

22.17
А 36



ПРИКЛАДНАЯ СТАТИСТИКА

ОСНОВЫ ЭКОНОМЕТРИКИ

в двух томах

том 2

С. А. Айвазян

ОСНОВЫ ЭКОНОМЕТРИКИ



ПРИКЛАДНАЯ
СТАТИСТИКА

ОСНОВЫ
ЭКОНОМЕТРИКИ

В двух томах

ИЗДАНИЕ ВТОРОЕ, ИСПРАВЛЕННОЕ

S. A. Aivazian

ESSENTIALS
of
ECONOMETRICS

Volume **2**

Textbook

С. А. Айвазян

ОСНОВЫ ЭКОНОМЕТРИКИ

Том **2**

*Рекомендовано Министерством общего
и профессионального образования Российской
Федерации в качестве учебника для студентов
экономических специальностей высших
учебных заведений*



Москва

2001

УДК 311+330.43(075.8)

ББК 60.6+65вб73

П85

Первое издание данного учебника подготовлено в рамках Проекта Tacis "Преподавание экономических и бизнес-дисциплин в средних школах, технических и классических университетах", реализованного Фондом экономических исследований Университета "Эразмус", Роттердам (SEOR/EUR) и Государственным университетом — Высшей школой экономики

Рецензенты:

кафедра статистики С.-Петербургского государственного университета экономики и финансов; проф. Э.Б. Ершов (ГУ ВШЭ) и проф. Я. Маэнус (Университет Тилбурга)

Главный редактор издательства *Н.Д. Эриашвили*

П85 **Прикладная статистика. Основы эконометрики: Учебник для вузов: В 2 т.**
2-е изд., испр. — Т. 2: *Айвазян С.А. Основы эконометрики.* — М.: ЮНИТИ-ДАНА, 2001. — 432 с.
ISBN 5-238-00305-6

Содержание и стиль изложения в учебнике соответствуют принятым Министерством образования РФ стандартам и учебным программам высших учебных заведений экономического профиля по дисциплине "Эконометрика". Представленные в учебнике *методы и модели регрессионного анализа, анализа временных рядов и систем одновременных уравнений* могут составить содержание двух семестровых курсов (по формуле 32 часа лекций и 32 часа упражнений каждый), один из которых ("Эконометрика-1") рекомендуется включать в учебные планы 3-го или 4-го года 1-й ступени высшего экономического образования (бакалавриата), а второй, более продвинутый ("Эконометрика-2"), — в учебные планы 1-го или 2-го года магистратуры (т.е. 5-го или 6-го года обучения).

В данном томе существенно используются обозначения, понятия и результаты 1-го тома данного издания ("Теория вероятностей и прикладная статистика"), однако возможно и его автономное усвоение.

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

ББК 60.6+65вб.73

Учебник

Айвазян Сергей Артемьевич

ОСНОВЫ ЭКОНОМЕТРИКИ

Корректор *Л.И. Ганина*

Оформление художника *А.В. Лебедева*

Лицензия серия ИД № 03562 от 19.12.2000 г. Подписано в печать 10.08.2001
Формат 70х100 1/16. Усл. печ. л. 35. Уч.-изд. л. 29. Тираж 20 000 экз. (1-й завод — 5000). Заказ 1644

ООО «ИЗДАТЕЛЬСТВО ЮНИТИ-ДАНА». Генеральный директор *В.Н. Закаидзе*

123298, Москва, ул. Ирины Левченко, 1—9. Тел. (095) 194-00-15. Тел/факс (095) 194-00-14
www.unity-dana.ru E-mail: unity@msm.ru

Отпечатано во ФГУП ИПК «Ульяновский Дом печати». 432980, г. Ульяновск, ул. Гончарова, 14

Качество печати соответствует предоставленным оригиналам

ISBN 5-238-00305-6

© С.А. Айвазян, 1998, 2001
© «ИЗДАТЕЛЬСТВО ЮНИТИ-ДАНА», 1998, 2001
Воспроизведение всей книги или любой ее части запрещается без письменного разрешения издательства

О Г Л А В Л Е Н И Е

Вступительное слово	10
Предисловие к первому изданию	11
Предисловие ко второму изданию	14
Глава 1. Эконометрика и эконометрическое моделирование: основные понятия и определения	17
1.1. Эконометрика и ее место в ряду математико-статистических и экономических дисциплин	17
1.2. Эконометрическая модель и проблемы эконометрического мо- делирования	20
1.2.1. От простых количественных взаимосвязей между экономическими переменными к эконометрической модели	20
1.2.2. Основные понятия эконометрического моделирования .	25
1.2.3. Основные проблемы эконометрического моделирования	31
1.3. Математико-статистический инструментарий эконометрики .	37
1.3.1. Традиционный состав математико-статистических ме- тодов эконометрики	37
1.3.2. Некоторые задачи социально-экономической теории и практики, решение которых требует методов приклад- ной статистики, выходящих за рамки традиционного эконометрического инструментария	38
Выводы	39

Глава 2. Модели и методы регрессионного анализа	42
2.1. Введение в регрессионный анализ (основные понятия и определения)	42
2.1.1. Результирующая (зависимая, эндогенная) переменная y	43
2.1.2. Объясняющие (предикторные, экзогенные) переменные X	43
2.1.3. Функция регрессии y по X	43
2.1.4. Уравнения регрессионной связи между y и X	45
2.1.5. Измеритель степени тесноты статистической связи между y и X	46
2.1.6. Исходные статистические данные	47
2.1.7. Основные задачи прикладного регрессионного анализа.	48
2.2. Классическая линейная модель множественной регрессии (КЛММР)	49
2.3. Оценивание неизвестных параметров КЛММР: метод наименьших квадратов и метод максимального правдоподобия	53
2.3.1. Метод наименьших квадратов (МНК)	54
2.3.2. Метод максимального правдоподобия (ММП)	57
2.3.3. Анализ вариации результирующего показателя y и выборочный коэффициент детерминации $\hat{R}_{y,x}^2$	59
2.3.4. Статистические свойства оценок параметров КЛММР.	62
2.4. Мультиколлинеарность и отбор наиболее существенных объясняющих переменных в КЛММР	74
2.4.1. Признаки и причины мультиколлинеарности	74
2.4.2. Методы устранения мультиколлинеарности	77
2.5. Ошибки спецификации модели	89
2.6. Обобщенная линейная модель множественной регрессии (ОЛММР) и обобщенный метод наименьших квадратов (ОМНК)	93
2.6.1. Обобщенная линейная модель множественной регрессии (ОЛММР)	93
2.6.2. Обобщенный метод наименьших квадратов (ОМНК)	97
2.7. ОЛММР с гетероскедастичными остатками	102
2.7.1. Сравнение ОМНК- и МНК-оценок в моделях регрессии с гетероскедастичными остатками	105
2.7.2. Некоторые практические рекомендации по анализу модели регрессии с гетероскедастичными остатками	107
2.8. ОЛММР с автокоррелированными остатками	111
2.8.1. Искажение характеристик точности МНК-оценок, обусловленное игнорированием автокоррелированности остатков	114
2.8.2. Проверка гипотезы о наличии/отсутствии автокоррелированности регрессионных остатков	117

2.8.3. Некоторые практические рекомендации по анализу модели регрессии с автокоррелированными остатками . . .	119
2.9. Практические рекомендации по построению, анализу и интерпретации регрессионной модели	120
2.9.1. Построение и анализ обобщенной ЛММР при неизвестной ковариационной матрице регрессионных остатков (практически реализуемый ОМНК)	120
2.9.2. Точечный и интервальный прогноз, основанный на моделях линейной регрессии	123
2.9.3. Исследование точности регрессионной модели в реалистической ситуации	132
2.9.4. Анализ эластичностей с использованием моделей регрессии	136
2.10. Линейные модели регрессии со стохастическими объясняющими переменными	138
2.10.1. Случайные остатки ϵ не зависят от предикторов X и оцениваемых коэффициентов регрессии Θ	140
2.10.2. Общий случай: стохастические предикторы X коррелированы с регрессионными остатками ϵ . Метод инструментальных переменных	144
2.10.3. Случайные ошибки в измерении значений объясняющих переменных	150
2.11. Линейные регрессионные модели с переменной структурой (построение линейной модели по неоднородным регрессионным данным)	155
2.11.1. Проблема неоднородных (в регрессионном смысле) данных	156
2.11.2. Введение «манекенов» (фиктивных переменных) в линейную модель регрессии	159
2.11.3. Проверка регрессионной однородности двух групп наблюдений (критерий Г. Чоу)	167
2.11.4. Построение КЛММР по неоднородным данным в условиях, когда значения сопутствующих переменных неизвестны	170
2.12. Нелинейные модели регрессии и линеаризация	172
2.12.1. Некоторые виды нелинейных зависимостей, поддающиеся непосредственной линеаризации	173
2.12.2. Подбор линеаризующего преобразования (подход Бокса-Кокса)	180
2.13. Дихотомические (бинарные) результирующие показатели и связанные с ними логит- и пробит-модели	187
Выводы	193

Глава 3. Анализ временных рядов (модели и прогнозирование)	199
3.1. Временной ряд (определения, примеры, формулировка основных задач)	201
3.2. Стационарные временные ряды и их основные характеристики	208
3.3. Неслучайная составляющая временного ряда и методы его сглаживания	220
3.3.1. Проверка гипотезы о неизменности среднего значения временного ряда	221
3.3.2. Методы сглаживания временного ряда (выделение неслучайной составляющей)	227
3.3.3. Подбор порядка аппроксимирующего полинома с помощью метода последовательных разностей	240
3.4. Модели стационарных временных рядов и их идентификация	244
3.4.1. Модели авторегрессии порядка p (АР(p)-модели)	246
3.4.2. Модели скользящего среднего порядка q (СС(q)-модели)	261
3.4.3. Авторегрессионные модели со скользящими средними в остатках (АРСС(p, q)-модели)	270
3.4.4. Простая и обобщенная модели авторегрессионных условно гетероскедастичных остатков	284
3.5. Модели нестационарных временных рядов и их идентификация	287
3.5.1. Модель авторегрессии — проинтегрированного скользящего среднего (АРПСС(p, q, k)-модель)	287
3.5.2. Модели рядов, содержащих сезонную компоненту	291
3.5.3. Регрессионные модели с распределенными лагами	296
3.6. Прогнозирование экономических показателей, основанное на использовании моделей временных рядов	312
3.6.1. Прогнозирование на базе АРПСС-моделей	313
3.6.2. Адаптивные методы прогнозирования	320
Выводы	327
Глава 4. Системы линейных одновременных уравнений	331
4.1. Модель спроса-предложения как пример системы одновременных уравнений	332
4.2. Условия идентифицируемости уравнений системы	339
4.3. Идентификация систем одновременных уравнений (статистическое оценивание неизвестных значений параметров системы)	348
4.3.1. Идентификация рекурсивных систем	349
4.3.2. Косвенный метод наименьших квадратов	353
4.3.3. Двухшаговый метод наименьших квадратов оценивания структурных параметров отдельного уравнения	355

4.3.4. Трехшаговый метод наименьших квадратов одновременной оценки всех параметров системы	360
4.3.5. Другие методы оценивания СОУ и некоторые общие рекомендации	364
4.4. Точечный и интервальный прогноз значений эндогенных переменных	366
4.5. Некоторые общие подходы к анализу точности оценивания и к сравнению методов и моделей	374
Выводы	380
Приложение 1. Таблицы математической статистики	384
Приложение 2. Необходимые сведения из матричной алгебры	404
Литература	426
Алфавитно-предметный указатель	428

ВСТУПИТЕЛЬНОЕ СЛОВО

Студенты, изучающие эконометрику, несомненно, почувствуют облегчение с выходом этого тома. Это первый истинно российский учебник по основам эконометрики, написанный на уровне мировых стандартов. Эконометрика, как известно, является обязательным компонентом в джентльменском наборе современного экономиста. Западных учебников по эконометрике, рассчитанных на различный контингент читателей, безбрежное море (см., например, список литературы в конце настоящего тома). Российский читатель находится в худшем положении. Имеется несколько относительно старых переводных книг, учебник Магнуса-Катышева-Пересецкого, первое издание Айвазяна-Мхитаряна, наконец вышедший в этом году учебник «Эконометрика» коллектива авторов под общей редакцией члена-корреспондента РАН Елисейевой.

Настоящий же том — это классическая эконометрика, но изложенная не в западном стиле, а так, как это привычнее российскому читателю, особенно получившему основательную подготовку по математике. Айвазян читает эконометрику уже несколько лет в лучших университетах Москвы, является признанным в этой области лидером в стране. Следует напомнить, что Айвазян — выходец из Российской школы теории вероятностей, ученик Колмогорова и Прохорова. Отпечаток традиций этой всемирно признанной школы чувствуется в изложении материала тома, что приятно отметить во времена, когда Российская математическая школа переживает не самые лучшие времена.

Это своеобразность подхода автора к изложению материала выражается, в частности, в следующем. 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 неизменными составными частями эконометрических учебников, а решающие точно те же задачи **методы дискриминантного анализа** (кстати при определенных условиях *функционально связанные* с логит- и пробит- моделями) 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. Автор предлагаемой вниманию читателей книги одинаково хорошо ориентируется на обеих упомянутых территориях, что позволяет ему, там где это естественно и необходимо, выходить за пределы жестко очерченных границ традиционного эконометрического инструментария.

Хочется еще отметить, что эконометрика представляет собой не только науку, но и искусство. Искусство работы с данными. Автор имеет огромный опыт в этом деле и кое-где в деликатной форме предупреждает читателя об осторожности обращения с данными и с выводами из их анализа. Искусством нельзя овладеть, только читая книгу. Надо практически возиться с данными, считать и пересчитывать. Подготавливаемый автором задачник по эконометрике, который выйдет в ближайшее время, специально акцентирует внимание на некоторых чисто практических приемах эконометрической работы.

Академик Валерий Леонидович МАКАРОВ,
ректор Российской экономической школы,
директор ЦЭМИ РАН

25 июля 2001 г.

ПРЕДИСЛОВИЕ К ПЕРВОМУ ИЗДАНИЮ

Эконометрика — одна из базовых (наряду с микро- и макроэкономикой) дисциплин экономического образования во всем мире. К сожалению, до начала 90-х годов эконометрика по существу не была признана в СССР и России, не включалась в учебные планы подготовки специалистов (студентов, аспирантов) экономического профиля. Объяснение этому найти нетрудно: из трех основных составляющих эконометрики — экономической теории, экономической статистики и математико-статистического инструментария две первые были представлены в нашей стране явно неудовлетворительно. Не было доброкачественной экономической теории, не было системы национальных счетов и необходимого информационного обеспечения эконометрического моделирования.

Теперь ситуация изменилась. Авторам предлагаемого вниманию читателя учебника довелось принять непосредственное участие в процессе «восстановления в своих законных правах» эконометрики: в формировании базовых положений концепции современного экономического образования в российской высшей школе, в составлении первых программ и чтении первых курсов лекций по этой дисциплине на экономическом факультете Московского государственного университета им. М. В. Ломоносова (МГУ), начиная с 1992 г., и в Московском государственном университете экономики, статистики и информатики (МЭСИ), — с 1993 г. В формировании своей позиции по данной проблеме авторы опирались на многолетний опыт исследовательской и педагогической работы в области разработки и практического использования методов эконометрического моделирования: один из них (С. А. Айвазян), работая с 1969 года в Центральном экономико-математическом институте Российской академии наук, занимается эконометрическим моделированием распределительных отношений в обществе, одновременно являясь автором учебных программ и постоянным лектором по всему спектру дисциплин эконометрического профиля — *теории вероятностей, математической статистике, многомерному статистическому анализу* (МГУ им. М. В. Ломоносова, МЭСИ, Российская экономическая школа — РЭШ), *методам прогнозирования в бизнесе* (Московское отделение Калифорнийского государственного университета, г. Хэйвард); другой (В. С. Мхитарян) использует эконо-

метрические методы в социально-экономических исследованиях и задачах статистического контроля качества изделий, а также много лет читает курсы лекций вероятностно-статистического профиля в МЭСИ.

Что побудило авторов написать этот учебник? Ведь богатые традиции европейских и северо-американских университетов в области эконометрической науки и ее преподавания в системе экономического образования всех уровней (undergraduate, graduate, post-graduate) отражены в разнообразной и интересной монографической и учебно-методической литературе по данному предмету. Достаточно назвать широко распространенные учебники Дж. Джонстона, Э. Маленво, А. Goldberger, R. S. Pindyck—D. L. Rubinfeld, W. Greene, C. Dougherty, E. R. Berndt (см. список литературы в конце книги). Почему бы не повторить переводы двадцатилетней давности первых двух учебников и не перевести с английского языка некоторые из других, здесь упомянутых? Мы полагаем, что в условиях острого дефицита на отечественном рынке эконометрической учебной литературы эта деятельность была бы крайне желательной, более того — необходимой. Однако издание нашей книги имеет свой замысел, свою специфику, и вот в чем они заключаются.

Во-первых, в предлагаемом учебнике отражено понимание содержания математико-статистического инструментария эконометрики, несколько отличающееся от общепринятого. По нашему мнению, современные достижения математико-статистической науки (особенно в многомерном статистическом анализе), с одной стороны, и существенное расширение круга экономических задач, требующих эконометрических методов решения, — с другой, обусловили *необходимость более широкого взгляда на математико-статистический инструментарий эконометрики* и, в частности, включения в него, помимо традиционных разделов по регрессионным моделям, анализу временных рядов и системам одновременных уравнений, таких разделов многомерного статистического анализа, как марковские цепи, классификация многомерных наблюдений и снижение размерности анализируемого факторного пространства. Говоря о широком спектре экономических задач, требующих выходящих за традиционные рамки эконометрических методов решения, мы имели в виду, в частности, статистическое исследование динамики структурных изменений (в демографии, в стратификационной структуре общества и т.п.), выявление скрытых (латентных) факторов, определяющих течение того или иного социально-экономического процесса, построение интегральных индикаторов качества или эффективности функционирования социально-экономической системы, типологизацию социально-экономических объектов и др.

Во-вторых, в ходе многолетнего опыта преподавания различных дисциплин вероятностно-статистического профиля в экономических вузах и на экономических факультетах университетов мы пришли к убеждению, что *необходимо строить учебный процесс таким образом, чтобы добиваться цельного, системного восприятия всего блока этих дисциплин*. Речь идет, в частности, о курсах по элементарным методам статистической обработки данных (или дескриптивной статистике), теории вероятностей, математической статистике, многомерному статистическому анализу (или многомерным статистическим методам), анализу временных рядов и, наконец, эконометрике. Очевидно, *реализации этой цели должен способствовать и учебник, одновременно содержащий взаимосвязанное изложение всех этих курсов*.

Другими словами, мы попытались написать такую книгу, которую нам бы хотелось иметь под рукой в процессе нашей преподавательской деятельности. К сожалению, среди многих прекрасных зарубежных книг по эконометрике книги, обладающей двумя вышеуказанными особенностями, не оказалось.

Заметим, что несмотря на наличие ряда иллюстративных примеров и упражнений, предлагаемый учебник *не решает проблемы задачника по эконометрике*. Поэтому для проведения полноценного учебного процесса он должен быть дополнен набором эконометрических задач и упражнений (например, в духе книги [Berndt, E. R.]).

Материал учебника и ответственность распределены между авторами следующим образом. В. С. Мхитарян принимал участие в написании глав 6, 7, 8 и 13, а также предложил большую часть задач, которыми снабжены главы учебника. Остальной материал (включая упомянутые главы) написаны С. А. Айвазяном. Им же выполнено общее научное редактирование учебника.

В заключение мы хотим выразить свою признательность. Мы благодарны, во-первых, руководителям образовательного проекта EU-TACIS «Преподавание экономических и бизнес дисциплин в средних школах, технических и классических университетах» профессору Соломону Кохену (S. Cohen, SEOR и Эразмус Университет, Роттердам, Голландия) и ректору Государственного университета — Высшей школы экономики — профессору Ярославу Ивановичу Кузьминову: финансовая поддержка этого проекта сыграла решающую роль в том, чтобы авторы смогли определить для себя *приоритетной* задачей написание данного учебника. Мы выражаем также свою признательность нашим коллегам: профессору эконометрики Тилбургского университета и Высшей экономической школы Лондона Яну Магнусу и профессору Государственного университета —

Высшей школы экономики — Эмилю Борисовичу Ершову. Их внимание к учебнику было постоянным (в процессе его написания), взыскательным и одновременно доброжелательным. Высказанные ими замечания и советы, бесспорно, способствовали улучшению качества рукописи. Существенное влияние на замысел и содержание книги оказал опыт исследовательской и педагогической работы авторов, их постоянные контакты с коллегами по ЦЭМИ РАН, по научному семинару «Многомерный статистический анализ и вероятностное моделирование реальных процессов». Без них и без наших главных критиков и генераторов вопросов — многих поколений студентов МГУ им. М. В. Ломоносова, МЭСИ, РЭШ, — эта книга вряд ли появилась бы на свет. Мы благодарны Алле Павловне и Галине Юрьевне Грохотовым за самоотверженный труд по подготовке оригинал-макета рукописи книги, а также Николаю Владимировичу Третьякову и Елене Владимировне Герасимовой за очень полезную и профессиональную консультационную поддержку Грохотовых в этом объемном и непростом производственном процессе. Что касается слабых мест и недостатков учебника, то за них всю ответственность несут, естественно, только авторы. Мы будем признательны читателям за их отзывы о книге и критические замечания, направленные в издательство или непосредственно нам.

Москва–Роттердам–
деревня «Плужково»
Московской области.
1996–1998 гг.

*С. А. Авазян
В. С. Мзитарян*

ПРЕДИСЛОВИЕ КО ВТОРОМУ ИЗДАНИЮ

Что побудило нас подготовить второе издание учебника? Во-первых, весь пятитысячный тираж первого издания разошелся за неполные два года несмотря на внушительный объем книги и относительно специфическую область знаний, к которой она относится. Во-вторых, — это наш опыт общения со студентами и коллегами-преподавателями. В том, что учебник пользуется спросом у студентов, убеждает не только статистика продаж, но и специальные анкетные обследования, и информация, размещаемая на студенческих интернетовских сайтах (см. www.sachok.ru). Что касается вузовских преподавателей

эконометрики, то прекрасную возможность общения с ними предоставила одному из авторов серия специально организованных для них семинаров в разных регионах России и бывших республик СССР, на которых отечественные и зарубежные специалисты (в их числе — С.А. Айвазян) представляли свои циклы лекций в рамках общей программы повышения квалификации преподавателей по эконометрике. С 1997-го по 2001 год такие семинары прошли в Москве (дважды), Санкт-Петербурге (дважды), Екатеринбурге, Нижнем Новгороде, Сочи, Владивостоке, Воронеже, Перми, Вильнюсе и других городах.

Второе издание выходит с т е р е о т и п н ы м, без серьезных доделок или изменений. Устранены лишь обнаруженные (к сожалению, в весьма большом количестве) опечатки и явные погрешности. Кроме того, мы предложили издательству выпустить учебник в д в у х т о м а х: том 1 — «Теория вероятностей и прикладная статистика» (введение и главы 1-13 первого издания); том 2 — «Основы эконометрики» (главы 14-17 первого издания). Поскольку материал второго тома подготовлен полностью С.А. Айвазяном, этот том выходит только под его авторством. Практика использования студентами и преподавателями первого издания учебника подсказала нам, что д в у х т о м н ы й вариант книги может оказаться более технологичным и удобным в использовании.

Отметим, что со временем несколько трансформируются представления специалистов о самом предмете эконометрики, пополняется багаж его методов, смещаются акценты. Не со всеми такими представлениями, принятыми, скажем, в научных кругах США, согласны авторы этого учебника. Там, например, принято включать в продвинутые курсы (и учебники) по эконометрике «Теорию больших выборок» (или «Асимптотическую теорию»), «Непараметрические и полупараметрические методы принятия статистических решений», развернутое изложение метода максимального правдоподобия. С нашей точки зрения, 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, главные компоненты и др.) по непонятным для нас причинам отсутствуют в эконометрических курсах и учебниках Северной Америки и Западной Европы.

Следует, однако, признать, что во втором томе предлагаемого издания представлены, конечно, далеко не все важнейшие разделы современной эконометрики (поэтому этот том и называется «О с н о в ы эконометрики»). В нем нет, например, *обобщенного метода моментов*, методов

анализа *панельных данных*, раздела, посвященного *моделям с урезанными и цензурированными выборками*, недостаточно внимания уделено проблемам исследования стационарности временного ряда и, в связи с этим, *приемам коинтеграции и анализу единичных корней характеристического уравнения временного ряда*. Подобным образом расширенный вариант нашего второго тома составит, по существу, новый учебник, создание которого входит в ближайшие планы одного из авторов.

Наконец, о *задачах и упражнениях*, которыми необходимо оснастить аудиторные занятия со студентами по прикладной статистике и эконометрике. Одновременно с выходом данного двухтомного издания издательство «Юнити-Дана» публикует нашу книгу «Прикладная статистика в задачах и упражнениях», которая является естественным дополнением (задачником) к первому тому. Аналогичное дополнение ко второму тому (задачник по эконометрике) один из авторов планирует представить в течение ближайшего года.

В заключение хотим поблагодарить профессора Гарвардского университета Дэйла Джоргенсона (Dale Jorgenson) за внимание к книге и очень полезные обсуждения, которые состоялись во время пребывания С.А. Айвазяна в Гарвардском университете в апреле 2001г. Мы благодарны также д-ру Джону Киммелу (John Kimmel) из Североамериканского отделения издательства «Шпрингер-Верлаг» за организацию рецензий западных специалистов на наш учебник. Мы искренне благодарны преподавателям статистики и эконометрики различных вузов России и Литвы — участникам упомянутых выше региональных семинаров по преподаванию эконометрики за полезные обсуждения отдельных фрагментов учебника. Наконец, мы благодарны коллективам и руководству Центрального экономико-математического института Российской академии наук и Московского государственного университета экономики, статистики и информатики, плодотворная профессиональная среда существенно помогла нам в работе над учебником.

Москва, июнь 2001 г.

ГЛАВА 1. ЭКОНОМЕТРИКА И ЭКОНОМЕТРИЧЕСКОЕ МОДЕЛИРОВАНИЕ: ОСНОВНЫЕ ПОНЯТИЯ И ОПРЕДЕЛЕНИЯ

1.1. Эконометрика и ее место в ряду математико-статистических и экономических дисциплин

Существуют различные варианты определения эконометрики — от чрезмерно расширительных (при которых к эконометрике относят *все, что связано с измерениями в экономике*), до узко инструментально ориентированных (при которых под эконометрикой понимают *определенный набор математико-статистических средств*, позволяющих верифицировать модельные соотношения между анализируемыми экономическими показателями и оценивать неизвестные значения параметров в этих соотношениях на базе исходных экономических данных).

Название «эконометрика» было введено в 1926 г. норвежским экономистом и статистиком Рагнарм Фришем.

В буквальном переводе этот термин означает «измерения в экономике» и поэтому формально отвечает упомянутому выше расширительному толкованию (сравните с биометрикой, наукометрикой, астрометрией и т.п.). Однако ныне устоялся и широко распространен более ограниченный взгляд на содержание и назначение эконометрики. Этот взгляд, в частности, отражен в следующем определении.

О п р е д е л е н и е 1.1. *Эконометрика — это самостоятельная научная дисциплина, объединяющая совокупность теоретических результатов, приемов, методов и моделей, предназначенных для того, чтобы на базе*

(i) **экономической теории,**

(ii) **экономической статистики и экономических измерений,**

(iii) **математико-статистического инструментария**

придавать конкретное количественное выражение общим (качественным) закономерностям, обусловленным экономической теорией.

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 определением 1.1 суть эконометрики — именно в *синтезе экономики, экономической статистики 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* (определение этих понятий см. в п. 1.2). При рассмотрении *экономической статистики* как составной части эконометрики нас будет интересовать лишь тот аспект этой самостоятельной дисциплины, 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 экономиче-

ская статистика и математическая статистика. Так, математическая экономика, которая на самом деле является математически сформулированной экономической теорией, изучает взаимосвязи между 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 прослеживания 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 (т.е. страны в целом), мезоуровень (регионы, отрасли, корпорации) и микроуровень (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 она поможет читателю лучше понять нашу точку зрения на эконометрику и на ее место в ряду математико-статистических и экономических дисциплин.

Заметим, что вынесенные в заглавие книги «Основы эконометрики» подчеркивают 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.

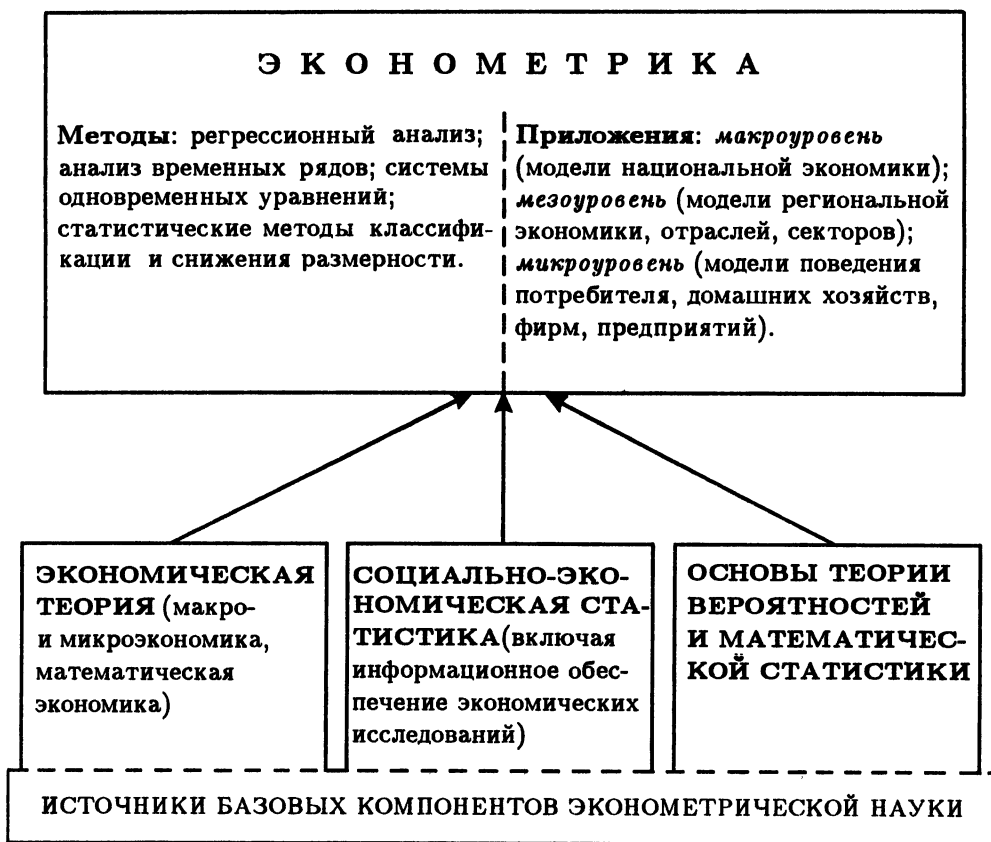


Рис. 1.1. Эконометрика и ее место в ряду других экономических и статистических дисциплин

1.2. Эконометрическая модель и проблемы эконометрического моделирования

1.2.1. 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*

ми 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. п. 10.5 в томе 1).

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. 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, т. е.

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

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 (т. е. $E\delta_t \equiv 0$). 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 развитии экономических теорий 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 (т. е. 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 удовлетворения сформулированным априорным требованиям теоретика. Какие соотношения выбрать между переменными — линейные или нелинейные? Если остановиться на нелинейных, 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 не-

скольких 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:

$$y_i^{(1)} = \alpha_0 + \alpha_1(y_i^{(3)} - x_i^{(1)}) + \delta_i^{(1)}, \quad (1.3)$$

$$y_i^{(2)} = \beta_1 y_{i-1}^{(3)} + \beta_2 \cdot x_i^{(2)} + \delta_i^{(2)}, \quad (1.4)$$

$$y_i^{(3)} = y_i^{(1)} + y_i^{(2)} + x_i^{(3)}, \quad (1.5)$$

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 $y_i^{(1)}$ обозначает потребление, $y_i^{(2)}$ — инвестиции, $y_i^{(3)}$ — национальный доход, $x_i^{(1)}$ — подоходный налог, $x_i^{(2)}$ — норму процента как инструмент государственного регулирования, $x_i^{(3)}$ — государственные закупки товаров и услуг, *измеренные в «момент времени» t*

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

1.2.2. Основные понятия эконометрического моделирования

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

экзогенные, т. е. задаваемые как бы «извне», автономно, в определенной степени управляемые (планируемые);

эндогенные, т. е. 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, т. е. 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*, т. е. 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 предыдущего пункта 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*.

При построении и анализе эконометрической модели следует различать ее *структурную* и *приведенную* формы. Для пояснения этих понятий условимся в дальнейшем обозначать латинской буквой X *все предопределенные переменные*, т. е. все экзогенные переменные и все участвующие в модели лаговые эндогенные переменные. Пусть общее число эндогенных переменных равно m , а общее число предопределенных пере-

менных — p . Примем (пока без объяснения)¹, что общее число уравнений и тождеств в эконометрической модели равно числу эндогенных переменных, т. е. равно m . И пусть из общего числа m соотношений модели мы имеем m_1 уравнений, включающих случайные остаточные компоненты, и m_2 тождеств ($m_1 + m_2 = m$). Разобьем вектор эндогенных переменных $Y_t = (y_t^{(1)}, y_t^{(2)}, \dots, y_t^{(m)})^T$ на два подвектора $Y_t^{(1)} = (y_t^{(1)}, y_t^{(2)}, \dots, y_t^{(m_1)})^T$ и $Y_t^{(2)} = (y_t^{(m_1+1)}, \dots, y_t^{(m_1+m_2)})^T$, при этом порядок, в котором перенумерованы эндогенные переменные, не имеет значения.

Тогда общий вид линейной эконометрической модели может быть представлен в форме

$$\begin{cases} B_1 Y_t^{(1)} + B_2 Y_t^{(2)} + C_1 X_t = \Delta_t \\ B_3 Y_t^{(1)} + B_4 Y_t^{(2)} + C_2 X_t = 0, \quad t = 1, 2, \dots, n, \end{cases} \quad (1.6)$$

где $B_1 = (\beta_{ij})_{i,j=\overline{1,m_1}}$ — матрица размерности $(m_1 \times m_1)$ из коэффициентов при $y_t^{(1)}, \dots, y_t^{(m_1)}$ в m_1 первых уравнениях; $B_2 = (\beta_{ij})_{\substack{i=\overline{1,m_1} \\ j=\overline{m_1+1,m_1+m_2}}}$ матрица размерности $m_1 \times m_2$ из коэффициентов при $y_t^{(m_1+1)}, \dots, y_t^{(m_1+m_2)}$ в m_1 первых уравнениях²; $X_t = (x_t^{(0)}, x_t^{(1)}, x_t^{(2)}, \dots, x_t^{(p)})^T$ — вектор-столбец предопределенных переменных (в нем $x_t^{(0)} \equiv 1$); $C_1 = (c_{ij})_{\substack{i=\overline{1,m_1} \\ j=\overline{0,p}}}$ — матрица размерности $m_1 \times (p+1)$ из коэффициентов при предопределенных переменных в первых m_1 уравнениях (очевидно, коэффициенты c_{i0} играют роль свободных членов уравнений); $B_3 = (\beta_{ij})_{\substack{i=\overline{m_1+1,m_1+m_2} \\ j=\overline{1,m_1}}}$ — матрица размерности $m_2 \times m_1$ из коэффициентов при $y_t^{(1)}, \dots, y_t^{(m_1)}$ в m_2 тождествах системы; $B_4 = (\beta_{ij})_{\substack{i=\overline{m_1+1,m_1+m_2} \\ j=\overline{m_1+1,m_1+m_2}}}$ — матрица размерности $m_2 \times m_2$ из коэффициентов при $y_t^{(m_1+1)}, \dots, y_t^{(m_1+m_2)}$ в m_2 тождествах системы; $C_2 = (c_{ij})_{\substack{i=\overline{m_1+1,m_1+m_2} \\ j=\overline{0,p}}}$ — матрица размерности $m_2 \times (p+1)$ из коэффициентов при предопределенных переменных в m_2 тождествах системы; $\Delta_t = (\delta_t^{(1)}, \delta_t^{(2)}, \dots, \delta_t^{(m_1)})^T$ — вектор-столбец размерности m_1 случайных остаточных составляющих m_1 первых уравнений системы и $0_{m_2} = (0, 0, \dots, 0)^T$ — вектор-столбец размерности m_2 , состоящий из нулей.

¹ Подробному анализу эконометрической модели, представленной в виде общей линейной системы одновременных уравнений, посвящена гл. 4.

² С помощью символов $\overline{i} = n_1, n_2$ кратко обозначается тот факт, что индекс i может принимать все целые значения от n_1 до n_2 , т. е. $i = n_1, n_1 + 1, \dots, n_2$.

Заметим, что *исходными статистическими данными*, необходимыми для проведения статистического анализа системы (1.6) (а именно для оценки неизвестных коэффициентов β_{ij} и c_{ij} , проверки статистических гипотез, например, о линейном характере исследуемых зависимостей и т.п.), являются матрицы

$$Y = \begin{pmatrix} Y_1^\top \\ \vdots \\ Y_n^\top \end{pmatrix} \quad \text{и} \quad X = \begin{pmatrix} X_1^\top \\ \vdots \\ X_n^\top \end{pmatrix} \quad (1.7)$$

соответственно размерностей $n \times t$ и $n \times (p + 1)$, а все элементы матриц B_3, B_4 и C_2 являются *известными* (их числовые значения определяются содержательным смыслом соответствующих тождеств системы).

Система (1.6) может быть записана также в виде

$$BY_t + CX_t = \bar{\Delta}_t, \quad t = 1, 2, \dots, n, \quad (1.6')$$

или в виде

$$Y \cdot B^\top + X \cdot C^\top = \bar{\Delta}, \quad (1.6'')$$

где

$$Y_t = \begin{pmatrix} Y_t^{(1)} \\ Y_t^{(2)} \end{pmatrix}, \quad B = \begin{pmatrix} B_1 & B_2 \\ B_3 & B_4 \end{pmatrix}, \quad C = \begin{pmatrix} C_1 \\ C_2 \end{pmatrix}, \quad \bar{\Delta}_t = \begin{pmatrix} \Delta_t \\ 0_{m_2} \end{pmatrix}, \quad \bar{\Delta} = \begin{pmatrix} \bar{\Delta}_1^\top \\ \vdots \\ \bar{\Delta}_n^\top \end{pmatrix},$$

а матрицы Y и X определены в (1.7).

О п р е д е л е н и е 1.2. Система уравнений и тождеств вида (1.6) (или эквивалентных ей записей (1.6') или (1.6'')) называется **структурной формой линейной эконометрической модели**. При этом предполагается, что коэффициент при i -й эндогенной переменной в i -м структурном стохастическом уравнении ($i = 1, 2, \dots, t$) равен единице (правило нормировки системы), а матрицы B_4 и B невырождены (допускаются и другие способы нормировки системы).

З а м е ч а н и е. Пользуясь тем, что эндогенные переменные $Y_t^{(2)}$ из (1.6) могут быть явно выражены через $Y_t^{(1)}$ и X_t , их можно исключить из общей системы и рассматривать систему, содержащую в качестве эндогенных переменных только переменные $Y_t^{(1)}$, а именно:

$$B^* Y_t^{(1)} + C^* X_t = \Delta_t, \quad t = 1, 2, \dots, n, \quad (1.6^a)$$

где $B^* = B_1 - B_2 B_4^{-1} B_3$ и $C^* = C_1 B_2 B_4^{-1} C_2$. Поэтому в дальнейшем, говоря о *статистическом анализе* системы соотношений, описывающих

линейную эконометрическую модель, можно рассматривать системы, содержащие только m_1 эконометрических стохастических уравнений и не содержащие тождеств. Правда, этот переход может приводить к определенной «перепараметризации» модели, т. е. к утрате автономности первоначально введенных параметров (см. ниже реализацию данного перехода в примере 1.1).

Поскольку при реализации конечных прикладных целей эконометрического моделирования (т. е. при прогнозе значений эндогенных переменных и при различных имитационных расчетах, см. п. 1.1) главный интерес представляют соотношения, позволяющие явно выразить все эндогенные переменные Y_t через предопределенные X_t , то одновременно со структурной формой имеет смысл рассмотреть так называемую *приведенную (редуцированную)* форму линейной эконометрической модели.¹ Требуемый результат мы получим, домножив слева обе части соотношений (1.6') на матрицу B^{-1} и уединив затем Y_t :

$$Y_t = -B^{-1}CX_t + B^{-1}\bar{\Delta}_t, \quad t = 1, 2, \dots, n. \quad (1.8)$$

Аналогично можно поступить и со структурной формой, записанной в виде (1.6^a):

$$Y_t^{(1)} = -(B^*)^{-1}C^*X_t + (B^*)^{-1}\Delta_t, \quad t = 1, 2, \dots, n. \quad (1.9)$$

В дальнейшем, говоря о приведенной (редуцированной) форме эконометрической модели, мы будем иметь в виду запись

$$Y_t = \Pi \cdot X_t + \varepsilon_t, \quad t = 1, 2, \dots, n \quad (1.8')$$

или

$$Y_t^{(1)} = \Pi^* X_t + \varepsilon_t^*, \quad t = 1, 2, \dots, n, \quad (1.9')$$

где $m \times (p+1)$ матрица Π , $m_1 \times (p+1)$ матрица Π^* и векторы остаточных случайных составляющих ε_t и ε_t^* определяются соотношениями

$$\Pi = -B^{-1}C, \quad (1.10)$$

$$\Pi^* = -(B^*)^{-1}C^*, \quad (1.11)$$

$$\varepsilon_t = B^{-1}\bar{\Delta}_t, \quad (1.12)$$

$$\varepsilon_t^* = (B^*)^{-1}\Delta_t. \quad (1.13)$$

¹ Более распространенный в русскоязычной литературе термин «приведенная форма» в действительности менее точно передает в переводе с английского смысл «reduced form model», так как основное отличие последней — в меньшем (редуцированном) числе содержащихся в ней параметров по сравнению со структурной формой (более подробно об этом см. ниже в п. 1.2.3).

О п р е д е л е н и е 1.3. Система соотношений (1.8') (или (1.9')), в которой все эндогенные переменные эконометрической модели явно линейно выражены через предопределенные переменные и случайные остаточные компоненты, называется **приведенной формой линейной эконометрической модели.**

Проиллюстрируем введенные понятия на примере (1.3)–(1.5) (см. п. 1.2.1).

В этом примере число эндогенных переменных, так же как и общее число всех соотношений модели, равно трем ($m = 3$). Среди этих соотношений мы имеем одно тождество (следовательно, $m_1 = 2$, $m_2 = 1$). Общее число предопределенных переменных $p = 4$, в том числе три экзогенные переменные ($x_t^{(1)}$, $x_t^{(2)}$, $x_t^{(3)}$) и одна лаговая эндогенная переменная ($y_{t-1}^{(3)}$), которую мы в соответствии с принятой договоренностью кодируем как $x_t^{(4)}$ (т. е. $y_{t-1}^{(3)} \equiv x_t^{(4)}$).

Структурная форма модели в данном примере задается соотношениями (1.3)–(1.5). В общих матричных обозначениях, использованных в формуле (1.6), имеем:

$$\mathbf{B}_1 = \begin{pmatrix} 1; & 0 \\ 0; & 1 \end{pmatrix}; \quad \mathbf{B}_2 = \begin{pmatrix} -\alpha_1 \\ 0 \end{pmatrix}; \quad \mathbf{C}_1 = \begin{pmatrix} -\alpha_0; & -\alpha_1; & 0; & 0; & 0 \\ 0; & 0; & -\beta_2; & 0; & -\beta_1 \end{pmatrix};$$

$$\mathbf{B}_3 = (-1; -1); \quad \mathbf{B}_4 = (1); \quad \mathbf{C}_2 = (0; 0; 0; -1; 0)$$

и

$$\Delta_t = (\delta_t^{(1)}, \delta_t^{(2)})^\top.$$

Если же структурная форма записана в виде (1.6'), то в данном примере участвующие в этой записи матрицы конкретизируются в виде

$$\mathbf{B} = \begin{pmatrix} 1 & 0 & -\alpha_1 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{pmatrix}; \quad \mathbf{C} = \begin{pmatrix} \alpha_0 & -\alpha_1 & 0 & 0 & 0 \\ 0 & 0 & -\beta_2 & 0 & -\beta_1 \\ 0 & 0 & 0 & -1 & 0 \end{pmatrix}. \quad (1.14)$$

и

$$\bar{\Delta}_t = (\delta_t^{(1)}, \delta_t^{(2)}, 0)^\top.$$

Отметим, что, во-первых, выполнено условие нормировки ($y_t^{(i)}$ входит в i -е уравнение системы, $i = 1, 2$, с коэффициентом единица); во-вторых, значения элементов матриц \mathbf{B}_3 , \mathbf{B}_4 и \mathbf{C}_2 известны, они определяются содержательным смыслом тождества; в третьих, требование невырожденности матриц \mathbf{B}_4 и \mathbf{B} соблюдено; и, наконец, в четвертых, матрицы \mathbf{B}_1 , \mathbf{B}_2 и \mathbf{C}_1 относительно «слабо заполнены» неизвестными (подлежащими статистическому оцениванию) коэффициентами: их всего четыре — α_0 , α_1 , β_1

и β_2 . Последняя особенность рассматриваемой эконометрической модели является, к счастью, достаточно общей отличительной чертой систем эконометрических уравнений. Если бы это было не так, т. е. если бы мы были вынуждены иметь дело с системами, «сильно заполненными» неизвестными коэффициентами, то задача статистического анализа таких систем оказывалась бы *принципиально неразрешимой*: имеющихся в нашем распоряжении исходных статистических данных просто не хватало бы для корректного проведения такого анализа. Ведь при построении и анализе систем эконометрических уравнений, описывающих макроэкономические модели, исследователю зачастую приходится иметь дело с десятками и сотнями эндогенных и экзогенных переменных!

Рассмотрим далее структурную форму в записи (1.6^а), при которой исключается часть эндогенных переменных $Y_i^{(2)}$ посредством их выражения через $Y_i^{(1)}$ из тождеств системы. В нашем примере эта структурная форма имеет вид:

$$\begin{cases} (1 - \alpha_1)y_i^{(1)} - \alpha_1 y_i^{(2)} - \alpha_0 + \alpha_1 x_i^{(1)} - \alpha_1 x_i^{(3)} = \delta_i^{(1)} \\ y_i^{(2)} - \beta_2 x_i^{(2)} - \beta_1 x_i^{(4)} = \delta_i^{(2)}, \end{cases}$$

т. е. матрицы B^* и C^* , участвующие в общей записи (1.2.6^а), в данном примере конкретизируются в форме

$$B^* = \begin{pmatrix} 1 - \alpha_1 & -\alpha_1 \\ 0 & 1 \end{pmatrix}, \quad C^* = \begin{pmatrix} -\alpha_0 & \alpha_1 & 0 & -\alpha_1 & 0 \\ 0 & 0 & -\beta_2 & 0 & -\beta_1 \end{pmatrix}.$$

И, наконец, *приведенная* форма модели (1.9'), в которой все эндогенные переменные выражаются через предопределенные, в данном примере имеет вид

$$\begin{cases} y_i^{(1)} = \frac{1}{1 - \alpha_1} (\alpha_0 - \alpha_1 x_i^{(1)} + \alpha_1 \beta_2 x_i^{(2)} + \alpha_1 x_i^{(3)} + \alpha_1 \beta_1 x_i^{(4)}) \\ \quad + \left(\delta_i^{(1)} + \frac{\alpha_1}{1 - \alpha_1} \delta_i^{(2)} \right) \\ y_i^{(2)} = \beta_2 x_i^{(2)} + \beta_1 x_i^{(4)} + \delta_i^{(2)}, \end{cases}$$

т. е. матрица Π^* и вектор случайных остаточных составляющих ϵ_i^* , участвующие в общей записи приведенной формы (1.9'), в данном примере конкретизируются в виде:

$$\Pi^* = \begin{pmatrix} \frac{\alpha_0}{1 - \alpha_1} & -\frac{\alpha_1}{1 - \alpha_1} & \frac{\alpha_1 \beta_2}{1 - \alpha_1} & \frac{\alpha_1}{1 - \alpha_1} & \frac{\alpha_1 \beta_1}{1 - \alpha_1} \\ 0 & 0 & \beta_2 & 0 & \beta_1 \end{pmatrix},$$

$$\varepsilon_i^* = \left(\delta_i^{(1)} + \frac{\alpha_1}{1 - \alpha_1} \delta_i^{(2)}, \delta_i^{(2)} \right)^T.$$

1.2.3. Основные проблемы эконометрического моделирования

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, т. е. выбор **общего вида модели**, в том числе состава и формы входящих в нее связей;

4-й этап (информационный) — сбор необходимой статистической информации, т. е. 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.

Последние три этапа (4-й, 5-й и 6-й) сопровождаются крайне трудоемкой **процедурой калибровки модели**. Дело в том, что при построении эконометрической модели исследователь, как правило, находится в ситуации, когда, с одной стороны, действует большое число «нормативных» (т. е. определенных содержательным смыслом анализируемых связей) ограничений на коэффициенты матриц **B** и **C**, а с другой стороны, ему приходится действовать в условиях определенной нечеткости (или неполноты) исходной статистической информации (1.7). Процедура калибровки модели заключается в переборе большого числа различных вариантов «нормативные ограничения — значения отдельных переменных» (что связано с многократными «вычислительными прогонами» модели) с целью получения совместной, непротиворечивой и идентифицируемой модели.

Математическая модель, в том числе математическая модель *экономического* явления или процесса, может быть сформулирована на общем (качественном) уровне, без настройки на конкретные статистические данные, т. е. 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:¹

- (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*);
- числовых значений отдельных элементов матриц B и C в структурной форме модели (1.6') или (1.6'').

Итак, спецификация модели — это первый и, быть может, *важнейший* шаг эконометрического исследования. От того, насколько удачно реше-

¹ Ниже описывается содержание проблемы спецификации эконометрической модели в предположении ее линейности, т. е. 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. Тому, как поступать в противном случае, посвящен п. 2.12 учебника.

на 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 элементов матриц \mathbf{B} и \mathbf{C} , решающим образом зависит успех всего эконометрического исследования.

Спецификация опирается на имеющиеся экономические теории, специальные знания или на интуитивные представления исследователя об анализируемой экономической системе. Эти априорные сведения определяют, в частности, природу матриц \mathbf{B} и \mathbf{C} . Например, информация (или предположение) о том, что определенные переменные непосредственно не участвуют в том или ином уравнении, означает равенство нулю соответствующих элементов в строках матриц \mathbf{B} и \mathbf{C} . Дополнительные сведения о системе могут иметь вид ограничений на комбинации элементов матриц \mathbf{B} и \mathbf{C} .

Проиллюстрируем сказанное на нашем примере из п. 1.2.1. Мы видим, что непосредственно из состава и смысла уравнений системы (1.3)–(1.5) (т. е. из решения части вопросов проблемы спецификации модели) непосредственно следует специальный вид матриц \mathbf{B} и \mathbf{C} (см. (1.14)) и, в частности, их слабая заполненность априори не известными элементами (из 24 элементов этих матриц нам предстоит статистически оценить лишь четыре: $\alpha_0, \alpha_1, \beta_1$ и β_2).

Априорные сведения о системе находят свое отражение при спецификации модели не только в определении конфигурации матриц \mathbf{B} и \mathbf{C} , 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, т. е. $\mathbf{E}\delta_t^{(j)} \equiv 0, t = 1, \dots, n$;
- не коррелируют друг с другом, т. е. $\mathbf{E}(\delta_t^{(j)} \cdot \delta_t^{(l)}) = 0, (j \neq l)$;
- не имеют автокорреляций, т. е. $\mathbf{E}(\delta_{t_1}^{(j)} \cdot \delta_{t_2}^{(j)}) = 0, t_1 \neq t_2$;
- не коррелируют ни с одной из предопределенных переменных.

Как правило, подобные допущения в процессе спецификации модели оказываются достаточно реалистичными. Однако обращаем внимание читателя на тот факт, что, к сожалению, *ничего подобного нельзя сказать о случайных остатках ε_t и ε_t^* приведенной формы*. Из (1.12), (1.13) видно, что эти случайные остатки зависят (линейно) от элементов (в том числе неизвестных) матрицы \mathbf{B} , и это, как мы увидим позже, создает свои трудности в статистическом анализе уравнений приведенной формы (1.8') и (1.9').

Проблема идентифицируемости. При анализе эконометрической модели, представленной системой уравнений вида (1.6) (или (1.6')), *исследователя в конечном счете интересует прежде всего поведение эндогенных переменных Y_t* . Из соответствующей приведенной формы модели (1.8) видно, что эндогенные переменные Y_t являются по своей природе случайными величинами, поведение которых определяется *внутренней структурой модели*, а именно, элементами матриц B и C и природой случайных остатков Δ_t . Возникает вопрос: а возможно ли, следуя в «обратном направлении», восстановить структурную форму (1.6') (т. е. все элементы матриц B и C), располагая знанием приведенной формы (1.8') (т. е. знанием числовых значений всех элементов матрицы Π и природы случайных остатков ε_t)? Именно этот вопрос и отражает сущность проблемы идентифицируемости эконометрической модели (не смешивать с *проблемой идентификации* модели, заключающейся в выборе и реализации методов статистического оценивания ее неизвестных параметров, см. ниже).

Ответ на поставленный вопрос в общем случае, очевидно, отрицательный: без дополнительных ограничений на внутреннюю структуру модели (т. е. без соблюдения некоторых *условий идентифицируемости*) по $m_1 \times (p+1)$ элементам матрицы Π невозможно восстановить гораздо большее число элементов матриц B и C (нетрудно подсчитать, что общее число коэффициентов β_{ij} и c_{ik} в структурной форме равно $m_1 \times (m_1 + m_2 + p + 1)$, хотя, конечно, общее число *коэффициентов, подлежащих статистическому оцениванию*, оказывается меньшим).

В эконометрической теории приняты следующие определения, связанные с проблемой идентифицируемости.

О п р е д е л е н и е 1.4. *Уравнение структурной формы эконометрической модели называется точно идентифицируемым, если все участвующие в нем неизвестные (т. е. априори не заданные) коэффициенты однозначно восстанавливаются по коэффициентам приведенной формы без каких-либо ограничений на значения последних.*

О п р е д е л е н и е 1.5. *Эконометрическая модель называется точно идентифицируемой, если все уравнения ее структурной формы являются точно идентифицируемыми.*

О п р е д е л е н и е 1.6. *Уравнение структурной формы называется сверхидентифицируемым, если все участвующие в нем неизвестные коэффициенты восстанавливаются по коэффициентам приведенной формы, причем некоторые из его коэффициентов могут принимать одновременно несколько (более одного) числовых значений, соответствующих одной и той же приведенной форме.*

О п р е д е л е н и е 1.7. *Уравнение структурной формы называется*

ся неидентифицируемым, если хотя бы один из участвующих в нем неизвестных коэффициентов не может быть восстановлен по коэффициентам приведенной формы. Соответственно модель называется неидентифицируемой, если хотя бы одно из уравнений ее структурной формы является неидентифицируемым.

Говоря о проблеме идентифицируемости модели, мы начали с того, что исследователя в конечном счете интересует поведение эндогенных переменных, и с этой точки зрения может показаться несущественной, более того, надуманной проблема «однозначного возврата» от приведенной формы к структурной. Однако в действительности проблема идентифицируемости крайне важна, в первую очередь с позиций выработки предложений по решению следующей проблемы — *проблемы идентификации эконометрической модели*, т. е. проблемы выбора и реализации методов статистического оценивания участвующих в ней неизвестных параметров.

Более подробное рассмотрение проблемы идентифицируемости эконометрической модели, включающее в себя, в частности, *формулировки необходимых условий идентифицируемости отдельного параметра, целого уравнения и всей системы уравнений структурной формы*, приведено в гл. 4, посвященной системам одновременных уравнений.

Проблема идентификации. Решение этой проблемы предусматривает «настройку» записанной в общей структурной форме (1.6') модели на реальные статистические данные (1.7). Другими словами, речь идет о выборе и реализации методов статистического оценивания неизвестных параметров модели (1.6) (т. е. той части элементов матриц B и C , значения которых не являются априори известными) по исходным статистическим данным (1.7). Описание необходимых методов статистического оценивания параметров в подобных моделях и в их отдельных фрагментах приводится в гл. 2–4.

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, т. е. статистическим оцениванием участвующих в ней неизвестных коэффициентов (параметров) b_{ij} и c_{lk} . После этого, однако, возникают вопросы: (а) насколько удачно удалось решить проблемы спецификации, идентифицируемости и идентификации модели, т. е. можно ли рассчитывать на то, что использование построенной модели в целях прогноза эндогенных переменных и имитационных расчетов, определяющих варианты социально-экономического развития анализируемой системы, даст результаты, достаточно адекватные реальной действительности? (б) какова точность (абсолютная, относи-

тельная) 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 гл. 2–4. 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*. Конкретная схема построения таких расчетов зависит от конечных прикладных целей моделирования, однако, их общая сущность заключается в следующем.

Предположим, мы строим эконометрическую модель с целью прогноза эндогенных переменных или имитационных расчетов *на τ временных тактов вперед*. Тогда исходные статистические данные (1.7) делятся на две части:

$$\text{обучающую выборку} \quad \left\{ \left(\begin{array}{c} Y_1^\top \\ \vdots \\ Y_{n-\tau}^\top \end{array} \right), \left(\begin{array}{c} X_1^\top \\ \vdots \\ X_{n-\tau}^\top \end{array} \right) \right\}, \quad (1.7^a)$$

$$\text{экзаменующую выборку} \quad \left\{ \left(\begin{array}{c} Y_{n-\tau+1}^\top \\ \vdots \\ Y_n^\top \end{array} \right), \left(\begin{array}{c} X_{n-\tau+1}^\top \\ \vdots \\ X_n^\top \end{array} \right) \right\}. \quad (1.7^b)$$

Далее все ранее принятые решения по проблемам спецификации, идентифицируемости и идентификации модели применяются только к наблюдениям обучающей выборки (1.7^a). Из полученной таким образом модели подстановкой в ее приведенную форму значений экзогенных переменных $X_{n-\tau+1}, X_{n-\tau+2}, \dots, X_n$ получают *модельные (ретроспективно прогнозные)* значения соответственно $\bar{Y}_{n-\tau+1}, \bar{Y}_{n-\tau+2}, \dots, \bar{Y}_n$. Сравнение этих модельных значений с соответствующими реальными значениями экзаменующей выборки $Y_{n-\tau+1}, Y_{n-\tau+2}, \dots, Y_n$ позволяет проанализировать и адекватность модельных выводов реальной действительности и их точность. Более подробно реализация этого принципа будет описана в связи с построением и анализом моделей регрессии (гл. 2), временных рядов (гл. 3) и систем одновременных уравнений (гл. 4).

1.3. Математико-статистический инструментарий эконометрики

В предисловии кратко упомянуто о том, что позиция авторов данного учебника относительно понимания содержания математико-статистического инструментария эконометрики несколько отличается от общепринятой. В частности, современные продвижения в математико-статистической науке (особенно в области многомерного статистического анализа), с одной стороны, и заметное расширение круга экономических задач, требующих эконометрического подхода в их решении, — с другой, создали все необходимые предпосылки для пересмотра сложившегося взгляда на математико-статистический инструментарий эконометрики в направлении его существенного пополнения.

1.3.1. Традиционный состав математико-статистических методов эконометрики

подавляющее большинство учебников и учебных пособий по эконометрике предлагает читателю стандартный набор математико-статистических методов, представленных в следующих пяти разделах:

- *классическая линейная модель множественной регрессии (ЛММР) и классический метод наименьших квадратов (МНК);*
- *обобщенная ЛММР и обобщенный МНК;*
- *некоторые специальные модели регрессии (со стохастическими объясняющими переменными, с переменной структурой, с дискретными зависимыми переменными, нелинейные);*
- *модели и методы статистического анализа временных рядов;*
- *анализ систем одновременных эконометрических уравнений.*

При этом некоторые учебники не содержат описания моделей и методов анализа временных рядов ([Джонстон Дж.], [Goldberger, A. S.]). И действительно, упомянутые пять разделов составляют математико-статистическую базу эконометрического моделирования, предоставляя исследователю разнообразные методы и модели статистического исследования зависимостей, существующих между экономическими показателя-

ми. Именно этим разделам посвящены гл. 2, 3 и 4 данного учебника¹.

1.3.2. Некоторые задачи социально-экономической теории и практики, решение которых требует методов прикладной статистики, выходящих за рамки традиционного эконометрического инструментария

Краткий перечень таких задач приведен в предисловии к данному учебнику. Останемся на некоторых из них подробнее.

1) *Типологизация и кластеризация социально-экономических объектов.* Моделирование и статистический анализ распределения по среднедушевому доходу, выявление основных типов потребительского поведения домашних хозяйств, задачи социально-экономической стратификации общества, межстрановой макроэкономический анализ и многие другие решаются сегодня с привлечением современного аппарата многомерного статистического анализа — методов дискриминантного анализа, моделей расщепления смесей распределений, методов кластер-анализа (см. гл. 12 в томе 1).

2) *Построение и анализ целевых функций и интегральных индикаторов.* Один из эффективных (и достаточно распространенных в теории и практике экономических исследований) подходов к описанию и анализу поведения хозяйствующего субъекта (индивидуума, домашнего хозяйства, фирмы, предприятия и т.п.) связан с построением соответствующей целевой функции, которая по-существу является некоторой сверткой ряда частных показателей его поведения. Аналогичные задачи возникают при построении и анализе комплексных, агрегатных показателей какого-либо сложного свойства — качества населения, качества жизни, научно-технического уровня производственной системы и т.п. Как правило, при решении подобных задач не удается обойтись привлечением только методов регрессионного анализа и анализа временных рядов. Чаще исследо-

¹ Современные продвинутые учебники по эконометрике уделяют, кроме того, внимание анализу так называемых панельных данных («пространственно-временных выборок» в терминологии п. 9.1 первого тома данного учебника), урезанных и цензурированных выборок, обобщенному методу моментов. В последние годы на Западе стало модным включать в учебники магистерского уровня образования в специализированном контексте эконометрических задач такие разделы фундаментального инструментария теории вероятностей и математической статистики как теория больших выборок (асимптотическая теория или избранные вопросы предельных теорем теории вероятностей), метод максимального правдоподобия, непараметрические и полупараметрические методы принятия статистических решений (см., например, (Hayashi F., 2000), (Horowitz J., 1998), (Ruud P., 2000) и др.).

вателю приходится обращаться к таким методам снижения размерности анализируемого факторного пространства, как главные компоненты, факторный анализ, многомерное шкалирование (см. гл. 13 в томе 1).

3) *Анализ динамики «состояний» объекта* (типологии потребительского поведения семей, социально-экономической или демографической структуры общества и т.п.). Эффективным средством решения задач подобного типа являются, как мы видели (см. гл. 5 в томе 1), модели марковских цепей.

Упомянутый в пп. 1)–3) инструментарий прикладной статистики, приспособленный к специфике экономических и социально-экономических задач, с полным правом может быть отнесен (наряду с указанными в п. 1.3.1 пятью разделами) к математико-статистическому инструментарию эконометрики. Однако, не столь важно, как назвать эти методы. Важно, чтобы ими был оснащен современный специалист по эконометрике!

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 *два типа конечных прикладных целей* (или один из них): (а) *прогноз* экономических и социально-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 *иераргический уровень и профиль*. Анализируемые задачи могут относиться к *макро-* (страна, межстрановой анализ), *мезо-* (регионы внутри страны) и *микро-* (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 анали-

зируемыми 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. Эконометрическая модель, выраженная системой одновременных уравнений, может быть записана в *структурной* или *приведенной форме*. Структурная форма отражает непосредственно результат формирования уравнений связи между эндогенными и экзогенными переменными, опирающегося на их экономическую сущность. При этом левая часть любого уравнения структурной формы может содержать и эндогенные, и экзогенные переменные *без какого-либо разделения их ролей*, а правая часть — остаточную случайную компоненту. Приведенная форма системы есть результат разрешения структурной формы относительно эндогенных переменных.

7. К основным проблемам эконометрического моделирования следует отнести:

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

Сущность этих четырех проблем описана в п. 1.2.3.

8. Содержание *основного математико-статистического инструментария* эконометрики в его традиционном понимании определяется следующими пятью разделами:

(1) классическая модель регрессии и классический метод наименьших квадратов;

(2) обобщенная линейная модель регрессии и обобщенный метод наименьших квадратов;

(3) некоторые специальные модели регрессии (со стохастическими объясняющими переменными, с переменной структурой, с дискретными зависимыми переменными, нелинейные);

(4) статистический анализ временных рядов;

(5) анализ систем одновременных уравнений.

Однако в современном понимании 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. Разделы (1)–(4) математического инструментария эконометрики описываются в последующих трех главах (см. гл. 2–4), а упомянутым здесь методам многомерного статистического анализа и прикладной статистики были посвящены гл. 9–13 первого тома данного учебника.

ГЛАВА 2. МЕТОДЫ И МОДЕЛИ РЕГРЕССИОННОГО АНАЛИЗА

2.1. Введение в регрессионный анализ (основные понятия и определения)

Регрессионный анализ занимает, бесспорно, центральное место во всем математико-статистическом инструментарии эконометрики. По существу мы значительно раньше этой главы начали обсуждение проблем регрессионного анализа. Ведь впервые *условное математическое ожидание* одной случайной величины ($\xi^{(i)}$) при условии, что вторая случайная величина $\xi^{(j)}$ зафиксирована на уровне $x^{(j)}$ (т.е. при условии $\{\xi^{(j)} = x^{(j)}\}$), обсуждается уже в п. 3.1.5 первого тома в связи с анализом условного распределения одной из компонент двумерного нормального распределения. Общие постановки задач статистического исследования зависимостей и основные типы регрессионных зависимостей между количественными признаками обсуждались в пп. 10.1, 10.2, 10.5 первого тома. Там же впервые дано определение *функции регрессии* (см. п. 10.1 тома 1). И, наконец, в гл. 11 первого тома 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 изучению этой главы, желательно освежить в памяти сведения, почерпнутые читателем в результате чтения гл. 10 и 11 тома 1. Самые необходимые из этих сведений мы приводим ниже в качестве *вспомогательной сводки понятий и результатов*, предваряющей систематическое изложение содержания данной главы.

2.1.1. Результирующая (зависимая, эндогенная) переменная y

Переменная (или признак), характеризующая *результат* или *эффективность функционирования* анализируемой экономической системы. Ее значения формируются *в процессе и внутри* функционирования этой системы под воздействием ряда других переменных и факторов, часть из которых поддается регистрации и, в определенной степени, управлению и планированию (эту часть принято называть объясняющими переменными, см. ниже). В регрессионном анализе результирующая переменная выступает в роли *функции*, значения которой определяются (правда, с некоторой случайной погрешностью) значениями упомянутых выше объясняющих переменных, выступающих в роли *аргументов*. Поэтому по природе своей результирующая переменная y всегда стохастична (случайна).

2.1.2. Объясняющие (предикторные, экзогенные) переменные $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$

Переменные (или признаки), поддающиеся регистрации, описывающие условия функционирования изучаемой реальной экономической системы и в существенной мере определяющие процесс формирования значений результирующих переменных. Как правило, часть из них поддается хотя бы частичному регулированию и управлению. Значения ряда объясняющих переменных могут задаваться как бы «извне» анализируемой системы. В этом случае их принято называть *экзогенными*. В регрессионном анализе они играют роль аргументов той функции, в качестве которой рассматривается анализируемый результирующий показатель y . По своей природе объясняющие переменные могут быть как случайными, так и неслучайными.

2.1.3. Функция регрессии y по $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$

Функция $f(X^*)$ называется *функцией регрессии y по X* (или просто — *регрессией y по X*), если она описывает изменение условного среднего значения результирующей переменной y (при условии, что значения объясняющих переменных X зафиксированы на уровнях X^*) в зависимости от изменения значений X^* объясняющих переменных. Соответственно математически это определение может быть записано в виде

$$f(X^*) = E(y | X = X^*). \quad (2.1)$$

В дальнейшем в целях упрощения обозначений правую часть (2.1) мы будем записывать просто $E(y | X^*)$ или даже $E(y | X)$ (в последнем варианте не различаются обозначения *самих объясняющих переменных* X и их *возможных значений* X^* , однако, с учетом данного пояснения, это не должно приводить к неправильному пониманию). Поэтому сокращенно функция регрессии может быть определена также соотношением

$$f(X) = E(y | X). \quad (2.1')$$

З а м е ч а н и е об этимологии слова «регрессия». Строго говоря, по своей смысловой нагрузке слово «регрессия» не имеет отношения к существу стохастических связей, для описания которых оно используется. Объяснение этому термину можно дать, лишь обратившись к истории исследований в области статистического анализа связей между признаками. Одним из первых примеров исследований такого рода была работа шведских статистиков, пытавшихся по наблюдениям значений пар признаков: x — отклонение от среднего уровня в росте отца; y — отклонение от среднего уровня в росте взрослого сына этого отца, — установить и описать стохастическую связь, существующую между x и y . В процессе исследования была подтверждена естественная гипотеза о наличии положительной статистической связи между ростом отца и сына («у высоких отцов в среднем высокие сыновья, и наоборот»), однако одновременно была подмечена *тенденция регрессии* (отступления, возврата) в росте сыновей к среднему уровню, а именно: «у очень высоких отцов сыновья в среднем высокие, но уже не такие высокие, как отцы, и наоборот: у очень маленьких по росту отцов сыновья в среднем низкорослые, но все-таки повыше, чем их отцы». Функцию, описывающую эту закономерность, авторы называли *функцией регрессии*, после чего этот термин и стали использовать применительно к *любой* функции, построенной аналогичными методами.

Как было упомянуто выше, в соотношениях (2.1)–(2.1') объясняющие переменные X могут быть как *случайными величинами*, так и *неслучайными параметрами*, от значений которых зависит закон распределения вероятностей случайной результирующей переменной y .

Пусть X — векторная *случайная* величина, а $p_{y,X}(\tilde{y}, \tilde{X})$ — функция плотности (или полигон вероятностей), задающая совместный закон распределения вероятностей многомерной случайной величины (y, X) в точке (\tilde{y}, \tilde{X}) (см. том 1 п. 2.5.3). Тогда, определив в соответствии с (2.13) (том 1) условную плотность $p_y(\tilde{y} | X = X^*)$ результирующей переменной y при условии, что объясняющие случайные переменные зафиксированы на уровнях X^* , мы получаем интерпретацию правой части (2.1) в виде ма-

тематического ожидания y , подсчитанного по *условному* распределению $p_y(\tilde{y} | X = X^*)$, т. е.

$$E(y | X = X^*) = \int \tilde{y} p_y(\tilde{y} | X = X^*) d\tilde{y} \quad (2.2)$$

(интегрирование в правой части (2.2) ведется по всем возможным значениям \tilde{y} результирующей переменной y).

Если же $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$ — ряд *неслучайных* параметров, от значений которых зависит одномерный закон распределения вероятностей $p_y(\tilde{y}; x^{(1)}, \dots, x^{(p)})$ результирующей переменной y , то правые части (2.1), (2.1') интерпретируются как математические ожидания y , подсчитанные по *безусловному* распределению $p_y(\tilde{y}; x^{(1)}, \dots, x^{(p)})$ при значениях параметров, равных соответственно X^* и X , т. е.

$$E(y | X = X^*) = \int \tilde{y} p_y(\tilde{y}; x^{(1)*}, x^{(2)*}, \dots, x^{(p)*}) d\tilde{y};$$

$$E(y | X) = \int \tilde{y} p_y(\tilde{y}; x^{(1)}, x^{(2)}, \dots, x^{(p)}) d\tilde{y}. \quad (2.2')$$

2.1.4. Уравнения регрессионной связи между y и X

Выше было отмечено, что в регрессионном анализе результирующая переменная y выступает в роли функции, значения которой определяются (*правда, с некоторой случайной погрешностью*) значениями объясняющих переменных $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$, выступающих в роли аргументов этой функции. Математически это может быть выражено в виде уравнений регрессионной связи (см. также выше, соотношения (10.3)–(10.4)):

$$\begin{cases} y(X) = f(X) + \varepsilon(X), \\ E\varepsilon(X) \equiv 0. \end{cases} \quad (2.3)$$

Присутствие случайной «остаточной» составляющей («*регрессионный остаток*») $\varepsilon(X)$ в первом соотношении уравнений (2.3) обусловлено причинами двойкой природы: во-первых, она отражает влияние на формирование значений y факторов, не учтенных в перечне объясняющих переменных X ; во-вторых, она может включать в себя случайную погрешность в измерении значения результирующего показателя y (даже в «идеальной» ситуации, когда по значениям объясняющих переменных X в принципе можно было бы *однозначно* восстановить значение анализируемой результирующей переменной).

Второе соотношение (тождество) в уравнениях (2.3) непосредственно следует из смысла функции регрессии $f(X) = \mathbb{E}(y | X)$, поскольку усреднение (вычисление математического ожидания) левых и правых частей первого из соотношений (2.3) при любом фиксированном значении X дает

$$\mathbb{E}(y(X) | X) = \mathbb{E}(f(X)) + \mathbb{E}(\varepsilon(X)).$$

А так как $\mathbb{E}(y(X) | X) = f(X)$ по определению и $\mathbb{E}(f(X)) = f(X)$ (поскольку величина $f(X)$ при фиксированных значениях X не является случайной), то $\mathbb{E}(\varepsilon(X)) = 0$ при любом фиксированном значении X .

Спецификация и способ статистического анализа моделей типа (2.3) зависят от конкретизации требований к виду функции $f(X)$, природе объясняющих переменных X и стохастических регрессионных остатков $\varepsilon(X)$. Ниже в данной главе рассматриваются различные варианты конкретизации этих требований.

2.1.5. Измеритель степени тесноты статистической связи между y и X

Выше, в п. 11.2.1 первого тома, с помощью соотношений (11.8)–(11.9) была введена универсальная характеристика *степени тесноты статистической связи* (с.т.с.с.), существующей между результирующей переменной y и объясняющими переменными X , — *коэффициент детерминации* $K_d(y, X)$. Из результатов, приведенных в п. 11.2, следовало, что в рамках исследования *линейных* регрессионных связей типа (2.3) (т.е. при функциях регрессии $f(X)$ вида $f(X) = \theta_0 + \theta_1 x^{(1)} + \dots + \theta_p x^{(p)}$, где $\theta_0, \theta_1, \dots, \theta_p$ — некоторые числовые параметры) коэффициент детерминации совпадает с квадратом множественного коэффициента корреляции $R_{y,X}^2$, определяемого соотношениями (11.34) и (11.35), том 1. Принимая во внимание, что *точность* (*прогностическая сила*) регрессионной модели определяется величиной условной дисперсии результирующего показателя $D(y | X)$ (она характеризует разброс индивидуальных значений $y(X)$ около значения, лежащего на линии регрессии $f(X)$, при каждом фиксированном значении X), а усредненная (по всем значениям X) условная дисперсия $\mathbb{E}_x(D(y | X))$ определяется формулой $\mathbb{E}_x(D(y | X)) = (1 - K_d(y, X))Dy$ (см. том 1 соотношения (11.8)–(11.9)), то можно сделать вывод, что *величина коэффициента детерминации $K_d(y, X)$ (или квадрата множественного коэффициента корреляции $R_{y,X}^2$ при исследовании линейных моделей) является решающей характеристикой прогностической силы анализируемой регрессионной модели.*

2.1.6. Исходные статистические данные

Все выводы в регрессионном анализе, так же как и в любом статистическом исследовании, строятся на основании имеющихся *исходных статистических данных*.

В регрессионном анализе используются данные *типа «объект-свойства»* (см. том 1, п. 9.1.2, соотношения (9.1) и (9.1')). Поскольку в регрессионном анализе принято обозначать результирующие переменные иначе, чем объясняющие (первые обычно обозначаются латинской буквой y , а вторые — x), то мы несколько видоизменим общие обозначения, принятые в п. 9.1.2.

Итак, будем полагать в дальнейшем (в данной главе), что мы располагаем результатами регистрации значений анализируемых объясняющих ($x^{(1)}, x^{(2)}, \dots, x^{(p)}$) и результирующей (y) переменных на n статистически обследованных объектах. Так что, если i — номер обследованного объекта, то имеющиеся исходные статистические данные состоят из n строк вида

$$(x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)}; y_i), \quad i = 1, 2, \dots, n, \quad (2.4)$$

где $x_i^{(j)}$ и y_i — значения соответственно j -й объясняющей переменной ($j = 1, 2, \dots, p$) и результирующего показателя, зарегистрированные на i -м обследованном объекте.

Из чисто технических соображений (смысл которых станет понятным несколько позже) данные (2.4) в регрессионном анализе обычно представляют в виде двух матриц вида:

$$X = \begin{pmatrix} 1 & x_1^{(1)} & \dots & x_1^{(p)} \\ 1 & x_2^{(1)} & \dots & x_2^{(p)} \\ \dots & \dots & \dots & \dots \\ 1 & x_n^{(1)} & \dots & x_n^{(p)} \end{pmatrix} \quad - \quad (2.4^a)$$

матрица размера $n \times (p + 1)$, составленная из наблюдаемых значений объясняющих переменных, и

$$Y = (y_1, y_2, \dots, y_n)^T \quad - \quad (2.4^b)$$

матрица размера $n \times 1$ (т.е. вектор-столбец высоты n), составленная из наблюдаемых значений результирующей переменной.

Возможны ситуации, когда данные регистрируются *на одном и том же объекте*, но в разные периоды («такты») времени. Тогда i будет означать номер периода времени, к которому «привязаны» соответствующие данные, а n — общее число тактов времени, в течение которых

собирались исходные данные (случай «временной» выборки в отличие от предыдущей — «пространственной»).

Наконец, возможна ситуация, когда «отслеживается» каждый из n объектов в течение N тактов времени («пространственно-временная» выборка, или «панельные данные»). В любой из упомянутых ситуаций исходные данные могут быть представлены в конечном счете в форме $(2.4^a) - (2.4^b)$, которую мы и примем за базовую в дальнейшем изложении.

2.1.7. Основные задачи прикладного регрессионного анализа

В п. 10.2 тома 1 описаны основные типы конечных прикладных целей, которые преследует регрессионный анализ. Резюмируя содержание этого пункта, отметим лишь, что в конечном счете анализ регрессионных зависимостей вида (2.3), базирующийся на исходных статистических данных $(2.4^a) - (2.4^b)$, нацелен на решение следующих основных задач:

Задача 1. Для любых заданных значений объясняющих переменных $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$ построить наилучшие в определенном смысле точечные и интервальные (с доверительной вероятностью P) оценки соответственно $\hat{f}(X)$ и $\Delta[f(X)]_P$ для неизвестной функции регрессии $f(X)$.

Задача 2. По заданным значениям объясняющих переменных $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$ построить наилучший в определенном смысле точечный и интервальный (с доверительной вероятностью P) прогноз соответственно $\hat{y}(X)$ и $\Delta[y(X)]_P$ для неизвестного значения результирующей переменной $y(X)$.

Задача 3. Пусть известно, что искомая функция регрессии принадлежит некоторому параметрическому семейству функций $\{f(X; \Theta)\}$, где $\Theta = (\theta_0, \theta_1, \dots, \theta_k)^T$ — векторный параметр, все или некоторые компоненты которого *допускают определенную экономическую интерпретацию*. Требуется построить наилучшие в определенном смысле точечные и интервальные оценки для неизвестных значений этих параметров.

Задача 4. Оценить удельный вес влияния каждой из объясняющих переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ на результирующий показатель $y(X)$ и, в частности, определить, какие из объясняющих переменных можно исключить из модели (2.3) как практически не влияющие на процесс формирования значений результирующего показателя.

* * *

Итак, собственно регрессионный анализ начинается с решения задачи 1 и, в частности, с конструирования по исходным данным вида $(2.4^a) -$

(2.4^б) оценки $\hat{f}(X)$ для неизвестной функции регрессии $f(X) = E(y | X)$. Исходным (и по существу ключевым) этапом в решении этой задачи следует признать выбор параметрического семейства функций $F = \{f(X; \Theta)\}$ — класса допустимых решений, в рамках которого предполагается вести поиск наилучшей (в определенном смысле) аппроксимации $\hat{f}(X)$ для $f(X)$. Этому этапу регрессионного анализа был посвящен специальный п. 10.7 первого тома данного учебника. В нем же приведены «некоторые общие рекомендации» (см. п. 10.7.4), среди которых есть тезис о том, что «не следует гнаться за чрезмерной сложностью функции, описывающей поведение искомой функции регрессии». Следуя этой логике, при подборе общего вида функции регрессии, как правило, идут от простого к сложному, т. е. *начинают с анализа возможности использовать простейшую линейную модель вида (10.22')*. Именно эта логика положена в основу изложения материала данной главы.

2.2. Классическая линейная модель множественной регрессии (КЛММР)

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).¹ В рамках КЛММР эти требования формулируются следующим образом:

$$\left\{ \begin{array}{l} y_i = \theta_0 + \theta_1 x_i^{(1)} + \dots + \theta_p x_i^{(p)} + \varepsilon_i, \quad i = 1, 2, \dots, n; \\ E\varepsilon_i = 0, \quad i = 1, 2, \dots, n; \\ E(\varepsilon_i \varepsilon_j) = \begin{cases} \sigma^2 & \text{при } i = j, \\ 0 & \text{при } i \neq j; \end{cases} \\ (x^{(1)}, x^{(2)}, \dots, x^{(p)}) \text{ — неслучайные переменные;} \\ \text{ранг матрицы } X = p + 1 < n \\ \text{(матрица } X \text{ определена соотношением (2.4}^a \text{))}. \end{array} \right. \quad (2.5)$$

Из (2.5) следует, что в рамках КЛММР рассматриваются только *ли-*

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. 1.2.3).

нейные функции регрессии, т. е.

$$f(X) = E(y | X) = \theta_0 + \theta_1 x^{(1)} + \dots + \theta_p x^{(p)}, \quad (2.6)$$

где объясняющие переменные $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ играют роль *неслучайных 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 10.1 из тома 1).

Кроме того, постулируется *взаимная некоррелированность* случайных регрессионных остатков ($E(\varepsilon_i \varepsilon_j) = 0$ 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 (2.4^а)–(2.4^б), т. е. о ситуациях, когда значения анализируемых переменных регистрируются на различных объектах (индивидуумах, семьях, 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, т. е. равнялся бы числу столбцов этой матрицы, которое в свою очередь должно быть меньше числа ее строк (т. е. общего числа имеющихся наблюдений). Случай $p + 1 \geq n$ 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 вы-

ражена 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 $X^T X$ (т. е. $\det(X^T X) = 0$), 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 (2.5). При этом кроме обозначений (2.4^а)–(2.4^б) введем также матрицы (векторы):

$$I_n = \begin{pmatrix} 1 & & & 0 \\ & 1 & & \\ & & \dots & \\ 0 & & & 1 \end{pmatrix} \quad - \quad (2.7)$$

единичная матрица размерности $n \times n$;

$$\Theta = (\theta_0, \theta_1, \dots, \theta_p)^T \quad - \quad (2.8)$$

вектор-столбец неизвестных значений параметров;

$$\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T \quad - \quad (2.9)$$

вектор-столбец регрессионных остатков;

$$0_n = (0, 0, \dots, 0)^T \quad - \quad (2.10)$$

вектор-столбец высоты n , состоящий из одних нулей;

$$\Sigma_\varepsilon = E(\varepsilon\varepsilon^T) = \begin{pmatrix} E(\varepsilon_1^2) & E(\varepsilon_1\varepsilon_2) & \dots & E(\varepsilon_1\varepsilon_n) \\ E(\varepsilon_2\varepsilon_1) & E(\varepsilon_2^2) & \dots & E(\varepsilon_2\varepsilon_n) \\ \dots & \dots & \dots & \dots \\ E(\varepsilon_n\varepsilon_1) & E(\varepsilon_n\varepsilon_2) & \dots & E(\varepsilon_n^2) \end{pmatrix} \quad - \quad (2.11)$$

ковариационная матрица размерности $n \times n$ вектора остатков;

$$\hat{\theta} = (\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_p)^T \quad - \quad (2.12)$$

вектор-столбец оценок неизвестных значений параметров;

$$\Sigma_{\hat{\theta}} = E[(\hat{\Theta} - \Theta)(\hat{\Theta} - \Theta)^T] = (\sigma_{ij}(\hat{\Theta})), \quad i, j = 0, 1, 2, \dots, p, \quad - \quad (2.13)$$

ковариационная матрица размерности $(p+1) \times (p+1)$ вектора несмещенных оценок $\hat{\Theta}$ неизвестных параметров Θ (в соотношении (2.13) $\sigma_{ij}(\Theta) = E[(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j)]$).

Тогда матричная форма записи КЛММР имеет вид:

$$\left\{ \begin{array}{l} Y = X\Theta + \epsilon, \\ E\epsilon = 0_n, \\ \Sigma_\epsilon = \sigma^2 I_n, \\ (x^{(1)}, x^{(2)}, \dots, x^{(p)}) \text{ — неслучайные переменные;} \\ \text{ранг матрицы } X = p + 1 < n. \end{array} \right. \quad (2.5')$$

Когда дополнительно к условиям (2.5) (или (2.5')) постулируют *нормальный* характер распределения регрессионных остатков $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)^T$ (что записывается в виде $\epsilon \in N_n(0; \sigma^2 I_n)$), то говорят, что y и X связаны *нормальной* КЛММР.

Пример 2.1. 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).

Исходные данные для 20 сельскохозяйственных районов области приведены в табл. 2.1.

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. 2.1. Вектор-столбец Y определяется 2-м столбцом табл. 2.1. 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 (2.5').

Таблица 2.1

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

2.3. Оценивание неизвестных параметров КЛММР: метод наименьших квадратов и метод максимального правдоподобия

Соотношения (2.5) и (2.5') 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 (2.4^а)–(2.4^б). Ниже описываются способы статистического оценивания параметров Θ и σ^2 в рамках КЛММР (метод наименьших квадратов) и в рамках *нормальной* КЛММР (метод максимального правдоподобия).

2.3.1. 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 *сглаженные* (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 сглаженного и наблюдаемого (в i -м наблюдении) значений результирующего показателя разность

$$\hat{\varepsilon}_i = y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i^{(1)} - \dots - \hat{\theta}_p x_i^{(p)} \quad (2.14)$$

(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

$$Q(\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_p) = \sum_{i=1}^n (y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i^{(1)} - \dots - \hat{\theta}_p x_i^{(p)})^2 = \sum_{i=1}^n \hat{\varepsilon}_i^2. \quad (2.15)$$

Очевидно, величина Q будет определяться при заданной системе наблюдений (2.4^а)–(2.4^б) конкретным выбором значений оценок параметров $\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_p$. *Оценки по методу наименьших квадратов (МНК-оценки) $\hat{\theta}_{0.мнк}, \hat{\theta}_{1.мнк}, \dots, \hat{\theta}_{p.мнк}$ как раз и подбираются таким образом, чтобы минимизировать величину Q , определенную соотношением (2.15), т. е.*

$$Q(\hat{\theta}_{0.мнк}, \hat{\theta}_{1.мнк}, \dots, \hat{\theta}_{p.мнк}) = \min_{\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_p} Q(\hat{\theta}_0, \hat{\theta}_1, \dots, \hat{\theta}_p) \quad (2.16)$$

или

$$\hat{\Theta}_{мнк} = \underset{\Theta}{\operatorname{arg\,min}} Q(\hat{\Theta}). \quad (2.16')$$

Опишем процедуру решения оптимизационной задачи (2.16'). Начнем с простейшего частного случая, когда рассматривается зависимость y от *единственной* объясняющей переменной x (т. е. $p = 1$). Этот случай в литературе обычно называют моделью *парной линейной регрессии*. Первое из уравнений связи (2.5) 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:

$$Q(\hat{\theta}_0, \hat{\theta}_1) = \sum_{i=1}^n (y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i)^2. \quad (2.15')$$

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:

$$\begin{cases} n\hat{\theta}_0 + \hat{\theta}_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i, \\ \left(\sum_{i=1}^n x_i \right) \hat{\theta}_0 + \hat{\theta}_1 \cdot \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i. \end{cases} \quad (2.17)$$

Система (2.17) 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:

$$\begin{aligned} \hat{\theta}_{1. \text{мнк}} &= \frac{n \sum_{i=1}^n x_i y_i - (\sum_{i=1}^n x_i) (\sum_{i=1}^n y_i)}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \\ \hat{\theta}_{0. \text{мнк}} &= \bar{y} - \hat{\theta}_{1. \text{мнк}} \bar{x}, \end{aligned} \quad (2.18)$$

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;

$$Q(\hat{\Theta}) = \sum_{i=1}^n (y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i^{(1)} - \dots - \hat{\theta}_p x_i^{(p)})^2 = (Y - X\hat{\Theta})^T (Y - X\hat{\Theta}) - \quad (2.19)$$

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 (2.19):

$$Q(\hat{\Theta}) = Y^T Y - 2\hat{\Theta}^T X^T Y + \hat{\Theta}^T X^T X \hat{\Theta}. \quad (2.19')$$

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 (2.16'), получаются дифференцированием правой части (2.19') 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. Приложение 2):

$$\frac{\partial Q(\hat{\Theta})}{\partial \hat{\Theta}} = -2X^T Y + 2X^T X \hat{\Theta} = \mathbf{0}_{p+1}. \quad (2.20)$$

В (2.20) $\mathbf{0}_{p+1}$ — это вектор-столбец размерности $p+1$, состоящий из одних нулей.

Разрешая систему уравнений (2.20) относительно $\hat{\Theta}$, получаем:

$$X^T X \hat{\Theta} = X^T Y \quad (2.21)$$

и следовательно:

$$\hat{\Theta}_{\text{мнк}} = (X^T X)^{-1} X^T Y. \quad (2.22)$$

В основной формуле метода наименьших квадратов (2.22) мы воспользовались невырожденностью матрицы $X^T X$, которая следует из требования максимального ранга для матрицы X , входящего в описание КЛММР (см. (2.5) или (2.5')).

Применим общую формулу (2.21) 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 (2.21), получаем:

$$\begin{cases} n\hat{\theta}_0 + \hat{\theta}_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i, \\ \left(\sum_{i=1}^n x_i \right) \hat{\theta}_0 + \hat{\theta}_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i, \end{cases}$$

т.е. ту же стандартную форму нормальных уравнений, с которой мы встретились в (2.17), анализируя эту же модель покомпонентно.

Пример 2.1 (продолжение). Применение формулы (2.22) к данным табл. 2.1 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:

$$\hat{f}(X) = 3,515 - 0,006x^{(1)} + 15,542x^{(2)} + 0,110x^{(3)} + 4,475x^{(4)} - 2,932x^{(5)}. \quad (2.23)$$

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.

2.3.2. 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 имеющихся выборочных данных (см. том 1, п. 7.5.1). Поэтому, если мы проводим регрессионный анализ в рамках *нормальной* КЛММП, т.е. если дополнительно

к условиям (2.5) 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. том 1, Свойство 5 в п. 11.2.2), можно выписать функцию правдоподобия в терминах остатков (2.14) (см. п. 7.2 в томе 1):

$$\begin{aligned} L(\hat{\varepsilon}_1, \hat{\varepsilon}_2, \dots, \hat{\varepsilon}_n \mid \Theta; \sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y_i - \theta_0 - \theta_1 x_i^{(1)} - \dots - \theta_p x_i^{(p)})^2} \\ &= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp \left[-\frac{1}{2\sigma^2} (Y - X\Theta)^T (Y - X\Theta) \right]. \end{aligned} \quad (2.24)$$

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 приравнением к нулю производных функции l по Θ и σ^2 :

$$\begin{aligned} l(\hat{\varepsilon}_1, \hat{\varepsilon}_2, \dots, \hat{\varepsilon}_n \mid \Theta; \sigma^2) &= -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} (Y - X\Theta)^T (Y - X\Theta); \\ \begin{cases} \frac{\partial l}{\partial \Theta} = -\frac{1}{2\sigma^2} \frac{\partial}{\partial \Theta} [(Y - X\Theta)^T (Y - X\Theta)] = \mathbf{0}_{p+1}, \\ \frac{\partial l}{\partial \sigma^2} = -\frac{n}{2} \cdot \frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2} (Y - X\Theta)^T (Y - X\Theta) = 0. \end{cases} \end{aligned} \quad (2.25)$$

Первая строка (2.25) после сокращения левой и правой частей на $-1/2\sigma^2$ повторяет систему уравнений (2.20) метода наименьших квадратов. Следовательно;

$$\hat{\Theta}_{\text{ММП}} = \hat{\Theta}_{\text{МНК}} = (X^T X)^{-1} X^T Y. \quad (2.26)$$

Вторая строка системы (2.25) позволяет вычислить ММП-оценку для σ^2 :

$$\hat{\sigma}_{\text{ММП}}^2 = \frac{1}{n} (Y - X\hat{\Theta})^T (Y - X\hat{\Theta}) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i^{(1)} - \dots - \hat{\theta}_p x_i^{(p)})^2, \quad (2.27)$$

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 (2.26), мы будем обозначать просто $\hat{\Theta}$ (без индексирования метода).

