достигается тогда и только тогда,

когда между компонентами анализируемой случайной величины ξ существует линейная функциональная зависимость. В подобных случаях говорят о **вырожденности** распределения случайной величины ξ . Мы будем рассматривать (если случайно не оговорено противное) *невырожденные* з.р.в., так что ковариационная матрица Σ в (П3.1) — положительно определена. А в этом случае $|\Sigma| \neq 0$ и, значит, существует обратная матрица Σ^{-1} (действительно, предположив что у положительно определенной матрицы Σ определитель $|\Sigma| = 0$, приходим к существованию нетривиального решения λ у системы уравнений $\Sigma\lambda = 0_p$; но тогда $\lambda^T \Sigma \lambda = 0$, что противоречит положительной определенности матрицы Σ).

Для обозначения p -мерного нормального з.р.в., имеющего вектор средних значений $a = (a^{(1)}, \dots, a^{(p)})$ и ковариационную матрицу $\Sigma = (\sigma_{ij})$ в книге используется сочетание символов $N_p(a; \Sigma)$, а утверждение «случайная величина ξ подчинена p -мерному нормальному з.р.в. с вектором средних значений a и ковариационной матрицей Σ » кодируется с помощью

$$\xi \in N_p(a; \Sigma).$$

2. Двумерный нормальный з.р.в. и его свойства. Рассмотрим специально случай $p = 2$. Выразим плотность (П3.1) в терминах средних $a^{(1)} = E\xi^{(1)}$ и $a^{(2)} = E\xi^{(2)}$, дисперсий $D\xi^{(1)} = \sigma_{11} = \sigma_1^2$ и $D\xi^{(2)} = \sigma_{22} = \sigma_2^2$ и коэффициента корреляции $r(\xi^{(1)}, \xi^{(2)}) = \sigma_{12}/\sqrt{\sigma_{11} \cdot \sigma_{22}} = \sigma_{12}/\sigma_1\sigma_2$:

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & r\sigma_1\sigma_2 \\ r\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}; \quad |\Sigma| = \sigma_1^2\sigma_2^2(1 - r^2),$$

$$\Sigma^{-1} = \frac{1}{1 - r^2} \begin{pmatrix} \frac{1}{\sigma_1^2} & -\frac{r}{\sigma_1\sigma_2} \\ -\frac{r}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix},$$

так что выражения правой части (П3.1) приводятся к виду:

$$\begin{aligned} & -\frac{1}{2}(X - a)^T \Sigma^{-1}(X - a) = \\ & = -\frac{1}{2}(x^{(1)} - a^{(1)}, x^{(2)} - a^{(2)}) \cdot \frac{1}{1 - r^2} \begin{pmatrix} \frac{1}{\sigma_1^2} & -\frac{r}{\sigma_1\sigma_2} \\ -\frac{r}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix} \begin{pmatrix} x^{(1)} - a^{(1)} \\ x^{(2)} - a^{(2)} \end{pmatrix} = \\ & = -\frac{1}{2(1 - r^2)} \left[\frac{(x^{(1)} - a^{(1)})^2}{\sigma_1^2} - 2r \frac{(x^{(1)} - a^{(1)})(x^{(2)} - a^{(2)})}{\sigma_1\sigma_2} + \right. \\ & \quad \left. + \frac{(x^{(2)} - a^{(2)})^2}{\sigma_2^2} \right]; \\ & \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1 - r^2}}. \end{aligned}$$

Подставляя эти выражения в правую часть (ПЗ.1), получаем:

$$\varphi_{\xi}(x^{(1)}, x^{(2)}; a^{(1)}, a^{(2)}; \sigma_1^2; \sigma_2^2; r) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \times \\ \times e^{-\frac{1}{2(1-r^2)} \left[\frac{(x^{(1)}-a^{(1)})^2}{\sigma_1^2} - 2r \frac{(x^{(1)}-a^{(1)})(x^{(2)}-a^{(2)})}{\sigma_1\sigma_2} + \frac{(x^{(2)}-a^{(2)})^2}{\sigma_2^2} \right]} \quad (\text{ПЗ.1}')$$

Анализ распределения (ПЗ.1') позволяет сформулировать следующие полезные свойства двумерного нормального закона.

С в о й с т в о 1. Частное распределение каждой из компонент $\xi^{(1)}$ и $\xi^{(2)}$ снова является нормальным, а именно: $\xi^{(j)} \in N_1(a^{(j)}; \sigma_j^2)$, — $j = 1, 2$. То есть:

$$\varphi_{\xi^{(j)}}(x^{(j)}; a^{(j)}; \sigma_j^2) = \frac{1}{\sqrt{2\pi}\sigma_j} e^{-\frac{1}{2\sigma_j^2}(x^{(j)}-a^{(j)})^2} \quad j = 1, 2. \quad (\text{ПЗ.2})$$

С в о й с т в о 2. Условное распределение случайной величины $\xi^{(2)}$ (при условии: $\xi^{(1)} = x^{(1)}$) является нормальным со средним значением

$$E(\xi^{(2)}|\xi^{(1)} = x) = a^{(2)} + r \frac{\sigma_2}{\sigma_1}(x^{(1)} - a^{(1)}) \quad (\text{ПЗ.3a})$$

и дисперсией

$$D(\xi^{(2)}|\xi^{(1)} = x) = \sigma_2^2(1 - r^2). \quad (\text{ПЗ.3б})$$

Подчеркнем два важных факта, следующих из свойства 2. Соотношение (ПЗ.3a) означает, что *функция регрессии переменной $\xi^{(2)}$ по $\xi^{(1)}$ является линейной* формой по $x^{(1)}$, коэффициенты определяются по параметрам двумерного нормального распределения $a^{(1)}, a^{(2)}, \sigma_1, \sigma_2$ и r . А из (ПЗ.3б) следует, что *условная дисперсия случайной величины ξ_2 «при условии $\xi^{(1)} = x^{(1)}$ » на самом деле не зависит от условия!*

Отметим, что поскольку случайные величины $\xi^{(1)}$ и $\xi^{(2)}$ симметрично входят в выражение (ПЗ.1) для плотности $\varphi_{\xi}(\cdot)$, то справедлива и симметричная формулировка свойства 2 для условного распределения случайной величины $\xi^{(1)}$ при условии $\xi^{(2)} = x^{(2)}$.

С в о й с т в о 3. Из взаимной некоррелированности компонент $\xi^{(1)}, \xi^{(2)}$ двумерной случайной величины $\xi = (\xi^{(1)}, \xi^{(2)})^T$ следует их взаимная статистическая независимость, то есть из $r = 0$ следует тождество

$$\varphi_{\xi}(x^{(1)}, x^{(2)}; a^{(1)}, a^{(2)}, \sigma_1^2, \sigma_2^2, r = 0) = \varphi_{\xi^{(1)}}(x^{(1)}; a^{(1)}; \sigma_1^2) \cdot \varphi_{\xi^{(2)}}(x^{(2)}; a^{(2)}; \sigma_2^2), \quad (\text{ПЗ.4})$$

где $\varphi_{\xi^{(j)}}(\cdot)$ определены соотношением (ПЗ.2).

Доказательство свойства 1. Воспользуемся известным правилом вычисления плотности частного распределения случайной величины $\xi^{(2)}$

по плотности совместного распределения двумерной случайной величины $\xi = (\xi^{(1)}, \xi^{(2)})^T$:

$$\varphi_{\xi^{(2)}}(x^{(2)}) = \int_{-\infty}^{\infty} \varphi_{\xi}(x^{(1)}, x^{(2)}) dx^{(1)}. \quad (\text{П3.5})$$

Сначала докажем свойство 1 для взаимнонекоррелированных компонент вектора $\xi = (\xi^{(1)}, \xi^{(2)})^T$, то есть при $r = 0$. В этом случае из (П3.5) и (П3.1') следует:

$$\begin{aligned} \varphi_{\xi^{(2)}}(x^{(2)}) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2\sigma_1^2}(x^{(1)}-a^{(1)})^2} \cdot \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2\sigma_2^2}(x^{(2)}-a^{(2)})^2} dx^{(1)} = \\ &= \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2\sigma_2^2}(x^{(2)}-a^{(2)})^2} \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2\sigma_1^2}(x^{(1)}-a^{(1)})^2} dx^{(1)} = \\ &= \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2\sigma_2^2}(x^{(2)}-a^{(2)})^2} \end{aligned}$$

(в силу условия нормировки плотности одномерного нормального закона).

Пусть теперь $r \neq 0$ (то есть $\sigma_{12} = \sigma_{21} \neq 0$). Применим с двумерной нормальной случайной величине $\xi = (\xi^{(1)}, \xi^{(2)})^T$ такое линейное преобразование C которое оставляет $\xi^{(2)}$ без изменения и обеспечивает, при этом, взаимную некоррелированность преобразованных компонент $\tilde{\xi} = C\xi$. Это достигается подбором такой константы c , при которой $\text{cov}(\xi^{(1)} + c\xi^{(2)}, \xi^{(2)}) = 0$. Из $\text{cov}(\xi^{(1)} + c\xi^{(2)}, \xi^{(2)}) = \text{cov}(\xi^{(1)}, \xi^{(2)}) + c \cdot \text{cov}(\xi^{(2)}, \xi^{(2)}) = \sigma_{12} + c \cdot \sigma_{22} = 0$ получаем $c = -\sigma_{12}/\sigma_{22}$, так что

$$C = \begin{pmatrix} 1 & -\frac{\sigma_{12}}{\sigma_{22}} \\ 0 & 1 \end{pmatrix}.$$

Преобразование C невырожденное (так как $|C| = 1$), так что преобразованная двумерная случайная величина снова нормальна (см. ниже п.5): $\tilde{\xi} \in N_2(Ca, C\Sigma C^T)$, причем ее компоненты $\tilde{\xi}^{(1)} = \xi^{(1)} - \frac{\sigma_{12}}{\sigma_{22}} \xi^{(2)}$ и $\tilde{\xi}^{(2)} = \xi^{(2)}$ взаимнонекоррелированы (по построению). А в этом случае, как было показано выше, частное распределение $\tilde{\xi}^{(2)} = \xi^{(2)}$ является нормальным с параметрами $a^{(2)} = E\xi^{(2)}$ и $\sigma_2^2 = D\xi^{(2)}$. В силу симметрии $\xi^{(1)}$ и $\xi^{(2)}$, поменяв их ролями в приведенных выше рассуждениях, получаем доказательство $(a^{(1)}, \sigma_1^2)$ -нормальности частного распределения случайной величины $\xi^{(1)}$. Доказательство свойства 1 завершено.

Доказательство свойства 2. В целях упрощения обозначений положим $\Delta_j = x^{(j)} - a^j$ ($j = 1, 2$) и воспользуемся общим правилом вычисле-

ния функции плотности *условного* распределения

$$\begin{aligned}
 \varphi_{\xi^{(2)}}(x^{(2)} | \xi^{(1)} = x^{(1)}) &= \frac{\varphi_{\xi}(x^{(1)}, x^{(2)})}{\varphi_{\xi^{(1)}}(x^{(1)})} = \\
 &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} e^{-\frac{1}{2(1-r^2)}\left(\frac{\Delta_1^2}{\sigma_1^2} - 2r\frac{\Delta_1}{\sigma_1}\frac{\Delta_2}{\sigma_2} + \frac{\Delta_2^2}{\sigma_2^2}\right)} = \\
 &= \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2}\frac{\Delta_1^2}{\sigma_1^2}} \cdot \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-r^2}} e^{-\frac{1}{2\sigma_2^2(1-r^2)}\left[\frac{\sigma_2^2}{\sigma_1^2}\Delta_1^2 - 2r\frac{\sigma_2}{\sigma_1}\Delta_1\Delta_2 + \Delta_2^2 - \frac{\sigma_2^2}{\sigma_1^2}(1-r^2)\Delta_1^2\right]} = \\
 &= \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-r^2}} e^{-\frac{1}{2\sigma_2^2(1-r^2)}\left(\Delta_2^2 - 2r\frac{\sigma_2}{\sigma_1}\Delta_1\Delta_2 + r^2\frac{\sigma_2^2}{\sigma_1^2}\Delta_1^2\right)} = \\
 &= \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-r^2}} e^{-\frac{1}{2\sigma_2^2(1-r^2)}\left(\Delta_2 - r\frac{\sigma_2}{\sigma_1}\Delta_1\right)^2} = \\
 &= \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-r^2}} e^{-\frac{1}{2\sigma_2^2(1-r^2)}\left[x^{(2)} - \left(a^{(2)} + r\frac{\sigma_2}{\sigma_1}(x^{(1)} - a^{(1)})\right)\right]^2},
 \end{aligned}$$

что и завершает доказательство свойства 2.

Доказательство свойства 3 следует непосредственно из подстановки в (ПЗ.1') значения $r = 0$:

$$\begin{aligned}
 \varphi_{\xi}(x^{(1)}, x^{(2)}; a^{(1)}, a^{(2)}, \sigma_1^2, \sigma_2^2, r = 0) &= \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left[\frac{(x^{(1)} - a^{(1)})^2}{\sigma_1^2} + \frac{(x^{(2)} - a^{(2)})^2}{\sigma_2^2}\right]} = \\
 &= \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{1}{2\sigma_1^2}(x^{(1)} - a^{(1)})^2} \cdot \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{1}{2\sigma_2^2}(x^{(2)} - a^{(2)})^2} = \\
 &= \varphi_{\xi^{(1)}}(x^{(1)}; a^{(1)}, \sigma_1^2) \cdot \varphi_{\xi^{(2)}}(x^{(2)}; a^{(2)}, \sigma_2^2),
 \end{aligned}$$

что и означает статистическую независимость случайных величин $\xi^{(1)}$ и $\xi^{(2)}$.

3. Общий случай ($p \geq 2$). Разобьем анализируемую p -мерную величину $\xi = (\xi^{(1)}, \dots, \xi^{(p)})^T$ на два подвектора

$$\xi = \begin{pmatrix} \xi^{(1)} \\ \xi^{(q)} \\ \xi^{(q+1)} \\ \vdots \\ \xi^{(p)} \end{pmatrix} = \begin{pmatrix} \xi^{(1)} \\ \xi^{(2)} \end{pmatrix} \in N_p(a; \Sigma),$$

где $\xi^{(1)} = (\xi^{(1)}, \dots, \xi^{(q)})^T$ и $\xi^{(2)} = (\xi^{(q+1)}, \dots, \xi^{(p)})^T$. Соответственно разбиваются вектор $X = (x^{(1)}, \dots, x^{(p)})^T$ возможных значений ξ , вектор

средних значений $\mathbf{a} = (a^{(1)}, \dots, a^{(p)})$ и ковариационная матрица $\Sigma = (\sigma_{ij})$:

$$X = \begin{pmatrix} X(1) \\ X(2) \end{pmatrix}, \quad \mathbf{a} = \begin{pmatrix} a(1) \\ a(2) \end{pmatrix} \quad \text{и} \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

Введенные таким образом подвектора векторов X и $\mathbf{a}$, а также блоки матрицы Σ имеют следующий смысл:

$X(1) = (x^{(1)}, \dots, x^{(q)})^T$ — возможные значения случайной величины $\xi(1)$;

$X(2) = (x^{(q+1)}, \dots, x^{(p)})^T$ — возможные значения случайной величины $\xi(2)$;

$\mathbf{a}(1) = (a^{(1)}, \dots, a^{(q)})^T$ — вектор средних значений случайной величины $\xi(1)$;

$\mathbf{a}(2) = (a^{(q+1)}, \dots, a^{(p)})^T$ — вектор средних значений случайной величины $\xi(2)$;

Σ_{11} — $(q \times q)$ -матрица ковариаций случайной величины $\xi(1)$;

Σ_{22} — $(p - q) \times (p - q)$ -матрица ковариаций случайной величины $\xi(2)$;

Σ_{12} и $\Sigma_{21} = \Sigma_{12}^T$, — соответственно, $q \times (p - q)$ - и $(p - q) \times q$ -матрицы ковариаций компонент вектора $\xi(1)$ с компонентами вектора $\xi(2)$.