2.3.3. Анализ вариации результирующего показателя y и выборочный коэффициент детерминации $\hat{R}_{y,x}^2$

Из структуры КЛММР (2.5) 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 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

$$\begin{aligned} \text{вар}(y) &= (Y - \bar{Y})^T (Y - \bar{Y}) = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{f}_i + \hat{f}_i - \bar{y})^2 \\ &= \sum_{i=1}^n (y_i - \hat{f}_i)^2 + \sum_{i=1}^n (\hat{f}_i - \bar{y})^2 + 2 \sum_{i=1}^n (y_i - \hat{f}_i)(\hat{f}_i - \bar{y}) \\ &= \sum_{i=1}^n \hat{\varepsilon}_i^2 + \text{вар}(\hat{f}) + 2 \sum_{i=1}^n \hat{\varepsilon}_i (\hat{f}_i - \bar{y}), \end{aligned} \quad (2.27^a)$$

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 (2.27^а) равна нулю. С этой целью докажем сначала, что невязки $\hat{\varepsilon}_i$ ортогональны ко всем объясняющим переменным (включая искусственную переменную $x_i^{(0)} \equiv 1$), т. е. что

$$\sum_{i=1}^n \hat{\varepsilon}_i = 0 \quad \text{и} \quad \sum_{i=1}^n \hat{\varepsilon}_i x_i^{(j)} = 0 \quad \text{для всех} \quad j = 1, 2, \dots, p. \quad (2.27^б)$$

Действительно:

$$\begin{aligned} \mathbf{X}^T \hat{\boldsymbol{\varepsilon}} &= \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1^{(1)} & x_2^{(1)} & \dots & x_n^{(1)} \\ \dots & \dots & \dots & \dots \\ x_1^{(p)} & x_2^{(p)} & \dots & x_n^{(p)} \end{pmatrix} \begin{pmatrix} \hat{\varepsilon}_1 \\ \hat{\varepsilon}_2 \\ \vdots \\ \hat{\varepsilon}_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n \hat{\varepsilon}_i \\ \sum_{i=1}^n \hat{\varepsilon}_i x_i^{(1)} \\ \vdots \\ \sum_{i=1}^n \hat{\varepsilon}_i x_i^{(p)} \end{pmatrix} = \mathbf{X}^T (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\Theta}}) \\ &= \mathbf{X}^T \mathbf{Y} - \mathbf{X}^T \mathbf{X} \hat{\boldsymbol{\Theta}} = \mathbf{X}^T \mathbf{Y} - (\mathbf{X}^T \mathbf{X}) (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} = \mathbf{0}_{p+1}, \end{aligned}$$

что и означает справедливость равенств (2.27^б) (при выводе этого соотношения МНК-оценка $\hat{\boldsymbol{\Theta}}$ была заменена ее выражением по формуле (2.26); $\mathbf{0}_{p+1}$, как и прежде, означает вектор-столбец размерности $p+1$, состоящий из одних нулей).

Теперь мы можем показать, что

$$\sum_{i=1}^n \hat{\varepsilon}_i (\hat{f}_i - \bar{y}) = 0. \quad (2.27^в)$$

Действительно, учитывая равенства (2.27^б), имеем:

$$\begin{aligned} \sum_{i=1}^n \hat{\varepsilon}_i (\hat{f}_i - \bar{y}) &= \sum_{i=1}^n \hat{\varepsilon}_i (\hat{\theta}_0 + \hat{\theta}_1 x_i^{(1)} + \dots + \hat{\theta}_p x_i^{(p)} - \bar{y}) \\ &= \hat{\theta}_0 \sum_{i=1}^n \hat{\varepsilon}_i + \hat{\theta}_1 \sum_{i=1}^n \hat{\varepsilon}_i x_i^{(1)} + \dots \\ &\quad + \hat{\theta}_p \sum_{i=1}^n \hat{\varepsilon}_i x_i^{(p)} - \bar{y} \sum_{i=1}^n \hat{\varepsilon}_i = 0. \end{aligned} \quad (2.27^г)$$

Возвращаясь к (2.27^а) и учитывая (2.27^г), имеем:

$$\text{вар}(y) = \text{вар}(f) + \sum_{i=1}^n \hat{\varepsilon}_i^2. \quad (2.27^д)$$

Поделив обе части (2.27^б) на $\text{вар}(y)$, получаем:

$$\frac{\text{вар}(f)}{\text{вар}(y)} = 1 - \frac{\sum \hat{\varepsilon}_i^2}{\text{вар}(y)} \quad (2.27^e)$$

или

$$\hat{R}_{y.X}^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i^{(1)} - \dots - \hat{\theta}_p x_i^{(p)})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}. \quad (2.27^ж)$$

Сравнив левые и правые части соотношений (2.27^е) и (2.27^ж), а последнее соотношение — с разложением (11.3) из тома 1 и ранее введенным коэффициентом детерминации (11.8), с учетом результата (11.37) можно сделать следующие выводы:

(i) Характеристика, определенная соотношением (2.27^ж), является выборочным аналогом коэффициента детерминации (11.8) и одновременно квадратом выборочного множественного коэффициента корреляции между y и X ;

(ii) Выборочный коэффициент детерминации $\hat{R}_{y.X}^2$, определенный соотношением (2.27^ж), учитывая соотношение (2.27^е), характеризует долю общей вариации результирующего признака y , объясненной поведением (вариацией) выборочной функции регрессии $\hat{f}(X) = \hat{\theta}_0 + \hat{\theta}_1 x^{(1)} + \dots + \hat{\theta}_p x^{(p)}$.

(iii) Принимая во внимание результат (11.37), том 1, для вычисления выборочного коэффициента детерминации $\hat{R}_{y.X}^2$ в рамках КЛММР можно использовать (помимо формулы (2.27^ж)) формулы (11.34) и (11.35), том 1, с заменой участвующих в них коэффициентов парной и частной корреляции их выборочными аналогами (оценками).

(iv) Соответственно определенный формулой (2.27^ж) выборочный коэффициент детерминации $\hat{R}_{y.X}^2$ обладает всеми описанными в п. 11.2.4 (см., в частности, соотношения (11.38)–(11.40), том 1) статистическими свойствами квадрата множественного коэффициента корреляции.

В заключение еще раз выпишем выведенное выше соотношение (2.27^а), но в *матричной форме*:

$$(Y - \bar{Y})^\top (Y - \bar{Y}) = (X\hat{\Theta} - \bar{Y})^\top (X\hat{\Theta} - \bar{Y}) + \hat{\varepsilon}^\top \hat{\varepsilon}. \quad (2.27^н)$$

2.3.4. Статистические свойства оценок параметров КЛММР

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 (т. е. для каждой выборки (2.4^а)–(2.4^б)) мы их будем вычислять по одним и тем же формулам, соответственно (2.26) и (2.27). Другими словами, наши оценки «ведут себя» как случайные величины (см. также п. 7.1.2, том 1). 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 $\hat{\sigma}_{\text{ММР}}^2$. В данном случае свойство состоятельности наших оценок (см. п. 7.1.3, том 1) определяется структурой матрицы X . Существуют различные формулировки условий (в терминах элементов матрицы X), при которых оценки $\hat{\Theta}$ и $\hat{\sigma}_{\text{ММР}}^2$ являются состоятельными. Пожалуй, наиболее удобное для приложений условие состоятельности оценок $\hat{\Theta}$ и $\hat{\sigma}_{\text{ММР}}^2$ — следующее (см., например, [Себер Дж., п. 3.2]):

оценки $\hat{\Theta}$ и $\hat{\sigma}_{\text{ММР}}^2$ 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 λ_{\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 $\hat{\sigma}_{\text{ММР}}^2$ для случая, например, парной линейной регрессионной зависимости (т. е. при $p = 1$) равносильно требо-

ванию:

$$\sum_{i=1}^n (x_i - \bar{x})^2 \asymp n \quad \text{при } n \rightarrow \infty, \quad (2.28)$$

где символ \asymp означает, что величины, стоящие слева и справа от него, ведут себя асимптотически одинаково при $n \rightarrow \infty$, т. е. :

$$\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} \rightarrow \text{const} \quad \text{при } n \rightarrow \infty.$$

Так, например, условие (2.28) 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 (см. п. 7.1.4, том 1), 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 (2.26), подставим в формулу (2.26) вместо Y его выражение из основного (первого) соотношения системы (2.5')

$$\hat{\Theta} = (X^T X)^{-1} X^T (X\Theta + \varepsilon) = \Theta + (X^T X)^{-1} X^T \varepsilon. \quad (2.29)$$

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 (2.29) с учетом того, что величины Θ и $(X^T X)^{-1} X^T$ неслучайны, а средние значения остатков равны нулю (т. е. $E\varepsilon = 0_n$), получаем:

$$E\hat{\Theta} = E\Theta + (X^T X)^{-1} X^T E\varepsilon = \Theta. \quad (2.30)$$

Тем самым показано, что МНК-оценки (они же ММП-оценки) $\hat{\Theta}$ неизвестных параметров КЛММР 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 (2.27), т. е. определить значение

$$E \left[\frac{1}{n} (Y - X\hat{\Theta})^\top (Y - X\hat{\Theta}) \right],$$

где $\hat{\Theta}$ определено формулой (2.26). С этой целью выразим вектор невязок $Y - X\hat{\Theta}$ через остаточные случайные компоненты модели ε :

$$\begin{aligned} Y - X\hat{\Theta} &= (X\Theta + \varepsilon) - X[(X^\top X)^{-1} X^\top (X\Theta + \varepsilon)] \\ &= \varepsilon - X(X^\top X)^{-1} X^\top \varepsilon = [I_n - X(X^\top X)^{-1} X^\top] \varepsilon = Z\varepsilon. \end{aligned} \quad (2.31)$$

В соотношении (2.31) использовано обозначение для $(n \times n)$ -матрицы

$$Z = I_n - X(X^\top X)^{-1} X^\top. \quad (2.31^a)$$

Отметим, что матрица Z является симметричной и *идемпотентной* (см. Приложение 2), т. е. $Z^\top = Z$ (симметричность) и $Z^2 = Z$ (идемпотентность). Оба свойства устанавливаются непосредственной проверкой. Действительно:

$$Z^\top = (I_n - X(X^\top X)^{-1} X^\top)^\top = I_n - X(X^\top X)^{-1} X^\top = Z$$

(при переходе к правой части использовались правило транспонирования произведения матриц, а также симметричность матрицы $(X^\top X)^{-1}$);

$$\begin{aligned} Z^2 &= ZZ = [I_n - X(X^\top X)^{-1} X^\top] [I_n - X(X^\top X)^{-1} X^\top] \\ &= I_n - X(X^\top X)^{-1} X^\top - X(X^\top X)^{-1} X^\top + [X(X^\top X)^{-1} X^\top] \\ &\quad \times [X(X^\top X)^{-1} X^\top] = I_n - X(X^\top X)^{-1} X^\top = Z. \end{aligned}$$

Теперь, воспользовавшись (2.31), а также симметричностью и идемпотентностью матрицы Z , мы можем вычислить величину

$$\begin{aligned} E[(Y - X\hat{\Theta})^\top (Y - X\hat{\Theta})] &= E[(Z\varepsilon)^\top (Z\varepsilon)] \\ &= E(\varepsilon^\top Z^\top Z\varepsilon) = E(\varepsilon^\top Z\varepsilon) = \sigma^2 \operatorname{tr} Z \\ &= \sigma^2 \operatorname{tr} [I_n - X(X^\top X)^{-1} X^\top] \\ &= \sigma^2 [\operatorname{tr} I_n - \operatorname{tr} X(X^\top X)^{-1} X^\top] \\ &= \sigma^2 [n - \operatorname{tr} (X^\top X)^{-1} (X^\top X)] \\ &= \sigma^2 (n - \operatorname{tr} I_{p+1}) = \sigma^2 [n - (p + 1)]. \end{aligned} \quad (2.32)$$

В ходе преобразований мы воспользовались следующими двумя свойствами (см. Приложение 2):

- (а) $E(\hat{\epsilon}^T A \epsilon) = \sigma^2 \text{tr} A$,
 (б) $\text{tr}(ABC) = \text{tr}(BCA) = \text{tr}(CAB)$.

Из соотношения (2.32) следует, что оценка $\hat{\sigma}_{\text{ММП}}^2$, определенная формулой (2.27), является смещенной, так как

$$\begin{aligned} E\hat{\sigma}_{\text{ММП}}^2 &= E \left[\frac{1}{n} (Y - X\hat{\Theta})^T (Y - X\hat{\Theta}) \right] \\ &= \frac{1}{n} E[(Y - X\hat{\Theta})^T (Y - X\Theta)] = \sigma^2 \left(1 - \frac{p+1}{n} \right). \end{aligned} \quad (2.33)$$

Воспользуемся вместо $\hat{\sigma}_{\text{ММП}}^2$ оценкой

$$\begin{aligned} \hat{\sigma}^2 &= \frac{n}{n-p-1} \hat{\sigma}_{\text{ММП}}^2 = \frac{1}{n-p-1} \Sigma (y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i^{(1)} - \dots - \hat{\theta}_p x_i^{(p)})^2 \\ &= \frac{1}{n-p-1} (Y - \hat{\Theta}X)^T (Y - \hat{\Theta}X). \end{aligned} \quad (2.34)$$

Очевидно, такой способ оценивания неизвестной дисперсии остатков ϵ_i (так называемой *остаточной дисперсии* σ^2) уже будет *несмещенным*.

Пример 2.1 (продолжение). Оценка остаточной дисперсии σ^2 по данным табл. 2.1 по формуле (2.34) с использованием результатов оценивания функции регрессии (2.23) дает

$$\hat{\sigma}^2 = 2,59. \quad (2.35)$$

Ковариационная матрица вектора $\hat{\Theta}$ МНК-оценок. Ковариационная матрица $\Sigma_{\hat{\Theta}}$ вектора $\hat{\Theta}$ была введена соотношением (2.13) при описании КЛММП в п.2.1. 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!). В п.2.6.6 был введен многомерный аналог дисперсии случайной величины — так называемая *обобщенная дисперсия*, количественно выражающая степень случайного разброса значений *векторной* случайной величины около своего среднего в соответствующем многомерном пространстве (см. формулу (2.30)).

Соответственно применительно к вектору оценок $\hat{\Theta}$ эта степень случайного разброса будет определяться величиной детерминанта ковариационной матрицы $\Sigma_{\hat{\Theta}}$, т. е.

$$D_{\text{о.о.}}\hat{\Theta} = \det \Sigma_{\hat{\Theta}}.$$

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. (2.33) и п. 11.2.2 в томе 1), который в свою очередь выражается через элементы ковариационной матрицы $\Sigma_{\hat{\Theta}} = (\sigma_{ij}(\hat{\Theta}))$, $-l, j = 0, 1, \dots, p$:

$$r(\hat{\theta}^{(l)}, \hat{\theta}^{(j)}) = \frac{\sigma_{lj}}{(\sigma_{ll}\sigma_{jj})^{1/2}}.$$

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 (2.4^a)–(2.4^b). Для этого при выражении разности $\hat{\Theta} - \Theta$ воспользуемся ранее полученным соотношением (2.29):

$$\begin{aligned} \Sigma_{\hat{\Theta}} &= \mathbf{E}[(\hat{\Theta} - \Theta)(\hat{\Theta} - \Theta)^T] \\ &= \mathbf{E}\{[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\varepsilon}] [(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\varepsilon}]^T\} \\ &= \mathbf{E}[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}] \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{E}(\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T) \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \sigma^2 \mathbf{I}_n \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1}. \end{aligned} \quad (2.36)$$

При выводе соотношения (2.36) мы воспользовались:

- правилом транспонирования произведения матриц (см. Приложение 2);
- специальным видом ковариационной матрицы вектора остатков $\Sigma_{\boldsymbol{\varepsilon}} = \mathbf{E}(\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T) = \sigma^2 \mathbf{I}_n$ (см. (2.5'));
- тем, что при умножении на единичную любая «подходящая»¹ матрица не меняется (в нашем случае $\mathbf{X}^T \mathbf{I}_n = \mathbf{X}^T$).

¹ Матрица \mathbf{A} как сомножитель, стоящий слева от \mathbf{I}_n , является «подходящей», если длина ее строки (т. е. число ее столбцов) равна размерности n единичной матрицы \mathbf{I}_n .

Заметим в заключение, что в статистической практике пользуются не самой матрицей $\Sigma_{\hat{\theta}}$, а ее оценкой

$$\hat{\Sigma}_{\hat{\theta}} = \hat{\sigma}^2 (\mathbf{X}^T \mathbf{X})^{-1}, \quad (2.36')$$

где $\hat{\sigma}^2$ определяется по формуле (2.34). Кроме того, в большинстве случаев ограничиваются выводом на печать и использованием только среднеквадратических ошибок

$$s_l = \hat{\sigma}_{ll}^{1/2}(\hat{\Theta}) = \hat{\sigma} \sqrt{a_{ll}} \quad (2.37)$$

оценок $\hat{\theta}_l$ ($l = 0, 1, \dots, p$); в соотношении (2.37) $\hat{\sigma}_{ll}(\hat{\Theta})$ — l -й диагональный элемент матрицы $\hat{\Sigma}_{\hat{\theta}}$ (см. (2.36')), $\hat{\sigma}$ определено формулой (2.34), а a_{ll} — l -й диагональный элемент матрицы $\mathbf{A} = (\mathbf{X}^T \mathbf{X})^{-1}$, $l = 0, 1, \dots, p$.

Пример 2.1 (продолжение). По формулам (2.34), (2.36'), (2.37) 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, имеет следующий вид:

$$\hat{f}(X) = \underset{(5,41)}{3,515} - \underset{(0,69)}{0,006} x^{(1)} + \underset{(21,59)}{15,542} x^{(2)} + \underset{(0,85)}{0,110} x^{(3)} + \underset{(1,54)}{4,475} x^{(4)} - \underset{(3,09)}{2,932} x^{(5)}. \quad (2.23')$$

В скобках под значениями оцененных коэффициентов регрессии $\hat{\theta}_l$ указаны их среднеквадратические ошибки s_l . Анализируя полученный результат, читатель даже на этой стадии подготовленности отметит для себя, что среднеквадратические ошибки в оценке всех коэффициентов, кроме θ_4 , 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 $E(\hat{\theta} - \theta)^2$ (см. п. 7.1.5). 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

$$E(\hat{\theta}_{\text{опт}} - \theta)^2 = \min_{\hat{\theta} \in M} E(\hat{\theta} - \theta)^2. \quad (2.38)$$

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 (2.38) 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 уместной по меньшей мере в двух из четырех основных задач прикладного регрессионного анализа (см. задачи 1 и 3 в начале гл. 2). 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 (2.5') ($C = (c_0, c_1, \dots, c_p)^T$ — 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 (2.38), т. е.*

$$E(C^T \hat{\Theta}_{\text{МНК}} - \theta_C)^2 \leq E(B^T Y - \theta_C)^2 \quad (2.39)$$

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 (2.26)), а во-вторых, она является несмещенной (так как $E(C^T \hat{\Theta}_{\text{МНК}}) = C^T E \hat{\Theta}_{\text{МНК}} = C^T \Theta$ в силу того, что $\hat{\Theta}_{\text{МНК}}$ является несмещенной оценкой параметра Θ , см. (2.30)). Наша цель — доказать справедливость неравенства (2.39), 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

$$B^T X = C^T. \quad (2.40)$$

Теперь вычислим и сравним средние квадраты ошибок для оценок $B^T Y$ и $C^T \hat{\Theta}_{\text{мнк}}$:

$$\begin{aligned} \mathbf{E}(B^T Y - C^T \Theta)^2 &= \mathbf{E}[B^T (X\Theta + \varepsilon) - C^T \Theta]^2 \\ &= \mathbf{E}(B^T X\Theta + B^T \varepsilon - C^T \Theta)^2 = \mathbf{E}(C^T \Theta + B^T \varepsilon - C^T \Theta)^2 \\ &= \mathbf{E}(B^T \varepsilon \varepsilon^T B) = B^T \mathbf{E}(\varepsilon \varepsilon^T) B = \sigma^2 (B^T B). \end{aligned} \quad (2.41)$$

При выводе соотношения (2.41) мы воспользовались равенством (2.40), специальным видом ковариационной матрицы вектора остатков ε ($\mathbf{E}(\varepsilon \varepsilon^T) = \sigma^2 \mathbf{I}_n$) и возможностью представления квадрата числа $B^T \varepsilon$ в виде $B^T \varepsilon \varepsilon^T B$.

$$\begin{aligned} \mathbf{E}(C^T \hat{\Theta}_{\text{мнк}} - C^T \Theta)^2 &= \mathbf{E}[C^T (\hat{\Theta}_{\text{мнк}} - \Theta)]^2 \\ &= \mathbf{E}[C^T (\hat{\Theta}_{\text{мнк}} - \Theta)(\hat{\Theta}_{\text{мнк}} - \Theta)^T C] \\ &= C^T \mathbf{E}[(\hat{\Theta}_{\text{мнк}} - \Theta)(\hat{\Theta}_{\text{мнк}} - \Theta)^T] C \\ &= \sigma^2 C^T (X^T X)^{-1} C = \sigma^2 B^T X (X^T X)^{-1} X^T B. \end{aligned} \quad (2.42)$$

При выводе соотношения (2.42) мы воспользовались возможностью представления квадрата числа $C^T (\hat{\Theta}_{\text{мнк}} - \Theta)$ в виде $C^T (\hat{\Theta}_{\text{мнк}} - \Theta)(\hat{\Theta}_{\text{мнк}} - \Theta)^T C$, знанием ковариационной матрицы вектора оценок $\hat{\Theta}_{\text{мнк}}$ (см (2.36)) и выражением C^T через $B^T X$ (см. (2.40)).

Остается показать, что разность правых частей (2.41) и (2.42) всегда неотрицательна. Действительно:

$$\begin{aligned} \mathbf{E}(B^T Y - C^T \Theta)^2 - \mathbf{E}(C^T \hat{\Theta}_{\text{мнк}} - C^T \Theta)^2 &= \sigma^2 (B^T B) \\ &\quad - \sigma^2 B^T X (X^T X)^{-1} X^T B = \sigma^2 B^T [\mathbf{I}_n - X (X^T X)^{-1} X^T] B \\ &= \sigma^2 B^T Z B, \end{aligned} \quad (2.43)$$

где матрица $Z = \mathbf{I}_n - X (X^T X)^{-1} X^T$ нам уже встречалась при анализе смещения оценки $\hat{\sigma}_{\text{ММП}}^2$ (см. (2.31^а) и (2.32)). Было установлено, что она — симметричная и идемпотентная. Кроме того, из хода преобразования в соотношении (2.32) следует, что матрица Z — неотрицательно

определенная, т. е. для любого вектора $B = (b_1, b_2, \dots, b_n)^T$ выражение $B^T ZB$ всегда неотрицательно (в ходе преобразований (2.32), показано, что сумма квадратов невязок, являющаяся по определению существенно неотрицательным числом, представима в виде $\varepsilon^T Z\varepsilon$, где вектор невязок $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$ может быть составлен из произвольных действительных компонент). А неотрицательность правой части (2.43) как раз и доказывает справедливость неравенства (2.39).

Свойства оценок, справедливые только при дополнительном условии нормальности регрессионных остатков (т. е. в рамках нормальной КЛММР). Если мы несколько сузим класс рассматриваемых моделей, добавив к условиям (2.5) КЛММР требование нормальности регрессионных остатков, то соответствующая нормальная КЛММР может быть кратко описана следующими условиями:

$$\left\{ \begin{array}{l} Y = X\Theta + \varepsilon, \\ \varepsilon \in N_n(0; \sigma^2 I_n), \\ (x^{(1)}, x^{(2)}, \dots, x^{(p)}) - \text{нечастные переменные;} \\ \text{ранг матрицы } X = p + 1. \end{array} \right. \quad (2.5'')$$

В (2.5'') $N_k(a; \Sigma)$, 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 показателя при заданных величинах объясняющих переменных (см. формулировки задач 1–3 в начале этой главы). *Необходимой базой для решения этих задач является знание законов распределения вероятностей используемых оценок.* Именно это знание доставляют нам следующие (справедливые в рамках нормальной КЛММР) результаты, являющиеся по существу многомерным обобщением известной теоремы Фишера о распределении выборочного среднего значения и выборочной дисперсии, построенных по выборке из нормальной генеральной совокупности (см. п. 6.2.8, соотношения (6.26)–(6.28) в томе 1):

- оценки $\hat{\Theta}_{\text{мнк}}$ подчиняются $(p+1)$ -мерному нормальному з.р.в. с вектором средних значений, равным истинным значениям анализируемых параметров Θ , и с ковариационной матрицей $\Sigma_{\hat{\Theta}}$, определяемой соотношением (2.36), т. е.

$$\hat{\Theta}_{\text{мнк}} \in N_{p+1}(\Theta; \sigma^2(\mathbf{X}^T \mathbf{X})^{-1}); \quad (2.44)$$

- случайная величина $(n-p-1)\hat{\sigma}^2/\sigma^2$ подчиняется χ^2 -распределению с $n-p-1$ степенями свободы, т. е.

$$(n-p-1)\frac{\hat{\sigma}^2}{\sigma^2} \in \chi^2(n-p-1); \quad (2.45)$$

- оценки $\hat{\Theta}_{\text{мнк}}$ и $\hat{\sigma}^2$ являются статистически независимыми. (2.45')

Доказательство справедливости (2.44) основано на следующих фактах: 1) среднее значение и ковариационная матрица вектора оценок $\hat{\Theta}_{\text{мнк}}$ были подсчитаны ранее (см. (2.30) и (2.36)); 2) оценки $\hat{\Theta}_{\text{мнк}}$ строятся как линейные комбинации нормально распределенных независимых случайных величин y_i или ϵ_i (это видно соответственно из (2.22) и (2.29)), а любая линейная комбинация независимых нормально распределенных случайных величин снова распределена по нормальному закону (что подтверждается, например, индуктивным применением формулы композиции, см. п. 4.4). Утверждение (2.45) и статистическая независимость оценок $\hat{\Theta}_{\text{мнк}}$ и $\hat{\sigma}^2$ следует, например, из теоремы 8.2.2 [Андерсон Т.].

Сформулируем одно важное следствие из приведенных выше результатов, которое позволит нам в дальнейшем воспользоваться t -распределением (см. том 1, п. 3.2.2) 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 (2.5''). Тогда статистика

$$\gamma_l = \frac{\hat{\theta}_{l, \text{мнк}} - \theta_l^0}{s_l} = \frac{\hat{\theta}_{l, \text{мнк}} - \theta_l^0}{\hat{\sigma} \sqrt{a_{ll}}}, \quad l = 0, 1, \dots, p, \quad (2.46)$$

распределена по закону Стьюдента (t -распределению) с $n-p-1$ степенями свободы. В (2.46) среднеквадратическая ошибка s_l оценки $\hat{\theta}_{l, \text{мнк}}$ определяется по формуле (2.37), $\hat{\sigma}^2$ — по формуле (2.34) и a_{ll} — это l -й диагональный элемент матрицы $\mathbf{A} = (\mathbf{X}^T \mathbf{X})^{-1}$.

$$\text{Правило проверки гипотез вида } H_0: \theta_l = \theta_l^0 \quad (2.47)$$

Теперь мы можем сформулировать основанное на (2.46) правило проверки гипотез H_0 . Пусть θ_l^0 — заданное гипотетическое значение l -го коэффициента регрессии. По заданному уровню значимости критерия α (см. том 1, п. 8.1.3 и п. 8.2) из таблиц процентных точек t -распределения (см. Приложение 1) находим $100\alpha/2\%$ -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 (2.46), больше значения $t_{\alpha/2}(n - p - 1)$, то проверяемую гипотезу (2.47) следует отвергнуть (вероятность ошибиться при этом равна α). В противном случае (т. е. 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 для неизвестного значения коэффициента регрессии θ_l

Приведенное выше следствие (2.46) дает нам возможность построить доверительный интервал для каждого «отдельно взятого» неизвестного коэффициента регрессии θ_l (о понятии доверительного интервала и подходе к их построению см. том 1, п. 7.4 и п. 7.5.4). В частности, из (2.46) следует, что с доверительной вероятностью $P = 1 - \alpha$ истинное значение θ_l l -го коэффициента регрессии должно удовлетворять неравенствам

$$\hat{\theta}_{l, \text{мнк}} - s_l t_{\alpha/2}(n - p - 1) \leq \theta_l \leq \hat{\theta}_{l, \text{мнк}} + s_l t_{\alpha/2}(n - p - 1). \quad (2.48)$$

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. 11.2.4 (том 1) при описании корреляционных характеристик линейной зависимости y от нескольких объясняющих переменных $x^{(1)}, x^{(2)}$,

..., $x^{(p)}$ был приведен результат, в соответствии с которым в предположении справедливости гипотезы

$$H_0: \theta_1 = \theta_2 = \dots = \theta_p = 0$$

статистика

$$\gamma = \frac{\widehat{R}_{y.X}^2}{1 - \widehat{R}_{y.X}^2} \cdot \frac{n - p - 1}{p} \quad (2.49)$$

должна подчиняться F -распределению с числами степеней свободы числителя и знаменателя, соответственно равными p и $n - p - 1$ (см. том 1, п. 3.2.3). Следовательно, гипотезу H_0 можно проверить следующим образом. По заданному уровню значимости критерия α определяем из таблиц Приложения 1 100 $\alpha\%$ -ную точку $F(p, n - p - 1)$ -распределения $v_\alpha^2(p, n - p - 1)$. Если окажется, что

$$\frac{\widehat{R}_{y.X}^2}{1 - \widehat{R}_{y.X}^2} \cdot \frac{n - p - 1}{p} > v_\alpha^2(p, n - p - 1), \quad (2.50)$$

то гипотеза об отсутствии линейной связи между y и X отвергается (с вероятностью ошибиться, равной α), и принимается — в противном случае.

Пример 2.1 (продолжение). Отдельная проверка статистически значимого отличия от нуля оценок коэффициентов $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_5$ в модели (2.23'), т. е. последовательное сравнение отношений $\hat{\theta}_i/s_i$ 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 $\widehat{R}_{y.X}^2$ для множественного коэффициента корреляции между урожайностью зерновых культур y и совокупностью пяти факторов сельскохозяйственного производства ($\widehat{R}_{y.X}^2 = 0,517$) и проверка гипотезы об отсутствии какой бы то ни была линейной связи между y и X с помощью статистики (2.49) ($\gamma = 0,517 \cdot 14 / 0,483 \cdot 5 = 7,238 / 2,415 = 3,00$; $v_{0.05}^2(5; 20) = 2,96$) говорят за то, чтобы продолжить изучение линейной связи между y и $(x^{(1)}, x^{(2)}, \dots, x^{(5)})$ в рамках КЛММП. 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 R (см. табл. 2.2), мы

вынуждены придти 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 2.2 Матрица 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

2.4. Мультиколлинеарность и отбор наиболее существенных объясняющих переменных в КЛММР

Речь идет о мультиколлинеарности (т.е. *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.

2.4.1. Признаки и причины мультиколлинеарности

Мультиколлинеарность в строгом смысле (или полная мультиколлинеарность) определяется условием нарушения одного из требований КЛММР, а именно, требования к рангу матрицы X (см. (2.5) или (2.5')): говорят, что объясняющие переменные модели $(x^{(1)}, x^{(2)}, \dots, x^{(p)})$ характеризуются свойством полной мультиколлинеарности, если ранг матрицы их наблюдаемых значений (2.4^а) меньше $p + 1$. Мы уже обсуждали эту ситуацию в начале главы, комментируя смысл условия «ранг матрицы $X = p + 1$ ». Напомним, что при нарушении этого условия между анализируемыми объясняющими переменными существует линейная *функциональная* связь (т.е. значения по меньшей мере одной из них могут быть выражены в виде линейной комбинации наблюдаемых значений остальных переменных), а *матрица* $X^T X$ *оказывается вырожденной*, т.е. ее определитель равен нулю (а значит не существует обратная матрица $(X^T X)^{-1}$,

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 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 R в примере 2.1 (см. табл. 2.2) 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 называемой *слабой обусловленности* матрицы $X^T X$, т. е. 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* λ_{\min} матрицы $X^T X$. Это объясняется двумя обстоятельствами. Во-первых, из близости к нулю λ_{\min} следует близость к нулю величины $\det(X^T X)$, и наоборот. Во-вторых, можно показать, что среднеквадратическая ошибка оценки $\hat{\theta}_{j, \min}$ обратно пропорциональна величине λ_{\min} (см., например, [Джонстон Дж., п. 5.7]). Поэтому, наряду с величиной $\det(X^T X)$ (или вместо нее), вычисляют и сравнивают с накапливающимися ошибками от округлений значение λ_{\min} , т. е. минимальный корень уравнения

$$|X^T X - \lambda I_{p+1}| = 0.$$

4) Анализ корреляционной матрицы \mathbf{R} позволяет лишь в первом приближении (и относительно поверхностно) судить о наличии/отсутствии мультиколлинеарности в наших исходных данных (2.4^а). 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. 10.2, соотношения (10.9)–(10.9')–(10.10) в томе 1);*
- *небольшое изменение исходных статистических данных (2.4^а)–(2.4^б) (добавление или изъятие небольшой порции наблюдений) 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 (2.46)), 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 (2.49) (именно с такой ситуацией мы столкнулись в примере 2.1).*

Подобные не поддающиеся содержательной интерпретации особенности модели можно понять и предвидеть, если обратиться к основным соотношениям (2.26) и (2.36), определяющим правила вычисления соответственно МНК-оценок $\hat{\Theta}_{\text{МНК}}$ и их ковариационной матрицы. И в той и в другой формуле присутствует матрица $(\mathbf{X}^T \mathbf{X})^{-1}$, элементы которой обратно пропорциональны величине определителя $\det(\mathbf{X}^T \mathbf{X})$. Если эта величина достаточно мала, то изъятие или дополнение одной-двух строк в матрице \mathbf{X} (что равносильно изъятию или дополнению одного-двух наблюдений в массив исходных данных (2.4^а)–(2.4^б)) может радикально (во много раз) изменить значение $\det(\mathbf{X}^T \mathbf{X})$, а следовательно, и зависящие от него значения $\hat{\Theta}_{\text{МНК}}$ и $\Sigma_{\hat{\Theta}}$. Одновременно малость величины $\det(\mathbf{X}^T \mathbf{X})$ влечет за собой непомерно большие значения диагональных элементов матрицы $\Sigma_{\hat{\Theta}}$

(т. е. дисперсий $s_j^2 = D(\hat{\theta}_j - \theta_j)^2$), что может приводить к статистически незначимым отклонениям от нуля критических значений $|\hat{\theta}_{j, \text{МНК}}|/s_j$.

2.4.2. Методы устранения мультиколлинеарности

I. Переход к смещенным методам оценивания. Как известно (см. в предыдущем пункте формулировку теоремы Гаусса–Маркова, соотношение (2.39)), 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 (см. рис. 2.1). Пусть оценка $\hat{\theta}_{\text{МНК}}$ — наилучшая среди *несмещенных* оценок неизвестного значения θ анализируемого параметра, т. е. $E(\hat{\theta}_{\text{МНК}} - \theta)^2 \leq E(\hat{\theta} - \theta)^2$ 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, т. е. если $|\hat{\theta} - \theta| \leq \Delta$, то оценка $\hat{\theta}$ считается «хорошей», а при $|\hat{\theta} - \theta| > \Delta$ — «плохой».

Простой визуальный анализ рис. 2.1 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 $f_{\hat{\theta}_{\text{МНК}}}(x)$, т. е. $E(\hat{\theta}_{\text{МНК}} - \theta)^2 = \int_{-\infty}^{\infty} (x - \theta)^2 f_{\hat{\theta}_{\text{МНК}}}(x) dx$)

будет превосходить средний квадрат ошибок, получаемых при оценивании с помощью смещенной оценки $\hat{\theta}_{\text{см}}$ (т. е. величину $E(\hat{\theta}_{\text{см}} - \theta)^2 = \int_{-\infty}^{\infty} (x - \theta)^2 f_{\hat{\theta}_{\text{см}}}(x) dx$).

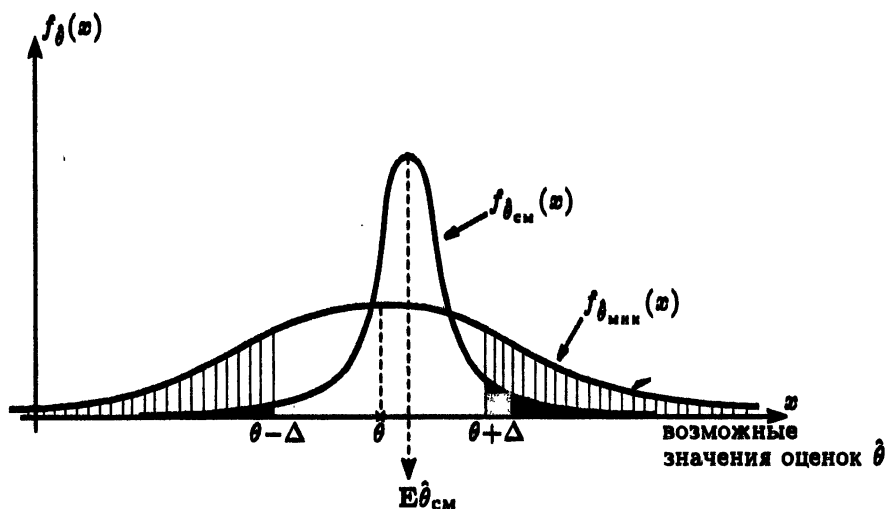


Рис. 2.1. Плотности распределения несмещенной ($f_{\hat{\theta}_{\text{МНК}}}(x)$) и смещенной ($f_{\hat{\theta}_{\text{см}}}(x)$) оценок истинного значения θ неизвестного параметра

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

$$\hat{\Theta}_{\tau} = (\mathbf{X}^T \mathbf{X} + \tau I_{p+1})^{-1} \mathbf{X}^T \mathbf{Y}. \quad (2.51)$$

Добавление к диагональным элементам матрицы $(\mathbf{X}^T \mathbf{X})$ «гребня», «хребта» τ (τ — некоторое небольшое положительное число) с одной стороны, делает получаемые при этом оценки смещенными, а с другой, — превращает матрицу $\mathbf{X}^T \mathbf{X}$ из «плохо обусловленной» в «хорошо обусло-

вленную». 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 (2.51) 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 11.3 и 11.4, том 1) 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. том 1, п. 13.2). Соответствующий прием известен в литературе под названием «*построение регрессии на главных компонентах*». Ссылка на п. 13.2 требует пояснений. В п. 13.2 (том 1) описывается метод построения главных компонент *многомерной случайной величины* $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$, 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 п. 13.2 построения и результаты, мы должны сделать, тем не менее, ряд пояснений:

(i) 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 п. 13.2, перейдем к *центрированным* наблюдениям $X_{\text{цл}} = (x_i^{(1)} - \bar{x}^{(1)}, x_i^{(2)} - \bar{x}^{(2)}, \dots, x_i^{(p)} - \bar{x}^{(p)})^T$,

где $\bar{x}^{(j)} = \sum_{i=1}^n x_i^{(j)}/n$, а вместо теоретической ковариационной матрицы $\Sigma = E[(X - EX)(X - EX)^T]$, относительно которой решается задача определения собственных чисел и собственных векторов, будем рассматривать $(p \times p)$ -матрицу $X_{\text{ц}}^T X_{\text{ц}}$, где $X_{\text{ц}}$ — это $(n \times p)$ -матрица *центрированных* наблюдений, т. е.

$$X_{\text{ц}} = \begin{pmatrix} x_1^{(1)} - \bar{x}^{(1)} & x_1^{(2)} - \bar{x}^{(2)} & \dots & x_1^{(p)} - \bar{x}^{(p)} \\ x_2^{(1)} - \bar{x}^{(1)} & x_2^{(2)} - \bar{x}^{(2)} & \dots & x_2^{(p)} - \bar{x}^{(p)} \\ \dots & \dots & \dots & \dots \\ x_n^{(1)} - \bar{x}^{(1)} & x_n^{(2)} - \bar{x}^{(2)} & \dots & x_n^{(p)} - \bar{x}^{(p)} \end{pmatrix}$$

(заметим, что матрица $X_{\text{ц}}^T X_{\text{ц}}$, поделенная на n , при случайных переменных X интерпретируется как выборочная ковариационная матрица вектора X).

(ii) Уравнение регрессии Y по X в терминах центрированных переменных $Y_{\text{ц}} = Y - \bar{Y}$ и $X_{\text{ц}} = X - \bar{X}$ примет вид:

$$E(Y - \bar{Y} | X - \bar{X}) = \theta_1(x^{(1)} - \bar{x}^{(1)}) + \theta_2(x^{(2)} - \bar{x}^{(2)}) + \dots + \theta_p(x^{(p)} - \bar{x}^{(p)}), \quad (2.52)$$

так что свободный член θ_0 из исходного уравнения (2.5) может быть выражен в виде

$$\theta_0 = \bar{y} - \sum_{j=1}^p \theta_j \bar{x}^{(j)}, \quad (2.53)$$

где $\bar{y} = \sum_{i=1}^n y_i/n$, а $\bar{Y} = (\bar{y}, \bar{y}, \dots, \bar{y})^T$ — 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 $X_{\text{ц}}^T X_{\text{ц}}$ ($\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$, см. п. 13.2.2, том 1).

2) Составляется матрица коэффициентов преобразования

$$L = \begin{pmatrix} l_{11} & l_{12} & \dots & l_{1p} \\ l_{21} & l_{22} & \dots & l_{2p} \\ \dots & \dots & \dots & \dots \\ l_{p1} & l_{p2} & \dots & l_{pp} \end{pmatrix}, \quad (2.54)$$

в которой j -й собственный вектор l_j играет роль j -й строки. Отметим, что из построения следует, что матрица L ортогональна, т. е. $L^T = L^{-1}$ и соответственно $L^T L = L L^T = I_p$.

3) С помощью матрицы \mathbf{L} переходят к вектору главных компонент

$$\mathbf{Z} = (z^{(1)}, z^{(2)}, \dots, z^{(p)})^T = \mathbf{L}\mathbf{X}_ц. \quad (2.55)$$

Соответственно i -е наблюдение вектора главных компонент определится соотношением

$$Z_i = (z_i^{(1)}, z_i^{(2)}, \dots, z_i^{(p)})^T = \mathbf{L}\mathbf{X}_{цi},$$

а матрица наблюдений главных компонент — соотношением

$$\mathbf{Z} = \mathbf{X}_ц\mathbf{L}^T. \quad (2.56)$$

Отметим, что из построения следует:

$$\mathbf{Z}^T\mathbf{Z} = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_p \end{pmatrix} \quad (2.57)$$

и соответственно

$$(\mathbf{Z}^T\mathbf{Z})^{-1} = \begin{pmatrix} 1/\lambda_1 & & & 0 \\ & 1/\lambda_2 & & \\ & & \ddots & \\ 0 & & & 1/\lambda_p \end{pmatrix}. \quad (2.58)$$

Отметим также, что вектор центрированных исходных переменных $\mathbf{X}_ц$ может быть выражен из (2.55):

$$\mathbf{X}_ц = \mathbf{X} - \bar{\mathbf{X}} = \mathbf{L}^{-1}\mathbf{Z} = \mathbf{L}^T\mathbf{Z}. \quad (2.59)$$

4) Возвращаясь к задаче регрессии, анализируем КЛММР $Y_ц$ по \mathbf{Z} :

$$\mathbf{E}(Y_ц | \mathbf{Z}) = c_1 z^{(1)} + c_2 z^{(2)} + \dots + c_p z^{(p)}. \quad (2.60)$$

В соответствии с основными соотношениями метода наименьших квадратов (2.26) и (2.36) получаем:

$$\begin{aligned} \hat{\mathbf{C}}_{\text{МНК}} &= (\hat{c}_{1\text{МНК}}, \hat{c}_{2\text{МНК}}, \dots, \hat{c}_{p\text{МНК}})^T = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T Y_ц, \\ \Sigma_{\hat{\mathbf{C}}_{\text{МНК}}} &= \hat{\sigma}_{\text{ГК}}^2 (\mathbf{Z}^T\mathbf{Z})^{-1}, \end{aligned}$$

где

$$\hat{\sigma}_{\text{ГК}}^2 = \frac{1}{n-p} \sum_{i=1}^n (y_{цi} - \hat{c}_{1\text{МНК}} z_i^{(1)} - \dots - \hat{c}_{p\text{МНК}} z_i^{(p)})^2.$$

С учетом (2.58) можно сделать вывод, что оценки $\hat{c}_{1\text{мнк}}, \hat{c}_{2\text{мнк}}, \dots, \hat{c}_{p\text{мнк}}$ взаимно некоррелированы и

$$\hat{c}_{j\text{мнк}} = \frac{1}{\lambda_j} \sum_{i=1}^n z_i^{(j)} y_{\text{ил}}, \quad (2.61)$$

$$E(\hat{c}_{j\text{мнк}} - c_j)^2 = \frac{\hat{\sigma}_{\text{гк}}^2}{\lambda_j}. \quad (2.62)$$

5) Наконец, пользуясь формулами (2.61) и (2.62), последовательно проверяем гипотезы (см. п. 2.2.3, (2.47))

$$H_{0j}: c_j = 0 \quad (j = 1, 2, \dots, p)$$

с помощью статистик

$$\gamma_j = \frac{|\frac{1}{\lambda_j} \sum_{i=1}^n z_i^{(j)} y_{\text{ил}}|}{\hat{\sigma}_{\text{гк}} / \sqrt{\lambda_j}} = \frac{|\sum_{i=1}^n z_i^{(j)} y_{\text{ил}}|}{\sqrt{\lambda_j} \hat{\sigma}_{\text{гк}}}.$$

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

$$y - \bar{y} = \sum_{j \in J_{\text{сущ}}} \hat{c}_j z^{(j)}. \quad (2.63)$$

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. (2.58)) 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 (2.63) 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 (2.5) получаются с учетом соотношений (2.52), (2.53), (2.55) и (2.63):

$$\hat{\theta}_{q\text{гк}} = \sum_{j \in J_{\text{сущ}}} \hat{c}_j l_{jq}, \quad q = 1, 2, \dots, p,$$

$$\hat{\theta}_{0\text{гк}} = \bar{y} - \sum_{q=1}^p \hat{\theta}_{q\text{гк}} \bar{x}^{(q)}.$$

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 С. А. и др., 1985, п. 8.5].

III. Отбор наиболее существенных объясняющих переменных в КЛММР. В п. 13.1 (том 1) 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, упоминалось *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 случайной, возможны две точки зрения на оценку уравнения регрессии, полученную после отбора существенных предсказывающих переменных.

Первая точка зрения исходит из того, что модель регрессии (2.5) является истинной, и несмещенная оценка коэффициентов регрессии получается с помощью метода наименьших квадратов, т. е. путем решения системы уравнений (2.21) (в условиях мультиколлинеарности эта оценка может быть неудовлетворительной, но тем не менее несмещенной). Тогда принудительное приравнение 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, т. е. мы приходим к классу смещенных оце-

нок, рассмотренных в п. 2.3.2. 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 $x^{(i_1^{(k)})}, x^{(i_2^{(k)})}, \dots, x^{(i_k^{(k)})}$, для которых коэффициент детерминации с результирующим показателем y был бы максимальным, т. е.

$$\hat{R}_{y, (x^{(i_1^{(k)})}, \dots, x^{(i_k^{(k)})})}^2 = \max_{\substack{1 \leq i_1, i_2, \dots, i_k \leq p \\ i_{q_1} \neq i_{q_2} \text{ при } q_1 \neq q_2}} \hat{R}_{y, (x^{(i_1)}, \dots, x^{(i_k)})}^2. \quad (2.64)$$

Таким образом, на *первом шаге* процедуры ($k = 1$) мы найдем *одну объясняющую переменную* $x^{(i_1^{(1)})}$, 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 (2.64) 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 тес-

ную 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, т. е. *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. 2.2. 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 (11.39) в томе 1). 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

$$R_{\min}^2 = \hat{R}^{*2}(k) - 2\sqrt{\frac{2k(n-k-1)}{(n-1)(n^2-1)}}(1 - \hat{R}^2(k)). \quad (2.65)$$

В соотношении (2.65) 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. том 1, формулу (11.39)). *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. 2.2).

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 (полуэвристически)* методов оптимизации критерия (2.64) (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.

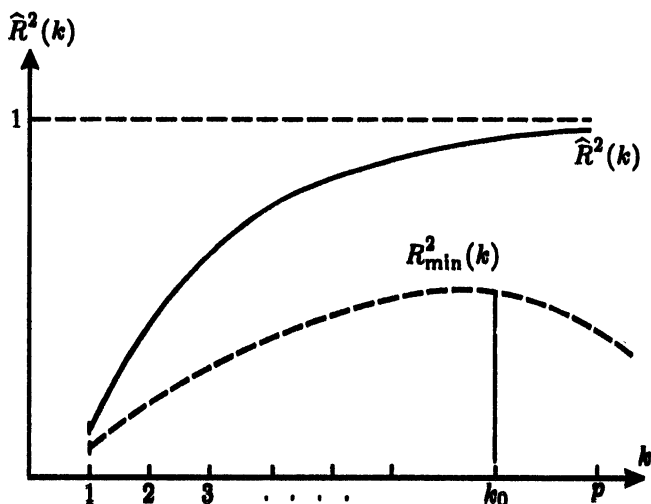


Рис. 2.2. Зависимость нижней доверительной границы коэффициента детерминации от числа предикторов (пунктирная кривая)

Версия пошагового отбора переменных. Эта версия отличается от предыдущей только тем, что на каждом следующем шаге (т. е. 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 (2.64) 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 переборov снижается с $2^p - 2$ до $p + (p - 1) + (p - 2) + \dots + 2 = (p + 2)(p - 1)/2$ (т. е. 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 (2.64)) 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 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 С. А. и др., 1985, п. 8.7].

Пример 2.1 (окончание). Напомним, что анализ матрицы парных корреляций (см. табл. 2.2) 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, т. е. 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 (2.64)) предиктор. Поскольку при $k = 1$ величина $\hat{R}_{y,x}^2$ 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 $r^{*2}(y, x^{(4)}) = \hat{R}_{y,x}^{*2}(1)$ и его нижней доверительной границы $R_{\min}^{(1)}$

соответственно по формулам (11.40) тома 1 и (2.65) (при $k = 1$) дает:

$$\widehat{R}^{*2}(1) = 0,299 \quad \text{и} \quad R_{\min}^2(1) = 0,207.$$

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 (2.64)) пара предикторов. Поскольку

$$\max_{\substack{1 \leq j \leq 5 \\ (j \neq 4)}} \widehat{R}_{y.(x^{(4)}, x^{(j)})}^2 = \widehat{R}_{y.(x^{(4)}, x^{(3)})}^2 = 0,483,$$

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 (11.34), том 1).

Подсчет подправленного (на несмещенность) значения коэффициента детерминации $\widehat{R}_{y.(x^{(4)}, x^{(3)})}^{*2} = \widehat{R}^{*2}(2)$ и его нижней доверительной границы $R_{\min}^2(2)$ соответственно по формулам (11.40) тома 1 и (2.65) дает:

$$\widehat{R}^{*2}(2) = 0,422 \quad \text{и} \quad R_{\min}^2(2) = 0,324.$$

Оценка уравнения регрессии урожайности зерновых культур y (ц/га) по $x^{(3)}$ и $x^{(4)}$, подсчитанная по формуле (2.26), имеет вид:

$$\hat{f}(x^{(3)}, x^{(4)}) = 7,29 + 3,48 x^{(3)} + 3,48 x^{(4)}.$$

(0,66) (0,13) (1,07)

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 (2.46); 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 $R_{\min}^2(2)$, т. к. $R_{\min}^2(1) < R_{\min}^2(2)$.

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).

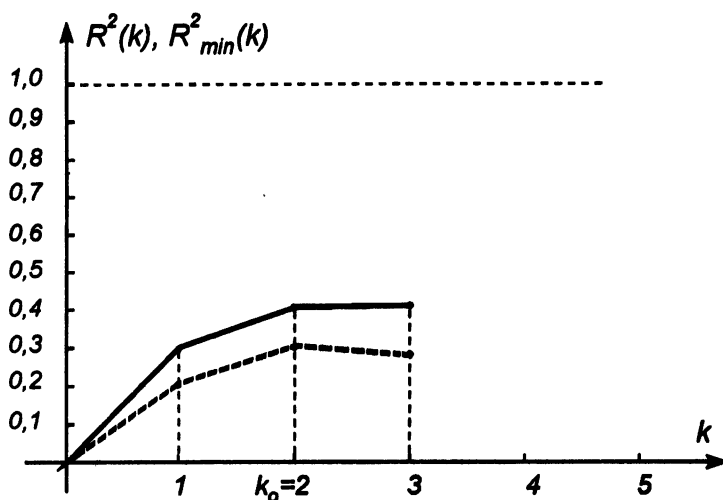


Рис. 2.3. График изменения $\hat{R}^2(k)$ (сплошная линия) и $R_{\min}^2(k)$ (пунктир) в примере 2.1

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, т. к. $R_{\min}^2(3) < R_{\min}^2(2)$ (см. также рис. 2.3). 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} .*

2.5. Ошибки спецификации модели

Выше (см. п. 1.2.3) было подробно описано содержание проблемы спецификации эконометрической модели. Напомним, что к вопросам специ-

фикации модели относится, в частности, выбор общего вида зависимости между анализируемыми переменными (см. также том 1, п. 10.7), 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 (2.5'), что истинное уравнение связи между y и X имеет вид

$$Y = X\Theta + \epsilon, \quad (2.66)$$

где, X , как и прежде, матрица наблюдаемых 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, т. е. часть объясняющих переменных истинной модели (2.66) 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 (т. е. не участвующим в модели (2.66)) 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

$$\bar{\Theta} = (\bar{X}^T \bar{X})^{-1} \bar{X}^T Y. \quad (2.67)$$

Подставим (2.66) в (2.67):

$$\bar{\Theta} = (\bar{X}^T \bar{X})^{-1} \bar{X}^T X\Theta + (\bar{X}^T \bar{X})^{-1} \bar{X}^T \epsilon,$$

так что

$$E\bar{\Theta} = B\Theta, \quad (2.68)$$

где

$$B = (\bar{X}^T \bar{X})^{-1} \bar{X}^T X. \quad (2.69)$$

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 $\beta_{\cdot j} = (\beta_{0j}, \beta_{1j}, \dots, \beta_{k-1j})^T$ ($j = 0, 1, \dots, p$), компоненты которых являются МНК-оценками коэффициентов регрессии j -й исходной (существенной) объясняющей переменной по всем k используемым исследователем предикторам \bar{X} .

Используем теперь соотношения (2.68)–(2.69) 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, т. е.

$$X = (X^{(0)} X^{(1)} \dots X^{(p)})$$

и

$$\bar{X} = (X^{(0)}, X^{(1)}, \dots, X^{(k)}).$$

Здесь, как и прежде (см. (11.1')), $X^{(j)}$ — вектор-столбец размерности n , состоящий из n наблюдений переменной j , т. е.: $X^{(0)} = (1, 1, \dots, 1)^T$ 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:

$$B = \begin{pmatrix} 1 & 0 & \dots & 0 & \beta_{0k} & \dots & \beta_{0p} \\ 0 & 1 & \dots & 0 & \beta_{1k} & \dots & \beta_{1p} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & \beta_{k-1k} & \dots & \beta_{k-1p} \end{pmatrix}. \quad (2.70)$$

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 (2.70) в (2.68) и получим выражение для величины смещения в оценке любого из оцениваемых с помощью $\bar{\Theta}$ коэффициентов:

$$E\bar{\theta}_j = \theta_j + \beta_{jk}\theta_k + \beta_{j,k+1}\theta_{k+1} + \dots + \beta_{jp}\theta_p. \quad (2.71)$$

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» (т. е., в действительности, не участвующих в истинной модели (2.66)). Тогда матрица \bar{X} отличается от X дополнительными m столбцами, а матрица B имеет порядок $(p+1+m) \times (p+1)$ и принимает вид:

$$B = \begin{pmatrix} I_{p+1} \\ 0_{m,p+1} \end{pmatrix}, \quad (2.72)$$

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 (2.72) в (2.68), убеждаемся, что оценки $\bar{\Theta}$ в этом случае будут *несмещенными*.

* . * * *

Из проведенного выше анализа можно сделать следующие выводы.

(i) Соотношения (2.70) и (2.71) 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 (2.34) остаточной дисперсии σ^2 . А это значит, что неточность выводов, основанных на оценках $\bar{\Theta}$, еще более усугубится.

(ii) 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 ослабевают точность статистических выводов, зависящая от отношения p/n). Кроме того, увеличение числа предикторов может приводить к мультиколлинеарности, с которой связаны свои «неприятности» (см. п. 2.3).

(iii) 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 (2.5) или (2.5')).

2.6. Обобщенная линейная модель множественной регрессии (ОЛММР) и обобщенный метод наименьших квадратов (ОМНК)

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 (2.5') оказываются нарушенными. Так, если в качестве исходных статистических данных (2.4^а)–(2.4^б) 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 (2.5') соотношением $\Sigma_\epsilon = \sigma^2 I_n$. 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

2.6.1. Обобщенная линейная модель 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 (2.4^а)–(2.4^б) (т. е. объем имеющейся в нашем распоряжении выборки). И пусть ковариационная матрица Σ_ϵ регрессионных остатков $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)^\top$ выражается через Σ_0 соотношением $\Sigma_\epsilon = \sigma^2 \Sigma_0$. 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 описывает-

ся системой следующих соотношений и условий:

$$\begin{cases} Y = X\Theta + \varepsilon, \\ E\varepsilon = 0_p, \\ \Sigma_\varepsilon = \sigma^2 \Sigma_0, \\ (x^{(1)}, x^{(2)}, \dots, x^{(p)}) - \text{нечслучайные переменные} \\ \text{ранг матрицы } X = p + 1 < n, \end{cases} \quad (2.73)$$

где участвующие в (2.73) векторы и матрицы определены ранее соотношениями (2.4^а)–(2.4^б), (2.7)–(2.12).

Сравнение (2.73) с (2.5') показывает, что ОЛММР отличается от КЛММР только видом ковариационной матрицы ошибок 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 (2.73) 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 (т. е. 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 10.1 (см. том 1, п. 10.1). 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 $D(\varepsilon | x)$ от значения x (см. табл. 2.3; данные составлены на основании информации, содержащейся в табл. 10.1, том 1).

Таблица 2.3

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 $y_{cp}(X) = E(y | X)$ (относительный разброс измеряется соответствующим коэффициентом вариации $V(X) = \sqrt{D(y | X)}/y_{cp}(X)$, см. п. 2.6.3).

Отметим, что при соблюдении условия взаимной некоррелированности регрессионных остатков ковариационная матрица вектора остатков ε в примере 10.1 может быть представлена в виде

$$\Sigma_{\varepsilon} = \sigma^2 \begin{pmatrix} 1/\lambda_1 & 0 \\ & \dots & \\ & & 1/\lambda_1 & & & & & & & & & & & & & & & & & & & \\ & & & 1/\lambda_2 & & & & & & & & & & & & & & & & & & \\ & & & & \dots & & & & & & & & & & & & & & & & & \\ & & & & & 1/\lambda_2 & & & & & & & & & & & & & & & & \\ & & & & & & 1/\lambda_3 & & & & & & & & & & & & & & & \\ & & & & & & & \dots & & & & & & & & & & & & & & \\ & & & & & & & & 1/\lambda_3 & & & & & & & & & & & & & \\ & & & & & & & & & 1/\lambda_4 & & & & & & & & & & & & \\ & & & & & & & & & & \dots & & & & & & & & & & & \\ & & & & & & & & & & & 1/\lambda_4 & & & & & & & & & & \\ & \\ & \\ & \\ & \\ 0 & \end{pmatrix},$$

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, т. е. рассматривая линейную модель *множественной* регрессии (ЛММР), условные дисперсии случайных остатков которой $D(\varepsilon | X)$ оказываются величинами, зависящими 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

$$\Sigma_{\epsilon} = \sigma^2 \begin{pmatrix} \lambda_1^{-1} & & & 0 \\ & \lambda_2^{-1} & & \\ & & \ddots & \\ 0 & & & \lambda_n^{-1} \end{pmatrix}, \quad (2.74)$$

где в соответствии с описанием ОЛММР (см. (2.73)) 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 $\mathbf{D}(\epsilon | X)$ от X , будут определены подходы и к статистическому оцениванию *априори неизвестных* значений $\lambda_i, i = 1, 2, \dots, n$).

2) ЛММР с автокоррелированными остатками.

Как уже упоминалось в начале п.2.6, 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:

$$r(\epsilon_i, \epsilon_j) = \rho^{|i-j|}, \quad (2.75)$$

где $r(\epsilon_i, \epsilon_j)$ — коэффициент корреляции между ϵ_i и ϵ_j , а ρ — некоторое число, по модулю меньшее единицы (очевидно, из (2.75) следует, что по своему вероятностному смыслу ρ — это коэффициент корреляции между соседними по времени остатками). Соотношение (2.75), в частности, означает:

(i) 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» (т. е., скажем, $r(\epsilon_1, \epsilon_5) = r(\epsilon_3, \epsilon_7)$);

(ii) корреляционная связь между ϵ_i и ϵ_j исчезает вовсе при $|i-j| \rightarrow \infty$, т. е. при неограниченном удалении остатков ϵ_i и ϵ_j друг от друга во

времени;

(iii) ковариации остатков $\sigma_{ij} = E(\varepsilon_i \varepsilon_j)$ имеют вид

$$\sigma_{ij} = \sigma_\varepsilon^2 \rho^{|i-j|}, \quad (2.76)$$

где $\sigma_\varepsilon^2 = D(\varepsilon | X_i)$ — условная дисперсия остатков (не зависящая от величины X_i в силу гомоскедастичности); так что ковариационная матрица остатков имеет в данном случае вид:

$$\Sigma_\varepsilon = \sigma_\varepsilon^2 \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \dots & \dots & \dots & \dots & \dots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-1} & \dots & 1 \end{pmatrix}. \quad (2.76')$$

Заметим, что соотношение (2.76) следует из формулы, связывающей $r(\varepsilon_i, \varepsilon_j)$ и $\sigma_{ij} = E(\varepsilon_i \varepsilon_j)$: действительно, по определению (см. (2.33)), $r(\varepsilon_i, \varepsilon_j) = \sigma_{ij} / \sqrt{D\varepsilon_i} \sqrt{D\varepsilon_j} = \sigma_{ij} / D\varepsilon$, откуда и следует (2.76).

2.6.2. Обобщенный метод наименьших квадратов (ОМНК)

Итак, нам предстоит провести статистический анализ зависимости, описываемой ОЛММР (2.73). Это значит, что на основании исходных статистических данных вида (2.4^а)–(2.4^б) мы должны уметь решать задачи 1–4, описанные в п. 2.1.7, и, в частности, 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 (2.26) и (2.36)?

Обычные МНК-оценки параметров ОЛММР. Можно показать, что определенные соотношениями (2.26) 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. (2.29), (2.30)).

Однако полученные ранее формулы (2.36) и (2.36') 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 (2.29) справедливо и для ОЛММР,

имеем:

$$\begin{aligned}
 \hat{\Theta}_{\text{МНК}} - \Theta &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\varepsilon}, \\
 \Sigma_{\hat{\Theta}_{\text{МНК}}} &= \mathbf{E}[(\hat{\Theta}_{\text{МНК}} - \Theta)(\hat{\Theta}_{\text{МНК}} - \Theta)^T] \\
 &= \mathbf{E}[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}] \\
 &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{E}(\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T) \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \\
 &= \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \Sigma_0 \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}.
 \end{aligned} \tag{2.77}$$

Сравнение правой части (2.77) с формулой (2.36), справедливой только в рамках КЛММР, позволяет установить факт существенного отличия *истинных* характеристик точности оценивания с помощью МНК параметров ОЛММР от тех, которые мы получили бы, воспользовавшись формулой (2.36) (ниже мы продемонстрируем существенность подобного расхождения на некоторых примерах). Мы не обсуждаем здесь свойств оценки $\hat{\sigma}^2$, поскольку в *общей* схеме ОЛММР параметр σ^2 не имеет четкой вероятностной интерпретации (в отличие от КЛММР (2.5') и ЛММР с автокоррелированными гомоскедастичными остатками, — см. (2.76'), где параметр σ^2 имеет смысл дисперсии случайных регрессионных остатков). Более детальный анализ (2.77) и ОЛММР показал, что *оценки* $\hat{\Theta}_{\text{МНК}}$ в условиях ОЛММР 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

$$\hat{\Theta}_{\text{ОМНК}} = (\mathbf{X}^T \Sigma_0^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma_0^{-1} \mathbf{Y}. \tag{2.78}$$

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 (2.73) оценки $\hat{\Theta}_{\text{ОМНК}}$, определенные соотношениями (2.78), являются оптимальными в смысле теоремы Гаусса–Маркова (см. 2.39).

Проведем доказательство этого утверждения.

Несмещенность оценок (2.78) устанавливается по той же схеме, что и несмещенность $\hat{\Theta}_{\text{МНК}}$. В частности:

$$\begin{aligned}
 \hat{\Theta}_{\text{ОМНК}} &= (\mathbf{X}^T \Sigma_0^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma_0^{-1} (\mathbf{X} \Theta + \boldsymbol{\varepsilon}) \\
 &= (\mathbf{X}^T \Sigma_0^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma_0^{-1} \mathbf{X} \Theta + (\mathbf{X}^T \Sigma_0^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma_0^{-1} \boldsymbol{\varepsilon} \\
 &= \Theta + (\mathbf{X}^T \Sigma_0^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma_0^{-1} \boldsymbol{\varepsilon}.
 \end{aligned} \tag{2.79}$$

Применение операции математического ожидания к левой и правой частям (2.79) с учетом $E\epsilon = 0_n$ дает:

$$E\hat{\Theta}_{\text{омнк}} = \Theta. \quad (2.80)$$

С целью доказательства оптимальных свойств оценок (2.78) воспользуемся знанием матрицы Σ_0 для того, чтобы преобразовать исходные данные (2.4^а)–(2.4^б) 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, т. е.

$$\Sigma_{\epsilon_{\text{пр}}} = \sigma^2 I_n.$$

Выполнить необходимое преобразование позволяет известный результат из матричной алгебры (см. Приложение 2), в соответствии с которым всякая положительно определенная симметричная ($m \times m$)-матрица A допускает представление в виде

$$A = CC^T, \quad (2.81)$$

где C — некоторая невырожденная ($m \times m$)-матрица (представление (2.81) не единственно, однако для нас это не имеет значения). Воспользуемся этим, чтобы представить в виде (2.81) матрицу Σ_0 . Итак, существует такая невырожденная ($n \times n$)-матрица C , что

$$\Sigma_0 = CC^T. \quad (2.81')$$

Заметим, что тогда

$$C^{-1}\Sigma_0(C^{-1})^T = I_n, \quad (2.82)$$

$$(C^{-1})^T C^{-1} = \Sigma_0^{-1} \quad (2.83)$$

(соотношение (2.82) получается из (2.81') домножением левой и правой частей последнего на матрицу C^{-1} слева и матрицу $(C^T)^{-1} = (C^{-1})^T$ справа, а (2.83) получается из (2.81') с учетом правила обращения произведения квадратных невырожденных матриц $(AB)^{-1} = B^{-1}A^{-1}$).

Теперь вернемся к ОЛММР (2.73) и, домножив все члены первого соотношения этой модели слева на матрицу C^{-1} , получим:

$$Y_{\text{пр}} = X_{\text{пр}}\Theta + \epsilon_{\text{пр}}, \quad (2.84)$$

где

$$Y_{\text{пр}} = C^{-1}Y, \quad X_{\text{пр}} = C^{-1}X \quad \text{и} \quad \epsilon_{\text{пр}} = C^{-1}\epsilon. \quad (2.85)$$

Нетрудно проверить, что (2.84) удовлетворяет всем требованиям классической модели (2.5'). Для этого достаточно убедиться в том, что $\Sigma_{\epsilon_{\text{пр}}} = \mathbf{E}(\epsilon_{\text{пр}} \epsilon_{\text{пр}}^T) = \sigma^2 \mathbf{I}_n$. Действительно,

$$\begin{aligned} \Sigma_{\epsilon_{\text{пр}}} &= \mathbf{E}(\epsilon_{\text{пр}} \epsilon_{\text{пр}}^T) = \mathbf{E}(C^{-1} \epsilon \epsilon^T (C^{-1})^T) = C^{-1} \mathbf{E}(\epsilon \epsilon^T) (C^{-1})^T \\ &= C^{-1} \sigma^2 \Sigma_0 (C^{-1})^T = \sigma^2 [C^{-1} \Sigma_0 (C^{-1})^T] = \sigma^2 \mathbf{I}_n \end{aligned}$$

(при переходе к правой части соотношения было использовано равенство (2.82)).

Следовательно, в соответствии с ранее полученными для КЛММР результатами наилучшей в классе всех линейных (относительно Y) несмещенных оценок параметра Θ будет оценка

$$\hat{\Theta} = (\mathbf{X}_{\text{пр}}^T \mathbf{X}_{\text{пр}})^{-1} \mathbf{X}_{\text{пр}}^T Y_{\text{пр}}, \quad (2.86)$$

ковариационная матрица которой определяется формулой

$$\Sigma_{\hat{\Theta}} = \sigma^2 (\mathbf{X}_{\text{пр}}^T \mathbf{X}_{\text{пр}})^{-1}. \quad (2.87)$$

Вернемся к исходным наблюдениям X и Y . Оценки (2.86), выраженные в исходных наблюдениях, назовем оценками обобщенного метода наименьших квадратов $\hat{\Theta}_{\text{омнк}}$:

$$\begin{aligned} \hat{\Theta}_{\text{омнк}} &= [\mathbf{X}^T (C^{-1})^T (C^{-1} \mathbf{X})]^{-1} \mathbf{X}^T (C^{-1})^T C^{-1} Y \\ &= (\mathbf{X}^T \Sigma_0^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma_0^{-1} Y \end{aligned} \quad (2.86')$$

$$\Sigma_{\hat{\Theta}_{\text{омнк}}} = \sigma^2 (\mathbf{X}^T \Sigma_0^{-1} \mathbf{X})^{-1}. \quad (2.87')$$

Как известно, МНК-оценки (2.86) являются по построению результатом минимизации по Θ критерия $Q(\Theta) = (Y_{\text{пр}} - \mathbf{X}_{\text{пр}} \Theta)^T (Y_{\text{пр}} - \mathbf{X}_{\text{пр}} \Theta)$ (см. (2.19)). Выпишем вид этого критерия в исходных наблюдениях ОЛММР:

$$\begin{aligned} Q(\Theta) &= [C^{-1}(Y - \mathbf{X}\Theta)]^T [C^{-1}(Y - \mathbf{X}\Theta)] \\ &= (Y - \mathbf{X}\Theta)^T (C^{-1})^T C^{-1} (Y - \mathbf{X}\Theta) = (Y - \mathbf{X}\Theta)^T \Sigma_0^{-1} (Y - \mathbf{X}\Theta). \end{aligned} \quad (2.88)$$

Поэтому ОМНК-оценки $\hat{\Theta}_{\text{омнк}}$ могут быть определены как «точка минимума» обобщенного критерия (2.88).

Анализ вариации результирующего показателя y и выборочный коэффициент детерминации $\hat{R}_{y,x}^2$ в ОЛММР. В п. 2.3.3 было получено разложение общей вариации результирующего показателя на две составляющие: вариацию, объясненную изменчивостью функции регрессии (т.е. изменением значений объясняющих переменных), и вариацию

случайных остатков, — справедливое в рамках *классической* ЛММР (см. (2.27^а) и (2.27^б)). Следовательно, в используемых в данном пункте обозначениях мы, *казалось бы*, можем записать это соотношение в виде

$$(Y_{\text{пр}} - \bar{Y}_{\text{пр}})^T (Y_{\text{пр}} - \bar{Y}_{\text{пр}}) = (X_{\text{пр}} \hat{\Theta}_{\text{омнк}} - \bar{Y}_{\text{пр}})^T (X_{\text{пр}} \hat{\Theta}_{\text{омнк}} - \bar{Y}_{\text{пр}}) + \hat{\varepsilon}_{\text{пр}}^T \hat{\varepsilon}_{\text{пр}}. \quad (2.89)$$

Тогда, возвращаясь с помощью (2.85) к *исходным* наблюдениям и параметрам *обобщенной* ЛММР, из (2.89) мы бы получили:

$$(Y - \bar{Y})^T \Sigma_0^{-1} (Y - \bar{Y}) = (X \hat{\Theta}_{\text{омнк}} - \bar{Y})^T \Sigma_0^{-1} (X \hat{\Theta}_{\text{омнк}} - \bar{Y}) + \hat{\varepsilon}^T \Sigma_0^{-1} \hat{\varepsilon}. \quad (2.89')$$

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:

$$\hat{R}_{y,x}^2 = 1 - \frac{\hat{\varepsilon}^T \Sigma_0^{-1} \hat{\varepsilon}}{(Y - \bar{Y})^T \Sigma_0^{-1} (Y - \bar{Y})}. \quad (2.90)$$

Однако, в действительности, мы не можем гарантировать справедливости соотношений (2.89) и, соответственно, (2.89'), а следовательно, — и ограниченности значений коэффициента $\hat{R}_{y,x}^2$ (определенного соотношением (2.90)) в пределах интервала $[0; 1]$. Причина в том, что базовое разложение (2.27^а) 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 (2.84). Поэтому определенный соотношением (2.90) коэффициент детерминации $\hat{R}_{y,x}^2$ 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. 2.3.4) и, в частности, соотношением (2.34) и свойствами (2.45)–(2.45'). 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 (2.85) дает:

- оценка

$$\hat{\sigma}^2 = \frac{1}{n-p-1} (Y - X\hat{\Theta}_{\text{ОМНК}})^T \Sigma_0^{-1} (Y - X\hat{\Theta}_{\text{ОМНК}}) \quad (2.91)$$

является несмещенной оценкой неизвестного параметра σ^2 модели (2.73);

- 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 (2.91) и (2.78), 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 (2.78), (2.87'), (2.77)) 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. (2.91)). 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 (2.73), оцениваемых по выборке $(2.4^a)-(2.4^b)$, 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 трех пунктах.

2.7. ОЛММР с гетероскедастичными остатками

В п. 2.6.1 описана линейная модель регрессии с гетероскедастичными регрессионными остатками как один из частных случаев ОЛММР (2.77), в котором ковариационная матрица остатков Σ_ε имеет вид (2.74). Это означает, что регрессионные остатки, как и в классической модели (2.5'), взаимно некоррелированы (т. е. $E(\varepsilon_i \varepsilon_j) = 0$ для $i \neq j$), но в отличие от условий (2.5') не обладают свойством гомоскедастичности (т. е., вообще

говоря, $D\epsilon_i \neq D\epsilon_j$ при $i \neq j$) или, как принято говорить, *они гетероскедастичны* (неоднородны по характеристике случайного разброса). Как было уже упомянуто в п. 2.6.1, гетероскедастичность остатков — вполне естественная ситуация для анализа *пространственных* выборок: для них чаще более оправданной является гипотеза постоянства *относительного* разброса остатков, выраженная, например, в виде

$$D(\epsilon | X_i) = \sigma^2 [E(y | X_i)]^2 \quad (2.92)$$

или

$$D(\epsilon | X_i) = \sigma^2 (a + bx_i^{(j)})^2, \quad (2.92')$$

где $x^{(j)}$ — та объясняющая переменная, пропорционально квадрату которой возрастает условная дисперсия регрессионных остатков.

Кстати, данные табл. 2.3 и построенный на основании этих данных рис. 2.4 свидетельствуют о том, что именно соотношение (2.92') оказывается подходящей формой допущения относительно поведения условной дисперсии $D(\epsilon | x_i^0)$ в примере 10.1 из тома 1.

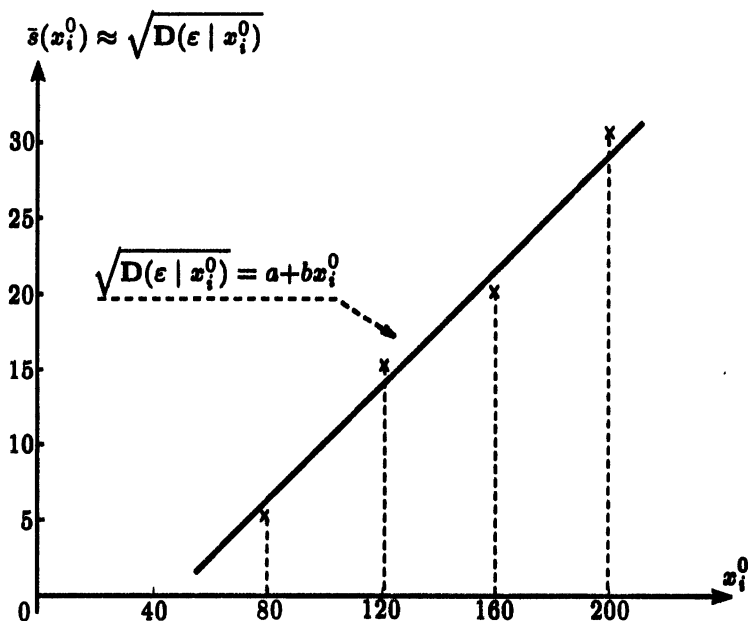


Рис. 2.4. Зависимость условного среднеквадратического отклонения остатков от значения объясняющей переменной в примере 10.1 из тома 1.

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 (2.85). Отправляясь от вытекающего из (2.74) вида матрицы Σ_0 , определим

$$C = \begin{pmatrix} 1/\sqrt{\lambda_1} & & & 0 \\ & 1/\sqrt{\lambda_2} & & \\ & & \ddots & \\ 0 & & & 1/\sqrt{\lambda_n} \end{pmatrix}. \quad (2.93)$$

Легко проверить, что в этом случае выполняются соотношения (2.81'), (2.82) и (2.83). Кроме того, определенный соотношением (2.88) критерий ОМНК при матрице Σ_ϵ вида (2.74) представим в форме

$$Q(\Theta) = (Y - X\Theta)^T \Sigma_0^{-1} (Y - X\Theta) = \sum_{i=1}^n \lambda_i (y_i - \theta_0 - \theta_1 x_i^{(1)} - \dots - \theta_p x_i^{(p)})^2. \quad (2.94)$$

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 (2.94), 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. (2.73)), определяемая формулой (2.91), в данном случае имеет вид

$$\hat{\sigma}^2 = \frac{1}{n - p - 1} \sum_{i=1}^n \lambda_i (y_i - \hat{\theta}_0 - \hat{\theta}_1 x_i^{(1)} - \dots - \hat{\theta}_p x_i^{(p)})^2. \quad (2.91')$$

При определении ОМНК-оценок $\hat{\Theta}_{\text{омнк}}$ для параметров регрессии Θ и ковариационной матрицы $\Sigma_{\hat{\Theta}_{\text{омнк}}}$ следует воспользоваться, соответственно, формулами (2.78) и (2.87') с заменой в формуле (2.87') неизвестной величины σ^2 ее оценкой (2.91') и с учетом того, что диагональный вид матрицы Σ_0 (а следовательно, и Σ_0^{-1}) существенно упрощает сам процесс вычисления.

2.7.1. 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 10.1 (том 1). Чтобы несколько упростить техническую сторону анализа, перенесем начало координат по оси абсцисс в точку $x = 40$ (ден. ед.), т. е. 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. 2.4, может быть описан упрощенным вариантом соотношения (2.92'), а именно формулой

$$D(\varepsilon | x) = \sigma^2 x^2 \quad (2.95)$$

(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 $p = 1$

$$\mathbf{X}^T \Sigma_0^{-1} \mathbf{X} = \begin{pmatrix} \sum_{i=1}^n \lambda_i & \sum_{i=1}^n \lambda_i x_i \\ \sum_{i=1}^n \lambda_i x_i & \sum_{i=1}^n \lambda_i x_i^2 \end{pmatrix},$$

то в соответствии с (2.78) получаем уравнения для определения ОМНК оценок параметров θ_0 и θ_1 :

$$\begin{cases} \theta_0 \sum_{i=1}^n \lambda_i + \theta_1 \sum_{i=1}^n \lambda_i x_i = \sum_{i=1}^n \lambda_i y_i \\ \theta_0 \sum_{i=1}^n \lambda_i x_i + \theta_1 \sum_{i=1}^n \lambda_i x_i^2 = \sum_{i=1}^n \lambda_i x_i y_i \end{cases} \quad (2.96)$$

Откуда, в частности, и следуют использованные нами ранее (том 1) формулы (10.7''') для ОМНК-оценок $\hat{\theta}_{0.OMНК}$ и $\hat{\theta}_{1.OMНК}$.

Приведенный выше конкретный вид матрицы $X^T \Sigma_0^{-1} X$ и формула (2.87') позволяют выписать дисперсию оценки $\hat{\theta}_{1.омнк}$:

$$D\hat{\theta}_{1.омнк} = E(\hat{\theta}_{1.омнк} - \theta_1)^2 = \sigma^2 \frac{\sum_{i=1}^n \lambda_i}{\left(\sum_{i=1}^n \lambda_i\right) \left(\sum_{i=1}^n \lambda_i x_i^2\right) - \left(\sum_{i=1}^n \lambda_i x_i\right)^2}. \quad (2.97)$$

Общая формула (2.22) для МНК-оценок и ее частный случай для $p = 1$ (см. (2.18)) дают нам способ вычисления *обычных* МНК-оценок $\hat{\theta}_{0.мнк}$ и $\hat{\theta}_{1.мнк}$. Воспользовавшись формулой (2.77), мы можем выписать, например, дисперсию оценки $\hat{\theta}_{1.мнк}$ и затем сравнить ее с дисперсией оценки $\hat{\theta}_{1.омнк}$. Итак, из (2.77) для случая $p = 1$ следует:

$$\begin{aligned} D\hat{\theta}_{1.мнк} &= E(\hat{\theta}_{1.мнк} - \theta_1)^2 \\ &= \sigma^2 \frac{\left(\sum_{i=1}^n \frac{1}{\lambda_i}\right) \left(\sum_{i=1}^n x_i\right)^2 - 2n \left(\sum_{i=1}^n \frac{x_i}{\lambda_i}\right) \left(\sum_{i=1}^n x_i\right) + n^2 \left(\sum_{i=1}^n \frac{x_i^2}{\lambda_i}\right)}{\left[n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2\right]^2}. \end{aligned} \quad (2.98)$$

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 (2.78), чем по формуле (2.22), поскольку ОМНК-оценки, найденные по формуле (2.78), 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 10.1 (см. пп. 10.1 и 2.6.1) при гипотезе (2.95) относительно характера гетероскедастичности регрессионных остатков. Эта гипотеза, в частности, означает, что числа λ_i , с помощью которых задается матрица Σ_0 (см. (2.74)–(2.73)), определяются соотношением:

$$\lambda_i = \frac{1}{x_i^2}, \quad (2.99)$$

что в нашем конкретном примере 10.1 означает следующее:

$$\begin{cases} \lambda_1 = \lambda_2 = \dots = \lambda_{10} = \frac{1}{(x_1^0)^2} = \frac{1}{40^2}; \\ \lambda_{11} = \lambda_{12} = \dots = \lambda_{20} = \frac{1}{(x_2^0)^2} = \frac{1}{80^2}; \\ \lambda_{21} = \lambda_{22} = \dots = \lambda_{30} = \frac{1}{(x_3^0)^2} = \frac{1}{(120)^2}; \\ \lambda_{31} = \lambda_{32} = \dots = \lambda_{40} = \frac{1}{(x_4^0)^2} = \frac{1}{(160)^2}. \end{cases} \quad (2.100)$$

Формулы (2.97) и (2.98) с учетом (2.95) принимают соответственно вид:

$$D\hat{\theta}_{1.омнк} = \sigma^2 \frac{\sum_{i=1}^n \left(\frac{1}{x_i^2}\right)}{n \sum_{i=1}^n \left(\frac{1}{x_i^2}\right) - \left[\sum_{i=1}^n \left(\frac{1}{x_i}\right)\right]^2}, \quad (2.97')$$

$$D\hat{\theta}_{1.мнк} = \sigma^2 \frac{\left(\sum_{i=1}^n x_i^2\right) \left(\sum_{i=1}^n x_i\right)^2 - 2n \left(\sum_{i=1}^n x_i^3\right) \left(\sum_{i=1}^n x_i\right) + n^2 \left(\sum_{i=1}^n x_i^4\right)}{\left[n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2\right]^2}. \quad (2.98')$$

Вычисление $D\hat{\theta}_{1.омнк}$ и $D\hat{\theta}_{1.мнк}$ с использованием формул (2.97') и (2.98') по данным примера 10.1 (том 1) и с учетом (2.100) 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 $D\hat{\theta}_{1.омнк}/D\hat{\theta}_{1.мнк} = 0,10\sigma^2/0,17\sigma^2 = 0,59$, т. е. эффективность обыкновенного метода наименьших квадратов составляет в данном конкретном случае всего лишь 59% эффективности обобщенного метода наименьших квадратов.

2.7.2. Некоторые практические рекомендации по анализу модели регрессии с гетероскедастичными остатками

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 $\Sigma_\epsilon = \sigma^2 \Sigma_0$ (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, т. е. 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 .

Проверка гипотез о гомо-/гетероскедастичности регрессионных остатков. В специальной литературе ([Джонстон Дж.], [Greene W. H.], [Goldberger A.], [Pindyck R., Rubinfeld D. L.]) описываются различные методы проверки гипотезы

$$H_0: D(\varepsilon | X) = \text{const} \quad (2.101)$$

при альтернативе, состоящей в том, что остатки $\varepsilon_1, \dots, \varepsilon_n$ гетероскедастичны. Опишем некоторые из них.

1) *Случай группированных данных и больших выборок.* Это — наиболее простой случай. Предполагается, что объем n имеющихся исходных данных (2.4^а)–(2.4^б) достаточно велик, и в частности, 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 10.1, в котором $k = 4$, а $n_1 = n_2 = n_3 = n_4 = 10$), либо принадлежат одному интервалу (гиперпараллелепипеду) группирования (см. том 1, п. 6.1). Если среднюю точку j -го интервала группирования (или общее значение объясняющей переменной в j -й подвыборке) обозначить X_j^0 , то проверка гипотезы (2.101) сведется к построению статистического критерия проверки гипотезы об однородности дисперсий

$$H_0: D(\varepsilon | X_1^0) = D(\varepsilon | X_2^0) = \dots = D(\varepsilon | X_k^0) \quad (2.101^a)$$

по величинам соответствующих выборочных дисперсий

$$s_j^2 = \frac{1}{n_j - 1} \sum_{i=1}^{n_j} (y_{ji} - \bar{y}_j)^2,$$

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 (см. том 1, п. 8.6.2, соотношение (8.32)). В случае отклонения гипотезы (2.101^а) значения s_j^2 могут быть использованы для комплектации оценки матрицы Σ_ε в качестве ее диагональных элементов.

2) *Критерий Глейсера.*¹ В данном подходе рассматривается регрессия абсолютных величин остатков $|\hat{\varepsilon}_i| = |y_i - \hat{\theta}_{0.\text{МНК}} - \hat{\theta}_{1.\text{МНК}} x_i^{(1)} - \dots - \hat{\theta}_{p.\text{МНК}} x_i^{(p)}|$

¹ Gleasier H. A New Test for Heteroscedasticity. Journ. Am. Stat. Assoc., 1969, vol. 64, pp. 316–323.

по некоторой функции от $x^{(j)}$, где $x^{(j)}$ — это та объясняющая переменная, от которой гипотетически зависит условная дисперсия остатков $D(\varepsilon | x^{(j)})$. На практике рассматриваются простые функции от $x^{(j)}$, например:

$$E(|\hat{\varepsilon}| | x^{(j)}) = a^{(1)} + a^{(2)} x^{(j)}, \quad (2.102^a)$$

$$E(|\hat{\varepsilon}| | x^{(j)}) = a^{(1)} + \frac{a^{(2)}}{x^{(j)}}, \quad (2.102^b)$$

$$E(|\hat{\varepsilon}| | x^{(j)}) = a^{(1)} (x^{(j)})^{a^{(2)}}. \quad (2.102^b)$$

Решение о гомоскедастичности регрессионных остатков ε принимается на основе проверки МНК-оценок коэффициентов $\hat{a}_{\text{МНК}}^{(1)}$ и $\hat{a}_{\text{МНК}}^{(2)}$ на их статистически значимое отличие от нуля (см. п. 2.3.4, (2.47)). Преимущество этого подхода состоит в том, что в случае отклонения гипотезы « $H_0: a^{(2)} = 0$ » исследователь одновременно получает основание к выбору формы параметризации матрицы Σ_0 . Так, случай статистически значимых коэффициентов $a^{(1)}$ и $a^{(2)}$ в схеме (2.102^a) соответствует предположению

$$\varepsilon_i = \delta_i (\hat{a}_{\text{МНК}}^{(1)} + \hat{a}_{\text{МНК}}^{(2)} x_i^{(j)}),$$

где вспомогательные случайные величины δ_i имеют нулевые средние значения и дисперсии $D\delta_i = \sigma^2$. Тогда оценка значения i -го диагонального элемента матрицы Σ_0 (т. е. величины λ_i^{-1} в обозначениях (2.74)) определится соотношением

$$\hat{\lambda}_i^{-1} = (\hat{a}_{\text{МНК}}^{(1)} + \hat{a}_{\text{МНК}}^{(2)} x_i^{(j)})^2. \quad (2.103)$$

Заметим, что регрессия $|\varepsilon|$ может строиться *не по одной* объясняющей переменной, а по нескольким или по определенной комбинации всех объясняющих переменных (например, по $x_i = x_i^{(1)2} + x_i^{(2)2} + \dots + x_i^{(p)2}$).

Состоятельное оценивание основных характеристик модели регрессии с гетероскедастичными остатками. Речь идет, конечно, о ситуациях, когда матрица Σ_0 , участвующая во всех основных формулах ОМНК, *неизвестна* (когда Σ_0 априори известна, ОМНК-оценки Θ , $\Sigma_{\hat{\Theta}_{\text{МНК}}}$ и $\Sigma_{\hat{\Theta}_{\text{ОМНК}}}$, определяемые соответственно формулами (2.78), (2.77) и (2.87') с заменой участвующего в них параметра σ^2 его оценкой (2.91), являются состоятельными).

1) *Оценка ковариационной матрицы $\Sigma_{\hat{\Theta}_{\text{МНК}}}$ обычных МНК-оценок $\hat{\Theta}_{\text{МНК}}$.* В предположении, что мы находимся в условиях модели с некор-

релированными и гетероскедастичными остатками, можно показать¹, что

$$\widehat{\Sigma}_{\widehat{\Theta}_{\text{МНК}}} = (\mathbf{X}^T \mathbf{X})^{-1} \left(\sum_{i=1}^n \widehat{\varepsilon}_i^2 \widetilde{X}_i \widetilde{X}_i^T \right) (\mathbf{X}^T \mathbf{X})^{-1}, \quad (2.104)$$

где $\widehat{\varepsilon}_i = y_i - \widehat{\Theta}_{\text{МНК}}^T \widetilde{X}_i$, а $\widetilde{X}_i = (1, x_i^{(1)}, \dots, x_i^{(p)})^T$, является состоятельной оценкой матрицы ковариаций обычных МНК-оценок коэффициентов регрессии $\theta_0, \theta_1, \dots, \theta_p$.

2) Оценка неизвестной ковариационной матрицы Σ_ε вектора остатков ε и ковариационной матрицы $\Sigma_{\widehat{\Theta}_{\text{МНК}}}$ вектора ОМНК-оценок коэффициентов регрессии Θ . Наиболее распространенный общий подход к оцениванию неизвестной ковариационной матрицы Σ_ε в условиях обобщенной линейной модели регрессии связан с экономной параметризацией этой матрицы с последующей оценкой небольшого числа параметров, в терминах которых она выражается (подробнее об этом подходе см. ниже, в п. 2.9). В условиях рассматриваемой модели с взаимно некоррелированными, но гетероскедастичными остатками этот подход может быть реализован, например, в рамках описанной выше общей схемы критерия Глейсера, когда определенная форма параметризации матрицы Σ_ε вытекает из избранного (гипотетичного) общего вида регрессии $|\widehat{\varepsilon}|$ по X (см. соотношения (2.102^а)–(2.102^в)). Примером может служить регрессия (2.102^а) и вытекающее из нее соотношение (2.103), при котором матрица Σ_0 определяется всего двумя параметрами $a^{(1)}$ и $a^{(2)}$:

$$\Sigma_0 = \begin{pmatrix} (a^{(1)} + a^{(2)} x_1^{(j)})^2 & & & 0 \\ & (a^{(1)} + a^{(2)} x_2^{(j)})^2 & & \\ & & \dots & \\ 0 & & & (a^{(1)} + a^{(2)} x_n^{(j)})^2 \end{pmatrix}. \quad (2.105)$$

Очевидно, возможны и другие формы регрессионной зависимости $|\widehat{\varepsilon}|$ от X : конкретный выбор этой формы будет определяться спецификой исходных данных (X, Y) и нашим умением адекватно подбирать общий вид регрессионной зависимости (см. методы, описанные в п. 10.7 тома 1). Вставляя полученную таким образом оценку $\widehat{\Sigma}_0$ матрицы Σ_0 в формулы (2.91) и (2.87'), получаем оценку $\widehat{\Sigma}_{\widehat{\Theta}_{\text{МНК}}}$ ковариационной матрицы $\Sigma_{\widehat{\Theta}_{\text{МНК}}}$.

¹ White H. A Heteroscedasticity-Consistent Matrix Estimator and a Direct Test for Heteroscedasticity. *Econometrica*, vol. 48 (1980), № 4, pp. 817–838.

2.8. ОЛММП с автокоррелированными остатками

Мы уже познакомились в общих чертах в п. 2.6.1 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», т. е. 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 порядка¹. Подобного рода зависимость между остатками описывается соотношениями

$$\varepsilon_i = \rho\varepsilon_{i-1} + \delta_i, \quad (2.106)$$

где ρ — некоторое число, по абсолютной величине меньшее единицы (т. е. $|\rho| < 1$), а случайные величины δ_i удовлетворяют требованиям, предъявляемым к регрессионным остаткам классической модели, т. е.

$$E\delta_i \equiv 0, \quad E(\delta_i\delta_j) = \begin{cases} \sigma_0^2 & \text{при } i = j; \\ 0 & \text{при } i \neq j. \end{cases} \quad (2.107)$$

При этом полагается, что соотношения (2.106) справедливы для любого «момента времени» i , сколь угодно удаленного в прошлое или будущее, т. е. i может «пробежать» все целочисленные значения от $-\infty$ до $+\infty$.