Имеют место следующие свойства многомерного нормального распределения, обобщающие свойства 1–3 на случай $p \geq 2$:

С в о й с т в о 1'. Частные распределения многомерных случайных величин $\xi(1)$ и $\xi(2)$ снова нормальны, а именно:

$$\begin{aligned} \xi(1) &\in N_q(\mathbf{a}(1); \Sigma_{11}), \\ \xi(2) &\in N_{p-q}(\mathbf{a}(2); \Sigma_{22}). \end{aligned} \quad (\text{ПЗ.2'})$$

С в о й с т в о 2'. Условное распределение многомерной случайной величины $\xi(2)$ (при условии $\xi(1) = X(1)$) является $(p - q)$ -мерным нормальным с вектором средних значений

$$\mathbf{E}(\xi(2) | \xi(1) = X(1)) = \mathbf{a}(2) + \Sigma_{21} \Sigma_{11}^{-1} (X(1) - \mathbf{a}(1)) \quad (\text{ПЗ.3'а})$$

и ковариационной матрицей

$$\Sigma_{(\xi(2)|\xi(1)=X(1))} = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}. \quad (\text{ПЗ.3'б})$$

И здесь следует подчеркнуть два важных факта, являющихся обобщениями отмеченных выше свойств двумерного нормального распределения на случай $p \geq 2$. Из свойства 2 следует, что если $(\xi(1)^T, \xi(2)^T)^T \in N_p(\mathbf{a}; \Sigma)$, то:

- регрессия любой из компонент вектора $\xi(2)$ по компонентам вектора $\xi(1)$ является линейной по $X(1)$;
- ковариационная матрица условного распределения случайного вектора $\xi(2)$ (при условии $\xi(1) = X(1)$) не зависит от условия, то есть от $X(1)$.

В частном случае $q = p - 1$, важном для рассматриваемых в книге линейных моделей регрессии, имеем:

$$(\xi^{(p)} \mid \xi^{(1)} = x^{(1)}, \xi^{(2)} = x^{(2)}, \dots, \xi^{(p-1)} = x^{(p-1)}) \in \\ \in N_1 \left(a^{(p)} + \Sigma_{21} \Sigma_{11}^{-1} (X(1) - a(1)); \sigma_{pp} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} \right),$$

где $X(1) = (x^{(1)}, x^{(2)}, \dots, x^{(p-1)})^T$.

С в о й с т в о 3'. Необходимым и достаточным условием взаимной статистической независимости случайных величин $\xi(1)$ и $\xi(2)$ является выполнение равенства

$$\Sigma_{12} = 0_{q,p-q} \quad \text{или, что то же:} \quad \Sigma_{21} = 0_{p-q,q},$$

где $0_{k,m}$ — означает $k \times m$ -матрицу, состоящую из нулей.

Отсюда, в частности, следует, что необходимым и достаточным условием взаимной статистической независимости всех компонент $\xi^{(1)}, \dots, \xi^{(p)}$ p -мерной нормальной случайной величины $\xi = (\xi^{(1)}, \dots, \xi^{(p)})^T$ является равенство нулю всех ковариаций σ_{ij} при $i \neq j$, то есть — диагональность ковариационной матрицы Σ .

Последнее свойство без особого труда может непосредственно проверить читатель самостоятельно, подставляя в формулу (ПЗ.1) матрицу

$$\Sigma = \begin{pmatrix} \sigma_{11} & & & 0 \\ & \sigma_{22} & & \\ & & \ddots & \\ 0 & & & \sigma_{pp} \end{pmatrix}$$

Доказательство остальных свойств мы опускаем (читатель может их найти, например, в книге [Андерсон (1963), пп. 2.4, 2.5]).

4. Статистическое оценивание параметров $N_p(a; \Sigma)$. Пусть мы располагаем случайной выборкой

$$(X_1, X_2, \dots, X_n), \quad X_i = (x_i^{(1)}, \dots, x_i^{(p)})^T \quad (\text{ПЗ.6})$$

из $(a; \Sigma)$ — нормальной p -мерной генеральной совокупности $N_p(a; \Sigma)$, и пусть по наблюдениям (ПЗ.6) требуется вычислить оценки методом максимального правдоподобия $\hat{a}_{\text{ММП}}$ и $\hat{\Sigma}_{\text{ММП}}$, соответственно, для вектора средних a и ковариационной матрицы Σ . Как известно, ММП-оценки

$\hat{\Theta}_{\text{ММП}}$ параметров Θ определяются как решения оптимизационной задачи вида

$$\hat{\Theta}_{\text{ММП}} = \arg \max_{\Theta} L(X_1, \dots, X_n | \Theta),$$

где $L(X_1, \dots, X_n | \Theta)$ — функция правдоподобия наблюдений $X_1, \dots, X_n$, вычисленная в предположении, что значения неизвестных параметров равны Θ . В нашем случае $\Theta = (\mathbf{a}; \Sigma)$, а функция правдоподобия L определяется соотношением:

$$\begin{aligned} L(X_1, \dots, X_n | \mathbf{a}; \Sigma) &= \prod_{i=1}^n \varphi_{\xi}(X_i; \mathbf{a}; \Sigma) = \\ &= \frac{1}{(2\pi)^{\frac{np}{2}} |\Sigma|^{\frac{n}{2}}} e^{-\frac{1}{2} \sum_{i=1}^n (X_i - \mathbf{a})^T \Sigma^{-1} (X_i - \mathbf{a})}. \end{aligned}$$

Переход к логарифмической функции правдоподобия $l = \ln L$ и ряд тождественных преобразований l , выполненных с учетом правил матричной алгебры (в частности, использовались правила: $|\mathbf{A}| = 1/|\mathbf{A}^{-1}|$, $\text{tr}(\sum_{i=1}^n \mathbf{A}_i) = \sum_{i=1}^n \text{tr} \mathbf{A}_i$, $\text{tr}(\mathbf{CD}) = \text{tr}(\mathbf{DC})$, где $\mathbf{C}$ и $\mathbf{D}$, соответственно, $p \times m$ - и $m \times p$ -матрицы), позволили представить l в виде:

$$\begin{aligned} l(X_1, \dots, X_n | \mathbf{a}; \Sigma) &= -\frac{pn}{2} \ln(2\pi) + \frac{n}{2} \ln |\Sigma^{-1}| - \frac{1}{2} \text{tr}(\Sigma^{-1} \mathbf{S}) - \\ &\quad - \frac{n}{2} \text{tr}[(\bar{X} - \mathbf{a})^T \Sigma^{-1} (\bar{X} - \mathbf{a})], \end{aligned} \quad (\text{ПЗ.7})$$

где $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$, а $\mathbf{S} = \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T$. Последний член правой части (ПЗ.7) всегда положителен (так как Σ^{-1} — положительно определенная матрица) и, следовательно, имеет абсолютно минимальное значение при $\mathbf{a} = \bar{X}$ (при любой положительно определенной матрице Σ^{-1}). Значит, максимизируя l , следует брать $\mathbf{a} = \bar{X}$, то есть $\hat{\mathbf{a}}_{\text{ММП}} = \bar{X}$. Остается максимизировать по Σ выражение

$$\tilde{l}(\Sigma) = \frac{n}{2} \ln |\Sigma^{-1}| - \frac{1}{2} \text{tr}(\Sigma^{-1} \mathbf{S}).$$

Удобнее решать эту задачу относительно $\mathbf{B} = \Sigma^{-1}$, то есть

$$\tilde{l}(\mathbf{B}) = \frac{n}{2} \ln |\mathbf{B}| - \frac{1}{2} \text{tr}(\mathbf{BS}) \rightarrow \max_{\mathbf{B}}.$$

Решение системы уравнений, получающейся от приравнивания к нулю производных функции $\tilde{l}(\mathbf{B})$ по элементам матрицы $\mathbf{B}$, дает:

$$\hat{\mathbf{B}}_{\text{ММП}}^{-1} = \frac{1}{n} \mathbf{S}.$$

Таким образом, ММП-оценки параметров α и Σ (α, Σ)-нормального p -мерного распределения имеют вид:

$$\hat{\alpha}_{\text{ММП}} = \frac{1}{n} \sum_{i=1}^n X_i, \quad (\text{ПЗ.8})$$

$$\hat{\Sigma}_{\text{ММП}} = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T \quad (\text{ПЗ.9})$$

или в покомпонентной записи:

$$\hat{\alpha}_{\text{ММП}}^{(j)} = \frac{1}{n} \sum_{i=1}^n x_i^{(j)} = \bar{x}^{(j)}, \quad (\text{ПЗ.8}')$$

$$(\hat{\sigma}_{jl})_{\text{ММП}} = \frac{1}{n} \sum_{i=1}^n (x_i^{(j)} - \bar{x}^{(j)})(x_i^{(l)} - \bar{x}^{(l)}) \quad (\text{ПЗ.9}')$$

(более подробное описание решения задачи $\tilde{l}(B) \rightarrow \max_B$ см., например, в [Андерсон (1963), п. 3.2]).

5. Линейное преобразование многомерной нормальной случайной величины. Пусть $\xi \in N_p(\alpha; \Sigma)$ и $C - q \times p$ -матрица преобразования ($q \leq p$) полного ранга (то есть $\text{ранг } C = q$). Тогда q -мерная случайная величина $\eta = C\xi$ также подчиняется нормальному закону с вектором средних значений $E\eta = C\alpha$ и ковариационной матрицей $C\Sigma C^T$, то есть $\eta \in N_q(C\alpha, C\Sigma C^T)$.

ПЗ.2. Дискриминантный анализ

1. Постановка задач. Пусть мы располагаем последовательностью независимых многомерных наблюдений

$$\{X_1, X_2, \dots, X_n\}, \quad X_i = (x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)})^T, \quad (\text{ПЗ.10})$$

о каждом из которых известно, что оно принадлежит одному из k классов (к какому именно – неизвестно!), где под j -м классом ($j = 1, 2, \dots, k$) понимается генеральная совокупность, описываемая законом распределения вероятностей с *одномодальной* функцией плотности $f_j(X)$. И пусть в нашем распоряжении имеются так называемые «обучающие выборки»

$$X_{j1}, X_{j2}, \dots, X_{jn_j}, \quad j = 1, 2, \dots, k, \quad (\text{ПЗ.11})$$

каждая (j -я) из которых определяет значения анализируемых признаков $X_{ji} = (x_{ji}^{(1)}, x_{ji}^{(2)}, \dots, x_{ji}^{(p)})$, $i = 1, 2, \dots, n_j$, — на n_j объектах, о которых априори известно, что все они принадлежат j -му классу (то есть каждый класс представлен своей порцией выборочных данных).

Задача дискриминантного анализа состоит в том, чтобы предложить наилучшее (в определенном смысле) правило отнесения каждого из наблюдений X_i последовательности (ПЗ.10) к одному из представленных в (ПЗ.11) классов.

При этом классифицируемые наблюдения (ПЗ.10) интерпретируются как случайная выборка из генеральной совокупности, описываемой смесью k одномодальных генеральных совокупностей с функцией плотности вероятности

$$f(X) = \sum_{j=1}^k \pi_j f_j(X), \quad (\text{ПЗ.12})$$

где π_j — априорная вероятность появления в этой выборке элемента из j -го класса (или, другими словами, π_j — это удельный вес элементов j -го класса в **общей** генеральной совокупности (ПЗ.12)). Отметим, что, вообще говоря, ни априорные вероятности π_j , ни плотности $f_j(X)$ ($j = 1, 2, \dots, k$) не являются известными в этой задаче, а последовательность классифицируемых наблюдений (ПЗ.10) может состоять и из единственного наблюдения (то есть возможно $n = 1$).

2. Оптимальное (байесовское) правило классификации. Введем понятие *решающего правила (процедуры)* $\delta(X)$ классификации наблюдения X . Определим его как функцию от X , которая может принимать только целые положительные значения $1, 2, \dots, k$, так что те X , при которых она принимает значение, равное j , мы будем относить к классу j . Тем самым p -мерное пространство $\Pi(X)$ возможных значений анализируемого признака X разбивается на k непересекающихся подпространств

$$S_j = \{X: \delta(X) = j\}, \quad j = 1, 2, \dots, k.$$

Другими словами, решающее правило $\delta(X)$ может быть задано разбиением

$$S = \{S_1, S_2, \dots, S_k\}$$

пространства $\Pi(X)$ на k непересекающихся областей.

Для того чтобы уточнить, в каком именно смысле искомое правило классификации должно быть наилучшим, введем **функцию потерь** $C(\delta)$ от неправильной классификации при использовании правила δ :

$$C(\delta) = \sum_{i=1}^k \pi_i \sum_{j=1}^k c(j|i) P_\delta(j|i), \quad (\text{ПЗ.13})$$

где $c(j|i)$ — потери, которые мы несем при отнесении объекта класса i к классу j , а $P_\delta(j|i)$ — вероятность отнесения объекта класса i к классу j при использовании правила δ . Можно показать (см., например, [Андерсон (1963)]), что наилучшее — в смысле минимизации потерь (ПЗ.13) — правило классификации $\delta_{\text{опт}}(X)$ задается разбиением

пространства $\Pi(X)$ на множества

$$S_j^{\text{опт}} = \left\{ X: \sum_{\substack{i=1 \\ i \neq j}}^k \pi_i f_i(X) c(j|i) = \min_{1 \leq l \leq k} \sum_{\substack{i=1 \\ i \neq l}}^k \pi_i f_i(X) c(l|i) \right\}. \quad (\text{ПЗ.14})$$

Другими словами, наблюдение X_ν ($\nu = 1, 2, \dots, n$) будет отнесено к классу j тогда и только тогда, когда средние удельные потери от его отнесения именно в этот класс окажутся минимальными по сравнению с аналогичными потерями, связанными с отнесением этого наблюдения в любой другой класс.

Относительно простой вид приобретает правило классификации (ПЗ.14) в случае равных потерь $c(j|i)$, то есть при

$$c(j|i) = \begin{cases} c_0 > 0 & \text{при } j \neq i \\ 0 & \text{при } j = i. \end{cases} \quad (\text{ПЗ.15})$$

В этом случае наблюдение X_ν будет отнесено к классу j тогда, когда

$$\pi_j f_j(X_\nu) = \max_{1 \leq l \leq k} \pi_l f_l(X_\nu), \quad (\text{ПЗ.14}')$$

то есть определяется тот класс, в рамках которого максимизируется «взвешенная правдоподобность» этого наблюдения (где в качестве весов выступают априорные вероятности π_j). Отметим, что если потери $c(j|i)$ от неправильной классификации определяются соотношением (ПЗ.15), то выражение для функции потерь $C(\delta)$ (см. (ПЗ.13)) сводится к виду

$$C(\delta) = c_0 \left(1 - \sum_{i=1}^k \pi_i P_\delta(i|i) \right), \quad (\text{ПЗ.13}')$$

так что минимизация (по δ) потерь $C(\delta)$ равносильна максимизации (по δ) средней вероятности правильной классификации, то есть выражения $\sum_{i=1}^k \pi_i P_\delta(i|i)$.

Однако соотношения (ПЗ.14) и (ПЗ.14') задают нам лишь *теоретическое* оптимальное правило классификации: для того чтобы его реально построить, необходимо знание априорных вероятностей $\pi_1, \pi_2, \dots, \pi_k$ и з.р.в. $f_1(X), f_2(X), \dots, f_k(X)$. В статистическом варианте решения этой задачи данные величины заменяются соответствующими оценками, построенными на базе обучающих выборок (ПЗ.11).

Априорные вероятности π_j ($j = 1, 2, \dots, k$) оцениваются просто, если ряд наблюдений, составленный из всех обучающих выборок (ПЗ.11), может быть классифицирован как случайная выборка объема $n_{\text{об}} = n_1 + n_2 + \dots + n_k$ из генеральной совокупности (ПЗ.12). Тогда вместо π_j используют их оценки

$$\hat{\pi}_j = \frac{n_j}{n_{\text{об}}},$$

где n_j — объем j -й обучающей выборки.

Впрочем, величины π_j часто определяются априори самой содержательной сущностью задачи.

Что касается задачи оценки з.р.в. $f_1(X), \dots, f_k(X)$, то ее удобно разбить на два случая:

1-й случай (параметрический дискриминантный анализ) характеризуется известным *общим видом* функций $f_j(X)$, то есть все классы описываются з.р.в. одного и того же параметрического семейства $\{f(X; \Theta)\}$: класс i отличается от класса j только значением параметра Θ , то есть

$$f_j(X) = f(X; \Theta_j), \quad j = 1, 2, \dots, k.$$

Тогда в качестве оценок $\hat{f}_j(X)$ неизвестных функций $f_j(X)$ используются функции $f_j(X; \hat{\Theta}_j)$, где $\hat{\Theta}_j$ — статистическая оценка неизвестного значения параметра Θ_j , полученная по наблюдениям j -й обучающей выборки (ПЗ.11). Пример реализации этой схемы приведен в следующем пункте.

2-й случай (непараметрический дискриминантный анализ) не предусматривает знания общего вида функций $f_j(X)$ ($j = 1, 2, \dots, k$). В этом случае приходится строить так называемые *непараметрические оценки* $\hat{f}_j(X)$ для функций $f_j(X)$, например, гистограммного или ядерного типа, либо пользоваться некоторыми специальными приемами (см., например, [Айвазян, Бежаева, Староверов (1974), с. 44–50]).

3. Параметрический дискриминантный анализ в случае нормальных классов. Предположим, что класс j идентифицируется p -мерной нормальной генеральной совокупностью с вектором средних значений α_j и ковариационной матрицей Σ (*общей для всех классов*), $j = 1, 2, \dots, k$. Тогда при реализации общих оптимальных правил классификации (ПЗ.14) и (ПЗ.14') в качестве функций $f_j(X)$ следует использовать p -мерные нормальные плотности $\varphi(X; \hat{\alpha}_j, \hat{\Sigma})$ (см. (ПЗ.1)), где оценки для векторов средних значений $\hat{\alpha}_j = (\hat{\alpha}_j^{(1)}, \hat{\alpha}_j^{(2)}, \dots, \hat{\alpha}_j^{(p)})^T$ и для ковариационной матрицы $\hat{\Sigma} = (\hat{\sigma}_{lq})$ получены с помощью метода максимального правдоподобия (см. (ПЗ.8), (ПЗ.9)) по обучающим выборкам (ПЗ.11); они имеют вид

$$\hat{\alpha}_j^{(l)} = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ji}^{(l)}, \quad \begin{matrix} l = 1, 2, \dots, p, \\ j = 1, 2, \dots, k; \end{matrix} \quad (\text{ПЗ.16})$$

$$\hat{\sigma}_{lq} = \frac{1}{n_{об} - k} \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ji}^{(l)} - \hat{\alpha}_j^{(l)})(x_{ji}^{(q)} - \hat{\alpha}_j^{(q)}), \quad (\text{ПЗ.17})$$

$$l, q = 1, 2, \dots, p,$$

$$n_{об} = n_1 + n_2 + \dots + n_k.$$

Правило классификации (ПЗ.14') наблюдений (ПЗ.10) эквивалентно следующему: наблюдение X_ν относится к классу с номером j_0 тогда и только тогда, когда

$$\frac{f_{j_0}(X_\nu)}{f_j(X_\nu)} \geq \frac{\pi_j}{\pi_{j_0}} \quad \text{для всех } j = 1, 2, \dots, k, \quad (\text{ПЗ.18})$$

или, что то же,

$$\ln \frac{f_{j_0}(X_\nu)}{f_j(X_\nu)} \geq \ln \frac{\pi_j}{\pi_{j_0}} \quad \text{для всех } j = 1, 2, \dots, k. \quad (\text{ПЗ.18}')$$

В случае *нормальных* плотностей $f_j(X)$, то есть функций $f_j(X)$, определяемых по формуле (ПЗ.1) с заменой α и Σ соответственно векторами $\hat{\alpha}_j$ и матрицей $\hat{\Sigma}$, определяемыми по формулам (ПЗ.16)–(ПЗ.17), соотношение (ПЗ.18') эквивалентно соотношению

$$\left[X_\nu - \frac{1}{2}(\hat{\alpha}_{j_0} + \hat{\alpha}_j) \right]^\top \hat{\Sigma}^{-1}(\hat{\alpha}_{j_0} - \hat{\alpha}_j) \geq \ln \frac{\pi_j}{\pi_{j_0}} \quad (\text{ПЗ.18}'')$$

для всех $j = 1, 2, \dots, k$.

Левая часть соотношения (ПЗ.18'') определяет *вид линейной дискриминантной функции Фишера в задаче различения нормальных классов при постоянных значениях потерь от неправильной классификации*. Особенно простым становится это правило классификации в случае двух классов ($k = 2$) и одинаковых априорных вероятностей ($\pi_1 = \pi_2 = 0,5$). В этом случае наблюдение X_ν следует относить к 1-му классу тогда и только тогда, когда

$$\left[X_\nu - \frac{1}{2}(\hat{\alpha}_1 + \hat{\alpha}_2) \right]^\top \hat{\Sigma}^{-1}(\hat{\alpha}_1 - \hat{\alpha}_2) \geq 0, \quad (\text{ПЗ.18}''')$$

и ко 2-му классу во всех остальных случаях. Легко понять, что при классификации на два класса *одномерных* ($p = 1$) нормальных наблюдений решение об отношении наблюдения x_ν к одному из двух классов будет определяться знаком произведения $[x_\nu - (\hat{\alpha}_1 + \hat{\alpha}_2)/2](\hat{\alpha}_1 - \hat{\alpha}_2)$.

Дискриминантный анализ нормальных классов основан на линейной дискриминантной функции Фишера (ПЗ.18'') лишь в случае, когда все различаемые классы имеют одну и ту же ковариационную матрицу Σ . Если же нормальные классы отличаются друг от друга и векторами средних значений α_j , и ковариационными матрицами Σ_j , то левая часть соотношений (ПЗ.18') сводится к виду

$$\frac{1}{2} \ln \frac{|\Sigma_j|}{|\Sigma_{j_0}|} - \frac{1}{2} [(X_\nu - \alpha_{j_0})^\top \Sigma_{j_0}^{-1} (X_\nu - \alpha_{j_0}) - (X_\nu - \alpha_j)^\top \Sigma_j^{-1} (X_\nu - \alpha_j)],$$

так что оптимальное (байесовское) правило классификации в этом случае формулируется следующим образом: *наблюдение X_ν относится к классу с номером j_0 тогда и только тогда, когда*

$$\begin{aligned} & -\frac{1}{2}[(X_\nu - a_{j_0})^\top \Sigma_{j_0}^{-1}(X_\nu - a_{j_0}) - (X_\nu - a_j)^\top \Sigma_j^{-1}(X_\nu - a_j)] \geq \\ & \geq \ln \frac{\pi_j}{\pi_{j_0}} + \frac{1}{2} \ln \frac{|\Sigma_{j_0}|}{|\Sigma_j|} \end{aligned} \quad (\text{ПЗ.19})$$

при всех $j = 1, 2, \dots, k$.

(левую часть соотношения (ПЗ.19) также называют *дискриминантной функцией*; как мы видим, в данном случае она является *квадратичной* по X_ν).

ПЗ.3. Кластерный анализ

1. Постановка задачи. В данном случае, так же, как и в п. ПЗ.2, речь идет о задаче классификации наблюдений (ПЗ.10). Однако **теперь мы не располагаем обучающими выборками (ПЗ.11)** и можем даже не иметь информации о числе классов k . Таким образом, вся исходная информация заключена в самих классифицируемых наблюдениях (ПЗ.10) и в некоторых априорных сведениях о природе классов и/или о критерии качества процедуры классификации (то есть о том, как измерять качество разбиения наблюдений (ПЗ.10) на классы).

В зависимости от наличия и характера априорных сведений о природе искоемых классов и от конечных прикладных целей исследования в кластерном анализе рассматриваются два подхода к решению задачи классификации наблюдений (ПЗ.10) — **параметрический** (основанный на параметрической версии модели смеси распределений (ПЗ.12)) и **непараметрический** (основанный на анализе геометрической структуры классифицируемых данных (ПЗ.10) и на заданном критерии качества классификации).

2. Параметрический кластерный анализ, основанный на модели смеси распределений. В данном подходе классифицируемые наблюдения (ПЗ.10) интерпретируются (так же, как и в параметрическом дискриминантном анализе) как случайная выборка из генеральной совокупности, являющейся смесью генеральных совокупностей типа (ПЗ.12) при заданном общем виде составляющих эту смесь функций $f_j(X) = f(X; \Theta_j)$, $j = 1, 2, \dots, k$ ¹. Тогда решение задачи классификации наблюдений (ПЗ.10) разбивается на 2 этапа. На первом этапе методом максимального правдоподобия решается задача оценивания параметров (π_j, Θ_j) — $j = 1, 2, \dots, k$ — в параметрической модели смеси (ПЗ.12) (число классов k полагается заданным). На втором этапе, уже располагая

¹Вообще говоря, допускается *различный* общий вид у компонентов смеси, то есть $f_j(X) = f_j(X; \Theta_j)$. Однако в приложениях эта схема используется крайне редко.

оценками $\hat{\pi}_j, \hat{\Theta}_j$, строится *оптимальное (байесовское) правило классификации* (см. выше п. ПЗ.2), с помощью которого и производится разбиение наблюдений (ПЗ.10) на классы. Отметим, что для реализации первого этапа мы должны располагать достаточным количеством классифицируемых наблюдений (в отличие от дискриминантного анализа, где n могло быть равным даже единице).

Таким образом, мы видим, что *главное отличие схемы параметрического дискриминантного анализа от данной схемы кластерного анализа, реализуемой в рамках модели смеси распределений, — в способе оценивания неизвестных параметров, от которых зависят функции плотности распределения, описывающие классы*. Но оценивание параметров в модели смеси — процесс существенно более сложный, чем оценивание параметров по обучающим выборкам. Весьма подробное описание процедур оценивания параметров смеси вероятностных распределений читатель найдет в книге [Айвазян и др. (1989)].

Наконец, следует упомянуть, что существуют схемы, обобщающие описанный выше подход на случай неизвестного числа классов k (см., например, [Aivazian (1996)]).

3. Непараметрический кластерный анализ. Задача, как и прежде, состоит в наилучшем (в определенном смысле) разбиении многомерных наблюдений (ПЗ.10) на определенное число (заданное или нет) однородных (в определенном смысле) классов. Однако в данном случае мы не располагаем ни «обучающими выборками», ни *априорной информацией о характере* распределения наблюдений X_i внутри каждого из классов (в предыдущем случае нам был известен общий параметрический вид этих распределений).

Для формализации проблемы классификации наблюдений (ПЗ.10) в данном случае (и в частности, для уточнения, в каком именно смысле мы ищем наилучшее разбиение и в каком именно смысле мы понимаем однородность наблюдений, принадлежащих одному классу) удобно интерпретировать анализируемые наблюдения в качестве точек в соответствующем p -мерном признаковом пространстве.

Естественно предположить, что геометрическая близость двух или нескольких точек в этом пространстве означает близость «физических» состояний соответствующих объектов, их однородность. Тогда проблема классификации состоит в разбиении анализируемой совокупности точек — наблюдений на сравнительно небольшое число (заранее известное или нет) классов таким образом, чтобы объекты, принадлежащие одному классу, находились бы на сравнительно небольших расстояниях друг от друга. Полученные в результате разбиения классы часто называют *кластерами* (таксонами, образами)². Именно поэтому методы их нахож-

²Cluster (англ.) — гроздь, пучок, скопление, группа элементов, характеризуемых каким-либо общим свойством. Тахон (англ.) — систематизированная группа любой

дения называют методами кластерного анализа, численной таксономией, распознаванием образов без обучения.

Однако, берясь за решение задачи классификации, исследователь с самого начала должен четко представлять, какую именно из двух задач он решает. Рассматривает ли он обычную задачу разбиения статистически обследованного (p -мерного) диапазона изменения значений анализируемых признаков на *интервалы (гиперобласти) группирования*, в результате решения которой исследуемая совокупность объектов разбивается на некоторое число групп так, что объекты такой одной группы находятся друг от друга на сравнительно небольшом расстоянии (многомерный аналог задачи построения интервалов группирования при обработке одномерных наблюдений). Либо он пытается определить *естественное расслоение* исходных наблюдений на четко выраженные кластеры, сгустки, лежащие друг от друга на некотором расстоянии, но не разбивающиеся на столь же удаленные части. В вероятностной интерпретации (то есть если интерпретировать классифицируемые наблюдения $X_1, X_2, \dots, X_n$ как выборку из некоторой многомерной генеральной совокупности, описываемой функцией плотности или полигоном распределения $f(X)$, как правило, не известными исследователю) вторая задача может быть сформулирована как задача выявления областей повышенной плотности наблюдений, то есть таких областей возможных значений анализируемого многомерного признака X , *которые соответствуют локальным максимумам функции плотности $f(X)$* .

Если первая задача — построения областей группирования — всегда имеет решение, то при второй постановке результат может быть и отрицательным: может оказаться, что множество исходных наблюдений не обнаруживает естественного расслоения на кластеры (например, образует один общий кластер).

В общем случае понятие **однородности** объектов определяется заданием правила вычисления величины ρ_{ij} , характеризующей либо расстояние $d(X_i, X_j)$ между объектами X_i и X_j из исследуемой совокупности, либо степень близости (сходства) $r(X_i, X_j)$ тех же объектов. Если задана функция $d(X_i, X_j)$, то близкие в смысле этой метрики объекты считаются однородными, принадлежащими к одному классу. Естественно, при этом необходимо сопоставление $d(X_i, X_j)$ с некоторым пороговым значением, определяемым в каждом конкретном случае по-своему.

При задании расстояний и мер близости нужно помнить о необходимости соблюдения следующих естественных требований: требования *симметрии* ($d(X_i, X_j) = d(X_j, X_i)$ и $r(X_i, X_j) = r(X_j, X_i)$); требования *максимального сходства объекта с самим собой* ($r(X_i, X_i) =$

категории (биологический термин). Название «кластер-анализ» для совокупности методов решения задач такого типа было впервые введено, по-видимому, Трайном в 1939 г. (Tryon R. C. Cluster Analysis//Ann. Arb., Edw. Brothers. 1939).

$= \max_{1 \leq j \leq n} r(X_i, X_j)$) и требования при заданной метрике *монотонного убывания* $r(X_i, X_j)$ по $d(X_i, X_j)$, то есть из $d(X_k, X_l) \geq d(X_i, X_j)$ должно с необходимостью следовать выполнение неравенства $r(X_k, X_l) \leq r(X_i, X_j)$.

Ниже приводятся расстояния, наиболее распространенные в практике статистических и эконометрических исследований, использующих методы кластерного анализа.

• **Обычное евклидово расстояние**

$$d_E(X_i, X_j) = \sqrt{\sum_{l=1}^p (x_i^{(l)} - x_j^{(l)})^2}.$$

К ситуациям, в которых использование этого расстояния можно признать оправданным, прежде всего относят следующие:

- а) наблюдения X извлекаются из генеральных совокупностей, описываемых многомерным нормальным законом с ковариационной матрицей вида $\sigma^2 \mathbf{I}$, то есть компоненты X взаимно независимы и имеют одну и ту же дисперсию;
- б) компоненты $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ вектора наблюдений X однородны по своему физическому смыслу, причем установлено, например с помощью опроса экспертов, что все они одинаково важны с точки зрения решения вопроса об отнесении объекта к тому или иному классу;
- в) признаковое пространство совпадает с геометрическим пространством нашего бытия, что может быть лишь в случаях $p = 1, 2, 3$, и понятие близости объектов соответственно совпадает с понятием геометрической близости в этом пространстве, например классификация попаданий при стрельбе по цели.