Отправляясь от (2.106) и (2.107), определим основные числовые характеристики (средние значения $E\varepsilon_i$ и ковариационную матрицу $\Sigma_\varepsilon = E(\varepsilon\varepsilon^T)$) вектора регрессионных остатков ε .

¹ Регрессионные 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.

Из (2.106) следует:

$$\begin{aligned}
 \varepsilon_i &= \rho\varepsilon_{i-1} + \delta_i = \rho(\rho\varepsilon_{i-2} + \delta_{i-1}) + \delta_i = \rho^2\varepsilon_{i-2} + \rho\delta_{i-1} + \delta_i \\
 &= \rho^2(\rho\varepsilon_{i-3} + \delta_{i-2}) + \rho\delta_{i-1} + \delta_i = \rho^3\varepsilon_{i-3} + \rho^2\delta_{i-2} + \rho\delta_{i-1} + \delta_i = \dots \\
 &= \delta_i + \rho\delta_{i-1} + \rho^2\delta_{i-2} + \dots = \sum_{k=0}^{\infty} \rho^k \delta_{i-k}.
 \end{aligned} \tag{2.108}$$

Из (2.108) с учетом (2.107) непосредственно следует:

$$\begin{aligned}
 \mathbf{E}\varepsilon_i &\equiv 0, \\
 \sigma^2 &= \mathbf{D}\varepsilon_i = \mathbf{E}\varepsilon_i^2 = \mathbf{E}\delta_i^2 + \rho^2\mathbf{E}\delta_{i-1}^2 + \rho^4\mathbf{E}\delta_{i-2}^2 + \dots \\
 &= \sigma_0^2(1 + \rho^2 + \rho^4 + \dots) = \frac{\sigma_0^2}{1 - \rho^2}.
 \end{aligned} \tag{2.109}$$

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 (2.108) в виде:

$$\begin{aligned}
 \varepsilon_i\varepsilon_{i-k} &= [(\delta_i + \rho\delta_{i-1} + \dots + \rho^{k-1}\delta_{i-(k-1)}) + \rho^k(\delta_{i-k} + \rho\delta_{i-k-1} + \dots)] \\
 &\quad \times (\delta_{i-k} + \rho\delta_{i-k-1} + \dots).
 \end{aligned}$$

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:

$$\text{cov}(\varepsilon_i, \varepsilon_{i-k}) = \mathbf{E}(\varepsilon_i\varepsilon_{i-k}) = \mathbf{E}[\rho^k(\delta_{i-k} + \rho\delta_{i-k-1} + \dots)^2] = \rho^k \mathbf{E}\varepsilon_{i-k}^2 = \sigma^2 \rho^k, \tag{2.110}$$

где дисперсия σ^2 определена соотношением (2.109).

Таким образом, мы пришли к тому, что автокорреляционная зависимость 1-го порядка (2.106), связывающая между собой регрессионные остатки ОЛМНР, в терминах ковариаций этих остатков эквивалентна соотношениям (2.109)–(2.110). Это позволяет записать линейную модель

множественной регрессии с автокоррелированными остатками в виде:

$$\left\{ \begin{array}{l} Y = X\Theta + \varepsilon, \\ E\varepsilon = 0_n, \\ \Sigma_\varepsilon = E(\varepsilon\varepsilon^\top) = \frac{\sigma_0^2}{1-\rho^2} \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-2} & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-3} & \rho^{n-2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & \rho & 1 \end{pmatrix} = \sigma^2 \Sigma_0, \\ (x^{(1)}, x^{(2)}, \dots, x^{(p)}) - \text{нелучайные переменные;} \\ \text{ранг матрицы } X = p + 1 < n. \end{array} \right.$$

Таким образом, $(n \times n)$ -матрица Σ_0 определяется в данной модели значением *единственного* параметра ρ по формуле

$$\Sigma_0 = \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-2} & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-3} & \rho^{n-2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & \rho & 1 \end{pmatrix}. \quad (2.112)$$

Понадобится нам и матрица Σ_0^{-1} , участвующая во всех основных формулах ОМНК. Поэтому приведем здесь ее вид:

$$\Sigma_0^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho & 0 & 0 & \dots & 0 & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & 0 & \dots & 0 & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & -\rho & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & -\rho & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & 0 & \dots & 0 & -\rho & 1 \end{pmatrix}. \quad (2.113)$$

Становится понятным теперь и вероятностный смысл формально введенного в (2.106) параметра ρ . Действительно, из определения коэффициента корреляции (2.33) и соотношений (2.109)–(2.110) следует, что коэффициент корреляции $r(\varepsilon_i, \varepsilon_{i\pm k})$ между регрессионными остатками, отстоящими друг от друга по времени на k единиц, равен:

$$r(\varepsilon_i, \varepsilon_{i\pm k}) = \rho^k, \quad k = 0, 1, 2, \dots \quad (2.114)$$

Именно в такой математической форме реализуется в модели (2.111) идея ослабления корреляционных связей между регрессионными остатками по мере их взаимного удаления во времени.

Как мы знаем (см. п. 2.6.2), 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

$$C^{-1} = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \dots & 0 & 0 & 0 \\ -\rho & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -\rho & 1 & \dots & 0 & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -\rho & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & -\rho & 1 \end{pmatrix}. \quad (2.115)$$

Соответствующие преобразованные исходные данные имеют при такой матрице преобразования C^{-1} следующий вид:

$$Y_{\text{пр}} = C^{-1}Y = (\sqrt{1-\rho^2}y_1, y_2 - \rho y_1, y_3 - \rho y_2, \dots, y_n - \rho y_{n-1})^T,$$

$$X_{\text{пр}} = C^{-1}X = \begin{pmatrix} \sqrt{1-\rho^2} & \sqrt{1-\rho^2}x_1^{(1)} & \dots & \sqrt{1-\rho^2}x_1^{(p)} \\ 1-\rho & x_2^{(1)} - \rho x_1^{(1)} & \dots & x_2^{(p)} - \rho x_1^{(p)} \\ \dots & \dots & \dots & \dots \\ 1-\rho & x_n^{(1)} - \rho x_{n-1}^{(1)} & \dots & x_n^{(p)} - \rho x_{n-1}^{(p)} \end{pmatrix}.$$

Путем непосредственного перемножения соответствующих матриц нетрудно убедиться в справедливости равенств (2.81'), (2.82) и (2.83) при матрице C^{-1} вида (2.113), а соответственно и в справедливости соотношения

$$E(C^{-1}\varepsilon\varepsilon^T C^{-1T}) = \sigma^2 I_n.$$

2.8.1. 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 (2.78), и соотношений (2.77) и (2.87'), позволяющих оценивать точность соответственно МНК- и ОМНК-оценок, пользуется соответствующими формулами *обычного* метода наименьших квадратов — (2.26) для $\hat{\Theta}_{\text{мнк}}$ и (2.36) для ковариационной матрицы оценок $\hat{\Theta}_{\text{мнк}}$.

В этом пункте мы хотим проанализировать меру искажения характеристик точности обычных МНК-оценок при использовании формулы (2.36) (позволяющей оценивать ковариационную матрицу МНК-оценок $\hat{\Theta}_{\text{МНК}}$ только в условиях *классической* ЛММР) вместо действующей в условиях *обобщенной* ЛММР формулы (2.77).

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 неустрашимые искажения при этом возникают в оценивании их характеристик точности, т. е. ковариационной матрицы $\Sigma_{\hat{\Theta}_{\text{МНК}}}$.

С целью анализа этих искажений рассмотрим модель (2.111) зависимости y от *единственной* объясняющей переменной x без свободного члена. Очевидно, в этом случае $X = x, \Theta = \theta$, т. е. $y_i = \theta x_i + \varepsilon_i$, 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 (2.26), в данном случае имеет вид

$$\hat{\theta}_{\text{МНК}} = (X^T X)^{-1} X^T Y = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}. \quad (2.116)$$

Подсчитаем дисперсию (средний квадрат ошибки) этой оценки параллельно двумя способами: 1) по формуле (2.36), справедливой только в рамках условий *классической* модели, т. е. игнорирующей автокоррелированность регрессионных остатков, и 2) по формуле (2.77), учитывающей эту автокоррелированность в рамках условий *обобщенной* модели (2.111).

Использование формулы (2.36) дает:

$$(D\hat{\theta}_{\text{МНК}})_{\text{прибл}} = \sigma^2 (X^T X)^{-1} = \frac{\sigma^2}{\sum_{i=1}^n x_i^2}. \quad (2.117)$$

Проведем расчет величины $D\hat{\theta}_{\text{мнк}}$ по *правильной* формуле (2.77):

$$\begin{aligned}
 D\hat{\theta}_{\text{мнк}} &= \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \Sigma_0 \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \\
 &= \sigma^2 \frac{1}{\sum_{i=1}^n x_i^2} (x_1, x_2 \dots x_n) \\
 &\times \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-2} & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-3} & \rho^{n-2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & \rho & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \frac{1}{\sum_{i=1}^n x_i^2} = \\
 &= \frac{\sigma^2}{\sum_{i=1}^n x_i^2} \left(1 + 2\rho \frac{\sum_{i=1}^{n-1} x_i x_{i+1}}{\sum_{i=1}^n x_i^2} + 2\rho^2 \frac{\sum_{i=1}^{n-2} x_i x_{i+2}}{\sum_{i=1}^n x_i^2} + \dots + 2\rho^{n-1} \frac{x_1 x_n}{\sum_{i=1}^n x_i^2} \right). \quad (2.118)
 \end{aligned}$$

Возьмем для сравнения результатов (2.118) и (2.117) их отношение:

$$\frac{D\hat{\theta}_{\text{мнк}}}{(D\hat{\theta}_{\text{мнк}})_{\text{прибл}}} = 1 + 2\rho \frac{\sum_{i=1}^{n-1} x_i x_{i+1}}{\sum_{i=1}^n x_i^2} + \dots + 2\rho^{n-1} \frac{x_1 x_n}{\sum_{i=1}^n x_i^2}. \quad (2.119)$$

Очевидно, правая часть (2.119) 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), т. е.

$$r(k) = r(x_i, x_{i+k}) = \frac{\sum_{i=1}^{n-k} x_i x_{i+k}}{\sum_{i=1}^n x_i^2}, \quad k = 1, 2, \dots, n-1.$$

Тогда, придав определенное числовое значение ρ и положив, например, $r(k) = \rho^k$, можно численно оценить правую часть (2.119). 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 (2.117) при подсчете среднего квадрата ошибки МНК-оценки $\hat{\theta}_{\text{МНК}}$ приводит к непропорциональному 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 (2.111) методами, игнорирующими автокорреляцию регрессионных остатков. Речь идет об оценивании параметра σ^2 (беря отношения сравниваемых характеристик (2.117) и (2.118), мы сумели элиминировать этот параметр из нашего рассмотрения). Проведенный аналогично только что выполненному анализ результатов правильного (с помощью формулы (2.91)) и некорректного (с помощью формулы (2.34)) 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 Дж., с. 246–248]).

2.8.2. Проверка гипотезы о наличии/отсутствии автокоррелированности регрессионных остатков

Как это следует из предыдущего пункта, 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

$$d = \frac{\sum_{i=2}^n (\hat{\varepsilon}_i - \hat{\varepsilon}_{i-1})^2}{\sum_{i=1}^n \hat{\varepsilon}_i^2}, \quad (2.120)$$

где $\hat{\varepsilon}_i = y_i - \hat{\theta}_{0.\text{мнк}} - \hat{\theta}_{1.\text{мнк}}x_i^{(1)} - \dots - \hat{\theta}_{p.\text{мнк}}x_i^{(p)}$ — невязки *обыкновенного* метода наименьших квадратов. В упомянутой выше работе исследовано распределение статистики d в предположении справедливости гипотезы

$$H_0: \rho = 0. \quad (2.121)$$

Поскольку, как оказалось, распределение статистики d (в предположении справедливости гипотезы H_0) *зависит от наблюдаемых значений объясняющих переменных* X , Дёрбину и Уотсону удалось установить (для двух заданных величин уровня значимости критерия $\alpha = 0,05$ и $\alpha = 0,01$) лишь такие пороговые значения $d_L(\alpha)$ и $d_U(\alpha)$ ($d_L(\alpha) < d_U(\alpha)$), которые позволяют построить следующие два варианта процедуры проверки гипотезы (2.121) (в зависимости от альтернативы о наличии в остатках *положительной* или *отрицательной* автокорреляции 1-го порядка):

а) *При $d < 2$ (альтернатива: существование в остатках положительной автокорреляции первого порядка):*

- по заданному α находим из таблиц (см. Приложение 1) пороговые значения $d_L(\alpha)$ и $d_U(\alpha)$;

- по формуле (2.120) подсчитываем значение критической статистики d ;

- если $d < d_L(\alpha)$, то гипотеза H_0 отвергается (с вероятностью ошибиться, равной α) в пользу гипотезы о положительной автокорреляции;

- если $d > d_U(\alpha)$, то гипотеза H_0 не отвергается;

- если $d_L(\alpha) \leq d \leq d_U(\alpha)$, то сделать определенный вывод по имеющимся исходным данным нельзя.

б) *При $d > 2$ (альтернатива: существование в остатках отрицательной автокорреляции первого порядка):*

- первые два действия — те же, что и п. а);

- если $4 - d < d_L(\alpha)$, то гипотеза H_0 отвергается (с вероятностью ошибиться, равной α) в пользу гипотезы об отрицательной автокорреляции;

- если $4 - d > d_U(\alpha)$, то гипотеза H_0 не отвергается;

- если $d_L(\alpha) \leq 4 - d \leq d_U(\alpha)$, то сделать определенный вывод по имеющимся исходным данным нельзя.

Грубо говоря, в случае отсутствия автокорреляции в регрессионных остатках значение критической статистики d должно «не слишком отличаться» от 2.

Отметим, что данный критерий, строго говоря, выведен лишь для *неслучайных* объясняющих переменных. Поэтому, если в составе объясняющих переменных присутствуют, например, так называемые лаговые значения результирующего показателя, то он должен быть соответствующим образом модифицирован (см. [Джонстон Дж., п. 10.3]).

2.8.3. Некоторые практические рекомендации по анализу модели регрессии с автокоррелированными остатками

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 (2.78), (2.77), (2.87) и (2.91).

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 R. S., Rubinfeld D. L., п. 6.2.1]). Приведем здесь описание одной из наиболее распространенных процедур подобного типа, известной в литературе под названием *процедуры Козрейна-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 (2.26);
- 2) подсчитываются невязки 1-й итерации $\hat{\epsilon}_i^{(1)} = y_i - \hat{\theta}_{0,\text{МНК}} - \hat{\theta}_{1,\text{МНК}} x_i^{(1)} - \dots - \hat{\theta}_{p,\text{МНК}} x_i^{(p)}$;
- 3) первое приближение $\hat{\rho}^{(1)}$ оценки $\hat{\rho}$ неизвестного параметра ρ определяется в качестве МНК-оценки коэффициента регрессии ρ в модели $\hat{\epsilon}_i^{(1)} = \rho \hat{\epsilon}_{i-1}^{(1)} + \delta_i^{(1)}$, где остатки $\delta_i^{(1)}$ удовлетворяют условиям классической модели (2.5);
- 4) вычисляются ОМНК-оценки $\hat{\Theta}_{\text{ОМНК}}(\hat{\rho}^{(1)})$ по формуле (2.78) с матрицей $\Sigma_0(\hat{\rho}^{(1)})$, определенной соотношением (2.112), в котором вместо ρ подставлены значения $\hat{\rho}^{(1)}$;
- 5) подсчитываются невязки 2-й итерации

$$\hat{\epsilon}_i^{(2)} = y_i - \hat{\theta}_{0,\text{ОМНК}}(\hat{\rho}^{(1)}) - \hat{\theta}_{1,\text{ОМНК}}(\hat{\rho}^{(1)}) x_i^{(1)} - \dots - \hat{\theta}_{p,\text{ОМНК}}(\hat{\rho}^{(1)}) \cdot x_i^{(p)};$$

6) второе приближение $\hat{\rho}^{(2)}$ оценки $\hat{\rho}$ неизвестного параметра ρ определяется в качестве МНК-оценки коэффициента регрессии ρ в модели $\hat{\epsilon}_i^{(2)} = \rho \hat{\epsilon}_{i-1}^{(2)} + \delta_i^{(2)}$, где остатки $\delta_i^{(2)}$ удовлетворяют условиям классической модели (2.5);

7) вычисляются ОМНК-оценки $\hat{\Theta}_{\text{омнк}}(\hat{\rho}^{(2)})$ по формуле (2.78) с матрицей $\Sigma_0(\hat{\rho}^{(2)})$, определенной соотношением (2.112), в котором вместо ρ подставлены значения $\hat{\rho}^{(2)}$ — и т. д.

Процедуру заканчивают при стабилизации получаемых значений $\hat{\rho}^{(k)}$, т. е. 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.

2.9. Практические рекомендации по построению, анализу и интерпретации регрессионной модели

2.9.1. Построение и анализ обобщенной ЛММР при неизвестной ковариационной матрице регрессионных остатков (практически реализуемый ОМНК)

В замечании, завершающем п. 2.6, 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: (2.78), (2.77), (2.87'), (2.90) и (2.91). 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 структура матрицы Σ_{ϵ}* (и соответственно Σ_0), т. е. форма ее функциональной зависимости от сравнительно небольшого коли-

чества 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 10.1 (том 1) удалось построить подобную схему с *единственным* неизвестным параметром σ^2 (т.е. при $m = 0$), так как в нем была использована структура матрицы Σ_ϵ , определяемая соотношением (2.74), где $\lambda_i = 1/x_i^2$.

В примере ОЛММР с автокоррелированными остатками (см. (2.76')) 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 Глейсера случай (2.103) соответствует двухпараметрической структуре матрицы Σ_0 (она определяется параметрами $a^{(1)}$ и $a^{(2)}$); следовательно, матрица Σ_ϵ будет определяться значениями трех параметров: $\sigma^2, a^{(1)}$ и $a^{(2)}$. 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 (2.91)) и $\hat{\Sigma}_\epsilon = \hat{\sigma}^2 \hat{\Sigma}_0$. При некоторых достаточно общих условиях¹ удается показать, что использование полученных таким образом оценок $\hat{\Sigma}_0$ и $\hat{\Sigma}_\epsilon$ в основных формулах ОМНК (2.77), (2.78) и (2.87') вместо неизвестных истинных значений Σ_0 и Σ_ϵ дает также состоятельные оценки соответственно для $\Sigma_{\hat{\theta}_{\text{ОМНК}}}, \hat{\theta}$ и $\Sigma_{\hat{\theta}_{\text{ОМНК}}}$. Подобный способ оценивания принято называть в специальной литературе *практически реализуемым обобщенным методом наименьших квадратов*.

Опишем общую схему процедуры практически реализуемого ОМНК, которая используется в рамках *нормальной* ОЛММР, т.е. в рамках условий модели (2.73), дополненных условием $(0_n; \sigma^2 \Sigma_0(a))$ — нормальности вектора регрессионных остатков ϵ .

В этом случае функция правдоподобия остатков $\epsilon = Y - X\theta$ имеет вид (сравнить с (2.24)):

$$L(\epsilon | \theta; \sigma^2; \Sigma_0(a)) = \frac{1}{(2\pi)^{\frac{n}{2}} (\sigma^2)^{\frac{n}{2}} |\Sigma_0(a)|^{\frac{1}{2}}} \times \exp \left[-\frac{1}{2\sigma^2} (Y - X\theta)^\top \Sigma_0^{-1}(a) (Y - X\theta) \right].$$

¹ В качестве таких условий используются, например, условия сходимости по вероятности (при $n \rightarrow \infty$): а) матрицы $\frac{1}{n} X^\top \Sigma_0^{-1} X$ — к некоторой конечной невырожденной матрице и б) вектора-столбца $\frac{1}{n} X^\top \Sigma_0^{-1} \epsilon$ — к вектору-столбцу 0_{p+1} , состоящему из одних нулей.

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 $\hat{a} = (\hat{a}_1, \dots, \hat{a}_m)$.

$$l(\varepsilon | \Theta; \sigma^2; \Sigma_0(a)) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) + \frac{1}{2} \ln(\Sigma_0^{-1}(a)) - \frac{1}{2\sigma^2} (Y - X\Theta)^\top \Sigma_0^{-1}(a) (Y - X\Theta);$$

$$\left\{ \begin{array}{l} \frac{\partial l}{\partial \Theta} = -\frac{1}{2\sigma^2} \frac{\partial}{\partial \Theta} [(Y - X\Theta)^\top \Sigma_0^{-1}(a) (Y - X\Theta)] = 0_{p+1}, \\ \frac{\partial l}{\partial \sigma^2} = -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2(\sigma^2)^2} (Y - X\Theta)^\top \Sigma_0^{-1}(a) (Y - X\Theta) = 0, \\ \frac{\partial l}{\partial a^{(j)}} = -\frac{(Y - X\Theta)^\top \omega_j (Y - X\Theta)}{(Y - X\Theta)^\top \Sigma_0(a) (Y - X\Theta)} + \frac{1}{n} \text{tr}(\omega_j \Sigma_0(a)) = 0, \end{array} \right. \quad (2.122)$$

где $\omega_j = \frac{\partial \Sigma_0^{-1}(a)}{\partial a^{(j)}}$, $j = 1, 2, \dots, m$.

Анализ системы уравнений (2.122) позволяет сделать следующие выводы:

1) Решение первого уравнения системы доставляет минимум критерию ОМНК (см. (2.88)), а следовательно, *даже не производя дифференцирования по Θ* , можно заключить, что оценки параметров Θ в схеме ОЛМ-МР по обобщенному методу наименьших квадратов совпадают с оценками по методу максимального правдоподобия (т.е. $\hat{\Theta}_{\text{ОМНК}} = \hat{\Theta}_{\text{ММП}}$, где оценка $\hat{\Theta}_{\text{ОМНК}}$ определена формулой (2.78)).

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, т.е.

$$\hat{\sigma}_{\text{ММП}}^2 = \frac{1}{n} (Y - X\Theta)^\top \Sigma_0^{-1}(a) (Y - X\Theta) = \frac{n-p-1}{n} \hat{\sigma}^2,$$

где оценка $\hat{\sigma}^2$ определена формулой (2.91).

3) Сложность решения системы уравнений относительно оценок вспомогательных параметров $a = (a^{(1)}, \dots, a^{(m)})$ (эта система описана третьим уравнением в системе (2.122), 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 $\varepsilon = Y - X\Theta$, то они заменяются соответствующими «невязками» $\hat{\varepsilon} = Y - X\hat{\Theta}$ в режиме обычной (для ОМНК) итерационной процедуры: сначала в качестве Θ в выражение $\varepsilon = Y - X\Theta$ 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 $\hat{\Sigma}_0$ для получения первого приближения для ОМНК-оценок $\hat{\Theta}_{\text{омнк}}$, по ним вычисляют «невязки» $\hat{\varepsilon} = Y - X\hat{\Theta}_{\text{омнк}}$ 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. 2.8.3) относится к этому же типу методов оценивания.

2.9.2. Точечный и интервальный прогноз, основанный на моделях линейной регрессии

Задачи, о которых идет речь в данном пункте, являются главным стимулятором развития и применения разнообразных методов регрессионного анализа. Потому с них мы и начали изложение материала данной главы (см. задачи 1 и 2 в п. 2.1.7). Действительно, изучая зависимость между результирующей (y) и объясняющими (X) переменными, *исследователь, в первую очередь, хочет уметь предсказывать значение y (или функции регрессии y по X) по заданному значению объясняющей переменной X .*

Построение прогноза для значения результирующего показателя $y(X)$ или его условного среднего значения $f(X) = E(y | X)$ по заданным величинам объясняющих переменных $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$ основано на решении задач анализа точности полученного уравнения регрессии $\hat{f}(X)$ (см. «об основных задачах анализа точности полученной регрессионной зависимости», описание этапа 7 в п. 10.6 первого тома).

Начнем с задачи *точечного* (т. е. оцениваемого *одним числом*) прогноза значения результирующего показателя y по заданной величине объясняющих переменных.

Точечный прогноз значения результирующего показателя в условиях ОЛММР. Пусть мы располагаем исходными статистическими данными (X, Y) (см. (2.4^а)–(2.4^б)) и нам известно, что статистическая связь между y и $\tilde{X} = (1, x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$ может быть

описана регрессионной моделью вида (2.73), т. е. удовлетворяет условиям обобщенной линейной модели множественной регрессии (ОЛММР). Соответственно известна (или оценена методами, описанными в предыдущем пункте) ковариационная матрица регрессионных остатков Σ_ε . Помимо упомянутых исходных данных (X, Y) нам заданы также значения объясняющих переменных $\tilde{X}_{n+1} = (1, x_{n+1}^{(1)}, \dots, x_{n+1}^{(p)})^T$, т. е. еще одна $(n+1)$ -я строка матрицы X , однако соответствующее значение результирующего показателя $y_{n+1} = y(\tilde{X}_{n+1})$ нам не известно. Известен, правда, характер ковариационных связей регрессионного остатка ε_{n+1} из уравнения $y_{n+1} = \tilde{X}_{n+1}^T \Theta + \varepsilon_{n+1}$ со всеми остальными регрессионными остатками $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$, описываемый вектор-столбцом

$$\sigma_\varepsilon^{(n+1)} = (\mathbf{E}(\varepsilon_1 \varepsilon_{n+1}), \mathbf{E}(\varepsilon_2 \varepsilon_{n+1}), \dots, \mathbf{E}(\varepsilon_n \varepsilon_{n+1}))^T,$$

а также то, что $\mathbf{E}\varepsilon_{n+1} = 0$ и $\mathbf{E}\varepsilon_{n+1}^2 = \Delta^2$. 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:

$$\mathbf{E}(y_{n+1} - C_0^T Y)^2 = \min_C \mathbf{E}(y_{n+1} - C^T Y)^2, \quad (2.123)$$

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} , т. е. они должны удовлетворять условию:

$$\mathbf{E}(C^T Y) = \mathbf{E}y_{n+1}. \quad (2.124)$$

Обозначим

$$\hat{y}_C(\tilde{X}_{n+1}) = C^T Y = C^T (X\Theta + \varepsilon) \quad (2.125)$$

линейный прогноз для $y_{n+1} = y(\tilde{X}_{n+1})$, удовлетворяющий условию несмещенности (2.124).

Чтобы получить решение поставленной задачи, выпишем и проанализируем средний квадрат ошибки прогноза (т. е. левую часть соотношения (2.123))

$$\begin{aligned} \hat{y}_C(X_{n+1}) - y_{n+1} &= C^T (X\Theta + \varepsilon) - \tilde{X}_{n+1}^T \Theta - \varepsilon_{n+1} \\ &= (C^T X - \tilde{X}_{n+1}^T) \Theta + (C^T \varepsilon - \varepsilon_{n+1}) \\ &= C^T \varepsilon - \varepsilon_{n+1}, \end{aligned} \quad (2.126)$$

так как из условия несмещенности (2.124) следует:

$$\mathbf{E}(C^T Y - y_{n+1}) = (C^T \mathbf{X} - \tilde{X}_{n+1}^T) \Theta + \mathbf{E}(C^T \boldsymbol{\varepsilon} - \varepsilon_{n+1}) = (C^T \mathbf{X} - \tilde{X}_{n+1}^T) \Theta = 0,$$

т. е. вектор C должен удовлетворять условию

$$C^T \mathbf{X} - \tilde{X}_{n+1}^T = \mathbf{0}. \quad (2.124')$$

Поэтому, возвращаясь к (2.126), имеем:

$$\begin{aligned} \mathbf{E}(\hat{y}_C(X_{n+1}) - y_{n+1})^2 &= \mathbf{E}(C^T \boldsymbol{\varepsilon} - \varepsilon_{n+1})^2 \\ &= \mathbf{E}[(C^T \boldsymbol{\varepsilon} - \varepsilon_{n+1})(C^T \boldsymbol{\varepsilon} - \varepsilon_{n+1})^T] \\ &= \mathbf{E}[(C^T \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^T C) + \varepsilon_{n+1}^2 - 2C^T \boldsymbol{\varepsilon} \varepsilon_{n+1}] \\ &= C^T \boldsymbol{\Sigma}_\varepsilon C + \mathbf{E}(\varepsilon_{n+1}^2) - 2C^T \sigma_\varepsilon^{(n+1)}. \end{aligned} \quad (2.127)$$

Далее минимизируем (2.127) при условии (2.124') по C . С этой целью выписываем соответствующую функцию Лагранжа $\varphi(\Lambda)$, дифференцируем ее по C и по Λ и приравниваем эти производные к нулю:

$$\varphi(\Lambda) = C^T \boldsymbol{\Sigma}_\varepsilon C + \Delta^2 - 2C^T \sigma_\varepsilon^{(n+1)} - (C^T \mathbf{X} - \tilde{X}_{n+1}^T) \Lambda, \quad (2.128)$$

где $\Lambda = (\lambda^{(0)}, \lambda^{(1)}, \dots, \lambda^{(p)})^T$.

Система уравнений

$$\begin{cases} \frac{\partial \varphi(\Lambda)}{\partial \Lambda} = \mathbf{0}_{p+1}, \\ \frac{\partial \varphi(\Lambda)}{\partial C} = \mathbf{0}_n \end{cases}$$

имеет вид

$$\begin{pmatrix} \boldsymbol{\Sigma}_\varepsilon & \mathbf{X} \\ \mathbf{X}^T & \mathbf{0}_{(p+1) \times (p+1)} \end{pmatrix} \begin{pmatrix} C \\ -\Lambda \end{pmatrix} = \begin{pmatrix} \sigma_\varepsilon^{(n+1)} \\ \tilde{X}_{n+1} \end{pmatrix} \quad (2.129)$$

($\mathbf{0}_{(p+1) \times (p+1)}$ — матрица размерности $(p+1) \times (p+1)$, состоящая из нулей).

Решение системы (2.129) относительно компонент вектора C и дает нам по построению коэффициенты C_0 для оптимального прогноза:

$$C_0 = \boldsymbol{\Sigma}_\varepsilon^{-1} [\mathbf{I}_n - \mathbf{X}(\mathbf{X}^T \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma}_\varepsilon^{-1}] \sigma_\varepsilon^{(n+1)} + \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{X}(\mathbf{X}^T \boldsymbol{\Sigma}_\varepsilon^{-1} \mathbf{X})^{-1} \tilde{X}_{n+1}. \quad (2.130)$$

Таким образом, наилучшим линейным несмещенным прогнозом значения $y_{n+1} = y(\tilde{X}_{n+1})$ будет

$$\begin{aligned} \hat{y}_{C_0}(\tilde{X}_{n+1}) &= C_0^T Y = \tilde{X}_{n+1}^T (X^T \Sigma_\varepsilon^{-1} X)^{-1} X^T \Sigma_\varepsilon^{-1} Y \\ &\quad + \sigma_\varepsilon^{(n+1)T} \Sigma_\varepsilon^{-1} Y - \sigma_\varepsilon^{(n+1)T} \Sigma_\varepsilon^{-1} X (X^T \Sigma_\varepsilon^{-1} X)^{-1} X^T \Sigma_\varepsilon^{-1} Y \\ &= \tilde{X}_{n+1}^T \hat{\Theta}_{\text{ОМНК}} + (\sigma_\varepsilon^{(n+1)})^T \Sigma_\varepsilon^{-1} (Y - X \hat{\Theta}_{\text{ОМНК}}) \\ &= \tilde{X}_{n+1}^T \hat{\Theta}_{\text{ОМНК}} + (\sigma_\varepsilon^{(n+1)})^T \Sigma_\varepsilon^{-1} \hat{\varepsilon}. \end{aligned} \quad (2.131)$$

В преобразованиях, приведших к соотношению (2.131), мы воспользовались правилом транспонирования правой части выражения (2.130), формулой (2.78) для $\hat{\Theta}_{\text{ОМНК}}$ и выражением вектора «невязок» $\hat{\varepsilon}$ в виде $Y - X \hat{\Theta}_{\text{ОМНК}}$.

Анализируя формулу (2.131), мы видим, что наилучший прогноз значения $y(X_{n+1})$ определяется оцененным с помощью ОМНК значением функции регрессии в точке X_{n+1} (т.е. величиной $\tilde{X}_{n+1}^T \hat{\Theta}_{\text{ОМНК}}$) 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 (2.131) тождественно равно нулю, так как $\sigma_\varepsilon^{(n+1)} = (0, 0, \dots, 0)^T$.

Посмотрим, какой результат дает общая формула (2.131) 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 $\tilde{X}_{n+1} = (1, x_{n+1}^{(1)}, \dots, x_{n+1}^{(p)})^T$ интерпретируются как прогнозные (или планируемые) значения предикторов в будущем такте времени « $n + 1$ ».

По условиям модели (2.76') регрессионные остатки ε подчиняются авторегрессионной схеме первого порядка, следовательно

$$\sigma_\varepsilon^{(n+1)} = \sigma^2 \begin{pmatrix} \rho^n \\ \rho^{n-1} \\ \vdots \\ \rho \end{pmatrix} = \rho \sigma^2 \begin{pmatrix} \rho^{n-1} \\ \rho^{n-2} \\ \vdots \\ 1 \end{pmatrix},$$

так что вектор $\sigma_\varepsilon^{(n+1)}$ может быть получен умножением на ρ последнего столбца матрицы Σ_ε , определенной формулой (2.76'). Но поскольку

$\Sigma_{\epsilon} \Sigma_{\epsilon}^{\tau 1} = \mathbf{I}_n$, то произведение $(\sigma_{\epsilon}^{(n+1)})^{\tau} \Sigma_{\epsilon}^{-1}$ дает умноженную на ρ последнюю строку матрицы \mathbf{I}_n и, следовательно:

$$(\sigma_{\epsilon}^{(n+1)})^{\tau} \Sigma_{\epsilon}^{-1} \hat{\epsilon} = \rho \hat{\epsilon}_n. \quad (2.132)$$

Возвращаясь к формуле (2.131) и принимая во внимание (2.132), имеем:

$$\hat{y}(\tilde{X}_{n+1}) = \tilde{X}_{n+1}^{\tau} \hat{\Theta}_{\text{ОМНК}} + \rho \hat{\epsilon}_n. \quad (2.133)$$

Строго говоря, использование прогноза (2.133) 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 величины ρ и $\hat{\Theta}_{\text{ОМНК}}$ их значениями, полученными на последней итерации *практически реализуемого ОМНК* (см. п. 2.8.3 и 2.9.1).

Интервальный прогноз значения результирующего показателя в условиях нормальной ОЛММР. Условия предыдущей задачи построения точечного прогноза должны быть усилены требованием $(\mathbf{0}_n; \Sigma_{\epsilon})$ — нормальности регрессионных остатков ϵ . *Требуется при заданной величине доверительной вероятности P построить основанный на наилучшем линейном несмещенном точечном прогнозе $\hat{y}(\tilde{X}_{n+1})$ интервальный прогноз для значения результирующего показателя $y(\tilde{X}_{n+1})$, т. е. указать такой интервал значений $[\hat{y}_{\min}^{(P)}(\tilde{X}_{n+1}), \hat{y}_{\max}^{(P)}(\tilde{X}_{n+1})]$, который с вероятностью, не меньшей P , накроет неизвестное значение $y(\tilde{X}_{n+1})$.*

По существу речь идет об оценке точности точечного прогноза, а точность определяется средним квадратом (или дисперсией) ошибки. Средний квадрат ошибки оптимального прогноза $\sigma_{\text{прогн}}^2(\tilde{X}_{n+1}) = E(\hat{y}_{C_0}(X_{n+1}) - y(X_{n+1}))^2$ получаем, подставив в правую часть соотношения (2.127) оптимальные значения коэффициентов C_0 , определенные формулой (2.130) (из-за громоздкости получающегося при этом выражения не будем приводить здесь его явный вид).

Поскольку точечный прогноз $\hat{y}_{C_0}(\tilde{X}_{n+1})$ 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 $\hat{y}_{C_0}(\tilde{X}_{n+1}) - y(\tilde{X}_{n+1})$ является нормально распределенной случайной величиной со средним значением, равным нулю (в силу (2.124), т. е. из-за свойства несмещенности прогноза), и с дисперсией $\sigma_{\text{прогн}}^2(\tilde{X}_{n+1})$, определяемой соотношениями (2.127) и (2.130). Следовательно, определив по таблицам квантилей стандартного нормального закона (см. Приложение 1)

квантиль $u_{\frac{1+P}{2}}$ уровня $(1+P)/2$, мы можем вычислить границы искомого интервального прогноза по формулам:

$$\begin{aligned} \hat{y}_{\min}^{(P)}(\tilde{X}_{n+1}) &= \hat{y}_{C_0}(\tilde{X}_{n+1}) - u_{\frac{1+P}{2}} \sigma_{\text{прогн.}}(\tilde{X}_{n+1}), \\ \hat{y}_{\max}^{(P)}(\tilde{X}_{n+1}) &= \hat{y}_{C_0}(\tilde{X}_{n+1}) + u_{\frac{1+P}{2}} \sigma_{\text{прогн.}}(\tilde{X}_{n+1}), \end{aligned} \quad (2.134)$$

где $\hat{y}_{C_0}(\tilde{X}_{n+1})$ вычисляется по формуле (2.131).

Другими словами, можно гарантировать (с заданной вероятностью P), что величина исследуемого 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 формулам (2.134), а участвующая в них величина $\hat{y}_{C_0}(\tilde{X}_{n+1})$ — по формуле (2.131).

Случай классической линейной модели регрессии. Полученный результат весьма нагляден и прост в применении к классической ЛММР. Действительно, в соответствии с (2.5') в данном случае имеем:

$$\begin{aligned} \Sigma_{\epsilon} &= \sigma^2 \mathbf{I}_n, \\ \Delta^2 &= \mathbf{D}\epsilon_{n+1} = \sigma^2, \\ \sigma_{\epsilon}^{(n+1)} &= \mathbf{0}_n. \end{aligned}$$

Соответственно, с учетом формул (2.130), (2.131) и (2.127), получаем:

$$\begin{aligned} C_0 &= \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \tilde{X}_{n+1}, \\ \hat{y}_{C_0}(\tilde{X}_{n+1}) &= \tilde{X}_{n+1}^T \hat{\Theta}_{\text{мнк}}, \\ \sigma_{\text{прогн.}}^2(\tilde{X}_{n+1}) &= \sigma^2 [\tilde{X}_{n+1}^T (\mathbf{X}^T \mathbf{X})^{-1} \tilde{X}_{n+1} + 1]. \end{aligned} \quad (2.135)$$

Применим полученный результат к построению интервального прогноза в классической модели *парной* регрессии (т.е. при $p=1$). Ранее (в п. 2.3.1) мы выписывали конкретный вид матриц \mathbf{X} , $\mathbf{X}^T \mathbf{X}$ и $\mathbf{X}^T \mathbf{Y}$ в данной схеме. В частности:

$$\mathbf{X}^T \mathbf{X} = \begin{pmatrix} n & n\bar{x} \\ n\bar{x} & \sum_{i=1}^n x_i^2 \end{pmatrix}, \quad \text{где } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Соответственно:

$$(\mathbf{X}^T \mathbf{X})^{-1} = \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n x_i^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix},$$

$$\begin{aligned} \tilde{X}_{n+1}^T (\mathbf{X}^T \mathbf{X})^{-1} \tilde{X}_{n+1} &= (1 \quad x_{n+1}) \begin{pmatrix} \frac{1}{n} \sum_{i=1}^n x_i^2 & -\bar{x} \\ -\bar{x} & 1 \end{pmatrix} \begin{pmatrix} 1 \\ x_{n+1} \end{pmatrix} \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ &= \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}. \end{aligned} \quad (2.136)$$

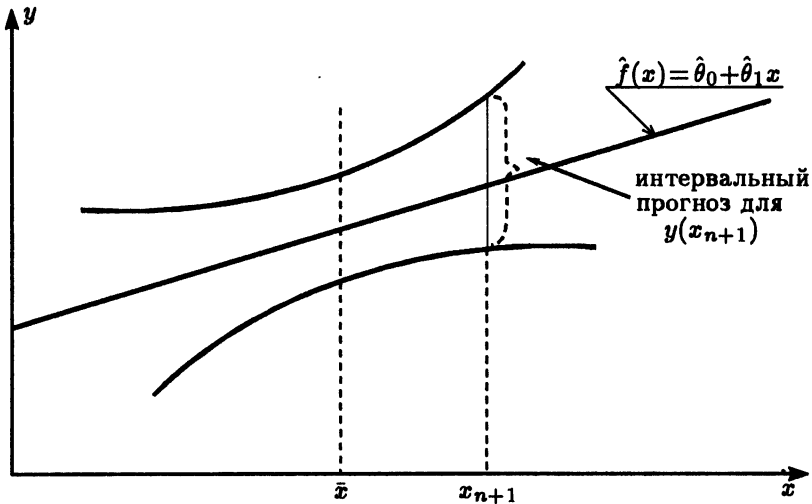


Рис. 2.5. Интервальный прогноз для значения зависимой переменной $y(x_{n+1})$ при заданной величине объясняющей переменной x_{n+1}

Подставляя полученное выражение в (2.135) и используя формулу (2.134), с учетом свойств классической модели имеем: прогнозируемое знание $y(X_{n+1})$ заключено (с вероятностью P) между величинами

$$\hat{y}_{\min}^{(P)}(\tilde{X}_{n+1}) = \hat{\theta}_{0.\text{мнк}} + \hat{\theta}_{1.\text{мнк}} x_{n+1} \mp u_{1+\frac{P}{2}} \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}},$$

где $\hat{\theta}_{0.\text{мнк}}$, $\hat{\theta}_{1.\text{мнк}}$ — оценки коэффициентов парной регрессии обычным методом наименьших квадратов (см. (2.18)), x_{n+1} — заданное значение объясняющей переменной, а $\sigma^2 = D\varepsilon$ — дисперсия регрессионных остатков в классической модели линейной регрессии.

Рис. 2.5 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} .

* * *

З а м е ч а н и е. Условия ОМНК, в рамках которых описаны процедуры построения точечных и интервальных прогнозов, предполагают априорное знание ковариационной матрицы остатков Σ_ϵ . В наших построениях считалась известной даже величина σ^2 (дисперсия остатков в КЛММР и числовой параметр в соотношении $\Sigma_\epsilon = \sigma^2 \Sigma_0$ — в ОЛММР). Однако описанные процедуры могут применяться и при неизвестных Σ_0 и σ^2 , если заменить в них значения соответствующими оценками, полученными на последней итерации *практически реализуемого ОМНК* (см. пп. 2.8.3 и 2.9.1). При этом, правда, придется воспользоваться свойством статистической независимости оценок $\hat{\sigma}^2$ и $\hat{\Theta}_{\text{омнк}}$ (см. (2.91) и (2.91')) и заменить квантили нормального распределения в формулах (2.134) на соответствующие процентные точки t -распределения с $n - p - 1$ степенями свободы.

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 $\tilde{X}_{n+1} = (1, x_{n+1}^{(1)}, \dots, x_{n+1}^{(p)})^\top$, а *условного среднего значения* $f(\tilde{X}_{n+1}) = E(y | \tilde{X}_{n+1})$. Так, в примере 10.1 из тома 1 (см. пп. 10.1, 2.6.1 и 2.7.1) 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} .

Ответ на вопрос о наилучшей *точечной* оценке значения линейной функции регрессии при заданном значении аргумента дается теоремой Гаусса–Маркова для классической модели (см. п. 2.3.4) и ее расширением на случай обобщенной модели (см. теорему Эйткена в п. 2.6.2): наилучшая точечная оценка (наилучший точечный прогноз) определяется значением эмпирической функции регрессии $\hat{f}(\tilde{X}_{n+1}) = \tilde{X}_{n+1}^\top \hat{\Theta}$ при заданном значении \tilde{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 $f(\tilde{X}) = \tilde{X}^T \Theta$ при заданном значении $\tilde{X} = \tilde{X}_{n+1}$ в условиях обобщенной ЛММР. Для этого вычислим сначала средний квадрат ошибки наилучшего точечного прогноза:

$$\begin{aligned} \sigma_f^2(\tilde{X}_{n+1}) &= \mathbf{E}(\tilde{X}_{n+1}^T \hat{\Theta}_{\text{ОМНК}} - \tilde{X}_{n+1}^T \Theta)^2 \\ &= \mathbf{E}[\tilde{X}_{n+1}^T (\hat{\Theta}_{\text{ОМНК}} - \Theta) (\hat{\Theta}_{\text{ОМНК}} - \Theta)^T \tilde{X}_{n+1}] \\ &= \tilde{X}_{n+1}^T \mathbf{E}[(\hat{\Theta}_{\text{ОМНК}} - \Theta) (\hat{\Theta}_{\text{ОМНК}} - \Theta)^T] \tilde{X}_{n+1} \\ &= \tilde{X}_{n+1}^T \Sigma_{\hat{\Theta}_{\text{ОМНК}}} \tilde{X}_{n+1} = \sigma^2 \tilde{X}_{n+1}^T (\mathbf{X}^T \Sigma_0^{-1} \mathbf{X})^{-1} \tilde{X}_{n+1}. \end{aligned} \quad (2.137)$$

При выводе соотношения (2.137) мы воспользовались возможностью представить $[\tilde{X}_{n+1}^T (\hat{\Theta}_{\text{ОМНК}} - \Theta)]^2$ в виде $[\tilde{X}_{n+1}^T (\hat{\Theta}_{\text{ОМНК}} - \Theta)][\tilde{X}_{n+1}^T (\hat{\Theta}_{\text{ОМНК}} - \Theta)]^T$, а также видом ковариационной матрицы оценок $\hat{\Theta}_{\text{ОМНК}}$ (см. формулу (2.87')).

Итак, наилучшая линейная несмещенная точечная оценка $\tilde{X}_{n+1}^T \hat{\Theta}_{\text{ОМНК}} = \tilde{X}_{n+1}^T (\mathbf{X}^T \Sigma_0^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma_0^{-1} \mathbf{Y}$ неизвестного значения функции регрессии $\tilde{X}_{n+1}^T \Theta$ — это нормально распределенная случайная величина (поскольку она является линейно комбинацией нормально распределенных случайных величин \mathbf{Y}) со средним значением, равным $\tilde{X}_{n+1}^T \Theta$, и с дисперсией $\sigma_f^2(\tilde{X}_{n+1})$, определяемой по формуле (2.137). Поэтому, определив по заданной величине P доверительной вероятности из таблиц (см. Приложение 1) значение $(1 + P)/2$ — квантиля $u_{(1+P)/2}$ стандартного нормального распределения, мы имеем основания утверждать, что неизвестное значение функции регрессии $f(\tilde{X}_{n+1}) = \tilde{X}_{n+1}^T \Theta$ с вероятностью P должно принадлежать интервалу

$$[\tilde{X}_{n+1}^T \hat{\Theta}_{\text{ОМНК}} - u_{(1+P)/2} \sigma_f(\tilde{X}_{n+1}), \tilde{X}_{n+1}^T \hat{\Theta}_{\text{ОМНК}} + u_{(1+P)/2} \sigma_f(\tilde{X}_{n+1})], \quad (2.138)$$

где величина среднеквадратической ошибки $\sigma_f(\tilde{X}_{n+1})$ оценки функции регрессии в точке \tilde{X}_{n+1} определяется по формуле (2.137).

Рекомендации для классической ЛММР получаются из (2.137)–(2.138) в качестве частного случая $\Sigma_0 = \mathbf{I}_n$. В частности, для классической модели парной регрессии (т. е. при $p = 1$) нами было подсчитано (см. (2.136)) выражение $\tilde{X}_{n+1}^T (\mathbf{X}^T \mathbf{X})^{-1} \tilde{X}_{n+1}$, необходимое для вычисления $\sigma_f^2(\tilde{X}_{n+1})$. Используя (2.136), с помощью формулы (2.137) получаем выражение для

$\sigma_f^2(X_{n+1})$ в данном простом случае:

$$\hat{\sigma}_f^2(X_{n+1}) = \sigma^2 \left[\frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right].$$

Подставляя это выражение в (2.138), 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. *Практическая реализация* рекомендаций по построению интервального прогноза для неизвестного значения функции регрессии в заданной точке \tilde{X}_{n+1} требует определенной модификации процедуры. В частности, априори известные (в рамках условий ОЛММР) значения σ^2 и Σ_0 заменяются их статистическими оценками $\hat{\sigma}^2$ и $\hat{\Sigma}_0$, полученными на последней итерации практически реализуемого варианта ОМНК (см. пп. 2.8.3 и 2.9.1). Из свойства статистической независимости полученных при этом оценок $\hat{\sigma}^2$ и $\hat{\Theta}_{\text{омнк}}$ (см. (2.91) и (2.91')) следует, что величина

$$\frac{\tilde{X}_{n+1}^\top (\hat{\Theta}_{\text{омнк}} - \Theta)}{\hat{\sigma} \sqrt{\tilde{X}_{n+1}^\top (\mathbf{X}^\top \hat{\Sigma}_0^{-1} \mathbf{X})^{-1} \tilde{X}_{n+1}}}$$

подчиняется распределению Стьюдента с $n - p - 1$ степенями свободы (или $t(n - p - 1)$ -распределению). Этот факт позволяет использовать в качестве интервального прогноза для функции регрессии в точке \tilde{X}_{n+1} (с доверительной вероятностью P) интервал, определяемый соотношением (2.138), с заменой в выражении (2.137) для $\sigma_f^2(\tilde{X}_{n+1})$ значений σ^2 и Σ_0 их оценками $\hat{\sigma}^2$ и $\hat{\Sigma}_0$, а $(1 + P)/2$ — квантиля стандартного нормального распределения $u_{(1+P)/2}$ $100(1 - P)/2$ — процентной точкой $t_{(1-P)/2}(n - p - 1)$ распределения Стьюдента с $n - p - 1$ степенями свободы.

2.9.3. Исследование точности регрессионной модели в реалистической ситуации

Неточный выбор общего вида функции регрессии, приводящий к нарушению базового допущения

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

на которое существенно опираются все выводы по оцениванию точности регрессионной модели, и в частности, по их использованию в задачах прогноза, может заключаться как в неполном или избыточном представлении

набора объясняющих переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$, так и в искажении *самой структуры модели*, т. е. в отклонении ее общего вида от линейного. Влияние неполного или избыточного представления набора объясняющих переменных на свойства оценок и соответственно на точность статистических выводов в регрессионном анализе (при правильном определении *структуры*, т. е. линейного характера зависимости между $E(y(X))$ и X) может быть учтено в рамках строгих математических конструкций (эти вопросы были рассмотрены в п. 2.5). Наиболее неприятные последствия влечет второй тип ошибок в спецификации 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 $f(X) = E(y | X)$. Это мы и называем «*реалистической ситуацией*» в регрессионном анализе. В подобных ситуациях необходимо руководствоваться следующими положениями при построении прогнозов и исследовании точности статистических выводов в регрессионном анализе:

1) при анализе точности аппроксимационных вариантов регрессионных моделей не следует претендовать на построение сколько-нибудь 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 $\eta(X)$ (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 $\hat{\varepsilon}_i = y_i - f_a(X_i; \hat{\Theta}^{(n_1)})$). Пусть мы произвели такое разбиение исходной выборки $\tilde{B}_n = (X, Y)$ на две подвыборки — обучающую $\tilde{B}_{n_1} = (X_j(1), Y_j(1))$ и экзаменующую $\tilde{B}_{n_2} =$

$(X_j(2), Y_j(2))$ k различными способами (т. е. $j = 1, 2, \dots, k$);

3) основной (и по существу единственной) характеристикой точности аппроксимационного варианта линейной регрессионной модели является оценка $\hat{\sigma}$ среднеквадратической ошибки аппроксимации σ , вычисляемая по формуле

$$\hat{\sigma}^2 = \frac{1}{\sum_{j=1}^k n_{2j} - p - 1} \sum_{j=1}^k (Y_j(2) - X_j(2)\hat{\Theta}_{\text{МНК}}^{(n_{1j})})^T (Y_j(2) - X_j(2)\hat{\Theta}_{\text{МНК}}^{(n_{1j})}) \quad (2.139)$$

при регрессионных остатках ϵ , специфицированных в соответствии с условиями классической модели (т. е. при $\Sigma_\epsilon = \sigma^2 I_n$), или по формуле

$$\hat{\sigma}^2 = \frac{1}{\sum_{j=1}^k n_{2j} - p - 1} \sum_{j=1}^k (Y_j(2) - X_j(2)\hat{\Theta}_{\text{ОМНК}}^{(n_{1j})})^T \Sigma_0^{-1} (Y_j(2) - X_j(2)\hat{\Theta}_{\text{ОМНК}}^{(n_{1j})}) \quad (2.139^a)$$

при регрессионных остатках ϵ , специфицированных в соответствии с условиями обобщенной модели (т. е. при $\Sigma_\epsilon = \sigma^2 \Sigma_0$). В соотношениях (2.139) и (2.139^a) оценки $\hat{\Theta}_{\text{МНК}}^{(n_{1j})}$ и $\hat{\Theta}_{\text{ОМНК}}^{(n_{1j})}$ вычислены в соответствии с формулами соответственно (2.28) и (2.78) 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 $\pm 2\hat{\sigma}$;

4) следует проявлять известную сдержанность и осторожность при использовании аппроксимационных вариантов регрессионных моделей для решения задач интерполяции и (особенно) экстраполяции, т. е. 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*.

Поясним подробнее реализацию положений 2) и 3) на примере исполь-

зования 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}^{06}) и экзаменуемую ($B_{2j}^{3к3}$) следующим образом:

$$\begin{cases} \tilde{B}_{n_{1j}}^{06} = \{(X_1, y_1), \dots, (X_{j-1}, y_{j-1}), (X_{j+1}, y_{j+1}), \dots, (X_n, y_n)\}; \\ B_{n_{2j}}^{3к3} = \{X_j, y_j\}, \quad j = 1, 2, \dots, n. \end{cases}$$

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 экзаменуемой выборки содержит единственное наблюдение — (X_j, y_j) . Применение к такой последовательности обучающих и экзаменуемых выборок формул (2.139) и (2.139^а) 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 обычных формул (2.34) и (2.91).

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 $f_a(X; \Theta) = \tilde{X}^T \Theta$ 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) :

$$\hat{\Theta}^{(n_{1i})} = \hat{\Theta} + \frac{y_i - \tilde{X}_i^T \hat{\Theta}}{1 - q_i} \cdot (X^T X)^{-1} \cdot \tilde{X}_i,$$

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 М. Дж., Стьюарт А., 1973]).

где $q_i = \tilde{X}_i^T \cdot (X^T X)^{-1} \cdot \tilde{X}_i^1$;

$$\tilde{X}_i^T \hat{\Theta}^{(n_{1i})} = \tilde{X}_i^T \hat{\Theta} + q_i \frac{y_i - \tilde{X}_i^T \hat{\Theta}}{1 - q_i};$$

$$\hat{\varepsilon}_i(\hat{\Theta}^{(n_{1i})}) = \frac{\hat{\varepsilon}_i(\hat{\Theta})}{1 - q_i}, \quad \text{где } \hat{\varepsilon}_i(\hat{\Theta}) = y_i - f_a(X_i; \hat{\Theta});$$

$$\sum_{i=1}^n \hat{\varepsilon}_i^2(\hat{\Theta}^{(n_{1i})}) = \sum_{i=1}^n \frac{1}{(1 - q_i)^2} \hat{\varepsilon}_i^2(\hat{\Theta}).$$

Эти соотношения позволяют избежать многократной вычислительной «прогонки» процедур метода наименьших квадратов на различных вариантах обучающей выборки за счет пересчета значений $\hat{\Theta}^{(n_{1i})}$, $f(X_i; \hat{\Theta}^{(n_{1i})}) = \tilde{X}_i^T \hat{\Theta}^{(n_{1i})}$ и т. д. по соответствующим характеристикам, подсчитанным по наблюдениям всей выборки \tilde{B}_n .

2.9.4. Анализ эластичностей с использованием моделей регрессии

Коэффициенты эластичности являются важным инструментом экономического анализа. Они определяются на базе анализа функциональных зависимостей, существующих между некоторой интересующей нас экономической характеристикой z (под z может пониматься, например, спрос на определенный вид товаров или услуг или объем произведенной продукции), с одной стороны, и набором объясняющих переменных $u^{(1)}, u^{(2)}, \dots, u^{(p)}$ (цены на рассматриваемый и другие товары, основные факторы производства) — с другой. Пусть эта зависимость описывается функцией

$$z = f(u^{(1)}, u^{(2)}, \dots, u^{(p)}).$$

Коэффициент эластичности $\Theta_{z/u^{(j)}}$ результирующего показателя z по объясняющей переменной $u^{(j)}$ формально определяется как логарифмическая производная z по $u^{(j)}$, т. е.

$$\Theta_{z/u^{(j)}} = \frac{\partial(\ln z)}{\partial(\ln u^{(j)})} = \frac{\partial(\ln f(u^{(1)}, \dots, u^{(p)}))}{\partial(\ln u^{(j)})}.$$

¹ При самых естественных ограничениях на структуру матрицы 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 $q_i \sim (p + 1)/n$.

Из этого определения нетрудно вывести содержательный смысл коэффициента эластичности. Действительно, анализируя процесс вычисления $\Theta_{z/u^{(j)}}$

$$\Theta_{z/u^{(j)}} = \frac{\partial(\ln z)}{\partial(\ln u^{(j)})} = \frac{(dz/z)}{(du^{(j)}/u^{(j)})} = \lim_{\Delta u^{(j)} \rightarrow 0} \frac{(\Delta z/z) \cdot 100\%}{(\Delta u^{(j)}/u^{(j)}) \cdot 100\%},$$

мы видим, что эластичность $\Theta_{z/u^{(j)}}$ показателя z по переменной $u^{(j)}$ приблизительно определяет, на сколько процентов изменится значение z при изменении объясняющей переменной на 1%.

Теперь предположим, что зависимость $f(u^{(1)}, \dots, u^{(p)})$ получена на основе функции регрессии y по $X = (x^{(1)}, \dots, x^{(p)})^T$, построенной по исходным данным (X_i, y_i) , $i = 1, 2, \dots, n$, где переменные y и X связаны определенными функциональными соотношениями соответственно с переменными z и $U = (u^{(1)}, \dots, u^{(p)})^T$. Можно ли по коэффициентам построенной функции регрессии что-либо сказать о значениях эластичностей z по $u^{(j)}$ ($j = 1, \dots, p$)?

Рассмотрим две простейшие ситуации:

1) Переменные X , y связаны с экономическими переменными U и z логарифмическим преобразованием, т. е.

$$y = \ln z \quad \text{и} \quad x^{(j)} = \ln u^{(j)}, \quad j = 1, 2, \dots, p.$$

Тогда оцененная по выборке линейная модель регрессии y по X в терминах U и z выразится в виде

$$\ln z = \hat{\theta}_0 + \hat{\theta}_1 \ln u^{(1)} + \dots + \hat{\theta}_p \ln u^{(p)}.$$

Откуда в соответствии с определением эластичности имеем

$$\Theta_{z/u^{(j)}} = \frac{\partial(\ln z)}{\partial(\ln u^{(j)})} = \hat{\theta}_j.$$

2) Переменные X и y , на основании наблюдений над которыми строилась линейная функция регрессии, как раз и есть те экономические характеристики, которые нас интересуют, т. е.

$$z = E(y | X) \quad \text{и} \quad u^{(j)} = x^{(j)}.$$

Тогда оценка функции регрессии y по X дает нам зависимость $f(u^{(1)}, \dots, u^{(p)})$ в виде

$$z = \hat{\theta}_0 + \hat{\theta}_1 u^{(1)} + \dots + \hat{\theta}_p u^{(p)}.$$

Вычисление эластичности z по $u^{(p)}$ на основе этой зависимости дает:

$$\Theta_{z/u^{(j)}} = \frac{\partial(\ln z)}{\partial(\ln u^{(j)})} = \frac{(dz/z)}{(du^{(j)}/u^{(j)})} = \frac{dz}{du^{(j)}} \cdot \frac{u^{(j)}}{z} = \hat{\theta}_j \cdot \frac{u^{(j)}}{z}.$$

Как видим, эта эластичность, в отличие от предыдущего случая, зависит от того, при каких конкретных значениях $u^{(j)}$ и z она вычисляется. Однако часто в целях упрощения в этом случае пользуются *приближенным* соотношением:

$$\Theta_{z/u^{(j)}} \approx \hat{\theta}_j \frac{\bar{u}^{(j)}}{\bar{z}} = \hat{\theta}_j \frac{\bar{x}^{(j)}}{\bar{y}},$$

где значения $\bar{x}^{(j)} = \sum_{i=1}^n x_i^{(j)}/n$ и $\bar{y} = \sum_{i=1}^n y_i/n$ — выборочные средние величины, соответственно, объясняющей переменной $x^{(j)}$ и результирующего показателя y нам известны (они подсчитаны в ходе статистического анализа рассматриваемой регрессионной модели).

2.10. Линейные модели регрессии со стохастическими объясняющими переменными

До сих пор мы изучали линейную модель множественной регрессии (ЛММР) в предположении, что объясняющие переменные $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$, 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 (2i2'). 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 10.1 (том 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 в смысле (2.2), т. е. 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 10.1, 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 случай-*

ными ошибками.

При этом, как и прежде, рассматривается ЛММР вида (обозначения определены соотношениями (2.4^а)–(2.4^б), (2.7)–(2.12))

$$Y = X\Theta + \varepsilon, \quad (2.140)$$

где

$$\text{ранг } X = p + 1 \text{ (с вероятностью единица)}^1. \quad (2.141)$$

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:

$$E\varepsilon = O_n, \quad (2.142)$$

$$\Sigma_\varepsilon = E(\varepsilon\varepsilon^\top) = \sigma^2 \Sigma_0, \quad (2.143)$$

где усреднение (операция вычисления математического ожидания E) производится в безусловном смысле, т.е. по совместному з.р.в. случайных величин ε и X . При этом, вообще говоря, $(n \times n)$ -матрица Σ_0 может зависеть от неизвестных параметров, в том числе и от параметров Θ .

2.10.1. Случайные остатки ε не зависят от предикторов X и оцениваемых коэффициентов регрессии Θ

Из взаимной независимости ε и X следует взаимная независимость ε и X . В этом случае безусловные соотношения (2.142) и (2.143) влекут за собой справедливость соответствующих соотношений для условных математических ожиданий

$$E(\varepsilon | X) = O_n, \quad (2.142')$$

$$\Sigma_\varepsilon(X) = E(\varepsilon\varepsilon^\top | X) = \sigma^2 \Sigma_0, \quad (2.143')$$

где матрица Σ_0 не зависит ни от X , ни от Θ .

Заметим, что условия (2.142') и (2.143') эквивалентны условиям

$$E(Y | X) = X\Theta, \quad (2.142'')$$

$$\Sigma_Y(X) = E\{(Y - X\Theta)(Y - X\Theta)^\top | X\} = \sigma^2 \Sigma_0 \quad (2.143'')$$

¹ Условие (2.141) сопровождается требованием его выполнения с вероятностью единица, поскольку матрица X в рамках данной постановки задачи формируется из элементов случайной выборки $x_i^{(j)}$ ($i = 1, 2, \dots, n$; $j = 1, 2, \dots, p$).

и что из общих свойств многомерного нормального распределения следует, в частности, выполнение этих условий в ситуациях, когда вектор анализируемых стохастических переменных $(x^{(1)}, x^{(2)}, \dots, x^{(p)}, y)$ подчинен $(p+1)$ -мерному нормальному закону (см. [Андерсон Т.], теорема 2.5.1).

Покажем, что к модели, специфицированной соотношениями (2.140)–(2.141)–(2.142')–(2.143'), 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 (2.143') полагается единичной, т. е.

$$\Sigma_\epsilon(X) = \sigma^2 I_n. \quad (2.143^a)$$

Мы будем предполагать также, что существуют следующие пределы (по вероятности):

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n} X^\top X \right) = \Sigma, \quad (2.144)$$

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n} X^\top \epsilon \right) = O_{p+1}, \quad (2.145)$$

где $(p+1) \times (p+1)$ -матрица Σ положительно определена (а значит существует матрица Σ^{-1}).

Заметим, что при интерпретации исходных статистических данных (X, Y) как случайной выборки из некоторой $(p+1)$ -мерной генеральной совокупности выполнение условий (2.144) и (2.145) обеспечивается действием закона больших чисел (см. том 1, п. 4.2). Действительно, в этом случае в соответствии с определением случайной выборки (см. п. 6.1) наблюдения X_1, X_2, \dots, X_n образуют последовательность независимых одинаково распределенных p -мерных случайных величин $(X_i = (x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)})^\top)$. То же самое можно сказать и о последовательности регрессионных остатков $\epsilon_1, \epsilon_2, \dots, \epsilon_n$. Далее, (k, j) -й элемент матрицы $(X^\top X)/n$ выражается

в виде суммы $\hat{s}_{kj}(n) = \sum_{i=1}^n x_i^{(k)} x_i^{(j)} / n$, а l -я компонента вектора-столбца $\mathbf{X}^T \boldsymbol{\varepsilon} / n$ — в виде суммы $\widehat{\text{cov}}(x^{(l)}, \boldsymbol{\varepsilon}) = \sum_{i=1}^n x_i^{(l)} \varepsilon_i / n$. А следовательно, при существовании конечных моментов, скажем, 4-го порядка у случайных величин $x^{(j)}$ ($j = 1, 2, \dots, p$) и с учетом ранее сделанных предположений о гомоскедастичности регрессионных остатков $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ (см. (2.143^а)) и их некоррелированности с компонентами вектора \mathbf{X} закон больших чисел обеспечивает сходимость по вероятности (при $n \rightarrow \infty$):

$$\hat{s}_{kj}(n) \rightarrow \sigma_{kj} = \mathbf{E}(x^{(k)} x^{(j)}), \quad k, j = 1, 2, \dots, p;$$

$$\widehat{\text{cov}}(x^{(l)}, \boldsymbol{\varepsilon}) \rightarrow \mathbf{E}(x^{(l)} \boldsymbol{\varepsilon}) = \text{cov}(x^{(l)}, \boldsymbol{\varepsilon}) = 0, \quad l = 1, 2, \dots, n.$$

Итак, рассмотрим МНК-оценку $\hat{\Theta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$ неизвестных параметров Θ модели (2.140)–(2.141)–(2.142')–(2.143^а) и проанализируем ее статистические свойства. Ее существование при любых выборочных данных \mathbf{X}, \mathbf{Y} обеспечивается условием (2.141).

Состоятельность МНК-оценки $\hat{\Theta}$ следует из ее представления в виде

$$\begin{aligned} \hat{\Theta} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{X}\Theta + \boldsymbol{\varepsilon}) = (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{X}^T \mathbf{X})\Theta + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\varepsilon} \\ &= \Theta + \left(\frac{1}{n} \mathbf{X}^T \mathbf{X} \right)^{-1} \left(\frac{1}{n} \mathbf{X}^T \boldsymbol{\varepsilon} \right). \end{aligned} \quad (2.146)$$

В соответствии с условиями (2.144)–(2.145) второе слагаемое правой части выражения (2.146) 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 (т. е. при зафиксированных значениях элементов матрицы \mathbf{X}), проведенный точно так же, как это было сделано в п. 2.3.4, дает:

$$\mathbf{E}(\hat{\Theta} | \mathbf{X}) = \Theta \quad (\text{см. (2.29) – (2.30)});$$

$$\Sigma_{\hat{\Theta}}(\mathbf{X}) = \mathbf{E}\{[(\hat{\Theta} - \Theta)(\hat{\Theta} - \Theta)^T] | \mathbf{X}\} = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \quad (\text{см. (2.36)});$$

$$\hat{\boldsymbol{\varepsilon}}(\mathbf{X}) = \mathbf{Y} - \mathbf{X}\hat{\Theta} = \mathbf{Z}\boldsymbol{\varepsilon} \quad (\text{см. (2.31) – (2.31^а)});$$

$$\mathbf{E}(\hat{\boldsymbol{\varepsilon}} | \mathbf{X}) = \mathbf{O}_n;$$

$$\Sigma_{\hat{\varepsilon}}(\mathbf{X}) = \mathbf{E}(\hat{\varepsilon}\hat{\varepsilon}^T | \mathbf{X}) = \mathbf{E}(\mathbf{Z}\varepsilon\varepsilon^T\mathbf{Z}^T | \mathbf{X}) = \sigma^2\mathbf{Z};$$

$$\mathbf{E}(\hat{\varepsilon}^T\hat{\varepsilon} | \mathbf{X}) = \sigma^2 \text{tr } \mathbf{Z} = \sigma^2(n - p - 1) \quad (\text{см. (2.32)});$$

$$\mathbf{E}(\hat{\sigma}^2 | \mathbf{X}) = \sigma^2 \quad (\text{см. (2.34)});$$

$$\mathbf{E}[\hat{\sigma}^2(\mathbf{X}^T\mathbf{X})^{-1} | \mathbf{X}] = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}.$$

Вывод этих соотношений опирается на тот факт, что сомножитель, зависящий от «условия» (т. е. от матрицы \mathbf{X}), можно выносить из-под знака *условного* математического ожидания.

С помощью известного метода последовательного (повторного) вычисления математического ожидания вида

$$\mathbf{E}_{(\xi, \eta)}\varphi(\xi, \eta) = \mathbf{E}_X\{\mathbf{E}_\eta[\varphi(\xi; \eta) | \xi = X]\}^1$$

легко доказывается и *безусловная* несмещенность оценок $\hat{\Theta}$ и $\hat{\sigma}^2$:

$$\mathbf{E}\hat{\Theta} = \mathbf{E}_X[\mathbf{E}_Y(\hat{\Theta} | \mathbf{X})] = \mathbf{E}_X(\Theta) = \Theta,$$

$$\mathbf{E}\hat{\sigma}^2 = \mathbf{E}_X[\mathbf{E}_Y(\hat{\sigma}^2 | \mathbf{X})] = \mathbf{E}_X(\sigma^2) = \sigma^2.$$

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 (2.38)–(2.39). При этом участвующие в*

¹ Нижний 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 $p_{(\xi, \eta)}(X, Y)$. Предположим, что существует $\mathbf{E}_{(\xi, \eta)}\varphi(\xi, \eta)$. Тогда, учитывая, что функция $p_{(\xi, \eta)}(X, Y)$ может быть представлена (см. (2.13')) 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.

этих соотношениях моменты должны быть заменены на соответствующие *условные* моменты (при условии, что матрица наблюдаемых значений объясняющих переменных равна X).

2.10.2. Общий случай: стохастические предикторы X коррелированы с регрессионными остатками ε . Метод инструментальных переменных

Из (2.146) видно, что если хотя бы одна из объясняющих переменных X асимптотически (по $n \rightarrow \infty$) коррелирует с регрессионными остатками ε , то $X^T \varepsilon/n$ не стремится (по вероятности) к нулевому вектору (т. е. не выполняется условие (2.145)) и, следовательно, МНК-оценки $\hat{\Theta}$ уже не будут состоятельными. При этом *всего один* ненулевой элемент вектора $\lim_{n \rightarrow \infty} (X^T \varepsilon/n)$ может приводить к несостоятельности *всех* компонент МНК-оценки $\hat{\Theta}$. Действительно, возвращаясь к (2.146), имеем¹:

$$\lim_{n \rightarrow \infty} \hat{\Theta} = \Theta + \lim_{n \rightarrow \infty} \left(\frac{1}{n} X^T X \right)^{-1} \lim_{n \rightarrow \infty} \left(\frac{1}{n} X^T \varepsilon \right).$$

Предположим, что в вектор-столбце $\lim_{n \rightarrow \infty} (X^T \varepsilon/n)$ на l -м месте стоит постоянная $r_l = \text{cov}(x^{(l)}, \varepsilon) \neq 0$, 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 (2.144). 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 анализа моделей такого типа используется альтернативный метод, известный в литературе как **метод инструментальных переменных**.

¹ Все 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 (см. том 1, п. 4.2).

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 $\tilde{X} = (1, x^{(1)}, \dots, x^{(p)})$; следовательно, наряду с $n \times (p+1)$ -матрицей значений \tilde{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 , т. е.

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n} U^T \varepsilon \right) = O_{p+1}; \quad (2.147)$$

(iii) перекрестные (смешанные) вторые моменты переменных U и X в пределе (по $n \rightarrow \infty$) не все равны нулю и образуют невырожденную (положительно определенную) $(p+1) \times (p+1)$ -матрицу Σ_{UX} , т. е.

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n} U^T X \right) = \Sigma_{UX}; \quad (2.148)$$

(iv) должна существовать положительно определенная матрица Σ_{UU} такая, что

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n} U^T U \right) = \Sigma_{UU}. \quad (2.149)$$

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 (2.140) определяются по формуле

$$\hat{\Theta}_U = (U^T X)^{-1} U^T Y. \quad (2.150)$$

В состоятельности оценки $\hat{\Theta}_U$ можно убедиться, подставив Y , выраженное соотношением (2.140), в (2.150), 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)

$$\begin{aligned}\lim_{n \rightarrow \infty} \hat{\Theta}_U &= \Theta + \lim_{n \rightarrow \infty} \left(\frac{1}{n} U^T X \right)^{-1} \lim_{n \rightarrow \infty} \left(\frac{1}{n} U^T \varepsilon \right) \\ &= \Theta + \Sigma_{UX}^{-1} \mathbf{0}_{p+1} = \Theta;\end{aligned}$$

2)

$$\begin{aligned}\mathbf{E}(\hat{\Theta}_U | X, U) &= \Theta \quad \text{и} \\ \mathbf{E}\hat{\Theta}_U &= \mathbf{E}_{X,U} \{ \mathbf{E}(\hat{\Theta}_U | X, U) \} = \mathbf{E}_{X,U} \Theta = \Theta,\end{aligned}$$

т.е. 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 $\Sigma_{\hat{\Theta}_U}(X, U) = \mathbf{E}\{(\hat{\Theta}_U - \Theta)(\hat{\Theta}_U - \Theta)^T | X, U\}$ может быть вычислена по формуле

$$\hat{\Sigma}_{\hat{\Theta}_U} = \hat{\sigma}^2 (U^T X)^{-1} (U^T U) (X^T U)^{-1}, \quad (2.151)$$

где

$$\hat{\sigma}^2 = \frac{1}{n - p - 1} (Y - X\hat{\Theta}_U)^T (Y - X\hat{\Theta}_U). \quad (2.152)$$

Анализ правой части (2.151) 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 (2.150')):

$$\hat{\theta}_{1U} = \frac{\sum_{i=1}^n (u_i - \bar{u})(y_i - \bar{y})}{\sum_{i=1}^n (u_i - \bar{u})(x_i - \bar{x})}, \quad (2.150')$$

а ее выборочная дисперсия (в соответствии с (2.151)–(2.152))

$$s_{\hat{\theta}_{1U}}^2 = \hat{\sigma}^2 \frac{\sum_{i=1}^n (u_i - \bar{u})^2}{\left[\sum_{i=1}^n (u_i - \bar{u})(x_i - \bar{x}) \right]^2}. \quad (2.151')$$

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 2.2. 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

$$y = \theta_0 + \Theta_1 x + \varepsilon. \quad (2.153)$$

В ходе выборочных обследований семей регистрировались значения следующих показателей:

- y_i — *удельные (т. е. 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 (т.е. $\text{cov}(x, \delta) = E[(x - Ex)\delta] = 0$) и с ε_i (т.е. $\text{cov}(\delta, \varepsilon) = E(\delta\varepsilon) = 0$);

(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 (т.е. $\text{cov}(x_i^0, \varepsilon) = \text{cov}(x_i^0, \delta) = 0$), 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 , т.е. $\text{cov}(x^*, \varepsilon) = E((x^* - Ex^*)\varepsilon) = 0$.

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 (2.153) значения x_i , выраженные через наблюдаемые величины x_i^* и будем анализировать модель (2.153) на основании данных (x_i^*, y_i) , $i = 1, 2, \dots, n$:

$$y_i = \theta_0 + \theta_1(x_i^* - \delta_i) + \varepsilon_i = \theta_0 + \theta_1 x_i^* + \varepsilon_i^*, \quad (2.153')$$

где $\varepsilon_i^* = \varepsilon_i - \theta_1 \delta_i$. Перепишем уравнение (2.153') в терминах *центрированных* переменных:

$$y_i - \bar{y} = \theta_1(x_i^* - \bar{x}^*) + \varepsilon_i^*. \quad (2.153'')$$

Далее можно было бы попытаться воспользоваться МНК-оценками параметров θ_0 и θ_1 по данным (x_i^*, y_i) , в частности (см. (2.18)):

$$\hat{\theta}_{1. \text{мнк}} = \frac{\sum_{i=1}^n (x_i^* - \bar{x}^*)(y_i - \bar{y})}{\sum_{i=1}^n (x_i^* - \bar{x}^*)^2}. \quad (2.154)$$

Однако легко убедиться в том, что объясняющая переменная x^* коррелирована с регрессионными остатками ε^* в модели (2.153'') и, следовательно, в соответствии с (2.146) МНК-оценки будут несостоятельными и обладать асимптотически неустрашимым смещением.

Действительно:

$$\begin{aligned} \text{cov}(x^*, \varepsilon^*) &= E((x^* - Ex^*)\varepsilon^*) = E((x^* - Ex^*)(\varepsilon - \theta_1 \delta)) \\ &= \text{cov}(x^*, \varepsilon) - \theta_1 \text{cov}(x^*, \delta) = -\theta_1 \text{cov}(x + \delta, \delta) = -\theta_1 \text{cov}(\delta, \delta) \\ &= -\theta_1 \sigma_\delta^2 > 0, \end{aligned}$$

где $\sigma_\delta^2 = D\delta$ — дисперсия ошибки δ . Т.е. x^* коррелирована с ε^* при $\theta_1 \neq 0$.

Вычислим смещение оценки (2.154):

$$\begin{aligned}\hat{\theta}_{1. \text{мнк}} &= \frac{\sum_{i=1}^n (x_i^* - \bar{x}^*) [\theta_1 (x_i^* - \bar{x}^*) + (\varepsilon_i - \theta \delta_i)]}{\sum_{i=1}^n (x_i^* - \bar{x}^*)^2} \\ &= \theta_1 + \frac{1}{\frac{1}{n} \sum_{i=1}^n (x_i^* - \bar{x}^*)^2} \left[\widehat{\text{cov}}(x^*, \varepsilon) - \theta_1 \cdot \frac{1}{n} \sum_{i=1}^n (x_i + \delta_i - \bar{x} - \bar{\delta}) \delta_i \right] \\ &= \theta_1 + \frac{1}{s_{x^*}^2} [\widehat{\text{cov}}(x^*, \varepsilon) - \theta_1 \widehat{\text{cov}}(x, \delta) - \theta_1 s_{\delta}^2].\end{aligned}$$

В выполненных тождественных преобразованиях мы использовали следующие обозначения:

$$s_{x^*}^2 = \frac{1}{n} \sum_{i=1}^n (x_i^* - \bar{x}^*)^2 \text{ — выборочная дисперсия переменной } x^*;$$

$\widehat{\text{cov}}(\xi, \eta) = \frac{1}{n} \sum_{i=1}^n (\xi_i - \bar{\xi})(\eta_i - \bar{\eta})$ — выборочная ковариация переменных ξ и η ;

$$s_{\delta}^2 = \widehat{\text{cov}}(\delta, \delta) = \frac{1}{n} \sum_{i=1}^n (\delta_i - \bar{\delta})^2 \text{ — выборочная дисперсия ошибки } \delta.$$

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:

$$\lim_{n \rightarrow \infty} \hat{\theta}_{1. \text{мнк}} = \theta_1 \left(1 - \frac{D\delta}{Dx^*} \right) = \theta_1 \left(1 - \frac{D\delta}{Dx + D\delta} \right) = \theta_1 \frac{1}{1 + \frac{D\delta}{Dx}}. \quad (2.155)$$

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 $\varepsilon^* = \varepsilon - \theta_1 \delta$ (т. е. x^0 некоррелирована 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

$$\hat{\theta}_{1.u} = \frac{\sum_{i=1}^n (x_i^0 - \bar{x}^0)(y_i - \bar{y})}{\sum_{i=1}^n (x_i^* - \bar{x}^*)(x_i^0 - \bar{x}^0)}.$$

2.10.3. 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 2.2 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

$$Y^* = Y + \Delta_Y. \quad (2.156)$$

Подставляя выраженные из (2.156) значения Y в модель (2.140), получаем:

$$Y^* = X\theta + (\varepsilon + \Delta_Y). \quad (2.140^*)$$

Анализ (по стандартной схеме, см. п. 2.3) полученных по наблюдениям (X, Y^*) МНК-оценок

$$\hat{T}_{\text{МНК}} = (X^T X)^{-1} X^T Y^* = (X^T X)^{-1} X^T (X\theta + \varepsilon + \Delta_Y), \quad (2.157)$$

проведенный с учетом упомянутых выше свойств вектора случайных ошибок Δ_Y ($E\Delta_Y = \mathbf{0}_n$, $E(\Delta_Y \varepsilon^T) = \mathbf{0}_{n,n}$, $E(\Delta_Y X^T) = \mathbf{0}_{n,p}$), подтвержда-

ет, что МНК-оценки (2.157) остаются несмещенными, состоятельными и оптимальными (в смысле (2.38)–(2.39)).

Теперь проанализируем, как влияют случайные ошибки в измерении значений объясняющих переменных на МНК-оценки параметров классической ЛММР (2.140). При этом мы будем предполагать:

$$\mathbf{X}^* = \mathbf{X} + \Delta_X; \quad (2.158)$$

$$\mathbf{E}\Delta_X = \mathbf{O}_{p+1}; \quad (2.159)$$

$$\mathbf{E}(\Delta_X \boldsymbol{\varepsilon}^\top) = \mathbf{O}_{p+1, n}; \quad (2.160)$$

$$\mathbf{E}(\Delta_X \tilde{\mathbf{X}}^\top) = \mathbf{O}_{p+1, p+1}; \quad (2.161)$$

$$\mathbf{E}(\tilde{\mathbf{X}}^* \boldsymbol{\varepsilon}^\top) = \mathbf{O}_{p+1, n}; \quad (2.162)$$

$$\boldsymbol{\Sigma}_{\Delta_X} = \mathbf{E}(\Delta_X \Delta_X^\top). \quad (2.163)$$

В соотношениях (2.158)–(2.163) $\Delta_X = (\Delta_X^{(0)}, \Delta_X^{(1)}, \dots, \Delta_X^{(p)})^\top$ — вектор случайных ошибок в измерении объясняющих переменных $\tilde{\mathbf{X}} = (1, x^{(1)}, \dots, x^{(p)})^\top$; Δ_X — $n \times (p+1)$ -матрица значений ошибок $\Delta_i^{(j)}$ в измерениях $x_i^{(j)}$; соответственно \mathbf{X}^* — $n \times (p+1)$ -матрица искаженных наблюдаемых значений объясняющих переменных, а $\tilde{\mathbf{X}}^* = (1, x^{*(1)}, \dots, x^{*(p)})^\top$ — вектор наблюдаемых (искаженных) значений объясняющих переменных; соотношения (2.160)–(2.161) свидетельствуют о некоррелированности ошибок измерения с регрессионными остатками и объясняющими переменными (откуда следует некоррелированность искаженных объясняющих переменных с регрессионными остатками, см. (2.162)); и, наконец, соотношением (2.163) определена ковариационная матрица вектора ошибок измерения Δ_X .

Итак, мы рассматриваем классическую ЛММР со стохастическими объясняющими переменными (2.140)–(2.142)–(2.143^а)–(2.144)–(2.145) в условиях случайных ошибок в измерении предикторов (2.158)–(2.163). Подставим в модель (2.140) значения \mathbf{X} , выраженные из (2.158) в терминах *наблюдаемых* (но искаженных) значений объясняющих переменных:

$$\mathbf{Y} = \mathbf{X}^* \boldsymbol{\Theta} + (\boldsymbol{\varepsilon} - \Delta_X \boldsymbol{\Theta}). \quad (2.140^b)$$

МНК-оценка для $\boldsymbol{\Theta}$ из (2.140^b) есть

$$\begin{aligned} \hat{\boldsymbol{\Theta}}_{\text{МНК}} &= (\mathbf{X}^{*\top} \mathbf{X}^*)^{-1} \mathbf{X}^{*\top} \mathbf{Y} = (\mathbf{X}^{*\top} \mathbf{X}^*)^{-1} \mathbf{X}^{*\top} [\mathbf{X}^* \boldsymbol{\Theta} + \boldsymbol{\varepsilon} - \Delta_X \boldsymbol{\Theta}] \\ &= \boldsymbol{\Theta} + \left(\frac{1}{n} \mathbf{X}^{*\top} \mathbf{X}^* \right)^{-1} \left[\frac{1}{n} \mathbf{X}^{*\top} \boldsymbol{\varepsilon} - \left(\frac{1}{n} \mathbf{X}^{*\top} \Delta_X \right) \boldsymbol{\Theta} \right]. \end{aligned} \quad (2.164)$$

Проанализируем поведение правой части (2.164) при $n \rightarrow \infty$.

1)

$$\begin{aligned} \lim_{n \rightarrow \infty} \left(\frac{1}{n} \mathbf{X}^{*\top} \mathbf{X}^* \right) &= \lim_{n \rightarrow \infty} \left(\frac{1}{n} (\mathbf{X} + \Delta_{\mathbf{X}})^{\top} (\mathbf{X} + \Delta_{\mathbf{X}}) \right) \\ &= \lim_{n \rightarrow \infty} \left(\frac{1}{n} \mathbf{X}^{\top} \mathbf{X} \right) + \lim_{n \rightarrow \infty} \left(\frac{1}{n} \Delta_{\mathbf{X}}^{\top} \mathbf{X} \right) + \lim_{n \rightarrow \infty} \left(\frac{1}{n} \mathbf{X}^{\top} \Delta_{\mathbf{X}} \right) \\ &\quad + \lim_{n \rightarrow \infty} \left(\frac{1}{n} \Delta_{\mathbf{X}}^{\top} \Delta_{\mathbf{X}} \right) = \Sigma + \Sigma_{\Delta_{\mathbf{X}}} \end{aligned} \quad (2.165)$$

(полученный результат опирается на закон больших чисел (З.Б.Ч) и соотношения (2.144), (2.161) и (2.163)).

2)

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n} \mathbf{X}^{*\top} \varepsilon \right) = \mathbf{O}_{p+1} \quad (2.166)$$

(полученный результат опирается на З.Б.Ч. и соотношение (2.162)).

3)

$$\begin{aligned} \lim_{n \rightarrow \infty} \left(\frac{1}{n} \mathbf{X}^{*\top} \Delta_{\mathbf{X}} \right) &= \lim_{n \rightarrow \infty} \left(\frac{1}{n} (\mathbf{X} + \Delta_{\mathbf{X}})^{\top} \Delta_{\mathbf{X}} \right) \\ &= \lim_{n \rightarrow \infty} \left(\frac{1}{n} \mathbf{X}^{\top} \Delta_{\mathbf{X}} \right) + \lim_{n \rightarrow \infty} \left(\frac{1}{n} \Delta_{\mathbf{X}}^{\top} \Delta_{\mathbf{X}} \right) = \Sigma_{\Delta_{\mathbf{X}}} \end{aligned} \quad (2.167)$$

(полученный результат опирается на З.Б.Ч. и соотношения (2.161) и (2.163)).

Переходя к пределу (по вероятности) при $n \rightarrow \infty$ в (2.164) и учитывая (2.165)–(2.167), имеем:

$$\lim_{n \rightarrow \infty} \hat{\Theta}_{\text{МНК}} = \Theta - (\Sigma + \Sigma_{\Delta_{\mathbf{X}}})^{-1} \Sigma_{\Delta_{\mathbf{X}}} \Theta = [\mathbf{I}_{p+1} - (\Sigma + \Sigma_{\Delta_{\mathbf{X}}})^{-1} \Sigma_{\Delta_{\mathbf{X}}}] \Theta. \quad (2.168)$$

Как видим из (2.168), оценка $\hat{\Theta}_{\text{МНК}}$ в данном случае не является ни состоятельной, ни несмещенной (ее асимптотически-неустраняемое смещение оказывается равным, в соответствии с (2.168), величине $-(\Sigma + \Sigma_{\Delta_{\mathbf{X}}})^{-1} \Sigma_{\Delta_{\mathbf{X}}} \Theta$). 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 (т.е. при $p = 1$) формула (2.168), естественно, повторяет ранее выведенную нами (при анализе примера 2.2) формулу (2.155).

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 $\Sigma_{\Delta x}$, то на основании (2.168) можно построить «подправленную МНК-оценку»

$$\hat{\Theta}_{\text{МНК}}^* = [I_{p+1} - (\Sigma + \Sigma_{\Delta x})^{-1} \Sigma_{\Delta x}]^{-1} \hat{\Theta}_{\text{МНК}}. \quad (2.169)$$

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 2.2 (см. модель (2.153)–(2.154)–(2.153')–(2.153'')). 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 (2.153''), записанной в терминах центрированных переменных:

$$X_{\text{ц}}^{\text{T}} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1^* - \bar{x}^* & x_2^* - \bar{x}^* & \dots & x_n^* - \bar{x}^* \end{pmatrix}$$

и

$$Y_{\text{ц}} = (y_1 - \bar{y}, y_2 - \bar{y}, \dots, y_n - \bar{y})^{\text{T}}.$$

Определим матрицу значений инструментальных переменных U следующим образом:

$$U^{\text{T}} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ -1 & +1 & \dots & -1 \end{pmatrix},$$

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 $X_{\text{ц}}^{\text{T}}$ выше или ниже медианы выборки ($x_1^* - \bar{x}^*, x_2^* - \bar{x}^*, \dots, x_n^* - \bar{x}^*$). В соответствии с (2.150) оценка по методу инструментальной переменной окажется равной в данном случае

$$\hat{\Theta}_U = \begin{pmatrix} n & 0 \\ 0 & \frac{n}{2}(\bar{x}_{\text{ц}2}^* - \bar{x}_{\text{ц}1}^*) \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \frac{n}{2}(\bar{y}_{\text{ц}2} - \bar{y}_{\text{ц}1}) \end{pmatrix}, \quad (2.170)$$

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 (2.170) получаем:

$$\hat{\theta}_{0U} = 0; \quad \hat{\theta}_{1U} = \frac{\bar{y}_{u2} - \bar{y}_{u1}}{\bar{x}_{u2}^* - \bar{x}_{u1}^*}. \quad (2.171)$$

Соотношения (2.170)–(2.171) определяют способ оценивания, впервые предложенный Вальдом.¹ Когда значение n нечетно, прежде чем приступить к вычислениям, нужно исключить среднее значение. При весьма общих предположениях оценка (2.171) 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.² Исключив центральные $n/3$ наблюдений, получим оценку для коэффициента наклона линии регрессии

$$\hat{\theta}'_{1U} = \frac{\bar{y}_3 - \bar{y}_1}{\bar{x}_3 - \bar{x}_1}, \quad (2.172)$$

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 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 $y(\tilde{X}_{n+1})$ по заданному значению объясняющей переменной $\tilde{X}_{n+1} = (1, x_{n+1}^{(1)}, \dots, x_{n+1}^{(p)})^T$ (см. п. 2.9.2). Пусть значение \tilde{X}_{n+1} задается со случайной ошибкой (что чаще всего и бывает в действительности, так как в качестве \tilde{X}_{n+1} берутся обычно *прогнозные* или *планируемые* значения предикторов), т. е. мы располагаем значениями

$$\tilde{X}_{n+1}^* = \tilde{X}_{n+1} + \Delta x,$$

¹ Wald A. The Fitting of Straight Lines if Both Variables are Subject to Error. Ann. Math. Stat., 1940, vol. 11, pp. 284–300.

² Bartlett M.S. Fitting a Straight Line when Both Variables are Subject to Error. Biometrics, 1949, vol. 5, pp. 207–212.

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 (т.е. в рамках модели (2.140)–(2.142')–(2.143^{*})–(2.144)–(2.145)), то наилучший линейный несмещенный прогноз $\hat{y}(\tilde{X}_{n+1})$ неизвестного значения $y(\tilde{X}_{n+1})$ задается формулой

$$\hat{y}(\tilde{X}_{n+1}) = \tilde{X}^{*T} \hat{\Theta}_{\text{мнк}}, \quad (2.173)$$

а его средний квадрат ошибки определяется соотношением:

$$E(\hat{y}(\tilde{X}_{n+1}) - y(\tilde{X}_{n+1}))^2 = \sigma^2 [1 + \tilde{X}_{n+1}^T (\mathbf{X}^T \mathbf{X})^{-1} \tilde{X}_{n+1} + \sigma_{\Delta}^2 \text{tr}(\mathbf{X}^T \mathbf{X})^{-1}] + \sigma_{\Delta}^2 \Theta^T \Theta. \quad (2.174)$$

В соотношении (2.173) оценки $\hat{\Theta}_{\text{мнк}}$ строятся по *наблюдаемой* матрице \mathbf{X}^* (т.е. $\hat{\Theta}_{\text{мнк}} = (\mathbf{X}^{*T} \mathbf{X}^*)^{-1} \mathbf{X}^{*T} Y$). При статистической оценке величины среднего квадрата ошибки прогноза значения $\sigma^2, \sigma_{\Delta}^2$ и Θ заменяются их статистическими оценками, а *ненаблюдаемые* вектор \tilde{X}_{n+1} и матрица \mathbf{X} — соответственно наблюдаемыми \tilde{X}_{n+1}^* и \mathbf{X}^* .

Сравнение правой части (2.174) с правой частью ранее выведенной формулы (2.135) 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).