• **Взвешенное евклидово расстояние**

$$d_{B.E}(X_i, X_j) = \sqrt{\sum_{l=1}^p \omega_l (x_i^{(l)} - x_j^{(l)})^2}.$$

Обычно применяется в ситуациях, в которых так или иначе удастся приписать каждой из компонент $x^{(l)}$ вектора наблюдений X некоторый неотрицательный «вес» ω_l , пропорциональный степени его важности с точки зрения решения вопроса об отнесении заданного объекта к тому или иному классу. Удобно полагать при этом

$$0 \leq \omega_l \leq 1, \quad l = \overline{1, p}, \quad \sum_{l=1}^p \omega_l = 1.$$

- **Хеммингово расстояние**

Используется как мера различия объектов, задаваемых *дихотомическими* признаками. Оно задается с помощью формулы

$$d_H(X_i, X_j) = \sum_{s=1}^p |x_i^{(s)} - x_j^{(s)}|$$

и, следовательно, равно числу v_{ij} несовпадений значений соответствующих признаков в рассматриваемых i -м и j -м объектах.

- **Расстояние махаланобисского типа**

Обычно используется, если наблюдения X_i извлечены из генеральных совокупностей, описываемых многомерными нормальными законами с одной и той же ковариационной матрицей Σ :

$$d_M(X_i, X_j) = \sqrt{(X_i - X_j)^\top \Sigma^{-1} (X_i - X_j)}.$$

При конструировании различных процедур классификации (кластер-процедур) в ряде ситуаций оказывается целесообразным введение понятия **расстояния между целыми группами объектов**. Приведем примеры наиболее распространенных расстояний, характеризующих взаимное расположение отдельных групп объектов. Пусть S_i — i -я группа (класс, кластер) объектов, n_i — число объектов, образующих группу S_i , вектор $\bar{X}(i)$ — среднее арифметическое векторных наблюдений, входящих в S_i (другими словами, $\bar{X}(i)$ — «центр тяжести» i -й группы), а $\rho(S_l, S_m)$ — расстояние между группами S_l и S_m .

Ниже приводятся наиболее употребительные и наиболее общие расстояния между классами объектов:

- **расстояние, измеряемое по принципу «ближнего соседа»** (“nearest neighbour”)

$$\rho_{\min}(S_l, S_m) = \min_{X_i \in S_l, X_j \in S_m} d(X_i, X_j);$$

- **расстояние, измеряемое по принципу «дальнего соседа»** (“furthest neighbour”)

$$\rho_{\max}(S_l, S_m) = \max_{X_i \in S_l, X_j \in S_m} d(X_i, X_j);$$

- **расстояние, измеряемое по «центрам тяжести» групп**

$$\rho(S_l, S_m) = d(\bar{X}(l), \bar{X}(m));$$

- **расстояние, измеряемое по принципу «средней связи»** определяется как арифметическое среднее всевозможных попарных расстояний между представителями рассматриваемых групп, то есть

$$\rho_{\text{ср}}(S_l, S_m) = \frac{1}{n_l n_m} \sum_{X_i \in S_l} \sum_{X_j \in S_m} d(X_i, X_j);$$

- **обобщенное (по Колмогорову) расстояние между классами, или обобщенное K -расстояние**, вычисляется по формуле

$$\rho_{\tau}^{(K)}(S_l, S_m) = \left[\frac{1}{n_l n_m} \sum_{X_i \in S_l} \sum_{X_j \in S_m} d^{\tau}(X_i, X_j) \right]^{\frac{1}{\tau}}.$$

В частности, при $\tau \rightarrow \infty$ при $\tau \rightarrow -\infty$ имеем

$$\rho_{\infty}^{(K)}(S_l, S_m) = \rho_{\max}(S_l, S_m); \quad \rho_{-\infty}^{(K)}(S_l, S_m) = \rho_{\min}(S_l, S_m).$$

Можно показать также, что

$$\rho_o^{(K)}(S_l, S_m) = \left(\prod_{X_i \in S_l, X_j \in S_m} d(X_i, X_j) \right)^{\frac{1}{n_l \cdot n_m}},$$

$$\rho_1^{(K)}(S_l, S_m) = \rho_{\text{ср}}(S_l, S_m).$$

- **расстояние Каллбэка** между нормальными классами $N_p(\mathbf{a}(l), \Sigma(l))$ и $N_p(\mathbf{a}(m), \Sigma(m))$:

$$\rho^2(S_l, S_m) = \frac{1}{2} (\bar{X}(l) - \bar{X}(m))^{\top} \left(\hat{\Sigma}^{-1}(l) + \hat{\Sigma}^{-1}(m) \right) (\bar{X}(l) - \bar{X}(m)) + \frac{1}{2} \text{tr} \left\{ \left(\hat{\Sigma}(l) - \hat{\Sigma}(m) \right) \left(\hat{\Sigma}^{-1}(l) - \hat{\Sigma}^{-1}(m) \right) \right\},$$

где $\hat{\Sigma}(q) = \frac{1}{n_q} \sum_{X_i \in S_q} (X_i - \bar{X}(q))(X_i - \bar{X}(q))^{\top}$ — оценка ковариационной матрицы $\Sigma(q)$ по наблюдениям q -го класса ($q = 1, 2, \dots, k$).

- **расстояние Махаланобиса** между нормальными классами $N_p(\mathbf{a}(l); \Sigma)$ и $N_p(\mathbf{a}(m); \Sigma)$, различающимися только векторами средних значений:

$$\rho^2(S_l, S_m) = (\bar{X}(l) - \bar{X}(m))^{\top} \hat{\Sigma}^{-1} (\bar{X}(l) - \bar{X}(m)),$$

где $\hat{\Sigma} = \frac{n_1 \hat{\Sigma}(1) + n_2 \hat{\Sigma}(2) + \dots + n_k \hat{\Sigma}(k)}{n_1 + n_2 + \dots + n_k - k}$ — оценка *общей* ковариационной матрицы Σ .

Понятие расстояния между группами элементов особенно важно в так называемых *агломеративных иерархических кластер-процедурах* (см. ниже), поскольку принцип работы таких алгоритмов состоит в последовательном объединении элементов, а затем и целых групп, сначала самых близких, а потом все более и более отдаленных друг от друга.

Для сравнения качества различных разбиений $S = \{S_1, S_2, \dots, S_k\}$ заданного множества наблюдений (ПЗ.10) на классы используются **критерии качества разбиения** $Q(S)$. Тогда под наилучшим разбиением $S^{\text{опт}}$ понимается то разбиение, на котором достигается экстремум (минимум или максимум, в зависимости от смысла критерия $Q(S)$) заданного критерия качества разбиения $Q(S)$. Выбор конкретной формы критерия $Q(S)$ опирается, как правило, на эмпирические и профессионально-интуитивные соображения, а не на какую-либо строгую формализованную схему. Одним из наиболее распространенных критериев качества разбиений (при заданном числе классов k) является критерий в **форме взвешенной суммы мер внутриклассового разброса наблюдений**, то есть

$$Q(S) = \sum_{j=1}^k \sum_{X_i \in S_j} d^2(X_i, \bar{X}(j)), \quad (\text{ПЗ.20})$$

где $\bar{X}(j) = \frac{1}{n_j} \sum_{X_i \in S_j} X_i$ — «центр тяжести» класса j в разбиении $S = \{S_1, S_2, \dots, S_k\}$, а $d^2(\cdot, \cdot)$ — квадрат расстояния между заданными точками анализируемого p -мерного пространства.

Более подробные сведения о критериях качества разбиений наблюдений на классы, в том числе, в случае априори неизвестного числа классов k , читатель может найти, например, в [Айвазян, Мхитарян (2001)].

Ниже дается краткое описание двух процедур непараметрического кластерного анализа, программное обеспечение которых представлено в ряде широко распространенных пакетов статистических программ (SPSS, Statistica и др.).

Метод k -средних (при априори заданном числе классов k) используется обычно в ситуациях, когда число n классифицируемых наблюдений (ПЗ.10) достаточно велико (от сотни и более). Процедура состоит из двух этапов. *На первом этапе* в итерационном режиме производится последовательное уточнение «центров тяжести» (так называемых «эталонных точек») искомых классов, а *на втором этапе* производится собственно разбиение наблюдений (ПЗ.10) на классы. Поясним это подробнее.

Первый этап. Пусть $e_j(\nu) = (e_j^{(1)}(\nu), e_j^{(2)}(\nu), \dots, e_j^{(p)}(\nu))^T$ — центр тяжести («эталон») j -го класса на ν -й итерации описываемого алгоритма ($j = 1, 2, \dots, k$) и $w_j(\nu)$ — вес, придаваемый эталону $e_j(\nu)$ на ν -й итерации. Соответственно, $E(\nu) = (e_1(\nu), e_2(\nu), \dots, e_k(\nu))$ определяет местоположение эталонов всех k классов на ν -й итерации алгоритма.

При определении нулевого приближения $E(0)$ эталоны $e_j(0)$ желательно выбирать наиболее «взаимоудаленными». По умолчанию нулевое приближение $E(0)$ строится с помощью случайно выбранных первых k точек исследуемой совокупности, то есть

$$\begin{aligned} e_i(0) &= X_i, \\ \omega_i(0) &= 1, \quad i = 1, 2, \dots, k. \end{aligned}$$

Затем на 1-м шаге «извлекается» точка X_{k+1} и выясняется, к какому из эталонов $e_i(0)$ она оказалась ближе всего. Именно этот, самый близкий к X_{k+1} , эталон заменяется эталоном, определяемым как центр тяжести старого эталона и присоединенной к нему точки X_{k+1} (с увеличением на единицу соответствующего ему веса), а все другие эталоны остаются неизменными (с прежними весами) и т. д. Таким образом, пересчет эталонов и весов на ν -м шаге, то есть при извлечении очередной точки $X_{k+\nu}$, происходит по следующему правилу:

$$e_i(\nu) = \begin{cases} \frac{\omega_i(\nu-1)e_i(\nu-1) + X_{k+\nu}}{\omega_i(\nu-1) + 1}, & \text{если} \\ \rho(X_{k+\nu}, e_i(\nu-1)) = \min_{1 \leq j \leq k} \rho(X_{k+\nu}, e_j(\nu-1)), & \\ e_i(\nu-1) & \text{в противном случае,} \end{cases}$$

$$\omega_i(\nu) = \begin{cases} \omega_i(\nu-1) + 1, & \text{если} \\ \rho(X_{k+\nu}, e_i(\nu-1)) = \min_{1 \leq j \leq k} \rho(X_{k+\nu}, e_j(\nu-1)), & \\ \omega_i(\nu-1) & \text{в противном случае,} \end{cases}$$

$$i = 1, 2, \dots, k.$$

При этом если обнаруживается несколько (по i) одинаковых минимальных значений $\rho(X_{k+\nu}, e_i(\nu-1))$, то можно условиться относить точку $X_{k+\nu}$ к эталону с минимальным порядковым номером.

При достаточно большом числе итераций или при достаточно больших объемах классифицируемых совокупностей n и при весьма широких ограничениях на природу исследуемых наблюдений дальнейший пересчет эталонных точек практически не приводит к их изменению, то есть имеет место сходимость (в определенном смысле) $E(\nu)$ к некоторому пределу $E = \{e_1, e_2, \dots, e_k\}$ при $\nu \rightarrow \infty$.

Если же в какой-то конкретной задаче исследователь не успел добраться до стадии практически устойчивых (по ν) значений эталонных точек, то пользуются одним из двух вспомогательных приемов. Либо «зацикливают» алгоритм, «прогоняя» его после рассмотрения последней точки $X_n = X_{k+(n-k)}$ снова через точку X_1 , затем X_2 , и т. д., либо производят многократное повторение алгоритма, используя в качестве

начального эталона $E(0)$ различные комбинации из k точек исследуемой совокупности и выбирая для дальнейшего наиболее повторяющийся (в некотором смысле) финальный эталон $E(n - k)$.

В т о р о й э т а п процедуры состоит непосредственно в разбиении наблюдений (ПЗ.10) на классы: наблюдение X_i относится к классу j , если $\rho(X_i, e_j) \leq \rho(X_i, e_l)$ при всех $l = 1, 2, \dots, k, l \neq j$. Если оказывается, что для некоторого l $\rho(X_i, e_l) = \rho(X_i, e_j)$, то точку X_i относят к тому из классов S_l или S_j , который обладает меньшим порядковым номером.

Проведенные исследования свойств метода k -средних (см., например, [Айвазян, Бежаева, Староверов (1974)]) говорят о том, что в достаточно общих ситуациях при больших объемах классифицируемых наблюдений этот алгоритм строит разбиение, близкое к наилучшему в смысле критерия (ПЗ.20).

Агломеративная иерархическая процедура. Принцип работы таких процедур состоит в последовательном объединении сначала отдельных элементов, а затем и групп элементов, наиболее близких друг к другу. На первом шаге алгоритма каждое наблюдение X_i ($i = 1, 2, \dots, n$) рассматривается как отдельный одноточечный кластер. Далее анализируется $(n \times n)$ -матрица попарных расстояний между элементами $D = (d(X_i, X_j))$ и объединяются в один кластер два самых близких элемента. После этого пересчитывается матрица попарных расстояний (с учетом образования кластера и *принятого способа вычисления расстояния между кластерами*) — ее размерность при этом, очевидно, снижается на единицу, — и таким образом продолжается процесс последовательного объединения двух самых близких друг к другу кластеров. Работа алгоритма заканчивается, когда все исходные наблюдения объединены в один класс. Очевидно одна агломеративная иерархическая процедура отличается от другой лишь выбором способа вычисления расстояния между кластерами (см. выше).

С некоторой точки зрения иерархические процедуры, по сравнению с другими кластер-процедурами, дают более полный и тонкий анализ структуры исследуемого множества наблюдений. Привлекательной стороной подобных алгоритмов является и возможность наглядной интерпретации проведенного анализа в форме так называемых *дендограмм*. Легко себе представить также использование иерархических процедур и для разбиения наблюдений на какое-то объективно обусловленное число классов, заданное или неизвестное. При решении задач первого типа для этого, очевидно, следует продолжать реализацию иерархического алгоритма до тех пор, пока число различных классов не станет равным априори заданному числу k . При решении задач второго типа естественно было бы подчинить правило остановки иерархической процедуры одному из критериев качества разбиения.

К недостаткам иерархических процедур следует отнести громоздкость их вычислительной реализации. Соответствующие алгоритмы на

каждом шаге требуют вычисления всей матрицы расстояний, а следовательно, емкой машинной памяти и большого времени. Поэтому реализация таких алгоритмов при числе наблюдений, большем нескольких сотен, оказывается нецелесообразной.

Кроме того, имеется широкий класс достаточно естественных примеров, в которых иерархические процедуры, даже подчиненные на каждом шаге некоторому критерию качества разбиения, приводят для любого наперед заданного числа кластеров k к разбиению, весьма далекому от оптимального в смысле того же самого критерия качества.

ПЗ.4. Метод главных компонент

1. Общая оптимизационная задача снижения размерности.

В исследовательской и практической статистической работе приходится сталкиваться с ситуациями, когда общее число p признаков $x^{(1)}, x^{(2)}, \dots, x^{(p)}$, регистрируемых на каждом из множества обследуемых объектов (стран, городов, предприятий, семей, пациентов, технических или экологических систем), очень велико — порядка ста и более. Тем не менее имеющиеся многомерные наблюдения

$$X_i = \begin{pmatrix} x_i^{(1)} \\ x_i^{(2)} \\ \vdots \\ x_i^{(p)} \end{pmatrix} \quad i = 1, 2, \dots, n \quad (\text{ПЗ.21})$$

следует подвергнуть статистической обработке, осмыслить либо ввести в базу данных для того, чтобы иметь возможность их использовать в нужный момент.