2.11. Линейные регрессионные модели с переменной структурой (построение линейной модели по неоднородным регрессионным данным)

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 (т.е. в

значениях коэффициентов регрессии $\theta_0, \theta_1, \dots, \theta_p$). Попробуем формализовать сказанное.

2.11.1. Проблема неоднородных (в регрессионном смысле) данных

Пусть функция плотности вероятности анализируемого результирующего показателя y зависит не только от значений объясняющих переменных (регрессоров) $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$, но и от того, на каких конкретных уровнях (градациях) зафиксированы так называемые *сопутствующие*, как правило, *качественные* переменные $Z = (z^{(1)}, z^{(2)}, \dots, z^{(k)})^T$, 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 $p_y(\tilde{y} | X; Z)$ 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 (т. е. при любых значениях сопутствующих переменных Z) функция регрессии y по X остается линейной, то естественно предположить, что коэффициенты этой линейной зависимости (т. е. параметры $\Theta = (\theta_0, \theta_1, \dots, \theta_p)^T$), вообще говоря, зависят от Z , т. е.

$$f(X | Z) = E(y | X, Z) = \theta_0(Z) + \theta_1(Z)x^{(1)} + \dots + \theta_p(Z)x^{(p)}. \quad (2.175)$$

Исходные статистические данные называются однородными (в регрессионном смысле), если все они зарегистрированы при одних и тех же условиях (т. е. при одних и тех же значениях сопутствующих переменных Z^), т. е.*

$$\tilde{B}_n = \{(X_1, y_1, Z^*), (X_2, y_2, Z^*), \dots, (X_n, y_n, Z^*)\}.$$

Если же эти данные объединяют в себе наблюдения, зарегистрированные при различных условиях (значениях сопутствующих переменных Z), то они, вообще говоря, могут быть **неоднородными**. Тогда соответственно:

$$\tilde{B}_n = \{(X_1, y_1, Z_1), (X_2, y_2, Z_2), \dots, (X_n, y_n, Z_n)\},$$

причем найдется по меньшей мере пара наблюдений (с номерами i_1 и i_2) таких, что $Z_{i_1} \neq Z_{i_2}$.

Проблема, которой посвящен данный параграф, 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 (2.175)? Казалось бы, ответ тривиален: разбить имеющиеся в нашем распоряжении исходные статистические данные \tilde{B}_n на заведомо однородные порции (т.е. 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 (2.175):

$$\hat{f}(X | Z^*) = \hat{\theta}_0(Z^*) + \hat{\theta}_1(Z^*)x^{(1)} + \dots + \hat{\theta}_p(Z^*)x^{(p)}. \quad (2.176)$$

При этом оценки $\hat{f}(X | Z^*)$ и $\hat{f}(X | Z^{**})$, полученные по двум различным однородным подвыборкам, *могут, вообще говоря, статистически значимо отличаться друг от друга*, т.е. могут представлять собой эмпирические версии двух различных теоретических функций регрессии.

Пример 2.3. Вернемся к условиям примера 2.2, 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 (2.153), будет существенно различным для беднейших и состоятельных социально-экономических страт общества: для первых величина θ_1 может достигать значений $0,6 \sim 0,7$, в то время как для вторых она колеблется в пределах $0,2 \sim 0,4$. Если же речь идет о потреблении, скажем, пива и прохладительных напитков, то даже в рамках одной и той же стра-

ты населения и одинаковых «склонностей к потреблению» θ_1 количество потребляемого продукта будет подвержено существенному *сезонному варьированию*: очевидно, в модели регрессии, построенной по *летним* исходным данным, оценка θ_0 будет статистически значимо превышать оценку $\hat{\theta}_0$, построенную, например, по *зимним* исходным данным. В дальнейшем мы еще вернемся к рассмотрению примера 2.3.

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 2.3, 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* (т. е. 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 п. 12.3 и 12.4 данного учебника.

б) Сопутствующие переменные Z наблюдаемы (т. е. известны значения $Z_i, i = 1, 2, \dots, n$), но прямое (т. е. по значениям Z) разбиение выборки \tilde{B}_n на регрессионно однородные подвыборки приводит к слишком малым подвыборкам $\tilde{B}_n^*(z^*)$, т. е. к подвыборкам таких объемов, которых оказывается недостаточно для статистически надежной оценки искомых функций регрессии (2.175). В этих случаях для оценки искомых функций регрессии привлекают подходы, связанные с введением так называемых *фиктивных переменных*¹

¹ В англоязычной литературе по эконометрике эти переменные называются «dummy variables». В соответствующей русскоязычной литературе исторически «привився» 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 $x^{(j)}$ *вполне реальна*, а слово «dummy» определяет лишь способ ее кодирования в искомом уравнении.

или с использованием аппарата *ковариационного анализа* (описание методов ковариационного анализа читатель найдет, например, в книге [Айвазян С. А., Енюков И. С., Мешалкин Л. Д., 1985, п. 13.5]).

2.11.2. 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 гипотезы 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 *дихотомических (бинарных, булевых)* переменных, т. е. таких переменных, которые могут принимать одно из двух возможных значений («ноль» или «единица»). При этом, если сопутствующая качественная переменная $z^{(j)}$ имеет k_j градаций (т. е. может принимать k_j возможных «значений»), то для отражения ее влияния на структуру искомой регрессионной связи необходимо ввести $k_j - 1$ дихотомических переменных $z^{(j,1)}, z^{(j,2)}, \dots, z^{(j,k_j-1)}$. Конкретная форма, в которой эти переменные будут представлены в анализируемом уравнении, будет зависеть от наших допущений о характере влияния сопутствующей переменной на коэффициенты исходной модели регрессии. Поясним это на примере 2.3.

Пример 2.3 (продолжение). Предположим, что $y_{(руб)}$ — это удельное (т. е. в расчете на одного члена семьи и один такт времени) потребление пива и прохладительных напитков. Тогда векторная сопут-

ствующая качественная переменная Z состоит из двух компонент, т.е. $Z = (z^{(1)}, z^{(2)})^T$, 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» (т.е. число градаций k_1 качественной переменной $z^{(1)}$ равно трем). Из определения $z^{(2)}$ следует, что число ее градаций (k_2) равно четырем (1 — зима, 2 — весна, 3 — лето, 4 — осень).

Чтобы учесть влияние сопутствующей переменной $z^{(1)}$ на структуру модели (2.153), введем $k_1 - 1 = 2$ фиктивные переменные:

$$z_i^{(1.1)} = \begin{cases} 1, & \text{если } i\text{-е наблюдение относилось к домашнему хозяйству,} \\ & \text{принадлежащему страте 2 (т.е. к } \textit{среднедоходным}); \\ 0 & \text{в противном случае;} \end{cases}$$

$$z_i^{(1.2)} = \begin{cases} 1, & \text{если } i\text{-е наблюдение относилось к домашнему хозяйству,} \\ & \text{принадлежащему страте 3 (т.е. к } \textit{высокодоходным}); \\ 0 & \text{в противном случае.} \end{cases}$$

Учет сезонности (т.е. влияния сопутствующей переменной $z^{(2)}$) на структуру модели (2.153) осуществляется введением $k_2 - 1 = 3$ 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 (2.153) 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* (т.е. на коэффициент θ_1).

Тогда анализируемое уравнение регрессии следует записать в виде:

$$y_i = \theta_0 + \theta_1 x_i + \theta_{01} z_i^{(1.1)} + \theta_{02} z_i^{(1.2)} + \theta_{03} z_i^{(2.1)} + \theta_{04} z_i^{(2.2)} + \theta_{05} z_i^{(2.3)} + \varepsilon_i. \quad (2.177)$$

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:

$$y_i = \hat{\theta}_0 + \hat{\theta}_1 x_i + \varepsilon_i \text{ — для семей 1-й страты в зимний период;}$$

$$y_i = (\hat{\theta}_0 + \hat{\theta}_{03}) + \hat{\theta}_1 x_i + \varepsilon_i \text{ — для семей 1-й страты в весенний период;}$$

$$y_i = (\hat{\theta}_0 + \hat{\theta}_{04}) + \hat{\theta}_1 x_i + \varepsilon_i \text{ — для семей 1-й страты в летний период;}$$

$$y_i = (\hat{\theta}_0 + \hat{\theta}_{05}) + \hat{\theta}_1 x_i + \varepsilon_i \text{ — для семей 1-й страты осенний период;}$$

$$y_i = (\hat{\theta}_0 + \hat{\theta}_{01}) + \hat{\theta}_1 x_i + \varepsilon_i \text{ — для семей 2-й страты в зимний период;}$$

$$y_i = (\hat{\theta}_0 + \hat{\theta}_{01} + \hat{\theta}_{03}) + \hat{\theta}_1 x_i + \varepsilon_i \text{ — для семей 2-й страты в весенний период;}$$

и т. д.

$$y_i = (\hat{\theta}_0 + \hat{\theta}_{02} + \hat{\theta}_{05}) + \hat{\theta}_1 x_i + \varepsilon_i \text{ — для семей 3-й страты в осенний период.}$$

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 (11.39) и (11.40) 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 (2.153): 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 2.3 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 (т. е. $n_1 = n_2 = \dots = n_{12} = 12$). Поэтому при *изолированной* оценке параметров θ_0 и θ_1 модели (2.153) *отдельно по каждой однородной выборке* отношение $n/(p+1)$ составит величину $12/2 = 6$, в то время как при оценке *семи* параметров модели (2.177) по всем 144 наблюдениям это отношение оказывается равным $144/7=20,57$, т. е. 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 (2.177):

$$y_i = \underset{(2,1)}{6,8} + \underset{(0,006)}{0,025} x + \underset{(1,2)}{3,1} z_i^{(1,1)} + \underset{(2,1)}{5,2} z_i^{(1,2)} + \underset{(1,20)}{0,80} z_i^{(2,1)} + \underset{(1,1)}{4,7} z_i^{(2,2)} - \underset{(11,8)}{1,2} z_i^{(2,3)} + \varepsilon_i.$$

Отсюда непосредственно следует, что фиктивные переменные $z^{(2,1)}$ и $z^{(2,3)}$ статистически не значимо влияют на y , т. е. 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 k_j градаций, следует вводить $k_j - 1$ фиктивных переменных. Предположим, что в примере 2.3 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

$$z_i^{(1.3)} = \begin{cases} 1, & \text{если } i\text{-е наблюдение относилось к домашнему хозяйству,} \\ & \text{принадлежащему низкодоходной страте;} \\ 0 & \text{в противном случае.} \end{cases}$$

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 низкодоходной страте, то столбцы матрицы наблюдений X , задающие наблюдения по переменным $z^{(1.1)}$, $z^{(1.2)}$ и $z^{(1.3)}$, будут представлены в виде (см. табл. 2.4):

Таблица 2.4

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

Просуммировав эти три столбца матрицы X , мы получим столбец, состоящий из одних единиц, т. е. столбец, повторяющий 1-й столбец общей матрицы X , соответствующий свободному члену анализируемого уравнения (2.177). 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. 2.4.1). 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.

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:

$$y_i = \theta_0 + \theta_1 x_i + \theta_{11}(z_i^{(1.1)} x_i) + \theta_{12}(z_i^{(1.2)} x_i) + \theta_{01} z_i^{(2.1)} + \theta_{02} z_i^{(2.2)} + \theta_{03} z_i^{(2.3)} + \varepsilon_i. \quad (2.177')$$

При этом очевидным образом меняются способ и результат комплектования общей $(n \times 7)$ -матрицы наблюдений X^1 и соответственно результаты метода наименьших квадратов. Легко видеть, что в данном случае в качестве оценок «склонности к потреблению» мы получим:

$$\begin{aligned} \text{для страты 1:} & \quad \hat{\theta}_1; \\ \text{для страты 2:} & \quad \hat{\theta}_1 + \hat{\theta}_{11}; \\ \text{для страты 3:} & \quad \hat{\theta}_1 + \hat{\theta}_{12}. \end{aligned}$$

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 $z^{(j.1)}, z^{(j.2)}, \dots, z^{(j.k_j-1)}$) и сопутствующей переменной $z^{(l)}$ (представленной в уравнении фиктивными переменными $z^{(l.1)}, z^{(l.2)}, \dots, z^{(l.k_l-1)}$) может быть учтен введением в анализируемое уравнение дополнительно $N = (k_j - 1)(k_l - 1)$ 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 $q = 1, 2, \dots, k_j - 1$ и $s = 1, 2, \dots, k_l - 1$ (напомним, что k_j и k_l — это

¹ Очевидно, 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 : $\tilde{x}_i^{(3)} = 0$).

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» $\tilde{z}^{(i)}$ могут оказаться статистически не значимо отличающимися от нуля, что будет означать отсутствие влияния соответствующих взаимодействий на структуру модели. Поясним это на примере.

Пример 2.4. Пусть y (руб.) — заработная плата работника; $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$ — набор количественных признаков, от которых может зависеть величина y (трудовой стаж; оценки, выставленные работнику по ряду профессиональных тестов, и т. п.);¹ $z^{(1)}$ — сопутствующая переменная, определяющая уровень образования работника ($k_1 = 3$: начальное, среднее и высшее) и $z^{(2)}$ — сопутствующая переменная, определяющая пол работника ($k_2 = 2$: мужской и женский). В соответствии с приведенными выше рекомендациями вводим фиктивные 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 2.5 отражает правила формирования элементов матрицы наблюдений X в той ее части, которая относится к «значениям» фиктивных переменных.

¹ 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 (см. ниже, п. 2.12).

Таблица 2.5

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:

$$y_i = \theta_0 + \theta_1 x_i^{(1)} + \dots + \theta_p x_i^{(p)} + \theta_{01} z_i^{(1.1)} + \theta_{02} z_i^{(1.2)} + \theta_{03} z_i^{(2.1)} + \theta_{04} \tilde{z}_i^{(1)} + \theta_{05} \tilde{z}_i^{(2)} + \varepsilon_i. \quad (2.178)$$

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 (т.е. после потенцирования уравнения (2.178)) 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 (2.178) позволяет выписать *специфицированные* (т.е. относящиеся только к однородным по $z^{(1)}$ и $z^{(2)}$ ситуациям) модели зависимости заработной платы от рассматриваемых объясняющих переменных, см. табл. 2.6.

Таблица 2.6

№ 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 не значимое 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.

2.11.3. Проверка регрессионной однородности двух групп наблюдений (критерий Г. Чоу)

Предположим, сопутствующая переменная z принимала в процессе сбора регрессионных наблюдений

$$\tilde{B}_n = \{(X_1, y_1), (X_2, y_2), \dots, (X_n, y_n)\} \quad (2.179)$$

всего два значения: z_1 и z_2 . Тогда общая выборка \tilde{B}_n может быть поделена

на две подвыборки

$$\tilde{B}_{n_1}^{(1)} = \{(X_{i_1}, y_{i_1}), (X_{i_2}, y_{i_2}), \dots, (X_{i_{n_1}}, y_{i_{n_1}}) \mid z = z_1\}, \quad (2.179^a)$$

$$\tilde{B}_{n_2}^{(2)} = \{(X_{j_1}, y_{j_1}), (X_{j_2}, y_{j_2}), \dots, (X_{j_{n_2}}, y_{j_{n_2}}) \mid z = z_2\} \quad (2.179^b)$$

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

$$H_0: \Theta^{(1)} = \Theta^{(2)}, D\varepsilon^{(1)} = D\varepsilon^{(2)} = \sigma^2,$$

где $\Theta^{(j)} = (\theta_0^{(j)}, \theta_1^{(j)}, \dots, \theta_p^{(j)})$ и $\varepsilon^{(j)}$ соответственно коэффициенты регрессии и случайные регрессионные остатки в КЛМНР (2.5), 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 $\hat{\Theta}^{(1)} = (\hat{\theta}_0^{(1)}, \hat{\theta}_1^{(1)}, \dots, \hat{\theta}_p^{(1)})^T$, 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 (2.5) 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 (2.5) 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 $F(n_2, n_1 - p - 1)$ (см. п. 3.2.3 данного учебника). 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).

2.11.4. Построение КЛММР по неоднородным данным в условиях, когда значения сопутствующих переменных неизвестны

Если воздействия сопутствующих качественных факторов на структуру КЛММР скрыты, т.е. 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 экономических исследованиях. Для подтверждения этого тезиса рассмотрим следующий пример.

Пример 2.5. В целях исследования зависимости интенсивности региональных эмиграционных процессов ($y\%$) от уровня (общей продолжительности) полученного образования (x лет) были собраны исходные статистические данные вида (2.179) за определенный промежуток времени в анализируемом регионе. Так что элемент (x_i, y_i) выборки (2.179) интерпретируется в данном случае следующим образом: x_i (лет) — общая продолжительность процесса обучения взрослого (т.е. 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 (2.179) 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. 2.6), естественно предположить, что наша выборка регрессионно не однородна 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 и 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 (2.179) на подвыборки (2.179^a) и (2.179^b) по этой сопутствующей пе-

ременной 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. 2.6) интенсивность эмиграции *падает* (по линейному закону) с ростом уровня образования, то для жителей с естественно-научно-техническим образованием (им соответствуют точки на рис. 2.6) мы наблюдаем обратную картину.

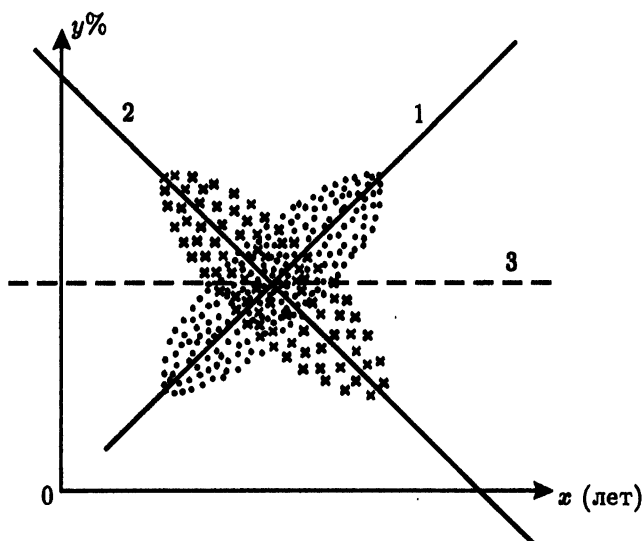


Рис. 2.6. Прямые 1, 2 и 3 — графики аппроксимирующих функций регрессии, построенных соответственно по наблюдениям подвыборок: 1 (точки), 2 (крестики) и по объединенной выборке, состоящей из тех и других наблюдений

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 (т. е. при $p = 3$) визуальный анализ геометрической структуры исходной выборки (2.179) *принципиально невозможен*. Поэтому с целью предварительного анализа геометрической структуры данных (2.179) и выделения в них регрессионно однородных подвыборок необходимо использовать методы кластер-анализа, в том числе методы расщепления смеси многомерных распределений вероятностей (см. том 1, пп. 12.3, 12.4). Подчеркнем, что разбиение исходных данных (2.179) на регрессионно одно-

родные подвыборки (кластеры) проводится в *объединенном* $(p+1)$ -мерном пространстве (X, y) (а не в пространстве только объясняющих переменных X). Причем процесс этот либо предшествует процессу построения регрессионных моделей, либо проводится в режиме итерационного взаимодействия с последним.

2.12. Нелинейные модели регрессии и линеаризация

До сих пор мы рассматривали лишь *линейные* модели регрессионной зависимости y от $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$. Направления обобщений рассматриваемых моделей касались только смягчения ограничений на природу регрессионных остатков ϵ (см. пп. 2.6–2.8). В то же время многие важные связи в экономике являются *нелинейными*. Примеры такого рода регрессионных моделей доставляет нам изучение так называемых *производственных функций* (зависимостей, существующих между объемом произведенной продукции и основными факторами производства — трудом, капиталом и т. п.), функций *спроса* (зависимостей, существующих между спросом на какой-либо вид товаров или услуг, с одной стороны, и доходом и ценами на этот и другие товары — с другой). Ниже мы подробнее обсудим возможный общий вид этих и других моделей. Как уже было отмечено выше (см. том 1, п. 10.7), *этап параметризации регрессионной модели*, т. е. выбора параметрического семейства функций $\{f(X; \Theta)\}$, в рамках которого производится дальнейший поиск неизвестной функции регрессии $f(X) = E(y | X)$, является одновременно наиболее важным и наименее формализованным и теоретически обоснованным этапом регрессионного анализа. Если же в результате реализации этого этапа экономист пришел к выводу, что функция $f(X; \Theta)$ нелинейна, то далее он обычно действует следующим образом:

- вначале он пытается подобрать такие преобразования к анализируемым переменным $y, x^{(1)}, \dots, x^{(p)}$, которые позволили бы представить искомую зависимость в виде *линейного* соотношения между преобразованными переменными; другими словами, если $\varphi_0, \varphi_1, \dots, \varphi_p$ — те самые искомые функции, которые определяют переход к преобразованным переменным, т. е. $\tilde{y} = \varphi_0(y), \tilde{x}^{(1)} = \varphi_1(x^{(1)}), \dots, \tilde{x}^{(p)} = \varphi_p(x^{(p)})$, то связь между y и $X = (x^{(1)}, \dots, x^{(p)})$ может быть представлена в виде *линейной* функции регрессии \tilde{y} по \tilde{X} , а именно:

$$\tilde{y}_i = \theta_0 + \theta_1 \tilde{x}_i^{(1)} + \dots + \theta_p \tilde{x}_i^{(p)} + \epsilon_i, \quad i = 1, 2, \dots, n;$$

эту часть исследования обычно называют **процедурой линеаризации модели**;

- в случае невозможности линеаризации модели приходится исследовать искомую регрессионную зависимость в терминах *исходных переменных*, а именно:

$$y_i = f(X_i; \Theta) + \varepsilon_i;$$

если спецификация регрессионных остатков ε_i соответствует условиям *классической* модели, то для вычисления МНК-оценок $\hat{\Theta}_{\text{МНК}}$ векторного параметра Θ решается оптимизационная задача вида

$$\hat{\Theta}_{\text{МНК}} = \arg \min_{\Theta} \sum_{i=1}^n (y_i - f(X_i; \Theta))^2;$$

о методах преодоления возникающих при этом вычислительных трудностей читатель может узнать, например, из книги [Айвазян С. А., Енюков И. С., Мешалкин Л. Д., 1985, гл. 9].

2.12.1. Некоторые виды нелинейных зависимостей, поддающиеся непосредственной линеаризации

Итак, пусть y и $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$ — *исходные* анализируемые переменные (соответственно результирующая и объясняющие), а ε — случайная остаточная компонента, участвующая в записи регрессионной зависимости, связывающей между собой y и X . За редким исключением вопросы линеаризации анализируемых связей решаются на основе рассмотрения *парных* зависимостей (и графически представляющих их парных корреляционных полей) типа $(x_i^{(j)}, y_i)$ и $(x_i^{(j)}, x_i^{(l)})$, $i = 1, 2, \dots, n$. Поэтому ниже будут представлены и проанализированы именно *парные* регрессионные зависимости, поддающиеся линеаризации.

Зависимости гиперболического типа

1) Предположим, что анализируемые переменные и случайные регрессионные остатки соответственно x , y и ε связаны между собой статистической зависимостью вида

$$y = \theta_0 + \theta_1 \frac{1}{x} + \varepsilon \quad (0 < x < \infty).$$

Соответствующая кривая регрессии $f(x; \Theta) = f(x; \theta_0, \theta_1) = \theta_0 + \theta_1/x$ (см. рис. 2.7) характеризуется двумя асимптотами (т. е. прямыми, к которым график функции неограниченно приближается, не достигая их) — горизонтальной ($y = \theta_0$) и вертикальной ($x = 0$). С помощью преобразова-

ния объясняющей переменной $\tilde{x} = 1/x$ (т.е. при переходе к новой объясняющей переменной \tilde{x}) эта зависимость приводится к линейному виду $y = \theta_0 + \theta_1 \tilde{x} + \varepsilon$. Соответственно при вычислении МНК-оценок параметров θ_0 и θ_1 второй столбец матрицы X должен быть сформирован из чисел $1/x_1, 1/x_2, \dots, 1/x_n$.

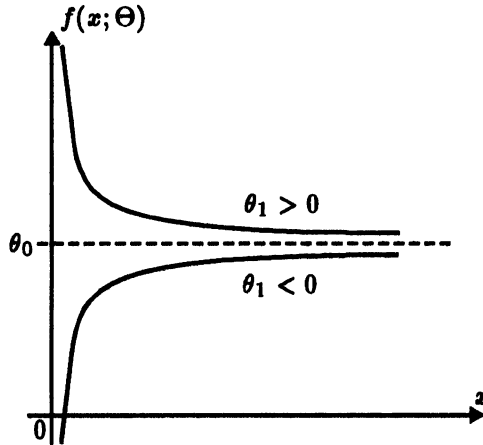


Рис. 2.7. График гиперболической зависимости вида $f(x; \Theta) = \theta_0 + \frac{\theta_1}{x}$

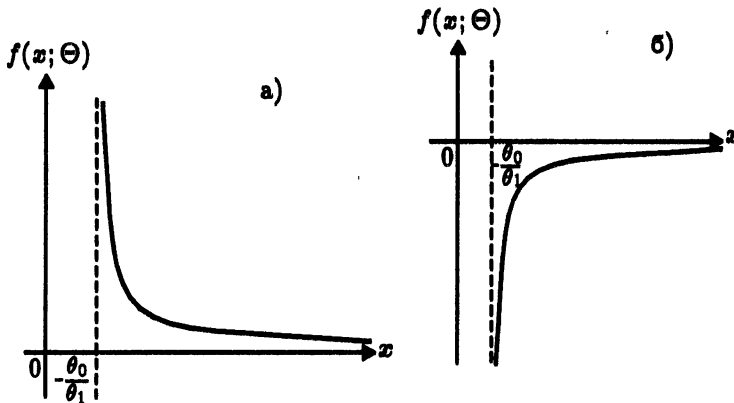


Рис. 2.8. График гиперболической зависимости вида $f(x; \Theta) = 1/(\theta_0 + \theta_1 x)$:
а) случай $\theta_0 < 0$; $\theta_1 > 0$ (для $x > -\theta_0/\theta_1$); б) случай $\theta_0 > 0$; $\theta_1 < 0$
(для $x > -\theta_0/\theta_1$)

2) Пусть переменные x , y и случайные регрессионные остатки ε связаны между собой статистической зависимостью вида

$$y = \frac{1}{\theta_0 + \theta_1 x + \varepsilon} \quad \left(-\frac{\theta_0}{\theta_1} < x < \infty \right)$$

(см. рис. 2.8). Очевидно, мы приходим к *линейной* модели $\tilde{y} = \theta_0 + \theta_1 x + \varepsilon$, если в качестве результирующего признака рассмотрим переменную $\tilde{y} = 1/y$. Следует не забыть только, что при вычислении МНК-оценок $\hat{\theta}_0$ и $\hat{\theta}_1$ надо использовать в качестве вектора наблюдаемых значений зависимой переменной вектор $\tilde{Y} = (1/y_1, 1/y_2, \dots, 1/y_n)^T$.

3) Если этап параметризации модели регрессии приводит нас к зависимости вида

$$y = \frac{x}{\theta_0 x + \theta_1 + x\varepsilon} \quad \left(-\frac{\theta_1}{\theta_0} < x < \infty \right)$$

(см. рис. 2.9), то линейаризацию исследуемой связи обеспечит переход к новым переменным $\tilde{y} = 1/y$ и $\tilde{x} = 1/x$. Легко видеть, что эти переменные будут связаны между собой зависимостью вида

$$\tilde{y} = \theta_0 + \theta_1 \tilde{x} + \varepsilon.$$

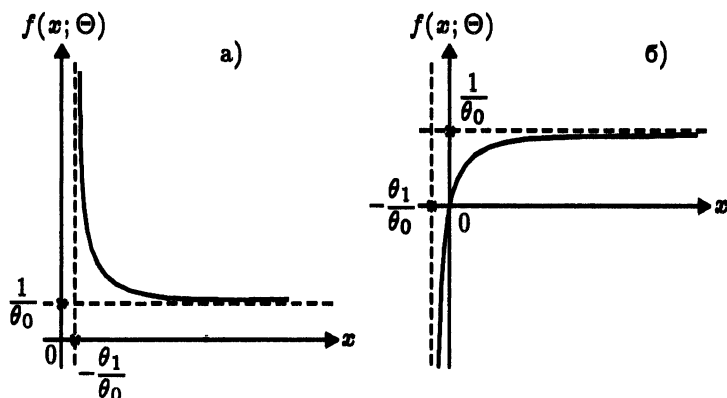


Рис. 2.9. График гиперболической зависимости вида $f(x; \Theta) = \frac{x}{\theta_0 x + \theta_1}$:
 а) случай $\theta_0 > 0$; $\theta_1 < 0$ (для $x > -\theta_1/\theta_0$); б) случай $\theta_0 > 0$; $\theta_1 > 0$ (для $x > -\theta_1/\theta_0$)

Очевидно, что матрицы \tilde{X} и \tilde{Y} , используемые в формулах метода наименьших квадратов при вычислении оценок $\hat{\theta}_0$ и $\hat{\theta}_1$, должны формиро-

ваться не из наблюдаемых значений, соответственно x_i и y_i , а из обратных к ним величин $\tilde{x}_i = 1/x_i$ и $\tilde{y}_i = 1/y_i$.

Заметим, что функции, изображенные на рис. 2.7 (вариант $\theta_1 < 0$) и 2.9 (вариант б)) используются в определенных ситуациях при построении так называемых *кривых Энгеля*, которые описывают зависимость спроса на определенный вид товаров или услуг (y) от уровня доходов (x) потребителей. При этом спрос определяется либо абсолютными либо относительными (по отношению к общим потребительским расходам) расходами на данный вид товаров или услуг. Функции, изображенные на рис. 2.7 (вариант $\theta_1 > 0$), 2.8. а) и 2.9. а) могут оказаться полезными при изучении спроса на товар (y) в зависимости от его цены (x).

Зависимости показательного (экспоненциального) типа

4) Достаточно широкий класс экономических показателей *характеризуется приблизительно постоянным темпом относительного прироста во времени*. Этому соответствует следующая форма зависимости этого показателя (y) от времени (x):

$$y = \theta_0 e^{\theta_1 x + \epsilon}.$$

Действительно, если пренебречь влиянием случайной остаточной компоненты ϵ (т. е. положить $\epsilon = 0$, см. рис. 2.10. а)), то непосредственные расчеты дают:

$$\frac{dy}{dx} = \theta_1 \theta_0 e^{\theta_1 x} = \theta_1 y,$$

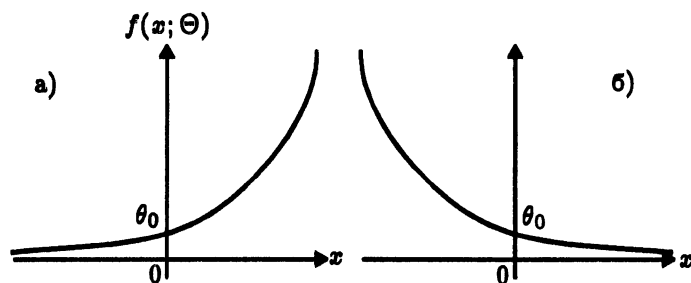


Рис. 2.10. График показательной (экспоненциальной) зависимости вида $f(x; \Theta) = \theta_0 e^{\theta_1 x}$: а) случай $\theta_1 > 0$; б) случай $\theta_1 < 0$

так что относительный прирост y за единицу времени (т. е. за единицу

«количества» x) определяется выражением

$$\frac{dy}{dx}/y = \theta_1 \text{ (в долях } y\text{)}.$$

Легко видеть, что переход к новой переменной $\tilde{y} = \ln y$ позволяет свести исследуемую зависимость к линейному виду:

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

где $\tilde{\theta}_0 = \ln \theta_0$. Располагая наблюдениями $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ и формируя вектор-столбец \tilde{Y} из $\ln y_1, \ln y_2, \dots, \ln y_n$, мы с помощью МНК можем построить оценки $\hat{\tilde{\theta}}_0$ и $\hat{\theta}_1$ параметров $\tilde{\theta}_0$ и θ_1 , а затем получить оценку $\hat{\theta}_0 = e^{\hat{\tilde{\theta}}_0}$ для параметра θ_0 исходного уравнения.

5) Если в результате параметризации модели мы пришли к необходимости исследовать экспоненциальную статистическую зависимость вида

$$y = \theta_0 e^{\frac{\theta_1}{x} + \varepsilon}$$

(см. рис. 2.11), то линейаризация искомой зависимости достигается с помощью следующих преобразований переменных: $\tilde{y} = \ln y$, $\tilde{x} = 1/x$. Очевидно, в терминах переменных (\tilde{x}, \tilde{y}) исследуемая зависимость будет иметь вид

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

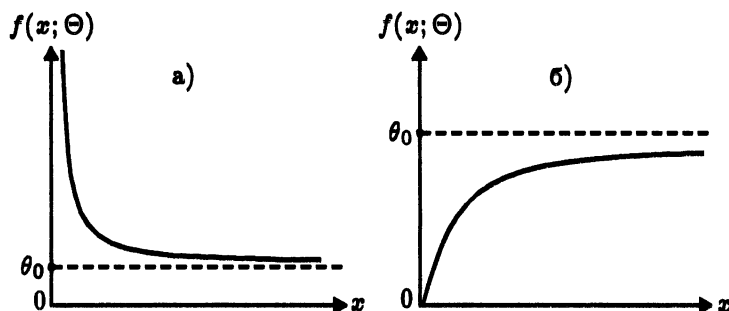


Рис. 2.11. График показательной (экспоненциальной) зависимости вида $f(x; \Theta) = \theta_0 e^{\theta_1/x}$: а) случай $\theta_1 > 0$; б) случай $\theta_1 < 0$

где $\tilde{\theta}_0 = \ln \theta_0$. Соответственно вектор-столбец \tilde{Y} и матрица \tilde{X} , участвующие в формулах МНК, определяются по исходным наблюдениям

$\{(x_i, y_i)\}_{i=1,2,\dots,n}$ следующим образом:

$$\tilde{Y} = (\ln y_1, \ln y_2, \dots, \ln y_n)^T; \quad \tilde{X} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1/x_1 & 1/x_2 & \dots & 1/x_n \end{pmatrix}^T.$$

6) Весьма гибкую форму параметризации искомой регрессионной зависимости представляет один из частных случаев так называемой логистической кривой (см. рис. 2.12)

$$y = \frac{1}{\theta_0 + \theta_1 e^{-x} + \varepsilon}, \quad (-\infty < x < +\infty).$$

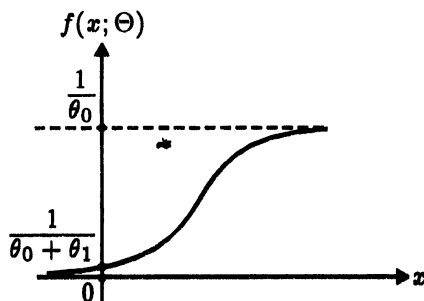


Рис. 2.12. График логистической кривой вида $f(x; \Theta) = 1/(\theta_0 + \theta_1 e^{-x})$. (Случай $\theta_1 > 0$)

Кривая $f(x; \Theta)$ имеет две горизонтальные асимптоты $y = 0$ и $y = 1/\theta_0$ и «точку перегиба» ($x_0 = \ln(\theta_1/\theta_0)$, $y_0 = 1/2\theta_0$). Линеаризация этой зависимости производится с помощью перехода к переменным $\hat{y} = 1/y$ и $\hat{x} = e^{-x}$. Соответственно, вектор-столбец \tilde{Y} и матрица \tilde{X} , участвующие в формулах МНК, определяются по исходным наблюдениям $\{(x_i, y_i)\}_{i=1,2,\dots,n}$ следующим образом:

$$\tilde{Y} = (1/y_1, 1/y_2, \dots, 1/y_n)^T; \quad \tilde{X} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ e^{-x_1} & e^{-x_2} & \dots & e^{-x_n} \end{pmatrix}.$$

Логистические кривые используются для описания поведения показателей, имеющих определенные «уровни насыщения», например, для описания зависимости спроса на товар (y) от дохода (x).

Зависимости степенного типа

7) Широко распространены в практике социально-экономических исследований так называемые *степенные зависимости*. Степенная модель

множественной регрессии имеет вид

$$y = \theta_0 (x^{(1)})^{\theta_1} (x^{(2)})^{\theta_2} \dots (x^{(p)})^{\theta_p} e^\varepsilon.$$

Очевидно, при переходе к переменным $\tilde{y} = \ln y$, $\tilde{x}^{(j)} = \ln x^{(j)}$ ($j = 1, 2, \dots, p$) можно представить эту зависимость в виде КЛММР, а именно:

$$\tilde{y} = \tilde{\theta}_0 + \theta_1 \tilde{x}^{(1)} + \dots + \theta_p \tilde{x}^{(p)} + \varepsilon,$$

где $\tilde{\theta}_0 = \ln \theta_0$. При оценке параметров $\tilde{\theta}_0, \theta_1, \dots, \theta_p$ участвующие в формулах МНК вектор-столбец \tilde{Y} и матрица \tilde{X} будут определяться по исходным наблюдениям $\{x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)}; y_i\}_{i=1,2,\dots,n}$ следующим образом:

$\tilde{Y} = (\ln y_1, \ln y_2, \dots, \ln y_n)^\top$, а $(j+1)$ -й столбец матрицы \tilde{X} есть $(\ln x_1^{(j)}, \ln x_2^{(j)}, \dots, \ln x_n^{(j)})^\top$, $j = 1, 2, \dots, p$ (первый столбец матрицы \tilde{X} , как обычно, составлен из одних единиц). Графики зависимостей данного типа для случая $p = 1$ представлены на рис. 2.13.

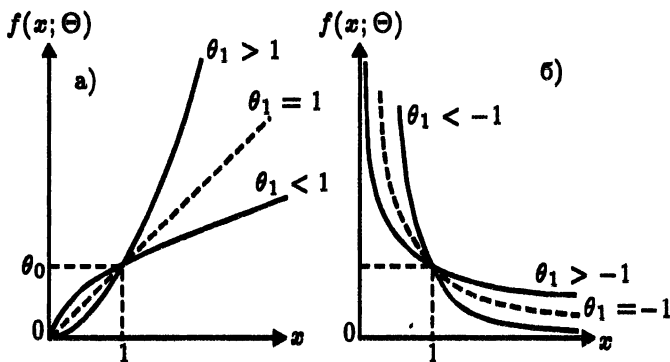


Рис. 2.13. График степенной зависимости вида $f(x; \Theta) = \theta_0 x^{\theta_1}$: а) случай $\theta_1 > 0$; б) случай $\theta_1 < 0$

Важную роль играют зависимости степенного типа в задачах построения и анализа *производственных функций* (y — объем произведенной продукции, $x^{(1)}, x^{(2)}, \dots$ — основные факторы производства: труд, капитал и т. д.). Достаточно часто используются степенные зависимости и при построении и анализе функций спроса (y — спрос на определенный вид товаров или услуг, $x^{(1)}$ — доход потребителя, $x^{(2)}, x^{(3)}, \dots$ — цены на данный и другие виды товаров).

Отметим, что при анализе степенных регрессионных зависимостей прозрачную содержательную интерпретацию получают коэффициенты

$\theta_1, \theta_2, \dots, \theta_p$, а именно: в соответствии с определением *коэффициента эластичности* признака y по объясняющей переменной $x^{(j)}$ (см. п. 2.9) величина $\theta_j = \partial \ln f(X; \Theta) / \partial \ln x^{(j)}$ есть не что иное, как коэффициент эластичности анализируемого результирующего показателя по j -й объясняющей переменной. Можно, кстати, показать, что если эластичность y по каждой из объясняющих переменных $x^{(j)}$ *постоянна* (т. е. не зависит от того, при каких именно значениях объясняющих переменных она вычисляется), то y и X могут быть связаны только зависимостью степенного типа.

Зависимости логарифмического типа

8) На рис. 2.14 представлены графики зависимостей логарифмического типа:

$$y = \theta_0 + \theta_1 \ln x + \varepsilon \quad (0 < x < \infty).$$

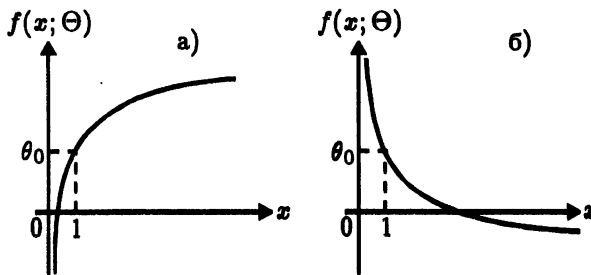


Рис. 2.14. График логарифмической зависимости вида $f(x; \Theta) = \theta_0 + \theta_1 \ln x$:
а) случай $\theta_1 > 0$; б) случай $\theta_1 < 0$

Кривые на рис. 2.14 проходят через точку $(1; \theta_0)$ и имеют в качестве вертикальной асимптоты ось y (т. е. прямую $x = 0$). Переход к линейному виду зависимости осуществляется с помощью логарифмического преобразования объясняющей переменной: $\tilde{x} = \ln x$. Соответственно, второй столбец матрицы \tilde{X} , участвующей в формулах МНК, будет иметь вид: $(\ln x_1, \ln x_2, \dots, \ln x_n)^T$.

2.12.2. Подбор линеаризующего преобразования (подход Бокса-Кокса)

В предыдущем пункте описан набор зависимостей, поддающихся линеаризации с помощью подходящих преобразований анализируемых переменных. Но решение вопроса о том, к какому именно из перечисленных *линеаризуемых* типов зависимостей следует отнести наш конкретный

случай, является задачей не простой. Можно, конечно, действовать методом «проб и ошибок»: последовательно построить по имеющимся у нас исходным статистическим данным (2.4) каждую из альтернативного набора линеаризуемых моделей, а затем выбрать из них наилучшую в смысле какого-то «критерия качества» (например, по максимальному значению подправленной на несмещенность оценки коэффициента детерминации \hat{R}^{*2} , см. формулу (11.40)).

Английские статистики Г. Бокс и Д. Кокс предложили более формализованную процедуру подбора линеаризующего преобразования¹. Их метод основан на предположении, что искомое преобразование принадлежит определенному *однопараметрическому семейству* преобразований вида

$$\tilde{y}_i(\lambda) = \frac{y_i^\lambda - 1}{\lambda}, \quad \tilde{x}_i^{(j)}(\lambda) = \frac{(x_i^{(j)})^\lambda - 1}{\lambda}, \quad i = 1, 2, \dots, n. \quad (2.180)$$

Точнее их гипотезу можно сформулировать следующим образом: *существует такое вещественное (положительное или отрицательное) число λ^* , что один из двух нижеследующих вариантов представления искомой регрессионной зависимости между наблюдаемыми переменными y и $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})$:*

$$\tilde{y}_i(\lambda^*) = \theta_0 + \theta_1 \tilde{x}_i^{(1)}(\lambda^*) + \dots + \theta_p \tilde{x}_i^{(p)}(\lambda^*) + \varepsilon_i \quad (2.181)$$

или

$$\tilde{y}_i(\lambda^*) = \theta_0 + \theta_1 x_i^{(1)} + \dots + \theta_p x_i^{(p)} + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (2.181')$$

будет удовлетворять всем требованиям нормальной классической линейной модели множественной регрессии (см. (2.5) и п. 2.3.2).

З а м е ч а н и е 1. Преобразования вида (2.180) применяются обычно к переменным, принимающим только *положительные* значения. Поэтому, если это не так, то вначале подбирают «сдвиговые» константы $c^{(0)}, c^{(1)}, \dots, c^{(p)}$, которые обеспечивают положительность значений $y_i + c^{(0)}$ и $x_i^{(j)} + c^{(j)}$ ($j = 1, 2, \dots, p$), а затем к сдвинутым значениям переменных применяют данное преобразование, т. е.:

$$\tilde{y}_i(\lambda) = \frac{(y_i + c^{(0)})^\lambda - 1}{\lambda}, \quad \tilde{x}_i(\lambda) = \frac{(x_i^{(j)} + c^{(j)})^\lambda - 1}{\lambda} \quad (i = 1, 2, \dots, n) \quad (2.180')$$

¹ См.: Box G.E.P. and Cox D.R. An Analysis of Transformations. «Journ. of Royal Statist. Soc.» , Series B, vol. 26 (1964), pp. 211–243.

З а м е ч а н и е 2. Семейство степенных преобразований вида (2.180) (или (2.180')) весьма широко и гибко. При $\lambda = 1$ модели (2.181) и (2.181') являются линейными относительно y_i и $x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)}$. При $\lambda = 0$ мы имеем степенную зависимость между y и X вида 7) (см. п. 2.12.1), поскольку $\tilde{y}_i(0) = \lim_{\lambda \rightarrow 0} (y_i^\lambda - 1)/\lambda = \ln y_i$ и $\tilde{x}_i^{(j)}(0) = \lim_{\lambda \rightarrow 0} [(x_i^{(j)})^\lambda - 1]/\lambda = \ln x_i^{(j)}$. При других значениях λ уравнения (2.181) и (2.181') будут связывать между собой какие-то степени исходных переменных.

Оценка неизвестного значения параметра λ . Таким образом, если исходить из справедливости сформулированной выше гипотезы, подбор линеаризующего преобразования анализируемых переменных сводится к оценке параметра λ в формулах (2.180) или (2.180') по имеющимся в нашем распоряжении исходным статистическим данным (2.4). Эта проблема решается с помощью метода максимального правдоподобия. Будем исходить для определенности из справедливости нашего допущения применительно к представлению искомой модели в форме (2.181'), т. е. в матричной записи при неизвестном значении параметра λ $\tilde{Y}(\lambda)$ и X связаны между собой уравнением

$$\tilde{Y}(\lambda) = X\Theta + \varepsilon, \quad (2.181'')$$

где $\tilde{Y}(\lambda) = (\tilde{y}_1(\lambda), \tilde{y}_2(\lambda), \dots, \tilde{y}_n(\lambda))^T$, $\tilde{y}_i(\lambda) = (y_i^\lambda - 1)/\lambda$ (мы предполагаем, что все y_i положительны), X — матрица размерности $n \times (p + 1)$ ранга $p + 1$ из наблюдаемых значений объясняющих переменных (см. (2.4^a)), а $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$ — вектор-столбец $(0, \sigma^2)$ -нормально распределенных и взаимнонезависимых регрессионных случайных остатков.

Для составления уравнений метода максимального правдоподобия относительно неизвестных параметров λ , Θ и σ^2 при заданных значениях Y и X выпишем вначале функцию правдоподобия $L(\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_n | X; \lambda, \Theta, \sigma^2)$ для преобразованных значений $\tilde{y}_i(\lambda)$, а затем, воспользовавшись правилом вычисления закона распределения вероятностей случайных величин, являющихся заданными функциями от известных случайных величин (см. п. 4.4, формулу (4.11)), определим нужную нам функцию правдоподобия $L(y_1, y_2, \dots, y_n | X; \lambda, \Theta, \sigma^2)$ для непосредственно наблюдаемых значений y_i .

Опираясь на результаты п. 2.3.2 (см. соотношения (2.24)–(2.27)) и с учетом взаимной независимости и $(\theta_0 + \theta_1 x_i^{(1)} + \dots + \theta_p x_i^{(p)}, \sigma^2)$ -нормальной

распределенности случайных величин y_i ($i = 1, 2, \dots, n$) имеем:

$$L(\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_n | X; \lambda, \Theta, \sigma^2) = \frac{1}{(2\pi)^{\frac{n}{2}} (\sigma^2)^{\frac{n}{2}}} \times \exp \left\{ -\frac{1}{2\sigma^2} (\tilde{Y} - X\Theta)^T (\tilde{Y} - X\Theta) \right\}. \quad (2.182)$$

Таблица 2.7

№ п/п	Смысл понятия или характеристики, используемой в формуле (4.11)	Обозначения формулы (4.11)	Обозначения, принятые в данном пункте
1	Размерность анализируемых случайных величин	p	n
2	Случайная величина, распределение которой задано	$\xi = (\xi^{(1)}, \dots, \xi^{(p)})$	$\tilde{Y}(\lambda) = (\tilde{y}_1(\lambda), \dots, \tilde{y}_n(\lambda))^T$
3	Случайная величина, распределение которой надо вычислить	$\eta = (\eta^{(1)}, \dots, \eta^{(p)})$	$Y = (y_1, \dots, y_n)^T$
4	Преобразование, связывающее исследуемые случайные величины	$\eta = g(\xi)$	$Y = g(\tilde{Y}(\lambda)),$ где $y_i = (\lambda \tilde{y}_i + 1)^{1/\lambda}$
5	Обратное преобразование, связывающее исследуемые величины	$\xi = g^{-1}(\eta) = (g_1^{-1}(\eta), \dots, g_n^{-1}(\eta))$	$\tilde{Y}(\lambda) = g^{-1}(Y) = (g_1^{-1}(Y), \dots, g_n^{-1}(Y)),$ где $g_i^{-1}(Y) = \tilde{y}_i = (y_i^\lambda - 1)/\lambda$
6	Определитель матрицы преобразования (якобиан)	$J = \left \det \left(\frac{\partial g_i^{-1}(Y)}{\partial y^{(l)}} \right) \right ,$ $i, l = 1, 2, \dots, p$	$J(\lambda) = \det(j_{il}) ,$ где $j_{il} = \begin{cases} y_i^{\lambda-1} & \text{при } i = l; \\ 0 & \text{при } i \neq l, \end{cases}$ $i, l = 1, 2, \dots, n$

Воспользуемся далее формулой (4.11), позволяющей перейти от известного распределения (2.182) многомерной случайной величины $\tilde{Y} = (\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_n)^T$ к распределению случайной величины Y , являющейся некоторой (заданной) функцией от \tilde{Y} . Нижеследующая таблица, устанавливающая соответствие между обозначениями формулы (4.11) и обозначениями, принятыми в данном пункте, облегчит читателю понимание этого перехода.

Итак, в соответствии с формулой (4.11) имеем:

$$L(y_1, y_2, \dots, y_n | X; \lambda, \Theta, \sigma^2) = \frac{J(\lambda)}{(2\pi)^{\frac{n}{2}} (\sigma^2)^{\frac{n}{2}}} \exp \left\{ -\frac{1}{2\sigma^2} (\tilde{Y} - X\Theta)^T (\tilde{Y} - X\Theta) \right\}$$

или (в терминах *логарифмической* функции правдоподобия $l = \ln L$)

$$l(y_1, y_2, \dots, y_n | X; \lambda, \Theta, \sigma^2) = \text{const} + \ln J(\lambda) - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} (\tilde{Y} - X\Theta)^T (\tilde{Y} - X\Theta), \quad (2.183)$$

где $J(\lambda) = \left(\prod_{i=1}^n y_i \right)^{\lambda-1} > 0$ (так как все $y_i > 0$), а const — некоторая постоянная величина, не зависящая от оцениваемых параметров λ , Θ и σ^2 .

Предположим, что значение параметра λ *закреплено*. Тогда дифференцирование (2.183) по Θ и σ^2 и приравнивание полученных частных производных к нулю (см. (2.25)–(2.27)) дает:

$$\hat{\Theta}(\lambda) = (X^T X)^{-1} X^T \tilde{Y}(\lambda), \quad (2.184)$$

$$\hat{\sigma}^2(\lambda) = \frac{1}{n} (\tilde{Y}(\lambda) - X\hat{\Theta}(\lambda))^T (\tilde{Y}(\lambda) - X\hat{\Theta}(\lambda)). \quad (2.185)$$

Для того, чтобы подобрать теперь оптимальное значение параметра λ , вернемся к соотношению (2.183), подставив в него оптимальные выражения (2.184) и (2.185) соответственно для $\Theta(\lambda)$ и $\sigma^2(\lambda)$. Обозначим полученное при этом значение l с помощью $l_{\max}(\lambda)$. Итак:

$$l_{\max}(\lambda) = l(y_1, \dots, y_n | X; \lambda, \hat{\Theta}, \hat{\sigma}^2) = \text{const} + (\lambda - 1) \sum_{i=1}^n \ln y_i - \frac{n}{2} \ln \hat{\sigma}^2(\lambda) \quad (2.186)$$

(при выводе (2.186) использован тот факт, что при оптимальных выражениях для $\hat{\Theta}(\lambda)$ и $\hat{\sigma}^2(\lambda)$ последний член в правой части (2.183) равен $n/2$, т. е. не зависит от λ).

Далее анализируется функция $l_{\max}(\lambda)$ и отыскивается такое значение λ^* , при котором $l_{\max}(\lambda^*) = \max_{\lambda} l_{\max}(\lambda)$. С этой целью определяется априорный диапазон $(\lambda_{\min}, \lambda_{\max})$ возможных значений λ (обычно достаточно рассмотреть в качестве области возможных значений λ отрезок от $\lambda_{\min} = -1$ до $\lambda_{\max} = 2$), на этом диапазоне выбирается сетка («решето») значений $\lambda_i = \lambda_{\min} + i(\lambda_{\max} - \lambda_{\min})/N$, $i = 0, 1, \dots, N$ и для каждого такого значения λ_i последовательно вычисляются $\hat{\Theta}(\lambda_i)$, $\hat{\sigma}^2(\lambda_i)$ и $l_{\max}(\lambda_i)$. То значение λ^* , при котором

$$l_{\max}(\lambda^*) = \max_{\lambda = \lambda_0, \lambda_1, \dots, \lambda_N} l_{\max}(\lambda_i),$$

и будет определять искомое линейризующее преобразование (2.180). Оценки λ^* , $\hat{\theta}(\lambda^*)$ и $\hat{\sigma}^2(\lambda^*)$ являются оценками метода максимального правдоподобия, а процедуру их поиска часто называют «решетчатой».

З а м е ч а н и е 3. Оценка параметра λ в случае, если преобразования (2.180) применяются одновременно к результирующей и к объясняющим переменным, производится тем же способом с единственным видоизменением процедуры: в формулах (2.181''), (2.192) ~ (2.186) матрицу X следует заменить на матрицу $\tilde{X}(\lambda)$ наблюдаемых значений преобразованных объясняющих переменных.

П р и м е р 2.6 (заимствован из [А. Зельнер, с. 182–184]). Изложенный выше подход к подбору линейризующего преобразования с использованием метода максимального правдоподобия был применен в приложении к анализу функции спроса на деньги. Результаты предварительного анализа показали, что функция спроса на деньги может быть записана в виде

$$\frac{y_i^\lambda - 1}{\lambda} = \theta_0 + \theta_1 \frac{(x_i^{(1)})^\lambda - 1}{\lambda} + \theta_2 \frac{(x_i^{(2)})^\lambda - 1}{\lambda} + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (2.187)$$

где индекс i обозначает, что значение, переменной относится к году i , y_i — денежная наличность, включая текущие и срочные депозиты (дефлятированные индексом цен), $x_i^{(1)}$ — измеренный доход (дефлятированный индексом цен), $x_i^{(2)}$ — средняя норма процента по коммерческим бумагам и ε_i — остаточная случайная компонента, удовлетворяющая всем требованиям нормальной КЛММР (исходные статистические данные $(x_i^{(1)}, x_i^{(2)}, y_i)$ представляют собой годовые наблюдения по экономике США за 1869–1963 гг., причем $i = 1$ соответствует 1869 году, $i = 95$ соответствует 1963 году, а общее число наблюдений $n = 95$).

При реализации «решетчатой» процедуры был определен диапазон возможных значений λ от $\lambda_{\min} = -0,90$ до $\lambda_{\max} = 1,30$ и «шаг», равный $\Delta = (\lambda_{\max} - \lambda_{\min})/N = 2,20/22 = 0,1$. Затем для каждого значения $\lambda_j = -0,90 + j \cdot 0,1$ ($j = 0, 1, \dots, 22$) были подсчитаны:

$$\tilde{Y}(\lambda_j) = (\tilde{y}_1(\lambda_j), \dots, \tilde{y}_n(\lambda_j)), \quad \text{где } \tilde{y}_i(\lambda_j) = \frac{y_i^{\lambda_j} - 1}{\lambda_j};$$

$$\tilde{X}(\lambda_j) = \begin{pmatrix} \tilde{x}_1^{(1)}(\lambda_j) & \tilde{x}_2^{(1)}(\lambda_j) & \dots & \tilde{x}_n^{(1)}(\lambda_j) \\ \tilde{x}_1^{(2)}(\lambda_j) & \tilde{x}_2^{(2)}(\lambda_j) & \dots & \tilde{x}_n^{(2)}(\lambda_j) \end{pmatrix}^T,$$

$$\text{где } \tilde{x}_i^{(1)}(\lambda_j) = \frac{(x_i^{(1)})^{\lambda_j} - 1}{\lambda_j} \quad \text{и} \quad \tilde{x}_i^{(2)}(\lambda_j) = \frac{(x_i^{(2)})^{\lambda_j} - 1}{\lambda_j},$$

а также значения $\hat{\Theta}(\lambda_j)$ и $\hat{\sigma}^2(\lambda_j)$, соответственно, по формулам (2.184) и (2.185) с заменой матрицы X матрицей $\tilde{X}(\lambda)$; и, наконец, величина $\tilde{l}_{\max}(\lambda_j) = (\lambda_j - 1) \sum_{i=1}^n \ln y_i - \frac{n}{2} \ln \hat{\sigma}^2(\lambda_j)$ по формуле (2.186) (с исключением константы).

График функции $\tilde{l}_{\max}(\lambda)$ изображен на рис. 2.15. Он позволяет вычислить $\lambda^* \approx 0,20$, т.е. такое значение λ , при котором функция $\tilde{l}_{\max}(\lambda)$ достигает своего максимума. Стандартный регрессионный анализ регрессии $\tilde{y}(0,2)$ по $\tilde{x}^{(1)}(0,2)$ и $\tilde{x}^{(2)}(0,2)$ дает:

$$\tilde{y}_i(0,2) = -1,055 + 1,112 \tilde{x}_i^{(1)}(0,2) - 0,097 \tilde{x}_i^{(2)}(0,2) + \varepsilon_i; \quad \hat{R}^{*2} = 0,72$$

(0,239) (0,016) (0,016)

(напомним, что в скобках под значениями оценок коэффициентов регрессии θ_0 , θ_1 и θ_2 указаны величины среднеквадратических ошибок этих оценок).

Следовательно, отправляясь от (2.187) (при $\lambda = 0,20$), можно показать, что функция регрессии спроса на деньги (y) по доходу ($x^{(1)}$) и средней норме процента ($x^{(2)}$) в исходных переменных будет иметь вид:

$$E(y | x^{(1)}, x^{(2)}) = \left(-0,226 + 1,112 \sqrt[5]{x^{(1)}} - 0,097 \sqrt[5]{x^{(2)}} \right)^5. \quad (2.188)$$

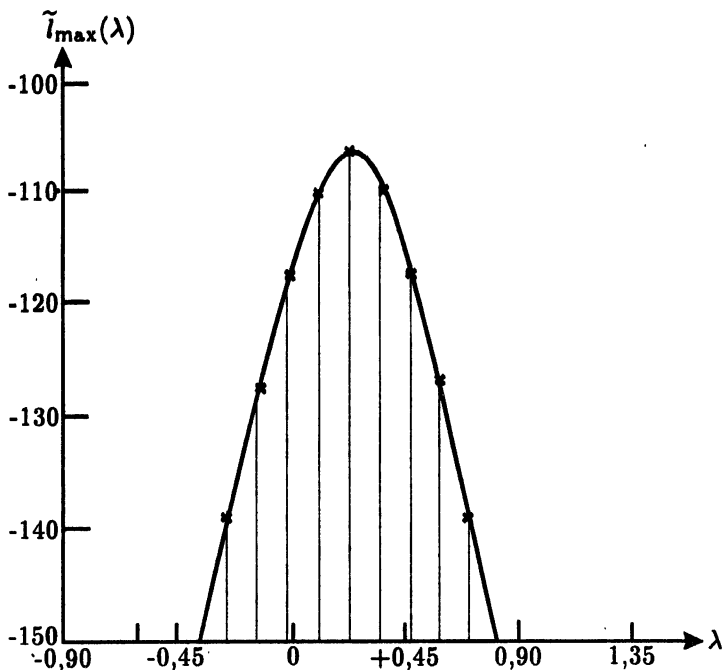


Рис. 2.15. «Решето» значения логарифмической функции правдоподобия (без учета величины константы)

Заметим, что близость к нулю оптимального значения λ ($\lambda^* = 0,2$) говорит о том, что и «логарифмический» вариант функции спроса на деньги (т. е. переход к переменным $\tilde{y} = \ln y$, $\tilde{x}^{(1)} = \ln x^{(1)}$ и $\tilde{x}^{(2)} = \ln x^{(2)}$) находится в приближенном согласии с имеющимися выборочными данными. Так что наряду с функцией (2.188) можно было бы рассчитать и конкурирующую функцию регрессии вида

$$E(y | x^{(1)}, x^{(2)}) = \theta_0 (x^{(1)})^{\theta_1} (x^{(2)})^{\theta_2}.$$

2.13. Дихотомические (бинарные) результирующие показатели и связанные с ними логит- и пробит-модели

Представим себе, что результирующий показатель y , «поведение» которого существенно зависит от количественных объясняющих перемен-

ных $\tilde{X} = (1, x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$, является *качественной* переменной, определяющей *одно из двух* возможных состояний характеризуемого ею объекта. Так, например, объясняющие переменные \tilde{X}_i могут описывать различные социально-демографо-экономические характеристики i -го индивидуума (его доход, возраст, стаж работы и т. п.), а результирующему показателю y_i приписывается значение единица, если i -й индивидуум 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 *диготомические (бинарные) признаки*, т. е. его компоненты y_i могут принимать только два значения: «0» или «1». Можно ли построить линейную регрессионную модель, описывающую зависимость y от \tilde{X} в данном случае?

Очевидно, непосредственное распространение описанной в предыдущих пунктах этой главы методики построения КЛММР или ОЛММР, вряд ли, пригодно: неясно, как интерпретировать в этом случае регрессионные значения $\hat{\Theta}^T \tilde{X}$, измеренные в *непрерывной количественной шкале*.

Поэтому для исследования статистической связи между y и \tilde{X} строят некоторую специальную регрессионную модель зависимости *вероятности* $P\{y = 1 \mid \tilde{X}\}$ от линейной формы $\Theta^T \tilde{X}$. Но описывать вероятность *непосредственно линейной функцией* $\Theta^T \tilde{X}$, 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 $P\{y = 1 \mid \tilde{X}\}$ подбирают функции, область значений которых определяется отрезком $[0, 1]$, а линейная форма $\Theta^T \tilde{X}$ играет роль аргумента этой функции, т. е.

$$P\{y = 1 \mid \tilde{X}\} = F(\Theta^T \tilde{X}), \quad (2.189)$$

причем функции $F(z)$ должны удовлетворять следующим требованиям:

$$\left. \begin{array}{l} \text{(а)} \quad F(z) \text{ монотонно возрастает по } z; \\ \text{(б)} \quad 0 \leq F(z) \leq 1; \\ \text{(в)} \quad F(z) \rightarrow 1 \text{ при } z \rightarrow \infty; \\ \text{(г)} \quad F(z) \rightarrow 0 \text{ при } z \rightarrow -\infty. \end{array} \right\} \quad (2.190)$$

График одной из функций данного вида приведен на рис. 2.16.

Модели типа (2.189) при ограничениях (2.190) называют обычно *моделями бинарного выбора*. Наиболее распространенными моделями бинарного выбора являются так называемые логит- и пробит-модели.

Логит-модель. Модель типа (2.189)–(2.190) называется логит-моделью, если в качестве $F(y)$ рассматривается *логистическая* функция, т. е.

$$P\{y_i = 1 \mid \tilde{X}_i\} = \frac{e^{\Theta^T \tilde{X}_i}}{1 + e^{\Theta^T \tilde{X}_i}}. \quad (2.189^a)$$

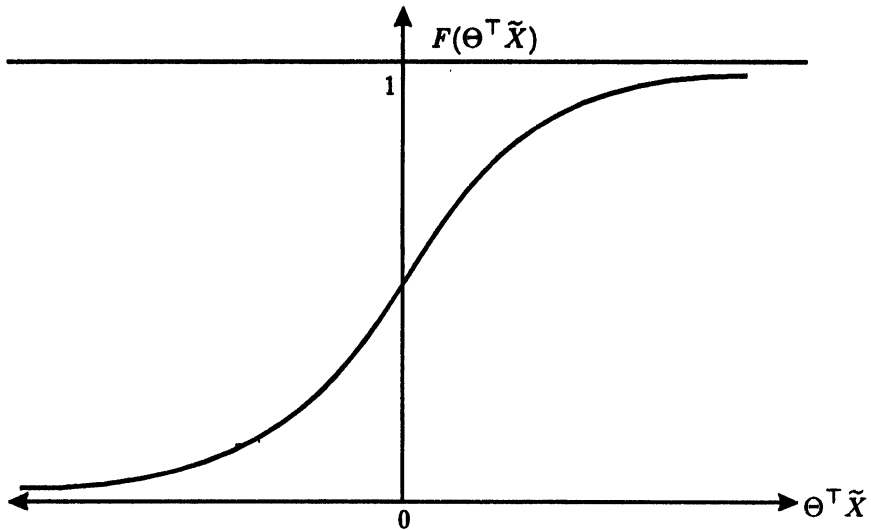


Рис. 2.16. График функции $F(\Theta^T \tilde{X})$, используемой в модели бинарного выбора

Нетрудно проверить, что эта функция удовлетворяет условиям (2.190). Непосредственной проверкой можно убедиться также, что эта функция симметрична относительно точки $\Theta^T \tilde{X} = 0$, т. е. $\Lambda(-z) = 1 - \Lambda(z)$, где

$$\Lambda(z) = \frac{e^z}{1 + e^z}. \quad (2.191)$$

Пробит-модель. Модель типа (2.189)–(2.190) называется пробит-моделью, если в качестве $F(z)$ рассматривается *стандартная нормальная вероятностная функция распределения* $\Phi(z)$, т. е.

$$P\{y_i = 1 \mid \tilde{X}_i\} = \Phi(\Theta^T \tilde{X}_i), \quad (2.189^b)$$

где $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-\frac{x^2}{2}} dx$.

С этой целью на первой итерации определяют оценку $\hat{\Theta}^{(1)}$ по *обычному* (а не взвешенному) методу наименьших квадратов, т. е. находят такое $\hat{\Theta}^{(1)}$, при котором

$$\sum_{j=1}^N \left(\hat{p}_j - \frac{e^{\hat{\Theta}^{(1)\top} \tilde{X}_j}}{1 + e^{\hat{\Theta}^{(1)\top} \tilde{X}_j}} \right)^2 = \min_{\Theta} \sum_{j=1}^N \left(\hat{p}_j - \frac{e^{\Theta^\top \tilde{X}_j}}{1 + e^{\Theta^\top \tilde{X}_j}} \right)^2.$$

Затем подставляют найденное значение $\hat{\Theta}^{(1)}$ в выражение «весов» $[\Lambda(\Theta^\top \tilde{X}_j) (1 - \Lambda(\Theta^\top \tilde{X}_j))]^{-1}$ в формуле (2.194) и находят следующие приближение $\hat{\Theta}^{(2)}$ для Θ , минимизируя (2.194), и т. д.

Аналогичная процедура оценивания параметров Θ в *пробит*-модели производится точно так же с заменой функции $\Lambda(z)$ на функцию $\Phi(z)$, определенную в (2.189^б).

Можно, однако, несколько упростить эту процедуру оценивания, применив к обеим частям (2.193) преобразование F^{-1} и разложив полученный результат в ряд Тейлора в окрестности точки $\varepsilon = 0$ (ограничившись при этом членом линейного порядка):

$$F^{-1}(\hat{p}_j) = F^{-1}(F(\Theta^\top \tilde{X}_j) + \varepsilon_j) \approx F^{-1}(F(\Theta^\top \tilde{X}_j)) + \left[\frac{dF^{-1}(F_j)}{dF_j} \right] \varepsilon_j. \quad (2.195)$$

Но $F^{-1}(F(\Theta^\top \tilde{X}_j)) = \Theta^\top \tilde{X}_j$, и

$$\frac{dF^{-1}(F_j)}{dF_j} = \frac{1}{dF(F_j)/dF_j} = \frac{1}{f(F_j)} = \frac{1}{f(\Theta^\top \tilde{X}_j)} = \frac{1}{f_j},$$

где $f_j = f(\Theta^\top \tilde{X}_j)$ — значение функции плотности з.р.в. F в точке $\Theta^\top \tilde{X}_j$. Поэтому, возвращаясь к (2.195), имеем:

$$\tilde{y}_j \approx \hat{\Theta}^\top \tilde{X}_j + \tilde{\varepsilon}_j, \quad (2.195')$$

где $\tilde{y}_j = F^{-1}(\hat{p}_j)$ — квантиль уровня \hat{p}_j функции распределения F , $\tilde{\varepsilon}_j = \varepsilon_j/f_j$, причем $E\tilde{\varepsilon}_j = 0$ и $D\tilde{\varepsilon}_j = F_j(1 - F_j)/n_j f_j^2$.

Таким образом, задача оценки параметра Θ в логит- и пробит-моделях сводится к анализу *линейной* функции регрессии (2.195') с гетероскедастичными остатками $\tilde{\varepsilon}_j$. При этом в качестве компонент вектора $\tilde{Y} = (\tilde{y}_1, \tilde{y}_1, \dots, \tilde{y}_N)^\top$ наблюдаемых значений зависимой переменной будут использоваться:

- в логит-модели: $\tilde{y}_j = \ln[\hat{p}_j/(1 - \hat{p}_j)]$, так как непосредственно проверяется, что именно \tilde{y}_j является решением уравнения $e^z/(1 + e^z) = \hat{p}_j$ относительно z , т. е. тот факт, что \tilde{y}_j — квантиль уровня \hat{p}_j распределения $\Lambda(z)$;
- в пробит-модели: \tilde{y}_j — табличное значение квантиля уровня \hat{p}_j стандартного нормального распределения.

Как известно (см. пп. 2.6.1 и 2.7), оценка параметров линейной регрессионной модели с гетероскедастичными остатками $\tilde{\epsilon}_j$ производится с помощью взвешенного МНК, в котором в качестве весов w_j используются величины, обратные $D\tilde{\epsilon}_j$. Соответственно для оценки параметров Θ в логит-модели мы должны решить оптимизационную задачу вида

$$\sum_{j=1}^N w_j^{(n)} \left(\ln \frac{\hat{p}_j}{1 - \hat{p}_j} - \Theta^\top \tilde{X}_j \right)^2 \rightarrow \min_{\Theta}, \quad (2.196)$$

где $w_j^{(n)} = n_j \Lambda(\Theta^\top \tilde{X}_j)(1 - \Lambda(\Theta^\top \tilde{X}_j))$.

Данное выражение для весов $w_j^{(n)}$ получено с учетом справедливости непосредственно проверяемого тождества:

$$\begin{aligned} \frac{1}{D\tilde{\epsilon}_j} &= \frac{n_j f_j^2}{F_j(1 - F_j)} = \frac{n_j \left(\frac{d\Lambda(\Theta^\top \tilde{X}_j)}{d(\Theta^\top \tilde{X}_j)} \right)^2}{\Lambda(\Theta^\top \tilde{X}_j)(1 - \Lambda(\Theta^\top \tilde{X}_j))} = \frac{n_j \Lambda_j^2 (1 - \Lambda_j)^2}{\Lambda_j(1 - \Lambda_j)} \\ &= n_j \Lambda_j(1 - \Lambda_j), \quad \text{где } \Lambda_j = \Lambda(\Theta^\top \tilde{X}_j). \end{aligned}$$

Оценка параметров Θ в пробит-модели сводится к решению оптимизационной задачи вида

$$\sum_{j=1}^N w_j^{(n)} (\tilde{y}_j - \Theta^\top \tilde{X}_j)^2 \rightarrow \min_{\Theta}, \quad (2.196')$$

где \tilde{y}_j — табличные значения квантилей уровня \hat{p}_j стандартного нормального распределения, «веса» $w_j^{(n)} = n_j \varphi^2(\Theta^\top \tilde{X}_j)/\Phi(\Theta^\top \tilde{X}_j)(1 - \Phi(\Theta^\top \tilde{X}_j))$, а $\varphi(z)$ и $\Phi(z)$ — значения функций соответственно плотности и распределения стандартного нормального закона в точке z .

Решение оптимизационных задач (2.196) и (2.196') осложнено тем обстоятельством, что «веса» $w_j^{(n)}(\Theta)$ и $w_j^{(n)}(\Theta)$ зависят от оцениваемого параметра Θ . Однако эти вычислительные трудности обходятся с помощью стандартной итерационной процедуры: на первом шаге подсчитываются значения $\hat{\Theta}^{(1)}$, являющиеся решением задачи (2.196) (или задачи

(2.196') при $w_j \equiv 1$. Полученные значения $\hat{\Theta}^{(1)}$ используются для подсчета весов $w_j^{(n)}(\hat{\Theta}^{(1)})$ (или $w_j^{(n)}(\hat{\Theta}^{(1)})$), после чего анализируемые оптимизационные задачи решаются уже с учетом этих весов. В результате получают значения $\hat{\Theta}^{(2)}$ и т. д.

Более подробное освещение регрессионных моделей бинарного выбора читатель может найти, например, в книге [Greene W., Ch. 21].

ВЫВОДЫ

1. Методы и модели регрессионного анализа занимают центральное место во всем математико-статистическом инструментарии эконометрики. Постулируя возможность восстановления (с некоторой случайной ошибкой ε) значения анализируемого *результатирующего показателя* y по заданным значениям так называемых *объясняющих переменных* $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$, эконометрист по-существу рассматривает модель вида

$$y = f(X) + \varepsilon.$$

Основная цель регрессионного анализа заключается в том, чтобы на базе имеющихся наблюдений анализируемых переменных $\{(x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)}; y_i)\}_{i=1,2,\dots,n}$ построить для заданных значений объясняющих переменных $X_0 = (x_0^{(1)}, x_0^{(2)}, \dots, x_0^{(p)})^T$, — наилучшие в определенном смысле точечные и интервальные оценки для неизвестных значений $f(X_0)$ и $y(X_0)$.

2. Следуя от простого к сложному, эконометрист рассматривает в первую очередь *линейные модели множественной регрессии* (ЛММР), т. е. модели, в которых функция регрессии $f(X)$ представима в виде

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

В зависимости от принятой в модели спецификации случайных ошибок (регрессионных остатков) ε ЛММР называется *классической* (при этом регрессионные остатки $\varepsilon_i = y_i - f(X_i)$, $i = 1, 2, \dots, n$, полагаются *взаимно некоррелированными и гомоскедастичными*), или *обобщенной* (при отказе хотя бы одного из вышеупомянутых свойств). Если в дополнение к упомянутым свойствам регрессионных остатков ε_i классической ЛММР постулируется их *нормальность*, то ЛММР называется *нормальной классической*.

3. В рамках *классической* ЛММР обычный метод наименьших квадратов (МНК) позволяет получить *состоятельные, несмещенные и в*

определенном смысле наиболее эффективные оценки для неизвестных параметров $\Theta = (\theta_0, \theta_1, \dots, \theta_p)^T$ функции регрессии. Если исследуется обобщенная ЛММР, то наилучшим способом оценивания неизвестных параметров является обобщенный метод наименьших квадратов (ОМНК), в формулах которого участвует ковариационная матрица регрессионных остатков Σ_ϵ . Правда, и обычные МНК-оценки остаются состоятельными и несмещенными для регрессионных коэффициентов обобщенной ЛММР, однако они утрачивают при этом свойство наибольшей эффективности.

4. 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 $\hat{\Theta}_{\text{мнк}} = (\hat{\theta}_{\text{омнк}}, \hat{\theta}_{\text{мнк}}, \dots, \hat{\theta}_{\text{р.мнк}})^T$ параметров Θ ; затем вычисляются «невязки» (первое приближение) $\hat{\epsilon}_i^{(1)} = y_i - \hat{\theta}_{0.\text{мнк}} - \hat{\theta}_{1.\text{мнк}} x_i^{(1)} - \dots - \hat{\theta}_{\text{р.мнк}} x_i^{(p)}$ и по ним строится первое приближение $\hat{\Sigma}_\epsilon^{(1)}$ оценки для неизвестной матрицы Σ_ϵ ; затем матрица $\hat{\Sigma}_\epsilon^{(1)}$ используется для вычисления обобщенных МНК-оценок $\hat{\Theta}_{\text{омнк}}^{(1)}$. После этого пересчитываются невязки $\hat{\epsilon}_i^{(2)} = y_i - \hat{\theta}_{0.\text{омнк}}^{(1)} - \hat{\theta}_{1.\text{омнк}}^{(1)} x_i^{(1)} - \dots - \hat{\theta}_{\text{р.омнк}}^{(1)} x_i^{(p)}$ (т.е. вычисляется их второе приближение) и т.д. При этом, как правило, неизвестная ковариационная матрица остатков Σ_ϵ каким-то образом параметризуется, т.е. 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.

5. Основной характеристикой прогностической силы модели является так называемый коэффициент детерминации R^2 , определяющий, какая доля общей вариации анализируемой результирующей переменной 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, т.е. мы имеем возможность практически без ошибок предсказывать значения y по заданным значениям объясняющих переменных $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$.

6. 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, см. (2.4^а); в близости к единице хотя бы одного из коэффициентов детерминации $R_{x^{(j)}, x^{(j)}}$ 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.

7. К основным методам устранения мультиколлинеарности в ЛММР 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.

8. 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).

9. Специального рассмотрения требует ЛММР, в которых объясняющие переменные $X = (x^{(1)}, x^{(2)}, \dots, x^{(p)})^T$ стохастичны по своей природе. К таким моделям относятся, например, ситуации, в которых значения объясняющих переменных могут измеряться только со случайной ошибкой, а также весьма распространенные в эконометрических исследованиях построения, в которых в качестве объясняющих переменных используются значения случайных по своей природе результирующих показателей в предыдущие моменты времени (так называемые лаговые результирую-

щие показатели). Если при этом случайные *объясняющие переменные* X *некоррелированы с регрессионными остатками* ϵ , то к рассматриваемой модели применимы все методы и результаты, полученные в рамках классической или обобщенной (в зависимости от природы остатков ϵ) ЛМНР. Правда, интерпретация этих результатов при *стохастических* регрессорах X базируется на понятии *условного* распределения регрессионных остатков ϵ , в котором в качестве условия используется *заданность* (фиксированность) элементов матрицы наблюдений X . Если же стохастические предикторы X коррелированы с регрессионными остатками ϵ , то МНК- и ОМНК-оценки неизвестных коэффициентов Θ перестают быть состоятельными и несмещенными. В этом случае необходимо использовать специальные методы, одним из наиболее распространенных среди которых является метод введения *инструментальных переменных*.

10. 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 (т. е. 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 ЛМНР по неоднородным данным*, а само исследование в зависимости от условий проводится по одной из трех нижеследующих схем. 1) 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*. 2) 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. 3) Качественные перемен-

ные, 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 *сначала используют методы кластер-анализа или расщепления смесей распределений в пространстве $(x^{(1)}, x^{(2)}, \dots, x^{(p)}, y)$ с целью получения регрессионно однородных кластеров, а затем, так же как и в случае 1), строят и анализируют искомую модель регрессии отдельно по наблюдениям каждого такого кластера.*

11. Если в результате этапа параметризации модели исследователь приходит к выводу, что искомая регрессионная зависимость $f(X)$ *нелинейна*, то далее он действует следующим образом. Во-первых, он обращается к *процедуре линеаризации* модели, т.е. пытается подобрать такие преобразования к анализируемым переменным $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ и y , которые позволили бы представить искомую зависимость в виде *линейного* соотношения между *преобразованными* переменными. Помимо определенного набора стандартных преобразований (гиперболических, экспоненциальных, степенных и т.п.) с этой целью используют также подход Бокса-Кокса, при котором необходимое линеаризирующее преобразование подбирается из некоторого однопараметрического семейства преобразований на базе метода максимального правдоподобия.

Если же линеаризирующее преобразование подобрать не удастся, то приходится исследовать искомую *нелинейную* регрессионную зависимость в терминах *исходных* переменных. Вычисление МНК- или ОМНК-оценок неизвестных параметров потребует в этом случае привлечения соответствующих методов нелинейной оптимизации.

12. Специальный класс *логит- и пробит-моделей* составляют зависимости *дизотомического (бинарного) результирующего показателя y от количественных объясняющих переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$* . К этим моделям приходится обращаться, когда анализируемый социально-экономический объект (индивид, домашнее хозяйство, фирма, предприятие и т.п.) в контексте проводимого исследования может находиться в *одном из двух* состояний в зависимости от значений ряда объясняющих переменных (например, $y_i = 1$, если i -й трудоспособный член общества относится к категории «занятых», и $y_i = 0$, если он относится к категории «безработных»; при этом переменные $x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)}$ характеризуют его возраст, уровень образования, имеющийся стаж работы и т.д.). В подобных случаях переходят к новому результирующему показателю

\tilde{y}_i , определяющему *вероятность* события $\{y_i = 1\}$ как функцию специального вида $F(z_i)$, в которой в роли аргумента z_i выступает линейная функция от объясняющих переменных, т. е. $z_i = \theta_0 + \theta_1 x_i^{(1)} + \dots + \theta_p x_i^{(p)}$. Если в качестве функции $F(z)$ используют так называемую *логистическую кривую*, то модель вида $\tilde{y}_i = F(z_i)$ называется «*логит-моделью*». Если же в качестве функции $F(z)$ используется функция *стандартного нормального распределения*, то соответствующая модель носит название «*пробит-модели*». Очевидно, оценка параметров $\theta_0, \theta_1, \dots, \theta_p$ в таких моделях связана с необходимостью решения специального класса нелинейных задач оптимизации.

ГЛАВА 3. АНАЛИЗ ВРЕМЕННЫХ РЯДОВ (МОДЕЛИ И ПРОГНОЗИРОВАНИЕ)

Всякий эконометрический анализ основывается на *исходных статистических данных* (и.с.д.). Их общая форма была описана в п. 9.1.2 (см. том 1, (9.1), (9.1') и (9.2)). При этом если процесс регистрации и.с.д. происходит *во времени* t и само время фиксируется наряду со значениями анализируемых характеристик $x_i^{(j)}(t_k)$ ($j = 1, 2, \dots, p$; $i = 1, 2, \dots, n$; $k = 1, 2, \dots, N$), то говорят о статистическом анализе так называемых *панельных данных*¹. Если зафиксировать номер переменной j и номер статистически обследуемого объекта i , то расположенную в хронологическом порядке последовательность значений

$$x_i^{(j)}(t_1), x_i^{(j)}(t_2), \dots, x_i^{(j)}(t_k) \quad (3.1)$$

называют *одномерным временным рядом*. Если же одновременно рассматривать p одномерных временных рядов вида (3.1), т.е. исследовать закономерности во *взаимосвязанном* поведении временных рядов (3.1) 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 использование в качестве своей статистической базы временных рядов тех или иных показателей.

Данная глава практически полностью посвящена *методам построения, идентификации (т.е. — статистического оценивания параметров)*

¹ Напомним, что в принятых выше обозначениях j — номер анализируемой количественной характеристики (или переменной), i — номер статистически обследованного объекта, а k — порядковый номер времени регистрации значения анализируемой переменной на i -м объекте.

и верификации (т. е. — 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 схем с прямой и обратной связями* (см., например, [Дж. Бокс, Г. Дженкинс]). Добавим к этому, что в данной главе будут рассматриваться лишь *дискретные (по времени наблюдения)* одномерные временные ряды для *равноотстоящих моментов наблюдения*, т. е. $t_2 - t_1 = t_3 - t_2 = \dots = t_N - t_{N-1} = \Delta$, 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

$$x(1), x(2), \dots, x(N), \quad (3.1')$$

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 (3.1') для прогнозирования значения $x(t)$ на один или несколько временных тактов вперед (т. е. — для прогнозной оценки значений $\hat{x}(N+l)$, $l = 1, 2, 3$) 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 (3.1'), 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*.

3.1. Временной ряд (определения, примеры, формулировка основных задач)

Выше, говоря о *расположенной в хронологическом порядке последовательности наблюдаемых значений* (3.1) или (3.1') *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 (3.1') могут быть в точности вычислены как значения некоторой *нестандартной функции* $f(t)$, т. е. $x(t) = f(t)$. Поэтому уточним понятие временного ряда, принятое в данной главе.

О п р е д е л е н и е 3.1. *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 (3.1').

Определение 3.1 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 (3.1') от введенной в п. 6.1 (том 1) последовательности наблюдений x_1, x_2, \dots, x_n , образующих случайную выборку? Этим отличиям 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*, т. е. $P\{x(t_1) < x\} \neq P\{x(t_2) < x\}$ 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 (т. е. 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 $f_{тр}(t)$, 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*

¹ Случайные 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 $f_{тр}(t), \varphi(t)$ 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*.

четырёх типов. Из приведенных ниже примеров мы увидим, что в одних случаях значения временного ряда формируются под воздействием факторов (А), (Б) и (Г) (см. пример 3.1), в других — под воздействием факторов (А), (В) и (Г) (см. пример 3.2) и, наконец, — исключительно под воздействием одних только случайных факторов (Г) (см. пример 3.3). Однако во всех случаях *предполагается непрерывное участие случайных (эволюционных) факторов (Г)*. Кроме того, мы примем (в качестве гипотезы) для определенности *аддитивную структурную схему* влияния факторов А, Б, В и Г на формирование значений $x(t)$, которая означает правомерность представления значений членов временного ряда в виде разложения:

$$x(t) = \chi(A) f_{\text{TP}}(t) + \chi(B) \varphi(t) + \chi(V) \psi(t) + \epsilon(t), \quad t = 1, 2, \dots, N, \quad (3.2)$$

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 (т.е. 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 3.1. В табл. 3.1 и на рис. 3.1 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. (т.е. $t = 1, 2, \dots, 96$; временной такт Δ равен одному месяцу).

Таблица 3.1. Расстояния, пройденные британскими авиалайнерами 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

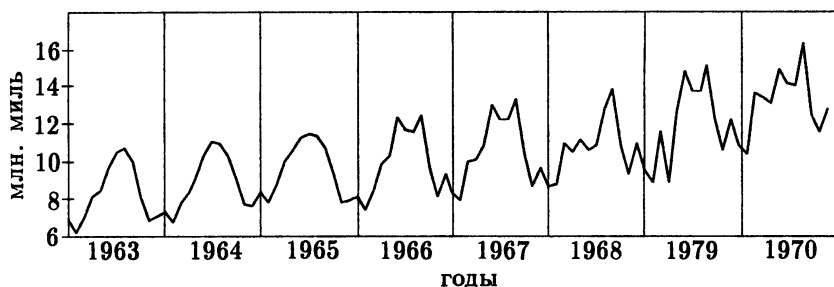


Рис. 3.1. График данных из табл. 3.1 (расстояния, пройденные авиалайнерами Соединенного Королевства за месяц)

Пример 3.2. В табл. 3.2 и на рис. 3.2 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 $N = 48$).

Таблица 3.2. 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. 3.2. График временного ряда, представленного данными табл. 3.2 (индекс FT, квартальные средние)

Пример 3.3. В табл. 3.3 и на рис. 3.3 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 3.3. Урожайность ячменя в Англии и Уэльсе с 1884 по 1939 гг. (в англ. центнерах на акр¹, 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		

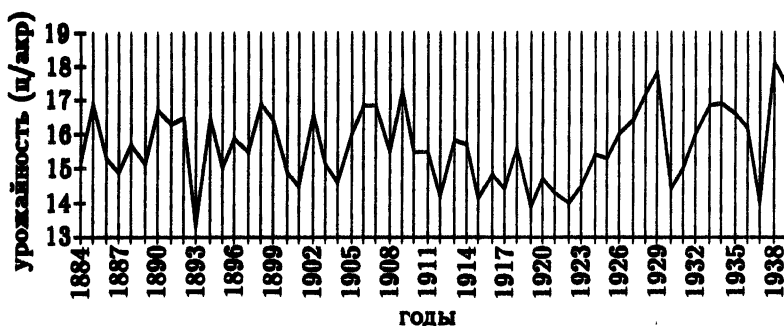


Рис. 3.3. График временного ряда, представленного данными табл. 3.3 (урожайность ячменя, ц/акр)

¹ Английский центнер равен 50,8 кг, один акр равен 0,405 га.

Попробуем проанализировать приведенные примеры временных рядов, опираясь на описанную выше общую схему их генезиса и вытекающее из нее разложение вида (3.2).

Данные об авиалайнерах (п р и м е р 3.1) 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 авиалайнеров, 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 п р и м е р е 3.2 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 3.3 (см. рис. 3.3), 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* (3.2) временного ряда $x(t)$, можно дать общую формулировку *базисной цели* его статистического анализа¹:

- по имеющейся траектории (3.1') анализируемого временного ряда $x(t)$ требуется:
 - (i) определить, какие из неслучайных функций $f_{\text{ТР}}(t)$, $\varphi(t)$ и $\psi(t)$ при-

¹ Мы 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.

существуют в разложении (3.2), т. е. определить значения индикаторов $\chi(C)$ ($C = A, B, V$) в разложении (3.2);

(ii) построить «хорошие» оценки для тех неслучайных функций, которые присутствуют в разложении (3.2);

(iii) подобрать модель, адекватно описывающую поведение «случайных остатков» $\varepsilon(t)$, и статистически оценить параметры этой модели.

Успешное решение задач (i) ~ (iii), 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.

3.2. 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 (iii)), 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 3.3).

О п р е д е л е н и е 3.2. 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:

$$\text{среднее значение } E x(t) = a \quad (3.3)$$

и

$$\text{дисперсия } D x(t) = E(x(t) - a)^2 = \sigma^2. \quad (3.4)$$

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 $x(1), x(2), \dots, x(N)$ ¹. В частности:

$$\hat{a} = \frac{1}{N} \sum_{t=1}^N x(t) - \text{оценка среднего значения}, \quad (3.3')$$

$$\hat{\sigma} = \frac{1}{N} \sum_{t=1}^n (x(t) - \hat{a})^2 - \text{оценка дисперсии}. \quad (3.4')$$

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:

$$\gamma(\tau) = \text{cov}(x(t), x(t + \tau)) = E[(x(t) - a)(x(t + \tau) - a)]. \quad (3.5)$$

При анализе изменения величины $\gamma(\tau)$ в зависимости от значения τ принято говорить об *автоковариационной функции $\gamma(\tau)$* . Значения авто-

¹ В общем случае (т.е. при отсутствии свойства стационарности) для оценки з.р.в. случайной величины $x(t)$ и ее основных числовых характеристик мы должны были 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 [М. Дж. Кендалл, А. Стьюарт, 1976]) оценивать среднее значение, дисперсию и ковариации стационарного временного ряда по его *единственной реализации* (3.1').

ковариационной функции могут быть *статистически оценены* по имеющимся наблюдениям временного ряда (3.1') по формуле

$$\hat{\gamma}(\tau) = \frac{1}{N - \tau} \sum_{t=1}^{N-\tau} (x(t) - \hat{a})(x(t + \tau) - \hat{a}), \quad (3.5')$$

где $\tau = 1, 2, \dots, N - 1$, а \hat{a} вычислено по формуле (3.3').

Очевидно, значение автоковариационной функции при $\tau = 0$ есть не что иное, как дисперсия временного ряда, т. е.

$$\gamma(0) = \sigma^2 = \mathbf{E}(x(t) - a)^2$$

и, соответственно,

$$\hat{\gamma}(0) = \hat{\sigma}^2 = \frac{1}{N} \sum_{t=1}^N (x(t) - \hat{a})^2. \quad (3.4'')$$

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. 3.1), в том, что члены временного ряда являются, вообще говоря, статистически *взаимозависимыми*. Степень тесноты статистической связи между двумя случайными величинами может быть измерена *парным коэффициентом корреляции* (см. гл. 11). Так что степень тесноты статистической связи между наблюдениями временного ряда, «разнесенными» (по времени) на τ единиц, определится величиной коэффициента корреляции

$$r(\tau) = \frac{\mathbf{E}[(x(t) - a)(x(t + \tau) - a)]}{[\mathbf{E}(x(t) - a)^2 \cdot \mathbf{E}(x(t + \tau) - a)^2]^{\frac{1}{2}}} = \frac{\gamma(\tau)}{\gamma(0)}, \quad (3.6)$$

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*, т. е. 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

$$\hat{r}(\tau) = \frac{\frac{1}{N-\tau} \sum_{t=1}^{N-\tau} (x(t) - \hat{a})(x(t+\tau) - \hat{a})}{\frac{1}{N} \sum_{t=1}^N (x(t) - \hat{a})^2} = \frac{\hat{\gamma}(\tau)}{\hat{\gamma}(0)}, \quad (3.6')$$

$$\tau = 1, 2, \dots, N - 1.$$

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 $x(t + \tau)$ (т. е. 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 (т. е. $r(\tau) \equiv 0$ для всех $\tau \geq \tau_0$).

В левых частях рис. 3.4, 3.5 и 3.6 приведены графики автокорреляционных функций для ряда моделей временных рядов, рассмотренных ниже (см. пп. 3.4–3.7). Обращаем внимание читателя на тот факт, что в отличие от теоретических автокорреляционных функций их выборочные (эмпирические) аналоги (3.6') 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 разделёнными τ тактами времени членами временного ряда $x(t)$ и $x(t + \tau)$, при *устраненном опосредованном влиянии на эту взаимозависимость всех промежуточных* (т. е. расположенных между $x(t)$ и $x(t + \tau)$) членов этого временного ряда. Идея «о ч и щ е н о й» от опосредованного влияния, т. е. *частной*, корреляции уже обсуждалась в этой книге (см. том 1, п. 11.2.4). Поэтому нам остается лишь применить понятия, результаты и формулы этого раздела корреляционного анализа к последовательности членов временного ряда. Так, частная автокорреля-

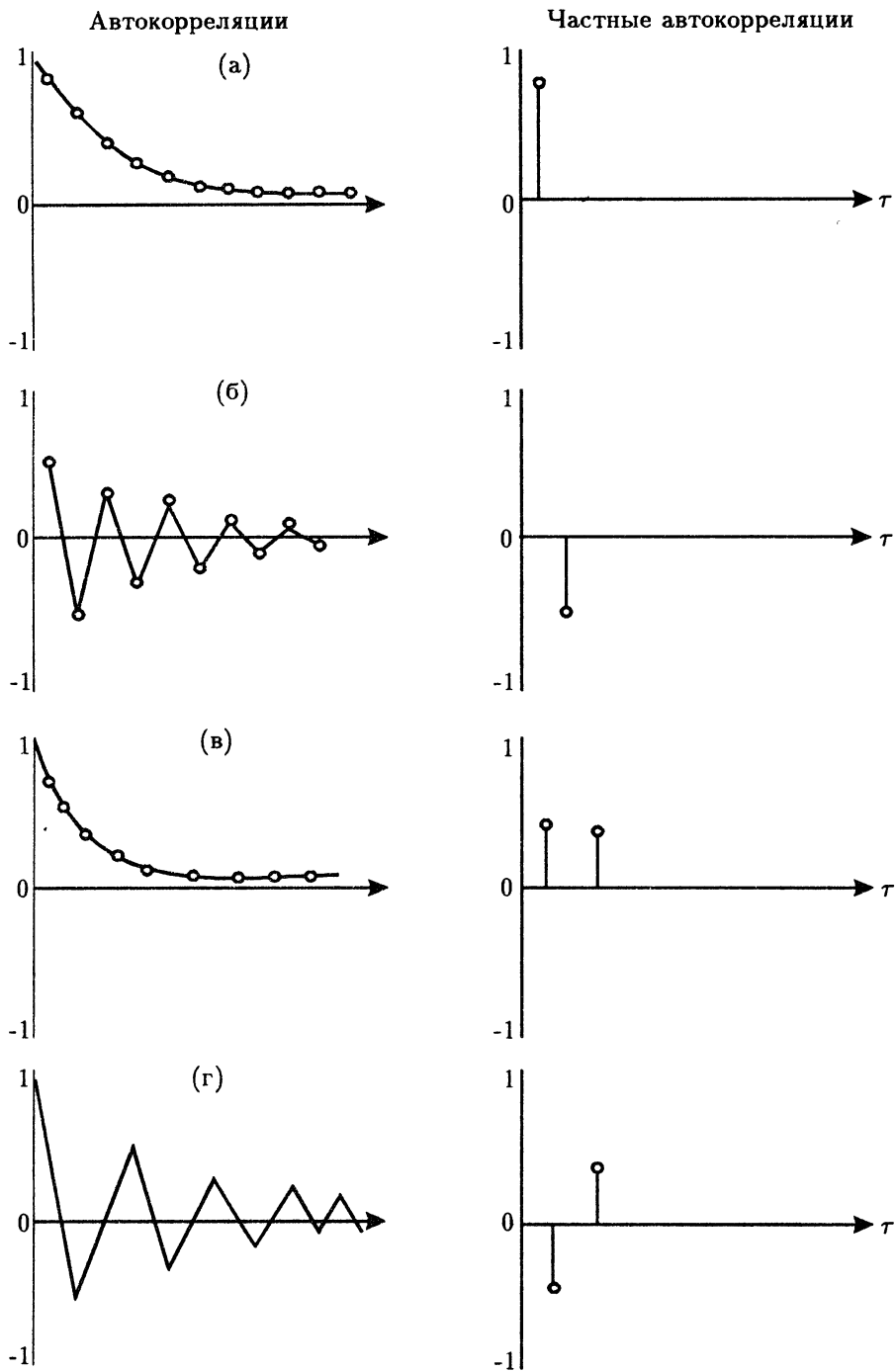


Рис. 3.4. Схематичные графики автокорреляционной и частной автокорреляционной функций для моделей авторегрессии 1-го ((а) и (б)) и 2-го ((в) и (г)) порядков

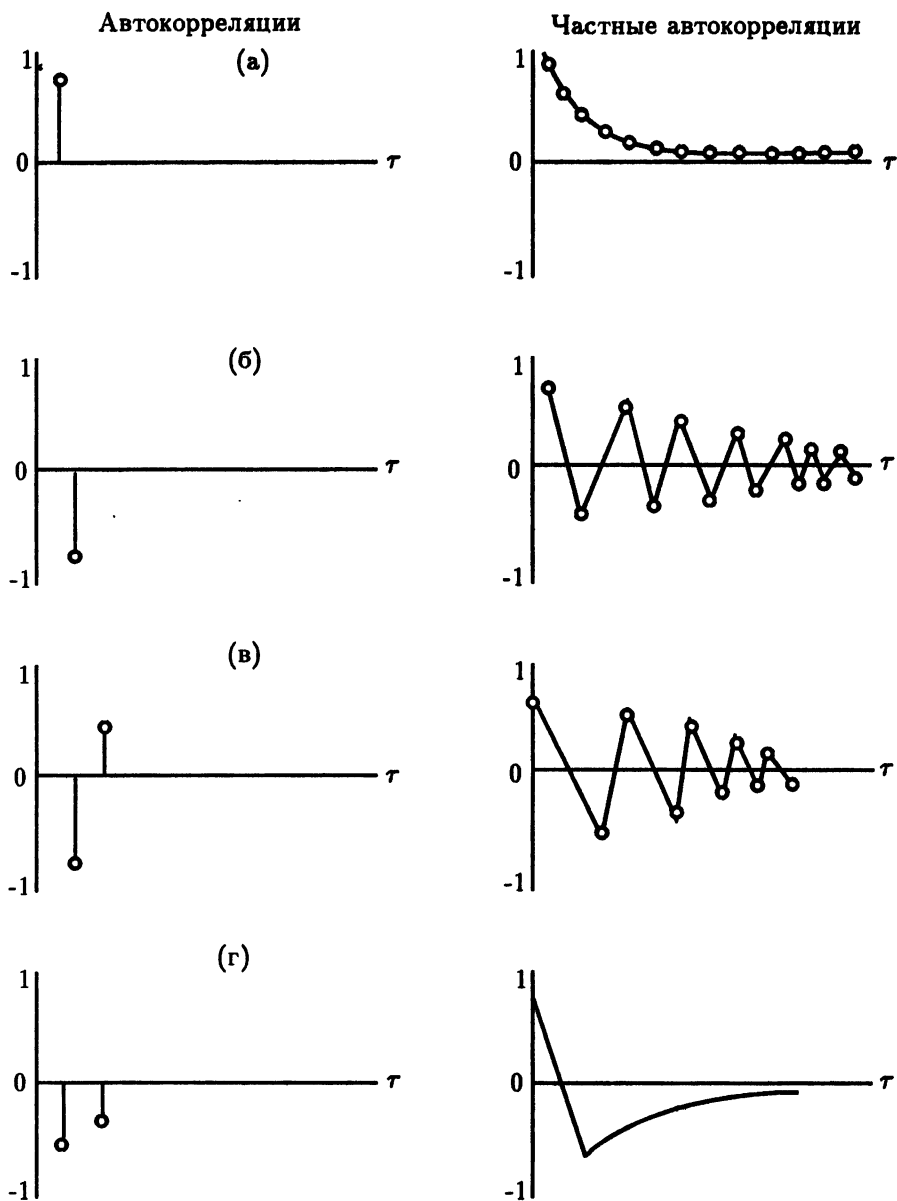


Рис. 3.5. Схематичные графики автокорреляционной и частной автокорреляционной функции для моделей скользящего среднего 1-го ((а) и (б)) и 2-го ((в) и (г)) порядков

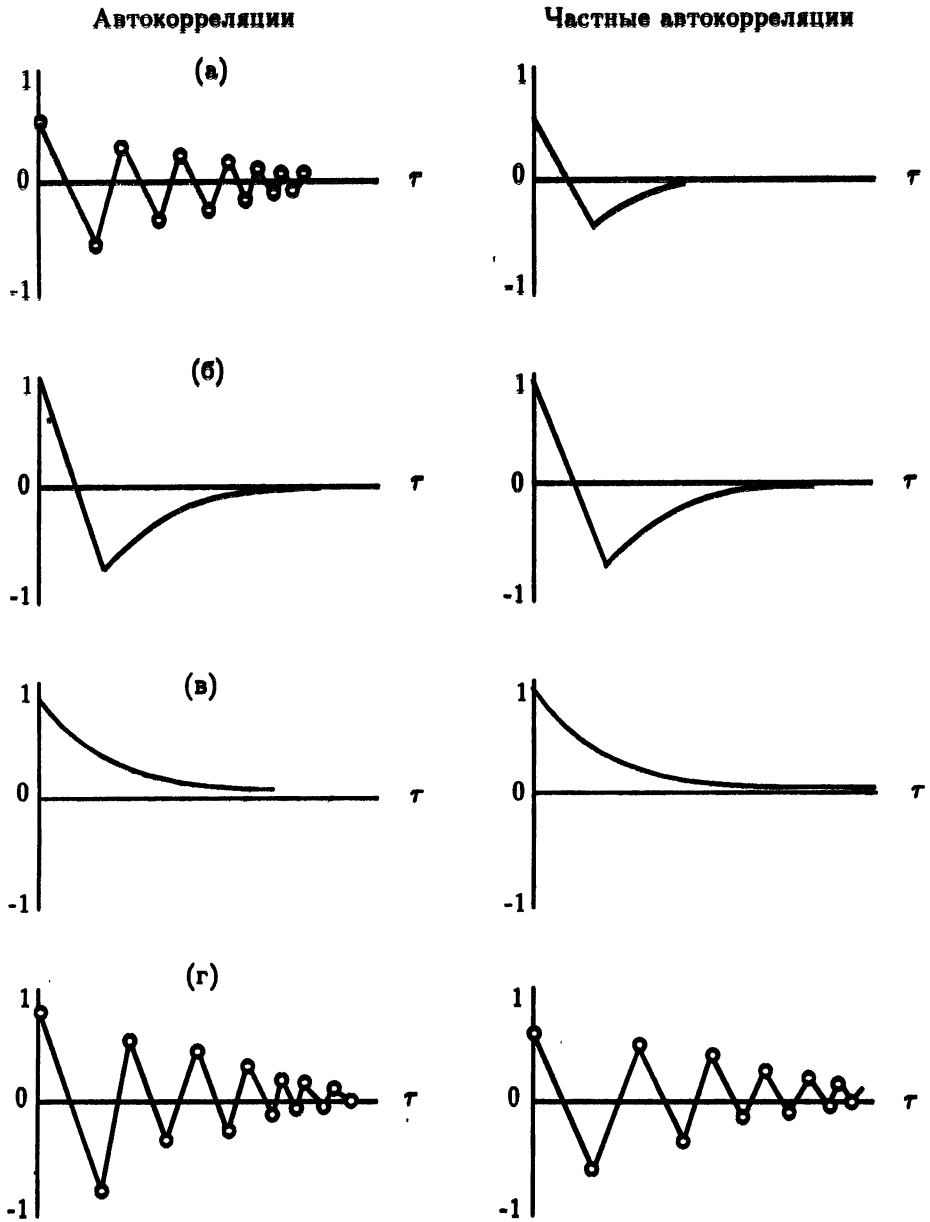


Рис. 3.6. Схематичные графики автокорреляционной и частной автокорреляционной функций для моделей вида $x(t) = \alpha x(t-1) + \delta(t) - \theta\delta(t-1)$ для различных сочетаний знаков коэффициентов α и θ

ция 1-го порядка может быть подсчитана с использованием соотношения (11.32) из тома 1:

$$\begin{aligned} r_{\text{част}}(2) &\equiv r(x(t), x(t+2) | x(t+1) = a) \\ &\equiv \frac{r(x(t), x(t+2)) - r(x(t), x(t+1))r(x(t+2), x(t+1))}{\sqrt{[1 - r^2(x(t), x(t+1))][1 - r^2(x(t+2), x(t+1))]} \\ &\equiv \frac{r(2) - r^2(1)}{1 - r^2(1)}, \end{aligned} \quad (3.7)$$

где a — среднее значение анализируемого стационарного процесса.

Частные автокорреляции более высоких порядков могут быть подсчитаны аналогичным образом с использованием либо рекуррентной формулы (11.32'), либо с помощью формулы (11.31) тома 1 по элементам общей корреляционной ($N \times N$) — матрицы 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* (т.е. 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. (11.32')):

$$r_{\text{част}}(3) = r(x(t), x(t+3) | x(t+1) = x(t+2) = a) = \frac{r_{03.1} - r_{02.1}r_{32.1}}{\sqrt{(1 - r_{02.1}^2)(1 - r_{32.1}^2)}}, \quad (3.8)$$

где

$$\begin{aligned} r_{03.1} &= r(x(t), x(t+3) | x(t+1) = a), \\ r_{02.1} &= r(x(t), x(t+2) | x(t+1) = a), \\ r_{32.1} &= r(x(t+3), x(t+2) | x(t+1) = a) \end{aligned}$$

и в соответствии с формулой (11.32), имеем:

$$\begin{aligned} r_{03.1} &= r(x(t), x(t+3) | x(t+1) = a) = \frac{r(3) - r(1)r(2)}{\sqrt{(1 - r^2(1))(1 - r^2(2))}}, \\ r_{02.1} &= r(x(t), x(t+2) | x(t+1) = a) = \frac{r(2) - r^2(1)}{1 - r^2(1)}, \\ r_{32.1} &= r(x(t+3), x(t+2) | x(t+1) = a) = \frac{r(1) - r(2)r(1)}{\sqrt{(1 - r^2(2))(1 - r^2(1))}}. \end{aligned}$$

Эмпирические (выборочные) версии частных автокорреляционных функций получаются с помощью тех же соотношений (3.7), (3.8) при за-

мене участвующих в них теоретических значений автокорреляций $r(\tau)$ их статистическими оценками $\hat{r}(\tau)$ (см. формулу (3.6')).

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. 3.4, 3.5 и 3.6 приведены схематические графики частных автокорреляционных функций для рассмотренных ниже (см. пп. 3.4~3.7) 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. 3.4), можно избежать прямого вычисления частных корреляционных функций по формулам (3.7)–(3.8), воспользовавшись более простой процедурой.

Спектральная плотность $p(\omega)$. Определим спектральную плотность стационарного временного ряда $x(t)$ через его автокорреляционную функцию $r(\tau)$ соотношением вида

$$p(\omega) = \sum_{\tau=-\infty}^{\infty} r(\tau)e^{i\tau\omega}, \quad (3.9)$$

где $i = \sqrt{-1}$ (мнимая единица). Благодаря тому, что $r(\tau) = r(-\tau)$, спектральная плотность $p(\omega)$ может быть записана в виде

$$p(\omega) = 1 + 2 \sum_{\tau=1}^{\infty} r(\tau) \cos(\tau\omega). \quad (3.9')$$

Мы видим, что функция (3.9') является *гармонической* и имеет период $2\pi^1$. Поскольку $\cos[(2\pi - \omega)\tau] = \cos(\omega\tau)$, то график спектральной плотности $p(\omega)$, называемый *спектром*, симметричен относительно $\omega = \pi$.

¹ Аргумент ω измеряется в радианах на единицу времени и называется *угловой частотой* (т. к. 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 (3.2) анализируемого временного ряда $x(t)$ по-своему отражается на поведении спектральной плотности $p(\omega)$.

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. Дженкинс, 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, т. к. *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 экономических рядов 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

$$r(k) = \frac{1}{\pi} \int_0^{\pi} p(\omega) e^{-ik\omega} d\omega = \frac{1}{\pi} \int_0^{\pi} \cos(k\omega) p(\omega) d\omega. \quad (3.10)$$

Этот результат получается, в частности, из соотношения (3.9') после его домножения на $\cos(k\omega)$ и почленного интегрирования получившегося 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 (3.1') (измеренного относительно его среднего, т. е. 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 автокорреляции $r(k)$. Однако, устремив $N \rightarrow \infty$ и переходя к пределу, мы получим:

$$E(I(\omega)) = \frac{\sigma^2}{\pi} \left[1 + 2 \sum_{k=1}^{\infty} r(k) \cos(k\omega) \right],$$

что, как это следует из (3.9'), эквивалентно

$$E(I(\omega)) = \frac{\sigma^2}{\pi} p(\omega) \quad (3.11)$$

(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 (3.2). 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.

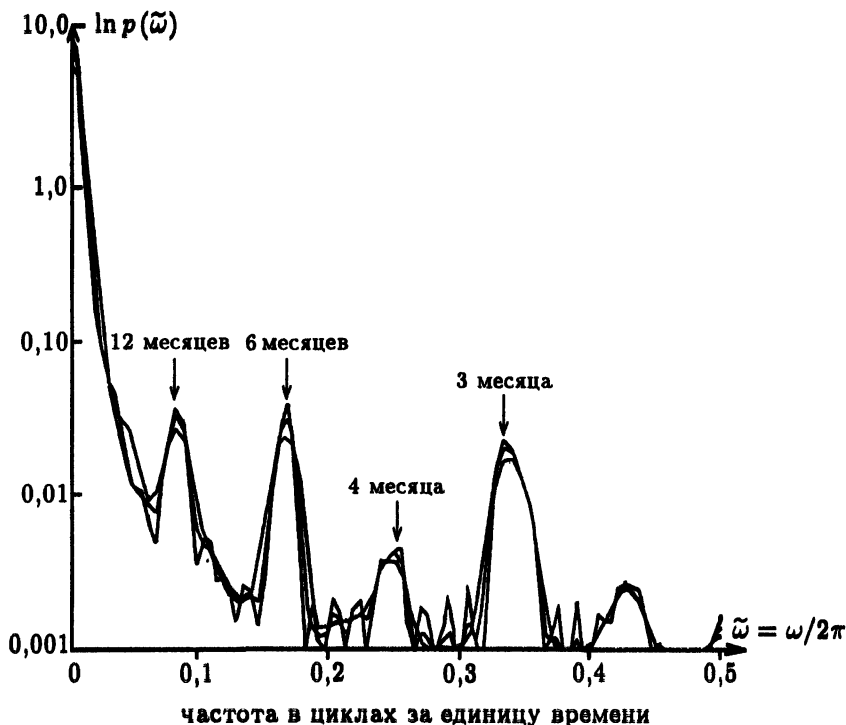


Рис. 3.7. Спектральная функция очищенных от тренда данных о банковском клиринге в США

Пример 3.4. На рис. 3.7 представлен пример графика спектральной плотности (т. е. *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.

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.

Определение 3.3. Ряд $x(t)$ называется **слабо стационарным (или стационарным в широком смысле)**, если его среднее значение, дисперсия и ковариации не зависят от t , т. е. если выполняются соотношения (3.3)–(3.4)–(3.5).

Очевидно, все строго стационарные (или стационарные в узком смысле, см. определение 3.2) временные ряды являются одновременно и стационарными в широком смысле, но не наоборот.

3.3. Неслучайная составляющая временного ряда и методы его сглаживания

В п. 3.1 были описаны основные задачи (i) ~ (iii) статистического анализа временного ряда (3.1'). Существенную роль в решении задач выявления и оценивания трендовой ($f_{\text{ТР}}(t)$), сезонной ($\varphi(t)$) и циклической ($\psi(t)$) составляющих в разложении (3.2) играет начальный этап анализа, на котором:

- *выявляется сам факт наличия/отсутствия неслучайной (и зависящей от времени t) составляющей в разложении (3.2); по существу, речь идет о статистической проверке гипотезы*

$$H_0: E x(t) = a = \text{const} \quad (3.12)$$

(включая утверждение о взаимной статистической независимости

членов анализируемого временного ряда (3.1')) при различных вариантах конкретизации альтернативных гипотез типа

$$H_1: \mathbf{E}x(t) \neq \text{const}; \quad (3.13)$$

- строится оценка (аппроксимация) для неизвестной интегральной неслучайной составляющей $f(t) = \chi(A)f_{\text{ТР}}(t) + \chi(B)\varphi(t) + \chi(C)\psi(t)$, т. е. решается задача сглаживания (элиминирования случайных остатков $\varepsilon(t)$) анализируемого временного ряда $x(t)$ (функции χ определены в (3.2)).

Данный пункт посвящен именно этому начальному этапу статистического анализа временного ряда.

3.3.1. Проверка гипотезы о неизменности среднего значения временного ряда

Критерий серий, основанный на медиане. Расположим члены анализируемого временного ряда (3.1') в порядке возрастания, т. е. образуем по наблюдениям (3.1') *вариационный ряд*

$$x_{(1)}, x_{(2)}, \dots, x_{(n)}.$$

Следуя рекомендациям п. 6.2.3, определим *выборочную медиану* $x_{\text{med}}^{(n)}$ по формуле

$$x_{\text{med}}^{(n)} = \begin{cases} x_{(\frac{n+1}{2})}, & \text{если } n \text{ нечетно,} \\ \frac{1}{2}(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}), & \text{если } n \text{ четно.} \end{cases}$$

После этого мы образуем «серии» из плюсов и минусов, на статистическом анализе которых основана процедура проверки гипотезы (3.12). В частности, возвращаясь к исходному временному ряду (3.1'), 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 (3.1') состоит из статистически независимых наблюдений, случайно варьирующих около некоторого *постоянного* уровня a (т. е. если справедлива гипотеза (3.12)), то чередование плюсов и минусов в построенной указанным выше способом последовательности должно быть более или менее случайным, т. е. 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 данной критерии целесообразно рассматривать *одновременно пару критических статистик* ($\nu(n)$; $\tau(n)$) (определение критической статистики см. в п. 8.2).

Тогда, следуя общей логической схеме построения статистического критерия (см. п. 8.2), мы должны были бы вывести и затабулировать двумерный закон распределения случайной величины ($\nu(n)$; $\tau(n)$), справедливый при условии, что проверяемая гипотеза (3.12) верна. Мы ограничимся здесь изложением *приближенного* критерия. Для его построения воспользуемся:

- $(\frac{n+2}{2}, \frac{n-1}{4})$ — 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 $[x]$ 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 (3.12):

если хотя бы одно из неравенств

$$\begin{aligned} \nu(n) &> \left[\frac{1}{2}(n+2 - 1,96\sqrt{n-1}) \right], \\ \tau(n) &< [1,43 \ln(n+1)] \end{aligned} \quad (3.14)$$

окажется нарушенным, то гипотеза (3.12) отвергается с вероятностью ошибки α , заключенной между 0,05 и 0,0975 (и, следовательно, подтвер-

ждается наличие зависящей от времени неслучайной составляющей в разложении (3.2) анализируемого временного ряда).

Пример 3.5. 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, т. е. 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 (3.14) 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 обследования) 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, т. е. 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:

$$\nu(n) > \left[\frac{1}{3}(2n - 1) - 1,96\sqrt{\frac{16n - 29}{90}} \right], \quad (3.15)$$

$$\tau(n) < \tau_0(n),$$

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:

$$\begin{array}{cccc} n & n \leq 26 & 26 < n \leq 153 & 153 < n \leq 1170 \\ \tau_0(n) & \tau_0 = 5 & \tau_0 = 6 & \tau_0 = 7. \end{array}$$

Если хотя бы одно из неравенств (3.15) окажется нарушенным, то гипотезу (3.12) следует отвергнуть (и, соответственно, признать, что в разложении (3.2) анализируемого временного ряда присутствует неслучайная, зависящая от времени t компонента).

Пример 3.6. Вернемся к данным примера 3.3 (см. выше табл. 3.3 и рис. 3.3) 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 (3.15), 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 (3.15) 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 подчиняется

нормальному закону распределения вероятностей, то для выяснения вопроса о возможном систематическом смещении среднего в ходе выборочного обследования целесообразнее воспользоваться критерием квадратов последовательных разностей¹.

Для проверки гипотезы (3.12) 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

$$\gamma(n) \leq \gamma_{\alpha}^{\min}(n), \quad (3.16)$$

то гипотеза (3.12) 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 нормированного нормального распределения.

Величины $\gamma_{\alpha}^{\min}(n)$ при $n \leq 60$ для трех наиболее употребительных значений уровня значимости α даны в табл. 4.9 книги [Большев Л. Н., Смирнов Н. В.].

¹ В этом случае данный критерий оказывается более мощным (см. гл. 8), чем предыдущий. Это означает, в частности, что если мы воспользуемся обоими этими критериями при данном объеме выборки n и заданном уровне значимости α (т. е. при заданной вероятности ошибочного отвержения гипотезы (3.12)), то вероятность ошибиться в другую сторону (т. е. принять гипотезу (3.12), в то время как на самом деле она является ошибочной) окажется меньшей в случае критерия квадратов последовательных разностей.

3.3.2. Методы сглаживания временного ряда (выделение неслучайной составляющей)

Методы выделения неслучайной составляющей в траектории, отражающей поведение анализируемого временного ряда (3.1'), можно условно разделить на два типа.

Методы первого типа (аналитические) основаны на допущении, что нам известен общий вид неслучайной составляющей в разложении (3.2)

$$f(t) = \chi(A)f_{\text{Тр}}(t) + \chi(B)\varphi(t) + \chi(C)\psi(t). \quad (3.17)$$

Например, рассматривая график временного ряда из примера 3.2 (см. рис. 3.2), исследователь может предположить, что неслучайная составляющая динамики курса анализируемых акций описывается *линейной функцией времени* t , т. е.

$$f(t) = \theta_0 + \theta_1 t,$$

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* (3.2), или, что то же, — *задача сглаживания временного ряда* (3.1')) 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 (3.17) 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 (т. е. заданное в виде явной формулы) представление функции (3.17).

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 (3.1'), 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 $\Theta = (\theta_0, \theta_1, \dots, \theta_m)^T$. Оценки $\hat{\Theta}$ параметров Θ строятся по наблюдениям $\{t; x(t)\}_{t=1,2,\dots,n}$ с помощью методов, описанных в гл. 2. 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 , т. е.

$$f(t; \Theta) = \theta_0 + \theta_1 t + \dots + \theta_p t^p, \quad (3.18)$$

и при этом длина временного ряда n существенно превышает степень этого полинома p^1 , 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. 2.3.1, формулу (2.22)):

$$\hat{\Theta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}. \quad (3.19)$$

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. 2.9.1); исследование моделей (3.18), в которых неслучайная компонента $f(t; \Theta)$ *нелинейна* относительно оцениваемых параметров Θ , потребует техники статистического анализа нелинейных моделей регрессии (см. п. 2.12) и т. д.

Алгоритмические методы выделения неслучайной составляющей временного ряда (методы скользящего среднего). В основе

¹ Обычно удовлетворительная надежность статистических оценок $\hat{\Theta}$ параметров Θ достигается уже при $n \geq 4p$, но, как минимум, требуется, чтобы $n > p + 1$.

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 N членов временного ряда $(x(1) + x(2) + \dots + x(N))/N$ около того же значения a будет характеризоваться гораздо меньшей величиной дисперсии, а именно дисперсией, равной σ^2/N . А уменьшение меры случайного разброса (дисперсии) и означает как раз *сглаживание* соответствующей траектории. Поэтому выбирают некоторую нечетную «длину усреднения» $N = 2m + 1$, измеренную в числе подряд идущих членов анализируемого временного ряда (конкретный выбор числа m 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 $x(t - m)$, $x(t - m + 1)$, ..., $x(t)$, $x(t + 1)$, ..., $x(t + m)$ по формуле

$$\hat{f}(t) = \sum_{k=-m}^m w_k x(t+k), \quad t = m+1, m+2, \dots, n-m, \quad (3.20)$$

где w_k ($k = -m, -m+1, \dots, m$) — некоторые «весовые» коэффициенты (или просто «веса»), в сумме равные единице, т. е. $\sum_{k=-m}^m w_k = 1$. Посколь-

ку, 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 (3.20) происходит замена только одного слагаемого $x(t-m)$ слагаемым $x(t+m+1)$), то и методы, основанные на формуле (3.20), 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 (т. е. 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* (т. е. $(m+1)$ -й) точке этого отрезка ряда, т. е. 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, т.е. $\hat{f}(m+2) = \hat{x}_2(m+2)$, 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 (т.е. 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 поведение сглаженной функции $f(t)$ описывается алгебраическим полиномом 1-й степени ($p = 1$), т.е. 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, т.е.

$$\sum_{t'=-m}^m (x(t') - \theta_0 - \theta_1 t')^2 \rightarrow \min_{\theta_0, \theta_1}.$$

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:

$$\begin{aligned} \hat{f}(m+1) &= \hat{\theta}_0 + \hat{\theta}_1 t']_{t'=0} = \hat{\theta}_0 = \frac{1}{2m+1} \sum_{t'=-m}^m x(t') \\ &= \frac{1}{2m+1} \sum_{t=1}^{2m+1} x(t). \end{aligned} \quad (3.21)$$

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:

$$\hat{f}(t) = \frac{1}{2m+1} \sum_{k=-m}^m x(t+k), \quad (3.21')$$

т. е. при линейном характере локальной аппроксимации траектории временного ряда в качестве его сглаженного значения в точке t следует брать среднее арифметическое из охватывающих 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 ($p=0$), т. е., 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 (3.21).

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 ($N = 2m+1 = 2 \cdot 2 + 1 = 5$) и что локальное поведение сглаженного временного ряда внутри каждого такого отрезка мы будем аппроксимировать параболой 2-го порядка, т. е.

$$f(t') = \theta_0 + \theta_1 t' + \theta_2 (t')^2, \quad t' = -2, -1, 0, 1, 2$$

(здесь t' — уже «переобозначенное» время, т. е. 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:

$$\left\{ \begin{array}{l} \frac{\partial Q(\Theta)}{\partial \theta_0} = -2 \left[\sum_{t'=-2}^2 x(t') - 5\theta_0 - \theta_1 \sum_{t'=-2}^2 t' - \theta_2 \sum_{t'=-2}^2 (t')^2 \right] = 0, \\ \frac{\partial Q(\Theta)}{\partial \theta_1} = -2 \left[\sum_{t'=-2}^2 t' x(t') - \theta_0 \sum_{t'=-2}^2 t' - \theta_1 \sum_{t'=-2}^2 (t')^2 - \theta_2 \sum_{t'=-2}^2 (t')^3 \right] = 0, \\ \frac{\partial Q(\Theta)}{\partial \theta_2} = -2 \left[\sum_{t'=-2}^2 (t')^2 x(t') - \theta_0 \sum_{t'=-2}^2 (t')^2 - \theta_1 \sum_{t'=-2}^2 (t')^3 - \theta_2 \sum_{t'=-2}^2 (t')^4 \right] = 0; \end{array} \right. \quad (3.22)$$

- решение системы уравнений МНК (с учетом того, что $\sum_{t'=-2}^2 t' = \sum_{t'=-2}^2 (t')^3 = 0$):

$$\begin{aligned} \hat{\theta}_0 &= \frac{\sum_{t'=-2}^2 (t')^4 \cdot \sum_{t'=-2}^2 x(t') - \sum_{t'=-2}^2 (t')^2 \sum_{t'=-2}^2 (t')^2 x(t')}{5 \sum_{t'=-2}^2 (t')^4 - \left(\sum_{t'=-2}^2 (t')^2 \right)^2} \\ &= \frac{34 \sum_{t'=-2}^2 x(t') - 10 \sum_{t'=-2}^2 (t')^2 x(t')}{170 - 100} \\ &= \frac{1}{35} [-3x(-2) + 12x(-1) + 17x(0) + 12x(1) - 3x(2)], \end{aligned}$$

$$\hat{\theta}_1 = \frac{\sum_{t'=-2}^2 t' x(t')}{\sum_{t'=-2}^2 (t')^2} = \frac{1}{10} [-2x(-2) - x(-1) + x(1) + 2x(2)],$$

$$\hat{\theta}_2 = \frac{5 \sum_{t'=-2}^2 (t')^2 x(t') - \sum_{t'=-2}^2 (t')^2 \sum_{t'=-2}^2 x(t)}{5 \sum_{t'=-2}^2 (t')^4 - \left(\sum_{t'=-2}^2 (t')^2 \right)^2}$$

$$= \frac{1}{14} [2x(-2) - x(-1) - 2x(0) - x(1) + 2x(2)].$$

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 $t' = 0$, т. е.

$$\hat{f}(t) = \hat{\theta}_0 + \hat{\theta}_1 t' + \hat{\theta}_2 (t')^2 \Big|_{t'=0} = \hat{\theta}_0$$

$$= \frac{1}{35} [-3x(t-2) + 12x(t-1) + 17x(t) + 12x(t+1) - 3x(t+2)].$$

Таким образом, мы снова получили формулу вида (3.20), 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 (3.20) имеет общий характер. Если по $2m + 1$ точкам подбирают с помощью МНК алгебраический полином степени p , то необходимо минимизировать критерий вида

$$Q(\Theta) = \sum_{t'=-m}^m (x(t') - \theta_0 - \theta_1 t' - \dots - \theta_p (t')^p)^2. \quad (3.23)$$

Минимизация $Q(\Theta)$ по Θ приводит к уравнениям относительно Θ , аналогичным (3.22), 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 (3.19) и вида вектора Y в данной задаче).

Оценкой $\hat{f}(t)$ неслучайной составляющей $f(t)$ анализируемого временного 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 табл. 3.4 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 3.4. Значения весовых коэффициентов в формуле метода скользящего среднего (3.20) для различных длин отрезков усреднения m и порядков аппроксимирующих полиномов p

m	p	w_{-m}	w_{-m+1}	...	w_0
m_0	0 или 1	$\frac{1}{2m_0+1}$	$\frac{1}{2m_0+1}$...	$\frac{1}{2m_0+1}$
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. (3.23)) и вытекающего из него вида θ_0 (определяющего значения весовых коэффициентов w_k в формуле (3.20)) можно вывести следующие важные свойства этих коэффициентов:

(1⁰) веса w_k симметричны относительно серединного значения w_0 , т. е. $w_k = w_{-k}$, $k = 1, 2, \dots, m$ (это следует из того факта, что они получены как функции сумм $\sum_{t'=-m}^m (t')^j x(t')$, которые сами симметричны);

(2⁰) сумма всех весов w_k равна единице, т. е. $\sum_{k=-m}^m w_k = 1$ (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 $\sum_{k=-m}^m w_k = 1$);

(3⁰) при одной и той же длине $N = 2m+1$ временного отрезка, по которому производится усреднение, веса w_k в формуле (3.20) для полиномов четной степени (т. е. для $p = 2l$, $l = 0, 1, 2, \dots$) будут теми же самыми, что и для полиномов степени, на единицу большей (т. е. для $p = 2l+1$); этот результат виден из того, что при составлении систем уравнений, аналогичных (3.22), 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 (т. е. 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:

$$\hat{f}(t) = \begin{cases} \hat{x}_1(t) = \hat{\theta}_0^{(1)} + \hat{\theta}_1^{(1)}t + \dots + \hat{\theta}_p^{(1)}t^p, & t = 1, 2, \dots, m; \\ \hat{x}_{n-2m}(t) = \hat{\theta}_0^{(n-2m)} + \hat{\theta}_1^{(n-2m)}t + \dots + \hat{\theta}_p^{(n-2m)}t^p, & t = n - m + 1, \dots, n. \end{cases} \quad (3.24)$$

В (3.24) $\hat{x}_1(t)$ — сглаживающий полином, построенный по 1-му интервалу сглаживания, т. е. — 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, т. е. — 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

$$t^* = \frac{k + (k + 2m) - 1}{2} = k + m - \frac{1}{2}, \quad k = 1, 2, \dots, n - 2m + 1. \quad (3.25)$$

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. (3.25)), 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 (т.е. для $t_1^* = k + m - \frac{1}{2}$ и $t_2^* = t_1^* + 1 = k + m + \frac{1}{2}$) и взять их среднее арифметическое, т.е.:

$$\hat{f}(k + m) = \frac{1}{2} \left[\hat{f} \left(k + m - \frac{1}{2} \right) + \hat{f} \left(k + m + \frac{1}{2} \right) \right].$$

Предположим, например, что имеются помесечные наблюдения с января по декабрь, «привязанные» к последнему дню каждого месяца. Простое скользящее среднее (т.е. МСС с $p = 0$ или 1) по двенадцати точкам (т.е. $N = 2m = 2 \cdot 6 = 12$) дает оценку значения неслучайной составляющей для *середины июля* (т.е. 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 (3.2)).

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 (3.17) 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. (3.2)) в ситуации, когда эти исходные случай-

ные остатки взаимно некоррелированы и гомоскедастичны, т. е.

$$\begin{aligned} \mathbf{E} \varepsilon(t) &\equiv 0, \\ \mathbf{E} (\varepsilon(t)\varepsilon(t+\tau)) &= \begin{cases} \sigma^2 & \text{при } \tau = 0, \\ 0 & \text{при } \tau \neq 0. \end{cases} \end{aligned} \quad (3.26)$$

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:

$$\begin{aligned} \mathbf{E} \tilde{\varepsilon}(t) &= E \tilde{\varepsilon}(t+m+1) = 0, \\ \mathbf{D} \tilde{\varepsilon}(t) &= \mathbf{D} \tilde{\varepsilon}(t+m+1) = \mathbf{D} \varepsilon(t) \sum_{k=1}^{2m+1} w_k^2 = \sigma^2 \sum_{k=1}^{2m+1} w_k^2, \end{aligned} \quad (3.27)$$

$$\begin{aligned} \text{cov} (\tilde{\varepsilon}(t+m+1), \tilde{\varepsilon}(t+m+1+\tau)) \\ = \mathbf{E} [w_1 \varepsilon(t+1) + w_2 \varepsilon(t+2) + \dots + w_{2m+1} \varepsilon(t+2m+1)] \\ \times [w_1 \varepsilon(t+1+\tau) + \dots + w_{2m+1} \varepsilon(t+2m+1+\tau)]. \end{aligned} \quad (3.28)$$

Учитывая взаимную некоррелированность исходных случайных остатков $\varepsilon(1), \varepsilon(2), \dots, \varepsilon(n)$ (см. (3.26)), соотношение (3.28) может быть преобразовано к виду

$$\text{cov} (\tilde{\varepsilon}(t+m+1), \tilde{\varepsilon}(t+m+1+\tau)) = \sigma^2 \sum_{j=1}^{2m+1-\tau} w_j w_{j+\tau}, \quad \tau = 1, 2, \dots, 2m. \quad (3.28')$$

Из (3.28') и (3.27), в частности, следует, что автокорреляция между преобразованными случайными остатками $\tilde{\varepsilon}(t)$, разнесенными друг от друга на τ единиц, равна

$$r_{\tilde{\varepsilon}} = \frac{\sum_{j=1}^{2m+1-\tau} w_j w_{j+\tau}}{\sum_{j=1}^{2m+1} w_j^2}, \quad \tau = 1, 2, \dots, 2m. \quad (3.29)$$

Отсюда следует, что последовательность сглаженных значений $\hat{f}(t)$ имеет ненулевые автокорреляции, определяемые формулой (3.29), вплоть до порядка $2m$ (то, что автокорреляции более высокого порядка равны нулю, следует непосредственно из (3.28)). Более того, для скользящих средних, наиболее часто применяемых на практике, величина $r_{\hat{f}}(1) = r_{\hat{f}}(1)$ 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 $\hat{\varepsilon}(t)$, будет меньше исходной дисперсии σ^2 , что следует из (3.27)), однако в нем могут появиться систематические колебания, обусловленные автокоррелированностью его последовательных значений (эффект Слущкого–Юла).

Метод экспоненциально взвешенного скользящего среднего (метод Брауна¹). 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 (3.23) 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

$$Q(f) = \sum_{k=0}^{t-1} \lambda^k (x(t-k) - f)^2 \rightarrow \min_f, \quad (3.30)$$

где λ — некоторое положительное число, меньшее единицы (т. е. $0 < \lambda < 1$).

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 (т. е. по мере роста k), — отсюда и название метода.

¹ См. Brown R. G. Smoothing, Forecasting and Prediction. Prentice-Hall, Englewood Cliffs, N.Y., 1963.

Решение оптимизационной задачи (3.30), т. е. дифференцирование $Q(f)$ по f , приравнивание полученной производной к нулю и решение образовавшегося таким образом уравнения относительно f дает:

$$\hat{f}(t) = \frac{1 - \lambda}{1 - \lambda^t} \sum_{k=0}^{t-1} \lambda^k x(t - k). \quad (3.31)$$

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 закреплен в точке $t = 1$) и, во-вторых, веса при $x(t - k)$ экспоненциально уменьшаются по мере «удаления в прошлое» (т. е. по мере роста k). Кроме того, формула (3.31) дает 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 (3.2) при переходе от исходных значений временного ряда $x(t)$ к сглаженным $\hat{f}(t)$ по формуле (3.31).

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. (3.31)), то очевидно:

$$\begin{aligned} \mathbf{E} \tilde{\varepsilon}(t) &= 0, \\ \mathbf{D} \tilde{\varepsilon}(t) &= \sigma^2 \frac{(1 - \lambda)^2}{(1 - \lambda^t)^2} \sum_{k=0}^{t-1} \lambda^{2k} = \sigma^2 \frac{(1 - \lambda)(1 + \lambda^t)}{(1 + \lambda)(1 - \lambda^t)}. \end{aligned} \quad (3.32)$$

Из (3.32) видно, что при значениях λ , не слишком близких к единице, и для достаточно удаленных от прошлого значениях t случайные остатки $\tilde{\varepsilon}(t)$ (а следовательно, и сглаженные значения $\hat{f}(t)$) подвержены существенно меньшему случайному разбросу, чем $x(t)$, т. е. ведут себя *более гладко*.

З а м е ч а н и е 4 (случай «бесконечно удаленного» прошлого).

Если анализируемый временной ряд достаточно длинный (т. е. n достаточно велико), то для представляющих главный интерес в прогнозе «свежих» значений временного ряда $x(t)$ (т. е. при t близких к n) можно приближенно допускать, что их прошлое как бы «уходит в бесконечность». Тогда соотношения (3.30), (3.31) и (3.32) могут быть заменены,

соответственно, соотношениями

$$Q(f) = \sum_{k=0}^{\infty} \lambda^k (x(t-k) - f)^2 \rightarrow \min_f, \quad (3.30')$$

$$\hat{f}(t) = (1 - \lambda) \sum_{k=0}^{\infty} \lambda^k x(t-k), \quad (3.31')$$

$$D\hat{E}(t) = \sigma^2 \frac{1 - \lambda}{1 + \lambda}. \quad (3.32')$$

Кроме того, в этом случае легко непосредственно проверить справедливость следующего полезного рекуррентного соотношения

$$\hat{f}(t) = \lambda \hat{f}(t-1) + (1 - \lambda)x(t). \quad (3.33)$$

3.3.3. Подбор порядка аппроксимирующего полинома с помощью метода последовательных разностей

Реализация *алгоритмически* 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. (3.18)). 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 разностям¹ $x(1), x(2), \dots, x(n)$, повторенный $p+1$ раз (т. е. переход к последовательным разностям порядка $p+1$), исключает*

¹ Если 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, т. е. $\Delta^2 x(t) = \Delta(\Delta x(t)) = \Delta x(t) - \Delta x(t-1) = (x(t) - x(t-1)) - (x(t-1) - x(t-2)) = x(t) - 2x(t-1) + x(t-2)$. 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

$$\Delta^k x(t) = x(t) - C_k^1 x(t-1) + C_k^2 x(t-2) - \dots + (-1)^k x(t-k), \quad (3.34)$$

где C_k^j — это число сочетаний из k элементов по j .

нелучайную составляющую (включая константу θ_0), оставляя элементы, выражающиеся только через остаточную случайную компоненту $\varepsilon(t)$.

Продemonстрируем справедливость этого утверждения на простых примерах.

1) Пусть $p = 1$, т. е.

$$x(t) = \theta_0 + \theta_1 t + \varepsilon(t), \quad t = 1, 2, \dots, n.$$

Тогда

$$\begin{aligned} \Delta x(t) &= x(t) - x(t-1) = \theta_1 + (\varepsilon(t) - \varepsilon(t-1)), \\ \Delta^2 x(t) &= \Delta x(t) - \Delta x(t-1) = \varepsilon(t) - 2\varepsilon(t-1) + \varepsilon(t-2). \end{aligned}$$

Мы видим, что если нелучайная составляющая $f(t)$ временного ряда выражается *линейной* функцией времени (т. е. является алгебраическим полиномом времени порядка $p = 1$), то последовательность *вторых* разностей $\Delta^2 x(t)$ анализируемого временного ряда уже не будет содержать нелучайной 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 $p = 2$, т. е.

$$x(t) = \theta_0 + \theta_1 t + \theta_2 t^2 + \varepsilon(t), \quad t = 1, 2, \dots, n.$$

Тогда

$$\begin{aligned} \Delta x(t) &= \theta_0 + \theta_1 + \theta_2 t^2 - \theta_0 - \theta_1(t-1) - \theta_2(t-1)^2 + (\varepsilon(t) - \varepsilon(t-1)) \\ &= \theta_1 - \theta_2 + 2\theta_2 t + (\varepsilon(t) - \varepsilon(t-1)), \\ \Delta^2 x(t) &= \Delta x(t) - \Delta x(t-1) = 2\theta_2 + (\varepsilon(t) - 2\varepsilon(t-1) + \varepsilon(t-2)), \\ \Delta^3 x(t) &= \Delta^2 x(t) - \Delta^2 x(t-1) = \varepsilon(t) - 3\varepsilon(t-1) + 3\varepsilon(t-2) - \varepsilon(t-3), \\ & \quad t = 4, 5, \dots, n. \end{aligned}$$

Мы вновь получили подтверждение сформулированного выше математического факта: при нелучайной составляющей временного ряда, выраженной в виде алгебраического полинома времени t порядка $p = 2$, последовательность *третьих* разностей $\Delta^3 x(t)$ уже не будет содержать нелучайной составляющей, а будет представлена только в терминах случайных остатков $\varepsilon(t)$ разложения (3.2).

Теперь рассмотрим общий случай. Пусть

$$x(t) = \theta_0 + \theta_1 t + \dots + \theta_p t^p + \varepsilon(t), \quad t = 1, 2, \dots, n, \quad (3.35)$$

где случайные остатки $\varepsilon(t)$ удовлетворяют требованиям (3.26).

Последовательно переходя к разностям порядка $p + 1$, с учетом (3.34) получаем:

$$\begin{aligned} \Delta^{p+1} x(t) &= \varepsilon(t) - C_{p+1}^1 \varepsilon(t-1) \\ &\quad + C_{p+1}^2 \varepsilon(t-2) - \dots + (-1)^{p+1} \varepsilon(t-p-1), \end{aligned} \quad (3.36)$$

$$t = p+2, p+3, \dots, n.$$

Вычислим среднее значение и дисперсию членов последовательности $\Delta^{p+1} x(t)$:

$$E(\Delta^{p+1} x(t)) = 0, .$$

$$\begin{aligned} D(\Delta^{p+1} x(t)) &= D x(t) [1 + (C_{p+1}^1)^2 + (C_{p+1}^2)^2 + \dots + (C_{p+1}^p)^2 + 1] \\ &= \sigma^2 C_{2(p+1)}^{p+1}. \end{aligned} \quad (3.37)$$

При получении выражения для $D(\Delta^{p+1} x(t))$ 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

$$\begin{aligned} (1+z)^{p+1} (1+z)^{p+1} &= (1 + C_{p+1}^1 z + \dots + C_{p+1}^p z^p + C_{p+1}^{p+1} z^{p+1}) \\ &\quad \times (1 + C_{p+1}^1 z + \dots + C_{p+1}^p z^p + z^{p+1}) \end{aligned}$$

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 $E\xi$ 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 $E\xi \neq 0$, 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 $D\xi$* . Возвращаясь к последовательному переходу к разностям $\Delta^k x(t)$, $k = 1, 2, \dots, p+1$, — в общем случае (3.35), 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 (3.37), а следовательно, при $k \geq p+1$ можно утверждать, что:

$$E(\Delta^k x(t)) = 0 \quad \text{и} \quad \sigma^2 = D(\Delta^k x(t))/C_{2(p+1)}^{p+1}. \quad (3.37')$$

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

$$\hat{\sigma}^2(k) = \frac{\frac{1}{n-k} \sum_{t=1}^{n-k} (\Delta^k x(t))^2}{C_{2k}^k}. \quad (3.38)$$

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 (3.38) 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, т. е. $p = k_0 - 1$.

Для упрощения использования описанного метода в табл. 3.5 приведены значения вспомогательных коэффициентов C_{2k}^k для $k = 1, 2, \dots, 10$.

Таблица 3.5. 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.

3.4. Модели стационарных временных рядов и их идентификация

В п. 3.2 был введен в рассмотрение *класс стационарных временных рядов*, в рамках которого подбирается модель, пригодная для описания поведения случайных остатков $\varepsilon(t)$ анализируемого временного ряда (3.2). В данном пункте мы займемся изучением определенного *набора линейных параметрических моделей* из этого класса, в том числе *методами идентификации этих моделей*, т. е. методами статистического оценивания участвующих в описании модели параметров по имеющейся в нашем распоряжении траектории (3.1') анализируемого временного ряда.

Таким образом, речь пока идет о моделировании не самих анализируемых временных рядов, а лишь их случайных остатков $\varepsilon(t)$, т. е. того, что мы получаем после элиминирования из исходного временного ряда $x(t)$ его неслучайной составляющей (3.17) (см. п. 3.3.2). 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. 2.9.2),

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, т. е. при всех значениях t , равно нулю, т. е. $E\varepsilon(t) \equiv 0$. Временные последовательности, образующие так называемый «белый шум», будем обозначать с помощью $\delta(t)$. Это значит, что

$$E\delta(t) \equiv 0, \quad (3.39)$$

$$E(\delta(t)\delta(t \pm \tau)) = \begin{cases} \sigma_0^2 & \text{при } \tau = 0 \\ 0 & \text{при } \tau \neq 0, \end{cases} \quad (3.40)$$

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 виде взвешенной суммы настоящего и прошлых значений белого шума, а именно:

$$\varepsilon(t) = \delta(t) + \beta_1\delta(t-1) + \beta_2\delta(t-2) + \dots = \sum_{k=0}^{\infty} \beta_k\delta(t-k), \quad (3.41)$$

$$\text{где } \beta_0 = 1 \text{ и } \sum_{k=0}^{\infty} \beta_k^2 < \infty.$$

Таким образом, белый шум можно рассматривать как серию импульсов, в широком классе реальных ситуаций генерирующих случайные остатки анализируемого временного ряда.

Существует *эквивалентная соотношению (3.41) 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), т. е.

$$\varepsilon(t) = \pi_1\varepsilon(t-1) + \pi_2\varepsilon(t-2) + \dots + \delta(t) = \sum_{k=1}^{\infty} \pi_k\varepsilon(t-k) + \delta(t). \quad (3.41')$$

При этом весовые коэффициенты π_1, π_2, \dots связаны определенными условиями, обеспечивающими стационарность ряда $\varepsilon(t)$. Переход от формы (3.41') к форме (3.41) осуществляется с помощью последовательной (и

повторенной бесконечное число раз!) подстановки в правую часть (3.41') вместо $\varepsilon(t-1), \varepsilon(t-2), \dots$ их выражений, вычисленных в соответствии с (3.41') для моментов времени $t-1, t-2$ и т. д.

Введем в рассмотрение, наряду с общим линейным процессом вида (3.41) или (3.41'), процесс *смешанного* типа, в представлении которого присутствуют как авторегрессионные члены самого процесса, так и скользящее суммирование элементов белого шума:

$$\varepsilon(t) = \sum_{k=1}^p \pi_k \varepsilon(t-k) + \delta(t) + \sum_{j=1}^q \beta_j \delta(t-j). \quad (3.41'')$$

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 (3.41'') (обоснование такого представления можно найти, например, в книге [Бокс Дж., Дженкинс Г., вып. 1], с. 67, 98–99):

$$p(\tilde{\omega}) = 2\sigma_0^2 \frac{|B(e^{-i2\pi\tilde{\omega}})|^2}{|A(e^{-i2\pi\tilde{\omega}})|^2}, \quad 0 \leq \tilde{\omega} \leq \frac{1}{2}, \quad (3.9^a)$$

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.

Теперь перейдем к рассмотрению *конкретных* линейных моделей стационарного временного ряда.

3.4.1. Модели авторегрессии порядка p (АР(p)-модели)¹

Начнем описание этого параметрического семейства моделей с анализа простейших частных случаев.

¹ В англоязычной специальной литературе эти модели известны как *AutoRegressive processes of order p* и кратко обозначаются *AR(p)-models*.

Модель авторегрессии 1-го порядка — АР(1) (марковский процесс). Эта модель представляет собой простейший вариант линейного авторегрессионного процесса типа (3.41'), когда все коэффициенты π ; кроме первого равны нулю. Соответственно, она может быть определена выражением

$$\varepsilon(t) = \alpha\varepsilon(t-1) + \delta(t), \quad (3.42)$$

где α — некоторый числовой коэффициент, не превосходящий по абсолютной величине единицу ($|\alpha| < 1$), а $\delta(t)$ — последовательность случайных величин, образующая белый шум (т. е. последовательность случайных величин, удовлетворяющая соотношениям (3.39)–(3.40)).

Внимательный читатель вспомнит, что мы уже имели дело с этой моделью. Ведь именно соотношением (3.42) были связаны между собой регрессионные остатки в *обобщенной модели множественной регрессии с автокоррелированными остатками* (см. п. 2.8, соотношения (2.106)–(2.107), в которых роль параметра α выполнял параметр ρ). В специальной литературе последовательности $\varepsilon(t)$, удовлетворяющие соотношению (3.42), называют также *марковским процессом*. Мы не будем повторять здесь выкладки (2.108)–(2.110), из которых следовало:

$$\mathbf{E}\varepsilon(t) \equiv 0, \quad (3.43^a)$$

$$r(\varepsilon(t), \varepsilon(t \pm k)) = \alpha^k, \quad (3.43^b)$$

$$\mathbf{D}\varepsilon(t) = \frac{\sigma_0^2}{1 - \alpha^2}, \quad (3.43^b)$$

$$\text{cov}(\varepsilon(t), \varepsilon(t \pm k)) = \alpha^k \mathbf{D}\varepsilon(t). \quad (3.43^f)$$

Проанализируем некоторые свойства временной последовательности, представимой в виде (3.42). Из представления $\varepsilon(t)$ в виде (2.108) следует, что $\varepsilon(t)$ зависит от $\delta(t)$ и всех предшествующих δ , но не зависит от будущих значений δ , т. е. от $\delta(t+1)$, $\delta(t+2)$ и т. д.

Заметим, что формула (2.108) 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 $\varepsilon(t)$, представимая в (2.108) как взвешенная сумма большого числа случайных слагаемых, должна быть распределена (в соответствии с центральной предельной теоремой, см. том 1, п. 4.3.1) *нормально*. Однако одно из условий Ц.П.Т. нарушено: веса в этом среднем — величины различного порядка.

Одно важное следствие (3.43^b) 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 (3.43^b) $\alpha = r(\varepsilon(t), \varepsilon(t \pm$

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. 3.8). Может быть вычислено среднее расстояние $d_{\text{ср}}$ между «пиками»¹ такой осциллирующей ломаной (см. [Кендэл М.], 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 (т.е. при $\alpha = 0$) это среднее расстояние оказывается равным 3,0 (т.к. $\arccos(-\frac{1}{2}) = \frac{2}{3}\pi$); для примеров, представленных на рис. 3.8, 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 (3.43), основные характеристики процесса авторегрессии 1-го порядка и попробуем оценить параметры этой модели (α и σ_0^2) по имеющейся в нашем распоряжении реализации (3.1') анализируемого временного ряда $x(t)$.

Условие стационарности ряда (3.42) определяется требованием к коэффициенту α :

$$|\alpha| < 1, \quad (3.42')$$

или, что то же, корень z_0 уравнения $1 - \alpha z = 0$ должен быть по абсолютной величине больше единицы².

¹ Диагностика модели по числу пиков в анализируемом временном ряду или по среднему расстоянию между пиками основана на том же подходе, который был использован в п. 3.3.1 при построениях критериев для проверки гипотезы о том, что исследуемый временной ряд представляет собой последовательность независимых и одинаково распределенных случайных величин (см. (3.12)). 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.

² Уравнение $1 - \alpha z = 0$ 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).

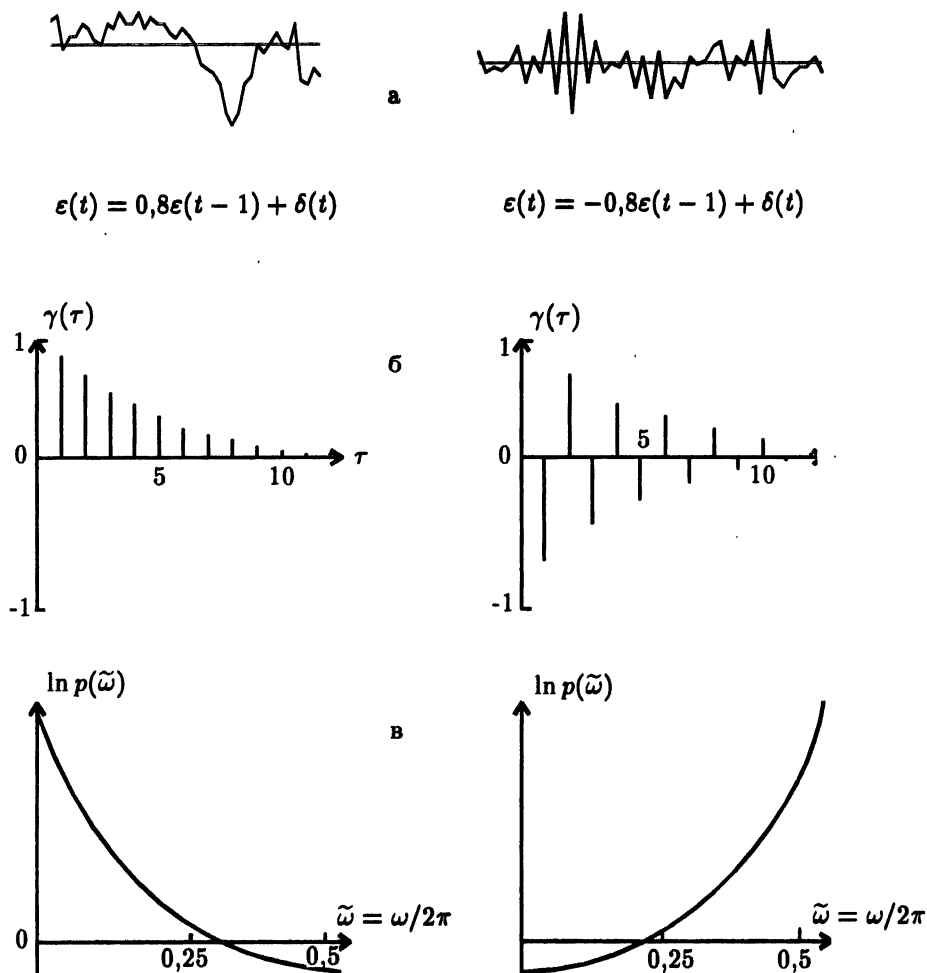


Рис. 3.8. Реализация процессов авторегрессии первого порядка (а), соответствующие им автокорреляционные функции (б) и логарифмы плотности (в)

Автокорреляционная функция марковского процесса определяется соотношением (3.43^б):

$$r(\tau) = r(\varepsilon(t), \varepsilon(t \pm \tau)) = \alpha^\tau. \quad (3.44)$$

Отсюда же, в частности, следует простая вероятностная интерпретация параметра α :

$$\alpha = r(1) = r(\varepsilon(t), \varepsilon(t \pm 1)),$$

т. е. значение α определяет величину коэффициента парной корреляции между двумя соседними членами ряда $\varepsilon(t)$.

Из (3.44) видно, что степень тесноты корреляционной связи между членами последовательности (3.42) 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 (3.7)–(3.8). 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 (3.42) может быть подсчитана либо по формуле (3.9^а), либо по формуле (3.9) 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. 3.8 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, т. е. статистическое оценивание ее параметров α и σ_0^2 по имеющейся реализации (3.1') *исходного* анализируемого временного ряда $x(t)$ (а не его остатков, которые являются *ненаблюдаемыми*), основана на соотношениях (3.43) (при $k = 0$ и $k = 1$) и может быть осуществлена с помощью метода моментов (см. п. 7.5.2). Для этого следует предварительно решить задачу выделения неслучайной составля-

ющей $\hat{f}(t)$ (см. п. 3.3), что позволит оперировать в дальнейшем остатками («невязками»)

$$\hat{\varepsilon}(t) = x(t) - \hat{f}(t). \quad (3.44')$$

Затем подсчитывается выборочная дисперсия $\hat{\gamma}(0)$ остатков по формуле

$$\hat{\gamma}(0) = \frac{1}{N} \sum_{t=1}^N (\hat{\varepsilon}(t) - \bar{\varepsilon})^2, \quad (3.45)$$

где $\bar{\varepsilon} = (\sum_{t=1}^N \hat{\varepsilon}(t))/N$, а «невязки» $\hat{\varepsilon}(t)$ вычислены по формуле (3.44').

Оценку $\hat{\alpha}$ параметра α получаем с помощью формулы (3.43^b), подставляя в нее вместо коэффициента корреляции $r(\varepsilon(t), \varepsilon(t \pm 1))$ его выборочное значение $\hat{r}(\varepsilon(t), \varepsilon(t + 1))$, т. е.

$$\hat{\alpha} = \frac{\frac{1}{N-1} \sum_{t=1}^{N-1} (\hat{\varepsilon}(t) - \bar{\varepsilon})(\hat{\varepsilon}(t+1) - \bar{\varepsilon})}{\hat{\gamma}(0)}. \quad (3.46)$$

Наконец, оценка $\hat{\sigma}_0^2$ параметра σ_0^2 основана на соотношении (3.43^a), 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» (т. е. $\bar{\varepsilon} = (\sum_{t=1}^N \hat{\varepsilon}(t))/N = 0$). В этих случаях в формулах (3.45) и (3.46) значение $\bar{\varepsilon}$ следует положить равным нулю.

Обращаем внимание читателя на следующий факт: *прямая* оценка значений автокорреляционной функции (т. е. оценка ее значений с помощью формулы (3.6')) 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 (3.6') 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 — AP(2) (процессы Юла). Эта модель, как и AP(1), представляет собой частный случай авторегрессионного процесса типа (3.41'), когда все коэффициенты π_j в

правой части (3.41'), кроме первых двух, равны нулю. Соответственно, она может быть определена выражением

$$\varepsilon(t) = \alpha_1 \varepsilon(t-1) + \alpha_2 \varepsilon(t-2) + \delta(t), \quad (3.47)$$

где последовательность случайных величин $\delta(1), \delta(2), \dots$ образует белый шум (см. выше).

Последовательно подставляя (бесконечное число раз) в правую часть (3.47) вместо $\varepsilon(t-1), \varepsilon(t-2), \dots$ их выражения, вычисленные по формуле (3.47), мы убедимся в том, что как и в $AR(1)$ -модели, $\varepsilon(t)$ зависит от $\delta(t), \delta(t-1), \delta(t-2), \dots$, но не зависит от будущих значений δ , т. е. от $\delta(t+1), \delta(t+2), \dots$.

Умножая обе части (3.47) 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 $\gamma(0)$:

$$\begin{cases} r(1) - \alpha_1 - \alpha_2 r(1) = 0, \\ r(2) - \alpha_1 r(1) - \alpha_2 = 0. \end{cases} \quad (3.48)$$

Разрешая эту систему относительно α_1 и α_2 , имеем:

$$\begin{aligned} \alpha_1 &= \frac{r(1)(1-r(2))}{1-r^2(1)}, \\ \alpha_2 &= \frac{r(2)-r^2(1)}{1-r^2(1)}. \end{aligned} \quad (3.49)$$

Уравнения (3.48) могут быть решены относительно $r(1)$ и $r(2)$:

$$\begin{aligned} r(1) &= \frac{\alpha_1}{1-\alpha_2}, \\ r(2) &= \alpha_2 + \frac{\alpha_1^2}{1-\alpha_2}. \end{aligned} \quad (3.50)$$

Обобщая использованный прием, домножим обе части (3.47) 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 $\gamma(0)$):

$$r(k) = \alpha_1 r(k-1) + \alpha_2 r(k-2). \quad (3.51)$$

Рекуррентная формула (3.51) 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 (3.47) 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

$$\sigma_0^2 = \gamma(0) - \alpha_1 \gamma(1) - \alpha_2 \gamma(2), \quad (3.52)$$

или в терминах $\gamma(0) = \mathbf{D}\varepsilon(t)$, α_1 , α_2 , $r(1)$ и $r(2)$:

$$\sigma_0^2 = \gamma(0)(1 - \alpha_1 r(1) - \alpha_2 r(2)) = \gamma(0) \frac{1 + \alpha_2}{1 - \alpha_2} [(1 - \alpha_2)^2 - \alpha_1^2]. \quad (3.52')$$

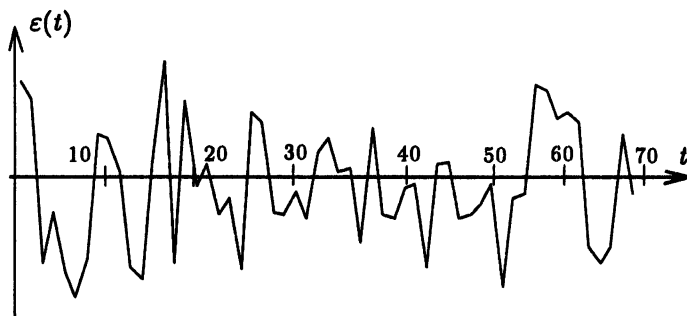


Рис. 3.9. Временной ряд, генерируемый моделью авторегрессии второго порядка

На рис. 3.9 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 $\varepsilon(t)$, бóльшим, чем два соседних значения), которое подсчитывается по значениям α_1 и α_2 по формуле (см. [Кендэл М.], 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. 3.9, 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* (3.47) 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 (3.50)):

$$\begin{cases} -1 < \frac{\alpha_1}{1 - \alpha_2} < 1 \\ -1 < \alpha_2 + \frac{\alpha_1^2}{1 - \alpha_2} < 1. \end{cases} \quad (3.53)$$

Отсюда получаем следующие необходимые и достаточные условия стационарности ряда (3.47):

$$\begin{cases} |\alpha_1| < 2, \\ \alpha_2 < 1 - |\alpha_1|. \end{cases} \quad (3.53')$$

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 (3.53') и выражений (3.50) 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 (3.50), а значения $r(\tau)$ для $\tau = 3, 4, \dots$ вычисляются с помощью рекуррентного соотношения (3.51).

На рис. 3.10 представлен график автокорреляционной функции для рассмотренного выше примера модели авторегрессии 2-го порядка (см. также рис. 3.9).

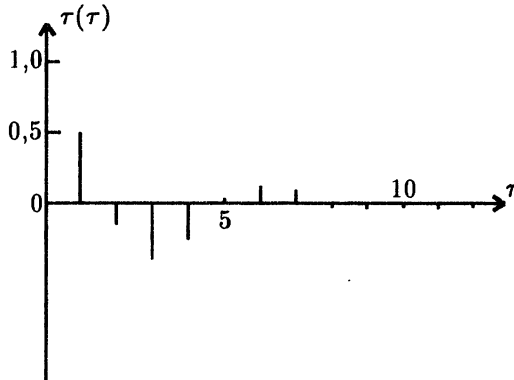


Рис. 3.10. Автокорреляционная функция процесса авторегрессии второго порядка $\varepsilon(t) = 0,75\varepsilon(t-1) - 0,50\varepsilon(t-2) + \delta(t)$

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 (3.7)–(3.8) для $\tau = 2, 3$ и с использованием рекуррентных формул (11.32') для $\tau > 3$. Действительно, подставляя в формулу (3.7) значения $r(1)$ и $r(2)$, выраженные через α_1 и α_2 с помощью (3.50), получаем

$$r_{\text{част}}(2) = \frac{r(2) - r^2(1)}{1 - r^2(1)} = \alpha_2.$$

Далее, числитель в формуле (3.8) 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 (3.50) и (3.51).

Спектральная плотность $p(\tilde{\omega})$ процесса Юла может быть вычислена с помощью общей формулы (3.9^а):

$$\begin{aligned} p(\tilde{\omega}) &= \frac{2\sigma_0^2}{|1 - \alpha_1 e^{-i2\pi\tilde{\omega}} - \alpha_2 e^{-i4\pi\tilde{\omega}}|^2} \\ &= \frac{2\sigma_0^2}{1 + \alpha_1^2 + \alpha_2^2 - 2\alpha_1(1 - \alpha_2) \cos(2\pi\tilde{\omega}) - 2\alpha_2 \cos(4\pi\tilde{\omega})}, \quad (3.54) \\ &0 \leq \tilde{\omega} \leq \frac{1}{2}. \end{aligned}$$

На рис. 3.11 представлен график спектральной плотности рассмотренного ранее (см. рис. 3.9) 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 (т. к. $T_0 = 1/\tilde{\omega}_0$). 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 .

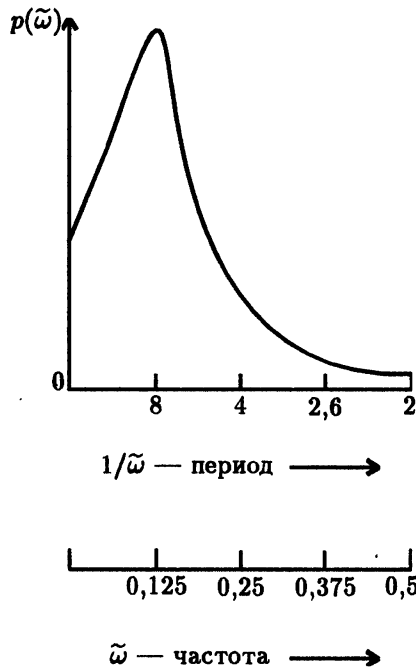


Рис. 3.11. Спектральная плотность процесса авторегрессии второго порядка

Идентификация модели авторегрессии 2-го порядка основана на соотношениях (3.49)~ (3.52'), 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 (3.4'') и (3.6'):

$$\hat{\gamma}(0) = \frac{1}{N} \sum_{t=1}^N (\hat{\varepsilon}(t) - \bar{\varepsilon})^2,$$

$$\hat{r}(k) = \frac{\frac{1}{N-k} \sum_{t=1}^{N-k} (\hat{\varepsilon}(t) - \bar{\varepsilon})(\hat{\varepsilon}(t+k) - \bar{\varepsilon})}{\hat{\gamma}(0)}, \quad k = 1, 2$$

(если из техники построения неслучайной составляющей $\hat{f}(t)$ автоматически следует равенство $\sum_{t=1}^N \hat{\varepsilon}(t) = 0$, то значение $\hat{\varepsilon}$ 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 (3.49) 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 (3.52'), 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 (3.51) на базе начальных значений $\hat{r}(1)$ и $\hat{r}(2)$.

Модели авторегрессии p -го порядка — $AR(p)$ ($p \geq 3$). Эти модели, образуя подмножество в классе *общих линейных моделей* (3.41'), сами составляют достаточно широкий класс моделей авторегрессии, включающий в себя в качестве частных случаев ранее рассмотренные $AR(1)$ - и $AR(2)$ -модели. Если в общей линейной модели (3.41') полагать все параметры π_j , *кроме первых p коэффициентов*, равными нулю, то мы приходим к определению $AR(p)$ -модели. Итак, авторегрессионной моделью порядка p называется модель вида:

$$\varepsilon(t) = \sum_{j=1}^p \alpha_j \varepsilon(t-j) + \delta(t), \quad (3.55)$$

где последовательность случайных величин $\delta(1), \delta(2); \dots$ образует белый шум, т. е. удовлетворяет условиям (3.39)–(3.40).

Условия стационарности процесса, генерируемого моделью (3.55), также формулируются в терминах корней его характеристического уравнения

$$1 - \alpha_1 z - \alpha_2 z^2 - \dots - \alpha_p z^p = 0. \quad (3.56)$$

Для стационарности процесса (3.55) необходимо и достаточно, чтобы все корни уравнения (3.56) лежали бы вне единичного круга, т. е. превосходили бы по модулю единицу.

Автокорреляционная функция процесса (3.55) так же, как и для $AR(2)$ -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 (3.55) на $\varepsilon(t-\tau)$ ($\tau > p$). Получим $\varepsilon(t-\tau)\varepsilon(t) = \alpha_1 \varepsilon(t-\tau)\varepsilon(t-1) + \alpha_2 \varepsilon(t-\tau)\varepsilon(t-2) + \dots + \alpha_p \varepsilon(t-\tau)\varepsilon(t-p) + \varepsilon(t-\tau)\delta(t)$. Переходя к

математическим ожиданиям величин, участвующих в этом соотношении, получаем:

$$\gamma(\tau) = \alpha_1 \gamma(\tau - 1) + \alpha_2 \gamma(\tau - 2) + \dots + \alpha_p \gamma(\tau - p). \quad (3.57)$$

Отметим, что $E(\varepsilon(t - k)\delta(t)) = 0$ при $k > 0$, т.к. $\varepsilon(t - k)$, выраженное через δ , может включать лишь импульсы $\delta(j)$ для $j \leq t - k$ (т.е. $\varepsilon(t)$ не зависит от будущих, по отношению к t , значений δ). Поделив все члены (3.57) на $\gamma(0)$, находим искомое рекуррентное соотношение, позволяющее последовательно вычислять любой элемент автокорреляционной функции процесса $\varepsilon(t)$ по первым p ее элементам $r(1), r(2), \dots, r(p)$:

$$r(\tau) = \alpha_1 r(\tau - 1) + \alpha_2 r(\tau - 2) + \dots + \alpha_p r(\tau - p), \quad (3.58)$$

$$\tau = p + 1, p + 2, \dots$$

Частная автокорреляционная функция $r_{\text{част}}(\tau) = r(\varepsilon(t), \varepsilon(t + \tau) | \varepsilon(t + 1) = \varepsilon(t + 2) = \dots = \varepsilon(t + \tau - 1) = 0)$ процесса (3.55) будет иметь ненулевые значения лишь при $\tau \leq p$; все значения $r_{\text{част}}(\tau)$ при $\tau > p$ будут нулевыми (доказательство этого факта см., например, в книге [Бокс Дж., Дженкинс Г., вып. 1], с. 81). Это свойство частной автокорреляционной функции $AR(p)$ -процесса используется, в частности, при подборе порядка p в модели авторегрессии для конкретных анализируемых временных рядов. Если, например, подсчитанные по исходным данным все частные коэффициенты автокорреляции, начиная с порядка k , статистически незначимо отличаются от нуля, то порядок модели авторегрессии, подбираемой для анализируемого временного ряда, естественно определить числом $p = k - 1$.

Спектральная плотность $p(\tilde{\omega})$ процесса авторегрессии p -го порядка определяется с помощью формулы (3.9^а):

$$p(\tilde{\omega}) = \frac{2\sigma_0^2}{|1 - \alpha_1 e^{-i2\pi\tilde{\omega}} - \alpha_2 e^{-i4\pi\tilde{\omega}} - \dots - \alpha_p e^{-i2\pi p\tilde{\omega}}|^2}, \quad (3.59)$$

$$0 \leq \tilde{\omega} \leq \frac{1}{2}$$

(здесь $i = \sqrt{-1}$ — мнимая единица).

Идентификация модели авторегрессии p -го порядка так же, как и при анализе моделей 1-го и 2-го порядков, основана на соотношениях, связывающих между собой неизвестные параметры модели и автокорреляции анализируемого ряда. Для вывода этих соотношений последовательно подставим в (3.58) значения $\tau = 1, 2, \dots, p$. Получим систему линейных

уравнений относительно $\alpha_1, \alpha_2, \dots, \alpha_p$:

$$\begin{cases} r(1) = \alpha_1 + \alpha_2 r(1) + \dots + \alpha_p r(p-1), \\ r(2) = \alpha_1 r(1) + \alpha_2 + \dots + \alpha_p r(p-2), \\ \dots \dots \dots \\ r(p) = \alpha_1 r(p-1) + \alpha_2 r(p-2) + \dots + \alpha_p. \end{cases} \quad (3.60)$$

Они обычно называются *уравнениями Юля-Уокера*¹. Мы уже рассматривали частные случаи этих уравнений в АР(1)- и АР(2)-моделях (см. (3.43⁶) и (3.48)). 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:

$$\alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{pmatrix}, \quad r = \begin{pmatrix} r(1) \\ r(2) \\ \vdots \\ r(p) \end{pmatrix},$$

$$R = \begin{pmatrix} 1 & r(1) & r(2) & \dots & r(p-1) \\ r(1) & 1 & r(1) & \dots & r(p-2) \\ \dots & \dots & \dots & \dots & \dots \\ r(p-1) & r(p-2) & r(p-3) & \dots & 1 \end{pmatrix}.$$

Тогда система (3.60) может быть представлена в форме

$$R\alpha = r, \quad (3.60')$$

а ее решение, соответственно, будет иметь вид:

$$\alpha = R^{-1}r. \quad (3.61)$$

Для вывода соотношения, необходимого при построении оценки $\hat{\sigma}_0^2$ параметра $\sigma_0^2 = D\delta(t)$, домножим обе части (3.55) на $\varepsilon(t)$:

$$\begin{aligned} \varepsilon^2(t) &= \alpha_1 \varepsilon(t-1)\varepsilon(t) + \alpha_2 \varepsilon(t-2)\varepsilon(t) + \dots + \alpha_p \varepsilon(t-p)\varepsilon(t) + \delta(t) \\ &\times (\alpha_1 \varepsilon(t-1) \\ &\quad + \alpha_2 \varepsilon(t-2) + \dots + \alpha_p \varepsilon(t-p) + \delta(t)). \end{aligned}$$

¹ См. Yule G.U. On a method of investigating periodicities in disturbed series. «Phil. Trans.», A226, p. 227, 1927. Walker G. On periodicity in series of related terms. «Proc. Royal Soc.», A131, p. 518, 1931.

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:

$$\sigma_0^2 = \gamma(0)(1 - \alpha_1 r(1) - \alpha_2 r(2) - \dots - \alpha_p r(p)). \quad (3.62)$$

Оценка $\hat{\sigma}_0^2$ для параметра σ_0^2 получается с помощью (3.62) заменой всех участвующих в правой части величин их оценками $(\hat{\gamma}(0), \hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_p)$.

3.4.2. Модели скользящего среднего порядка q (СС(q)-модели)

Рассмотрим частный случай общего линейного процесса (3.41), когда только первые q из весовых коэффициентов β_j ненулевые. В этом случае процесс имеет вид

$$\varepsilon(t) = \delta(t) - \theta_1 \delta(t - 1) - \theta_2 \delta(t - 2) - \dots - \theta_q \delta(t - q), \quad (3.63)$$

где символы $-\theta_1, -\theta_2, \dots, -\theta_q$ используются для обозначения *конечного* набора параметров β , участвующих в (3.41). Процесс (3.63) называется *моделью скользящего среднего порядка q* и сокращенно обозначается СС(q)¹. Особенно распространены в практике статистического моделирования СС-модели первого ($q = 1$) и второго ($q = 2$) порядка.

$$\text{СС (1): } \varepsilon(t) = \delta(t) - \theta \delta(t - 1), \quad (3.64)$$

$$\text{СС (2): } \varepsilon(t) = \delta(t) - \theta_1 \delta(t - 1) - \theta_2 \delta(t - 2). \quad (3.65)$$

Двойственность в представлении АР- и СС-моделей и понятие обратимости СС-модели. В начале п. 3.4 были приведены две эквивалентные формы записи *общего линейного процесса*. Из (3.41) и (3.41') видно, что один и тот же общий линейный процесс может быть

¹ Термин «скользящее среднее» здесь употреблен в несколько ином смысле, чем в п. 3.3.2, 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 МА (q)-models.

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. (2.108)), где АР(1)-модель $\varepsilon(t) = \alpha\varepsilon(t-1) + \delta(t)$ может быть представлена также в виде

$$\varepsilon(t) = \delta(t) + \alpha\delta(t-1) + \alpha^2\delta(t-2) + \dots,$$

т. е. 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).

Рассмотрим теперь СС(1)-модель и попробуем выразить $\delta(t)$ через текущее и все прошлые значения $\varepsilon(t)$. Уединяя $\delta(t)$ в соотношении (3.64), имеем:

$$\begin{aligned} \delta(t) &= \varepsilon(t) + \theta\delta(t-1) = \varepsilon(t) + \theta[\varepsilon(t-1) + \theta\delta(t-2)] \\ &= \varepsilon(t) + \theta\varepsilon(t-1) + \theta^2[\varepsilon(t-2) + \theta\delta(t-3)] \\ &= \varepsilon(t) + \theta\varepsilon(t-1) + \theta^2\varepsilon(t-2) + \theta^3[\varepsilon(t-3) + \theta\delta(t-4)] \\ &\dots\dots\dots \\ &= \varepsilon(t) + \theta\varepsilon(t-1) + \theta^2\varepsilon(t-2) + \theta^3\varepsilon(t-3) + \dots \end{aligned} \quad (3.66)$$

Соотношение (3.66) можно переписать в виде

$$\varepsilon(t) = - \sum_{k=1}^{\infty} \theta^k \varepsilon(t-k) + \delta(t). \quad (3.66')$$

Это означает, что ряд $\varepsilon(t)$, сгенерированный СС(1)-моделью (3.64), может быть представлен также в *виде модели авторегрессии бесконечного порядка*. Заметим, что, в отличие от АР(p)-моделей, в СС(q)-моделях не требуется накладывать никаких ограничений на параметры $\theta_1, \theta_2, \dots, \theta_q$ для обеспечения их стационарности: *ряды вида (3.63) стационарны при любых вещественных значениях параметров θ_j ($j = 1, 2, \dots, q$)*. Однако, если, например, в модели СС(1) параметр θ по абсолютной величине больше или равен единице ($|\theta| \geq 1$), то текущее значение анализируемого

ряда $\varepsilon(t)$ в соответствии с (3.66') 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 (3.66') образовывали бы сходящийся ряд, т. е. чтобы $|\theta| < 1$. Это требование равносильно условию, сформулированному в терминах *характеристического уравнения модели* (3.64) $1 - \theta z = 0$ и заключающемуся в том, чтобы корень этого характеристического уравнения был бы по модулю больше единицы («лежал бы вне единичного круга»). Итак, *СС(1)-модель называется обратимой, если в обратном разложении (3.66') 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 СС-моделей не имеют никаких связей с условиями стационарности временного ряда.

Обобщим этот результат на случай СС(q)-модели при произвольном конечном значении q . Уединяя $\delta(t)$ в соотношении (3.63), имеем:

$$\delta(t) = \varepsilon(t) + \theta_1 \delta(t-1) + \theta_2 \delta(t-2) + \dots + \theta_q \delta(t-q). \quad (3.63')$$

Действуя аналогично тому, как мы это делали при выводе соотношения (3.66), т. е. заменяя бесконечное число раз $\delta(t-1), \delta(t-2), \dots, \delta(t-q)$ в правой части (3.63') их выражениями, вычисленными по формуле (3.63'), получим разложение вида

$$\delta(t) = \varepsilon(t) - \pi_1 \varepsilon(t-1) - \pi_2 \varepsilon(t-2) - \dots, \quad (3.67)$$

в котором коэффициенты π_j ($j = 1, 2, \dots$) определенным образом выражаются через параметры $\theta_1, \theta_2, \dots, \theta_p$. Соотношение (3.67) может быть записано в виде модели авторегрессии бесконечного порядка (т. е. в виде *обращенного разложения*)

$$\varepsilon(t) = \sum_{j=1}^{\infty} \pi_j \varepsilon(t-j) + \delta(t). \quad (3.67')$$

Можно показать (см., например, [Бокс Дж., Дженкинс Г., вып. 1], с. 84), что *условие обратимости СС(q)-модели* (т. е. условие сходимости ряда $\sum_{j=1}^{\infty} \pi_j$) формулируется в терминах характеристического уравнения

модели (3.63) следующим образом:

$$\left. \begin{array}{l} \bullet \text{ все корни } z_1, z_2, \dots, z_q \text{ характеристического} \\ \text{уравнения } 1 - \theta_1 z - \theta_2 z^2 - \dots - \theta_q z^q = 0 \\ \text{должны лежать вне единичного круга, т.е.} \\ |z_j| > 1 \text{ для всех } j = 1, 2, \dots, q. \end{array} \right\} \quad (3.68)$$

Основные характеристики процесса $CC(q)$. 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 (3.63)

$$\varepsilon(t - k) = \delta(t - k) - \theta_1 \delta(t - k - 1) - \dots - \theta_q \delta(t - k - q), \quad (3.63^a)$$

перемножим почленно выражения (3.63) и (3.63^a) 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 $\gamma(\tau) = E(\varepsilon(t)\varepsilon(t - \tau))$:

$$\gamma(\tau) = \begin{cases} \sigma_0^2(1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) & \text{при } \tau = 0; \\ \sigma_0^2(-\theta_\tau + \theta_1\theta_{\tau+1} + \theta_2\theta_{\tau+2} + \dots + \theta_{q-\tau}\theta_q) & \text{при } \tau = 1, 2, \dots, q; \\ 0 & \text{при } \tau > q \end{cases} \quad (3.69)$$

(при этом, естественно, полагается, что $\theta_j = 0$ при $j > q$).

Автокорреляционная функция процесса $CC(q)$ получается непосредственно из (3.69):

$$r(\tau) = \begin{cases} \frac{-\theta_\tau + \theta_1\theta_{\tau+1} + \theta_2\theta_{\tau+2} + \dots + \theta_{q-\tau}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} & \text{при } \tau = 1, 2, \dots, q; \\ 0 & \text{при } \tau > q. \end{cases} \quad (3.70)$$

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 (3.9^a) с учетом того, что участвующий в нем оператор $B(z)$ для модели (3.63) имеет вид

$$B(z) = 1 - \theta_1 z - \theta_2 z^2 - \dots - \theta_q z^q.$$

Соответственно имеем:

$$p(\tilde{\omega}) = 2\sigma_0^2 |1 - \theta_1 e^{-i2\pi\tilde{\omega}} - \theta_2 e^{-i4\pi\tilde{\omega}} - \dots - \theta_q e^{-i2\pi q\tilde{\omega}}|^2, \quad (3.71)$$

$$0 \leq \tilde{\omega} \leq \frac{1}{2}.$$

Идентификация модели СС(q) производится на базе соотношений (3.70), а именно: 1) по значениям $\hat{\varepsilon}(t) = x(t) - \hat{f}(t)$ с помощью формулы

$$\hat{r}(\tau) = \frac{\frac{1}{N-\tau} \sum_{t=1}^{N-\tau} (\hat{\varepsilon}(t) - \bar{\varepsilon})(\hat{\varepsilon}(t+\tau) - \bar{\varepsilon})}{\frac{1}{N} \sum_{t=1}^N (\varepsilon(t) - \bar{\varepsilon})^2}, \quad \tau = 1, 2, \dots, q, \quad (3.6'')$$

подсчитываются значения $\hat{r}(1), \hat{r}(2), \dots, \hat{r}(q)$; 2) в соотношения (3.70) 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 (3.69) 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 (3.60), выведенных в связи с задачей оценивания параметров процесса авторегрессии, уравнения для определения оценок параметров СС(q)-модели, полученные на базе соотношений (3.70), *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 Дж., Дженкинс Г., вып. 1], с. 225).

Рассмотрим теперь специально два наиболее важных частных случая СС(q)-модели.

Модель скользящего среднего 1-го порядка (СС(1)-модель).

Эта модель описывается соотношением (3.64) и, как уже было показано, *стационарна* при любом значении параметра θ и *обратима* при условии $|\theta| < 1$.

Автокорреляционная функция СС(1)-модели определяется соотноше-

ниями (3.70) при $q = 1$, т. е.

$$r(\tau) = \begin{cases} \frac{-\theta}{1 + \theta^2} & \text{при } \tau = 1, \\ 0 & \text{при } \tau \geq 2. \end{cases} \quad (3.72)$$

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

$$r_{\text{част}}(\tau) = -\theta^\tau \frac{1 - \theta^2}{1 - \theta^{2(\tau+1)}} \quad (3.73)$$

(вывод этого соотношения см., например, в книге [Бокс Дж., Дженкинс Г., вып. 1], 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 $r(1)$ 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 (3.71) при $q = 1$:

$$\begin{aligned} p(\tilde{\omega}) &= 2\sigma_0^2 |1 - \theta e^{-i2\pi\tilde{\omega}}|^2 \\ &= 2\sigma_0^2 (1 + \theta^2 - 2\theta \cos(2\pi\tilde{\omega})), \quad 0 \leq \tilde{\omega} \leq \frac{1}{2}. \end{aligned} \quad (3.74)$$

Из (3.72) и (3.74) видно, что когда θ отрицательно, то $r(1)$ положительно и в спектре доминируют низкие частоты (т. е. 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

$$\theta^2 + \frac{1}{\hat{r}(1)} \theta + 1 = 0. \quad (3.75)$$

Из двух решений этого квадратного уравнения выбирается то, которое удовлетворяет условию обратимости $CC(1)$ -модели (т. е. 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 (3.75), 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 (3.65) и стационарна при любых значениях параметров θ_1 и θ_2 . Однако, как это следует из (3.68), 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, т. е.

$$\begin{cases} |\theta_1| < 2, \\ \theta_2 < 1 - |\theta_1|. \end{cases} \quad (3.68^a)$$

Заметим, что эти условия аналогичны условиям (3.53') стационарности процесса АР(2).

Автокорреляционная функция процесса СС(2) определяется соотношениями (3.70) при $q = 2$, т. е.

$$\begin{aligned} r(1) &= \frac{-\theta_1(1 - \theta_2)}{1 + \theta_1^2 + \theta_2^2}, \\ r(2) &= \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2}, \\ r(\tau) &= 0 \quad \text{при всех } \tau \geq 3. \end{aligned} \quad (3.76)$$

Из (3.68^a) и (3.76) следует, что первые две автокорреляции обратимого процесса СС(2) должны лежать на плоскости $r(1)0r(2)$ внутри площади, ограниченной отрезками линий

$$\begin{cases} r(2) + r(1) = -0,5, \\ r(2) - r(1) = -0,5, \\ r^2(1) = 4r(2)(1 - 2r(2)), \end{cases} \quad (3.77)$$

см. рис. 3.12. Рисунок позволяет оценить, согласуется ли вычисленная по экспериментальным данным с помощью формулы (3.6'') пара значений $\hat{r}(1), \hat{r}(2)$ с гипотезой, что анализируемый временной ряд $\varepsilon(t)$ может быть описан моделью СС(2).

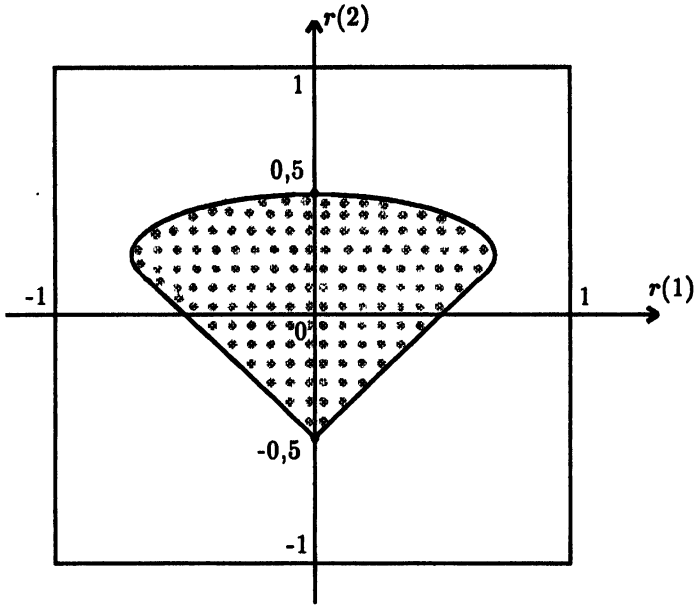


Рис. 3.12. Допустимые области значений $r(1)$ и $r(2)$ для обратимого процесса $CC(2)$

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 (3.71) 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 (3.54) 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:

$$\begin{cases} \hat{r}(1) = \frac{-\theta_1(1 - \theta_2)}{1 + \theta_1^2 + \theta_2^2}, \\ \hat{r}(2) = \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2}. \end{cases} \quad (3.78)$$

Взаимосвязь процессов АР(p) и СС(q). Результаты, изложенные в пунктах 3.4.1 и 3.4.2, позволяют сделать ряд заключений о взаимосвязях, существующих между процессами авторегрессии и скользящего среднего.

1) Для конечного процесса авторегрессии порядка p $\delta(t)$ может быть представлено как *конечная* взвешенная сумма предшествующих ϵ , или $\epsilon(t)$ может быть представлено как бесконечная сумма предшествующих δ . В то же время, в конечном процессе скользящего среднего порядка q $\epsilon(t)$ может быть представлено как *конечная* взвешенная сумма предшествующих δ или $\delta(t)$ — как бесконечная взвешенная сумма предшествующих ϵ .

2) Конечный процесс СС имеет автокорреляционную функцию, обращаящуюся в нуль после некоторой точки, но так как он эквивалентен бесконечному процессу АР, его частная автокорреляционная функция бесконечно протяженная. Главную роль в ней играют затухающие экспоненты и (или) затухающие синусоиды. И наоборот, процесс АР имеет частную автокорреляционную функцию, обращаящуюся 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 СС не должны удовлетворять каким-нибудь условиям для того, чтобы процесс был стационарным. Однако для того чтобы процесс СС был обратимым, корни его характеристического уравнения должны лежать вне единичного круга.

4) Спектр процесса скользящего среднего обратен спектру соответствующего процесса авторегрессии.

3.4.3. Авторегрессионные модели со скользящими средними в остатках (АРСС (p, q)-модели)

В п. 3.4.2 обсуждалась *двойственность* в представлении АР- и СС-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. (2.108)).