Желание статистика представить каждое из наблюдений (ПЗ.21) в виде вектора Z некоторых *вспомогательных* показателей $z^{(1)}, z^{(2)}, \dots, z^{(p')}$ с существенно меньшим (чем p) числом компонент p' бывает обусловлено в первую очередь следующими причинами:

- необходимостью *наглядного представления* (визуализации) исходных данных (ПЗ.21), что достигается их проецированием на специально подобранное трехмерное пространство ($p' = 3$), плоскость ($p' = 2$) или числовую прямую ($p' = 1$);
- стремлением к *лаконизму исследуемых моделей*, обусловленному необходимостью упрощения счета и интерпретации полученных статистических выводов;
- необходимостью *существенного сжатия объемов хранимой статистической информации* (без видимых потерь в ее информативности), если речь идет о записи и хранении массивов типа (ПЗ.21) в специальной базе данных;

- стремлением свести анализируемую многокритериальную схему к однокритериальной (в этом случае задаются значением $p' = 1$); именно такого рода задачи возникают при стремлении построить скалярный индикатор для некоторой латентной (то есть не поддающейся непосредственному измерению) синтетической категории по значениям набора характеризующих эту категорию статистических показателей (частных критериев) $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ (речь может идти о таких латентных синтетических категориях, как *уровень коррупции в регионе, уровень человеческого потенциала, уровень материального благосостояния населения региона* и т.п.).

При этом новые (вспомогательные) признаки $z^{(1)}, z^{(2)}, \dots, z^{(p')}$ могут выбираться из числа исходных или определяться по какому-либо правилу по совокупности исходных признаков, например как их линейные комбинации. При формировании новой системы признаков к последним предъявляются разного рода требования, такие, как наибольшая информативность (в определенном смысле), взаимная некоррелированность, наименьшее искажение геометрической структуры множества исходных данных и т.п. В зависимости от варианта формальной конкретизации этих требований приходим к тому или иному алгоритму снижения размерности. Имеется, по крайней мере, три основных типа принципиальных предпосылок, обуславливающих возможность перехода от большого числа p исходных показателей состояния (поведения, эффективности функционирования) анализируемой системы к существенно меньшему числу p' наиболее информативных переменных. Это, во-первых, *дублирование информации, доставляемой сильно взаимосвязанными признаками*; во-вторых, *неинформативность признаков, мало меняющихся при переходе от одного объекта к другому* (малая «вариабельность» признаков); в-третьих, *возможность агрегирования*, то есть простого или «взвешенного» суммирования, по некоторым признакам.

Формально задача перехода (с наименьшими потерями в информативности) к новому набору признаков $\tilde{z}^{(1)}, \tilde{z}^{(2)}, \dots, \tilde{z}^{(p')}$ может быть описана в виде оптимизационной задачи следующим образом. Пусть $Z = Z(X) = (z^{(1)}, z^{(2)}, \dots, z^{(p')})^T$ — некоторая p' -мерная вектор-функция от исходных переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ ($p' \leq p$) и пусть $I_{p'}(Z(X))$ — определенным образом заданная мера информативности p' -мерной системы признаков $Z(X) = (z^{(1)}(X), \dots, z^{(p')}(X))^T$. Конкретный выбор функционала $I_{p'}(Z)$ зависит от специфики решаемой реальной задачи и опирается на один из возможных критериев: критерий *автоинформативности*, нацеленный на максимальное сохранение информации, содержащейся в исходном массиве $\{X_i\}_{i=\overline{1,n}}$ относительно самих исходных признаков; и критерий *внешней информативности*, нацеленный на максимальное «выжимание» из $\{X_i\}_{i=\overline{1,n}}$ информации, содержащейся в этом массиве относительно некоторых других (внешних) показателей.

Задача заключается в определении такого набора признаков $\tilde{Z}$, найденного в классе $\mathbf{F}(X)$ допустимых преобразований исходных показателей $x^{(1)}, \dots, x^{(p)}$, что³

$$I_{p'}(\tilde{Z}(X)) = \max_{Z \in \mathbf{F}(X)} \{I_{p'}(Z(X))\}. \quad (\text{ПЗ.22})$$

Тот или иной вариант конкретизации этой постановки (определяющий конкретный выбор меры информативности $I_{p'}(Z)$ и класса допустимых преобразований) приводит к конкретному методу снижения размерности: к методу главных компонент, факторному анализу, экстремальной группировке параметров и т. д. (см., например, [Айвазян, Мхитарян (2001)]).

Математическая модель, лежащая в основе построения того или иного метода снижения размерности, включает в себя обычно три основных компонента.

Форма задания исходной информации. Речь идет об ответе на следующие вопросы: а) в каком виде задана описательная информация об объектах? б) имеется ли среди исходных статистических данных *обучающая* информация, то есть какие-либо сведения об анализируемом результирующем свойстве? в) если обучающая информация присутствует в исходных статистических данных, то в какой именно форме она представлена? Это могут быть, в частности, в привязке к объекту O_i ($i = 1, 2, \dots, n$): значения «зависимой» количественной переменной («отклика») y_i в моделях регрессии; номер однородного по анализируемому свойству класса, к которому относится объект O_i в задаче классификации; порядковый номер (ранг) объекта O_i в ряду всех объектов, упорядоченных по степени проявления рассматриваемого свойства, в задачах анализа предпочтений и построения упорядоченных типологизаций; наконец, значения $Y_i = (y_i^{(1)}, \dots, y_i^{(q)})^T$ набора результирующих признаков, характеризующих анализируемое в классификационной задаче свойство⁴.

Тип оптимизируемого критерия $I_{p'}(Z)$ информативности искомого набора признаков $Z = (z^{(1)}, \dots, z^{(p')})^T$. Как уже отмечалось, критерий информативности может быть ориентирован на достижение разных целей.

Следует выделить целый класс *критериев автоинформативности*, то есть критериев, оптимизация которых приводит к набору вспомогательных переменных $Z = (z^{(1)}, \dots, z^{(p')})^T$, позволяющих максимально

³Речь идет о нахождении (в виде функции от X) такого вектора $\tilde{Z}(X) = (\tilde{z}^{(1)}(X), \dots, \tilde{z}^{(p')}(X))^T$, который обращает в максимум или минимум (в зависимости от конкретного содержательного смысла оптимизируемого критерия информативности) значение $I_{p'}(Z)$. Поэтому справа в данном соотношении записано extr («экстремум»).

⁴Перечисленные варианты не исчерпывают всех возможных форм представления обучающей информации.

точно воспроизводить (в том или ином смысле, в зависимости от конкретного вида критерия) информацию, содержащуюся в описательном массиве данных типа (ПЗ.21).

Будем называть *критериями внешней информативности* (имеется в виду информативность, внешняя по отношению к информации, содержащейся в описательном массиве (ПЗ.21)) такие критерии $I_{p'}(Z)$, которые нацелены на поиск экономных наборов вспомогательных переменных $Z(X) = (z^{(1)}(X), \dots, z^{(p')}(X))^T$, обеспечивающих максимально точное воспроизведение (по значениям Z , а значит в конечном счете по значениям X) информации, относящейся к результирующему признаку.

Класс $F(X)$ допустимых преобразований исходных признаков X . Вспомогательные признаки $Z = (z^{(1)}, \dots, z^{(p')})^T$ в случае представления исходной описательной информации в форме (ПЗ.21) конструируются в виде функций от X , то есть $Z = Z(X)$. Как обычно в таких ситуациях, чтобы обеспечить содержательность и конструктивную реализуемость решения оптимизационной задачи (ПЗ.22), следует предварительно договориться об ограниченном классе допустимых решений $F(X)$, в рамках которого эта оптимизационная задача будет решаться. Очевидно, от выбора $F(X)$ будет существенно зависеть и получаемое решение $\tilde{Z}(X) = (\tilde{z}^{(1)}(X), \dots, \tilde{z}^{(p')}(X))^T$ упомянутой оптимизационной задачи.

2. Главные компоненты: основные понятия и определения. Во многих задачах анализа обработки многомерных наблюдений исследователя интересуют в первую очередь лишь те признаки, которые обнаруживают наибольшую изменчивость (наибольший разброс) при переходе от одного объекта к другому.

С другой стороны, не обязательно для описания состояния объекта использовать какие-то из исходных, непосредственно замеренных на нем признаков. Так, например, для определения специфики фигуры человека при покупке одежды достаточно назвать значения двух признаков (размер-рост), являющихся производными от измерений ряда параметров фигуры. При этом, конечно, теряется какая-то доля информации (портной измеряет до одиннадцати параметров на клиенте), как бы огрубляются (при агрегировании) получающиеся при этом классы. Однако, как показали исследования, к вполне удовлетворительной классификации людей с точки зрения специфики их фигуры приводит система, использующая три признака, каждый из которых является некоторой комбинацией от большого числа непосредственно замеряемых на объекте параметров.

Именно эти принципиальные установки заложены в сущность того линейного преобразования исходной системы признаков, которое приводит к главным компонентам. Формализуются же эти установки следующим образом.

Следуя общей оптимизационной постановке задачи снижения размерности (ПЗ.22) и полагая анализируемый признак X p -мерной случайной величиной с вектором средних значений $\mathbf{a} = (a^{(1)}, \dots, a^{(p)})$ и ковариационной матрицей $\Sigma = (\sigma_{ij})$ ($i, j = 1, 2, \dots, p$), вообще говоря, неизвестными, определим в качестве класса $\mathbf{F}(X)$ допустимых преобразований исследуемых признаков $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ их всевозможные *линейные центрированные и нормированные комбинации*, то есть

$$\mathbf{F} = \left\{ Z: z^{(j)} = \sum_{\nu=1}^p c_{j\nu} (x^{(\nu)} - a^{(\nu)}), j = 1, 2, \dots, p \right\},$$

где условия нормировки задаются соотношениями

$$\sum_{\nu=1}^p c_{j\nu}^2 = 1 \quad \text{и} \quad \sum_{\nu=1}^p c_{j\nu} c_{k\nu} = 0 \quad (\text{ПЗ.23})$$

для $j = 1, 2, \dots, p$ и $k = 1, 2, \dots, p$, но $j \neq k$, а в качестве критерия (меры) информативности p' -мерной системы показателей $Z(X) = (z^{(1)}(X), z^{(2)}(X), \dots, z^{(p')}(X))$ используется выражение

$$I_{p'}(Z(X)) = \frac{\mathbf{D} z^{(1)} + \dots + \mathbf{D} z^{(p')}}{\mathbf{D} x^{(1)} + \dots + \mathbf{D} x^{(p)}}. \quad (\text{ПЗ.24})$$

Тогда при любом фиксированном $p' = 1, 2, \dots, p$ вектор искомых вспомогательных переменных $\tilde{Z}(X) = (\tilde{z}^{(1)}(X), \dots, \tilde{z}^{(p')}(X))^T$ определяется как линейная комбинация

$$\tilde{Z} = \mathbf{L}X, \quad (\text{ПЗ.25})$$

где

$$\mathbf{L} = \begin{pmatrix} l_{11} & \dots & l_{1p} \\ \dots & \dots & \dots \\ l_{p'1} & \dots & l_{p'p} \end{pmatrix}$$

— матрица, строки которой удовлетворяют условию ортонормированности, такая, что

$$I_{p'}(\tilde{z}^{(1)}(X), \dots, \tilde{z}^{(p')}(X)) = \max_{Z(X) \in \mathbf{F}} I_{p'}(Z(X)).$$

Полученные таким образом переменные $\tilde{z}^{(1)}(X), \dots, \tilde{z}^{(p)}(X)$ и называют главными компонентами вектора X . Отсюда вытекает следующее определение главных компонент.

Первой главной компонентой $\tilde{z}^{(1)}(X)$ *исследуемой системы показателей* $X = (x^{(1)}, \dots, x^{(p)})^T$ *называется такая нормированно-центрированная линейная комбинация этих показателей, которая среди всех*

прочих нормированно-центрированных линейных комбинаций переменных $x^{(1)}, \dots, x^{(p)}$ обладает наибольшей дисперсией.

k -й главной компонентой $\tilde{z}^{(k)}(X)$ ($k = 2, 3, \dots, p$) исследуемой системы показателей $X = (x^{(1)}, \dots, x^{(p)})^T$ называется такая нормированно-центрированная линейная комбинация этих показателей, которая не коррелирована с $k - 1$ предыдущими главными компонентами и среди всех прочих нормированно-центрированных и некоррелированных с предыдущими $k - 1$ главными компонентами линейных комбинаций переменных $x^{(1)}, \dots, x^{(p)}$ обладает наибольшей дисперсией.

З а м е ч а н и е 1 (переход к центрированным переменным). Поскольку, как увидим ниже, решение задачи (а именно вид матрицы линейного преобразования L) зависит только от элементов ковариационной матрицы Σ , которые в свою очередь не изменяются при замене исходных переменных $x^{(j)}$ переменными $x^{(j)} - c^{(j)}$ ($c^{(j)}$ — произвольные постоянные числа), то в дальнейшем будем считать, что исходная система показателей уже *центрирована*, то есть что $E x^{(j)} = 0$, $j = 1, 2, \dots, p$. В статистической практике этого добиваются, переходя к наблюдениям $\tilde{x}_i^{(j)} = x_i^{(j)} - \bar{x}^{(j)}$, где $\bar{x}^{(j)} = \sum_{i=1}^n x_i^{(j)} / n$ (для упрощения обозначений волнистую черту над центрированной переменной и над главной компонентой в дальнейшем ставить не будем).

З а м е ч а н и е 2 (переход к выборочному варианту). Поскольку в реальных статистических задачах располагаем *лишь оценками* $\hat{a}$ и $\hat{\Sigma}$ соответственно вектора средних a и ковариационной матрицы Σ , то во всех дальнейших рассуждениях под $\hat{a}^{(j)}$ понимается $\bar{x}^{(j)}$, а под $\hat{\sigma}_{kj}$ — выборочная ковариация $\hat{\sigma}_{kj} = \sum_{i=1}^n (x_i^{(k)} - \bar{x}^{(k)})(x_i^{(j)} - \bar{x}^{(j)}) / n$ ($j, k = 1, 2, \dots, p$).

З а м е ч а н и е 3. Использование главных компонент оказывается наиболее естественным и плодотворным в ситуациях, в которых все компоненты $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ исследуемого вектора X имеют *общую физическую природу* и соответственно *измерены в одних и тех же единицах*. К таким примерам можно отнести исследование структуры бюджета времени индивидуумов (все $x^{(i)}$ измеряются в единицах времени), исследование структуры потребления семей (все $x^{(i)}$ измеряются в денежных единицах), исследование общего развития и умственных способностей индивидуумов с помощью специальных тестов (все $x^{(i)}$ измеряются в баллах), разного рода антропологические исследования (все $x^{(i)}$ измеряются в единицах меры длины) и т. д. Если же признаки $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ измеряются в различных единицах, то результаты исследования с помощью главных компонент будут существенно зависеть от выбора масштаба и природы единиц измерения. Поэтому в подобных ситуациях исследователь предварительно переходит к вспомогательным безразмерным

признакам $x^{*(i)}$, например, с помощью нормирующего преобразования

$$x_{\nu}^{*(i)} = \frac{x_{\nu}^{(i)}}{\sqrt{\hat{\sigma}_{ii}}} \quad \left(\begin{array}{l} i = 1, 2, \dots, p \\ \nu = 1, 2, \dots, n \end{array} \right),$$

где $\hat{\sigma}_{ii}$ соответствует ранее введенным обозначениям, а затем строит главные компоненты относительно этих вспомогательных признаков X^* и их ковариационной матрицы $\hat{\Sigma}_{X^*}$, которая, как легко видеть, является одновременно выборочной *корреляционной* матрицей $\hat{R}$ исходных наблюдений X_i .