Но если мы анализируем процесс действительно типа СС(1), то его представление в виде процесса авторегрессии неэкономично с точки зрения формы его параметризации. Аналогично процесс АР(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

$$\varepsilon(t) = \alpha_1 \varepsilon(t-1) + \dots + \alpha_p \varepsilon(t-p) + \delta(t) - \theta_1 \delta(t-1) - \dots - \theta_q \delta(t-q) \quad (3.79)$$

и называется *процессом авторегрессии — скользящего среднего порядка (p, q)*. Примем для него сокращенное обозначение АРСС (p, q)¹.

Отметим, что, последовательно выражая бесконечное число раз в правой части (3.79) величины $\varepsilon(t-1), \varepsilon(t-2), \dots, \varepsilon(t-p), \dots$ по формуле (3.79), мы убеждаемся в том, что $\varepsilon(t)$ не зависит от будущих значений δ , т. е. от $\delta(t+1), \delta(t+2), \dots$

¹ Модель (3.79) 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.

Стационарность и обратимость АРСС (p, q)-процессов. Записывая процесс (3.79) в виде

$$\varepsilon(t) = \sum_{j=1}^p \alpha_j \varepsilon(t-j) + \bar{\delta}_q(t), \quad (3.79')$$

где $\bar{\delta}_q(t) = \delta(t) - \theta_1 \delta(t-1) - \dots - \theta_q \delta(t-q)$, мы можем провести анализ стационарности (3.79') точно по той же схеме, что и для АР (p)-процессов. При этом различие «костатков» $\bar{\delta}_q(t)$ и $\delta(t)$ никак не повлияет на выводы, определяющие условия стационарности процесса авторегрессии. Поэтому процесс (3.79) является стационарным тогда и только тогда, когда все корни характеристического уравнения АР (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 (3.79) в виде

$$\bar{\varepsilon}_p(t) = \delta(t) - \theta_1 \delta(t-1) - \dots - \theta_q \delta(t-q), \quad (3.79'')$$

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 (3.79) на $\varepsilon(t-\tau)$ и перейдем к математическим ожиданиям получившегося выражения:

$$\gamma(\tau) = \alpha_1 \gamma(\tau-1) + \dots + \alpha_p \gamma(\tau-p) + \gamma_{\varepsilon\delta}(\tau) - \theta_1 \gamma_{\varepsilon\delta}(\tau-1) - \dots - \theta_q \gamma_{\varepsilon\delta}(\tau-q), \quad (3.80)$$

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 (3.80) следует:

$$\gamma(\tau) = \alpha_1 \gamma(\tau-1) + \alpha_2 \gamma(\tau-2) + \dots + \alpha_p \gamma(\tau-p) \quad \text{при } \tau \geq q+1$$

и, соответственно (после деления всех членов на $\gamma(0)$),

$$r(\tau) = \alpha_1 r(\tau - 1) + \alpha_2 r(\tau - 2) + \dots + \alpha_p r(\tau - p) \quad \text{при } \tau \geq q + 1. \quad (3.81)$$

Анализ соотношений (3.80) и (3.81) позволяет сделать следующие выводы.

1) Из соотношений (3.80) 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 $\gamma_{\epsilon\delta}(\tau), \gamma_{\epsilon\delta}(\tau - 1), \dots, \gamma_{\epsilon\delta}(\tau - q)$ 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 $\gamma_{\epsilon\delta}(-k) = E(\epsilon(t+k)\delta(t))$; в произведении $\epsilon(t+k)\delta(t)$ с помощью $(k+1)$ -кратной последовательной подстановки первого сомножителя по формуле (3.79) он заменяется линейной комбинацией $\epsilon(t-1)$, элементов белого шума δ и параметров модели, что после применения к получившемуся произведению операции усреднения E дает выражение, зависящее только от параметров модели (поскольку $E(\epsilon(t-1)\delta(t)) = 0$).

2) Значения автокорреляционной функции $r(\tau)$ для $\tau \geq q+1$ вычисляются по рекуррентному соотношению (3.81), которое в точности повторяет аналогичное рекуррентное соотношение (3.58) для автокорреляционной 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 $AP(p)$, т. е. 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 $r_{\text{част}}(\tau)$ 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 вы-

Стационарность и обратимость. В соответствии с общей теорией АРСС-процессов, процесс АРСС(1,1) стационарен, если корень характеристического уравнения АР(1)-модели, т. е. 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, т. е. 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 (3.80), 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 (3.84') 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,

$$\begin{aligned} r(1) &= \frac{(1 - \alpha\theta)(\alpha - \theta)}{1 + \theta^2 - 2\alpha\theta}, \\ r(\tau) &= \alpha r(\tau - 1) = \alpha^{\tau-1} r(1) \quad \text{для } \tau \geq 2. \end{aligned} \tag{3.85}$$

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 (3.85) и условий стационарности и обратимости следует, что $r(1)$

и $r(2)$ должны удовлетворять соотношениям:

$$\begin{aligned} |r(2)| &< |r(1)|, \\ r(2) &> r(1)(2r(1) + 1) \quad \text{при } r(1) < 0, \\ r(2) &> r(1)(2r(1) - 1) \quad \text{при } r(1) > 0. \end{aligned}$$

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 $r_{\text{част}}(\tau)$ 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 $(\alpha - \theta)$. При отрицательном θ величина $r_{\text{част}}(\tau)$ убывает экспоненциально-знакопеременно.

Спектральная плотность $p(\tilde{\omega})$ определится в соответствии с общим соотношением (3.82), в котором следует положить $p = q = 1$:

$$p(\tilde{\omega}) = 2\sigma_0^2 \frac{1 + \theta^2 - 2\theta \cos(2\pi\tilde{\omega})}{1 + \alpha^2 - 2\alpha \cos(2\pi\tilde{\omega})}, \quad 0 \leq \tilde{\omega} \leq \frac{1}{2}. \quad (3.86)$$

Идентификация АРСС(1,1)-модели может быть проведена в соответствии с общей двухэтапной процедурой, описанной выше (см. соотношения (3.79^a), (3.83) и т. д.). Система (3.83) 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

$$\hat{\alpha} = \frac{\hat{r}(2)}{\hat{r}(1)}. \quad (3.87)$$

На 2-м этапе выписываем соотношения

$$\varepsilon(t) - \hat{\alpha}\varepsilon(t-1) = \delta(t) - \theta\delta(t-1), \quad (3.88)$$

$$\varepsilon(t+1) - \hat{\alpha}\varepsilon(t) = \delta(t+1) - \theta\delta(t). \quad (3.88')$$

Возводя в квадрат (3.88), перемножая почленно соотношения (3.88) и (3.88'), 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 .

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:

$$\begin{aligned} F_+\varepsilon(t) &= \varepsilon(t+1); \\ F_-\varepsilon(t) &= \varepsilon(t-1). \end{aligned}$$

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 $\varepsilon(t-1)$), т.е. 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:

(i) $F_+ = F_-^{-1}$.

Действительно, $F_+(F_-\varepsilon(t)) = F_+F_-\varepsilon(t) = F_+\varepsilon(t-1) = \varepsilon(t)$, т.е. $F_+F_- = 1$.

(ii) $F_+^k\varepsilon(t) = \varepsilon(t+k)$ 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.

(iii) $(c_0 + c_1F_- + \dots + c_mF_-^m)\varepsilon(t) = c_0\varepsilon(t) + c_1\varepsilon(t-1) + \dots + c_m\varepsilon(t-m)$ и аналогично

$$(\bar{c}_0 + c_1F_+ + \dots + c_mF_+^m)\varepsilon(t) = c_0\varepsilon(t) + c_1\varepsilon(t+1) + \dots + c_m\varepsilon(t+m)$$

(в этих соотношениях c_0, c_1, \dots, c_m — некоторые постоянные числа).

В частности, АРСС(p, q)-модель может быть записана в виде

$$A_p(F_-, \alpha)\varepsilon(t) = B_q(F_-\theta)\delta(t), \quad (3.79^b)$$

где

$$A_p(F_-, \alpha) = 1 - \alpha_1 F_- - \alpha_2 F_-^2 - \dots - \alpha_p F_-^p \quad (3.89)$$

и

$$B_q(F_-, \theta) = 1 - \theta_1 F_- - \theta_2 F_-^2 - \dots - \theta_q F_-^q. \quad (3.89')$$

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

$$A_p(z, \alpha) = 0, \quad (3.90)$$

а $\tilde{z}_1(\theta), \tilde{z}_2(\theta), \dots, \tilde{z}_q(\theta)$ — корни характеристического уравнения СС(q)-модели

$$B_q(z, \theta) = 0, \quad (3.90')$$

где полиномы $A_p(z, \alpha)$ и $B_q(z, \theta)$ определяются соотношениями, соответственно, (3.89) и (3.89'). И пусть все эти корни удовлетворяют условиям стационарности процесса АР(p) и обратимости процесса СС(q). Тогда с помощью несложных рассуждений, используя представимость полимов (3.89) и (3.89') в виде

$$\begin{aligned} A_p(z, \alpha) &= (z - z_1(\alpha))(z - z_2(\alpha)) \dots (z - z_p(\alpha)), \\ B_q(z, \theta) &= (z - \tilde{z}_1(\theta))(z - \tilde{z}_2(\theta)) \dots (z - \tilde{z}_q(\theta)), \end{aligned}$$

переходим от (3.79^b) к соотношению

$$\prod_{i=1}^p \left(1 - \frac{1}{z_i(\alpha)} F_-\right) \varepsilon(t) = \prod_{j=1}^q \left(1 - \frac{1}{\tilde{z}_j(\theta)} F_-\right) \delta(t), \quad (3.79^b)$$

представляющему собой еще одну форму записи АРСС(p, q)-процесса.

Наконец, отметим связь, существующую между оператором F_- и оператором введенной в п. 3.3.3 *последовательной разности* Δ . Из определения Δ (напомним: $\Delta\varepsilon(t) = \varepsilon(t) - \varepsilon(t-1)$) и F_- непосредственно следует:

$$\Delta = 1 - F_-, \quad (3.91)$$

т. к. $\Delta\varepsilon(t) = (1 - F_-)\varepsilon(t) = \varepsilon(t) - \varepsilon(t-1)$.

Неоднозначность в определении параметров АРСС-моделей и «прогноз назад». Как известно (см. том 1, п. 2.5.3), 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, следовательно, каждый из них полностью 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 (3.75) 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, (3.90) и (3.90'). 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 Дж., Дженкинс Г., вып. 1], с. 218–221).

1) Все линейные параметрические модели вида

$$\prod_{i=1}^p \left(1 - \frac{1}{z_i(\alpha)} F_{+}^{-} \right) \varepsilon(t) = \prod_{j=1}^q \left(1 - \frac{1}{\tilde{z}_j(\theta)} F_{+}^{-} \right) \delta(t) \quad (3.92)$$

имеют *идентичную ковариационную структуру*. При этом допускается любая комбинация знаков «+» и «-» в качестве нижних индексов оператора F в левых и правых частях (3.92).

2) Существует *единственное* представление (3.79^в) (или равносильные ему представления (3.79), (3.79^а), (3.79^б)), связывающее в рамках модели АРСС(p, q) значение $\varepsilon(t)$ *только с прошлым этого ряда*.

3) Беря в левой и правой частях соотношения (3.92) оператор сдвига F со знаком «+», приходим к соотношению

$$\prod_{i=1}^p \left(1 - \frac{1}{z_i(\alpha)} F_+\right) \varepsilon(t) = \prod_{j=1}^q \left(1 - \frac{1}{\bar{z}_j(\theta)} F_+\right) \delta(t), \quad (3.92')$$

или, что то же,

$$\varepsilon(t) = \alpha_1 \varepsilon(t+1) + \dots + \alpha_p \varepsilon(t+p) + \delta(t) - \theta_1 \delta(t+1) - \dots - \theta_q \delta(t+q). \quad (3.92'')$$

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 (3.92'') (которое часто называют *возвратным*) кажется искусственным, странным, однако оно оказывается полезным в ситуациях, когда требуется восстанавливать неизвестные (потерянные, «стертые», незарегистрированные) прошлые значения анализируемого процесса, т. е. *строить «прогноз назад»*.

Проблема «перепараметризации» при подборе модели. Если отправляться в представлении АРСС(p, q)-модели от (3.79^б)–(3.79^в), то очевидно, что модель (3.79^б) 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 (3.79^в)

$$(1 - 0,5F_-)(1 - 0,8F_-) \varepsilon(t) = (1 - 0,5F_-) \delta(t).$$

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 (3.79^В), сколько ситуация, когда имеются *почти одинаковые* множители. Например, пусть истинная модель имеет вид

$$(1 - 0,4F_-)(1 - 0,8F_-)\varepsilon(t) = (1 - 0,5F_-)\delta(t). \quad (3.93)$$

Если делается попытка подогнать эту модель по экспериментальным данным, то можно ожидать крайнюю нестабильность в оценках параметров¹ из-за близости множителей $(1 - 0,4F_-)$ и $(1 - 0,5F_-)$ в разных частях (3.93). В подобных ситуациях следует идти на упрощение модели. Это 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 (3.79^В). В нашем примере (3.93) это привело бы к рассмотрению АР(1)-модели

$$(1 - 0,8F_-)\varepsilon(t) = \delta(t)$$

вместо АРСС(2,1)-модели (3.93), поскольку эти две модели на практике оказались *статистически неразличимы*.

Можно подойти к этому вопросу с несколько иной точки зрения, «загоняя» множители вида $(1 - F_-/z_j(\theta))$ правой части (3.79^В) в знаменатель левой части этого соотношения и формально разлагая получившийся таким образом оператор левой части

$$\frac{\prod_{i=1}^p \left(1 - \frac{1}{z_i(\alpha)} F_-\right)}{\prod_{j=1}^q \left(1 - \frac{1}{z_j(\theta)} F_-\right)}$$

по степеням $(F_-/z_j(\theta))$ (при этом полагается, что $|F_-/z_j(\theta)| < 1$, используется формула суммы бесконечно убывающей геометрической прогрессии и в получившемся бесконечном разложении ограничиваются двумя-тремя

¹ Под *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).

первыми членами). Так, например, реализация этой процедуры применительно к (3.93) дает

$$\begin{aligned} \frac{(1 - 0,4F_-)(1 - 0,8F_-)}{1 - 0,5F_-} &= (1 - 0,4F_-)(1 - 0,8F_-) \\ &\times (1 + 0,5F_- + 0,25F_-^2 + 0,125F_-^3 + \dots) \\ &= 1 - 0,700F_- - 0,030F_-^2 - 0,015F_-^3 \\ &\quad - 0,008F_-^4 - \dots \end{aligned}$$

Ограничиваясь первыми двумя членами этого разложения, получаем АР(1)-модель

$$(1 - 0,7F_-)\varepsilon(t) = \delta(t),$$

также статистически неразличимую с истинной моделью (3.93).

Из сказанного следует, что окончательный выбор *структурных параметров* модели (т. е. 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 (3.90) и (3.90') (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 $\tilde{z}_j(\theta)$, то в представлении (3.79^а) 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.

Векторные модели авторегрессии — скользящего среднего. В эконометрической литературе и приложениях обсуждаются и используются также *многомерные (или векторные) модели* АРСС и, как их частные случаи, векторные АР- и СС-модели. В сущности эта тема больше относится к проблематике *многомерной регрессии и системе одновременных уравнений* (см. гл. 1 и 4), т. е. к ситуациям, когда статистику-эконометристу приходится одновременно отслеживать поведение многих показателей и исследовать существующие между этими показателями статистические связи. Поэтому мы ограничимся здесь лишь краткой справкой о векторных АРСС-моделях.

Введем в рассмотрение многомерные временные ряды

$$\varepsilon(t) = \begin{pmatrix} \varepsilon^{(1)}(t) \\ \varepsilon^{(2)}(t) \\ \vdots \\ \varepsilon^{(m)}(t) \end{pmatrix}, \quad \delta(t) = \begin{pmatrix} \delta^{(1)}(t) \\ \delta^{(2)}(t) \\ \vdots \\ \delta^{(m)}(t) \end{pmatrix},$$

а также последовательности матриц коэффициентов

$$\alpha_k = \begin{pmatrix} \alpha_{11}^{(k)} & \alpha_{12}^{(k)} & \dots & \alpha_{1m}^{(k)} \\ \alpha_{21}^{(k)} & \alpha_{22}^{(k)} & \dots & \alpha_{2m}^{(k)} \\ \dots & \dots & \dots & \dots \\ \alpha_{m1}^{(k)} & \alpha_{m2}^{(k)} & \dots & \alpha_{mm}^{(k)} \end{pmatrix}, \quad k = 1, 2, \dots, p,$$

$$\Theta_j = \begin{pmatrix} \theta_{11}^{(j)} & \theta_{12}^{(j)} & \dots & \theta_{1m}^{(j)} \\ \theta_{21}^{(j)} & \theta_{22}^{(j)} & \dots & \theta_{2m}^{(j)} \\ \dots & \dots & \dots & \dots \\ \theta_{m1}^{(j)} & \theta_{m2}^{(j)} & \dots & \theta_{mm}^{(j)} \end{pmatrix}, \quad j = 1, 2, \dots, q.$$

Мы полагаем, как и прежде, что компоненты $\delta^{(l)}(t)$ вектора $\delta(t)$ образуют временные последовательности, именуемые белым шумом (см. (3.39)–(3.40)). Однако это не значит, что $\delta^{(l_1)}(t)$ и $\delta^{(l_2)}(s)$ взаимно некоррелированы ни при каких t и s , когда $l_1 \neq l_2$.

Тогда векторная АРСС(p, q)-модель может быть записана в виде

$$\varepsilon(t) = \alpha_1 \varepsilon(t-1) + \dots + \alpha_p \varepsilon(t-p) + \delta(t) - \Theta_1 \delta(t-1) - \dots - \Theta_q \delta(t-q), \quad (3.94)$$

или в покомпонентной форме

$$\varepsilon^{(i)}(t) = \sum_{l=1}^m \alpha_{il}^{(1)} \varepsilon^{(l)}(t-1) + \sum_{l=1}^m \alpha_{il}^{(2)} \varepsilon^{(l)}(t-2) + \dots + \sum_{l=1}^m \alpha_{il}^{(p)} \varepsilon^{(l)}(t-p) \\ + \delta^{(i)}(t) - \sum_{l=1}^m \theta_{il}^{(1)} \delta^{(l)}(t-1) - \dots - \sum_{l=1}^m \theta_{il}^{(q)} \delta^{(l)}(t-q), \\ i = 1, 2, \dots, m. \quad (3.94')$$

Однако по причине слишком большого числа участвующих в (3.94') параметров в эконометрических приложениях используется обычно упрощенный вариант этой модели, в котором отсутствует скользящее суммирование элементов белого шума, т.е. векторная авторегрессионная модель порядка p (сокращенно-векторная АР(p)-модель, или VAR(p)-модель):

$$\varepsilon(t) = \alpha_1 \varepsilon(t-1) + \dots + \alpha_p \varepsilon(t-p) + \delta(t), \quad (3.95)$$

или в покомпонентной записи

$$\varepsilon^{(i)}(t) = \sum_{l=1}^m \alpha_{il}^{(1)} \varepsilon^{(l)}(t-1) + \dots + \sum_{l=1}^m \alpha_{il}^{(p)} \varepsilon^{(l)}(t-p) + \delta^{(i)}(t), \quad i = 1, 2, \dots, m. \quad (3.95')$$

Если остатки $\delta^{(i)}(t)$ и $\delta^{(j)}(t)$ взаимно не коррелированы при $i \neq j$, то параметры m уравнений (3.96') могут оцениваться отдельно для каждого уравнения с помощью обычного метода наименьших квадратов. В общем случае (т. е. при взаимной коррелированности остатков) уравнения (3.95') можно рассматривать как один из типов систем одновременных уравнений (см. гл. 1 и 4) и применять к ним соответствующие методы статистического анализа.

Более подробные сведения о векторных моделях авторегрессии можно найти, например, в книгах [Greene W. H.], [Hayashi].

3.4.4. 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.¹, 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, т. е. гипотеза $D\varepsilon(t) = \gamma(0) = \text{const}$ не противоречила имеющимся экспериментальным данным. Однако в рамках моделей АРСС удовлетворительно объяснить этот феномен не удавалось. Требовалась определенная модификация известных моделей.

Такая модификация была предложена впервые Р. Энглom в 1982 г. (см. работу 1982 г. в сноске). Он рассматривал остатки $\varepsilon(t)$ как *условно гетероскедастичные*, связанные друг с другом простейшей авторегресси-

¹ См. 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.

онной зависимостью, а именно:

$$\begin{cases} [\varepsilon(t) | \varepsilon(t-1)] \in N(0; \sigma_t^2), \\ \text{где } \sigma_t^2 = \mathbf{D}(\varepsilon(t) | \varepsilon(t-1)) = \theta_0 + \theta_1 \varepsilon^2(t-1), \end{cases} \quad (3.96)$$

или, что то же,

$$\varepsilon(t) = \delta(t)[\theta_0 + \theta_1 \varepsilon^2(t-1)], \quad (3.96')$$

где последовательность $\delta(t)$, $t = 1, 2, \dots$, — образует *стандартизованный нормальный* белый шум (т. е. $\delta(t_1)$ 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 *закреплено* (задано). Соответственно, ее поведение будет описываться *условным* законом распределения вероятностей.

Следуя установившейся терминологии, будем называть модель (3.96) *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.], pp. 439–440; получающиеся оценки оказываются более эффективными, чем даже наиболее эффективные, в *классе линейных оценок*, МНК-оценки).

Естественное обобщение моделей типа (3.96) было предложено Р. Энглем 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):

$$\begin{cases} [\varepsilon(t) | \varepsilon(t-1)] \in N(0; \sigma_t^2), \\ \text{где } \sigma_t^2 = \theta_0 + \theta_1 \varepsilon^2(t-1) + \dots + \theta_q \varepsilon^2(t-q), \end{cases} \quad (3.96^a)$$

а параметры $\theta_0, \theta_1, \dots, \theta_q$ связаны некоторыми ограничениями, обеспечивающими *безусловную* гомоскедастичность остатков $\varepsilon(t)$.

Модели (3.96^a) называются моделями АРУГ порядка q (сокращенно АРУГ(q)). Очевидно, модель (3.96) является АРУГ(1)-моделью и соот-

ветствует частному случаю (3.96^а) при $q = 1$. Содержательно переход к $q > 1$ в моделях (3.96^а) 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 (3.96а) 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

$$\left\{ \begin{array}{l} [\varepsilon(t) | \psi(t)] \in N(0; \sigma_t^2), \\ \text{где условная дисперсия } \sigma_t^2 = D(\varepsilon(t) | \psi(t)) \text{ имеет вид} \\ \sigma_t^2 = \alpha_1 \sigma_{t-1}^2 + \alpha_2 \sigma_{t-2}^2 + \dots + \alpha_p \sigma_{t-p}^2 + \theta_0 + \theta_1 \varepsilon(t-1) + \dots + \theta_q \varepsilon(t-q). \end{array} \right. \quad (3.97)$$

В соотношениях (3.97) под $\psi(t)$ подразумевается вся информация о процессе $\varepsilon(t)$, которой мы располагаем к моменту времени t (т.е. — 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 (3.97), 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 (т.е. 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 «Обзорение прикладной и промышленной математики», серия «Финансовая и страховая математика», т. 3 (1996), вып. 6.

3.5. Модели нестационарных временных рядов и их идентификация

В предыдущем пункте мы занимались анализом и моделированием случайных остатков $\varepsilon(t)$, с которыми приходится иметь дело как в обобщенных линейных моделях множественной регрессии (2.73), так и в моделях временных рядов (3.2). Удовлетворительные (в прикладном смысле) результаты в моделировании случайных остатков можно получить, оставаясь в рамках класса моделей *стационарных* временных рядов. Однако сами реальные временные ряды $x(t)$, встречающиеся в экономике, финансах, торговле, маркетинге, за редкими исключениями являются *нестационарными*. Правда, их нестационарность чаще всего проявляется лишь в наличии зависящей от времени t неслучайной составляющей $f(t)$. В подобных случаях говорят о *нестационарности на уровне первых моментов*, или о *нестационарных однородных временных рядах*. Иначе говоря, временной ряд $x(t)$ называется нестационарным однородным, если его случайный остаток $\varepsilon(t)$, получающийся вычитанием из ряда $x(t)$ его неслучайной составляющей $f(t)$, представляет собой стационарный (в широком смысле) временной ряд. Моделям именно таких временных рядов $x(t)$ и посвящен данный пункт учебника.

3.5.1. Модель авторегрессии-проинтегрированного скользящего среднего (АРСС(p, q, k)-модель)

Эта модель предложена Дж. Боксом и Г. Дженкинсом (см. [Бокс Дж., Дженкинс Г., вып. 1], 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 следующими свойствами:

- (i) анализируемый временной ряд включает в себя (аддитивно) составляющую $f(t)$, имеющую вид *алгебраического полинома* (от параметра времени t) некоторой степени $k - 1$ ($k \geq 1$); при этом коэффициенты этого полинома могут быть как стохастической, так и нестохастической природы;
- (ii) 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. 3.3.3), может быть описан моделью АРСС(p, q).

Это означает, что АРСС(p, q, k)-модель анализируемого процесса

$x(t)$, $t = 1, 2, \dots, N$, может быть записана в виде

$$x_k(t) = \alpha_1 x_k(t-1) + \alpha_2 x_k(t-2) + \dots + \alpha_p x_k(t-p) + \delta(t) - \theta_1 \delta(t-1) - \dots - \theta_q \delta(t-q), \quad (3.98)$$

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 (3.79^б) и связи между операторами Δ и F_- (см. (3.91)) модель АРПСС(p, q) может быть записана в форме

$$A_p(F_-, \alpha)(1 - F_-)^k x(t) = B_q(F_-, \theta)\delta(t), \quad (3.98')$$

или, что то же,

$$A_p(F_-, \alpha)\Delta^k x(t) = B_q(F_-, \theta)\delta(t), \quad (3.98'')$$

где полиномы $A_p(F_-, \alpha)$ и $B_q(F_-, \theta)$ определены соотношениями, соответственно, (3.89) и (3.89'), 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 $\varphi(F_-, \alpha)$, т. е. полагая

$$\varphi(F_-, \alpha) = A_p(F_-, \alpha)(1 - F_-)^k, \quad (3.99)$$

представление АРПСС(p, q, k)-модели можно записать в виде

$$\varphi(F_-, \alpha)x(t) = B_q(F_-, \theta)\delta(t). \quad (3.98^a)$$

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 (3.99) следует, что $\varphi(F_-, \alpha)$ есть полином от F_- степени $p + k$ со свободным членом, равным единице, т. е.

$$\varphi(F_-, \alpha) = (1 - \alpha_1 F_- - \alpha_2 F_-^2 - \dots - \alpha_p F_-^p)(1 - F_-)^k \\ = 1 - \varphi_1 F_- - \varphi_2 F_-^2 - \dots - \varphi_{p+k} F_-^{p+k}, \quad (3.99')$$

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 (3.89) и (3.99)

$$\varphi(F_-, \alpha) = (1 - \alpha F_-)(1 - F_-) = 1 - (1 + \alpha)F_- + \alpha F_-^2,$$

так что форма (3.98^a) для модели АРПСС(1,1,1) будет иметь вид

$$[1 - (1 + \alpha)F_- + \alpha F_-^2]x(t) = (1 - \theta F_-)\delta(t), \quad (3.100)$$

или, что то же,

$$x(t) = (1 + \alpha)x(t-1) - \alpha x(t-2) + \delta(t) - \theta\delta(t-1). \quad (3.100')$$

Для многих целей, и в частности для вычисления прогнозов (см. п. 3.6), уравнения типа (3.100') 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:

$$S = \Delta^{-1} = (1 - F_-)^{-1} = 1 + F_- + F_-^2 + \dots \quad (3.101)$$

Соответственно определяются степени оператора S :

$$S^2 x(t) = S(Sx(t)) = Sx(t) + Sx(t-1) + \dots = \sum_{l=-\infty}^t \sum_{\tau=-\infty}^l x(\tau),$$

$$S^3 x(t) = S(S^2 x(t)) = \sum_{\nu=-\infty}^t \sum_{l=-\infty}^{\nu} \sum_{\tau=-\infty}^l x(\tau) \text{ и т.д.}$$

Из (3.101) 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, т.е.

$$x(t) = S^k(\Delta^k x(t)).$$

Применяя эту операцию к обеим частям АРПСС-процесса, записанного в форме (3.98''), имеем

$$A_p(F_-, \alpha)x(t) = S^k(B_q(F_-, \theta)\delta(t)), \quad (3.98^{\delta})$$

т.е. 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. 3.5.2) 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. (3.100'));
- 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 эвристического критерия описан в п. 3.3.3. Он основан на отслеживании поведения величины $\hat{\sigma}^2(k)$ (см. (3.38)) 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, q, k), 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 (т. е. 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 (т. е. 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, т. е. 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 п. 3.4.3.

Коинтеграция временных рядов в регрессионном анализе. Временной ряд $x(t)$ называется *интегрируемым порядка k* , если он становится впервые стационарным после k -кратного применения к нему

разностного оператора Δ . В регрессионном анализе обычно одновременно рассматривается *несколько* временных рядов. Очевидно, если $x(t)$ — интегрируемый временной ряд порядка k_1 и $y(t)$ — интегрируемый временной ряд порядка k_2 , причем $k_2 > k_1$, то при *любом* значении параметра θ (в том числе при $\theta = \hat{\theta}_{\text{МНК}}$, где $\hat{\theta}_{\text{МНК}}$ — МНК-оценка коэффициента регрессии в модели парной регрессии y по x) случайный остаток $\varepsilon(t) = y(t) - \theta x(t)$ будет интегрируемым временным рядом порядка k_2 . Если же $k_1 = k_2 = k$, то константа θ может быть подобрана так, что $\varepsilon(t)$ будет стационарным (или интегрируемым порядка 0) с *нулевым* средним. При этом вектор $(1; -\theta)$ (или любой другой, отличающийся от этого сомножителем) называется *коинтегрирующим*. При регрессионном анализе временных рядов $x(t)$ и $y(t)$ их *коинтеграция* (согласование порядков их интегрируемости) производится обычно по следующей схеме: 1) рассматривается модель $y(t) = \theta x(t) + \varepsilon(t)$ и строится МНК-оценка $\hat{\theta}_{\text{МНК}}$ для параметра θ ; 2) ряд $\hat{\varepsilon}(t) = y(t) - \hat{\theta}_{\text{МНК}} x(t)$ анализируется на стационарность в рамках одной из моделей АРСС(p, q); например, в рамках АР(1)-модели проверяется гипотеза $|\alpha| < 1$ в представлении $\hat{\varepsilon}(t) = \alpha \hat{\varepsilon}(t-1) + \delta(t)$; 3) если результат отрицательный, то возвращаются к спецификации исходной модели, пробуя в качестве зависимой и объясняющей переменных различные варианты $\Delta^{k_1} y(t)$ и $\Delta^{k_2} x(t)$.

Подробнее с проблемой коинтеграции временных рядов можно познакомиться в книгах [Greene W. H.], [Hayashi], [Ruud].

3.5.2. Модели рядов, содержащих сезонную компоненту

В разделе п. 3.1, 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 3.1 и на рис. 3.1 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, приводит 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 3.13 в п. 3.6.2). 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, т. е. *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 АРПСС-моделей с помощью так называемых «упрощающих 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, q)-модели, представленной уравнением (3.98^а), 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 успешно удовлетворяют потребности эконометрического моделирования в достаточно широком спектре приложений.

Модель сезонных рядов № 1. Уравнение модели имеет вид

$$\Delta^k \nabla_T^K x(t) = (1 - \theta F_-)(1 - \theta^* F_-^T) \delta(t), \quad (3.102)$$

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:

$$(1 - aF_-^T)\Delta^k \nabla_T^K x(t) = (1 - \theta F_-)(1 - \theta^* F_-^T)\delta(t), \quad (3.103)$$

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 $\sigma_0^2 = D\delta(t)$, a , θ и θ^* :

$$\gamma(0) = \sigma_0^2(1 + \theta^2) \left(1 + \frac{(\theta^* - a)^2}{1 - a^2} \right);$$

$$\gamma(1) = -\sigma_0^2\theta \left(1 + \frac{(\theta^* - a)^2}{1 - a^2} \right);$$

$$\gamma(T-1) = \gamma(T+1) = \sigma_0^2\theta \left(\theta^* - a - \frac{a(\theta^* - a)^2}{1 - a^2} \right);$$

$$\gamma(T) = -\sigma_0^2(1 + \theta^2) \left(\theta^* - a - \frac{a(\theta^* - a)^2}{1 - a^2} \right);$$

$$\gamma(\tau) = a\gamma(\tau - T) \quad \text{при } \tau \geq T + 2;$$

Модель сезонных рядов № 3. Уравнение модели имеет вид:

$$\Delta^k \nabla_T^K x(t) = (1 - \theta F_- - \theta^* F_-^T) \delta(t), \quad (3.104)$$

или, что то же,

$$\Delta^k \nabla_T^K x(t) = \delta(t) - \theta \delta(t-1) - \theta^* \delta(t-T), \\ T \geq 3.$$

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 $\sigma_0^2 = D\delta(t)$, θ и θ^* :

$$\gamma(0) = \sigma_0^2(1 + \theta^2 + \theta^{*2});$$

$$\gamma(1) = -\sigma_0^2\theta;$$

$$\gamma(T-1) = \sigma_0^2\theta\theta^*;$$

$$\gamma(T) = -\sigma_0^2\theta^*;$$

все остальные ковариации равны нулю.

Модель сезонных рядов № 4. Уравнение модели имеет вид:

$$(1 - aF_-^T) \Delta^k \nabla_T^K x(t) = (1 - \theta F_- - \theta^* F_-^T) \delta(t), \quad (3.105)$$

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 , $\sigma_0^2 = D\delta(t)$, 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:

$$\gamma(0) = \sigma_0^2 \left(1 + \frac{\theta^2 + (\theta^* - a)^2}{1 - a^2} \right);$$

$$\gamma(1) = -\sigma_0^2 \theta \left(1 - a \frac{(\theta^* - a)^2}{1 - a^2} \right);$$

$$\gamma(T-1) = \sigma_0^2 \frac{\theta(\theta^* - a)^2}{1 - a^2};$$

$$\gamma(T) = \sigma_0^2 \frac{a\theta^2 - (\theta^* - a)(1 - a\theta^*)}{1 - a^2};$$

$$\gamma(\tau) = a\gamma(\tau - T) \quad \text{при} \quad \tau \geq T + 1.$$

Схематично процедура построения сезонных моделей, основанных на АРСС-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 АРСС- или модифицированных (в правой части) АРСС-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 Дж., Дженкинс Г., вып. 1].

3.5.3. Регрессионные модели с распределенными лагами

Вернемся теперь к проблематике гл. 2, т. е. к *построению и анализу линейных регрессионных моделей*. Уместность рассмотрения некоторых *специальных* вопросов этой проблематики в главе, посвященной временным рядам, объясняется тем, что в качестве исходных статистических данных мы располагаем наблюдениями *двух временных рядов*

$$\begin{aligned} x(1), x(2), \dots, x(N), \\ y(1), y(2), \dots, y(N). \end{aligned} \quad (3.106)$$

Нашей целью является построение линейной регрессионной модели, позволяющей с наименьшими (в определенном смысле) ошибками восстанавливать и прогнозировать значения $y(t)$ по значениям $x(t)$, $x(t-1), \dots, x(t-T)$ для $t \geq T+1$ (при этом предполагается, конечно, что $T < N$). Иначе говоря, мы будем рассматривать модели вида

$$y(t) = c_0 + \sum_{k=0}^T \theta_k x(t-k) + \delta(t), \quad t = T+1, T+2, \dots, \quad (3.107)$$

где $\delta(t)$, $t = 1, 2, \dots, N$, как и прежде, последовательность гомоскедастичных и взаимно не коррелированных (и не коррелированных с $x(t), x(t-1), \dots, x(t-T)$) регрессионных остатков, а $c_0, \theta_0, \theta_1, \dots, \theta_T$ и $\sigma_0^2 = D\delta(t)$ — неизвестные параметры модели. При этом, для «очень длинных» (теоретически-бесконечных) временных рядов (3.106) в анализируемой модели (3.107) допускается случай $T = \infty$, т. е. суммирование в правой части (3.107) ведется по k от 0 до ∞ .

Подобные модели оказываются естественными (и, соответственно, правильно специфицированными) в ситуациях, когда две переменные x и y связаны так, что воздействие единовременного изменения одной из них (x) на другую (y) сказывается в течение достаточно продолжительного периода времени (T), т. е. наблюдается *распределенный во времени эффект воздействия*. В частности, такие связи возникают, в первую очередь, между регистрируемыми во времени входными и выходными характеристиками процессов накопления и распределения ресурсов (например, процессов преобразования доходов населения в его расходы) или процессов трансформации затрат в результаты (например, процессов воспроизводства основных доходов). Рассмотрим эти примеры.

Пример 3.7. *Зависимость общих расходов населения ($y(t)$) от его наблюдаемых доходов ($x(t)$).* При такой интерпретации участвующих в модели (3.107) переменных $x(t)$ и $y(t)$ коэффициенты регрессии θ_k

имеют прозрачный содержательный смысл, а именно: θ_k это, грубо говоря, доля дохода, которая тратится через k лет после его приобретения. Можно было бы обойтись без «грубо говоря», если бы в качестве $x(t)$ мы располагали бы величиной *истинного* (а не наблюдаемого) дохода, полученного в году t . В действительности же наблюдаемый доход в среднем несколько меньше истинного, поэтому вместо естественного соотношения $\sum_{k=0}^T \theta_k = 1$ (при $0 \leq \theta_k < 1$) мы будем иметь обычно $\sum_{k=0}^T \theta_k > 1$. Обращаем внимание читателя на тот факт, что вытекающие из содержательного смысла коэффициентов θ_k ограничения на их значения задают некоторую их *априорную структуру*.

Пример 3.8. *Зависимость объемов введенных основных фондов ($y(t)$) от капитальных вложений ($x(t)$).* В данном случае значения коэффициентов регрессии $\theta_0, \theta_1, \dots, \theta_T$ показывают, какими долями реализуются капитальные вложения $x(\tau)$, соответственно, в году $\tau, \tau+1, \dots, \tau+T$. В силу существования определенной доли «нефондообразующих» капитальных вложений (которая идет на постройки временного типа, содержание управленческого аппарата, обучение персонала в период строительства и т. п.) в данном случае мы будем иметь

$$\beta = \sum_{k=0}^T \theta_k < 1, \quad 0 \leq \theta_k < 1,$$

где β мы можем интерпретировать как долю фондообразующих капитальных вложений.

Прежде чем перейти к систематическому анализу подобных моделей, зададимся двумя вопросами:

1) Почему модель типа (3.107) требует *специального рассмотрения*, а не может быть проанализирована и идентифицирована в рамках *классической линейной модели множественной регрессии* (КЛММР, см. п. 2.2) или модели со стохастическими предикторами, не коррелированными с регрессионными остатками (см. п. 2.10.1), в которой роль $T+1$ объясняющих переменных играют члены временного ряда $x(t), x(t-1), \dots, x(t-T)$?

2) В чем заключается общая специфика того подкласса КЛММР, который принято называть моделями с *распределенными лагами* и которому посвящен данный пункт учебника?

Формально при сделанных предположениях о природе регрессионных остатков $\delta(t)$ в (3.107) и о некоррелированности $\delta(t)$ и $x(t), x(t-1), \dots, x(t-T)$ модель (3.107), действительно, может быть отнесена к КЛММР (при неслучайном характере временного ряда $x(t)$) или к линейным регрессионным моделям со стохастическим предикторами, не коррелированными с регрессионными остатками.

лированными с регрессионными остатками. И в том и в другом случае статистический анализ этой модели может быть произведен с помощью обычного метода наименьших квадратов (МНК). Однако при *практической реализации* этого метода в данном случае возникают принципиальные трудности. Так, величина T , определяющая число включенных в модель объясняющих переменных, как правило, относится к *неизвестным* параметрам модели. Чтобы определить значение T , приходится, выбрав вначале его достаточно большим, исследовать статистическую значимость получающихся при этом оценок коэффициентов регрессии θ_k для различных значений k . Но здесь нас подстерегают две серьезные (взаимосвязанные) «неприятности»: высокая корреляция между объясняющими переменными (и, следовательно, высокая степень мультиколлинеарности) и слабая статистическая достоверность наших выводов, недостаточная их точность (из-за низких значений отношения числа имеющихся в нашем распоряжении наблюдений к числу оцениваемых параметров модели). Именно эти обстоятельства стимулируют поиск некоторых специальных подходов к анализу моделей типа (3.107).

Что касается второго вопроса, то ответ на него как раз и подсказывает направление этого поиска. Общая специфика моделей (3.107) заключается в том, что из их содержательной сущности, как правило, вытекают определенные априорные сведения о значениях и взаимосвязях, существующих между весовыми коэффициентами $\theta_0, \theta_1, \dots, \theta_T$, или, иначе говоря, *об их структуре*. Так, например, в некоторых ситуациях коэффициенты θ_k экспоненциально убывают по мере роста k , т. е. $\theta_k = c\theta_0^k$, где $0 < \theta_0 < 1$, а это значит, что вместо $T+1$ неизвестных параметров нам придется оценивать всего два: c и θ_0 !

Таким образом, **главная идея, на которой базируется общий подход к анализу и построению моделей вида (3.107), может быть сформулирована следующим образом:**

- *отправляясь от содержательной сущности моделируемой зависимости и смысла весовых коэффициентов θ_k ($k = 0, 1, 2, \dots$), определить их структурные связи с помощью введения небольшого числа параметров $\alpha_1, \dots, \alpha_m$ ($m \ll T$), по значениям которых можно восстановить значения всех неизвестных коэффициентов регрессии θ_k (т. е. речь идет об экономической параметризации последовательности $\theta_0, \theta_1, \theta_2, \dots$); после этого задача сводится к оценке параметров $\alpha_1, \alpha_2, \dots, \alpha_m$.*

Модели (3.107), рассматриваемые в рамках этого общего подхода, называются, как мы уже упоминали, **регрессионными моделями с распределенными лагами** или просто **моделями с распределенными лага-**

ми. Поясним это название.

Если переменная (эндогенная или экзогенная) участвует в записи анализируемой модели, будучи измеренной в один из *прошлых* (по отношению к текущему моменту времени t) временных тактов $t - k$ ($k > 0$), то эту переменную называют *лаговой* или *запаздывающей*, а число единиц времени запаздывания (k) — *лагом* (*запаздыванием*). А поскольку эти лаги k в модели (3.107) *распределены* по объясняющим переменным с весами θ_k , то естественно называть такие модели моделями с *распределенными лагами* (ниже мы увидим, что при некоторых способах параметризации весов θ_k их пронормированные значения могут интерпретироваться как *элементы закона распределения вероятностей*, что снова подтверждает правомерность использования слова «*распределенные*» в названии соответствующих моделей). Последовательность весовых коэффициентов $\theta_0, \theta_1, \dots$ называют *структурой лага* (*конечной* или *бесконечной* в зависимости от конечности или бесконечности их числа T). Если все $\theta_j \geq 0$ ($j = 0, 1, 2, \dots$), то последовательность коэффициентов w_0, w_1, w_2, \dots , где $w_j = \theta_j / \sum_{j=0}^T \theta_j$, называют *нормированной структурой лага* модели (3.107) (очевидно, $\sum_{j=0}^T w_j = 1$).

Нормированная структура лага как распределение вероятностей. Можно воспользоваться формальным сходством нормированной структуры лага и закона распределения вероятностей дискретной случайной величины (см. п. 2.5). Для этого введем случайную величину τ («*время задержки*») с законом распределения вероятностей

$$P\{\tau = k\} = w_k, \quad k = 0, 1, 2, \dots, T, \quad (3.108)$$

где T может принимать и бесконечные значения.

Подобная интерпретация нормированной структуры лага открывает широкие возможности в построении экономичной параметризации последовательности весов w_k с помощью различных широко известных моделей законов распределения для дискретных случайных величин (см. пп. 3.1.1~3.1.3). Кстати, интерпретация весов w_k как вероятностей в ряде ситуаций оказывается вполне оправданной. Так, в примере 3.7 случайная величина τ в (3.108) интерпретируется как число тактов времени, прошедших с момента получения дохода до момента расходования одной случайно выбранной из него единицы. Тогда вероятность $P\{\tau = k\} = w_k$ определится, очевидно, отношением $y_{t,k}/\tilde{x}_t$, где \tilde{x}_t — *истинный* доход, полученный в t -м такте времени, а $y_{t,k}$ — та его часть, которая израсходована в $(t+k)$ -м

такте времени. Точно так же вероятностная интерпретация механизма зависимости (3.107) в примере 3.8 может быть сформулирована следующим образом: некоторая случайным образом выбранная единица фондообразующих капитальных вложений будет освоена на объекте в том же году с вероятностью w_0 , в следующем году — с вероятностью w_1 , через 2 года — с вероятностью w_2 и т. д. (здесь $w_k = \theta_k / \sum_{k=1}^T \theta_k$). При такой интерпретации вполне определенный смысл приобретают и основные характеристики вероятностных распределений. Так, среднее значение $E\tau$ в примере 3.8 будет задавать средний срок реализуемости фондообразующих капитальных вложений, дисперсия $D\tau$ будет характеризовать точность в определении среднего срока с помощью $E\tau$ и т. д.

Итак, из описанного выше следует, что одна типовая модель распределенных лагов отличается от другой способом параметризации весовых коэффициентов $\theta_0, \theta_1, \dots$, т. е. *способом параметризации своей лаговой структуры*. Опишем несколько наиболее распространенных в практике эконометрического моделирования способов параметризации лаговых структур.

Полиномиальная лаговая структура Ширли Алмон (см. Almon S. The Distributed Lag between Capital Appropriations and Expenditures. «Econometrica», 30 (1965), pp. 178–196). Мы рассмотрим здесь простейший вариант этой модели.

Подход основан на *полиномиальной форме параметризации конечной лаговой структуры* $\theta_0, \theta_1, \dots, \theta_T$. А именно, опираясь на теорему Вейерштрасса (которая утверждает, что непрерывная на замкнутом интервале функция может быть приближена на всем отрезке многочленом подходящей степени от ее аргумента, отличающимся от этой функции в любой точке меньше, чем на любое заданное число) и рассматривая весовые коэффициенты θ_k как функции k , автор предложила выразить их в виде полиномов невысокой степени m ($m \leq 3$) от k , т. е.

$$\theta_k = \alpha_0 + \alpha_1 k + \alpha_2 k^2 + \dots + \alpha_m k^m, \quad k = 0, 1, 2, \dots, T, \quad (3.109)$$

где $\alpha_0, \alpha_1, \dots, \alpha_m$ — некоторые неизвестные параметры, которые определяются из условия наиболее точной (в определенном смысле) подгонки модели (3.107).

Подставляя последовательно в (3.109) $k = 0, 1, 2, \dots, T$, получаем:

$$\begin{cases} \theta_0 = \alpha_0, \\ \theta_1 = \alpha_0 + \alpha_1 + \dots + \alpha_m, \\ \theta_2 = \alpha_0 + 2\alpha_1 + \dots + 2^m \alpha_m, \\ \dots\dots\dots \\ \theta_T = \alpha_0 + T\alpha_1 + \dots + T^m \alpha_m. \end{cases} \quad (3.109')$$

Возвращаемся к анализируемой модели (3.107), заменяя в ней коэффициенты θ_k их выражениями по формулам (3.109')

$$\begin{aligned} y(t) &= c_0 + \sum_{k=0}^T \theta_k x(t-k) + \delta(t) = c_0 \\ &+ \alpha_0 x(t) \\ &+ \alpha_0 x(t-1) + \alpha_1 x(t-1) + \dots + \alpha_m x(t-1) \\ &+ \alpha_0 x(t-2) + 2\alpha_1 x(t-2) + \dots + 2^m \alpha_m x(t-2) \\ &\dots\dots\dots \\ &+ \alpha_0 x(t-T) + T\alpha_1 x(t-T) + \dots + T^m \alpha_m x(t-T) + \delta(t). \end{aligned} \quad (3.110)$$

Суммируя c_0 и остальные слагаемые правой части (3.110) по столбцам, получаем:

$$\begin{aligned} y(t) &= c_0 + \alpha_0 [x(t) + x(t-1) + \dots + x(t-T)] \\ &+ \alpha_1 [x(t-1) + 2x(t-2) + \dots + Tx(t-T)] \\ &\dots\dots\dots \\ &+ \alpha_m [x(t-1) + 2^m x(t-2) + \dots + T^m x(t-T)] \\ &+ \delta(t). \end{aligned} \quad (3.110')$$

Обозначая первую квадратную скобку в правой части (3.110') как $\tilde{x}^{(1)}(t')$, вторую — как $\tilde{x}^{(2)}(t')$, ..., $(m+1)$ -ю — как $\tilde{x}^{(m+1)}(t')$, где «новое» время t' «привязано» к моменту времени $t-T$ (т.е. $t' = t - T$), получаем:

$$y(t'+T) = c_0 + \alpha_0 \tilde{x}^{(1)}(t') + \alpha_1 \tilde{x}^{(2)}(t') + \dots + \alpha_m \tilde{x}^{(m+1)}(t') + \delta(t'+T), \quad (3.110'')$$

$$t' = 1, 2, \dots, N - T.$$

В результате мы свели задачу оценивания $T + 2$ неизвестных весовых коэффициентов $c_0, \theta_0, \theta_1, \dots, \theta_T$ к статистическому анализу *стандартной линейной модели множественной регрессии всего с $m + 1$ ($m \leq 3$) неизвестными параметрами* (при этом предполагается, конечно, что длина исходных временных рядов N много больше, чем $T + m$). Так что оценки \hat{c}_0 и $\hat{\alpha}_j$ параметров c_0 и α_j ($j = 1, 2, \dots, m$) получаются с помощью обычного МНК (см. п. 2.3.1), после чего по формулам (3.109') вычисляются оценки $\hat{\theta}_k$ ($k = 0, 1, \dots, T$).

Заметим, что мы полагали в данной схеме максимальную величину лага T *известной*. В действительности она, как правило, определяется статистически. Обычно проводят описанные выше расчеты для нескольких предположительных значений T и окончательный выбор между ними производят на основании диагностики полученных моделей, т. е. — путем сравнения различных характеристик их точности.

Геометрическая лаговая структура Койка (см. *Koyck L.M. Distributed Lags and Investment Analysis. North-Holland Publishing Company, Amsterdam, 1954*). В данном подходе рассматривается *бесконечная* лаговая структура (т. е. в (3.107) полагается $T = \infty$), поэтому он применим лишь к достаточно длинным временным рядам (3.106). Общим (и естественным!) допущением при анализе бесконечных лаговых структур является *требование сходимости* ряда $\beta = \sum_{k=0}^{\infty} \theta_k$, т. е.

$\lim_{T \rightarrow \infty} \sum_{k=0}^T \theta_k = \beta < \infty$, и, следовательно, $\lim_{k \rightarrow \infty} \theta_k = 0$. Это означает, что влияние $x(t)$ на $y(t + k)$ уменьшается до нуля по мере неограниченного увеличения временного интервала k , что естественно, т. к. текущее значение y практически не должно зависеть от поведения x в бесконечно далеком прошлом. Койк в своем подходе конкретизировал и усилил это допущение. В частности, он постулировал, что все нормированные веса $w_k = \theta_k / \sum_{j=0}^{\infty} \theta_j$, являясь положительными, убывают с ростом k по геометрической прогрессии, т. е.

$$w_k = (1 - \lambda)\lambda^k, \quad \text{где } 0 < \lambda < 1 \quad (3.111)$$

(множитель $(1 - \lambda)$ в соотношении (3.111) нужен для того, чтобы обеспечить условие нормировки $\sum_{k=1}^{\infty} w_k = 1$).

Как мы сейчас увидим, это допущение приводит к огромным упрощениям модели (3.107), т. к. вместо оценивания *бесконечного ряда* весовых

коэффициентов $\theta_0, \theta_1, \theta_2, \dots$ нам придется оценить лишь два (!) параметра: λ и $\beta = \sum_{k=0}^{\infty} \theta_k$.

Действительно, возвращаясь к (3.107), имеем:

$$\begin{aligned} y(t) &= c_0 + \sum_{k=0}^{\infty} \theta_k x(t-k) + \delta(t) = c_0 + \beta \sum_{k=0}^{\infty} (1-\lambda)\lambda^k x(t-k) + \delta(t) \\ &= c_0 + \beta(1-\lambda) \left(\sum_{k=0}^{\infty} \lambda^k F_-^k \right) x(t) + \delta(t), \end{aligned} \quad (3.112)$$

где F_- , как и прежде (см. п. 3.4.3), оператор сдвига назад на единицу, т. е. $F_- x(t) = x(t-1)$. Произведя формальные операции с операторами сдвига, получаем

$$\begin{aligned} y(t) &= c_0 + \beta(1-\lambda)[1 + \lambda F_- + \lambda^2 F_-^2 + \dots]x(t) + \delta(t) \\ &= c_0 + \beta(1-\lambda) \frac{x(t)}{1 - \lambda F_-} + \delta(t), \end{aligned}$$

или, что то же

$$(1 - \lambda F_-)y(t) = (1 - \lambda F_-)c_0 + \beta(1 - \lambda)x(t) + (1 - \lambda F_-)\delta(t).$$

Применяя, в соответствии с этим соотношением, оператор сдвига F_- к c_0 , $y(t)$ и $\delta(t)$, получаем

$$y(t) = (1 - \lambda)c_0 + \beta(1 - \lambda)x(t) + \lambda y(t-1) + (\delta(t) - \lambda\delta(t-1)). \quad (3.112')$$

В результате мы получили уравнение регрессии $y(t)$ по объясняющим переменным $x(t)$ и $y(t-1)$ всего с двумя неизвестными коэффициентами β и λ (не считая свободного члена c_0 и дисперсии остатков $\sigma_0^2 = D\delta(t)$). Однако поскольку в правой части уравнения остаточная случайная компонента $\varepsilon(t) = \delta(t) - \lambda\delta(t-1)$ зависит от оцениваемого параметра λ и, вообще говоря, коррелирована, по крайней мере, с объясняющей переменной $y(t-1)$, метод оценивания параметров β и λ нестандартен и зависит от дополнительных предположений относительно природы остатков $\delta(t)$ и/или $\varepsilon(t)$.

Процедуры состоятельного оценивания параметров модели (3.112') при нескольких вариантах специфицирующих условий, касающихся природы случайных остатков $\delta(t)$ и $\varepsilon(t)$, описаны, например, в [Джонстон Дж.], с. 303–320.

З а м е ч а н и е 1. *Финальный* вид (3.112') модели Койка может быть получен и без помощи операций с оператором сдвига F_- . Для этого выпишем *исходный* вид модели для двух текущих моментов времени t и $t - 1$:

$$y(t) = c_0 + \beta(1 - \lambda) \sum_{k=0}^{\infty} \lambda^k x(t - k) + \delta(t),$$

$$y(t - 1) = c_0 + \beta(1 - \lambda) \sum_{k=0}^{\infty} \lambda^k x(t - 1 - k) + \delta(t - 1).$$

Умножив второе уравнение на λ и вычтя полученный результат из первого уравнения, приходим к (3.112').

Рассмотрим две хорошо известные динамические модели экономических процессов, сводящиеся к модели Койка, хотя их базовые априорные допущения *прямо* не формулируются в виде (3.111).

Модель частичного приспособления (или — «*частичной корректировки*») ¹. Предположим, что *желаемое* (или *оптимальное, целевое*) значение $y^*(t)$ некоторого экономического показателя определяется уравнением

$$y^*(t) = \tilde{\theta}_0 + \tilde{\theta}_1 x(t) + \tilde{\delta}(t), \quad (3.113)$$

где регрессионные остатки $\tilde{\delta}(t)$ удовлетворяют условиям (3.39)–(3.40), а $x(t)$ — переменная, выполняющая роль объясняющей, не коррелирована с $\tilde{\delta}(t)$. Однако *желаемое* значение исследуемой результирующей переменной не всегда является *наблюдаемым*. Экономический объект, характеризующий этой переменной, может не иметь возможности сразу (т. е. в *точности* к моменту времени t) «*выходить*» на *желаемое* значение $y^*(t)$. *Так что фактическое (наблюдаемое) значение $y(t)$ этого показателя будет со временем как бы «подтягиваться» к желаемому $y^*(t)$ в соответствии с правилом, формализуемым с помощью соотношения*

$$y(t) = y(t - 1) + \gamma(y^*(t) - y(t - 1)) + \delta(t), \quad (3.114)$$

$$0 \leq \gamma \leq 1,$$

где $\delta(t)$ удовлетворяет условиям (3.39)–(3.40). Из (3.114) следует, что на каждом следующем временном тажке *наблюдаемое* значение $y(t)$ будет

¹ Впервые подобная схема модели была описана в работах:

- 1) Nerlove M. Estimates of the elasticities of supply of selected agricultural commodities. «Journ farm econ.» , 38 (1956), pp. 496–509;
- 2) Nerlove M. The dynamics of supply: estimation of farmers response to price. The Johns Hopkins Press. Baltimore, 1958.

«подправляться» в направлении целевого значения $y^*(t)$ на величину, пропорциональную разнице между оптимальным и текущим уровнями результирующего показателя. Соотношение (3.114) может быть переписано в виде

$$y(t) = \gamma y^*(t) + (1 - \gamma)y(t - 1) + \delta(t), \quad (3.114')$$

откуда следует, что наблюдаемое значение исследуемой результирующей переменной есть (с точностью до регрессионного остатка $\delta(t)$) взвешенное среднее желаемого уровня (на данный момент времени) и фактического значения в предыдущем такте времени. Подставляя модельное оптимальное значение (3.113) в (3.114'), имеем

$$y(t) = \gamma \tilde{\theta}_0 + \gamma \tilde{\theta}_1 x(t) + (1 - \gamma)y(t - 1) + (\delta(t) + \gamma \tilde{\delta}(t)). \quad (3.114'')$$

Сравнивая это соотношение с (3.112'), мы видим, что исследуемая зависимость относится по своему типу к геометрической структуре Койка. В этом можно еще раз убедиться, «разматывая» (3.114'') в обратную сторону по сравнению с тем, как мы это делали при выводе финального соотношения структуры Койка. Действительно, выразим $y(t - 1)$, руководствуясь соотношением (3.114'')

$$y(t - 1) = \gamma \tilde{\theta}_0 + \gamma \tilde{\theta}_1 x(t - 1) + (1 - \gamma)y(t - 2) + (\delta(t - 1) + \gamma \tilde{\delta}(t - 1)).$$

Подставив это выражение $y(t - 1)$ в (3.114''), получим

$$y(t) = \gamma(2 - \gamma)\tilde{\theta}_0 + \gamma \tilde{\theta}_1 x(t) + \gamma(1 - \gamma)\tilde{\theta}_1 x(t - 1) + (1 - \gamma)^2 y(t - 2) + \tilde{\varepsilon}(t),$$

где

$$\tilde{\varepsilon}(t) = [\delta(t) + (1 - \gamma)\delta(t - 1)] + \gamma[\tilde{\delta}(t) + (1 - \gamma)\tilde{\delta}(t - 1)].$$

Неограниченно продолжая подобную подстановку, т. е. последовательно выражая $y(t - 2)$, $y(t - 3)$ и т. д. по формуле (3.114'') и подставляя их в соответствующее выражение для $y(t)$, получим в конечном счете

$$y(t) = \tilde{\theta}_0 + \theta_0 [x(t) + (1 - \gamma)x(t - 1) + (1 - \gamma)^2 x(t - 2) + \dots] + \varepsilon(t), \quad (3.115)$$

где свободный член получен в результате бесконечного суммирования величин $\gamma \tilde{\theta}_0 + (1 - \gamma)\gamma \tilde{\theta}_0 + (1 - \gamma)^2 \gamma \tilde{\theta}_0 + \dots$, коэффициент $\theta_0 = \gamma \tilde{\theta}_1$, а случайные остатки $\varepsilon(t)$ получаются как бесконечные скользящие средние исходных остатков $\delta(t)$ и $\tilde{\delta}(t)$, а именно:

$$\varepsilon(t) = \sum_{k=0}^{\infty} \left\{ (1 - \gamma)^k [\delta(t - k) + \gamma \tilde{\delta}(t - k)] \right\}.$$

Сравнив (3.115) со средним выражением в соотношении (3.112), убеждаемся в том, что модель частичного приспособления действительно относится к классу геометрических структур Койка (с точностью до условий, специфицирующих случайные остатки).

Схема «частичного приспособления» имеет довольно широкий спектр экономических приложений. Упомянем о некоторых из них.

Пусть $x(t)$ в соотношении (3.113) определяет уровень продаж некоторого товара, а $y^*(t)$ — соответствующая оптимальная (с позиций равновесия) реакция со стороны производства (т.е. — оптимальный объем предложения этого товара). Если принимать в расчет затраты двух типов — потери, связанные с нарушением равновесия, и затраты, связанные с изменениями состояния производителя, — и если полагать, что оба типа потерь, составляющие суммарные потери L , пропорциональны квадратам соответствующих отклонений, т.е.

$$L(y(t)) = c_1(y(t) - y^*(t))^2 + c_2(y(t) - y(t-1))^2, \quad c_1, c_2 > 0,$$

то задача состоит в выборе (при заданных величинах $y(t-1)$ и $y^*(t)$) значения $y(t)$, минимизирующего потери $L(y(t))$. Приравняем к нулю производную

$$\frac{dL(y(t))}{dy(t)} = 2c_1(y(t) - y^*(t)) + 2c_2(y(t) - y(t-1)) = 0.$$

Отсюда определяем

$$y(t) = y(t-1) + \frac{c_1}{c_1 + c_2} (y^*(t) - y(t-1)),$$

что в точности соответствует основному принципу (3.114) частичного приспособления.

По совершенно аналогичной схеме может быть рассмотрена модель, связывающая между собой доход $x(t)$ с оптимальной ($y^*(t)$) и реальной, т.е. наблюдаемой ($y(t)$), величинами потребительских расходов. Ведь потребитель не располагает всей необходимой информацией о своем «пространстве потребностей», так что не может *мгновенно оптимально среагировать* в своих расходах на изменение (в большую или меньшую сторону) своего дохода. Этой мгновенной реакции препятствуют также определенная инерционность потребительского процесса, наличие определенных (пусть неформальных) обязательств, связанных с прежним уровнем его дохода.

Наконец, модель частичного приспособления используется при корректировке (в динамике) доли прибыли, выплачиваемой компанией в виде

доходов своим акционерам. При росте прибыли $x(t)$ выплачиваемые дивиденды $y(t)$ также увеличиваются, но, как правило, не в той же пропорции θ , которая определена компанией в качестве целевой долгосрочной доли выплат. Это объясняется рядом причин, в том числе осторожностью руководства компании: возрастание прибыли может оказаться временным, поэтому если при этом сохранить долю θ , то в случае последующего понижения прибыли дивиденды придется сокращать, а это — серьезный удар по репутации фирмы. Как следствие, динамика выплаты дивидендов $y(t)$ в зависимости от текущей прибыли $x(t)$ и *желаемого* (в долгосрочной перспективе) объема выплат дивидендов $y^*(t) = \theta x(t)$ регулируется во многих случаях соотношением (3.114).

«Узкое место» модели частичного приспособления состоит в том, что иногда предположение о зависимости оптимального значения $y^*(t)$ *только от текущего* значения $x(t)$ оказывается не адекватным действительности. Другими словами, часто решающим мотивом для принятия ответственных решений не может служить *единственное* значение объясняющей переменной. Один из способов преодоления этой ограниченности отражен в описываемой ниже модели адаптивных ожиданий.

Модель адаптивных ожиданий. Моделирование закономерностей *с учетом ожидаемых ситуаций* — одна из важнейших проблем прикладной экономики. Это, в первую очередь, верно для *макроуровня*, на котором инвестиции, сбережения и спрос на активы оказываются особенно чувствительными к ожиданиям относительно будущего. Если в модели частичного приспособления в роли корректируемой величины выступала *зависимая* переменная (результатирующий показатель) $y(t)$, то в модели адаптивных ожиданий корректируется *объясняющая* переменная $x^*(t+1)$, которая определяет *ожидаемое* на момент $t+1$ (но экспертно формируемое в момент t) значение аргумента в исследуемой зависимости вида

$$y(t) = \tilde{\theta}_0 + \tilde{\theta}_1 x^*(t+1) + \delta(t), \quad (3.116)$$

где возмущающее воздействие $\delta(t)$ удовлетворяет условиям (3.39)–(3.40) и не коррелировано с наблюдаемым значением аргумента $x(t)$. В соответствии с основным допущением модели механизм формирования ожидаемого значения $x^*(t+1)$ описывается соотношением

$$\begin{aligned} x^*(t+1) &= x^*(t) + \gamma(x(t) - x^*(t)), \\ 0 &\leq \gamma \leq 1, \end{aligned} \quad (3.117)$$

или, что то же,

$$x^*(t+1) = \gamma x(t) + (1 - \gamma)x^*(t). \quad (3.117')$$

Это означает, что значение объясняющей переменной, ожидаемое в момент времени $t + 1$, формируется в момент времени t как взвешенное среднее ее реального и ожидаемого значений в текущий момент времени. От значения γ зависит скорость адаптации ожидаемых значений к реальности. Мы видим, что, в отличие от процесса частичного приспособления, который базируется на инерции и *прошлой* динамике показателей, процесс адаптивных ожиданий направлен в *будущее*. Другими словами, мы *формируем значение результирующего показателя на текущий момент времени с учетом будущего значения объясняющей переменной*.

Покажем, что и процесс адаптивных ожиданий вкладывается в общую схему моделей с распределенными лагами, имеющих геометрическую структуру Койка. Для этого перепишем (3.117') в виде

$$x^*(t+1) - (1 - \gamma)x^*(t) = \gamma x(t),$$

или, что то же,

$$[1 - (1 - \gamma)F_-]x^*(t+1) = \gamma x(t),$$

где с помощью F_- , как и прежде, обозначен оператор сдвига функции времени на один временной такт назад. Выражая отсюда $x^*(t+1)$ и подставляя это выражение в (3.116), получаем

$$y(t) = \tilde{\theta}_0 + \tilde{\theta}_1 \frac{\gamma x(t)}{1 - (1 - \gamma)F_-} + \delta(t). \quad (3.118)$$

После домножения всех членов этого выражения на $1 - (1 - \gamma)F_-$ и применения этого оператора к $y(t)$, $\tilde{\theta}_0$ и $\delta(t)$ получаем

$$y(t) = \gamma \tilde{\theta}_0 + (1 - \gamma)y(t-1) + \gamma \tilde{\theta}_1 x(t) + (\delta(t) - (1 - \gamma)\delta(t-1)). \quad (3.118')$$

Сравнение модели (3.118') с финальной формой (3.112') модели Койка свидетельствует об их эквивалентности.

Модель гиперинфляции Кагана и модель потребления Фридмана (основанная на «гипотезе о перманентном доходе») представляют собой наиболее известные примеры эконометрических приложений модели адаптивных ожиданий¹. В модели гиперинфляции исследуется соотношение между спросом на денежные остатки и *ожидаемым* изменением уровня инфляции. В несколько упрощенном (по сравнению с первоисточником)

¹ См. *Cagan P. The Monetary Dynamics of Hyperinflation. In: «Studies in the Quantity Theory of Money» . Chicago, University of Chicago Press, 1956.*

Friedman M. A Theory of the Consumption Function. Princeton, N.J.: Princeton University Press, 1957.

варианте модель может быть описана следующим образом. Изменение спроса на денежные остатки в момент времени t определяется показателем

$$y(t) = \ln \left[\frac{M(t)}{P(t)} \right],$$

где $M(t)$ — индекс изменения объема денег в обращении, а $P(t)$ — индекс цен. Зависимость изменения спроса на денежные остатки в момент t от *ожидаемого* в момент $t + 1$ уровня инфляции $x^*(t + 1)$ определяется уравнением

$$y(t) = -\theta_0 - \theta_1 x^*(t + 1) + \delta(t), \quad \theta_0 > 0, \theta_1 > 0.$$

Адаптивные ожидания уровня инфляции определялись формулой (3.117). Модель рассчитывалась по месячным данным, характеризующим семь инфляционных периодов в США в промежутке 1921–1956 гг. (как отдельно для каждого периода, так и агрегированно по совокупности всех семи периодов).

Модель потребления Фридмана основана на двух постулатах. Один из них выделяет эту модель среди других многочисленных моделей, исследующих зависимость объема потребления (y) от дохода (x) (в данном учебнике эта задача тоже рассматривалась в п. 2.10.2 в связи с методом инструментальных переменных, используемым в проблеме построения регрессионной модели со стохастическими предикторами). В соответствии с этим постулатом фактический объем потребления $y(t)$ и фактический уровень дохода $x(t)$ складываются из перманентной (постоянной) и временной (случайной) составляющих, т. е.

$$\begin{aligned} y(t) &= y^*(t) + y_{\text{вр}}(t), \\ x(t) &= x^*(t) + x_{\text{вр}}(t). \end{aligned}$$

Предполагается, что временная составляющая потребления $y_{\text{вр}}(t)$ и временная составляющая дохода $x_{\text{вр}}(t)$ являются случайными переменными со средними значениями, равными нулю, и постоянными значениями дисперсий и что распределены они независимо от постоянного дохода, постоянного потребления и друг от друга.

Второй постулат не отличается от базовых допущений большинства других аналогичных моделей и состоит в допущении, что $x^*(t)$ и $y^*(t)$ связаны между собой классической линейной регрессионной моделью, т. е.

$$y^*(t) = \theta_0 + \theta_1 x^*(t) + \delta(t). \quad (3.119)$$

Однако величина постоянного дохода $x^*(t)$ (впрочем, так же, как и величина $y^*(t)$) в модели (3.119) ненаблюдаема. Эта проблема обходится с

помощью предположения, что изменение постоянного дохода $x^*(t)$ подчиняется закону адаптивных ожиданий (3.117). Другими словами, домашние хозяйства (индивиды) корректируют свое представление об ожидаемом постоянном доходе по мере роста (убывания) фактического дохода, но не на полное значение прироста (убывания), сознавая, что изменения фактического дохода частично объясняются вариацией временной его составляющей.

Возвращаясь к (3.119) и подставляя вместо $y^*(t)$ разность $y(t) - y_{вп}(t)$, а вместо $x^*(t)$ его выражение, полученное из соотношения (3.117), имеем

$$y(t) = \theta_0 + \theta_1 \frac{\gamma x(t)}{1 - (1 - \gamma)F_-} + (\delta(t) + y_{вп}(t)). \quad (3.120)$$

Сравнивая это выражение с (3.118) и учитывая эквивалентность (3.118), (3.118') и (3.112'), мы видим, что модель потребления Фридмана также может быть отнесена к моделям распределенных лагов с геометрической структурой.

З а м е ч а н и е 2. И в моделях частичной корректировки, и в моделях адаптивных ожиданий мы сталкиваемся с проблемой оценки величины γ . Один из распространенных подходов к решению данной задачи — это использование эвристического по своей природе общего метода *перекрестного анализа дееспособности модели* (в англоязычной литературе он известен как «*cross validation method*»). Применительно к данной проблеме это означает следующую последовательность действий. Все имеющиеся в нашем распоряжении исходные статистические данные $M(1, N) = \{x(t), y(t); t = 1, 2, \dots, N\}$ разбиваются на две части: «обучающую» $M_{об}(1, N_1) = \{x(t), y(t); t = 1, 2, \dots, N_1 (N_1 < N)\}$ и «экзаменующую» $M_{экз}(N_1 + 1, N) = \{x(t), y(t); t = N_1 + 1, N_1 + 2, \dots, N\}$. Задаются сеткой значений параметра γ на интервале $[0; 1]$. Для каждого значения γ из этой сетки по «обучающим» данным $M_{об}(1, N_1)$ строится модель (3.120). Затем по «экзаменующим» данным $M_{экз}(N_1 + 1, N)$ вычисляется значение *критерия качества* полученной модели (роль такого критерия в данной задаче обычно играет соответствующий коэффициент детерминации R^2 , см. п. 2.3.3). Выбирают то значение γ , при котором значение критерия качества модели максимально.

Лаговые структуры, основанные на вероятностной параметризации. Мы уже упоминали, что одним из способов экономической параметризации лаговых структур является их интерпретация в терминах вероятностных распределений (см. (3.108)). Приведем в заключение два примера такой параметризации: один — для бесконечной лаговой структуры (структура Паскаля), другой — для конечной лаговой структуры.

Лаговая структура Паскаля была предложена в 1960 г. Р. Солоу¹. Она основана на так называемом отрицательном биномиальном распределении (см. п. 3.1.1), т. е. элементы w_k нормированной бесконечной лаговой структуры Паскаля ($k = 0, 1, 2, \dots$) определяются в соответствии с (3.108) с помощью соотношений

$$w_k = (1 - p)^M C_{M+k-1}^k p^k, \quad k = 0, 1, 2, \dots, \quad (3.121)$$

где p ($0 < p < 1$) и M (M — любое целое положительное число) — два параметра, определяющие (наряду с $\beta = \sum_{k=0}^{\infty} \theta_k$) конкретную лаговую структуру в данном параметрическом семействе. Отметим, что из свойств отрицательного биномиального распределения следует:

- (i) элементы w_k нормированной лаговой структуры при $M > 1$ сначала возрастают (при $k < (pM - 1)/(1 - p)$), а затем убывают (при $k > (pM - 1)/(1 - p)$);
- (ii) среднее значение лага ($E\tau$) и его дисперсия ($D\tau$) являются возрастающими функциями как M , так и p , а именно:

$$E\tau = \frac{pM}{1 - p};$$

$$D\tau = \frac{pM}{(1 - p)^2}.$$

Лаговая структура, основанная на биномиальном законе распределения вероятностей, является естественным аналогом структуры Паскаля в классе конечных моделей с распределенными лагами. Ее нормированные элементы w_k задаются с помощью соотношений

$$w_k = C_T^k p^k (1 - p)^{T-k}, \quad k = 0, 1, \dots, T. \quad (3.122)$$

Мы видим, что данный класс конечных лаговых структур описывается *однопараметрическим* семейством (параметр p — некоторое число между 0 и 1). Из свойств биномиального распределения (см. п. 3.1.1) следует, что последовательность (3.122) образует так же, как и в структуре Паскаля, *униmodalный* ряд, причем $E\tau = pT$ и $D\tau = p(1 - p)T$.

Описанные вероятностные лаговые структуры применимы в ситуациях, когда из содержательного смысла анализируемых зависимостей вида (3.107) следует, что весовые коэффициенты θ_k (а значит, и w_k) начинают

¹ См. Solow R. M. On a family of lag distributions. «Econometrica», 28 (1960), pp. 393–406.

монотонно убывать не сразу, а только после некоторого k_0 . С подобными ситуациями исследователь имеет дело, например, при анализе зависимостей, связывающих между собой *входные и выходные характеристики процессов накопления и распределения ресурсов* (см. выше примеры 3.7 и 3.8).

3.6. Прогнозирование экономических показателей, основанное на использовании моделей временных рядов

Познакомившись в п. 3.3 с различными способами сглаживания временного ряда, а в пп. 3.4 и 3.5 — 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 (как это и имеет место, например, в регрессионных моделях с распределенными лагами, см. п. 3.5.3), остаются без внимания. Это объясняется просто: единственный тип не авторегрессионных моделей, рассмотренный в данной главе (см. п. 3.5.3), как мы видели, сводится к относительно простым классическим или обобщенным моделям регрессии, а задача прогноза, основанная на этих моделях, была обсуждена в п. 2.9.

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, т. е. 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 измене-

ний 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*.

3.6.1. Прогнозирование на базе АРСС-моделей

Из результатов пп. 3.4–3.5 следует, что АРСС-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. 3.5.2). Поэтому начнем обсуждение проблемы прогнозирования временных рядов с методов, основанных на использовании АРСС-моделей. Мы говорим об АРСС-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 АРСС-модели для анализируемого временного ряда, включая идентификацию этой модели. Поэтому в дальнейших наших рассуждениях предполагается, что все параметры модели уже оценены, т. е. *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. (3.98)–(3.98')), что анализируемый в рамках АРПСС(p, q, k)-модели ряд $x(\tau)$ представим (при любом $\tau \geq k + 1$) уравнением вида

$$\begin{aligned} & (1 - \alpha_1 F_- - \dots - \alpha_p F_-^p)(x(\tau) - C_k^1 x(\tau - 1) + \dots + (-1)^k x(\tau - k)) \\ & = \delta(\tau) - \theta_1 \delta(\tau - 1) - \dots - \theta_q \delta(\tau - q), \end{aligned} \quad (3.98^b)$$

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 (3.98^b) 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:

$$\begin{aligned} x(t - q) &= \left(\sum_{j=1}^p \alpha_j F_-^j \right) x(t - q) + \left(1 - \sum_{j=1}^p \alpha_j F_-^j \right) \\ &\quad \times (C_k^1 x(t - q - 1) - \dots + (-1)^k x(t - q - k)) \\ &\quad + \delta(t - q) - \theta_1 \delta(t - q - 1) - \dots - \theta_q \delta(t - 2q), \end{aligned} \quad (3.123)$$

$$\begin{aligned} x(t - q + 1) &= \left(\sum_{j=1}^p \alpha_j F_-^j \right) x(t - q + 1) + \left(1 - \sum_{j=1}^p \alpha_j F_-^j \right) \\ &\quad \times (C_k^1 x(t - q) - \dots + (-1)^k x(t - q + 1 - k)) + \delta(t - q + 1) \\ &\quad - \theta_1 \delta(t - q) - \dots - \theta_q \delta(t - 2q + 1), \end{aligned} \quad (3.123')$$

.....

$$\begin{aligned} x(t + 1) &= \left(\sum_{j=1}^p \alpha_j F_-^j \right) x(t + 1) - \left(1 - \sum_{j=1}^p \alpha_j F_-^j \right) \\ &\quad \times (C_k^1 x(t) - \dots + (-1)^k x(t + 1 - k)) \\ &\quad + \delta(t + 1) - \theta_1 \delta(t) - \dots - \theta_q \delta(t + 1 - q), \end{aligned} \quad (3.123'')$$

.....

$$\begin{aligned} x(t + l) &= \left(\sum_{j=1}^p \alpha_j F_-^j \right) x(t + l) - \left(1 - \sum_{j=1}^p \alpha_j F_-^j \right) \\ &\quad \times (C_k^1 x(t + l - 1) - \dots + (-1)^k x(t + l - k)) + \delta(t + l) \\ &\quad - \theta_1 \delta(t + l - 1) - \dots - \theta_q \delta(t + l - q). \end{aligned} \quad (3.123''')$$

Обращаем внимание на тот факт, что правые части этих соотноше-

ний 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, т. к. 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 (3.123)–(3.123''') 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, Винером¹).

Условное математическое ожидание $E(x(t + l) | x(1), x(2), \dots, x(t))$ получается применением операции условного усреднения к обеим частям (3.123''') с учетом следующих соотношений:

$$E(x(t - j) | x(1), \dots, x(t)) = x(t - j) \quad \text{при всех } j = 0, 1, 2, \dots, t - 1; \quad (3.124^a)$$

$$E(x(t + j) | x(1), \dots, x(t)) = \hat{x}(t; j) \quad \text{при всех } j = 1, 2, \dots; \quad (3.124^b)$$

$$E(\delta(t + j) | x(1), \dots, x(t)) = 0 \quad \text{при всех } j = 1, 2, \dots; \quad (3.124^b)$$

$$E(\delta(t - j) | x(1), \dots, x(t)) = x(t - j) - \hat{x}(t - j - 1; 1) \quad \text{при всех } j = 0, 1, 2, \dots; \quad (3.124^f)$$

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 (3.123), (3.123'), ... для вычисления *ретроспективных* прогнозов $\hat{x}(t - q - 1; 1)$, $\hat{x}(t - q; 1)$ и т. д. до

¹ См.: 1) Wold H. O. A Study in the Analysis of Stationary Time Series. Almqvist and Wiksell, 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.

$\hat{x}(t-1; 1)$ 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 (3.123)–(3.123')–... 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 (3.124^Г), 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 (3.123^В)) мы могли бы подсчитать в правой части (3.123^В) значения $E(\delta(t-j) | x(1), \dots, x(t))$ ($j = 0, 1, \dots, q-1$) по формулам (3.124^Г);

2) на 2-ом этапе используются формулы (3.123^В), ..., (3.123^М) и правила (3.124^А)–(3.124^Г) подсчета условных математических ожиданий участвующих в правых частях (3.123^В), ..., (3.123^М) выражений для вычисления прогнозных значений $\hat{x}(t; 1), \hat{x}(t; 2), \dots, \hat{x}(t; l)$.

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 (3.123^В); 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 (3.124^Г), аппроксимируются нулями.

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 (3.98^Б) непосредственно следует, что описываемый АРПСС(p, q, k)-моделью временной ряд $x(\tau)$ может быть представлен в форме модели авторегрессии вида

$$x(\tau) = \sum_{j=1}^{p+k} \pi_j x(\tau-j) + \varepsilon(\tau), \quad (3.125)$$

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 (3.25) при $\tau = t+1$, можно было бы вычислить прогнозное

значение $\hat{x}(t; 1)$ для $x(t+1)$, по нему — прогнозное значение $\hat{x}(t+1; 1)$ 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 ; по той причине, что регрессионные остатки $\varepsilon(\tau)$ в модели (3.125) оказываются существенно коррелированными с объясняющими переменными $x(\tau - j)$ (см. п. 2.10.2).

Рассмотрим ряд примеров.

Пример 3.9. *Прогнозирование ряда, описываемого моделью АРПСС(1,0,1).* Анализ и идентификация имеющегося временного ряда $x(\tau)$ привели нас к АРПСС(1,0,1)-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 (3.123''), (3.123''') в данном случае имеют вид

$$\begin{aligned} x(t+1) &= 1,8x(t) - 0,8x(t-1) + \delta(t+1), \\ \dots\dots\dots & \\ x(t+l) &= 1,8x(t+l-1) - 0,8x(t+l-2) + \delta(t+l). \end{aligned} \quad (3.126)$$

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 (3.126) фиксируются значения $x(\tau)$, известные к моменту времени t (т.е. $x(1), \dots, x(t)$), 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 (3.124^r) из-за нехватки прогнозных значений $\hat{x}(t-j-1; 1)$, вставляются нули, а остальные прошлые значения δ подсчитываются по формулам (3.124^r). В результате от (3.126) 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 3.10. *Прогнозирование ряда, описываемого моделью АРПСС(0,1,1)¹.* Анализ ежедневных данных о биржевых ценах акций

¹ Пример заимствован из [Бокс Дж., Дженкинс Г.], с. 188–190.

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 (3.123^{''}), (3.123^{'''}) 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 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» (т.е. — 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. 3.6.2 в связи с сопоставлением описываемого метода прогнозирования с методом экспоненциально взвешенного скользящего среднего Брауна.

Пример 3.11. *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 (3.123^{'''}) 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 3.12. *Прогнозирование ряда, описываемого моделью АРСС(1,1,1).* Подчиняющийся закономерностям этой модели временной ряд представим, как известно, в форме

$$(1 - \alpha F_-) \Delta x(\tau) = \delta(\tau) - \theta\delta(t-1),$$

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 (3.123'''):

$$x(t+l) = (1 + \alpha)x(t+l-1) - \alpha x(t+l-2) + \delta(t+l) - \theta\delta(t+l-1).$$

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

$$\begin{aligned}\hat{x}(t; 1) &= (1 + \alpha)x(t) - \alpha x(t-1) - \theta\delta(t), \\ \hat{x}(t; l) &= (1 + \alpha)\hat{x}(t; l-1) - \alpha\hat{x}(t; l-2) \quad \text{при } l \geq 2.\end{aligned} \tag{3.127}$$

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 (3.127) является эквивалентное им представление прогноза $\hat{x}(t; l)$ в виде

$$\hat{x}(t; l) = \left[1 - \frac{\theta - \alpha}{1 - \alpha} (1 - \alpha^l) \right] x(t) + \frac{\theta - \alpha}{1 - \alpha} (1 - \alpha^l) \bar{x}_\theta(t-1), \tag{3.127'}$$

где $\bar{x}_\theta(t-1)$ — экспоненциально взвешенное скользящее среднее (с параметром θ) — см. выше п. 3.3.2 — анализируемого «длинного» ряда, просуммированного от бесконечно удаленного прошлого до момента времени $t-1$, т. е.

$$\bar{x}_\theta(t-1) = (1 - \theta) \sum_{j=1}^{\infty} \theta^{j-1} x(t-j). \tag{3.128}$$

Из (3.127') 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 $\lambda(l) = (\theta - \alpha)(1 - \alpha^l)/(1 - \alpha)$, т. е. $\hat{x}(t; l) = (1 - \lambda(l))x(t) + \lambda(l)\bar{x}_\theta(t-1)$ (напомним, что $|\alpha| < 1$ и $|\theta| < 1$, т. к. это обеспечивает условия, соответственно, стационарности и обратимости анализируемого ряда, см. п. 3.4.3).

Мы не касались здесь важных вопросов *оценки точности получающихся прогнозов*. Теоретические аспекты этой проблемы рассмотрены, например, в [Бокс Дж., Дженкинс Г.]. Некоторые общие эвристические и имитационные подходы к оценке точности «подгонки» модели по экспериментальным данным и точности получающихся с ее помощью прогнозов будут рассмотрены в конце гл. 4 (см. п. 4.5).

3.6.2. 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 бóльший 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, т. к. 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 (Брауна¹). Простейший вариант этого метода был описан в п. 3.3.2 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

¹ См.: Brown R.G. Smoothing, Forecasting and Prediction of Discrete Time-Series. Prentice-Hall, New Jersey, 1962.

в виде

$$x(\tau) = a_0 + \varepsilon(\tau), \quad (3.129)$$

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. 3.3.2), 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

$$\bar{x}_\lambda(t) = \frac{1 - \lambda}{1 - \lambda^t} \sum_{j=0}^{t-1} \lambda^j x(t - j), \quad (3.130)$$

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 также как коэффициент дисконтирования, характеризующий меру обесценивания наблюдения за единицу времени.

Для рядов с «бесконечным прошлым» формула (3.130) сводится к виду

$$\bar{x}_\lambda(t) = (1 - \lambda) \sum_{j=0}^{\infty} \lambda^j x(t - j). \quad (3.130')$$

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

$$\hat{x}(t; 1) = \bar{x}_\lambda(t), \quad (3.131)$$

где значение $\bar{x}_\lambda(t)$ определено формулой (3.130) (для короткого временного ряда) или формулой (3.130') (для длинного временного ряда).

Формула (3.131) 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

$$\bar{x}_\lambda(t + 1) = \lambda \bar{x}_\lambda(t) + (1 - \lambda)x(t + 1). \quad (3.132)$$

Для вывода этого соотношения выпишем (3.130') 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 (3.132). Для конечных временных рядов соотношение (3.132) имеет приближенный смысл, поскольку в аналогичных выкладках мы должны будем пренебречь разницей в знаменателях правых частей (3.130) (легко видеть, что приближение это высокой точности, т. к. 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, т. е. на ситуации, когда вместо (3.129) постулируется

$$x(t+\tau) = a_0 + a_1\tau + \dots + a_k\tau^k + \varepsilon(\tau), \quad (3.129')$$

где $k \geq 1$. В соотношении (3.129') 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 (3.129') при $\tau = 1$ и (3.131):

$$\hat{x}(t; 1) = \hat{x}(t+1) = \hat{a}_0^{(k)}(t; \lambda) + \hat{a}_1^{(k)}(t; \lambda) + \dots + \hat{a}_k^{(k)}(t; \lambda), \quad (3.132')$$

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, т. е.

$$\sum_{j=0}^{\infty} \lambda^j [x(t-j) - a_0 - a_1j - \dots - a_kj^k]^2 \rightarrow \min_{a_0, a_1, \dots, a_k}. \quad (3.133)$$

Решение задачи (3.133) не представляет принципиальных трудностей (см. описание взвешенного метода наименьших квадратов в п. 2.7). Мы выпишем здесь решения задачи (3.133) для случаев $k = 1$ и $k = 2$ и для рядов с «бесконечным прошлым».

1) *Случай линейного тренда, т. е. $E x(t + \tau) = a_0 + a_1 \tau$.*

$$\hat{a}_0^{(1)}(t; \lambda) = \bar{x}_\lambda(t) = (1 - \lambda) \sum_{j=0}^{\infty} \lambda^j x(t - j),$$

$$\hat{a}_1^{(1)}(t; \lambda) = (1 - \lambda) \bar{x}_\lambda(t) + (1 - \lambda)^2 \sum_{j=0}^{\infty} (j + 1) \lambda^j x(t - j). \quad (3.134)$$

Так что прогноз $\hat{x}(t; 1)$ будущего значения $x(t + 1)$ при линейном тренде ряда будет вычисляться в соответствии с (3.132')

$$\hat{x}(t; 1) = \hat{a}_0^{(1)}(t; \lambda) + \hat{a}_1^{(1)}(t; \lambda),$$

где коэффициенты $\hat{a}_0^{(1)}(t; \lambda)$ и $\hat{a}_1^{(1)}(t; \lambda)$ определяются по формулам (3.134).

Пересчет коэффициентов $\hat{a}_0^{(1)}$ и $\hat{a}_1^{(1)}$ при появлении следующего $(t + 1)$ -го члена ряда производится по формулам

$$\hat{a}_0^{(1)}(t + 1; \lambda) = (1 - \lambda^2) x(t + 1) + \lambda^2 \hat{a}_0^{(1)}(t) + \lambda^2 \hat{a}_1^{(1)}(t),$$

$$\hat{a}_1^{(1)}(t + 1; \lambda) = (1 - \lambda)^2 x(t + 1) - (1 - \lambda)^2 \hat{a}_1^{(1)}(t) - (1 - \lambda)^2 \hat{a}_0^{(1)}(t) + \hat{a}_1^{(1)}(t).$$

2) *Случай квадратичного тренда, т. е. $E x(t + \tau) = a_0 + a_1 \tau + a_2 \tau^2$.*

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 (3.133) при $t = s$, когда появляется очередное $(s + 1)$ -е наблюдение временного ряда $x(s + 1)$:

$$\hat{a}_0^{(2)}(s + 1; \lambda) = \hat{a}_0^{(2)}(s; \lambda) + \hat{a}_1^{(2)}(s; \lambda) + \hat{a}_2^{(2)}(s; \lambda) + (1 - \lambda^3) \hat{\varepsilon}(s + 1),$$

$$\hat{a}_1^{(2)}(s + 1; \lambda) = \hat{a}_1^{(2)}(s; \lambda) + 2\hat{a}_2^{(2)}(s; \lambda) + \frac{3}{2}(1 - \lambda)(1 - \lambda^2) \hat{\varepsilon}(s + 1),$$

$$\hat{a}_2^{(2)}(s + 1; \lambda) = \hat{a}_2^{(2)}(s; \lambda) + \frac{1}{2}(1 - \lambda)^3 \hat{\varepsilon}(s + 1),$$

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 (3.133), 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 3.13. Мы возвращаемся к примеру 3.10 (см. п. 3.6.1), 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. 3.14 наглядно представлено сравнение двух методов, а в табл. 3.6 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 (т. к. $\hat{x}(t; l) = 0,9x(t) + 0,1\hat{x}(t - 1; l)$), что тем не менее оказывается значительно точнее, чем прогнозы, полученные с помощью существенно более сложной методики Брауна.

Таблица 3.6. 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. 3.14. 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* . По этому поводу мы могли бы только в точности повторить наше «Замечание 2» из п. 3.5.3, в котором формулируются рекомендации по подбору параметра γ в

моделях с распределенными лагами, обладающих геометрической структурой Койка. Но даже при оптимальном подборе параметра λ модель Брауна существенно уступает в точности прогноза АРСС(0,1,1)-модели.

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 ($0 < \lambda_1 < 1, 0 < \lambda_2 < 1$)¹. 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 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, т. е. в одном году

¹ См.: *Holt C.C. Forecasting Seasonals and Trends by Exponentially Weighted Moving Averages. Carnegie Inst. Tech. Res. Mem. № 52, 1957.*

(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

$$x(\tau) = a_0(\tau) + \omega(\tau) + \delta(\tau),$$

$$a_0(\tau) = a_0(\tau-1) + a_1(\tau),$$

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 по-

мощью следующих формул обновления:

$$\hat{a}_0(\tau) = \hat{a}_0(\tau - 1) + \hat{a}_1(\tau - 1) + \lambda_1 [x(\tau) - \hat{x}(\tau - 1; 1)];$$

$$\hat{a}_1(\tau) = \hat{a}_1(\tau_1) + \lambda_1 \lambda_2 [x(\tau) - \hat{x}(\tau - 1; 1)];$$

$$\hat{\omega}(\tau) = \hat{\omega}(\tau - T) + (1 - \lambda_1) \lambda_3 [x(\tau) - \hat{x}(\tau - 1; 1)].$$

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 3.13, 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: запаваются 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 главах 6–13 первого тома данного учебника, 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 описании 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 (т. е. 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 «чисто случайными» 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. 3.3.1).

5. Если в результате проверки такой статистической гипотезы выяснилось, что имеющиеся наблюдения взаимно зависимы и неодинаково распределены (т. е. 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 используются, в основном, для описания поведения «случайных остатков») и классом *нестационарных (но однородных) временных рядов*.

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 , т. е. остаются постоянными при изменении момента времени t , для которого они подсчитываются (см. п. 3.2). Взаимозависимости, существующие между членами стационарного временного ряда, как правило, 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. 3.4.

7. Ряд $x(t)$ называется нестационарным *однородным*, если существует такой порядок k , что последовательные разности $\Delta^k x(t)$ этого ряда порядка k образуют стационарный временной ряд. Другими словами, нестационарный однородный ряд характеризуется тем, что после вычитания из него локального среднего уровня он становится стационарным. Поведение нестационарных однородных временных рядов, в том числе рядов, содержащих сезонную компоненту, в эконометрических прикладных задачах достаточно успешно описывают с помощью *моделей авторегрессии* — *проинтегрированного скользящего среднего порядка p, q и k* (АРСС(p, q, k)-моделей) или некоторых их модификаций (см. пп. 3.5.1–3.5.2).

8. 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*.

9. В ситуациях, когда нестационарные однородные временные ряды $\{x(t)\}$ и $\{y(t)\}$, $t = 1, 2, \dots, N$, являются исходными данными для построения регрессии y по x , причем, воздействие одновременного изменения одной из них (x) на другую (y) растянуто (*распределено*) во времени, большой прикладной интерес представляют так называемые *модели с распределенными лагами*. В рамках этого специального класса моделей проводится, в частности, интересный эконометрический анализ таких важных экономических явлений, как «*процесс частичного приспособления*», «*модели адаптивных ожиданий*» и др. (см. п. 3.5).

10. Важную роль в системах поддержки принятия экономических решений играет прогнозирование экономических показателей. В гл. 2 мы рассматривали методы прогнозирования, основанные на моделях регрессии. В данной главе внимание было сконцентрировано в основном 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.

11. Наиболее эффективный подход к решению задач кратко- и среднесрочного автопрогноза — это прогнозирование, основанное на использовании «подогнанных» (идентифицированных) моделей типа АРСС(p, q, k),

включая, в качестве частных случаев, и модели АР-, СС- и АРСС.

12. 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.

ГЛАВА 4. СИСТЕМЫ ЛИНЕЙНЫХ ОДНОВРЕМЕННЫХ УРАВНЕНИЙ

Изучение взаимосвязей между экономическими переменными — стержневая, базисная цель всяких эконометрических методов, моделей, приложений вне зависимости от того, что является *конечной* прикладной целью исследования — прогноз, управление или сценарные расчеты. До сих пор мы осваивали методы и модели, предназначенные для автономного анализа и оценивания *одного* уравнения, отражающего интересующие нас взаимосвязи, будь то корреляционно-регрессионные связи (т.1, гл. 10, 11), зависимости, позволяющие решать задачи типологизации объектов (т.1, гл. 12) или отбора наиболее существенных показателей (т.1, гл. 13), уравнение регрессии (т.2, гл. 2) или модель временных рядов (т.2, гл. 3).

Теперь мы расширим рамки нашего анализа. Во-первых, интересующие нас зависимости будут описываться *целой системой взаимосвязанных соотношений*, что ставит новые вопросы, связанные с правильной (т.е. непротиворечивой, идентифицируемой) комплектацией такой системы. Во-вторых, мы должны выяснить, можно ли применять технические приемы и методы, используемые в анализе и оценивании *отдельного* уравнения, последовательно к каждому из соотношений *системы*. Или же специфика взаимосвязей соотношений, из которых составлена система, *требует разработки специальных методов* для оценки одновременно всех неизвестных параметров этой системы?

Напомним, что *системой одновременных уравнений (СОУ)* называется набор взаимосвязанных регрессионных моделей, в которых одни и те же переменные могут одновременно играть роль (в различных уравнениях системы) *результатирующих показателей (объясняемых переменных)* и *предикторов (объясняющих переменных)*, и что введению в круг сформулированных выше проблем была посвящена гл. 1. Поэтому читателю, принимающему сразу за изучение данной главы, мы рекомендуем обратиться предварительно к гл. 1 и, в частности, внимательно ознакомиться в ней со «сквозным» примером 1.1 (с. 23 и далее), определениями

эндогенных, экзогенных и предопределенных переменных модели (с. 25), определением и общими формами записи (структурной и приведенной) системы одновременных уравнений (с. 27–29), формулировками проблем спецификации, идентифицируемости и верификации эконометрической модели, представленной в форме системы одновременных уравнений — СОУ (с. 32–36), а также с понятиями неидентифицируемой, идентифицируемой и свергентифицируемой СОУ (с. 34–35).

Мы же сосредоточим внимание в данной главе на проблеме идентификации СОУ, понимаемой в широком смысле, т. е. при включении в нее и проблемы идентифицируемости (отдельных уравнений и всей системы), и проблемы статистического оценивания неизвестных параметров системы. В заключительном пункте главы мы собираемся рассмотреть задачи точечного и интервального прогноза значений эндогенных переменных в рамках моделей СОУ.

Начнем рассмотрение указанных выше проблем с анализа относительно простого примера — «модели спроса-предложения».

4.1. Модель спроса-предложения как пример системы одновременных уравнений

Мы располагаем наблюдениями $(\tilde{x}_t, \tilde{y}_t^{(1)}, \tilde{y}_t^{(2)})$, $t = 1, 2, \dots, n$, характеризующими динамику среднедушевого дохода (\tilde{x}_t) , предложения некоторого товара $(\tilde{y}_t^{(1)})$, которое совпадает со спросом на этот товар в условиях равновесия и цены на него $(\tilde{y}_t^{(2)})$. Нам будет удобнее в дальнейшем оперировать не с самими наблюдаемыми переменными $(\tilde{x}, \tilde{y}^{(1)}$ и $\tilde{y}^{(2)})$, а с их отклонениями от своих средних уровней, т. е., соответственно, с наблюдениями $x_t = \tilde{x}_t - \bar{\tilde{x}}$, $y_t^{(1)} = \tilde{y}_t^{(1)} - \bar{\tilde{y}}^{(1)}$ и $y_t^{(2)} = \tilde{y}_t^{(2)} - \bar{\tilde{y}}^{(2)}$ (черта сверху означает усреднение соответствующей переменной по всем ее наблюдаемым значениям). Для *центрированных* таким образом переменных известные *линейные уравнения спроса и предложения* запишутся в виде

$$\begin{cases} y_t^{(1)} = b_1 y_t^{(2)} + \delta_t^{(1)} & - \text{уравнение предложения,} \\ y_t^{(1)} = b_2 y_t^{(2)} + c x_t + \delta_t^{(2)} & - \text{уравнение спроса,} \end{cases} \quad (4.1)$$

$$t = 1, 2, \dots, n,$$

где b_1 , b_2 и c — некоторые неизвестные коэффициенты, а каждая из последовательностей $\{\delta_t^{(1)}\}$ и $\{\delta_t^{(2)}\}$ ($t = 1, 2, \dots, n$) представляет собой ряд взаимно не коррелированных случайных величин, имеющих нулевые средние значения и дисперсии, соответственно, σ_1^2 и σ_2^2 , не зависящие от t . Кроме

того, предполагается, что регрессионные остатки $\delta_t^{(1)}$ и $\delta_t^{(2)}$ не коррелированы между собой, а также не коррелированы с переменной x_t .

Заметим, что в модели (4.1)~(4.2) значения цены $y^{(2)}$ и спроса-предложения $y^{(1)}$ формируются как бы «внутри» нее, в процессе взаимодействия связей, описываемых этой моделью. В то же время значения среднедушевого дохода x поступают в систему как характеристики *внешней* среды, в которой взаимодействуют эти анализируемые связи. Поэтому в данной модели спрос-предложение и цена отнесены к эндогенным переменным, а доход — к экзогенным.

Отправляясь от описанной выше спецификации модели, проведем ее анализ по схеме: а) основные структурные характеристики системы (число уравнений, эндогенных и экзогенных переменных, структурная и приведенные формы); б) идентифицируемость уравнений системы; в) статистическое оценивание параметров системы (включая анализ свойств обычных МНК-оценок, понятие о косвенном и двухшаговом методах наименьших квадратов и связь с методом инструментальных переменных).

Основные структурные характеристики модели. Анализируемая модель описана системой (4.1)–(4.2), состоящей из *двух* уравнений (в обозначениях гл. 1 $m = 2$) с *двумя* эндогенными и *одной* ($p = 1$) предопределенной (она же — экзогенная) переменными. Балансовых тождеств система не содержит (т. е. $m_2 = 0$ и, следовательно, $m_1 = m$). *Структурная форма* модели (см. определение 1.2) задана уравнениями (4.1)–(4.2). Решая эти уравнения относительно эндогенных переменных $y_t^{(1)}$ и $y_t^{(2)}$, получаем *приведенную форму* СОУ (см. определение 1.3):

$$\begin{aligned} y_t^{(1)} &= \pi_1 x_t + \varepsilon_t^{(1)}, \\ y_t^{(2)} &= \pi_2 x_t + \varepsilon_t^{(2)}, \quad t = 1, 2, \dots, n, \end{aligned} \quad (4.3)$$

где

$$\pi_1 = \frac{b_1 c}{b_1 - b_2}, \quad \pi_2 = \frac{c}{b_1 - b_2}, \quad (4.3')$$

$$\varepsilon_t^{(1)} = \frac{b_1 \delta_t^{(2)} - b_2 \delta_t^{(1)}}{b_1 - b_2}, \quad \varepsilon_t^{(2)} = \frac{\delta_t^{(2)} - \delta_t^{(1)}}{b_1 - b_2}. \quad (4.3'')$$

В рамках сделанных выше предположений о постоянстве дисперсий $D\delta_t^{(1)}$ и $D\delta_t^{(2)}$ и о некоррелированности $\delta_t^{(1)}$ и $\delta_t^{(2)}$ между собой и с экзогенной переменной x_t из (4.3'') следует, что *случайные остатки* $\varepsilon_t^{(1)}$ и $\varepsilon_t^{(2)}$ *приведенной формы также имеют не зависящие от t дисперсии, не*

коррелированы между собой и с экзогенной переменной x_i в уравнениях (4.3).

Идентифицируемость уравнений системы. Следуя определениям 1.4~1.7, попробуем ответить на вопрос об идентифицируемости параметров b_1, b_2 и c структурной формы, т. е. о возможности их однозначного выражения через параметры π_1 и π_2 приведенной формы. Поделив первое из уравнений (4.3') на второе, получаем

$$b_1 = \frac{\pi_1}{\pi_2}. \quad (4.4)$$

Однако соотношений (4.3') оказывается недостаточно для того, чтобы определить из них значения коэффициентов b_2 и c второго уравнения структурной формы. Это значит, что первое уравнение структурной формы анализируемой СОУ идентифицируемо, а второе и вся система — неидентифицируемы.

Статистическое оценивание параметров СОУ. Попробуем сначала формально воспользоваться обычным методом наименьших квадратов, примененным в отдельности к каждому из уравнений (4.1) и (4.2), и проанализируем состоятельность полученных оценок.

1) *Несостоятельность обычных МНК-оценок параметров уравнений структурной формы.* Начнем с уравнения (4.1) и воспользуемся формулой (2.18) для вычисления МНК-оценки в простейшем варианте модели регрессии с единственной объясняющей переменной (в уравнении (4.1) ее роль выполняет цена товара $y^{(2)}$):

$$\hat{b}_{1\text{ мнк}} = \frac{\sum_{i=1}^n y_i^{(1)} y_i^{(2)}}{\sum_{i=1}^n (y_i^{(2)})^2}. \quad (4.5)$$

Как обычно это делается при исследовании свойств несмещенности и состоятельности оценки, подставим в правую часть (4.5) вместо переменной $y_i^{(1)}$ ее модельное выражение из (4.1):

$$\hat{b}_{1\text{ мнк}} = \frac{\sum_{i=1}^n (b_1 y_i^{(2)} + \delta_i^{(1)}) y_i^{(2)}}{\sum_{i=1}^n (y_i^{(2)})^2} = b_1 + \frac{\frac{1}{n} \sum_{i=1}^n \delta_i^{(1)} y_i^{(2)}}{\frac{1}{n} \sum_{i=1}^n (y_i^{(2)})^2}. \quad (4.6)$$

Исследуем поведение (при $n \rightarrow \infty$) 2-го слагаемого в правой части (4.6), поскольку от него зависят свойства несмещенности и состоятельно-

сти оценки $\hat{b}_{1\text{МНК}}$. Если принять естественное допущение о существовании пределов (при $n \rightarrow \infty$) числителя и знаменателя этого отношения, то в силу закона больших чисел будем иметь:

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n} \sum_{i=1}^n \delta_i^{(1)} y_i^{(2)} \right) = E(\delta_i^{(1)} y_i^{(2)}) = \text{cov}(\delta_i^{(1)}, y_i^{(2)}), \quad (4.7)$$

$$\lim_{n \rightarrow \infty} \left[\frac{1}{n} \sum_{i=1}^n (y_i^{(2)})^2 \right] = E(y_i^{(2)})^2 = D y_i^{(2)} \quad (4.8)$$

(в этих соотношениях учтено, что по построению $E\delta_i^{(1)} = E y_i^{(2)} = 0$ при всех t , а сходимость пределов понимается «по вероятности»). Остается выразить значения $\text{cov}(\delta_i^{(1)}, y_i^{(2)})$ и $D y_i^{(2)}$ в терминах параметров модели. Для этого воспользуемся выражением $y_i^{(2)}$ из второго соотношения приведенной формы (4.3) анализируемой СОУ (с учетом (4.3') и (4.3'')):

$$\begin{aligned} \text{cov}(\delta_i^{(1)}, y_i^{(2)}) &= E \left[\delta_i^{(1)} \left(\pi_2 x_i + \frac{\delta_i^{(2)} - \delta_i^{(1)}}{b_1 - b_2} \right) \right] = \frac{c}{b_1 - b_2} \text{cov}(\delta_i^{(1)}, x_i) \\ &+ \frac{\text{cov}(\delta_i^{(1)}, \delta_i^{(2)}) - D\delta_i^{(1)}}{b_1 - b_2} = -\frac{\sigma_1^2}{b_1 - b_2}; \end{aligned} \quad (4.9)$$

$$D y_i^{(2)} = E \left(\frac{c x_i + (\delta_i^{(2)} - \delta_i^{(1)})}{b_1 - b_2} \right)^2 = \frac{1}{(b_1 - b_2)^2} (c^2 D x_i + D\delta_i^{(1)} + D\delta_i^{(2)}). \quad (4.10)$$

Теперь мы можем вычислить предельное выражение (при $n \rightarrow \infty$) для $\hat{b}_{1\text{МНК}}$, вернувшись к (4.6) и учтя (4.7)–(4.10):

$$\lim_{n \rightarrow \infty} \hat{b}_{1\text{МНК}} = b_1 + \frac{-\frac{\sigma_1^2}{b_1 - b_2}}{\frac{c^2 D x_i + \sigma_1^2 + \sigma_2^2}{(b_1 - b_2)^2}} = b_1 - (b_1 - b_2) \frac{\sigma_1^2}{c^2 D x_i + \sigma_1^2 + \sigma_2^2}. \quad (4.11)$$

Мы видим, что обычная МНК-оценка $\hat{b}_{1\text{МНК}}$ параметра b_1 имеет в общем случае (т. е. при $b_1 \neq b_2$ и $\sigma_1^2 \neq 0$) асимптотически неустранимое смещение, а следовательно, она не является ни несмещенной, ни (что более важно) состоятельной. Поскольку результат — отрицательный, оценивание 2-го уравнения системы с помощью обычного МНК, очевидно, не имеет смысла.

2) *Использование косвенного метода наименьших квадратов.* При переходе к приведенной форме (4.3)–(4.3')–(4.3'') было подмечено, что ре-

грессионные остатки в уравнениях (4.3) удовлетворяют условиям классической модели регрессии. Следовательно, оценки параметров π_1 и π_2 могут быть получены с помощью обычного метода наименьших квадратов, и при этом они будут несмещенными и состоятельными. Но в случае идентифицируемости параметров структурной формы они, по определению, могут быть выражены через параметры приведенной формы. Поэтому если в эти выражения мы вместо параметров π_j подставим их МНК-оценки $\hat{\pi}_j$, то получим (в соответствии с теоремой Слуцкого) состоятельные оценки для соответствующих идентифицируемых параметров структурной формы. Такой способ оценивания идентифицируемых параметров структурной формы называют *косвенным методом наименьших квадратов*. В нашем примере идентифицируемым является только 1-е уравнение. В соответствии с (4.4) получаем состоятельную оценку косвенного метода наименьших квадратов для параметра b_1 :

$$\hat{b}_{1\text{МНК}} = \frac{\hat{\pi}_{1\text{МНК}}}{\hat{\pi}_{2\text{МНК}}}, \quad (4.12)$$

где $\hat{\pi}_{j\text{МНК}}$ — обычные МНК-оценки параметров π_j в уравнениях (4.3) приведенной формы анализируемой СОУ.

3) *Использование метода инструментальных переменных*. Причина непригодности обычного метода наименьших квадратов в оценивании параметров 1-го уравнения структурной формы, как мы знаем (см. п. 2.10.2), — в коррелированности объясняющей переменной (в нашем случае — $y_t^{(2)}$) с регрессионными остатками (в нашем случае — с $\delta_t^{(1)}$). В соответствии с рекомендациями п. 2.10.2 в уравнении (4.1) в качестве инструментальной переменной естественно использовать доход x , поскольку:

- x_t не коррелирован по условию с $\delta_t^{(1)}$;
- x_t по построению достаточно сильно коррелирован с $y_t^{(2)}$ (что, в частности, подтверждается осмысленностью рассмотрения 2-го уравнения приведенной формы (4.3)).

Тогда в соответствии с (2.150') оценка \hat{b}_{1x} параметра b_1 по методу инструментальных переменных будет

$$\hat{b}_{1x} = \frac{\sum_{t=1}^n y_t^{(1)} x_t}{\sum_{t=1}^n y_t^{(2)} x_t}. \quad (4.13)$$

Если принять во внимание, что МНК-оценки $\hat{\pi}_j$ параметров π_j урав-

нений приведенной формы (4.3) имеют вид

$$\hat{\pi}_1 = \frac{\sum_{t=1}^n y_t^{(1)} x_t}{\sum_{t=1}^n x_t^2}, \quad \hat{\pi}_2 = \frac{\sum_{t=1}^n y_t^{(2)} x_t}{\sum_{t=1}^n x_t^2},$$

то сравнение оценки (4.12), полученной с помощью косвенного метода наименьших квадратов, с оценкой (4.13), полученной с помощью метода инструментальных переменных, приводит к выводу, что в *данном случае эти оценки совпадают*.

4) *Двухшаговый метод наименьших квадратов*. В ходе применения косвенного метода наименьших квадратов мы выяснили, что 2-е уравнение анализируемой системы (т.е. уравнение (4.2)) неидентифицируемо, и, соответственно, параметры b_2 и c остались не оцененными. Прямому применению к этому уравнению *обычного* метода наименьших квадратов препятствует коррелированность значений объясняющей переменной $y_t^{(2)}$ с регрессионными остатками $\delta_t^{(2)}$. Чтобы обойти эту трудность, предлагается следующая двухэтапная процедура оценивания. *На 1-м этапе*, чтобы избавиться от $y_t^{(2)}$ в правой части уравнения (4.2), мы заменяем эту переменную ее аппроксимацией с помощью линейной функции от x . Аппроксимацию строим с помощью *обычного* метода наименьших квадратов, который, с учетом центрированности обеих переменных, дает:

$$y_t^{(2)} = \hat{\theta}_{y^{(2)}/x} x_t + \varepsilon_t, \quad (4.14)$$

где МНК-оценка параметров $\hat{\theta}_{y^{(2)}/x}$ парной регрессии $y^{(2)}$ по x , как известно, имеет вид

$$\hat{\theta}_{y^{(2)}/x} = \frac{\sum_{t=1}^n y_t^{(2)} x_t}{\sum_{t=1}^n x_t^2}, \quad (4.15)$$

а остатки ε_t в соответствии со свойствами МНК не коррелированы с предиктором x_t .

Теперь, *на 2-м этапе оценивания*, мы можем возвратиться к оцениваемому уравнению (4.2) и заменить в нем объясняющую переменную $y_t^{(2)}$ ее аппроксимацией (4.14). В результате вместо (4.2) получаем уравнение

$$y_t^{(1)} = b_2(\hat{\theta}_{y^{(2)}/x} x_t + \varepsilon_t) + c x_t + \delta_t^{(2)} = (b_2 \hat{\theta}_{y^{(2)}/x} + c) x_t + (\delta_t^{(2)} + b_2 \varepsilon_t). \quad (4.2')$$

В этом уравнении остаток $\delta_t^{(2)} + b_2 \varepsilon_t$ по условию и по построению не коррелирован с предиктором x_t , а поэтому мы можем использовать обычный МНК для оценки параметра $\theta_{y^{(1)}/x} = b_2 \hat{\theta}_{y^{(2)}/x} + c$:

$$\hat{\theta}_{y^{(1)}/x} = \frac{\sum_{t=1}^n y_t^{(1)} x_t}{\sum_{t=1}^n x_t^2}. \quad (4.16)$$

Таким образом мы получаем линейное соотношение, связывающее между собой оцениваемые параметры b_2 и c :

$$b_2 \hat{\theta}_{y^{(2)}/x} + c = \hat{\theta}_{y^{(1)}/x}, \quad (4.17)$$

в котором величины $\hat{\theta}_{y^{(2)}/x}$ и $\hat{\theta}_{y^{(1)}/x}$ определены соотношениями (4.15) и (4.16). Для того, чтобы отсюда извлечь оценки \hat{b}_2 и \hat{c} , надо вернуться к соотношениям (4.3'), связывающим между собой параметры приведенной и структурной форм:

$$\begin{cases} \pi_1 = \frac{b_1 c}{b_2 - b_1}, \\ \pi_2 = \frac{c}{b_2 - b_1}, \\ b_2 \hat{\theta}_{y^{(2)}/x} + c = \hat{\theta}_{y^{(1)}/x}. \end{cases} \quad (4.18)$$

Решая эти три уравнения относительно параметров структурной формы b_1 , b_2 и c и воспользовавшись значениями уже подсчитанных оценок $\hat{\pi}_1$, $\hat{\pi}_2$, $\hat{\theta}_{y^{(1)}/x}$, $\hat{\theta}_{y^{(2)}/x}$, а также известной теоремой Слуцкого, получаем

$$\hat{b}_1 = \frac{\hat{\pi}_1}{\hat{\pi}_2}; \quad \hat{b}_2 = \frac{\hat{\theta}_{y^{(1)}/x} + \hat{\pi}_1}{\hat{\theta}_{y^{(2)}/x} + \hat{\pi}_2}; \quad \hat{c} = \frac{\hat{\theta}_{y^{(1)}/x} \hat{\pi}_2 - \hat{\theta}_{y^{(2)}/x} \hat{\pi}_1}{\hat{\theta}_{y^{(2)}/x} + \hat{\pi}_2}.$$

Описанный (для данного частного случая) способ получения оценок параметров *отдельного уравнения системы* носит название *двухшагового метода наименьших квадратов*. Ниже (см. п. 4.3.3) мы опишем этот метод для *общего* случая СОУ.

З а м е ч а н и е. Заменяя в ходе 1-го шага процедуры объясняющую переменную $y_t^{(2)}$ ее аппроксимацией $\hat{\theta}_{y^{(2)}/x} x_t$, мы тем самым сознательно идем на *чистую мультиколлинеарность* (см. п. 2.4.1) в получающейся модели

$$y_t^{(1)} = b_2 (\hat{\theta}_{y^{(2)}/x} x_t) + c x_t + \varepsilon(t).$$

Так что мы, конечно, не можем претендовать на получение состоятельных МНК-оценок *отдельно* для коэффициентов b_2 и c : в силу вырожденности матрицы наблюдений объясняющих переменных $x_t^{(1)} = \hat{\theta}_{y^{(2)}/x} x_t$ и $x_t^{(2)} = x_t$ получающаяся в результате применения МНК система нормальных уравнений будет содержать *два одинаковых уравнения* для определения оценок параметров b_2 и c . Однако с учетом уже имеющихся у нас соотношений (4.3'), связывающих между собой параметры структурной и приведенной форм системы, мы получаем как раз ту связь между этими параметрами, которой не хватало для определения оценок параметров неидентифицируемого уравнения.

Итак, мы обсудили и предварительно «прочувствовали» на примере модели спроса-предложения основные проблемы, возникающие при анализе СОУ: спецификацию, идентифицируемость, статистическое оценивание параметров модели. Теперь мы приступаем к общему описанию этих проблем и методов их решения.

4.2. Условия идентифицируемости уравнений системы

Сама проблема идентифицируемости СОУ, ее отдельных уравнений и даже отдельных параметров была сформулирована в гл. 1 (см. с. 35–36). Эта проблема логически предшествует статистическому оцениванию параметров системы, поскольку от ее решения зависит и выбор методов оценивания. Отсутствие идентифицируемости СОУ означает, что существует бесконечно много моделей, не противоречащих имеющимся в нашем распоряжении наблюдениям. В результате анализа проблемы идентифицируемости конкретного структурного параметра системы мы приходим к *одной из трех принципиально возможных ситуаций*: 1) этот параметр может быть однозначно выражен через коэффициенты приведенной системы; 2) структурный параметр допускает несколько разных вариантов определения с помощью косвенного МНК; 3) структурный параметр не может быть выражен через параметры приведенной формы. Очевидно, идентифицируемость или неидентифицируемость системы, ее уравнений и параметров зависят только от *внутренней структуры модели* (числа уравнений, соотношения количеств эндогенных и предопределенных переменных в системе и в каждом уравнении, мультиколлинеарности анализируемых переменных, некоторых свойств матриц структурных коэффициентов), но никак не связаны со статистическими свойствами исходных наблюдений, например с их количеством.

Следует, в основном, обозначениям, введенным в гл. 1 (см. (1.6), (1.7)),

(1.6'), (1.6''), (1.6^а) и т. д.), мы будем полагать, что балансовые тождества уже исключены из анализируемой системы (см. замечание на с. 27), так что ее структурная форма имеет вид:

$$BY_t + CX_t = \Delta_t, \quad t = 1, 2, \dots, n, \quad (4.19)$$

где $B = (\beta_{ij})$ — матрица размерности $\bar{m} \times m_Y$ (т. е. $i = 1, 2, \dots, \bar{m}$; $j = 1, 2, \dots, m_Y$) коэффициентов при m_Y эндогенных переменных $Y_t = (y_t^{(1)}, y_t^{(2)}, \dots, y_t^{(m_Y)})^T$; $C = (c_{ij})$ — матрица размерности $\bar{m} \times p$ (т. е. $i = 1, 2, \dots, \bar{m}$; $j = 1, 2, \dots, p$) коэффициентов при p предопределенных переменных $X_t = (x_t^{(1)}, x_t^{(2)}, \dots, x_t^{(p)})^T$, в состав которых, если необходимо, включен и свободный член; а вектор-столбец случайных остатков $\Delta_t = (\delta_t^{(1)}, \delta_t^{(2)}, \dots, \delta_t^{(m)})^T$ удовлетворяет в общем случае следующим условиям: $E\Delta_t \equiv 0_{\bar{m}}$; ковариационная матрица остатков $\Sigma_\Delta = E(\Delta_t \Delta_t^T)$ положительно определена и не зависит от t ; векторы Δ_{t_1} и Δ_{t_2} взаимно не коррелированы при $t_1 \neq t_2$, а $\delta_t^{(i)}$ не коррелированы со всеми предопределенными переменными системы, т. е. $E[\delta_t^{(i)}(x_t^{(j)} - Ex_t^{(j)})] = 0$ при $i = 1, 2, \dots, \bar{m}$, $j = 1, 2, \dots, p$. Коэффициенты β_{ij} пронормированы с помощью условия $\beta_{ii} = 1$.

Таким образом рассматривается система из \bar{m} уравнений, содержащая m_Y эндогенных и p предопределенных (т. е. экзогенных и лаговых эндогенных) переменных.

Прежде всего, для того, чтобы мы могли уединить эндогенные переменные Y_t , выразив их через предопределенные переменные X_t (т. е. — осуществить переход к приведенной форме и установить соотношения, связывающие коэффициенты структурной и приведенной форм), необходимо потребовать, чтобы матрица B была *квадратной* и *обратимой* (т. е. невырожденной). Так мы приходим к формулировке 1-го (необходимого) условия идентифицируемости:

1-е условие (необходимое) идентифицируемости: *число уравнений системы (\bar{m}) должно быть равно числу анализируемых эндогенных переменных (m_Y), т. е. $\bar{m} = m_Y = m$, а матрица B должна быть невырожденной.*

Тогда приведенная форма анализируемой системы будет иметь вид:

$$Y_t = \Pi X_t + \varepsilon_t, \quad t = 1, 2, \dots, n, \quad (4.19^a)$$

где

$$\Pi = -B^{-1}C \quad - \quad (4.20)$$

матрица размерности $m \times p$ коэффициентов приведенной формы, а $\varepsilon_t = (\varepsilon_t^{(1)}, \varepsilon_t^{(2)}, \dots, \varepsilon_t^{(m)})^T = B^{-1}\Delta_t$ — вектор-столбец остатков.

Очевидно, необходимым условием возможности оценить все коэффициенты π_{ij} ; матрицы Π параметров приведенной формы (4.19^а) является требование, чтобы предопределенные переменные X_t не были мультиколлинеарны. Итак,

2-е условие (необходимое) идентифицируемости: матрица наблюдений предопределенных переменных $X = (x_t^{(j)})$, $t = 1, 2, \dots, n$; $j = 1, 2, \dots, p$, — размерности $n \times p$ должна иметь полный ранг p (очевидно, при этом число наблюдений n должно существенно превышать общее число анализируемых переменных $m + p$).

На интуитивном уровне достаточно очевидна «перегруженность» модели (4.19) неизвестными значениями параметров β_{ij} и c_{ik} ($i = 1, 2, \dots, m$; $k = 1, 2, \dots, p$). Более того, можно доказать (см., например, [Джонстон Дж.], с. 357), что если не дополнить модель (4.19) никакими априорными ограничениями относительно числовых значений этих параметров, то ни одно из уравнений не может быть идентифицируемо. При этом априорные ограничения носят чаще всего *исключающий* характер, т. е. они определяют в каждом (i -м) уравнении «адреса» (i, j) и (i, k) тех коэффициентов, которые априори считаются нулевыми (априорные ограничения могут быть сформулированы и в виде определенных линейных связей, априори существующих между оцениваемыми коэффициентами). Поставим в соответствие каждому (i -му) уравнению системы (4.19) ($m + p$)-мерный булевский (т. е. состоящий только из нулей и единиц) вектор $\gamma_i = (\gamma_i^{(1)}, \dots, \gamma_i^{(m)}; \gamma_i^{(m+1)}, \dots, \gamma_i^{(m+p)})$, задающий исключающие априорные ограничения для параметров этого уравнения по следующему правилу: нулевые значения компонент $\gamma_i^{(k)}$ определяют «адреса» отсутствующих в уравнении переменных, или, что то же, априори равных нулю параметров i -го уравнения структурной формы.

3-е условие (необходимое) идентифицируемости: среди *исключающих априорных ограничений* $\gamma_i = (\gamma_i^{(1)}, \dots, \gamma_i^{(m+p)})$, $i = 1, 2, \dots, m$, не должно быть одинаковых.

Пример 4.1. Чтобы пояснить, к чему может приводить нарушение этого условия, вернемся к рассмотренному в предыдущем пункте примеру (4.1)–(4.2). Очевидно, анализируемая модель спроса-предложения может быть записана в следующем виде, эквивалентном (4.1)–(4.2):

$$y_t^{(1)} = b_1 y_t^{(2)} + \delta_t^{(1)}, \quad (4.1')$$

$$y_t^{(3)} = b_2 y_t^{(2)} + c x_t + \delta_t^{(2)}, \quad (4.2')$$

$$y_t^{(1)} = y_t^{(3)}. \quad (4.2'')$$

По сравнению с прежними обозначениями изменено только обозначение

спроса: в данной записи для спроса введена специальная переменная $y_i^{(3)}$ и, соответственно, появилось одно тождество (4.2'').

Желая несколько упростить модель, положим, что спрос $y_i^{(3)}$, так же, как и предложение $y_i^{(1)}$, зависит *только* от цены товара $y_i^{(2)}$ (т.е. не зависит от дохода). Тем самым мы воспользовались *двумя одинаковыми исключаящими априорными ограничениями* — и в первом, и во втором уравнениях системы «обнулили» коэффициент при доходе x . В результате получаем систему

$$\begin{cases} y_i^{(1)} = b_1 y_i^{(2)} + \delta_i^{(1)}, \\ y_i^{(3)} = b_2 y_i^{(2)} + \delta_i^{(2)}, \\ y_i^{(1)} = y_i^{(3)}. \end{cases}$$

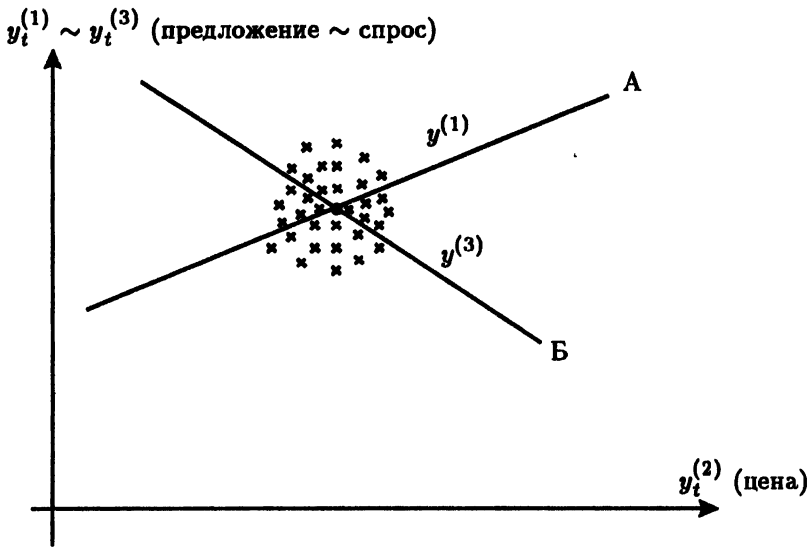


Рис. 4.1. Линии предложения (А), спроса (Б) и наблюдения (крестики), произведенные в условиях равновесия

На плоскости, где по оси ординат откладываются значения предложения и спроса, а на оси абсцисс — значения цены, равновесие представляется как пересечение линий предложения (А) и спроса (Б), см. рис. 4.1. Поскольку мы собираем исходные статистические данные $(y_i^{(2)}, y_i^{(1)}, y_i^{(3)})$, $t = 1, 2, \dots, n$, в условиях равновесия, то различие в наблюдаемых значениях обусловлено только случайными остатками $\delta_i^{(1)}$ и $\delta_i^{(2)}$, что, кстати,

подтверждается и приведенной формой этой модели

$$y_i^{(1)} = \frac{1}{b_1 - b_2} (b_1 \delta_i^{(2)} - b_2 \delta_i^{(1)}),$$

$$y_i^{(2)} = \frac{1}{b_1 - b_2} (\delta_i^{(2)} - \delta_i^{(1)}),$$

так как правые части уравнений выражаются только в терминах случайных остатков модели. Очевидно, располагая только разбросанными около точки равновесия значениями случайных остатков $\varepsilon_i^{(1)}$ и $\varepsilon_i^{(2)}$ (см. (4.3') и (4.3'')), мы ничего не можем сказать о самих функциях предложения (А) и спроса (Б). А это и значит, что мы не можем оценить параметры b_1 и b_2 в уравнениях (4.1') и (4.2'), т. е. оба анализируемых уравнения неидентифицируемы.

Заметим, что к точно такому же результату мы пришли бы, если бы пошли по линии некоторого усложнения (а не упрощения) модели (4.1)–(4.2), введя в 1-е уравнение системы доход x . Это снова означало бы использование *идентичных* векторов априорных ограничений и привело бы к неидентифицируемости системы.

Выведем еще два важных условия идентифицируемости отдельного уравнения системы (4.19). Для этого выпишем i -е уравнение этой системы, полагая (без ограничения общности), что анализируемые переменные системы перенумерованы таким образом, что участвующие в этом уравнении p_i предопределенных переменных ($p_i \leq p$) и m_i эндогенных переменных ($m_i \leq m$) идут первыми в общих перечнях, соответственно, предопределенных и эндогенных переменных. Тогда это уравнение может быть записано в виде

$$(\beta_{i1} \dots \beta_{im_i} 0 \dots 0)(y_i^{(1)} \dots y_i^{(m_i)} y_i^{(m_i+1)} \dots y_i^{(m)})^T + (c_{i1} \dots c_{ip_i} 0 \dots 0)(x_i^{(1)} \dots x_i^{(p_i)} x_i^{(p_i+1)} \dots x_i^{(p)})^T = \delta_i^{(i)}, \quad (4.21)$$

или, что то же,

$$\beta^T(i)Y_t(i) + c^T(i)X_t(i) = \delta_t^{(i)}, \quad (4.21')$$

где

$$\beta(i) = (\beta_{i1}, \dots, \beta_{im_i})^T, \quad c(i) = (c_{i1}, \dots, c_{ip_i})^T, \quad (4.21'')$$

$$Y_t(i) = (y_t^{(1)}, \dots, y_t^{(m_i)})^T, \quad X_t(i) = (x_t^{(1)}, \dots, x_t^{(p_i)})^T. \quad (4.21''')$$

Нас будут интересовать соотношения, связывающие между собой коэффициенты структурной и приведенной форм именно i -го уравнения системы (4.19) с точки зрения возможности однозначно выразить коэффициенты структурной формы $\beta(i)$ и $c(i)$ по известным значениям коэффициентов приведенной формы (т. е. по элементам матрицы Π из (4.19^а)).

Мы получили систему уравнений, связывающих элементы i -й строки структурной формы СОУ с параметрами приведенной формы. Эта система, как мы видим, распалась на две подсистемы: (4.25) и (4.24). По-видимому, сначала надо попытаться решить подсистему (4.25) относительно коэффициентов $\beta(i)$, а затем, подставив найденные решения в (4.24), решить эту подсистему относительно параметров $c(i)$.

Поэтому определим, в первую очередь, условия, при которых подсистема (4.25) имеет хотя бы одно решение. Соотношения (4.25) представляют собой систему из $p - p_i$ уравнений относительно $\beta_{i1}, \beta_{i2}, \dots, \beta_{im_i}$ с $m_i - 1$ неизвестными (поскольку в соответствии с условием нормировки один из коэффициентов β_{ij} равен единице). Для того чтобы параметры $\beta(i)$ в подсистеме (4.25) можно было бы выразить через элементы матрицы $\bar{\Pi}_X(i)$, необходимо, чтобы число уравнений в (4.25) было бы не меньше числа неизвестных, т. е.

$$p - p_i \geq m_i - 1.$$

Таким образом, мы пришли к еще одному необходимому условию идентифицируемости уравнения системы.

4-е условие (необходимое) идентифицируемости (правило порядка): число исключенных (при спецификации модели) из i -го уравнения системы предопределенных переменных (т. е. число $(p - p_i)$) должно быть не меньше числа включенных в него эндогенных переменных, уменьшенного на единицу. Сформулированное условие называется правилом порядка. Заметим, что выполнение условия $p - p_i = m_i - 1$ является необходимым условием точной идентификации i -го уравнения, в то время как при $p - p_i > m_i - 1$ уравнение будет *сверидентифицируемым*. Из анализа подсистемы уравнений (4.25) можно извлечь также необходимое и достаточное условие идентифицируемости i -го уравнения СОУ. Известно, в частности, что для разрешимости системы (4.25) относительно $\beta(i)$ необходимо и достаточно, чтобы матрица $\bar{\Pi}_X(i)$ имела бы ранг, равный числу неизвестных (т. е. $m_i - 1$). Таким образом получаем

5-е условие идентифицируемости отдельного уравнения системы (условие называется ранговым и является необходимым и достаточным): ранг матрицы $\bar{\Pi}_X(i) = m_i - 1$.

Следует отметить, что в отличие от правила порядка (см. 4-е условие), которое может проверяться на стадии спецификации системы, соблюдение рангового условия мы можем проверить в общем случае только после вычисления матрицы $\bar{\Pi}_X(i)$, т. е. после применения обычного МНК к приведенной форме (4.22) анализируемой системы. Правда, в отдельных простых случаях это можно сделать и до применения МНК, просто анализируя элементы матрицы $\bar{\Pi}_X(i)$, выраженные через параметры струк-

турной формы β_{ij} и c_{kl} (что мы и продемонстрируем сейчас на примере).

Пример 4.2. Снова несколько видоизменим пример, рассмотренный в п. 4.1. Полагая, что и ставка процента $\bar{x}_t^{(2)}$ может влиять на спрос, мы решили дополнить наши наблюдения за среднедушевым доходом (\bar{x}_t), предложением некоторого товара ($\bar{y}_t^{(1)}$) и ценой на него ($\bar{y}_t^{(2)}$) регистрацией во времени этой величины. Тогда, как и прежде, переходя от значений самих анализируемых величин к их *отклонениям от средних* (соответственно $x_t, y_t^{(1)}, y_t^{(2)}$ и $x_t^{(2)}$) и вводя в модель еще одну экзогенную переменную ($x_t^{(2)}$) в виде аргумента функции спроса (т. е. в уравнение (4.2)), мы получаем СОУ:

$$y_t^{(1)} = b_1 y_t^{(2)} + \delta_t^{(1)} \quad - \text{уравнение предложения}; \quad (4.1^a)$$

$$y_t^{(1)} = b_2 y_t^{(2)} + c_1 x_t^{(1)} + c_2 x_t^{(2)} + \delta_t^{(2)} \quad - \text{уравнение спроса}, \quad (4.2^a)$$

$$t = 1, 2, \dots, n,$$

в которой для удобства доход (в отклонениях от среднего) и коэффициент при нем мы теперь обозначаем, соответственно, $x_t^{(1)}$ и c_1 . *Структурные параметры этой системы*: число уравнений равно общему числу эндогенных переменных и равно $m = 2$; общее число предопределенных переменных $p = 2$; число эндогенных и предопределенных переменных, включенных в 1-е уравнение, соответственно, $m_1 = 2, p_1 = 0$; число эндогенных и предопределенных переменных, включенных во 2-е уравнение, соответственно, $m_2 = 2, p_2 = 2$; матрица **B** коэффициентов при эндогенных переменных (при канонической записи системы)

$$B = \begin{pmatrix} 1 & -b_1 \\ -\frac{1}{b_2} & 1 \end{pmatrix}.$$

Таким образом, *1-е условие* идентифицируемости будет выполненным, если потребовать $\det B \neq 0$, т. е. $b_2 \neq b_1$ (и $b_2 \neq 0$).

Для проверки *2-го условия* надо убедиться в том, что наблюдения $x_t^{(1)}$ и $x_t^{(2)}$ ($t = 1, 2, \dots, n$) не связаны пропорциональной зависимостью, т. е. что $x_t^{(2)} \neq \lambda x_t^{(1)}$.

Очевидно выполнение *3-го общего условия* идентифицируемости системы, т. к. векторы исключающих ограничений первого (γ_1) и второго (γ_2) уравнений системы различны, а именно: $\gamma_1 = (1, 1, 0, 0)$; $\gamma_2 = (1, 1, 1, 1)$.

Теперь сосредоточимся на проверке идентифицируемости *отдельного* (например, 1-го) уравнения системы. *Правило порядка* (4-е условие) выполняется со *строгим* знаком неравенства, т. к. число исключенных в 1-м

уравнении предопределенных переменных равно двум ($p - p_1 = 2 - 0$), а число включенных в него эндогенных переменных, уменьшенное на единицу, равно единице ($m_1 - 1 = 2 - 1$). Это означает, что 1-е уравнение системы является *сверхидентифицируемым*, т. е. его структурные параметры (в данном случае это один параметр b_1) допускают *неоднозначное* выражение через параметры приведенной формы. Этот факт в анализируемом примере легко проверить непосредственно. Действительно, решение системы (4.1^а)–(4.2^а) относительно $y_i^{(1)}$ и $y_i^{(2)}$ дает нам приведенную форму

$$\begin{aligned} y_i^{(1)} &= \pi_{11} x_i^{(1)} + \pi_{12} x_i^{(2)} + \varepsilon_i^{(1)}, \\ y_i^{(2)} &= \pi_{21} x_i^{(1)} + \pi_{22} x_i^{(2)} + \varepsilon_i^{(2)}, \end{aligned}$$

где

$$\begin{aligned} \pi_{11} &= \frac{b_1 c_1}{b_1 - b_2}, & \pi_{12} &= \frac{b_1 c_2}{b_1 - b_2}, \\ \pi_{21} &= \frac{c_1}{b_1 - b_2}, & \pi_{22} &= \frac{c_2}{b_1 - b_2}, \end{aligned}$$

а $\varepsilon_i^{(1)}$ и $\varepsilon_i^{(2)}$ такие же, как в (4.3''). Мы видим, что структурный коэффициент b_1 может быть определен по параметрам приведенной формы двумя различными способами, которые дают, вообще говоря, два разных результата.

$$1\text{-й вариант: } b_1 = \frac{\pi_{11}}{\pi_{21}};$$

$$2\text{-й вариант: } b_1 = \frac{\pi_{12}}{\pi_{22}}.$$

Для полноты картины проверим выполнение 5-го (*необходимого и достаточного*) условия идентифицируемости 1-го уравнения так называемого *рангового условия*, хотя мы уже знаем, что 1-е уравнение сверхидентифицируемо.

В нашем случае матрица $\bar{\Pi}_X(1)$, имеющая размерность $m_1 \times (p - p_1) = 2 \times 2$, совпадает со всей матрицей Π , т. е.

$$\bar{\Pi}_X = \begin{pmatrix} \frac{b_1 c_1}{b_1 - b_2} & \frac{b_1 c_2}{b_1 - b_2} \\ \frac{c_1}{b_1 - b_2} & \frac{c_2}{b_1 - b_2} \end{pmatrix}.$$

«Невооруженным глазом» видно, что строки этой матрицы связаны пропорциональной зависимостью (2-я строка получается из 1-й делением на b_1), так что ранг матрицы $\bar{\Pi}_X(1)$ равен единице, что и требуется в ранговом условии (т. к. в нашем случае $m_1 = 2$ и, следовательно, ранг $\bar{\Pi}_X(1) = m_1 - 1 = 1$).

4.3. Идентификация систем одновременных уравнений (статистическое оценивание неизвестных значений параметров системы)

В данном пункте речь идет только о той части проблемы идентификации СОУ, которая связана с методами статистического оценивания неизвестных значений параметров β_{ij} и c_{ik} ($i = 1, 2, \dots, m$; $j = 1, 2, \dots, m$; $k = 1, 2, \dots, p$) в уравнениях (4.19) по имеющимся в нашем распоряжении наблюдениям

$$Y = \begin{pmatrix} y_1^{(1)} & y_1^{(2)} & \dots & y_1^{(m)} \\ y_2^{(1)} & y_2^{(2)} & \dots & y_2^{(m)} \\ \dots & \dots & \dots & \dots \\ y_n^{(1)} & y_n^{(2)} & \dots & y_n^{(m)} \end{pmatrix}, \quad X = \begin{pmatrix} x_1^{(1)} & x_1^{(2)} & \dots & x_1^{(p)} \\ x_2^{(1)} & x_2^{(2)} & \dots & x_2^{(p)} \\ \dots & \dots & \dots & \dots \\ x_n^{(1)} & x_n^{(2)} & \dots & x_n^{(p)} \end{pmatrix}. \quad (4.26)$$

Мы будем предполагать, что число наблюдений n «много больше», чем $m + p$ ($n \gg m + p$), матрица B в (4.19) невырождена и что предопределенные переменные $x_i^{(1)}, \dots, x_i^{(p)}$ немультиколлинеарны, т.е. матрица X имеет полный ранг p . Среди столбцов матрицы X может присутствовать столбец, состоящий из одних единиц (если в правую часть модели (4.19) включен свободный член). Другие структурные предположения о модели (4.19), и в частности, допущения о природе случайных остатков Δ_t , будут формулироваться в ходе изложения того или иного метода статистического оценивания.

Мы уже убедились в том, что независимо от того, хотим ли мы оценить *только одно* из уравнений системы (4.19) или же намерены оценить *все* уравнения этой системы, в общем случае мы оказываемся в ситуации, когда известные нам по гл. 2 методы регрессионного анализа (МНК, различные версии ОМНК) не обеспечивают удовлетворительную процедуру оценивания. Причина подобной ситуации заключается в том, что при отсутствии специальных предположений о структуре СОУ (например, о ее *рекурсивном характере*, см. ниже) в анализируемом уравнении среди объясняющих переменных присутствует одна или несколько эндогенных переменных, *которые коррелируют с регрессионными остатками*. Правда, существует достаточно простой — так называемый *косвенный метод наименьших квадратов*, когда с помощью обычного МНК, примененного к каждому отдельному уравнению приведенной формы (4.19^а), сначала оцениваются элементы π_{ij} матрицы Π , а затем, используя соотношение (4.20) (в которое вместо матрицы Π подставляется ее оценка $\hat{\Pi}$), определяются значения оценок $\hat{\beta}_{ij}$ и \hat{c}_{ik} параметров структурной формы. Но

этот метод применим лишь к точно идентифицируемым уравнениям.

Переходя к описанию специальных методов оценивания структурных параметров модели (4.19), разделим их на два класса: методы, позволяющие оценивать каждое отдельное уравнение системы (4.19) поочередно, и методы, предназначенные для оценивания всех уравнений сразу. Правда, существует специальный тип СОУ — так называемые *рекурсивные системы*, — для которых при определенном выборе порядка и взаимосвязей оцениваемых отдельных уравнений системы процедура МНК приводит к оцениванию всех ее уравнений. С точки зрения задач статистического оценивания этот тип СОУ является простейшим, поэтому мы с него и начнем.

4.3.1. Идентификация рекурсивных систем

Во многом потому, что *рекурсивные СОУ* относительно просты для решения задачи статистического оценивания структурных параметров, в большом числе прикладных работ (может быть, в большей их части!) при спецификации модели стараются построить ее так, чтобы она удовлетворяла свойству рекурсивности. Для этого действуют следующим образом¹. В качестве 1-го уравнения системы специфицируют соотношение, в котором присутствует *только одна эндогенная переменная* $y^{(1)}$ (соответственно, и индексируют ее первым номером). Так что 1-е уравнение системы содержит $m_1 = 1$ эндогенных переменных и какое-то количество p_1 предопределенных переменных. Второе уравнение системы может содержать *не более двух* эндогенных переменных; это, если необходимо, эндогенная переменная $y^{(1)}$ («участница» 1-го уравнения) и, кроме нее, *еще только одну* эндогенную переменную (обозначим ее $y^{(2)}$). При комплектации состава и числа p_2 предопределенных переменных руководствуются содержательными соображениями и «правилом порядка» (см. правило идентифицируемости № 4 в предыдущем пункте). В третье уравнение, кроме уже участвовавших во 2-м уравнении эндогенных переменных $y^{(1)}$ и $y^{(2)}$,

¹ Следует, правда, признать, что в ряде публикаций ведущих эконометристов (в том числе, например, в работах Х. Волда) делаются серьезные попытки дать *содержательное (исходящее из сущности реальных экономических систем)* обоснование правомерности использования именно рекурсивных СОУ. Главные доводы их аргументации основаны на тезисе, что большинство реальных механизмов формирования рассматриваемых в модели экономических показателей функционирует в *рекурсивном (а не одновременном) режиме*. Так, например, трудно представить себе рыночные механизмы, *одновременно* формирующие цены и количества предлагаемых товаров (см. ниже пример 4.3).

можно включить еще *опять только одну эндогенную переменную* $y^{(3)}$ и т. д. В результате мы получим модель вида (4.19), в которой матрица \mathbf{B} является *нижней треугольной матрицей*, т. е. $\beta_{ij} = 0$ при $j > i$ для всех $i = 1, 2, \dots, m$ (при сохранении условия нормировки $\beta_{ii} = 1$). Оказывается, если для систем такого вида дополнительно потребовать взаимную некоррелированность случайных остатков $\delta_t^{(i_1)}$ и $\delta_t^{(i_2)}$ при всех t и для всех $i_1 \neq i_2$ ($i_1, i_2 = 1, 2, \dots, m$), то *оценки структурных параметров в каждом отдельном уравнении системы с помощью прямого метода наименьших квадратов будут состоятельными, а при нормальности $\delta_t^{(i)}$ — и асимптотически эффективными*. Под прямым МНК понимается следующая процедура, последовательно примененная к i -му уравнению системы ($i = 1, 2, \dots, m$): с помощью обычного МНК строятся оценки коэффициентов регрессии $y_t^{(i)}$ по всем включенным в это уравнение эндогенным и предопределенным переменным.

Достаточно естественно выглядит приведенная выше схема спецификации модели в виде СОУ при описании процесса формирования равновесных цен и количеств предлагаемых на рынке товаров. Рассмотрим пример.

Пример 4.3. Пусть $y_t^{(1)}$ — цена некоторого товара в момент времени t , а $y_t^{(2)}$ — объем продаж этого товара в тот же момент времени. Естественно предположить, что объем продаж $y_t^{(2)}$ зависит от цены $y_t^{(1)}$ и от объема продаж в предыдущий момент времени $y_{t-1}^{(2)}$. В свою очередь, цена товара $y_t^{(1)}$ зависит от объема его продаж в предыдущий момент времени (т. е. от $y_{t-1}^{(2)}$). В данной схеме цена $y_t^{(1)}$ и объем продаж $y_t^{(2)}$ играют роль эндогенных переменных, а *лаговой* переменной $y_{t-1}^{(2)}$ естественно отвести роль единственной предопределенной переменной x_t (т. е. $x_t = y_{t-1}^{(2)}$).

В результате анализируемая модель будет описана системой

$$\begin{cases} y_t^{(1)} + c_{11}x_t = \delta_t^{(1)}, \\ \beta_{21}y_t^{(1)} + y_t^{(2)} + c_{21}x_t = \delta_t^{(2)}, \end{cases} \quad t = 1, 2, \dots, n. \quad (4.27)$$

Первое уравнение системы без дополнительных ограничений является идентифицируемым, т. к. приведенная форма модели (4.27) имеет вид

$$\begin{cases} y_t^{(1)} = -c_{11}x_t + \delta_t^{(1)} = \pi_{11}x_t + \delta_t^{(1)}, \\ y_t^{(2)} = (\beta_{21}c_{11} - c_{21})x_t + (\delta_t^{(2)} - \beta_{21}\delta_t^{(1)}) = \pi_{21}x_t + \varepsilon_t^{(2)}, \end{cases} \quad (4.27^a)$$

где $c_{11} = -\pi_{11}$, а $\beta_{21}c_{11} - c_{21} = \pi_{21}$.

Однако если дополнительно потребовать некоррелированность остатков $\delta_t^{(1)}$ и $\delta_t^{(2)}$, т. е. диагональности ковариационной матрицы остатков

$$\Sigma_{\Delta} = \begin{pmatrix} \text{cov}(\delta_t^{(1)}, \delta_t^{(1)}) & \text{cov}(\delta_t^{(1)}, \delta_t^{(2)}) \\ \text{cov}(\delta_t^{(2)}, \delta_t^{(1)}) & \text{cov}(\delta_t^{(2)}, \delta_t^{(2)}) \end{pmatrix},$$

то коэффициенты β_{21} и c_{21} 2-го уравнения (4.27) могут быть состоятельно оценены с помощью обычного МНК, примененного к уравнению регрессии $y_t^{(2)}$ по $y_t^{(1)}$ и x_t . Действительно, перепишем 2-е уравнение (4.27) в виде

$$y_t^{(2)} = -\beta_{21}y_t^{(1)} - c_{21}x_t + \delta_t^{(2)}, \quad t = 1, 2, \dots, n.$$

Поскольку объясняющая переменная $y_t^{(1)} = -c_{11}x_t + \delta_t^{(1)}$ в правой части этого уравнения не коррелирована с $\delta_t^{(2)}$ (т. к. ни x_t , ни $\delta_t^{(1)}$ не коррелированы по условию с $\delta_t^{(2)}$), то в соответствии с общей теорией КЛММР (см. пп. 2.2–2.3) МНК-оценки параметров β_{21} и c_{21} будут состоятельными.

З а м е ч а н и е. Очень важным моментом в правильной спецификации модели, рассмотренной в примере 4.3, является *выбор продолжительности такта времени*. Действительно, продавец устанавливает цены, а покупатель на них реагирует. При этом торговые запасы будут либо накапливаться, либо рассасываться. Продавец среагирует на эту динамику и т. д. Если выбрать в качестве такта времени *один день*, то сделанные нами в модели примера 4.3 допущения выглядят естественными, поскольку последовательность причинных связей $y_{t-1}^{(2)} \rightarrow y_t^{(1)} \rightarrow y_t^{(2)}$ является *линейной* цепью (т. е. не содержит никаких петель обратной связи). Это позволяет нам, в частности, предположить возмущения, влияющие на спрос ($\delta_t^{(2)}$) и предложение ($\delta_t^{(1)}$), независимыми. Однако если бы публиковались, скажем, только недельные данные о *средней* недельной цене ($\bar{y}_t^{(1)}$) и *среднем* объеме дневных продаж ($\bar{y}_t^{(2)}$), то вынужденное агрегирование соответствующих возмущений $\bar{\delta}_t^{(1)}$ и $\bar{\delta}_t^{(2)}$ в системе (4.27) делает эти возмущения взаимно коррелированными, а саму модель — неидентифицированной. Введение же в модель дополнительных переменных с целью достижения идентифицируемости модели (если бы такие переменные существовали), как правило, превращает рекурсивную модель в обычную СОУ со всеми вытекающими отсюда проблемами ее оценивания.

Перейдем к описанию процедуры статистического оценивания структурных параметров *общей* рекурсивной модели СОУ. Пусть анализируется система из m стохастических уравнений с m эндогенными переменными $Y_t = (y_t^{(1)}, y_t^{(2)}, \dots, y_t^{(m)})^T$ и p экзогенными переменными $X_t =$

$$(x_t^{(1)}, x_t^{(2)}, \dots, x_t^{(p)})^T$$

$$BY_t + CX_t = \Delta_t, \quad t = 1, 2, \dots, n, \quad (4.28)$$

где B — $(m \times m)$ -матрица коэффициентов β_{ij} при эндогенных переменных ($i, j = 1, 2, \dots, m$; $\beta_{ii} = 1$ — условие нормировки), C — $(m \times p)$ -матрица коэффициентов c_{ij} при predetermined переменных ($i = 1, 2, \dots, m$; $j = 1, 2, \dots, p$), а вектор $\Delta_t = (\delta_t^{(1)}, \delta_t^{(2)}, \dots, \delta_t^{(m)})^T$ — вектор-столбец возмущений (или случайных остатков), не коррелированных с predetermined переменными.

О п р е д е л е н и е 4.1. Система одновременных уравнений (4.28) называется *чисто рекурсивной*, если:

(i) матрица B является нижней треугольной, т. е.

$$\beta_{ij} = 0 \quad \text{при } j > i \quad \text{для всех } i = 1, 2, \dots, m-1;$$

(ii) ковариационная матрица Σ_{Δ} случайных остатков Δ_t диагональна и не зависит от t , т. е.

$$\Sigma_{\Delta} = E(\Delta_t \Delta_t^T) = \begin{pmatrix} \sigma_{11} & 0 & 0 & \dots & 0 & 0 \\ 0 & \sigma_{22} & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & \sigma_{mm} \end{pmatrix}.$$

Иногда при спецификации модели удобнее ориентироваться на эквивалентное (ii) условие:

(ii^a) случайные возмущения $\delta_t^{(1)}, \delta_t^{(2)}, \dots, \delta_t^{(m)}$ обладают тем свойством, что $\delta^{(k)}$ не коррелирует со всеми «предшествующими» эндогенными переменными, т. е. с $y^{(1)}, y^{(2)}, \dots, y^{(k-1)}$ ($k = 2, 3, \dots, m$).

Если выполняется только условие (i), то система называется *формально рекурсивной*, или просто *рекурсивной*.

Выпишем теперь явный вид МНК-оценок структурных коэффициентов i -го уравнения системы (4.28). Пусть $j_1 < j_2 < \dots < j_{m_i-1}$ и $k_1 < k_2 < \dots < k_{p_i}$ — номера, соответственно, эндогенных (кроме переменной $y^{(i)}$) и predetermined переменных, участвующих в i -м уравнении системы (в силу ее рекурсивности, очевидно, $m_i \leq i$). Тогда i -е уравнение можно записать в виде

$$\begin{aligned} y_t^{(i)} = & -\beta_{ij_1} y_t^{(j_1)} - \beta_{ij_2} y_t^{(j_2)} - \dots - \beta_{ij_{m_i-1}} y_t^{(j_{m_i-1})} \\ & - c_{ik_1} x_t^{(k_1)} - \dots - c_{ik_{p_i}} x_t^{(k_{p_i})} + \delta_t^{(i)}, \end{aligned} \quad (4.29)$$

$$t = 1, 2, \dots, n.$$

Обозначим $Y^{(l)}$ и $X^{(q)}$, соответственно, l -й столбец матрицы Y и q -й столбец матрицы X (см. (4.26)). Пусть $\Delta(i) = (\delta_1^{(i)}, \delta_2^{(i)}, \dots, \delta_n^{(i)})^T$, $\Theta(i) = -(\beta_{ij_1}, \dots, \beta_{ij_{m_i-1}}; c_{ik_1}, \dots, c_{ik_{p_i}})^T$ — вектор-столбец, составленный из коэффициентов правой части (4.29), а

$$Z(i) = (Y^{(j_1)} \dots Y^{(j_{m_i-1})} X^{(k_1)} \dots X^{(k_{p_i})}) -$$

матрица размерности $n \times (m_i - 1 + p_i)$, составленная из столбцов, отобранных из матриц Y и X . Тогда уравнение (4.29) может быть представлено в виде

$$Y^{(i)} = Z(i)\Theta(i) + \Delta(i), \quad (4.29')$$

а МНК-оценки коэффициентов $\Theta(i)$ в соответствии с общим результатом (2.22) (см. с. 56) можно вычислить по формуле

$$\hat{\Theta}(i) = (Z^T(i)Z(i))^{-1}Z^T(i)Y^{(i)}. \quad (4.30)$$

При этом, конечно, необходимо, чтобы матрица $Z(i)$ имела бы полный ранг (т. е. $\text{rang } Z(i) = m_i - 1 + p_i$), или, что то же, матрица $Z^T(i)Z(i)$ должна быть невырожденной.

4.3.2. Косвенный метод наименьших квадратов

Сущность и реализация косвенного метода наименьших квадратов (КМНК) были продемонстрированы на примере модели спроса-предложения, рассмотренной в п. 4.1. Мы лишь кратко повторим здесь процедуру КМНК применительно к *общему случаю* СОУ.

Косвенный МНК предназначен для оценивания структурных параметров *отдельного* уравнения системы и может дать результат (без сочетания с другими методами, например, с двухшаговым методом наименьших квадратов) только в применении к *точно идентифицируемому* уравнению.

Итак, рассмотрим систему (4.19) и ее приведенную форму

$$Y_t = \Pi X_t + \varepsilon_t, \quad t = 1, 2, \dots, n, \quad (4.31)$$

в которой

$$\Pi = -B^{-1}C \quad (4.32)$$

матрица коэффициентов π_{ij} предопределенных переменных ($i = 1, 2, \dots, m; j = 1, 2, \dots, p$), *переменные* X_t -*немультиколлинеарны*, а

$$\varepsilon_t = B^{-1}\Delta_t, \quad t = 1, 2, \dots, n, - \quad (4.33)$$

вектор-столбец случайных остатков, не коррелированных с предопределенными переменными $X_t = (x_t^{(1)}, x_t^{(2)}, \dots, x_t^{(p)})^T$ по условию (т. к. все компоненты вектора Δ_t не коррелированы с X_t).

Наша цель — оценить структурные параметры i -го уравнения системы (4.19). Для анализа i -го уравнения воспользуемся общей схемой (4.21)–(4.25), которая была рассмотрена в п. 4.2 при выводе 4-го и 5-го условий идентифицируемости отдельного уравнения системы.

Процедуру статистического оценивания структурных параметров i -го уравнения системы (4.19) разобьем на 2 этапа.

На 1-м этапе оцениваем с помощью обычного МНК все параметры π_{ij} ($i = 1, 2, \dots, m$; $j = 1, 2, \dots, p$) приведенной формы (4.31), последовательно (и автономно) решая эту задачу для каждого отдельного уравнения системы (4.31). Хорошие свойства получаемых МНК-оценок $\hat{\pi}_{ij}$ (их состоятельность, а в случае нормальности $\delta_t^{(k)}$ — и их эффективность) обеспечиваются тем, что в модели (4.31)–(4.33) выполняются все условия классической линейной модели множественной регрессии.

На 2-м этапе используются соотношения (4.24)–(4.25), связывающие структурные параметры i -го уравнения системы (4.28) с параметрами π_{ij} приведенной формы. В случае точной (или сверх-) идентифицируемости i -го уравнения структурной формы (т. е. при выполнении условий 4 и 5 или только условия 4 со знаком *строгого* неравенства) его параметры $\beta(i)$ и $c(i)$ однозначно (или неоднозначно) определяются из системы (4.24)–(4.25) по значениям π_{ij} . Подставив в эти соотношения вместо π_{ij} их оценки $\hat{\pi}_{ij}$ и решив систему уравнений (4.24)–(4.25) относительно $\beta(i)$ и $c(i)$, мы получим состоятельные оценки $\hat{\beta}(i)$ и $\hat{c}(i)$ структурных параметров i -го уравнения системы (4.19).

В случае неидентифицируемости анализируемого уравнения структурной формы (т. е. при невыполнении 4-го условия идентифицируемости) число взаимно независимых связей между $\beta(i)$, $c(i)$ и π_{ij} , содержащихся в системе (4.24)–(4.25), будет меньше общего числа $m_i + p_i - 1$ неизвестных. Поэтому без дополнительной информации мы не сможем из (4.24)–(4.25) определить значения структурных коэффициентов $\beta(i)$ и $c(i)$. Один из самых распространенных способов получения необходимой дополнительной информации в данном случае, который в результате позволяет получать оценки структурных параметров неидентифицируемого уравнения, — это *двухшаговый метод наименьших квадратов*. К его рассмотрению мы и переходим.

4.3.3. Двухшаговый метод наименьших квадратов оценивания структурных параметров отдельного уравнения

Снова вернемся к рассмотренному выше примеру (4.1)–(4.2), в котором анализировалась модель спроса и предложения (см. п. 4.1). Напомним реализацию «шагов» в двухшаговом методе наименьших квадратов (2 МНК), использованном в этом примере. Чтобы обойти главное препятствие в применении обычного МНК к *отдельному неидентифицируемому уравнению* (в данном примере это было уравнение (4.2), описывающее зависимость спроса $y_i^{(1)}$ от цены $y_i^{(2)}$ и дохода x_i : $y_i^{(1)} = b_2 y_i^{(2)} + c x_i + \delta_i^{(2)}$) — коррелированность играющей роль предиктора эндогенной переменной $y_i^{(2)}$ со случайными остатками $\delta_i^{(2)}$, — на 1-м шаге строится регрессия эндогенной переменной-предиктора $y_i^{(2)}$ по предопределенной переменной x_i . В результате получаем оцененное регрессионное уравнение $y_i^{(2)} = \hat{\theta}_{y^{(2)}/x} x_i + \varepsilon_i$. На 2-м шаге в правую часть анализируемого уравнения (4.2) вместо эндогенной переменной $y_i^{(2)}$ вставлялось ее регрессионное выражение через x_i , в результате чего получали уравнение

$$y_i^{(1)} = (b_2 \hat{\theta}_{y^{(2)}/x} + c) x_i + (\delta_i^{(2)} + b_2 \varepsilon_i). \quad (\text{см. (4.2)'})$$

После этого к данному уравнению применялся обычный МНК с целью состоятельного оценивания коэффициента $\theta_{y^{(1)}/x} = b_2 \hat{\theta}_{y^{(2)}/x} + c$ (очевидно, вести речь об *отдельном* оценивании коэффициентов b_2 и c в уравнении $y_i^{(1)} = b_2 x_i^{(2)} + c x_i + (\delta_i^{(2)} + b_2 \varepsilon_i)$, в котором $x_i^{(2)} = \hat{\theta}_{y^{(2)}/x} x_i$, не имеет смысла, т. к. переменные \tilde{x}_i и x_i связаны чистой мультиколлинеарностью по построению). Получив оценку $\hat{\theta}_{y^{(1)}/x}$ параметра $\theta_{y^{(1)}/x} = b_2 \hat{\theta}_{y^{(2)}/x} + c$, мы к имеющимся уже двум соотношениям, связывающим параметры структурной и приведенной форм (вытекающим из общей формулы (4.32)), добавили еще одно: $\hat{\theta}_{y^{(1)}/x} = b_2 \hat{\theta}_{y^{(2)}/x} + c$. После этого, решая систему (4.18), мы смогли получить оценки параметров b_2 и c неидентифицируемого уравнения.

Точно такая же процедура 2 МНК предлагается и в общем случае. Правда, в отличие от упрощенной модели спроса-предложения (4.1)–(4.2), содержащей *единственную* предопределенную переменную (что и явилось причиной мультиколлинеарности переменных $x_i^{(2)}$ и x_i), в *реалистичных* эконометрических моделях мы будем крайне редко сталкиваться с проблемой чистой мультиколлинеарности объясняющих переменных на 2-м шаге процедуры, т. к. состав предопределенных переменных в каждом отдельном уравнении почти никогда не исчерпывает всего их априорного набора.

Итак, пусть, как и прежде, i -е уравнение системы представлено в форме (4.29), и пусть

$$Y(i) = (Y^{(j_1)} Y^{(j_2)} \dots Y^{(j_{m_i-1})}) - \quad (4.34)$$

матрица размерности $n \times (m_i - 1)$, составленная из соответствующих j_{m_i-1} столбцов матрицы Y (см. (4.26)),

$$X(i) = (X^{(k_1)} X^{(k_2)} \dots X^{(k_{p_i})}) - \quad (4.35)$$

матрица размерности $n \times p_i$, составленная из соответствующих столбцов матрицы X (см. (4.26)), а $X_{\text{ост}}(i)$ — матрица размерности $n \times (p - p_i)$, составленная из тех столбцов матрицы X , которые остались после отбора из нее столбцов для матрицы $X(i)$. Тогда анализируемое (i -е) уравнение системы (4.29) можно представить в виде

$$Y(i) = Y(i) b_-(i) + X(i) c_-(i) + \Delta(i), \quad (4.36)$$

где $b_-(i) = (-\beta_{ij_1}, -\beta_{ij_2}, \dots, -\beta_{ij_{m_i-1}})^T$, $c_-(i) = (-c_{ik_1}, -c_{ik_2}, \dots, -c_{ik_{p_i}})^T$ и, как и прежде, $\Delta(i) = (\delta_1^{(i)}, \delta_2^{(i)}, \dots, \delta_n^{(i)})^T$.

1-й шаг 2 МНК. С помощью обычного МНК строится регрессия выступающих в роли предикторов в уравнении (4.36) *эндогенных* переменных $Y(i)$ по *всем* предопределенным переменным X . В результате получаем регрессионную аппроксимацию $\hat{Y}(i)$ эндогенных переменных $Y(i)$ в виде линейной комбинации всех предопределенных переменных в форме

$$\hat{Y}(i) = X \hat{\Theta}_{Y_i/X}, \quad (4.37)$$

где $\hat{\Theta}_{Y_i/X} = (X^T X)^{-1} X^T Y(i)$, а $X = (X(i) : X_{\text{ост}}(i))$ — матрица размерности $n \times p$ наблюдаемых значений *всех* предопределенных переменных, составленная из матриц $X(i)$ и $X_{\text{ост}}(i)$.

2-й шаг 2 МНК. С помощью обычного МНК строится регрессия эндогенной переменной $Y^{(i)}$ по $\hat{Y}(i)$ и $X(i)$. Другими словами, в уравнении (4.36) объясняющие переменные $Y(i)$ заменяются на их аппроксимации $\hat{Y}(i)$, выраженные формулой (4.37). Реализация МНК в данном случае, проведенная в соответствии с общей теорией классической линейной модели множественной регрессии (см. п. 2.3, и в частности соотношения (2.21) на с. 56), приводит к следующей системе нормальных уравнений относительно оценок $\hat{b}_-(i)$ и $\hat{c}_-(i)$ неизвестных параметров, соответственно,

$\hat{b}_-(i)$ и $\hat{c}_-(i)$:

$$\begin{pmatrix} \hat{Y}^T(i) \cdot \hat{Y}(i) & \vdots & \hat{Y}^T(i) \cdot X(i) \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots \\ X^T(i) \cdot \hat{Y}(i) & \vdots & X^T(i) \cdot X(i) \end{pmatrix} \begin{pmatrix} \hat{b}_-(i) \\ \hat{c}_-(i) \end{pmatrix} = \begin{pmatrix} \hat{Y}^T(i) \cdot Y^{(i)} \\ X^T(i) \cdot Y^{(i)} \end{pmatrix}. \quad (4.38)$$

Заметим, что квадратная матрица A , являющаяся первым сомножителем в левой части соотношения (4.38) (она составлена из матриц $\hat{Y}^T(i)\hat{Y}(i)$, $\hat{Y}^T(i)X(i)$, $X^T(i)\hat{Y}(i)$ и $X^T(i)X(i)$), имеет размерность $(m_i - 1 + p_i) \times (m_i - 1 + p_i)$ и может оказаться вырожденной в силу мультиколлинеарности, возникающей (по построению) между объясняющими переменными $\hat{Y}(i)$ и $X(i)$ в случае $X(i) \equiv X$ (как это было в нашем примере). В этих случаях явные выражения оценок $\hat{b}_-(i)$ и $\hat{c}_-(i)$ из (4.38) получить невозможно. Однако так же, как и в рассмотренном выше примере модели спроса-предложения, объединение соотношений (4.38) и системы (4.24)–(4.25) даст нам систему линейных уравнений относительно $\hat{b}_-(i)$ и $\hat{c}_-(i)$ (т.е. относительно неизвестных параметров i -го уравнения структурной формы), которая позволит вычислить значения всех этих параметров. В остальных же случаях матрица A обратима и, следовательно, мы имеем возможность получить явные выражения оценок, а именно:

$$\begin{pmatrix} \hat{b}_-(i) \\ \hat{c}_-(i) \end{pmatrix} = A^{-1} \begin{pmatrix} \hat{Y}^T(i)Y^{(i)} \\ X^T(i)Y^{(i)} \end{pmatrix}, \quad (4.39)$$

где $(m_i - 1 + p_i) \times (m_i - 1 + p_i)$ — матрица A имеет вид

$$A = \begin{pmatrix} \hat{Y}^T(i)\hat{Y}(i) & \vdots & \hat{Y}^T(i)X(i) \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots \\ X^T(i)\hat{Y}(i) & \vdots & X^T(i)X(i) \end{pmatrix}, \quad (4.40)$$

или, что то же,

$$A = \begin{pmatrix} Y^T(i)X(X^T X)^{-1}X^T Y(i) & \vdots & \hat{Y}^T(i)X(i) \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots \\ X^T(i)\hat{Y}(i) & \vdots & X^T(i)X(i) \end{pmatrix}. \quad (4.40')$$

Заметим, что первые $m_i - 1$ компонент вектора-столбца, являющегося вторым сомножителем в правой части (4.39), могут быть также представлены в терминах исходных наблюдений, а именно:

$$\hat{Y}^T(i)Y^{(i)} = Y^T(i)X(X^T X)^{-1}X^T Y^{(i)}. \quad (4.41)$$

В задачах анализа точности 2 МНК-оценок отдельного уравнения СОУ важную роль играет ковариационная матрица Σ этих оценок. Можно показать (см., например, [Джонстон Дж.], с. 383), что при большом числе наблюдений (т. е. асимптотически по $n \rightarrow \infty$) матрица Σ может быть оценена (в случае невырожденности матрицы A) с помощью матрицы

$$\hat{\Sigma} = s^2 A^{-1}, \quad (4.42)$$

где матрица A определяется соотношениями (4.40) или (4.40'), а

$$s^2 = \frac{1}{n - m_i - 1 - p_i} (Y^{(i)} - Y(i)\hat{b}_-(i) - X(i)\hat{c}_-(i))^T \times (Y^{(i)} - Y(i)\hat{b}_-(i) - X(i)\hat{c}_-(i)). \quad (4.43)$$

Использование метода главных компонент в 2 МНК. В двухшаговом методе наименьших квадратов, как мы видим, на 1-м шаге вычисляется регрессия $Y(i)$ по *всем* (!) p предопределенным переменным, т. е. в *каждом* из $m_i - 1$ регрессионных уравнений (4.37) *приходится оценивать p параметров*. В моделях небольшой размерности, т. е. при относительно малых значениях p , это не создает трудностей, но в средних и больших моделях общее число предопределенных переменных (p) может быть одного порядка с общим числом наблюдений (n), и тогда статистическая надежность выводов становится неудовлетворительной (бывают случаи, когда $p > n$; однако при этом состоятельное оценивание параметров модели становится в принципе невозможным). В подобных ситуациях возникает необходимость снизить общее число предопределенных переменных, через которые выражается $Y(i)$ (т. е. число p участвующих в правой части выражения (4.37) предопределенных переменных X). Клоек и Меннес предложили с этой целью использовать в качестве аргументов аппроксимирующей функции $\hat{Y}(i)$ *не все предопределенные переменные $X(i)$ и $X_{\text{ост}}(i)$* (что предусмотрено в (4.37)), а лишь те, которые участвуют в i -м уравнении (т. е. $X(i)$), плюс небольшое количество главных компонент, построенных по прочим предопределенным переменным $X_{\text{ост}}^1$. Что касается переменных $X(i)$, то, с одной стороны, их не следует трогать, т. к. они включались в i -е уравнение и по содержательному смыслу, и с учетом выполнения условия 4 идентифицируемости (см. выше), а с другой стороны, ими нельзя ограничиваться при построении аппроксимации $\hat{Y}(i)$ для

¹ См.: Kloek T., Mennes L.B.M. Simultaneous Equation Estimation Based on Principal Components Variables. «Econometrica», vol. 28 (1960), pp. 45–61.

$Y(i)$, т. к. в этом случае мы столкнулись бы со строгой мультиколлинеарностью.

Поэтому предлагается следующая процедура.

1) Матрица $X_{\text{ост}}(i)$ размерности $n \times (p - p_i)$, состоящая из наблюдений предопределенных переменных, не вошедших в анализируемое (i -е) уравнение, *стандартизируется*, т. е. из каждого наблюдения переменной вычитается ее выборочное среднее значение и результат делится на ее выборочное среднеквадратическое отклонение. В результате получаем матрицу $\tilde{X}_{\text{ост}}(i)$.

2) По наблюдениям стандартизированной матрицы $\tilde{X}_{\text{ост}}(i)$ строятся главные компоненты $U^{(l)} = (u_1^{(l)}, u_2^{(l)}, \dots, u_n^{(l)})^T$, $l = 1, 2, \dots, p - p_i$ (см. п. 13.2).

3) Из $p - p_i$ построенных главных компонент $U^{(l)}$ отбирается сравнительно небольшое число k ($k \leq 4$) наиболее информативных с точки зрения точности аппроксимации значений $Y^{(i)}$ в виде функции от $X(i)$ и $U^{(l_1)}, U^{(l_2)}, \dots, U^{(l_k)}$. В качестве критерия информативности можно использовать, например, величину коэффициента детерминации R^2 , определяющего степень тесноты связи между $Y^{(i)}$, с одной стороны, и набором $X(i), U^{(l_1)}, \dots, U^{(l_k)}$, — с другой.

Иначе говоря, главные компоненты $U^{(l_1)}, \dots, U^{(l_k)}$ отбираются из общего числа $p - p_i$ в соответствии с условием¹

$$U^{(l_1)}, U^{(l_2)}, \dots, U^{(l_k)} \rightarrow \max_{l_1, l_2, \dots, l_k} R_{Y^{(i)}, (X(i), U^{(l_1)}, \dots, U^{(l_k)})}^2.$$

4) В качестве 1-го шага 2 МНК строится регрессия $Y(i)$ по $X(i), U^{(l_1)}, U^{(l_2)}, \dots, U^{(l_k)}$ (а не по $X(i), X_{\text{ост}}(i)$, как это рекомендовано в обычном 2 МНК).

З а м е ч а н и е. Иногда строят главные компоненты по *всей* матрице X и тогда аппроксимация $Y(i)$ осуществляется только по k главным компонентам этой матрицы, причем k должно подбираться с учетом условия 4 идентифицируемости уравнения. Тем самым обходится неудобство

¹ В эконометрической литературе и практике предлагались и использовались и другие критерии отбора главных компонент: критерий минимизации мультиколлинеарности между $X(i)$ и $U^{(l_1)}, \dots, U^{(l_k)}$; критерий максимальной дисперсии, когда отбираются главные компоненты, соответствующие максимальным характеристическим числам (см., например, [Джонстон Дж.], с. 393–395). Однако предложенный нами критерий представляется более предпочтительным.

этого метода, возникающее при необходимости оценивания сразу нескольких (а тем более — всех) уравнений системы. В вышеизложенном варианте тогда придется строить главные компоненты для каждого уравнения в отдельности. Однако существуют работы, демонстрирующие преимущества (в точности) именно изложенного выше метода по сравнению с подходом, в котором главные компоненты строятся по всей матрице X для всех уравнений системы.

В заключение отметим, что эконометрическая теория и практика свидетельствуют о том, что 2 МНК являются, пожалуй, наиболее важным и широко применяемым методом оценивания структурных параметров *отдельного уравнения* СОУ.

4.3.4. Трехшаговый метод наименьших квадратов одновременной оценки всех параметров системы

Трехшаговый метод наименьших квадратов был предложен впервые Зельнером и Тейлом¹ в качестве метода статистического оценивания одновременно всех уравнений модели (4.28) с *учетом возможной взаимной коррелированности регрессионных остатков различных уравнений этой системы*. Этот метод оказывается более эффективным, чем 2 МНК, если случайные остатки $\delta_i^{(1)}, \delta_i^{(2)}, \dots, \delta_i^{(m)}$ различных уравнений системы (4.28) взаимно коррелированы, т. е. если их ковариационная матрица $\Sigma_{\Delta} = (E(\delta_i^{(i)} \delta_i^{(l)}))_{i,l=1,2,\dots,m}$ отлична от диагональной. Хотя отметим, что и в этой ситуации 2 МНК-оценки структурных параметров системы *остаются состоятельными*.

В трехшаговом методе наименьших квадратов (3 МНК) сохранены первые два шага 2 МНК. Однако полученные в результате этих двух шагов, автономно для каждого отдельного (i -го) уравнения, оценки структурных параметров $\hat{\beta}_{2\text{МНК}}(i) = -\hat{b}_-(i)$ и $\hat{c}_{2\text{МНК}}(i) = -\hat{c}_-(i)$, $i = 1, 2, \dots, m$ ($\hat{b}_-(i)$ и $\hat{c}_-(i)$ определены соотношениями (4.39)), не являются окончательными, а *пересчитываются на 3-м шаге* следующим образом: оценки $\hat{\beta}_{2\text{МНК}}$ и $\hat{c}_{2\text{МНК}}$ используются для подсчета выборочной ковариационной матрицы случайных остатков $\hat{\Sigma}_{\Delta}$, а последняя, в свою очередь, используется для одновременного вычисления оценок всех структурных параметров B и C системы (4.28) с помощью *обобщенного метода наименьших квадратов* в рамках соответствующим образом построенной обобщенной линейной модели множественной регрессии (см. п. 2.6).

¹ См.: Zellner A., Theil H. Three-stage Least-squares: Simultaneous Estimation of Simultaneous Equations. «Econometrica», vol. 30 (1962), pp. 54–78.

Рассмотрим подробнее, как реализуется описанная выше общая схема 3 МНК. Можно было бы начать сразу с 3-го шага, т. к. первые два, как было отмечено, совпадают с соответствующими шагами 2 МНК. Однако мы вернемся вначале к 2 МНК, чтобы описать прием, эквивалентный двум шагам этого метода, поскольку он приводит к тем же самым оценкам (4.39), что и 2 МНК.

Замечание о 2 МНК-оценках. Итак, снова рассмотрим отдельное (i -е) уравнение (4.36) анализируемой СОУ (4.28). Так же, как и при анализе рекурсивных систем, введем матрицу

$$\mathbf{Z}(i) = \begin{pmatrix} \mathbf{Y}(i) & \vdots & \mathbf{X}(i) \end{pmatrix}, \quad (4.44)$$

составленную из матриц $\mathbf{Y}(i)$ и $\mathbf{X}(i)$, и вектор-столбец неизвестных параметров

$$\Theta(i) = (-\beta_{ij_1}, \dots, -\beta_{ij_{m_1-1}}; -c_{ik_1}, \dots, -c_{ik_p})^T, \quad (4.45)$$

составленный из векторов $b_-(i)$ и $c_-(i)$. Тогда мы можем записать уравнение (4.36) в виде

$$Y^{(i)} = \mathbf{Z}(i) \Theta(i) + \Delta(i). \quad (4.36')$$

В 2 МНК мы обходили неприятности, связанные с взаимной коррелированностью объясняющих переменных $\mathbf{Y}(i)$ и остатков $\Delta(i)$, подстановкой в уравнение (4.36) вместо $\mathbf{Y}(i)$ их регрессионного выражения через \mathbf{X} (см. формулу (4.37)). Оказывается, можно добиться того же и прийти к 2 МНК оценкам (4.39), действуя формально несколько иным способом. А именно, домножим все члены уравнения (4.36') слева на матрицу \mathbf{X}^T :

$$\mathbf{X}^T Y^{(i)} = (\mathbf{X}^T \mathbf{Z}(i)) \Theta(i) + \mathbf{X}^T \Delta(i), \quad (4.36'')$$

или, что то же,

$$\tilde{Y}^{(i)} = \tilde{\mathbf{Z}}(i) \Theta(i) + \tilde{\Delta}(i), \quad (4.36''')$$

где

$$\tilde{Y}^{(i)} = \mathbf{X}^T Y^{(i)}, \quad (4.46)$$

$$\tilde{\mathbf{Z}}(i) = \mathbf{X}^T \mathbf{Z}(i), \quad (4.47)$$

$$\tilde{\Delta}(i) = \mathbf{X}^T \Delta(i). \quad (4.48)$$

Можно показать, что в уравнении (4.36''') объясняющие переменные $\tilde{\mathbf{Z}}(i)$ не коррелированы с остатками $\tilde{\Delta}(i)$ и что ковариационная матрица

$\Sigma_{\tilde{\Delta}(i)}$ остатков $\tilde{\Delta}(i)$ имеет вид

$$\Sigma_{\tilde{\Delta}(i)} = E(\tilde{\Delta}(i)\tilde{\Delta}^T(i)) = \sigma_{ii}(X^T X), \quad (4.49)$$

где $\sigma_{ii} = D\delta_i^{(i)}$.

Соотношение (4.36''') (и эквивалентное ему (4.36'')) можно рассматривать как *обобщенную линейную модель множественной регрессии* (2.73), в которой $\tilde{Y}^{(i)}$ играет роль вектора наблюдений зависимой переменной, $\tilde{Z}(i)$ — матрица наблюдений объясняющих переменных, $\Theta(i)$ — вектор-столбец оцениваемых параметров, а $\tilde{\Delta}(i)$ — вектор-столбец регрессионных остатков с заданной (с точностью до постоянного множителя σ_{ii}) ковариационной матрицей¹. Тогда в соответствии с формулой (2.78) ОМНК-оценок получаем:

$$\begin{aligned} \hat{\Theta}_{\text{омнк}}(i) &= [\tilde{Z}^T(i)(X^T X)^{-1}\tilde{Z}(i)]^{-1}\tilde{Z}^T(i)(X^T X)^{-1}\tilde{Y}^{(i)} \\ &= [Z^T(i)X(X^T X)^{-1}X^T Z(i)]^{-1}Z^T(i)X(X^T X)^{-1}X^T Y^{(i)}. \end{aligned} \quad (4.50)$$

Сравнение правых частей формул (4.50) и (4.39) с учетом соотношений (4.40'), (4.44) и (4.45) приводит к заключению об эквивалентности 2 МНК-оценок и оценок (4.50).

3-й шаг 3 МНК. От *отдельных* уравнений (4.36'') вернемся ко всей СОУ (4.28). С этой целью, составив уравнения (4.36'') для $i = 1, 2, \dots, t$, объединим их в *одной обобщенной линейной модели множественной регрессии* вида

$$\begin{pmatrix} X^T Y^{(1)} \\ X^T Y^{(2)} \\ \vdots \\ X^T Y^{(m)} \end{pmatrix} = \begin{pmatrix} X^T Z(1) & 0 & 0 & \dots & 0 \\ 0 & X^T Z(2) & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & X^T Z(m) \end{pmatrix} \begin{pmatrix} \Theta(1) \\ \Theta(2) \\ \vdots \\ \Theta(m) \end{pmatrix} + \begin{pmatrix} X^T \Delta(1) \\ X^T \Delta(2) \\ \vdots \\ X^T \Delta(m) \end{pmatrix}. \quad (4.51)$$

¹ Если в состав predetermined переменных X входят помимо *неслучайных экзогенных* переменных лаговые *эндогенные* переменные (которые по своей природе являются *стохастическими*), то, как известно (см. п. 2.10.1), обобщенный метод наименьших квадратов будет продолжать давать тем не менее состоятельные оценки параметров, если эти стохастические предикторы не коррелированы с регрессионными остатками (что и имеет место в нашем случае).

Матрица ковариаций $\Sigma_{\tilde{\Delta}}$ вектора случайных остатков $\tilde{\Delta} = (\tilde{\Delta}^T(1), \tilde{\Delta}^T(2), \dots, \tilde{\Delta}^T(m))^T = (\Delta^T(1)X : \Delta^T(2)X : \dots : \Delta^T(m)X)^T$ будет иметь вид

$$\Sigma_{\tilde{\Delta}} = \begin{pmatrix} \sigma_{11}X^T X & \sigma_{12}X^T X & \dots & \sigma_{1m}X^T X \\ \sigma_{21}X^T X & \sigma_{22}X^T X & \dots & \sigma_{2m}X^T X \\ \dots & \dots & \dots & \dots \\ \sigma_{m1}X^T X & \sigma_{m2}X^T X & \dots & \sigma_{mm}X^T X \end{pmatrix} = \Sigma_{\Delta} \otimes X^T X, \quad (4.52)$$

где через σ_{ij} обозначена ковариация случайных остатков i -го и j -го уравнений структурной формы (она по условию не зависит от t , т. е. $\sigma_{ij} = \text{cov}(\Delta_t^{(i)}, \Delta_t^{(j)}) = E(\delta_t^{(i)}\delta_t^{(j)})$), $\Sigma_{\Delta} = (\sigma_{ij})_{i,j=1,2,\dots,m}$, а символ \otimes означает *кронекерово перемножение матриц* (см. Приложение 2).

Чтобы воспользоваться обобщенным методом наименьших квадратов в одновременном оценивании параметров $\Theta(1), \Theta(2), \dots, \Theta(m)$ обобщенной линейной модели множественной регрессии (4.51), необходимо знать ковариационную матрицу остатков $\Sigma_{\tilde{\Delta}}$. Из (4.52) видно, что для этого *нам необходимо знать матрицу Σ_{Δ} ковариаций случайных возмущений анализируемой структурной формы* (4.28). Сущность 3-го шага трехшагового метода наименьших квадратов как раз и заключается в том, что для построения оценки $\hat{\Sigma}_{\tilde{\Delta}}$ неизвестной матрицы Σ_{Δ} используются 2 МНК-оценки $\hat{\Theta}_{2\text{мнк}}(i)$ параметров $\Theta(i)$, а именно: (i, j) -й элемент σ_{ij} этой матрицы оценивается по формуле

$$\hat{\sigma}_{ij} = \frac{1}{n} (Y^{(i)} - Z(i)\hat{\Theta}_{2\text{мнк}}(i))^T (Y^{(j)} - Z(j)\hat{\Theta}_{2\text{мнк}}(j)), \quad (4.53)$$

где $Z(l)$ и $\hat{\Theta}_{2\text{мнк}}(l)$ определены соотношениями, соответственно, (4.44) и (4.39) (или (4.50)), а $Y^{(l)} = (y_1^{(l)}, y_2^{(l)}, \dots, y_n^{(l)})^T$, $l = 1, 2, \dots, m$.

После этого мы можем воспользоваться формулой (2.78) обобщенного метода наименьших квадратов, примененного к модели (4.51):

$$\hat{\Theta}_{3\text{мнк}} = (\tilde{Z}^T \hat{\Sigma}_{\tilde{\Delta}}^{-1} \tilde{Z})^{-1} \tilde{Z}^T \hat{\Sigma}_{\tilde{\Delta}}^{-1} \tilde{Y}, \quad (4.54)$$

где

$$\hat{\Theta} = \begin{pmatrix} \hat{\Theta}(1) \\ \hat{\Theta}(2) \\ \vdots \\ \hat{\Theta}(m) \end{pmatrix}, \quad \tilde{Y} = \begin{pmatrix} X^T Y^{(1)} \\ X^T Y^{(2)} \\ \vdots \\ X^T Y^{(m)} \end{pmatrix},$$

$$\hat{\Sigma}_{\tilde{\Delta}} = \begin{pmatrix} \hat{\sigma}_{11}X^T X & \hat{\sigma}_{12}X^T X & \dots & \hat{\sigma}_{1m}X^T X \\ \hat{\sigma}_{21}X^T X & \hat{\sigma}_{22}X^T X & \dots & \hat{\sigma}_{2m}X^T X \\ \dots & \dots & \dots & \dots \\ \hat{\sigma}_{m1}X^T X & \hat{\sigma}_{m2}X^T X & \dots & \hat{\sigma}_{mm}X^T X \end{pmatrix},$$

$\hat{\sigma}_{ij}$ определены формулой (4.53), а

$$\tilde{\mathbf{Z}} = \begin{pmatrix} \mathbf{X}^T \mathbf{Z}(1) & 0 & 0 & \dots & 0 \\ 0 & \mathbf{X}^T \mathbf{Z}(2) & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \mathbf{X}^T \mathbf{Z}(m) \end{pmatrix}$$

(напомним, что $\mathbf{Z}(i)$ и $\Theta(i)$, $i = 1, 2, \dots, m$ определены соотношениями, соответственно, (4.44) и (4.45)).