3. Вычисление главных компонент. Из определения главных компонент следует, что для вычисления первой главной компоненты необходимо решить оптимизационную задачу вида (ПЗ.22), то есть в данном случае:

$$\begin{cases} D(l_1 X) \rightarrow \max_{l_1}; \\ l_1 l_1^T = 1, \end{cases} \quad (\text{ПЗ.26})$$

где l_1 — первая строка матрицы L (см. (ПЗ.25)). Учитывая центрированность переменной X (то есть $E X = 0$) и то, что $E(X X^T) = \Sigma$, имеем

$$D(l_1 X) = E(l_1 X)^2 = E(l_1 X X^T l_1^T) = l_1 \Sigma l_1^T.$$

Следовательно, задача (ПЗ.26) может быть записана

$$\begin{cases} l_1 \Sigma l_1^T \rightarrow \max_{l_1}; \\ l_1 l_1^T = 1, \end{cases} \quad (\text{ПЗ.26}')$$

Вводя функцию Лагранжа $\varphi(l_1, \lambda) = l_1 \Sigma l_1^T - \lambda(l_1 l_1^T - 1)$ и дифференцируя ее по компонентам вектор-столбца l_1^T , имеем (см. п. П2.1 в Приложении 2):

$$\frac{\partial \varphi}{\partial l_1^T} = 2 \Sigma l_1^T - 2 \lambda l_1^T,$$

что дает систему уравнений для определения l_1 :

$$(\Sigma - \lambda I) l_1^T = 0 \quad (\text{ПЗ.27})$$

(здесь $0 = (0, 0, \dots, 0)^T$ — p -мерный вектор-столбец из нулей).

Для того чтобы существовало ненулевое решение системы (ПЗ.27) (а оно должно быть ненулевым, так как $l_1 l_1^T = 1$), матрица $\Sigma - \lambda I$ должна быть *вырожденной*, то есть

$$|\Sigma - \lambda I| = 0. \quad (\text{ПЗ.28})$$

Этого добиваются подбором соответствующего значения λ . Уравнение (ПЗ.28) (относительно λ) называется *характеристическим* для матрицы Σ . Известно (см. Приложение 2), что при симметричности и неотрицательной определенности матрицы Σ (каковой она и является как

всякая ковариационная матрица) это уравнение имеет p вещественных неотрицательных корней $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$, называемых *характеристическими* (или *собственными*) *значениями* матрицы Σ .

Учитывая, что $D\tilde{z}^{(1)} = D(l_1 X) = l_1 \Sigma l_1^T$ и $l_1 \Sigma l_1^T = \lambda$ (последнее соотношение следует из соотношения (П3.27) после его умножения слева на l_1 , с учетом $l_1 l_1^T = 1$), получаем $D\tilde{z}^{(1)}(X) = \lambda$.

Поэтому для обеспечения *максимальной* величины дисперсии переменной $\tilde{z}^{(1)}$ нужно выбрать из p собственных значений матрицы Σ *наибольшее*, то есть

$$D\tilde{z}^{(1)}(X) = \lambda_1.$$

Подставляем λ_1 в систему уравнений (П3.27) и, решая ее относительно $l_{11}, \dots, l_{1p}$, определяем компоненты вектора l_1 .

Таким образом, *первая главная компонента* получается как *линейная комбинация* $\tilde{z}^{(1)}(X) = l_1 X$, где l_1 — *собственный вектор* матрицы Σ , *соответствующий* *наибольшему собственному числу* этой матрицы.

Далее аналогично можно показать, что $\tilde{z}^{(k)}(X) = l_k X$, где l_k — *собственный вектор* матрицы Σ , *соответствующий* k -му по величине *собственному значению* λ_k этой матрицы ($k = 2, 3, \dots, p$).

Таким образом, соотношения для определения всех p главных компонент вектора X могут быть представлены в виде (П3.25), где $\tilde{Z} = (\tilde{z}^{(1)}, \dots, \tilde{z}^{(p)})^T$, $X = (x^{(1)}, \dots, x^{(p)})^T$, а матрица L состоит из строк $l_j = (l_{j1}, \dots, l_{jp})$, $j = \overline{1, p}$, являющихся собственными векторами матрицы Σ , соответствующими собственным числам λ_j . При этом сама матрица L в соответствии с условиями (П3.23) является *ортогональной*, то есть

$$LL^T = L^T L = I.$$

В дальнейшем в целях упрощения обозначений мы будем опускать «тильду» над переменными главных компонент, то есть обозначать главные компоненты просто $Z = (z^{(1)}, z^{(2)}, \dots, z^{(p')})$.

4. Основные числовые характеристики главных компонент. Определим основные числовые характеристики (средние значения, дисперсии, ковариации) главных компонент в терминах основных числовых характеристик исходных переменных и собственных значений матрицы Σ :

а) $E Z = E(L X) = L \cdot E X = 0;$

б) ковариационная матрица вектора главных компонент:

$$\begin{aligned} \Sigma_Z &= E(Z Z^T) = E((L X)(L X)^T) = E(L X X^T L^T) = \\ &= L \cdot E(X X^T) \cdot L^T = L \cdot \Sigma \cdot L^T. \end{aligned}$$

Умножая слева соотношения

$$(\Sigma - \lambda_k \mathbf{I}) l_k^\top = \mathbf{0}, \quad k = \overline{1, p},$$

на l_j ($j = \overline{1, p}$), получаем, что

$$\mathbf{L} \Sigma \mathbf{L}^\top = \Sigma_Z = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_p \end{pmatrix} \quad (\text{ПЗ.29})$$

Из (ПЗ.29), в частности, следует подтверждение взаимной некоррелированности главных компонент, а также $\mathbf{D} z^{(k)} = \lambda_k$ ($k = \overline{1, p}$);

- в) сумма дисперсий исходных признаков равна сумме дисперсий всех главных компонент. Действительно,

$$\begin{aligned} \sum_{k=1}^p \mathbf{D} z^{(k)} &= \text{tr } \Sigma_Z = \text{tr } (\mathbf{L} \Sigma \mathbf{L}^\top) = \text{tr } ((\mathbf{L} \Sigma) \mathbf{L}^\top) = \\ &= \text{tr } (\mathbf{L}^\top (\mathbf{L} \Sigma)) = \text{tr } ((\mathbf{L}^\top \mathbf{L}) \Sigma) = \text{tr } \Sigma = \sum_{k=1}^p \mathbf{D} x^{(k)}; \end{aligned}$$

- г) обобщенная дисперсия исходных признаков X (то есть $\det \Sigma_X$) равна обобщенной дисперсии главных компонент. Действительно, обобщенная дисперсия вектора Z равна

$$\begin{aligned} \det \Sigma_Z &= \det (\mathbf{L} \Sigma \mathbf{L}^\top) = \det ((\mathbf{L} \Sigma) \mathbf{L}^\top) = \det (\mathbf{L}^\top (\mathbf{L} \Sigma)) = \\ &= \det ((\mathbf{L} \mathbf{L}^\top) \Sigma) = \det (\Sigma). \end{aligned}$$

Следствие 1. Из б) и в), в частности, следует, что критерий информативности метода главных компонент (ПЗ.24) может быть представлен в виде

$$I_{p'}(Z(X)) = \frac{\lambda_1 + \dots + \lambda_{p'}}{\lambda_1 + \dots + \lambda_p}, \quad (\text{ПЗ.24}')$$

где $\lambda_1, \lambda_2, \dots, \lambda_p$ — собственные числа ковариационной матрицы Σ вектора X , расположенные в порядке убывания.

Кстати, представление $I_{p'}(Z(X))$ в виде (ПЗ.24') дает исследователю некоторую основу, опорную точку зрения, при вынесении решения о том, сколько последних главных компонент можно без особого ущерба изъять из рассмотрения, сократив тем самым размерность исследуемого пространства.

Действительно, анализируя с помощью (ПЗ.24') изменение относительной доли дисперсии, вносимой первыми p' главными компонентами,

в зависимости от числа этих компонент, можно разумно определить число компонент, которое целесообразно оставить в рассмотрении. Так, при изменении $I_{p'}$, изображенном на рис. ПЗ.1, очевидно, целесообразно было бы сократить размерность пространства с $p = 10$ до $p' = 3$, так как добавление всех остальных семи главных компонент может повысить суммарную характеристику рассеяния не более чем на 10%.

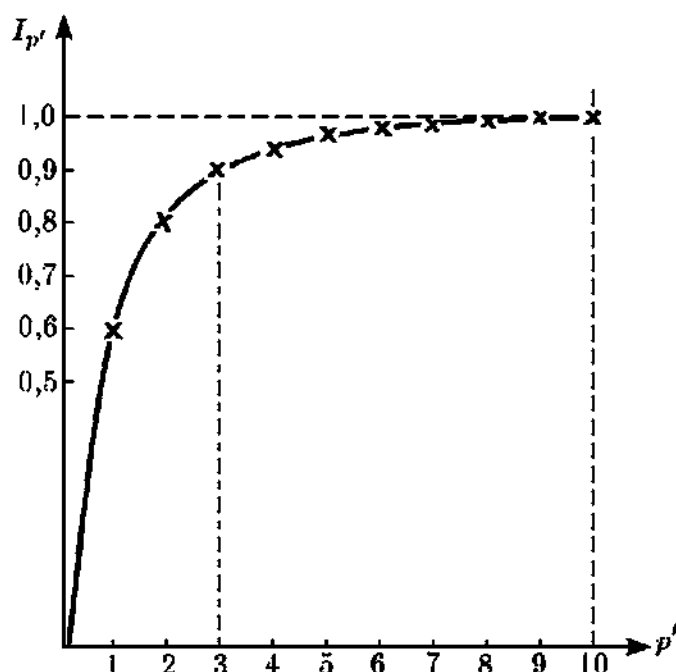


Рис. ПЗ.1 Изменение относительной доли суммарной дисперсии исследуемых признаков, обусловленной первыми p' главными компонентами, в зависимости от p' (случай $p = 10$)

Следствие 2. Если X^* — вектор нормированных признаков $x^{*(1)}, \dots, x^{*(p)}$, то есть $E x^{*(j)} = 0$ и $D x^{*(j)} = 1$ для $j = \overline{1, p}$, то согласно замечанию 3 ковариационная и корреляционные матрицы совпадают (то есть $\Sigma_{X^*} = R$) и из б) и в) следует

$$\text{tr } \Sigma_Z = \text{tr } \Sigma_{X^*} = \text{tr } R = p$$

или

$$\sum_{k=1}^p \lambda_k = p.$$

Тогда критерий информативности (ПЗ.24') может быть представлен в виде

$$I_{p'}(Z(X)) = \frac{\lambda_1 + \dots + \lambda_{p'}}{p}.$$

5. Геометрическая интерпретация главных компонент. Всякий переход к меньшему числу (p') переменных $z^{(1)}, \dots, z^{(p')}$, осуществляемый с помощью ортогонального линейного преобразования (матрицы) $C = (c_{ij})$, $i = 1, 2, \dots, p'$, $j = 1, 2, \dots, p$, можно рассматривать как

проекцию исследуемых p -мерных наблюдений $X_1, X_2, \dots, X_n$ в пространство размерности p' , натянутое на оси $Oz^{(1)}, Oz^{(2)}, \dots, Oz^{(p')}$, где

$$z^{(i)} = \sum_{j=1}^p c_{ij} x^{(j)}, \quad i = 1, 2, \dots, p'.$$

При этом проекциями p -мерных исходных наблюдений X_i ($i = 1, 2, \dots, n$) будут p' -мерные точки

$$Z_i = C X_i, \quad i = 1, 2, \dots, n.$$

Для пояснения сущности того линейного преобразования исходной системы признаков, которое приводит к главным компонентам, рассмотрим его геометрическую интерпретацию на примере двумерной системы наблюдений $(x_i^{(1)}, x_i^{(2)})$, $i = 1, 2, \dots, n$, извлеченной из нормальной генеральной совокупности со средним значением $a = (a^{(1)}, a^{(2)})$ и ковариационной матрицей

$$\Sigma = \begin{pmatrix} \sigma_1^2 & r\sigma_1\sigma_2 \\ r\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \quad |r| \leq 1, \quad \sigma_1 > 0, \quad \sigma_2 > 0.$$

Здесь σ_1^2 и σ_2^2 — дисперсии компонент соответственно $x^{(1)}$ и $x^{(2)}$, а r — коэффициент корреляции между ними. Геометрически это означает, что точки $(x_i^{(1)}, x_i^{(2)})$ будут располагаться примерно в очертаниях эллипсоидов рассеивания вида (см. рис. ПЗ.2а)

$$\frac{1}{1-r^2} \left[\left(\frac{x^{(1)} - a^{(1)}}{\sigma_1} \right)^2 - 2r \left(\frac{x^{(1)} - a^{(1)}}{\sigma_1} \right) \times \right. \\ \left. \times \left(\frac{x^{(2)} - a^{(2)}}{\sigma_2} \right) + \left(\frac{x^{(2)} - a^{(2)}}{\sigma_2} \right)^2 \right] = c^2.$$

В этом случае для изучения расположения точек $(x_i^{(1)}, x_i^{(2)})$ удобно перейти к новым координатам $(z^{(1)}, z^{(2)})$ с помощью преобразования:

$$\begin{aligned} z_i^{(1)} &= (x_i^{(1)} - a^{(1)}) \cos \alpha + (x_i^{(2)} - a^{(2)}) \sin \alpha, \\ z_i^{(2)} &= -(x_i^{(1)} - a^{(1)}) \sin \alpha + (x_i^{(2)} - a^{(2)}) \cos \alpha, \end{aligned}$$

где

$$\operatorname{tg} 2\alpha = \frac{2r\sigma_1\sigma_2}{\sigma_1^2 - \sigma_2^2}.$$

После этого преобразования точки $(z_i^{(1)}, z_i^{(2)})$ также будут распределены нормально, но компонента $z^{(1)}$ уже не будет зависеть от $z^{(2)}$. Кроме

того, если выбрать направления так, что $D z^{(1)} \geq D z^{(2)}$, то геометрически это будет означать следующее: сначала производится перенос начала координат в точку $(a^{(1)}, a^{(2)})$, а затем оси поворачиваются на угол α так, чтобы ось $z^{(1)}$ шла вдоль главной оси эллипсоида рассеивания (рис. П3.2а). Чем ближе $|r|$ к единице, тем теснее группируются наблюдения около главной оси эллипсоида рассеивания (то есть около новой оси $z^{(1)}$) и тем менее значащим для исследователя является разброс точек в направлении оси $z^{(2)}$, а следовательно, и сама эта координата. В предельном случае $|r| = 1$, исследуемые наблюдения в координатах $(z^{(1)}, z^{(2)})$ вообще не отличаются по координате $z^{(2)}$ (см. рис. П3.2б).

6. Оптимальные свойства главных компонент. Описываемые ниже свойства первых p' главных компонент во многом объясняют их широкую распространенность в практике статистических, в том числе эконометрических, исследований. Оказывается, к главным компонентам можно прийти, решая оптимизационные задачи, на первый взгляд не имеющие ничего общего с оптимизационной задачей типа (П3.22).

Свойство наименьшей ошибки «автопрогноза» или наилучшей самовоспроизводимости. Можно показать, что с помощью первых p' главных компонент $z^{(1)}, z^{(2)}, \dots, z^{(p')}$ ($p' < p$) исходных признаков $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ достигается наилучший прогноз этих признаков среди всех прогнозов, которые можно построить с помощью p линейных комбинаций набора из p' произвольных признаков.

Поясним и уточним сказанное. Пусть требуется заменить исходный исследуемый p -мерный вектор наблюдений X на вектор $Z = (z^{(1)}, z^{(2)}, \dots, z^{(p')})^T$ меньшей размерности p' , в котором каждая из компонент являлась бы линейной комбинацией p исходных (или каких-либо других, вспомогательных) признаков, теряя при этом не слишком много информации. Информативность нового вектора Z зависит от того, в какой степени p' введенных вспомогательных переменных дают возможность «реконструировать» p исходных (измеряемых на объектах) признаков с помощью подходящих линейных комбинаций $z^{(1)}, z^{(2)}, \dots, z^{(p')}$. Естественно полагать, что ошибка прогноза X по Z (обозначим ее σ) будет определяться так называемой остаточной дисперсионной матрицей вектора X при вычитании из него наилучшего прогноза по Z , то есть матрицей $\Delta = (\Delta_{ij})$, где

$$\Delta_{ij} = E \left\{ \left(x^{(i)} - \sum_{l=1}^{p'} b_{il} z^{(l)} \right) \left(x^{(j)} - \sum_{l=1}^{p'} b_{jl} z^{(l)} \right) \right\}.$$

Здесь $\sum_{l=1}^{p'} b_{il} z^{(l)}$ — наилучший, в смысле метода наименьших квадратов, прогноз $x^{(i)}$ по компонентам $z^{(1)}, z^{(2)}, \dots, z^{(p')}$ (см. главу 4). Ошибка прогноза X по Z задается как некоторая определенная функция

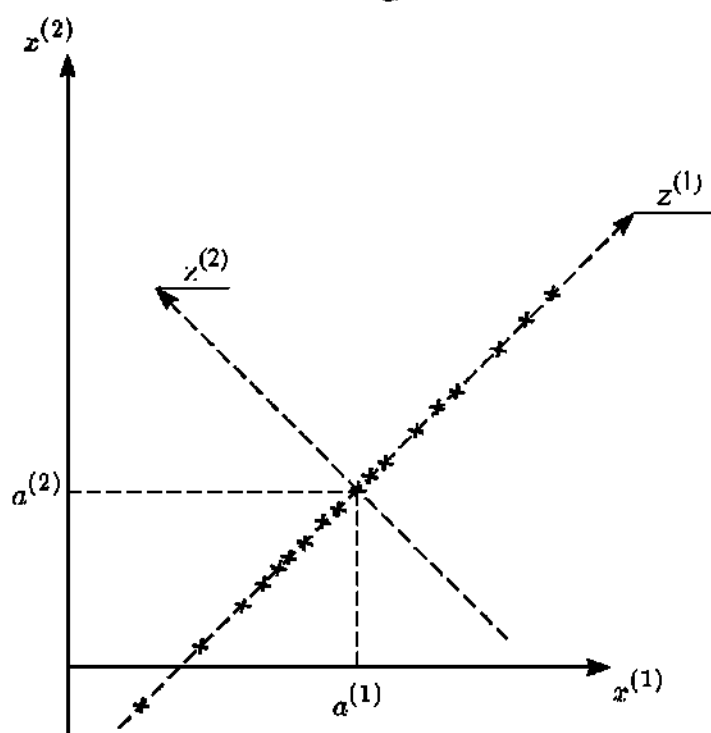
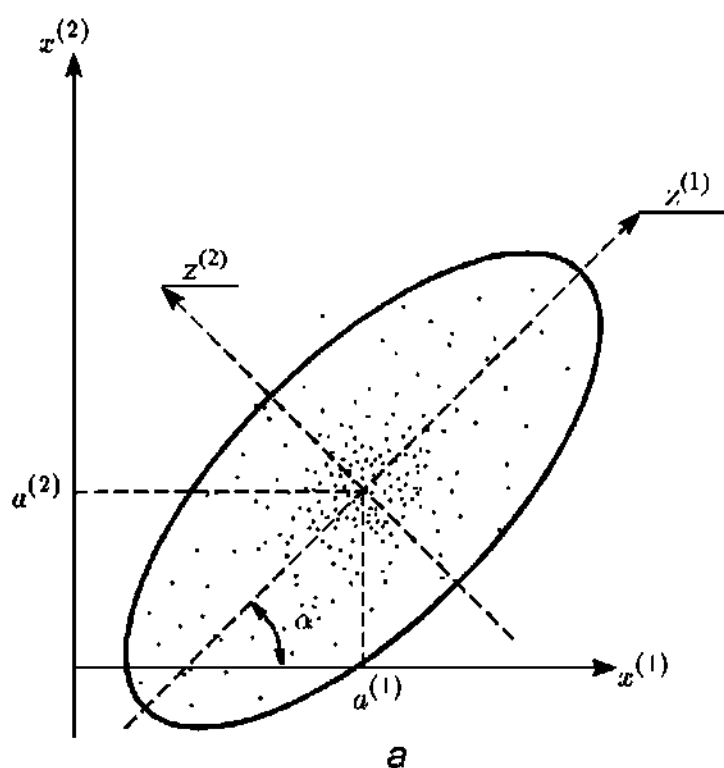


Рис. ПЗ.2 Эллипс рассеяния исследуемых наблюдений и направление координатных осей главных компонент $z^{(1)}$ и $z^{(2)}$: а) умеренный разброс точек; б) отсутствие разброса точек в направлении второй главной компоненты (вырожденный случай)

от элементов матрицы Δ , то есть $\sigma = f(\Delta)$, где $f(\Delta)$ определяет некоторый критерий качества предсказания.

Рассмотрим следующие естественные меры ошибки прогноза:

$$f(\Delta) = \text{tr}(\Delta) = \Delta_{11} + \Delta_{22} + \dots + \Delta_{pp}; \quad (\text{ПЗ.30})$$

$$f(\Delta) = \|\Delta\| = \sqrt{\sum_{i=1}^p \sum_{j=1}^p \Delta_{ij}^2}. \quad (\text{ПЗ.31})$$

Здесь $\text{tr}(\Delta)$ и $\|\Delta\|$ — соответственно *след* и *евклидова норма* матрицы Δ . Доказано, что функции (ПЗ.30) и (ПЗ.31) одновременно достигают минимума тогда и только тогда, когда в качестве $z^{(1)}, z^{(2)}, \dots, z^{(p')}$ выбраны первые p' главных компонент вектора X , причем величина ошибки прогноза σ явным образом выражается через последние $p - p'$ собственных чисел исходной ковариационной матрицы Σ или *приближенно* — через последние $p - p'$ собственных чисел $\lambda_{p'+1}, \dots, \lambda_p$ выборочной ковариационной матрицы $\hat{\Sigma}$, построенной по наблюдениям $X_1, X_2, \dots, X_n$. В частности,

$$\text{при } f(\Delta) = \text{tr}(\Delta): \sigma \approx \lambda_{p'+1} + \lambda_{p'+2} + \dots + \lambda_p;$$

$$\text{при } f(\Delta) = \|\Delta\|: \sigma \approx \sqrt{\lambda_{p'+1}^2 + \lambda_{p'+2}^2 + \dots + \lambda_p^2}.$$

Именно свойство *наименьшей ошибки автопрогноза* первой главной компоненты (то есть при $p' = 1$) делает оправданным и обоснованным ее использование в качестве *интегрального измерителя* (индикатора) такой латентной синтетической категории, которая характеризуется статистическими показателями $x^{(1)}, x^{(2)}, \dots, x^{(p)}$.

Свойства наименьшего искажения геометрической структуры множества исходных p -мерных наблюдений при их проектировании в пространство p' первых главных компонент. Речь идет о следующих трех оптимальных свойствах главных компонент (формулируются без доказательства).

Свойство 1. Сумма квадратов расстояний от исходных точек наблюдений $X_1, X_2, \dots, X_n$ до пространства, натянутого на первые p' главных компонент, наименьшая относительно всех других подпространств размерности p' , полученных с помощью произвольного линейного преобразования исходных координат⁵.

Наглядным пояснением этого свойства может служить рис. ПЗ.2а, на котором ось $z^{(1)}$ соответствует подпространству, натянутому на первую главную компоненту (то есть $p = 2$ и $p' = 1$), а сумма квадратов расстояний до этого подпространства есть сумма перпендикуляров, опущенных из точек, изображающих наблюдения $X_i = (x_i^{(1)}, x_i^{(2)})$, на эту ось.

⁵Гиперплоскость, натянутую на первые p' главных компонент, часто называют p' -мерной плоскостью ортогональной регрессии переменных X .

Свойство 2. Среди всех подпространств заданной размерности p' ($p' < p$), полученных из исследуемого признакового пространства с помощью произвольного линейного преобразования исходных координат $x^{(1)}, x^{(2)}, \dots, x^{(p)}$, в подпространстве, натянутом на первые p' главных компонент, наименее искажается сумма квадратов расстояний между всевозможными парами рассматриваемых точек-наблюдений.

В данном свойстве за критерий наименьшего искажения геометрической структуры совокупности исходных наблюдений $X_1, X_2, \dots, X_n$ принимается величина

$$M_p(X) = \sum_{i=1}^n \sum_{j=1}^n (X_i - X_j)^T (X_i - X_j),$$

то есть сумма квадратов евклидовых расстояний между всевозможными парами имеющихся наблюдений.

После проецирования точек X_i в p' -мерное пространство, определяемое матрицей преобразования C , мы получим точки-проекции $Z_i = C X_i$ ($i = 1, 2, \dots, n$) и соответствующую им сумму квадратов евклидовых расстояний

$$M_{p'}(Z(C)) = \sum_{i=1}^n \sum_{j=1}^n (Z_i - Z_j)^T (Z_i - Z_j).$$

Можно показать, что при $p' < p$ всегда $M_p(X) \geq M_{p'}(Z(C))$.

Так вот, на преобразовании L , с помощью которого получают первые p' главных компонент, достигается минимум разности $M_p(X) - M_{p'}(Z(C))$, то есть

$$M_p(X) - M_{p'}(Z(L)) = \min_C [M_p(X) - M_{p'}(Z(C))].$$

Свойство 3. Среди всех подпространств заданной размерности p' ($p' < p$), полученных из исследуемого признакового пространства с помощью произвольного линейного преобразования исходных координат $x^{(1)}, \dots, x^{(p)}$, в пространстве, натянутом на первые p' главных компонент, наименее искажаются расстояния от рассматриваемых точек-наблюдений до их общего «центра тяжести», а также углы между прямыми, соединяющими всевозможные пары точек-наблюдений с их общим «центром тяжести».

Литература

- Айвазян С. А.* Статистическое исследование зависимостей. — М. : Металлургия, 1968. — 228 с.
- Айвазян С. А.* Об опыте применения экспертно-статистического метода построения неизвестной целевой функции / Многомерный статистический анализ в социально-экономических исследованиях. — М. : Наука, 1974. с. 56—86.
- Айвазян С. А.* Конфлюэнтный анализ / Математическая энциклопедия. М., 1979. Т. 2, с. 1083.
- Айвазян С. А.* Прикладная статистика и основы эконометрики. — 2-е изд. Том 2: Основы эконометрики. — М. : Юнити, 2001. — 432 с.
- Айвазян С. А.* Анализ качества и образа жизни населения (эконометрический подход). — М. : Наука, 2010. — 280 с.
- Айвазян С. А., Бежаева З. И., Староверов О. В.* Классификация многомерных наблюдений. — М. : Статистика, 1974. — 240 с.
- Айвазян С.А., Енюков И.С., Мешалкин Л. Д.* Прикладная статистика. Исследование зависимостей. — М. : Финансы и статистика, 1985. — 488 с.
- Айвазян С. А., Колеников С. О.* Уровень бедности и дифференциация по расходам населения России. — EERC, WP N1, 2001. — 64 с.
- Айвазян С. А., Мхитарян В.С.* Прикладная статистика и основы эконометрики. — 2-е изд. Том 1: Теория вероятностей и прикладная статистика. — М. : Юнити, 2001. — 656 с.
- Андерсон Т.* Введение в многомерный статистический анализ : пер. с англ. — М. : Физматгиз, 1963. — 500 с.
- Берндт Э.* Практика эконометрики: классика и современность : пер. с англ. М. : Юнити, 2005. — 848 с.

- Бокс Дж., Дженкинс Г.* Анализ временных рядов. Выпуск 1: Прогноз и управление : пер. с англ. — М. : Мир, 1974. — 406 с.
- Большев Л. Н., Смирнов Н. В.* Таблицы математической статистики. — М. : Наука, 1965. — 464 с.
- Вербик М.* Путеводитель по современной эконометрии : пер. с англ. — М. : Научная книга, 2008. — 616 с.
- Джонстон Дж.* Эконометрические методы : пер. с англ. — М. : Статистика, 1980. — 446 с.
- Дидэ Э. и др.* Методы анализа данных: подход, основанный на методе динамических сгущений : пер. с франц. — М. : Финансы и статистика, 1985. — 360 с.
- Езекиэл М., Фокс К.* Методы анализа корреляций и регрессий : пер. с англ. — М. : Статистика, 1966. — 380 с.
- Кендалл М. Дж., Стьюарт А.* Статистические выводы и связи : пер. с англ. — М. : Наука, 1973. — 900 с.
- Кендалл М. Дж., Стьюарт А.* Многомерный статистический анализ и временные ряды : пер. с англ. — М. : Наука, 1976. — 736 с.
- Кендэл М.* Временные ряды : пер. с англ. — М. : Финансы и статистика, 1981. — 246 с.
- Магнус Я. Р., Катышев П. К., Пересецкий А. А.* Эконометрика. Начальный курс (7-е изд.). — М. : Дело, 2005. — 504 с.
- Магнус Я. Р., Нейдеккер Х.* Матричное дифференциальное исчисление с приложениями к статистике и эконометрии : пер. с англ. — М. : Физматлит, 2002. — 496 с.
- Носко В. П.* Эконометрика. Введение в регрессионный анализ временных рядов. — М. : Московский физико-технический институт (университет), 2002. — 284 с.
- Себер Дж.* Линейный регрессионный анализ : пер. с англ. — М. : Мир, 1980. — 456 с.
- Суслов В. И., Ибрагимов Н. М., Тальшиева Л. П., Цытлаков А. А.* Эконометрия. — Новосибирск : СО РАН, 2005. — 744 с.
- Тутубалин В.* Эконометрика: образование, которое нам не нужно. — М. : Фазис, 2004. — 168 с.

- Aivazian S.* Mixture-Model Cluster Analysis Using the Projection Pursuit Method. — In: Computational Learning and Probabilistic Reasoning. — John Wiley and Sons, 1996. Pp. 277–286.
- Ameniya T.* Tobit Models: A Survey. — Journal of Econometrics, 24, 1984. Pp. 3–61,
- Beaton A.* Understanding Consumption. — Clarendon Lectures in Economics. Oxford University Press, 1992. — 312 p.
- Dickey D. A., Fuller W.A.* Distribution of the Estimators for Autoregressive Time Series with a Unit Root. — Journal of the American Statistical Association, vol. 74, 1979. Pp. 427–431.
- Enders W.* Applied Econometric Time Series Analysis. — New York : John Wiley and Sons, 1995. — 446 p.
- Greene W. H.* Econometric Analysis (fourth ed.). — Prentice Hall International, Inc., 2000. — 1004 p.
- Hamilton J.* Time Series Analysis. — Princeton University Pless, 1994. — 820 p.
- Hatanaka M.* Time-Series-Based Econometrics. — New York: Oxford University Press, 1996. — 418 p.
- Hayashi F.* Econometrics. — Princeton and Oxford : Princeton University Press, 2000. — 684 p.
- Heckman J.J.* Varieties of Selection Bias. — American Economic Review, 80, 1990. Pp. 313–318.
- Johnston J., DiNardo J.* Econometric Methods (fourth ed.). — McGraw-Hill Comp., Inc., 1997. — 532 p.
- Maddala G.* Introduction to Econometrics. — New York : McGraw-Hill Companies, Inc., 1988. — 462p.
- Manski C. F.* Anatomy of the Selection Problem. — The Journal of Human Resources, 24, 1989. Pp. 243–260.
- Manski C. F.* Identification Problem in Econometrics. — Cambridge : Harvard University Press, 1995. — 308 p.
- Pindyck R., Rubinfeld D.L.* Econometric Models and Economic Forecasts. — Tokio : McGraw-Hill Kogakusha, Ltd, 1976. — 470 p.
- RLMS.* The Russia Longitudinal Monitoring Survey: “Family Questionnaire” and “Sample of Russian Federation Rounds V and VI” (Technical Report), 1996. — 336 p.

RLMS. The Russia Longitudinal Monitoring Survey 1992–1996. *Mroz T., Popkin B., Mancini D., Glinskaya N., Lokshin V.* Monitoring Economic Conditions in the Russian Federation. — Report Submitted to the U.S. Agency for International Development, University of North Caroline at Chapel Hill, 1997. — 286 p.