Как уже было упомянуто в начале этого пункта, процедура 3 МНК обеспечивает лучшую по сравнению с двухшаговым методом эффективность оценок лишь в случае, когда матрица Σ_{Δ} не является диагональной, т. е. когда случайные остатки $\delta_i^{(i)}$, входящие в различные структурные уравнения, коррелируют друг с другом. Если матрица ковариаций Σ_{Δ} для структурных остатков может быть приведена (соответствующим упорядочением уравнений системы) к *блочно-диагональному виду*, то всю процедуру трехшагового оценивания лучше применять *отдельно к каждой группе уравнений, соответствующих одному блоку*. Тем самым существенно снижается трудоемкость вычислительной реализации 3 МНК.

4.3.5. Другие методы оценивания СОУ и некоторые общие рекомендации

Помимо методов статистического оценивания СОУ, описанных выше, в эконометрической литературе и практике существуют и другие процедуры. Так, если дополнительно постулировать *нормальность структурных возмущений* $\delta_i^{(i)}$, то для оценки отдельного уравнения наряду с 2 МНК может быть использован *метод максимального правдоподобия с ограниченной информацией*, или, что то же, *метод наименьшего дисперсионного отношения*¹. Для одновременной оценки всех параметров системы в рамках тех же условий нормальности $\delta_i^{(i)}$ разработан также *метод максимального правдоподобия с полной информацией*². В расширении класса методов оценивания СОУ можно идти также по линии введения *целых параметрических семейств методов*, когда при конкретном выборе значения свободного параметра этого семейства мы приходим к тому или

¹ Метод впервые предложен в работе: Anderson T. W., Rubin H. Estimation of the Parameters of a Singl Equation in a Complete System of Stochastic Equations. «Ann. Math. Statistics», 20 (1949), pp. 46–63.

² См. Hood W. C., Koopmans T. C. (eds). Studies in Econometric Method. Wiley, New York, 1953.

иному уже известному методу как к частному случаю. Именно такого рода семейство процедур представляют собой так называемые «оценки класса k », которые содержат в себе в качестве частных случаев процедуры оценивания отдельного уравнения структурной формы по прямому методу наименьших квадратов (при $k = 0$) и по двухшаговому методу наименьших квадратов (при $k = 1$)¹.

Мы оставили эти методы за рамками учебника, поскольку трудоемкость их вычислительной реализации, по меньшей мере, не уступает описанным выше косвенному, двухшаговому и трехшаговому методам наименьших квадратов. В то же время они требуют дополнительных модельных ограничивающих предположений. Поэтому, наверное, они реже других методов применяются в эконометрической практике.

В заключение приведем несколько общих рекомендаций по построению и статистическому анализу СОУ.

1) Прежде всего следует провести процедуру исключения из СОУ всех уравнений, являющихся тождествами.

2) В ходе этапа «спецификация модели» надо стремиться к обеспечению выполнения условий 1~5 идентифицируемости отдельных уравнений и системы в целом. Если среди конечных прикладных целей моделирования ставится задача *интервального* прогноза значений эндогенных переменных (см. ниже, п. 4.4), то следует стремиться к обеспечению нормальности и взаимной независимости (по t) случайных возмущений $\delta_t^{(i)}$.

3) Приступая к статистическому оцениванию СОУ, следует применить, в первую очередь, 2 МНК последовательно ко всем уравнениям системы и использовать полученные оценки структурных параметров для оценки ковариационной матрицы Σ_{Δ} структурных остатков (см. формулу (4.53)).

4) По полученной оценке $\hat{\Sigma}_{\Delta}$ ковариационной матрицы Σ_{Δ} проверить гипотезу о ее диагональной (или блочно-диагональной) структуре. Если гипотеза о диагональности Σ_{Δ} *не отвергается*, то использовать полученные ранее 2 МНК-оценки структурных параметров \hat{B} и \hat{C} для вычисления оценок $\hat{\Pi}$ параметров приведенной формы по формуле $\hat{\Pi} = -\hat{B}^{-1}\hat{C}$ и далее — для прогноза и имитационных расчетов.

5) Если гипотеза о диагональности матрицы Σ_{Δ} отвергается, то следует использовать полученную ранее оценку $\hat{\Sigma}_{\Delta}$ для реализации процедуры 3 МНК.

¹ См. Теял Г. Экономические прогнозы и принятие решений. (Пер. с англ.). М.: Статистика, с 281–282.

б) Если в процессе статистического анализа оценки $\hat{\Sigma}_\Delta$ выяснится, что блочно-диагональная структура матрицы Σ_Δ не противоречит имеющимся наблюдениям, то процедуру 3 МНК следует применить отдельно к каждой группе уравнений СОУ, соответствующих тому или иному блоку матрицы Σ_Δ .

4.4. Точечный и интервальный прогноз значений эндогенных переменных

Можно выделить два основных типа конечных прикладных целей, преследуемых при построении эконометрических моделей в виде СОУ. Один из них связан с получением сведений о структурных параметрах модели и (или) коэффициентах приведенной формы. Это бывает обычно в ситуациях, когда интересующие исследователя параметры модели имеют четкую экономическую интерпретацию (они могут играть роль различных эластичностей, мультипликаторов, характеристик «склонности к сбережениям» и т. п.). Другой тип конечных прикладных целей заключается в стремлении осуществить с помощью анализируемой СОУ *условный прогноз эндогенных переменных* (при определенных условиях, накладываемых на значения предопределенных переменных), произвести многовариантные (в соответствии с различными вариантами условий) *сценарные расчеты*, показывающие, как себя «будут вести» эндогенные переменные при различных условиях, касающихся значений предопределенных переменных.

Если интерес сосредоточен на *структурных* параметрах B и C , то следует воспользоваться состоятельными методами их оценивания (2 МНК, 3 МНК) и по той же исходной информации *оценить точность* примененных процедур (либо с помощью *аналитических* методов, т. е. вывода и статистически оценивания ковариационную матрицу используемых оценок параметров, см., например, (4.42)–(4.43), либо с помощью разного рода *компьютерно-имитационных методов*, о последних см. ниже, п. 4.5).

Если же нас могут удовлетворить коэффициенты Π *приведенной* формы (сами по себе или как средство получения *точечного* прогноза эндогенных переменных по формуле $\hat{Y}_{n+\tau} = \hat{\Pi}X_{n+\tau}$, где $X_{n+\tau}$ — заданные для момента времени $n + \tau$ значения предопределенных переменных, а τ — «глубина» прогноза), то анализ точности полученных оценок и (или) прогноза можно проводить одним из двух способов. В первом из них (более сложном) отправным пунктом служат состоятельные оценки \hat{B} и \hat{C} (структурных параметров B и C) и их ковариационные матрицы $\Sigma_{\hat{B}}$ и $\Sigma_{\hat{C}}$, а точнее, — состоятельные оценки $\hat{\Sigma}_{\hat{B}}$ и $\hat{\Sigma}_{\hat{C}}$ этих матриц. По ним

строят оценки $\hat{\Pi} = -\hat{B}^{-1}\hat{C}$ (см. (4.20)) и стараются вывести общий вид состоятельной оценки $\hat{\Sigma}_{\hat{\Pi}}$ ковариационной матрицы $\Sigma_{\hat{\Pi}}$. Решение последней задачи составляет *аналитическую* базу для решения задачи построения *интервального* прогноза для эндогенных переменных Y и *интервальных* оценок для параметров приведенной формы Π . Этот путь анализа точности точечного прогноза $\hat{Y}_{n+\tau}$ и точечных оценок $\hat{\Pi}$ считается более предпочтительным в ситуациях, когда есть основания полагать, что спецификация модели в ее структурной форме осуществлена правильно¹. Однако сопутствующие этому пункту громоздкость вывода вида ковариационной матрицы $\Sigma_{\hat{\Pi}}$ и связанные с этим аналитико-вычислительные сложности побудили нас не включать подробное описание данного подхода в учебник (интересующийся читатель может обратиться непосредственно к работам: 1) *Goldberger A. S., Nagar A. L., Odeh H. S.* The Covariance Matrices of Reduced-Form Coefficients and of Forecasts for a Structural Econometric Model. «Econometrica», 29 (1961), pp. 556–573; 2) *Humans S. H.* Simultaneous Confidence Intervals in Econometric Forecasting. «Econometrica», 36 (1968), pp. 18–30).

Мы же остановимся здесь подробнее на относительно более простом пути построения интервальных оценок для эндогенных переменных и для отдельных параметров приведенной формы. Этот путь основан на непосредственном вычислении оценок $\hat{\Pi}$ коэффициентов приведенной формы Π с помощью *обыкновенного метода наименьших квадратов*, примененного к каждому уравнению системы (4.19^a). Состоятельность и несмещенность этих оценок обеспечиваются стандартными модельными допущениями, относящимися к системе (4.19), а их ковариационная матрица выводится на базе фундаментальных соотношений классического МНК.

Итак, по исходным статистическим данным (4.26) анализируется приведенная форма СОУ вида

$$Y_t = \Pi X_t + \varepsilon_t, \quad t = 1, 2, \dots, n, \quad (4.55)$$

где $Y_t = (y_t^{(1)}, y_t^{(2)}, \dots, y_t^{(m)})^T$ и $X_t = (x_t^{(1)}, x_t^{(2)}, \dots, x_t^{(p)})^T$ — векторы наблюдаемых в момент времени t значений, соответственно, эндогенных и предопределенных переменных, $\Pi = (\pi_{ij})_{\substack{i=1,2,\dots,m; \\ j=1,2,\dots,p}}$ — матрица коэффициентов приведенной формы, а случайные остатки $\varepsilon_t = (\varepsilon_t^{(1)}, \varepsilon_t^{(2)}, \dots, \varepsilon_t^{(m)})^T$ имеют нулевые средние значения, постоянные (не зависящие от времени) дисперсии и не коррелированы с предопределенными переменными X_t .

¹ См. *Klein L. R.* The Efficiency of Estimation in Econometric Models. Cowles Foundation Paper, № 157, Yale University, 1960.

Заметим, что модель (4.55) состоит из m обобщенных линейных моделей множественной регрессии вида

$$Y^{(i)} = X \Pi(i) + \varepsilon(i), \quad i = 1, 2, \dots, m, \quad (4.55')$$

где $Y^{(i)} = (y_1^{(i)}, y_2^{(i)}, \dots, y_n^{(i)})^T$, $\Pi(i) = (\pi_{i1}, \pi_{i2}, \dots, \pi_{ip})^T$ — i -й столбец матрицы Π^T , $\varepsilon(i) = (\varepsilon_1^{(i)}, \varepsilon_2^{(i)}, \dots, \varepsilon_n^{(i)})^T$, а X — матрица наблюдений предопределенных переменных (см. (4.26)).

Даже при взаимной коррелированности остатков $\varepsilon_1^{(i)}, \varepsilon_2^{(i)}, \dots, \varepsilon_n^{(i)}$ (а они могут быть таковыми в нашем случае, если не потребовать некоррелированности структурных возмущений $\delta_{i_1}^{(i)}$ и $\delta_{i_2}^{(j)}$ разных уравнений для разных моментов времени) обычные оценки наименьших квадратов $\hat{\pi}_{\text{МНК}}(i)$ параметров $\pi(i)$ в уравнении (4.55') являются *состоятельными* (см. пп. 2.6, 2.9, 2.10.1). Применяя известную формулу (2.22) для вычисления МНК-оценок, получаем

$$\hat{\Pi}_{\text{МНК}}(i) = (X^T X)^{-1} X^T Y^{(i)}, \quad i = 1, 2, \dots, m. \quad (4.56)$$

Матрицу, составленную из столбцов $\hat{\Pi}_{\text{МНК}}(1), \hat{\Pi}_{\text{МНК}}(2), \dots, \hat{\Pi}_{\text{МНК}}(m)$, очевидно, можно записать в виде

$$\left(\hat{\Pi}_{\text{МНК}}(1) \quad \hat{\Pi}_{\text{МНК}}(2) \quad \dots \quad \hat{\Pi}_{\text{МНК}}(m) \right) = (X^T X)^{-1} X^T Y, \quad (4.57)$$

где матрица Y составлена из столбцов $Y^{(1)}, Y^{(2)}, \dots, Y^{(m)}$, т.е. является *матрицей наблюдаемых значений эндогенных переменных*, представленной в (4.26). Но матрица, стоящая в левой части (4.57), есть не что иное, как МНК-оценка для Π^T , следовательно, вместо (4.57) мы можем записать

$$\hat{\Pi}_{\text{МНК}}^T = (X^T X)^{-1} X^T Y, \quad (4.57')$$

или, транспонируя обе части этого соотношения (воспользовавшись правилом транспонирования произведения матриц, см. Приложение 2), получаем:

$$\hat{\Pi}_{\text{МНК}} = Y^T X (X^T X)^{-1}. \quad (4.58)$$

Точечный прогноз эндогенных переменных $\hat{Y}_{n+\tau}$ на τ тактов времени вперед, т.е. точечная оценка значений $Y_{n+\tau}$ по заданным значениям предопределенных переменных $X_{n+\tau}$, опирающийся на наблюдаемые значения этих переменных до момента времени $t = n$ (т.е. на исходные данные (4.26)), строится с помощью формулы

$$\hat{Y}_{n+\tau} = \hat{\Pi}_{\text{МНК}} X_{n+\tau}, \quad (4.59)$$

где матрица МНК-оценок $\hat{\Pi}_{\text{МНК}}$ коэффициентов приведенной формы модели (4.19) определена соотношением (4.58). Заметим, что *истинные* значения эндогенных переменных в прогнозируемый период будут равны в соответствии с (4.55)

$$Y_{n+\tau} = \Pi X_{n+\tau} + \varepsilon_{n+\tau}.$$

Следовательно, ошибка $\hat{\varepsilon}(\tau)$ прогноза (4.59) может быть выражена в виде

$$\hat{\varepsilon}(\tau) = \hat{Y}_{n+\tau} - Y_{n+\tau} = (\hat{\Pi}_{\text{МНК}} - \Pi)X_{n+\tau} - \varepsilon_{n+\tau}. \quad (4.60)$$

А поскольку $E\hat{\Pi}_{\text{МНК}} = \Pi$ (в силу свойства несмещенности МНК-оценок в обобщенной линейной модели множественной регрессии, см. п. 2.6.2) и $E\varepsilon_{n+\tau} = \mathbf{0}_m$, то, применяя к (4.60) операцию усреднения E , убеждаемся в *несмещенности прогноза* (4.59), т. е. в том, что $E\hat{\varepsilon}(\tau) = \mathbf{0}_m$.

Интервальный прогноз эндогенных переменных Y_t или построение *совместных* доверительных интервалов для различных сочетаний компонент этого вектора на τ тактов времени вперед возможны, если мы будем знать (сможем оценить) ковариационную матрицу ошибок прогноза $\hat{\varepsilon}(\tau) = \hat{Y}_{n+\tau} - Y_{n+\tau}$, т. е. матрицу

$$\Sigma_{\hat{\varepsilon}(\tau)} = E(\hat{\varepsilon}(\tau)\hat{\varepsilon}^T(\tau)) = E[(\hat{Y}_{n+\tau} - Y_{n+\tau})(\hat{Y}_{n+\tau} - Y_{n+\tau})^T]. \quad (4.61)$$

В условиях соблюдения стандартных модельных допущений, сформулированных, в частности, при описании модели (4.19), удастся получить несмещенную оценку $\hat{\Sigma}_{\hat{\varepsilon}(\tau)}$ для ковариационной матрицы $\Sigma_{\hat{\varepsilon}(\tau)}$ в виде

$$\hat{\Sigma}_{\hat{\varepsilon}(\tau)} = [1 + X_{n+\tau}^T(X^T X)^{-1}X_{n+\tau}]\hat{\Sigma}_{\varepsilon}, \quad (4.62)$$

где матрица $\hat{\Sigma}_{\varepsilon}$ является несмещенной оценкой ковариационной матрицы остатков приведенной формы $\Sigma_{\varepsilon} = E(\varepsilon_t \varepsilon_t^T) = (s_{ij})_{i,j=1,2,\dots,m}$ и задается формулой

$$\hat{\Sigma}_{\varepsilon} = \frac{1}{n-p} (Y - X\hat{\Pi}_{\text{МНК}})^T (Y - X\hat{\Pi}_{\text{МНК}}) \quad (4.63)$$

(вывод соотношения (4.62) дается в конце данного пункта).

Теперь, если *дополнительно предположить нормальный характер распределения случайных возмущений структурной* ($\delta_t^{(i)}$), а значит, и *приведенной* ($\varepsilon_t^{(i)}$) *форм анализируемой модели*, мы располагаем всем необходимым для построения интервальных (как *совместных* — одновременно по набору эндогенных переменных, — так и *покоординатных* — для каждой эндогенной переменной в отдельности) прогнозов.

Отклонения значений $\hat{Y}_{n+\tau} = (\hat{y}_{n+\tau}^{(1)}, \hat{y}_{n+\tau}^{(2)}, \dots, \hat{y}_{n+\tau}^{(m)})^\top$ эндогенных переменных, *спрогнозированных* на τ тактов времени вперед, от их *истинных* будущих значений $Y_{n+\tau}$ с заданной доверительной вероятностью $P = 1 - \alpha$ должны концентрироваться внутри эллипсоида рассеяния, определяемого уравнением

$$(\hat{Y}_{n+\tau} - Y_{n+\tau})^\top \hat{\Sigma}_\varepsilon^{-1} (\hat{Y}_{n+\tau} - Y_{n+\tau}) \leq [1 + X_{n+\tau}^\top (X^\top X)^{-1} X_{n+\tau}] \times \frac{(n-p)m}{n-p-m+1} F_\alpha(m; n-p-m+1), \quad (4.64)$$

где $X_{n+\tau} = (x_{n+\tau}^{(1)}, x_{n+\tau}^{(2)}, \dots, x_{n+\tau}^{(p)})^\top$ — заданные (на прогнозный период) значения предопределенных переменных, а $F_\alpha(m; n-p-m+1)$ — $100\alpha\%$ -ная точка F -распределения с числами степеней свободы числителя и знаменателя, равными m и $n-p-m+1$.

С практической точки зрения бывает удобнее знать гарантированные (*одновременно для всех эндогенных переменных*) пределы варьирования каждой из компонент $y_{n+\tau}^{(i)}$ ($i = 1, 2, \dots, n$), обозначенные на своей координатной оси $Oy^{(i)}$. Геометрически эти покоординатные интервалы определяются проекциями эллипсоида (4.64) на каждую из координатных осей $Oy^{(i)}$. Хьюманс показал², что эти проекции могут быть заданы соотношениями вида (одновременно выполняющимися для всех $y_{n+\tau}^{(i)}$ с вероятностью $P = 1 - \alpha$):

$$\left\{ \begin{array}{l} \hat{y}_{n+\tau}^{(i)} - \sqrt{c_\alpha \hat{s}_{ii}} \leq y_{n+\tau}^{(i)} \leq \hat{y}_{n+\tau}^{(i)} + \sqrt{c_\alpha \hat{s}_{ii}}, \\ i = 1, 2, \dots, m, \end{array} \right. \quad (4.64')$$

где

$$c_\alpha = [1 + X_{n+\tau}^\top (X^\top X)^{-1} X_{n+\tau}] \frac{(n-p)m}{n-p-m+1} F_\alpha(m; n-p-m+1),$$

а \hat{s}_{ii} — i -й диагональный элемент матрицы $\hat{\Sigma}_\varepsilon$, определенной соотношением (4.63).

¹ Этот вывод опирается на результат Хупера и Зельнера, которые показали, что статистика $\frac{n-p-m+1}{(n-p)m} (\hat{Y}_{n+\tau} - Y_{n+\tau})^\top \hat{\Sigma}_{\hat{Y}}^{-1} (\hat{Y}_{n+\tau} - Y_{n+\tau})$ подчиняется $F(m; n-p-m+1)$ -распределению. См. Hooper J. W., Zellner A. The Error of Forecast for Multivariate Regression Models. «Econometrica», 29 (1961), pp. 544–555.

² См. Humans S. H. Simultaneous Confidence Intervals in Econometric Forecasting. «Econometrica», 36 (1968), pp. 18–30.

Соотношения (4.64) и (4.64') задают совместные доверительные области (интервалы) для всех анализируемых эндогенных переменных одновременно. Если же нас интересует *интервальный прогноз одной отдельно взятой эндогенной переменной* $y^{(i)}$, выполняющийся с заданной вероятностью $P = 1 - \alpha$, то он определяется неравенствами

$$\left| y_{n+\tau}^{(i)} - \hat{y}_{n+\tau}^{(i)} \right| \leq t_{\frac{\alpha}{2}}(n-p) \sqrt{\hat{s}_{ii} \cdot [1 + X_{n+\tau}^T (X^T X)^{-1} X_{n+\tau}]}, \quad (4.65)$$

где $t_{\frac{\alpha}{2}}(n-p) = \sqrt{F_{\alpha}(1; n-p)}$ — 100 $\frac{\alpha}{2}$ %-ная точка распределения Стьюдента с $n-p$ степенями свободы. Из общих соображений следует ожидать (и это можно доказать строго математически), что ширина интервального прогноза (4.65) для *одной отдельно взятой эндогенной переменной* всегда существенно меньше, чем ширина интервального прогноза для той же самой переменной, получающаяся в рамках задачи построения *совместных доверительных интервалов* (4.64').

Вывод формулы (4.62) для несмещенной оценки ковариационной матрицы ошибок прогноза. Вернемся к задаче вычисления ковариационной матрицы ошибок прогноза $\Sigma_{\varepsilon}(\tau)$ (см. (4.61)) и построения ее несмещенной оценки $\hat{\Sigma}_{\varepsilon}(\tau)$.

1) Подставим в правую часть (4.61) вместо ошибки прогноза $\hat{Y}_{n+\tau} - Y_{n+\tau}$ ее выражение из (4.60):

$$\begin{aligned} \Sigma_{\varepsilon}(\tau) &= \mathbf{E} \{ [(\hat{\Pi}_{\text{мнк}} - \Pi) X_{n+\tau} - \varepsilon_{n+\tau}] [(\hat{\Pi}_{\text{мнк}} - \Pi) X_{n+\tau} - \varepsilon_{n+\tau}]^T \} \\ &= \mathbf{E} [(\hat{\Pi}_{\text{мнк}} - \Pi) X_{n+\tau} X_{n+\tau}^T (\hat{\Pi}_{\text{мнк}} - \Pi)^T] + \mathbf{E}(\varepsilon_{n+\tau} \varepsilon_{n+\tau}^T) \\ &\quad - \mathbf{E}[\varepsilon_{n+\tau} X_{n+\tau}^T (\hat{\Pi}_{\text{мнк}} - \Pi)^T] \\ &\quad - \mathbf{E}[(\hat{\Pi}_{\text{мнк}} - \Pi) X_{n+\tau} \varepsilon_{n+\tau}^T]. \end{aligned} \quad (4.61')$$

Два последних члена в правой части этого выражения равны нулю. Действительно, рассмотрим, например, произведение $\varepsilon_{n+\tau} \cdot [X_{n+\tau}^T (\hat{\Pi}_{\text{мнк}} - \Pi)^T]$. Его первый сомножитель ($\varepsilon_{n+\tau}$) не коррелирован со вторым, т. к. $\varepsilon_{n+\tau}$ не коррелированы и с предопределенными переменными $X_{n+\tau}^T$ (по условию), и с $(\hat{\Pi} - \Pi)^T$, т. к. последнее выражение зависит только от *прошлых* (по отношению к моменту времени $n + \tau$) наблюдаемых значений $x_i^{(j)}$ и $y_i^{(l)}$ ($j = 1, 2, \dots, p$; $l = 1, 2, \dots, m$; $t = 1, 2, \dots, n$), см. формулу (4.58). А по условию случайные возмущения приведенной формы не коррелированы ни с предопределенными, ни с лагированными (т. е. измеренными в прошлом) значениями эндогенных переменных. В точности те же доводы относятся и к последнему члену правой части соотношения (4.61'). Таким образом,

получаем:

$$\Sigma_{\hat{\epsilon}(\tau)} = \mathbf{E}[(\hat{\Pi}_{\text{мнк}} - \Pi)X_{n+\tau}X_{n+\tau}^T(\hat{\Pi}_{\text{мнк}} - \Pi)^T] + \Sigma_{\epsilon}, \quad (4.66)$$

где $\Sigma_{\epsilon} = \mathbf{E}(\epsilon_t \epsilon_t^T) = \mathbf{E}(\epsilon_{n+\tau} \epsilon_{n+\tau}^T)$ — ковариационная матрица остатков приведенной формы (она по условию не зависит от t); $s_{ij} = \text{cov}(\epsilon_t^{(i)} \epsilon_t^{(j)}) = \mathbf{E}(\epsilon_t^{(i)} \epsilon_t^{(j)})$ — (i, j) -й элемент этой матрицы.

2) Займемся теперь приведением первого слагаемого правой части (4.66) к «работоспособному» виду, т. е. постараемся выразить его в терминах имеющихся у нас наблюдений (4.26). С этой целью вернемся к оценке $\hat{\Pi}_{\text{мнк}}$ и подставим в правую часть соотношения (4.58) матрицу \mathbf{Y} , выраженную в терминах приведенной формы анализируемой модели, т. е. $\mathbf{Y} = \mathbf{X}\Pi^T + \epsilon$, где матрица ϵ составлена из столбцов $\epsilon(i) = (\epsilon_1^{(i)}, \epsilon_2^{(i)}, \dots, \epsilon_n^{(i)})^T$ ($i = 1, 2, \dots, m$):

$$\begin{aligned} \hat{\Pi}_{\text{мнк}} &= \mathbf{Y}^T \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} = (\mathbf{X}\Pi^T + \epsilon)^T \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \\ &= \Pi \mathbf{X}^T \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} + \epsilon^T \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} = \Pi + \epsilon^T \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1}, \end{aligned}$$

откуда

$$\hat{\Pi}_{\text{мнк}} - \Pi = \epsilon^T \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1}. \quad (4.67)$$

Подставим это выражение в первое слагаемое правой части (4.66) и попробуем упростить получившееся выражение с учетом представления матрицы ϵ в виде $\epsilon = (\epsilon(1) : \epsilon(2) : \dots : \epsilon(m))$:

$$\begin{aligned} &\mathbf{E}[(\hat{\Pi}_{\text{мнк}} - \Pi)X_{n+\tau}X_{n+\tau}^T(\hat{\Pi}_{\text{мнк}} - \Pi)^T] \\ &= \mathbf{E}[\epsilon^T \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} X_{n+\tau} X_{n+\tau}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \epsilon] \\ &= \mathbf{E} \left(\begin{array}{cccc} \epsilon^T(1)M^T M \epsilon(1) & \epsilon^T(1)M^T M \epsilon(2) & \dots & \epsilon^T(1)M^T M \epsilon(m) \\ \epsilon^T(2)M^T M \epsilon(1) & \epsilon^T(2)M^T M \epsilon(2) & \dots & \epsilon^T(2)M^T M \epsilon(m) \\ \dots & \dots & \dots & \dots \\ \epsilon^T(m)M^T M \epsilon(1) & \epsilon^T(m)M^T M \epsilon(2) & \dots & \epsilon^T(m)M^T M \epsilon(m) \end{array} \right), \end{aligned} \quad (4.68)$$

где $M = X_{n+\tau}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ — строка длины n (т. е. матрица размерности $1 \times n$). Заметим, что все элементы матрицы, стоящей в правой части (4.68), являются *числами*, или, формально, матрицами размерности 1×1 . А такие матрицы равны своему следу. Мы воспользуемся этим при вычислении элементов правой части (4.68), а также известным свойством коммутативности при вычислении следа произведения квадратных

матриц (т. е. $\text{tr}(AB) = \text{tr}(BA)$):

$$\begin{aligned} \mathbf{E}(\varepsilon^\top(i)\mathbf{M}^\top\mathbf{M}\varepsilon(j)) &= \mathbf{E}[\text{tr}(\varepsilon^\top(i)\mathbf{M}^\top\mathbf{M}\varepsilon(j))] \\ &= \mathbf{E}[\text{tr}(\mathbf{M}\varepsilon(j)\varepsilon^\top(i)\mathbf{M}^\top)] = \text{tr}\{\mathbf{M}[\mathbf{E}(\varepsilon(j)\varepsilon^\top(i))]\mathbf{M}^\top\} \\ &= \text{tr}(\mathbf{M}\text{cov}(\varepsilon^{(i)}, \varepsilon^{(j)})\mathbf{I}_n\mathbf{M}^\top) = \text{cov}(\varepsilon^{(i)}, \varepsilon^{(j)}) \text{tr}(\mathbf{M}\mathbf{M}^\top) \\ &= s_{ij}X_{n+\tau}^\top(\mathbf{X}^\top\mathbf{X})^{-1}X_{n+\tau}. \end{aligned} \quad (4.69)$$

При выводе (4.69) мы воспользовались также следующими фактами:

(i) значения \mathbf{M} неслучайны, т. к. выражаются только через наблюдаемые в прошлом и заданные на прогнозируемый момент значения предикторных переменных;

(ii) $\mathbf{M}\mathbf{M}^\top = [X_{n+\tau}^\top(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top][\mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}X_{n+\tau}] = X_{n+\tau}^\top(\mathbf{X}^\top\mathbf{X})^{-1}X_{n+\tau}$;

(iii)
$$\mathbf{E}(\varepsilon(j)\varepsilon^\top(i)) = \mathbf{E} \begin{pmatrix} \varepsilon_1^{(j)}\varepsilon_1^{(i)} & \varepsilon_1^{(j)}\varepsilon_2^{(i)} & \dots & \varepsilon_1^{(j)}\varepsilon_n^{(i)} \\ \varepsilon_2^{(j)}\varepsilon_1^{(i)} & \varepsilon_2^{(j)}\varepsilon_2^{(i)} & \dots & \varepsilon_2^{(j)}\varepsilon_n^{(i)} \\ \dots & \dots & \dots & \dots \\ \varepsilon_n^{(j)}\varepsilon_1^{(i)} & \varepsilon_n^{(j)}\varepsilon_2^{(i)} & \dots & \varepsilon_n^{(j)}\varepsilon_n^{(i)} \end{pmatrix}$$

$$= \begin{pmatrix} \text{cov}(\varepsilon^{(j)}, \varepsilon^{(i)}) & 0 & \dots & 0 \\ 0 & \text{cov}(\varepsilon^{(j)}, \varepsilon^{(i)}) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \text{cov}(\varepsilon^{(j)}, \varepsilon^{(i)}) \end{pmatrix} = s_{ij}\mathbf{I}_n,$$

где \mathbf{I}_n — единичная матрица $n \times n$ (в данных выкладках использовались модельные допущения о равенстве нулю ковариаций вида $\text{cov}(\varepsilon_{t_1}^{(i)}, \varepsilon_{t_2}^{(j)})$ при $t_1 \neq t_2$, а также независимость ковариаций $\text{cov}(\varepsilon_t^{(i)}, \varepsilon_t^{(j)})$ от времени t).

3) Теперь вернемся к матрице (4.66), подставим в нее правую часть (4.68) (с учетом (4.69)) и получим окончательный вид для матрицы $\Sigma_{\hat{\varepsilon}(\tau)}$. Действительно, после подстановки в правую часть (4.68) элементов этой матрицы, выраженных соотношениями (4.69), имеем:

$$\mathbf{E}[(\hat{\Pi}_{\text{мнк}} - \Pi)X_{n+\tau}X_{n+\tau}^\top(\hat{\Pi}_{\text{мнк}} - \Pi)^\top] = X_{n+\tau}^\top(\mathbf{X}^\top\mathbf{X})^{-1}X_{n+\tau}\Sigma_{\varepsilon}, \quad (4.70)$$

где матрица Σ_{ε} составлена из элементов $s_{ij} = \text{cov}(\varepsilon^{(i)}, \varepsilon^{(j)})$. Подставляя (4.70) в правую часть (4.66), получаем:

$$\Sigma_{\hat{\varepsilon}(\tau)} = [1 + X_{n+\tau}^\top(\mathbf{X}^\top\mathbf{X})^{-1}X_{n+\tau}]\Sigma_{\varepsilon}. \quad (4.71)$$

4) Несмещенная оценка $\hat{\Sigma}_\varepsilon(\tau)$ матрицы $\Sigma_\varepsilon(\tau)$ получается подстановкой в правую часть (4.71) вместо теоретической ковариационной матрицы Σ_ε ее несмещенной оценки (4.63). Вывод формулы (4.62) закончен.

4.5. Некоторые общие подходы к анализу точности оценивания и к сравнению методов и моделей

Подавляющее большинство методов и моделей эконометрики базируется, как мы видели, на математическом аппарате *многомерного статистического анализа*. Соответственно, «болевые точки», «узкие места», словом, главные трудности этого аппарата как бы ретранслируются на эконометрический инструментарий. Среди этих главных трудностей, в первую очередь, следует упомянуть о проблемах, вынесенных в заглавие данного пункта учебника. Действительно, мы неоднократно видели, как непросто достаются результаты, связанные с *аналитическим* выводом характеристик точности статистически построенных методов и моделей, их прогностической силы (вывод формулы (4.62) в конце предыдущего пункта — последнее тому доказательство). Более того, даже тогда, когда нам удастся аналитический способ исследования качества методов статистического оценивания неизвестных значений параметров модели и построенных на их основе прогнозов, следует помнить, что, как правило, полученные результаты имеют *асимптотический* (по $n \rightarrow \infty$) характер (свойство состоятельности оценок, например, или асимптотический вид ковариационных матриц и т. п.). В эконометрической же практике работают с *конечными* выборками (т. е. при существенно ограниченных n), и поэтому *важно уметь проанализировать свойства различных способов оценивания и различных методов прогнозирования в условиях относительно малых выборок*. К сожалению, как правило, это не удается сделать с помощью аналитических методов и приходится прибегать к разного рода *имитационно-компьютерным экспериментам*. Мы уже упоминали о подобных подходах: в п. 3.5.3 — о так называемом методе «*перекрестного анализа дееспособности*» (модели или метода) в связи с подбором конкретного значения для свободного параметра дисконтирования λ в геометрической лаговой структуре Койка и в п. 3.1.1 — в связи с оценкой качества и сравнением различных прогнозных методов, основанных на тех или иных моделях временных рядов.

Ниже мы кратко остановимся на описании трех наиболее распространенных в эконометрической практике подходах такого типа.

Метод Монте-Карло (статистических испытаний). Сущность

метода статистических испытаний заключается в следующем. В компьютер закладываются все параметры анализируемой стохастической модели, после чего компьютер с помощью специальных датчиков случайных чисел *генерирует требуемое число выборок заданного объема*, которые, соответственно, интерпретируются как *наблюдения, произведенные над данной моделью*. Затем эти выборки используются в той статистической процедуре (в статистическом оценивании параметров анализируемой модели, в построении различного рода прогнозов и т. п.), качество которой мы хотим исследовать. Наличие большого числа сгенерированных компьютером выборок заданного объема n позволяет исследовать поведение интересующих нас выходных характеристик анализируемой процедуры (оценок $\hat{\Theta}(n)$ или прогнозов \hat{y}_n) и, в первую очередь, оценить их смещение и среднеквадратическую ошибку. Конечно, результаты такого исследования как бы «условно-локальны», т. е. зависят от тех численных значений параметров модели, которые мы заложили в компьютер и которые, соответственно, определяют специфику сгенерированных «наблюдений». Однако подобные эксперименты, проведенные в большом объеме и в широком диапазоне значений закладываемых в компьютер параметров, способны доставить исследователю ценную информацию.

Конкретизируем описанную процедуру применительно к анализу методов оценивания параметров СОУ и точности построенных на основании этой модели прогнозов. Итак, зададимся:

- 1) структурными параметрами СОУ, т. е. числом уравнений и эндогенных переменных m , а также числом предопределенных переменных p ;
- 2) конкретными значениями структурных параметров B и C ;
- 3) вероятностными распределениями структурных возмущений $\delta_t^{(1)}, \delta_t^{(2)}, \dots, \delta_t^{(m)}$;
- 4) требуемым объемом генерируемых выборок n ;
- 5) конкретными «наблюденными» значениями предопределенных переменных $x_t^{(1)}, x_t^{(2)}, \dots, x_t^{(p)}$, $t = 1, 2, \dots, n$;
- 6) значениями предопределенных переменных в прогнозный период времени (при анализе качества прогноза);

После этого:

- 7) с помощью компьютерного датчика случайных чисел моделируются случайные остатки $\delta_t^{(1)}, \delta_t^{(2)}, \dots, \delta_t^{(m)}$, $t = 1, 2, \dots, n$;
 - 8) в соответствии с соотношениями приведенной формы $Y_t = -B^{-1}CX_t + B^{-1}\Delta_t$ вычисляются значения Y_t , $t = 1, 2, \dots, n$.
- В результате пп. 1)–8) сгенерирована 1-я выборка объема n : $\{X(1), Y(1)\}$.

Затем пп. 7)–8) повторяются еще $N - 1$ раз, в результате чего получаем еще $N - 1$ выборки того же самого объема из той же самой генеральной совокупности (т. е. для той же самой модели).

Таким образом, в нашем распоряжении оказываются N выборок объема n , сгенерированных в рамках одной и той же модели:

$$\{X(1), Y(1)\}, \{X(2), Y(2)\}, \dots, \{X(N), Y(N)\}. \quad (4.72)$$

По «наблюдениям» каждой (j -й) выборки строим интересующие нас выходные характеристики анализируемой процедуры (оценки структурных параметров $\hat{B}(j), \hat{C}(j)$, прогнозы $\hat{Y}_{n+\tau}(j)$ и т. п.). Сравниваем полученные оценки $\hat{B}(j), \hat{C}(j)$ и прогнозы $\hat{Y}_{n+\tau}(j)$ с имеющимися у нас истинными значениями, соответственно, B, C и $Y_{n+\tau}$: оцениваем среднюю (по всем выборкам) величину смещения, среднеквадратическую ошибку и другие интересующие нас характеристики качества анализируемой процедуры.

Проиллюстрируем применение метода Монте-Карло на примере.

Пример 4.4. (заимствован из: *Simister L. T. Monte Carlo Studies of Simultaneous Equations System. Ph. D Thesis, University of Manchester, 1969*). Автор этой работы исследовал с помощью метода Монте-Карло влияние различных способов спецификации структурных остатков $\delta_t^{(i)}$ системы из двух уравнений ($i = 1, 2$) на точность прогноза, осуществляемого разными методами. Из методов, описанных выше в данной главе, сравнивались три процедуры:

- *МНК*, непосредственно примененный для оценивания параметров β и с каждого из двух структурных уравнений; затем на основании соотношения (4.20) определялись оценки параметров приведенной формы и строился прогноз для $y^{(1)}$ и $y^{(2)}$ по формуле (4.59);
- *МНК без ограничений (МНК БО)* применялся непосредственно для оценивания параметров π приведенной формы, после чего строился прогноз для $y^{(1)}$ и $y^{(2)}$ по формуле (4.59);
- *2 МНК*, с помощью которого оценивались параметры структурной формы, по ним (с помощью (4.20)) — параметры приведенной формы, а затем строился прогноз по формуле (4.59).

В качестве возможных спецификаций структурных возмущений $\delta_t^{(1)}$ и $\delta_t^{(2)}$ рассматривались 4 варианта:

- (i) $\delta_t^{(1)}$ и $\delta_t^{(2)}$ ($t = 1, 2, \dots, n$) образуют белый шум;
- (ii) $\delta_t^{(1)}$ связаны автокорреляцией 1-го порядка с параметром $\rho_1 = r(1) = 0,9$, а $\delta_t^{(2)}$ связаны автокорреляцией 1-го порядка с параметром $\rho_2 = r(1) = 0,225$;

- (iii) на стандартные возмущения (i) накладываются ошибки в измерении переменных модели (вариации ошибок составляли 10% и более от вариаций соответствующих переменных);
- (iv) на случайные возмущения марковского типа (ii) накладываются ошибки (iii) в измерении анализируемых переменных.

В табл. 4.1 приведены результаты оценки с помощью метода Монте-Карло среднеквадратических ошибок прогнозов значений переменных $y^{(1)}$ и $y^{(2)}$ для различных сочетаний «метод оценивания — вариант спецификации структурных возмущений модели». Величина среднеквадратической ошибки прогноза, получаемой при оценивании параметров модели с помощью 2 МНК при спецификации остатков типа (i), принята за эталонную единицу в сравнении различных методов.

Таблица 4.1 Среднеквадратические ошибки прогнозов $\hat{y}_{n+\tau}^{(1)}$ и $\hat{y}_{n+\tau}^{(2)}$ при различных методах оценивания параметров модели

Метод оценивания	При прогнозе $y^{(1)}$				При прогнозе $y^{(2)}$			
	Вариант спецификации возмущений $\delta_t^{(1)}$				Вариант спецификации возмущений $\delta_t^{(2)}$			
	(i)	(ii)	(iii)	(iv)	(i)	(ii)	(iii)	(iv)
2 МНК	Δ	1,47 Δ	1,83 Δ	2,25 Δ	Δ'	1,86 Δ'	3,1 Δ'	2,98 Δ'
МНК	3,59 Δ	1,36 Δ	2,39 Δ	2,74 Δ	12,32 Δ'	1,64 Δ'	8,4 Δ'	8,12 Δ'
МНК БО	1,13 Δ	2,98 Δ	3,86 Δ	2,36 Δ	1,17 Δ'	2,44 Δ'	5,53 Δ'	3,06 Δ'

Из анализа табл. 4.1 следует, что 2 МНК оказался самым точным во всех ситуациях, кроме случая автокоррелированных возмущений $\delta_t^{(1)}$ и $\delta_t^{(2)}$. В последнем случае несколько неожиданно лучше других проявил себя обыкновенный метод наименьших квадратов, примененный в отдельности к каждому из двух уравнений структурной формы.

Бутстреп метод «тиражирования» наблюдений. Этот метод получил весьма широкое распространение в последние два десятилетия в задачах статистического анализа, основанного на относительно малых выборках. Заложенная в его основание идея внешне выглядит несколько странной: метод предлагает тиражировать уже имеющиеся в нашем распоряжении наблюдения, генерируя их, как и в методе Монте-Карло, с помощью специальных компьютерных датчиков случайных чисел. Однако в методе Монте-Карло эти датчики «настраиваются» на экзогенно заданные параметры модели (там мы «на старте» не имели никаких наблюдений), в то время как в *бутстреп-процедурах* «настройка» датчиков

определяется структурой и спецификой уже имеющейся в нашем распоряжении выборки. Грубо говоря, схема действия бутстреп-процедуры заключается в следующем: по имеющейся выборке строится некоторая оценка (параметрическая или непараметрическая) закона распределения анализируемой генеральной совокупности, а затем генерируются (в необходимом количестве) наблюдения, подчиняющиеся этому закону распределения вероятностей. Как обрабатываются полученные таким образом «наблюдения» и какие с их помощью решаются задачи оценки качества модели или прогноза — это предмет для специального обсуждения, и мы оставляем его за рамками данного учебника¹.

Перекрестный анализ дееспособности модели (или ПАД-процедура)². Идеи этого подхода были сформулированы достаточно давно, однако высокую прикладную значимость и популярность различные версии ПАД-процедур обрели лишь с появлением современных вычислительных мощностей. Их сущность заключается в следующем.

Пусть мы решаем проблему подбора модели по имеющимся в нашем распоряжении исходным статистическим данным, но не уверены в ее общем виде, в ее структуре. Так, если речь идет о подборе регрессионной модели, то мы не знаем, является ли она линейной, степенной или еще какой-нибудь, нужно ли проводить линеаризующие преобразования переменных, каким выбрать значение «гребневого» параметра в методе ридж-регрессии (см. (2.51), с.) и т. д. Если речь идет о модели временных рядов, то мы обычно испытываем затруднения в выборе структурных параметров p, q и k АРСС-моделей, параметров адаптации (сглаживания) $\lambda_1, \lambda_2, \lambda_3$ в различных процедурах экспоненциального

¹ Интересующемуся этой проблематикой читателю мы можем порекомендовать, например, книгу: Б. Эфрон. Нетрадиционные методы многомерного статистического анализа. (Перев. с англ.) М.: Финансы и статистика, 1988.

² Здесь предложен вариант перевода английского названия этого подхода, впервые достаточно полно описанного как «*Cross-validation method*», по-видимому, в англоязычной специальной литературе (см., например, Stone C. Cross-validated choice and assessment of statistical predictions. «*Journ. Roy. Stat. Soc.*», Ser. B, 36 (1974), pp. 111–147). К сожалению, нет единой точки зрения на состав методов, объединяемых этим названием. Некоторые специалисты включают в рамки этого подхода и бутстреп-метод, и различные варианты метода «скользящего экзамена», о котором мы упоминали в п. 2.9.3 (см. Hinkley D.V. Jackknife, bootstrap and other cross-validation methods. London–New York: Chapman and Hall, 1983). Нам представляется, что из методических соображений удобнее несколько сузить столь широкое толкование «*Cross-validation*», акцентируя внимание на многократных вычислительных «прогонах» уже оцененных моделей при различных вариантах значений из свободных (т. е. не поддающихся статистическому оцениванию) параметров.

сглаживания или параметров дисконтирования λ или γ в моделях распределенных лагов с геометрической структурой. Аналогичные вопросы, связанные с подбором значений некоторых *свободных* (или *структурных*) параметров модели, возникают при построении и исследовании моделей дискриминантного или факторного анализа, систем одновременных уравнений и т. д. При этом под *свободными* (*структурными*) параметрами модели $\Lambda = (\lambda_1, \lambda_2, \dots, \lambda_m)$ мы в общем случае будем понимать параметры, от которых зависит общий вид (структура) модели, но *для которых не существует математически корректных процедур статистического оценивания*. И пусть существует некоторый экзогенно заданный *критерий качества модели*

$$K(\Lambda; \hat{\Theta}; X_{\text{экз}}), \quad (4.73)$$

значение которого зависит от значений использованных в модели свободных параметров Λ , статистических оценок неизвестных параметров модели $\hat{\Theta}$ и состава «экзаменующей выборки» $X_{\text{экз}}$, т. е. от состава тех экспериментальных (наблюдаемых) данных, на которых была проведена верификация построенной модели. В моделях регрессии и временных рядов в качестве критерия K чаще других используется среднеквадратическая ошибка прогноза, в моделях дискриминантного и кластерного анализов — доля неправильно расклассифицированных наблюдений из состава $X_{\text{экз}}$ и т. п.

Тогда *общую схему процедуры перекрестного анализа дееспособности модели* можно описать в виде реализации следующих двух этапов.

1-й этап. Все имеющиеся в нашем распоряжении наблюдения X (исходные статистические данные, результаты эксперимента) случайным образом разбиваются на две части $X_{\text{об}}$ и $X_{\text{экз}}$. Первая из них ($X_{\text{об}}$) используется для подбора («настройки») модели. Этот подбор обязательно включает в себя определение значений свободных параметров Λ и, если необходимо, статистическое оценивание параметров Θ . Поэтому эту часть исходных статистических данных называют обычно *«обучающей выборкой»* (хотя, строго говоря, это может не совпадать с классическим понятием обучающей выборки, принятым в математической статистике). Данная стадия процедуры, в частности, включает в себя выдвижение и предварительную проверку рабочих гипотез о структуре модели, перебор различных вариантов моделей с целью «нащупывания» относительно устойчивых результатов и т. п. При этом перебор различных вариантов модели обычно включает в себя задание некоторой сетки возможных значений структурных параметров $\lambda_1, \lambda_2, \dots, \lambda_m$ и проведение полного цикла всех необходимых вычислений по идентификации модели для каждого фиксированного сочетания этих значений (т. е. оценку параметров Θ и т. п.).

2-й этап. На этой стадии каждый из вычисленных на предыдущем этапе вариантов модели *верифицируется* (с помощью критерия (4.73) на данных «экзаменующей выборки» $X_{экз}$). Поскольку модели «настраивались» на данных $X_{об}$, очевидно, значения критерия качества на данных $X_{экз}$ будут существенно менее оптимистичными, чем на данных $X_{об}$. Из всех возможных значений структурного параметра Λ выбирается такое (Λ_0), при котором значение критерия $K(\Lambda_0; \hat{\Theta}; X_{экз})$ оптимально. Заметим, что в современных процедурах метода ПАД при делении массива исходных данных X на $X_{об}$ и $X_{экз}$ действуют так же, как и в методах скользящего экзамена (и в частности, в методе «складного ножа» — «jackknife», см. п. 2.9.3): в составе $X_{экз}$ оставляют *единственный* элемент, по остальным данным подбирают модель и вычисляют значение критерия K в этой единственной точке; затем в качестве $X_{экз}$ берут другой элемент из X и т. д. до тех пор, пока роль единственной «экзаменующей» точки не исполнят поочередно все элементы из массива исходных статистических данных X . «Настройка» модели на обучающих данных $X_{об}$ и последующая *перекрестная «перепроверка»* (верификация) модели на других (экзаменующих) данных $X_{экз}$ с неизменным многовариантным перебором значений структурных параметров на 1-й стадии процедуры — в этом и состоит сущность метода перекрестного анализа дееспособности модели (кстати, в некоторых переведенных с английского на русский язык работах, относящихся к данной тематике, метод «Cross-validation» переведен как «перепроверка», см. книгу [Эфрон Б.]).

ВЫВОДЫ

1. Эконометрическая модель, выраженная *системой одновременных уравнений* (СОУ), служит для объяснения поведения эндогенных (т. е. формирующихся в процессе и внутри функционирования описываемой социально-экономической системы) переменных в зависимости от значений экзогенных (задаваемых извне) и лаговых эндогенных переменных. СОУ широко используются в проведении многовариантных сценарных расчетов, касающихся социально-экономического развития анализируемой системы, а также в задачах прогноза экономических и социально-экономических показателей.

2. *Общая проблема идентификации СОУ* является, наряду со спецификацией модели, центральной в построении и анализе модели. Она включает в себя *проверку соблюдения условий идентифицируемости* каждого отдельного уравнения и всей системы в целом (см. п. 4.2), а также реализацию *процедуры статистического оценивания неизвестных значе-*

ний параметров системы (см. п. 4.3).

3. Проверка и соблюдение условий идентифицируемости производится как на стадии составления уравнений системы (1-е, 3-е и 4-е условия) и формирования массива исходных статистических данных (2-е условие), так и на стадии анализа результатов статистического оценивания неизвестных параметров системы (1-е условие в части, касающейся обратимости матрицы B , а также 5-е условие).

4. Методы статистического оценивания параметров СОУ подразделяются на два класса: 1) методы, предназначенные для оценки параметров *одного отдельно взятого* уравнения системы (МНК, косвенный МНК, 2МНК, метод максимального правдоподобия с ограниченной информацией, «оценки класса k »); 2) методы, предназначенные для одновременного оценивания параметров всех уравнений системы с учетом их взаимосвязей (3МНК, метод максимального правдоподобия с полной информацией).

5. Если уравнения структурной формы модели могут быть расположены в таком порядке, что i -е уравнение ($i = 1, 2, \dots, m$) может содержать в качестве объясняющих эндогенных переменных только переменные $y^{(1)}, y^{(2)}, \dots, y^{(i-1)}$ (или часть из них), а случайное возмущение $\delta_i^{(i)}$ этого уравнения не коррелирует со всеми этими эндогенными переменными, то такая система называется *рекурсивной*, и последовательное применение к каждому уравнению такой системы обычного МНК дает состоятельные оценки ее структурных параметров (см. п. 4.3.1). Класс рекурсивных систем является *простейшим* с точки зрения решения задачи оценивания структурных параметров СОУ.

6. Если исследователя интересуют только параметры приведенной формы и задача прогноза эндогенных переменных, то он может ограничиться применением обычного метода наименьших квадратов к каждому отдельному уравнению приведенной формы (с последующей оценкой, если это необходимо, *идентифицируемых* параметров структурной формы). Такой образ действий называют *косвенным методом наименьших квадратов*, или *методом наименьших квадратов без ограничений*, а оценки, полученные с его помощью, будут состоятельными (см. п. 4.3.2).

7. В ситуациях, когда среди уравнений системы имеются неидентифицируемые, так же как и в случаях, когда оценивание и анализ параметров *структурной* формы представляют для исследователя самостоятельный интерес, рекомендуется применять *двухшаговый метод наименьших квадратов* (2МНК). Этот метод предназначен для оценивания параметров *отдельного уравнения* структурной формы, а его последовательное применение к каждому из уравнений структурной формы СОУ позволяет получить состоятельные оценки всех структурных параметров (хотя 2МНК

и не учитывает возможные взаимосвязи между уравнениями системы).

8. Сущность двух шагов 2МНК заключается в следующем. На 1-м шаге для каждой эндогенной переменной, играющей роль объясняющей в анализируемом уравнении структурной формы, с помощью обычного МНК строится регрессия на все predetermined переменные X . На 2-м шаге эта эндогенная переменная заменяется в рассматриваемом уравнении ее регрессионным выражением через X , после чего в правой части этого уравнения остаются только predetermined переменные и к нему применяется обычный МНК. В моделях с большим числом predetermined переменных в целях снижения размерности рекомендуется на 1-м шаге строить регрессию предикторной эндогенной переменной не на все predetermined переменные, а лишь на небольшое число их главных компонент (см. п. 4.3.3).

9. Если структурные случайные возмущения $\delta_i^{(i)}$ различных уравнений системы взаимно коррелированы, то для оценивания структурных параметров рекомендуется применять *трехшаговый метод наименьших квадратов* (3МНК). Этот метод предназначен для одновременного оценивания структурных параметров всех уравнений системы и дает их состоятельные оценки, по эффективности превосходящие в данном случае оценки (тоже состоятельные) 2МНК.

10. 3МНК использует полученные на первых двух шагах 2МНК оценки структурных параметров для вычисления оценки ковариационной матрицы возмущений различных уравнений структурной формы. Затем на 3-м шаге оценки структурных параметров системы *пересчитываются* с помощью *обобщенного* МНК в рамках соответствующей схемы обобщенной линейной модели множественной регрессии, в которой в качестве ковариационной матрицы остатков используется полученная ранее оценка ковариационной матрицы возмущений (см. п. 4.3.4).

11. В ряде ситуаций могут оказаться полезными и другие методы статистического оценивания параметров СОУ. Для оценивания параметров *одного отдельно взятого уравнения* — это метод максимального правдоподобия с ограниченной информацией (требующий, правда, дополнительного априорного предположения о нормальном характере распределения структурных возмущений модели), «оценки класса k »; для одновременной оценки всех структурных параметров системы — это метод максимального правдоподобия с полной информацией. Однако эти методы из-за их относительно сложной вычислительной реализации и дополнительных априорных допущений существенно реже используются в эконометрических приложениях.

12. Одна из главных конечных прикладных целей построения и ана-

лиза эконометрических моделей в виде СОУ — это точечный и интервальный *прогноз эндогенных переменных* по заданным значениям предопределенных переменных и связанная с этим задача проведения *многовариантных сценарных расчетов*, показывающих, как будут «себя вести» эндогенные переменные при различных сочетаниях значений предопределенных переменных. «*Точечное решение*» этих задач основано на подсчете значений эндогенных переменных с помощью статистически оцененной приведенной формы СОУ. Для получения «*интервальных*» вариантов решения необходимо уметь оценивать ковариационную матрицу ошибок точечного прогноза, что является задачей аналитически достаточно сложной (см. п. 4.4).

13. Важные средства анализа качества построенных моделей, точности получаемых с их помощью прогнозов, наконец, решения задачи выбора наилучшей модели предоставляют исследователю разного рода *имитационно-компьютерные системы экспериментирования*. Среди них следует выделить метод *Монте-Карло*, *бутстреп-метод*, а также схему *перекрестного анализа дееспособности модели* (или метода).

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 $\phi\left(\frac{x_0 - a}{\sigma}\right)$ 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 $q = 0,9$, т. е. $u_{0,9} = 1,281552$.

З а м е ч а н и е 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

$$u_q = u_{q_1} + \frac{q - q_1}{q_2 - q_1} (u_{q_2} - u_{q_1}).$$

З а м е ч а н и е 2. При нахождении q -квантилей для значений $q < 0,5$ следует воспользоваться соотношением $u_q = -u_{1-q}$. Например, $u_{0,4} = -u_{1-0,4} = -u_{0,6} = -0,25335$.

З а м е ч а н и е 3. При отыскании 100 Q %-ных точек w_Q следует воспользоваться соотношением $w_Q = u_{1-Q}$. Например, $w_{0,05} = u_{0,95} = 1,64485$.

Таблица П1.4. Значения 100Q%-ных точек $\chi^2(v)$ χ^2 -распределения с v степенями свободы.

v	Q									
	0.995	0.990	0.975	0.950	0.900	0.100	0.050	0.025	0.010	0.005
1	392704 · 10 ⁻¹⁰	157088 · 10 ⁻⁹	982069 · 10 ⁻⁹	393214 · 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	10.0652	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	8.90555	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 $v^2_Q(v_1, v_2)$ F -распределения с числом степеней свободы числителя v_1 и знаменателя v_2 .

v_1	v_2														∞				
	1	2	3	4	5	6	7	8	9	10	12	15	20	24		30	40	60	120
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

$Q=0,1$

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	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18	2,13

Продолжение таблицы III.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

v ₁	Q=0,01												∞						
	1	2	3	4	5	6	7	8	9	10	12	15		20	24	30	40	60	120
1	4052	999,5	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	26,50	26,41	26,32	26,22	26,13	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

П р и м е ч а н и е. При вычислении 100%-ных точек для $Q \geq 0.9$ следует использовать тождество $v_Q^2(\mu_1, \mu_2) = (v_Q^2 - Q(\mu_2, \mu_1))^{-1}$.

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 П.1.7. Преобразование Фишера (x -преобразование) выборочного коэффициента корреляции r ($x = \operatorname{arctgh} r$).

r	,000	,002	,004	,006	,008	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 табл. 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

Указание. При работе с отрицательными значениями \hat{r} и \hat{z} используйте свойство нечетности функций $\text{th } \hat{z}$ и $\text{arcth } \hat{r}$, т.е. $\text{arcth } (-\hat{r}) = -\text{arcth } \hat{r}$.

П р и м е р ы.

1. Дано $\hat{r} = 0,206$. Определить $z = \text{arcth } 0,206$.

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, $z = \text{arcth } 0,2090$.

2. Дано $\hat{r} = -0,515$. Определить $z = \text{arcth } (-0,515)$.

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 $\text{arcth } 0,515$ находится как среднее арифметическое двух чисел данной строки, расположенных в столбцах, соответствующих верхним индексам 0,006 и 0,004, т.е.

$$\text{arcth } 0,515 = \frac{0,5709 + 0,5682}{2} = 0,56955.$$

Соответственно, $z = \text{arcth } (-0,515) = -\text{arcth } 0,515 = -0,56955$.

3. Дано $z = 0,8752$. Определить $\hat{r} = \text{th } z$.

Находим в таблице число, равное 0,8752, и определяем, какому значению \hat{r} оно соответствует. В нашем случае $\hat{r} = 0,704$.

П р и м е ч а н и е. В тех случаях, когда в таблице не найдется в точности заданного, 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 П.1.8. 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	0,92507	0,92877	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

Примечание. 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, $0,94692 = 0,9999692$.

Пример. 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 r 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

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	0,0064
45	0,0017	0,141	107	0,0033
49		0,123	109	0,0021
51		0,107	113	0,0014
53		0,093	117	0,0048
57		0,075	125	0,0030
59		0,067		

Продолжение

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

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	0,0090
44		0,077	0,230	130	0,0087
46		0,068	0,218	132	0,0073
48		0,054	0,197	134	0,0065
50		0,052	0,194	136	0,0040
52		0,036	0,163	138	0,0036
54		0,033	0,155	140	0,0028
56		0,019	0,127	144	0,0024
58		0,014	0,114	146	0,0022
62		0,012	0,108	148	0,0012
64		0,0069	0,089	150	0,0095
66		0,0062	0,088	152	0,0062
68		0,0027	0,073	154	0,0046
70		0,0027	0,066	158	0,0024
72		0,0016	0,060	160	0,0016
74		0,0094	0,056	162	0,0012
76		0,0094	0,043	164	0,0080
78		0,0094	0,041	170	0,0024
80		0,0072	0,037	180	0,0013

Продолжение

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	0,0090
20	0,682	42	0,253	64	0,045	90	0,0069

Таблица П1.11^б. Критические значения статистики S при уровне значимости $\alpha = 0,05$ для проверки статистической значимости выборочного значения коэффициента конкордации $\tilde{W}(m)$.

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 Дербина-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	0,96	1,54	0,82	1,75	0,69	1,97	0,58	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 П1.12^б. Значения статистик d_L и d_U критерия Дарбина-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	1,98
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 это и классификация многомерных данных, и всевозможные конструкции факторного анализа, включая метод главных компонент, и множественная регрессия, и, конечно, системы одновременных уравнений, и все, что связано с решением систем алгебраических уравнений (а сюда относятся и корреляционный анализ, и анализ временных рядов), — *требуют, для своего компактного и эффективного описания и анализа, использования аппарата матричной алгебры*. И если при описании этих моделей мы еще, как правило, 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 приводимых здесь фактов из линейной алгебры будут, как правило, опускаться.

П2.1. Виды матриц и действия с ними

Матрицей называется прямоугольная таблица чисел (элементов) вида

$$A = (a_{ij})_{\substack{i=1,2,\dots,n; \\ j=1,2,\dots,m}} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix}. \quad (\text{П2.1})$$

В общем обозначении элемента a_{ij} матрицы 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 A содержит nt элементов, расположенных на n строках и t столбцах, то говорят, что мы имеем $n \times t$ -матрицу A или матрицу A размерности $n \times t$.

Квадратной матрицей называется матрица, у которой число строк равно числу столбцов, т.е. $n \times n$ — матрица (при любом целом $n \geq 1$). Элементы $a_{11}, a_{22}, \dots, a_{nn}$ квадратной матрицы образуют ее *главную диагональ*.

Среди квадратных матриц выделим:

- **диагональную матрицу D** , у которой все элементы, кроме элементов, стоящих на главной диагонали, равны нулю, т.е.

$$D = \begin{pmatrix} d_{11} & 0 & 0 & \dots & 0 & 0 \\ 0 & d_{22} & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & d_{nn} \end{pmatrix} = \text{diag}(d_{11}, d_{22}, \dots, d_{nn});$$

заметим, что ковариационная матрица $\Sigma = (\sigma_{ij})$ любой последовательности $\xi^{(1)}, \xi^{(2)}, \dots, \xi^{(n)}$ *взаимно не коррелированных* случайных величин является диагональной, т.к. $\sigma_{ij} = E[(\xi^{(i)} - E\xi^{(i)})(\xi^{(j)} - E\xi^{(j)})] = 0$ при любых $i \neq j$.

- **единичную матрицу I_n** , которая является частным случаем $n \times n$ диагональной матрицы, поскольку все ее диагональные элементы равны единице:

$$I_n = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 0 & 1 \end{pmatrix};$$

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, т.е. $a_{ij} = a_{ji}$ для всех $i, j = 1, 2, \dots, n$; заметим, что всякая ковариационная матрица $\Sigma = (\sigma_{ij})$ является симметрической по определению, т.к. $\sigma_{ij} = E[(\xi^{(i)} - E\xi^{(i)})(\xi^{(j)} - E\xi^{(j)})] = E[(\xi^{(j)} - E\xi^{(j)})(\xi^{(i)} - E\xi^{(i)})] = \sigma_{ji}$.

Равенство матриц. Две матрицы $A = (a_{ij})$ и $B = (b_{ij})$ 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 $1 \times m$ ($m \geq 1$), т. е. *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-столбец

$$X_t = (x_t^{(1)}, x_t^{(2)}, \dots, x_t^{(p)}). \quad (\text{П2.2})$$

Вектор-столбец — это матрица размерности $n \times 1$ ($n \geq 1$), т. е. *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*, т. е. если A — симметрична, 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 , т. е. $(A^T)^T = A$.

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

$$\lambda A = (\lambda a_{ij})_{\substack{i=1,2,\dots,n_1 \\ j=1,2,\dots,m}}$$

т. е. 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 B умножается на матрицу A слева»). 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 , т. е. как сумма попарных произведений элементов i -й строки первой матрицы на соответствующие (т. е. 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, т. е.*

$$AB \neq BA. \quad (\text{П2.8})$$

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 $n \times n$ -матрицы A имеют место тождества

$$AI_n = I_n A = A. \quad (\text{П2.11})$$

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 *непрерывным «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

$$X^T X = \sum_{i=1}^n x_i^2 - \tag{П2.13}$$

это 1×1 -матрица, т. е. *число*, а

$$X X^T = \begin{pmatrix} x_1^2 & x_1 x_2 & \dots & x_1 x_n \\ x_2 x_1 & x_2^2 & \dots & x_2 x_n \\ \dots & \dots & \dots & \dots \\ x_n x_1 & x_n x_2 & \dots & x_n^2 \end{pmatrix} - \tag{П2.14}$$

это матрица размерности $n \times n$.

Матрицы типа (П2.13) и (П2.14) играют заметную роль в регрессионном анализе и в системах одновременных уравнений (см. гл. 1, 2, 4). 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

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1m}B \\ a_{21}B & a_{22}B & \dots & a_{2m}B \\ \dots & \dots & \dots & \dots \\ a_{n1}B & a_{n2}B & \dots & a_{nm}B \end{pmatrix}, \tag{П2.15}$$

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. ниже в п. П2.8).

Отметим, что некоторые эконометрические построения (в первую очередь, связанные с вычислительными проблемами *систем регрессионных уравнений*, см. гл. 4) значительно упрощаются, если использовать понятие прямого (кронекерова) произведения.

Матричное дифференцирование. Достаточно полное изложение аппарата векторного и матричного дифференциального исчисления читатель найдет в книге [Magnus J.R., Neudecker H.]¹. Мы же остановимся лишь на тех определениях и результатах, которые непосредственно использованы в нашем учебнике.

Пусть $\Theta = (\theta_1, \theta_2, \dots, \theta_m)^T$ — $m \times 1$ -матрица (т.е. вектор-столбец длины m), компоненты которой играют роль неизвестных параметров эконометрической модели, а $A(\Theta) = (a_1(\Theta), a_2(\Theta), \dots, a_n(\Theta))^T$ — $n \times 1$ -матрица (т.е. 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 $m \times n$ -матрица

$$\frac{\partial A(\Theta)}{\partial \Theta} = \begin{pmatrix} \frac{\partial a_1(\Theta)}{\partial \theta_1} & \frac{\partial a_2(\Theta)}{\partial \theta_1} & \dots & \frac{\partial a_n(\Theta)}{\partial \theta_1} \\ \frac{\partial a_1(\Theta)}{\partial \theta_2} & \frac{\partial a_2(\Theta)}{\partial \theta_2} & \dots & \frac{\partial a_n(\Theta)}{\partial \theta_2} \\ \dots & \dots & \dots & \dots \\ \frac{\partial a_1(\Theta)}{\partial \theta_m} & \frac{\partial a_2(\Theta)}{\partial \theta_m} & \dots & \frac{\partial a_n(\Theta)}{\partial \theta_m} \end{pmatrix}. \quad (\text{П2.16})$$

Матрицы вида (П2.16) используются, в частности, в качестве матриц преобразований («якобианов») в теории преобразований случайных величин (см., например, п. 4.4).

Рассмотрим важные для эконометрических приложений частные случаи функции $A(\Theta)$:

1⁰. $A(\Theta) = X\Theta$, где X — симметрическая матрица размерности $n \times m$.

Тогда $\frac{\partial A(\Theta)}{\partial \Theta} = \frac{\partial (X\Theta)}{\partial \Theta} = X$.

2⁰. $A(\Theta) = \Theta^T B \Theta$, где B — симметрическая квадратная матрица размерности $m \times m$.

Тогда $\frac{\partial A(\Theta)}{\partial \Theta} = \frac{\partial (\Theta^T B \Theta)}{\partial \Theta} = 2\Theta^T B$.

¹ Готовится русский перевод этой книги.

3⁰. $A(\Theta) = X^T \Theta$, где X — вектор-столбец длины m .
Тогда $\frac{\partial A(\Theta)}{\partial \Theta} = \frac{\partial (X^T \Theta)}{\partial \Theta} = X^T$.

П2.2. Основные числовые характеристики квадратной матрицы: определитель, след

Любой квадратной матрице A можно сопоставить некоторый набор ее числовых характеристик. В данном пункте остановимся на двух из них. Одна из таких характеристик называется *определителем* (*детерминантом*) матрицы A и обозначается $\det A$ или $|A|$, а другая определяется как *след матрицы* A и обозначается $\text{tr}(A)$.

Определитель (детерминант) $n \times n$ -матрицы A вычисляется по формуле

$$\det A = \sum_{j_1=1}^n \sum_{j_2=1}^n \dots \sum_{j_n=1}^n (-1)^{\nu(j_1, j_2, \dots, j_n)} a_{1j_1} a_{2j_2} \dots a_{nj_n}, \quad (\text{П2.17})$$

где суммирование ведется по всем возможным комбинациям *различных* столбцов (т.е. по всем возможным перестановкам вторых индексов), а $\nu(j_1, j_2, \dots, j_n)$ — это минимальное число инверсий (т.е. 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:

а) $n = 2$: $\det A = \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = a_{11} a_{22} - a_{12} a_{21}$;

б) $n = 3$: $\det A = \det \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$
 $= a_{11} a_{22} a_{33} + a_{13} a_{21} a_{32} + a_{12} a_{23} a_{31} - a_{13} a_{22} a_{31} - a_{11} a_{23} a_{32} - a_{12} a_{21} a_{33}$.

Приведем здесь важнейшие свойства определителя:

- 1) $\det(A B) = \det(B A) = \det A \cdot \det B$;
- 2) $\det(\lambda A) = \lambda^n \det A$ (λ — число, n — размерность матрицы A);
- 3) $\det[\text{diag}(a_{11}, a_{22}, \dots, a_{nn})] = a_{11} a_{22} \dots a_{nn}$;
- 4) $\det I_n = 1$;
- 5) $\det A^T = \det A$;

- 6) $\det A = 0$, если в матрице A есть две одинаковые строки (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 $A_i = (a_{i1}, a_{i2}, \dots, a_{im})$ — i -ую строку $n \times m$ 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 R исследуемого многомерного признака (см. п. 11.2).

Отметим, что если квадратная матрица имеет отличный от нуля определитель, то она называется невырожденной.

След квадратной $n \times n$ -матрицы A (обозначается $\text{tr } A$, от английского слова «trace») определяется как сумма ее диагональных элементов, т. е.

$$\text{tr } A = a_{11} + a_{22} + \dots + a_{nn}. \quad (\text{П2.18})$$

Основные свойства следа матрицы:

- (i) $\text{tr}(AB) = \text{tr}(BA)$;
- (ii) $\text{tr} I_n = n$;
- (iii) $\text{tr}(\lambda A) = \lambda \text{tr} A$, где λ — число;
- (iv) $\text{tr} A^T = \text{tr} A$;

- (v) $\text{tr}(A + B) = \text{tr}A + \text{tr}B$;
- (vi) в качестве частного случая свойства (i) особо отметим ситуации, в которых роль A играет вектор-столбец $X = (x_1, x_2, \dots, x_n)^T$, а роль B — X^T ; и хотя ни для X , ни для X^T след не определен, их произведения XX^T и $X^T X$ являются *квадратными*, соответственно, $n \times n$ - и 1×1 -матрицами и для них действует правило (i), т. е.

$$\text{tr}(XX^T) = \text{tr}(X^T X),$$

где, обращаем внимание читателя, матрица $X^T X = x_1^2 + x_2^2 + \dots + x_n^2$ — это число.

Заметим, что и определитель, и след матрицы имеют четкую вероятностную интерпретацию, например, когда в качестве A рассматривается *ковариационная матрица* Σ_ξ многомерной случайной величины $\xi = (\xi^{(1)}, \xi^{(2)}, \dots, \xi^{(p)})^T$. 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 (см. п. 2.6.6).

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* I_n , поскольку при умножении на 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 A — *квадратная невырожденная* (т. е. $\det A \neq 0$) матрица; тогда матрица A^{-1} называется *обратной*, если $AA^{-1} = A^{-1}A = I$.

Можно показать, что такое определение обратной матрицы $A^{-1} = (a_{ij}^{\text{обп}})$ приводит к следующей формуле для вычисления ее элементов $a_{ij}^{\text{обп}}$:

$$a_{ij}^{\text{обп}} = \frac{(-1)^{i+j} \det A_{ji}}{\det A}, \tag{П2.19}$$

где A_{ji} — как и прежде, матрица, получающаяся из матрицы A вычеркиванием из нее j -й строки и i -го столбца (т. е. 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 , т. е. 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;

$$2) \det A^{-1} = (\det A)^{-1};$$

$$3) (A^T)^{-1} = (A^{-1})^T;$$

$$4) (A^{-1})^{-1} = A;$$

$$5) (AB)^{-1} = B^{-1}A^{-1},$$

$$(ABC)^{-1} = C^{-1}B^{-1}A^{-1} \text{ и т. д.}$$

(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 $A^T A = AA^T = I$. Нетрудно также вывести, что определитель ортогональной матрицы всегда равен по абсолютной величине единице, т. е. $|\det A| = 1$. 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 $\det (AA^T) = \det A \det A^T = (\det A)^2 = \det I = 1$. 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 ее 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 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 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 $\text{ранг } A$) определяется 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, подразумевается, 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 $\text{ранг } A < 3$. И хотя большинство миноров 2-го порядка тоже равно нулю, все-таки существуют миноры этого порядка, отличные от нуля. А это значит, что $\text{ранг } A = 2$.

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 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 $A - \lambda I_n$ была вырожденной, т. е. необходимо потребовать, чтобы

$$\det(A - \lambda I_n) = 0. \quad (\text{П2.24})$$

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 A . Приходим к определению:

- **Характеристическими (собственными) числами $n \times n$ -матрицы 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

$$(A - \lambda_i I_n)X = 0_{n,1}. \quad (\text{П2.24}')$$

Существование нетривиального решения $X(i)$ этого уравнения обеспечивается равенством нулю определителя $\det(A - \lambda_i I_n)$. 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 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 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 A (будем для краткости их называть далее, соответственно, спо- и

сно-матрицами)¹.

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*, т. е.

$$X^T(i)X(j) = 0 \quad \text{при} \quad \lambda_i \neq \lambda_j.$$

Поскольку собственный вектор определяется с точностью до коэффициента пропорциональности, то можно пронормировать собственные векторы $X(1), X(2), \dots, X(n)$ так, что все они будут образовывать *ортонормированную систему*, т. е.

$$X^T(i)X(j) = \begin{cases} 1 & \text{при } i = j, \\ 0 & \text{при } i \neq j \end{cases} \quad \text{для всех } i, j = 1, 2, \dots, n.$$

Заметим, что $n \times n$ -матрица A , составленная из столбцов $X(i)$ ($i = 1, 2, \dots, n$), т. е. матрица

$$X = \left(X(1) \quad ; \quad X(2) \quad ; \quad \dots \quad ; \quad X(n) \right)$$

будет *ортогональной*, т. е. $X^T X = I_n$.

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, т. е.

$$X^T A X = \text{diag} (\lambda_1, \lambda_2, \dots, \lambda_n).$$

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*

¹ Симметрическая $n \times n$ -матрица A называется *положительно (неотрицательно) определенной*, если для любого ненулевого вектора $X = (x_1, x_2, \dots, x_n)^T$ 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. П 2.6.

определенная матрица размерности n , т. е. 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 (т. е. для вектора-столбца длины n)

$$X^T (B^T B) X = (B X)^T B X = Y^T Y = \sum_{i=1}^n y_i^2 \geq 0,$$

т. к. 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 (т. е. $\text{ранг } B = m$), 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 (т. к. для любого вектора-столбца X длины n $X^T I_n X = X^T X = \sum_{i=1}^n x_i^2 > 0$).

5) $\det A > 0$, т. е. определитель любой споматрицы 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, т. е.

$$\text{tr} A = \lambda_1 + \lambda_2 + \dots + \lambda_n.$$

где, напомним, A — споматрица размерности $n \times n$, а $\lambda_1, \lambda_2, \dots, \lambda_n$ — ее собственные числа.

7) Всякая споматрица A может быть представлена в виде

$$A = C C^T, \quad (\text{П2.25}^a)$$

где C некоторая невырожденная $(m \times m)$ -матрица.

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, т. е.

$$\begin{aligned} E x_i &= 0, \quad i = 1, 2, \dots, n; \\ D x_i &= E x_i^2 = \sigma^2, \quad i = 1, 2, \dots, n; \\ \text{cov}(x_i, x_j) &= E(x_i x_j) = 0 \quad \text{при } i \neq j, \end{aligned}$$

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

$$E Y = E(C^T X) = C^T E X = 0_{n,1}$$

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 симметрическая¹ матрица A , обладающая тем свойством, что она совпадает со своим квадратом, т. е.**

$$A = A^2. \quad (\text{П2.28})$$

Перечислим основные свойства идемпотентных матриц.

1) Любая целая положительная степень k идемпотентной матрицы A дает снова исходную матрицу A , т. е. $A^k = A$ (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.

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)

¹ Иногда требование симметричности не включают в определение идемпотентной матрицы.

будет приводить квадратичную форму $X^T A X$ относительно n независимых $(0, \sigma^2)$ -нормально распределенных случайных величин к виду

$$X^T A X \equiv y_1^2 + y_2^2 + \dots + y_{n-k}^2, \quad (П2.29)$$

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 $\sigma^2 \chi^2(n-k)$ -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 $\Sigma_{YY}, \Sigma_{YX}, \Sigma_{XY}$ и Σ_{XX} , а именно:

$$\Sigma_Z = \begin{pmatrix} \Sigma_{YY} & \vdots & \Sigma_{YX} \\ \dots\dots\dots & & \\ \Sigma_{XY} & \vdots & \Sigma_{XX} \end{pmatrix}. \quad (П2.30)$$

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* (т. е. размерности блоков A_{ij} и B_{ij} совпадают), то их сложение производится по правилу

$$A + B = \begin{pmatrix} A_{11} + B_{11} & \vdots & A_{12} + B_{12} & \vdots & \dots & \vdots & A_{1k_2} + B_{1k_2} \\ A_{21} + B_{21} & \vdots & A_{22} + B_{22} & \vdots & \dots & \vdots & A_{2k_2} + B_{2k_2} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ A_{k_11} + B_{k_12} & \vdots & A_{k_12} + A_{k_12} & \vdots & \dots & \vdots & A_{k_1k_2} + B_{k_1k_2} \end{pmatrix}.$$

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

$$A = \begin{pmatrix} A_{11} & \vdots & A_{12} \\ \dots & \dots & \dots \\ A_{21} & \vdots & A_{22} \end{pmatrix} \quad (\text{П2.31})$$

таким образом, что блоки A_{11} и A_{22} являются *квадратными* матрицами (как это было в нашем примере, см. (П2.30)). Тогда можно воспользоваться следующей общей формулой подсчета определителя матрицы A , во многих случаях существенно упрощающей вычисления:

$$\begin{aligned} \det \begin{pmatrix} A_{11} & \vdots & A_{12} \\ \dots & \dots & \dots \\ A_{21} & \vdots & A_{22} \end{pmatrix} &= \det A_{11} \cdot \det (A_{22} - A_{21}A_{11}^{-1}A_{12}) \\ &= \det A_{22} \cdot \det (A_{11} - A_{12}A_{22}^{-1}A_{21}). \end{aligned}$$

В частном случае *блочно-диагональной* матрицы A (т. е. 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 (П2.31) 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) $\text{tr} (A \otimes B) = \text{tr} A \cdot \text{tr} B.$

ЛИТЕРАТУРА

Абрамовиц М. Справочник по специальным функциям. /Пер. с англ./ — М.: Наука, 1979.

Абазян С. А. (1968). Статистическое исследование зависимостей. М.: Металлургия.

Абазян С. А. (1974). Об опыте применения экспертно-статистического метода построения неизвестной целевой функции. — В кн.: Многомерный статистический анализ в социально-экономических исследованиях. М.: Наука, С. 56–86.

Абазян С. А. (1979). Конфлюэнтный анализ. — В кн.: Математическая энциклопедия. М. Т. 2. С. 1083.

Абазян С. А. (1980). Вероятностно-статистическое моделирование распределительных отношений в обществе. — В кн.: Статистическое моделирование экономических процессов. Ученые записки по статистике. М., Статистика, Т.37.

Абазян С. А., Енюков И. С., Мешалкин Л. Д. Прикладная статистика. Исследование зависимостей. — М.: Финансы и статистика, 1985.

Абазян С. А., Бушитабер В. М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика. Классификация и снижение размерности. — М.: Финансы и статистика, 1989.

Андерсон Т. Введение в многомерный статистический анализ. М., Физматгиз, 1963.

Бокс Дж., Дженкинс Г. Анализ временных рядов. Прогноз и управление. /Пер. с англ./ — М.: Мир, 1974, вып. 1 и 2.

Большев Л. Н., Смирнов Н. В. Таблицы математической статистики. М.: Наука, 1965.

Боровков А. А. Курс теории вероятностей. М.: Наука, 1972.

Бушитабер В. М. Метод многомерной развертки в анализе временных рядов. Обзорные прикладной и промышленной математики. Т.4 (1997): № 4.

Вальд А. Последовательный анализ. /Пер. с англ./ — М.: Физматгиз, 1960.

Джонстон Дж. Эконометрические методы. /Пер. с англ./ — М.: Статистика, 1980.

Дидэ Э. и др. Методы анализа данных: подход, основанный на методе динамических сгущений. /Пер. с франц./ — М.: Финансы и статистика, 1985.

Езекиэл М., Фокс К. Методы анализа корреляций и регрессий. /Пер. с англ./ — М.: Статистика, 1966.

Кемени Дж., Снелл Дж. Конечные цепи Маркова. /Пер. с англ./ — М.: Наука, 1970.

Кендэл М. Временные ряды. /Пер. с англ./ — М.: Финансы и статистика, 1981.

Кендалл М. Дж., Стьюарт А. (1966). Теория распределений. /Пер. с англ./ — М.: Наука.

Кендалл М. Дж., Стьюарт А. (1973). Статистические выводы и связи. /Пер. с англ./ — М.: Наука.

Кендалл М. Дж., Стьюарт А. (1976). Многомерный статистический анализ и временные ряды. /Пер. с англ./ — М.: Наука.

Колмогоров А. Н. Основные понятия теории вероятностей. М. — Л., ОНТИ, 1936.

Крамер Г. Математические методы статистики. — 2-е изд. Пер. с англ. — М.: Мир, 1975.

Лумельский В. Я. Агрегирование объектов на основе квадратичной матрицы. Автоматика и телемеханика, 1970. № 1. С.133–143.

Магнус Я. Р., Катыйшев П. К., Пересецкий А. А. Эконометрика. Начальный курс. (3-е изд.) — М.: Дело, 2000

Майстров Л. Е. Развитие понятия вероятности. — М.: Наука, 1980.

Маленко Э. Статистические методы в эконометрии: Пер. с франц. — М.: Статистика, 1975, вып.1; 1976, вып.2.

Митоян А. А. Потребительское поведение семей: дифференциация, динамика, классификация. — М.: Экономика, 1990.

Феллер В. Введение в теорию вероятностей и ее приложения: Пер. с англ. — 2-е изд. М.: Мир, 1964.

Харман Г. Современный факторный анализ: Пер. с англ. — М.: Статистика, 1972. Мир, 1983.

Эфрон Б. Нетрадиционные методы многомерного статистического анализа: Пер. с англ. — М.: Финансы и статистика, 1988.

Anderson T. W., Rubin H. Statistical inferences in factor analysis. Proc. 3 Berkeley Symp. Math. Statist. and Probab. — Univ. Calif. Press, 1956. Pp. 11–50.

Berndt E. R. The practice of econometrics. Classic and contemporary. Addison-Wesley Publishing Company. Reading-Massachusetts-Menlo Parc-California, 1990.

Dougherty C. Introduction to econometrics. Oxford University Press. New York-Oxford, 1992.

Goldberger A. A course in Econometrics. Cambridge-Mass.: Harvard University Press, 1990.

Green W. H. Econometric analysis. Macmillan Publishing Company, New York, 1993.

Hayashi F. Econometrics. — Princeton University Press, Princeton and Oxford, 2000.

Horowitz J. Semiparametric Methods in Econometrics. — Springer-Verlag New York, Inc, 1998.

Johnston J. and Di Nardo J. Econometric Methods, 4th edition. — Mc Graw-Hill, 1997.

Jorgenson D. Welfare. Vol.1: Aggregate Consumer Behavior. Vol.2: Measuring Social Welfare. — MIT Press, Cambridge, Massachusetts, 1997.

Kennedy P. A Guide to Econometrics, 4th edition. — Blackwell Publishers, 1998.

Magnus J. R., Neudecker H. Matrix Differential Calculus with Applications in Statistics and Econometrics. New York, John Wiley, 1988.

Pindyck R., Rubinfeld D. L. Econometric models and economic forecasts. McGraw-Hill Kogakusha Ltd, Tokio, 1976.

Ruud P. An Introduction to Classical Econometric Theory. — Oxford University Press. New York-Oxford, 2000.

АЛФАВИТНО-ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

- А**
- Автоковариационная функция 209
 - Автокорреляционная функция 210
 - — марковского процесса 247
 - — процесса Юла 252–253
 - — частная 211
 - Автопрогноз 312
 - Авторегрессия 1-го порядка 111, 245
 - Адаптивные методы прогнозирования 320
 - Аддитивная линейная форма 21
 - Аддитивная модель сезонности 326
 - Аддитивное разложение временного ряда 203, 207
 - Алгебраическое дополнение 412
 - Алгебраический полином 287
 - АР (p)-модели 246–261
 - АРСС (p, q, \dots)-модели 287–291
 - АРСС (p, q)-модели 270
 - АРУГ (p)-модель 285
- Б**
- Белый шум 245
 - — стандартизированный нормальный 285
 - Бинарные (булевы, дихотомические) переменные 159
 - Блочно-диагональная матрица 426
 - Блочная структура матриц 357, 362–364, 424–426
 - Бутстреп-метод 377
- В**
- Вариация выборочной функции регрессии 59–61
 - Векторные АР-модели 283
 - Векторные АРСС-модели 282
 - Верификация модели 35, 200, 377
 - Временной ряд 199, 201
 - Временной такт 200
 - Выбор общего вида модели 374–380
 - Выборка — временная 47–48
 - пространственная 47–48
 - пространственно-временная 48
 - сгенерированная 374–378
- Г**
- Гармоника 219
 - Гармоническая составляющая 217
 - Генезис наблюдений 200, 202
 - Генерирование данных 374–378
 - Геометрическая лаговая структура 302
 - Гетероскедастичность 94, 102
 - Главные компоненты 79–83, 358
 - Гомоскедастичность 50, 108
 - Гребневая (ридж-) регрессия 78
- Д**
- Детерминант (определитель) 411
 - Доверительные интервалы (область) 70, 371
- И**
- Идентификация модели 35, 40
 - — авторегрессии 244, 250, 257
 - — скользящего среднего 265, 266, 268

- — APCC 276
- — АРПСС 290
- — СОУ 348–366

Идентифицируемости проблема 18, 34, 339

Имитационно-компьютерные эксперименты 374–380

Инструментальные переменные 144–147, 150, 153

Интегральный индикатор 38

Информационное обеспечение 18

Информационный этап исследования 31

Исходные статистические данные 47–48

К

Ковариационная матрица 51, 66, 110, 369 Ковариационная структура 279

Ковариационный анализ 159

Коинтеграция временных рядов 290

Корни характеристического уравнения 278

Коррелограмма 210

Корреляционный анализ 42

Коэффициент автокорреляции 210

— адаптации (сглаживания) 321

— детерминации 46, 59, 84, 100

— дисконтирования 321

— регрессии 71–73

— эластичности 137–138

— Дербина–Уотсона 117–118

Критерий качества аппроксимации (адекватности модели) 379

Критерий Чоу однородности двух групп наблюдений 167–169

Л

Лаговые переменные 25, 40

Лаговая структура 299

— — геометрическая (Койка) 302

— — нормированная 299

— — основанная на вероятностной параметризации 310

— — Паскаля 311

— — полиномиальная (Алмон) 300

Линеаризация 173

Линеаризующие преобразования 173–182

Линейная модель множественной регрессии 37, 49, 93, 193

Логит-модель 189

М

Макроуровень 19

«Манекены» (переменные) 158–159

Массив исходных статистических данных 47–48

Матрица 404

— квадратная 405

— идемпотентная 64, 423

— невырожденная 412

— неотрицательно определенная 419

— обратная 413

— полного ранга 420

— положительно определенная 419

— симметрическая 405

— транспонированная 406

Мезоуровень 19

Метод ветвей и границ 85

— всех возможных регрессий 84

— максимального правдоподобия 57–58

— Монте-Карло 374–377

— наименьшего дисперсионного отношения 364

— наименьших квадратов (МНК) 37, 54–57

— — — взвешенный 104

— — — двухшаговый (2 МНК) 355

— — — косвенный 353

— — — обобщенный 37, 97

— — — трехшаговый 360

— последовательных разностей 240

— пошагового отбора переменных (пошаговая регрессия) 86

Методы сглаживания 227

— — алгоритмические 228

— — аналитические 227

— — экспоненциально взвешенные 238

Микроуровень 19

МНК-оценки 54–56

Многомерный статистический анализ 374

Модели нестационарных временных рядов 287

— сезонных рядов 291, 293–297

Модель авторегрессии см. АР-модели

— — со скользящими средними в остатках, см. АРСС-модели

— адаптивных ожиданий 307

— бинарного выбора 187

— Бокса–Дженкинса, см. АРСС-модели

— гиперинфляции (Кагана) 308–309

— потребления (Фридмана) 308–309

— скользящего среднего, см. СС-модели

— частичного приспособления (или

частичной корректировки) 304

— эконометрическая 20, 25, 40

Мультиколлинеарность 74

Н

Невязки 54

Неидентифицируемость 34–35

Нелинейные модели регрессии 172

Неслучайная составляющая временного ряда 220

Нестационарный (однородный) временной ряд 287

Нормальных уравнений система 55, 356

Нормированная структура лага 299

О

ОАРУГ (p, q)-модель 286

Обобщенный метод наименьших квадратов (ОМНК) 97

Обращение матрицы 413

— — блочной 426

Объясняемая (результатирующая) переменная 43

Объясняющая переменная 43

Одномерный временной ряд 199, 201

ОМНК-оценки 97

Оператор авторегрессии 288

— сдвига по времени 277

— скользящего среднего 288

— упрощающий 292

Оптимальность МНК-оценок 67

Ошибки спецификации модели 89

П

ПАД-процедуры 378

Параметр (коэффициент) адаптации

321

Парная линейная регрессия 55
Передаточная функция системы 200
Перекрестный анализ дееспособности модели 378
Перепараметризации проблема 280
Подбор линеаризующего преобразования 173, 180
Последовательные разности 240
Предикторы в модели регрессии 43
Предопределенные переменные 25
Пробит-модель 189
Проверка однородности исходных данных 168
Прогноз долгосрочный 200, 312
— значений результирующего показателя 123, 132, 312, 366
— краткосрочный 200, 312
Производственные функции 172, 179
Прямое (кронекерово) произведение матриц 363, 409

Р

Ранг матрицы 415–416
Реалистическая ситуация (в анализе регрессионной модели) 132–134
Регрессионная зависимость 43–45
— неоднородность данных 156–157
Регрессионные модели с распределенными лагами 296
— остатки 45
Регрессионный анализ 42
Регрессия 43, 44
Результирующая переменная 43
Рекуррентная формула 253, 259, 326
Рекурсивные системы 349, 352
Ретроспективный прогноз 36, 315

Ридж-регрессия, см. Гребневая (ридж) регрессия
Ряд временной, см. Временной ряд

С

Сверхидентифицируемость 34, 345
Сглаживание 220–221
Сезонная составляющая 202
Симметрическая матрица 405
Система одновременных уравнений (СОУ) 331
— — — идентифицируемая 34, 334, 339
— — — неидентифицируемая 35, 334
— — — приведенная форма 25, 28, 29
— — — рекурсивная 349, 352
— — — структурная форма 27
След матрицы 412
Сно-матрицы 419
Собственное (характеристическое) значение (число) матрицы 418
Собственный (характеристический) вектор матрицы 418
Спектр (спектральная плотность) 216
Спектральный анализ временного ряда 217
Спецификация модели 18, 32
Спо-матрицы 418
Стохастическая зависимость 21
Стохастические объясняющие переменные 139–140
Структурная форма СОУ – см. Система одновременных уравнений
Структурные параметры (характеристики) модели 333, 339, 366, 375

Сценарные (многовариантные) расчеты 366, 381

Т

Тейла-Вейджа модель — см. Аддитивная модель сезонности

Теорема Гаусса-Маркова 68

Точность регрессионной модели 132

Транспонирование матрицы 406

— произведения матриц 408

Тренд временного ряда 202

У

Упрощающие операторы 292

Уравнение регрессионной связи 45

— характеристическое матрицы 418

— — модели авторегрессии 254-255, 258

— — скользящего среднего 263-264

Уравнения Юла-Уокера 260

Условия идентифицируемости уравнений (системы) 339-346

— обратимости временного ряда 263-264, 265, 267

— стационарности временного ряда 247, 253, 258

Условно гетероскедастичные остатки — см АРУГ-модели

Условное математическое ожидание 42

Ф

Факторы, формирующие значения временного ряда 202

Фиктивные переменные («манекены») 158-159

Функция регрессии 43-44

Х

Характеристические (собственные) векторы и числа — см. Собственные векторы и числа

Ц

Циклическая (конъюнктурная) составляющая временного ряда 202

Э

Экзаменуемая выборка 36, 378-379

Экзогенные переменные 25

Эконометрика 17-18

Эконометрическая модель 25-26

Экономическая статистика 18

— теория 18

Экспертные оценки 200, 313

Экспоненциальное сглаживание 238-240

Эндогенные переменные 25