Алфавитно-предметный указатель

А

Автоинформативность 478, 479
Автоковариационная функция 303
Автокорреляционная функция 304, 372
— — марковского процесса 339, 340
— — процесса Юла 344, 346
— — частная 305, 341, 347
Автопрогноз 395, 488
Авторегрессия 1-го порядка 338
Агломеративная иерархическая процедура 476
Адаптивные методы прогнозирования 402
Аддитивная линейная форма 23
Аддитивная модель сезонности 408
Аддитивное разложение временного ряда 297, 301
Алгебраический полином 56
Алгебраическое дополнение 88, 440
АР(p)-модели 338, 349
АРПСС(p, q, λ)-модели 388–391
АРСС(p, q)-модели 359
АРУГ(p)-модель 376

Б

Байесовское правило классификации 464
Белый шум 336
— — стандартизированный нормальный 376

Бинарные (булевы, дихотомические) переменные 255, 273, 472
Блочная структура матриц 451

В

Вариация выборочной функции регрессии 141
Верификация модели 30
Вероятность ошибочной классификации 464
Взвешенное евклидово расстояние 471
Визуализация данных 477
Временной ряд 293
Временной такт 294
Выбор общего вида модели 38, 42, 55–64
Выборочной селективности проблема 277, 290

Г

Гармоника 312
Гармоническая составляющая 310
Гаусса–Маркова теорема 136
Генезис наблюдений 295, 296
Геометрическая структура исходных данных 56, 58
Гетероскедастичность 180, 188
Главные компоненты 480, 481
Гомоскедастичность 28, 123
Гребневая (ридж-) регрессия 149

Д

Двумерный нормальный закон
распределения 456
Дендограмма 476
Детерминант (определитель) 439
Дискриминантная функция 467
Дискриминантный анализ 463
— — параметрический 466–468
— — непараметрический 466
Дихотомические (бинарные)
переменные 255, 273, 472
— — — результирующие 273
Доверительная вероятность 140
— область 214, 215
Доверительный интервал 140
ДС-ряд 378, 379

З

Зависимость детерминированная
51
— марковского типа 338–343
— регрессионная 42, 51–52, 122
— статистическая 33–35
— структурного типа 54

И

Идентификация модели 27, 29, 338
— — авторегрессии 342–343,
348–349
— — АРПСС 390–391
— — АРСС 362, 365
— — скользящего среднего
355–358
Идентифицируемости проблема 29
Иерархические процедуры
классификации 476
Инструментальные переменные
239, 246, 249
Интегральный индикатор 490

Интервальное оценивание 140, 215
Информационная характеристика
связи 114–115
Информационное расстояние
Каллбэка 473

К

Категоризованные переменные 96,
111
Класс (кластер) 463, 469–470
Классификация 285–286
— без обучения 468–476
— с обучением 463–468
Классическая линейная модель
множественной регрессии
(КЛММР) 121, 124
Кластер 469–470
Кластер-анализ 266, 468
Ковариационная матрица 441,
455–456
— — вектора остатков 124
— — МНК-оценок 134
Ковариационная структура 368
Ковариационный анализ 269
Конфлюентный анализ 54
Корни характеристического
уравнения 340, 346, 349, 353, 354,
446
Коррелограмма 304
Корреляционная матрица 88, 440
Корреляционное отношение 82
Корреляционный анализ 67–68
Коэффициент автокорреляции 304
— адаптации (сглаживания) 403
— детерминации 70–74, 140–144
— дисконтирования 403
— квадратичной сопряженности
114
— конкордации (согласованности)
107
— корреляции 75–79
— — выборочный 77

- — множественный 90–94
- — ранговый (см. «Ранговый коэффициент корреляции»)
- — частный 87–89
- Критерий автоинформативности 479–480
- Бартлетта 196
- Бреуша–Пагана 194–196
- Вальда 170–172
- внешней информативности 480
- Дёрбина–Уотсона 201
- Дики–Фуллера (DF-тест) 385
- Дики–Фуллера расширенный (ADF-тест) 387
- информативности 479
- качества аппроксимации (адекватности модели) 35, 40
- качества разбиения на классы 473
- логарифма отношения правдоподобия 171
- множителей Лагранжа 170–171, 173–175
- однородности дисперсий 196
- отношения правдоподобия 171, 173
- Чоу однородности двух групп наблюдений 264

Л

- Лаговые переменные 31
- Латентный показатель 18, 283
- Линейная модель множественной регрессии 23, 28
- — — — классическая 121–124
- — — — обобщенная 179–180
- Линейные комбинации нормально распределенных величин 463
- Линейные ограничения на параметры модели 162–170

- Линейные ортогональные нормальные комбинации 481
- Логарифмическая функция правдоподобия 130, 462
- Логит-модель 273

М

- Макроуровень 21
- «Манекены» (переменные) 254–255
- Матрица 433
- единичная 434
- идемпотентная 450
- квадратная 433–434
- невырожденная 440
- неотрицательно определенная 446–447
- обратная 441
- полного ранга 445
- положительно определенная 446–447
- симметрическая 434
- транспонированная 435
- Махаланобиса расстояние 473
- Мезоуровень 21
- Метод всех возможных регрессий 154
- главных компонент, см.
- Главные компоненты
- инструментальных переменных 239
- k -средних 267, 474
- максимального правдоподобия 129
- моментов 338
- наименьших квадратов (МНК) 126
- — — взвешенный 188–189
- — — обобщенный 183–184
- последовательных разностей 332

— пошагового отбора переменных (пошаговая регрессия) 156
— — — практически реализуемый 188, 193, 204
Методы сглаживания 314, 319
— — алгоритмические 321–330
— — аналитические 320
— — экспоненциально взвешенные 330–332
Микроуровень 21
Многомерное наблюдение 67
— шкалирование 18
Многомерный статистический анализ 455
Множественная линейная регрессия 121–122, 179–180
Множественный коэффициент корреляции 90–94
Модели нестационарных временных рядов 378
— сезонных рядов 391, 408
— стационарных временных рядов 336
Модель авторегрессии, см. АР-модели
АР-модели
— — со скользящими средними в остатках, см. АРСС-модели
— бинарного выбора 273
— Бокса–Дженкинса, см. АРПСС-модели
АРПСС-модели
— множественного выбора 272, 282–284 — — с переменной структурой 251–252
— с дискретно-непрерывной зависимой переменной 287–291
— скользящего среднего, см. СС-модели
СС-модели
— смеси распределений 464, 468
— эконометрическая 22, 28
Мультиколлинеарность 145

Н

Невязки 126
Независимость случайных величин 457, 461
Непараметрический дискриминантный анализ 466
Неразличимые («связные», «объединенные») ранги 97
Неслучайная составляющая временного ряда 314, 319
Нестационарный временной ряд 378, 388
Нормальность регрессионных остатков 130
Нормальный (гауссовский) закон распределения 455
Нормальных уравнений система 127
Нормирование 45

О

ОАРУГ (p, q) -модель 377
Обобщенный коэффициент корреляции 104
Обобщенный метод наименьших квадратов 184
Обращение матрицы 441
— — блочной 453
Обучающая выборка 232, 285, 463
Объясняемая (результатирующая) переменная 34, 42
Объем выборки 35
Объясняющая переменная 34, 42
Обычное евклидово расстояние 471
Одномерный временной ряд 294
Однородность объектов 470
Оператор авторегрессии 366, 389
— сдвига по времени 366
— скользящего среднего 366–367, 389

— упрощающий 391–392
Определитель (детерминант)
блочной матрицы 452
— — матрицы 439
Оптимальная (байесовская)
процедура классификации 464
Оптимальность МНК-оценок 136
Основные типы зависимостей
50–55
Ошибки спецификации модели 159

П

Параметр (коэффициент)
адаптации 403
Параметрические регрессионные
схемы 55
Параметрический
дискриминантный анализ 466
Парная линейная регрессия 127
Показатели качественные
(порядковые, ординальные) 96
— классификационные
(номинальные) 112
— количественные 69
Последовательные разности 332
Постоянство отдачи от масштаба
производства 162
Предикторы в модели регрессии
34
Предопределенные переменные 27
Пробит-модель 273–274
Проблема группового выбора
(упорядочения) 98–100
Проверка регрессионной
однородности исходных данных
263
— статистической независимости
последовательности наблюдений
314
Прогноз 43, 46, 213
— долгосрочный 294, 395

— значений результирующего
показателя 213
— интервальный 213, 220
— краткосрочный 395
— линейный 216
— ретроспективный 229
— среднесрочный 395
— точечный 213, 214
Производственные функции 162
Процедуры классификации
463–476
Прямое (кронекерово)
произведение матриц 438

Р

Равноотстоящие наблюдения вре-
менного ряда 294
Ранг матрицы 444
— объекта 96
Ранговая корреляция 98
— переменная 96
Ранговый коэффициент
корреляции 100, 102
— — — Кендалла 102
— — — случай связанных
рангов 103
— — — Спирмена 100
— — — случай связанных
рангов 100–101
Распределения вероятностей закон
— — — нормальный
(гауссовский) 413–415
— — — — двумерный 456
— — — — многомерный 124,
138, 139, 455, 459–461
— — — Стьюдента (t) 140, 422
— — — F (Фишера) 168, 170, 418
— — — «хи-квадрат» (χ^2) 139,
168, 416
Рассеяние внутриклассовое 474

- Расстояние между классами объектов 472
 — — — — «ближнего соседа» 472
 — — — — «дальнего соседа» 472
 — — — — информационное (Каллбэка) 473
 — — — — Махаланобиса 473
 — — — — обобщенное (по Колмогорову) 473
 — — — — «средний связи» 473
 — — — — «центров тяжести» 472
 — — объектами 470
 — — — взвешенное евклидово 471
 — — — евклидово 471
 — — — хеммингово 472
 Расщепление смеси распределений 468–469
 Реалистическая ситуация (в анализе регрессионной модели) 226
 Регрессионная зависимость 42, 65
 — неоднородность данных 251–252
 Регрессионные остатки 39, 64, 121–124
 Регрессия 42
 Результирующая переменная 34, 42
 Рекуррентная формула 344, 349–350
 Ретроспективная оценка ошибки прогноза 229
 Решающее правило 464
 Ридж-регрессия 149
 Ряд временной, см. Временной ряд
- С**
- Сбор статистических данных 27
 Сглаживание 314, 319
 Сезонная составляющая 391
 Сжатие массивов информации 477
 Симметрическая матрица 434
 Система одновременных уравнений 19, 31, 50
 Систематическая ошибка 24
 Ситуационный анализ 50
 След матрицы 440
 Смесь распределений вероятностей 464, 468
 — нескольких регрессионно однородных подвыборок 265
 Снижение размерности исследуемого пространства 477
 Сно-матрицы 446
 Собственное (характеристическое) значение (число) матрицы 445–446
 Собственный (характеристический) вектор матрицы 446
 Сопутствующие переменные 251–252
 Спектр (спектральная плотность) 309–310
 Спектральный анализ временного ряда 310
 Спецификация модели 28, 122
 Спо-матрицы 446
 Средняя мера внутриклассового рассеяния 474
 Стандартная нормальная функция распределения 414
 — форма нормальных уравнений 127
 Статистическая независимость случайных величин 457, 461
 Статистическая проверка гипотез 140, 167, 263, 314, 385, 387
 Статистическое исследование зависимостей 33–34
 Стационарность слабая (в широком смысле) 313
 — строгая (в узком смысле) 302
 Степень согласованности мнений группы экспертов 99–100

— тесноты статистической связи
68, 70

Стохастические объясняющие
переменные 233–234

Структурные параметры
(характеристики) модели 370

Т

Таблица (матрица)

«объект–свойство» 97

Таблица сопряженности 112

Тейла–Вейджа модель, см.

Аддитивная модель сезонности

Тобит-модель, см. Модель с
дискретно-непрерывной зависимой
переменной

Точечный прогноз

результатирующего показателя
213–214

Точность регрессионной модели
214–216

Транспонирование матрицы 435

— произведения матриц 437

Тренд временного ряда

— детерминированный 296, 378

— стохастический 378–379

TS-ряд 378

У

Уайта оценки стандартных
ошибок МНК-оценивания 197

Упрощающие операторы 391–392

Уравнение характеристическое
матрицы 446

— — модели авторегрессии 340,
346, 349

— — скользящего среднего 352,
353, 354

Уравнения Юла–Уокера 350

Усеченное нормальное
распределение 288–289

Условия обратимости временного
ряда 349

Условно гетероскедастичные
остатки, см. АРУГ-модели

Условно-оптимизационная задача
163, 165

Ф

Факторы, формирующие значения
временного ряда 296

Фиктивные переменные
(«манекены») 254–255

Функционал (критерий) качества
метода или модели 35, 40, 126, 186,
474

Функциональная зависимость 51,
93

Функция потерь 464

— правдоподобия 130, 171

— — логарифмическая 130, 462

— регрессии 42, 65

Х

Характеристические
(собственные) векторы 446

— — числа (значения) 445–446

Характеристические уравнения
340, 346, 349, 353, 354, 446

Ц

Циклическая (конъюнктурная)
составляющая временного ряда
296

Ч

Частный коэффициент
корреляции 87
— (маржинальный) закон
распределения вероятностей 457,
460

Ш

Шкалирование многомерное, см.
Многомерное шкалирование

Э

Экзаменуемая выборка 227, 232
Экзогенные переменные 26, 31
Эконометрика 19, 20, 30
Эконометрическая модель 22
Экономическая статистика 19, 20
— теория 19
Экспертное упорядочение
объектов 99–100
Экспертные оценки 99
Экспоненциальное сглаживание
330
Экстраполяция временного ряда
Экстремальные (оптимальные)
свойства главных компонент 488
Эндогенные переменные 26, 31
Этап параметризации
регрессионной модели 27
Этимология слова «регрессия» 42
Эффективность МНК-оценок 136

Сведения об авторе



АЙВАЗЯН Сергей Артемьевич

Известный российский ученый, доктор физико-математических наук, профессор, заслуженный деятель науки РФ, автор ряда фундаментальных научных работ, учебников по статистике и эконометрике.

Сфера научных интересов: теоретическая и прикладная эконометрика, многомерный статистический анализ, вероятностно-статистическое моделирование механизмов социально-экономических явлений и процессов (анализ и измерение латентных синтетических категорий качества жизни населения, распределительных отношений в обществе, межстрановой и межрегиональный

эконометрический анализ).

Звания, ученые степени, премии, награды

Академик (иностранный член) Национальной академии наук Республики Армения (2008).

Лауреат Премии Совета министров СССР (1986); Премия и медаль Французского национального конгресса статистиков (1986); медаль Европейского эконометрического общества (1988); Премия Международного научного фонда экономических исследований им. Н. П. Федоренко «За выдающийся вклад в развитие экономической науки России» (2007), орден «За вклад в просвещение» (2010).

Почетный профессор Российско-Армянского (Славянского) государственного университета (2005), заслуженный профессор МГУ имени М. В. Ломоносова (2010).

Научно-организационная и педагогическая деятельность

Заместитель директора ЦЭМИ РАН, научный руководитель Отделения эконометрики и прикладной статистики, Лаборатории вероятностно-статистических методов и моделей в экономике ЦЭМИ РАН.

Заведующий кафедрой эконометрики и математических методов экономики Московской школы экономики МГУ имени М. В. Ломоносова; заведующий кафедрой эконометрики Московской финансово-промышленной академии; профессор Российской экономической школы (1992—2004); профессор Государственного университета — Высшей школы экономики; приглашенный профессор Экономико-математического института университета Дижона (Франция, 1980) и Лондонского Холлоувэй Королевского университета (1996—1998).

Главный редактор журнала «Прикладная эконометрика», заместитель главного редактора журнала «Экономика и математические методы», член редакционного совета журнала «Теория вероятностей и ее применения»; научный руководитель постоянно действующего (с 1969 г.) семинара «Многомерный статистический анализ и вероятностное моделирование реальных процессов».

Учебное издание

Айвазян Сергей Артемьевич

МЕТОДЫ ЭКОНОМЕТРИКИ

Учебник