Я.Р. Магнус
П.К. Катышев
А.А. Пересецкий

ЭКОНОМЕТРИКА

Начальный курс

Рекомендовано Министерством общего и профессионального образования Российской Федерации в качестве учебника для студентов высших учебных заведений, обучающихся по экономическим специальностям

Академия народного хозяйства
при Правительстве Российской Федерации

Москва
Издательство "ДЕЛО"
2004

УДК 330.43(075.8)
ББК 65в6я73
М13

Рецензенты:

Елисеева И.И., доктор экономических наук, профессор,
член-корреспондент РАН, заслуженный деятель науки РФ,
зав. кафедрой статистики и эконометрики
Санкт-Петербургского государственного университета экономики и финансов;

Суворов Б.П., доктор экономических наук, заслуженный деятель науки РФ,
профессор экономического факультета МГУ им. М.В. Ломоносова

Магнус Я.Р., Катышев П.К., Пересецкий А.А.
М13 Эконометрика. Начальный курс: Учеб. — 6-е изд., перераб. и доп. — М.: Дело, 2004. — 576 с.

ISBN 5-7749-0055-X

Учебник содержит систематическое изложение основ эконометрики и написан на основе лекций, которые авторы в течение ряда лет читали в Российской экономической школе и Высшей школе экономики. Подробно изучаются линейные регрессионные модели (метод наименьших квадратов, проверка гипотез, гетероскедастичность, автокорреляция ошибок, спецификация модели). Отдельные главы посвящены системам одновременных уравнений, методу максимального правдоподобия в моделях регрессии, моделям с дискретными и ограниченными зависимыми переменными.

В шестое издание книги добавлены три новые главы. Глава «Панельные данные» дополняет книгу до полного списка тем, традиционно включаемых в современные базисные курсы эконометрики. Добавлены также главы «Предварительное тестирование» и «Эконометрика финансовых рынков», которые будут полезны тем, кто интересуется соответственно теоретическими и прикладными аспектами эконометрики. Значительно увеличено количество упражнений. Включены упражнения с реальными данными, доступными для читателя на web-сайте книги.

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

УДК 330.43(075.8)
ББК 65в6я73

ISBN 5-7749-0055-X

© Я.Р. Магнус, П.К. Катышев, А.А. Пересецкий, 2000
© Издательство "Дело", оформление, 2004

Оглавление

Вступительное слово 10

Предисловие к первому изданию 13

Предисловие к третьему изданию 18

Предисловие к шестому изданию 23

1. Введение 26
- 1.1. Модели . . . 26
- 1.2. Типы моделей . . . 28
- 1.3. Типы данных . . . 30

2. Модель парной регрессии 32
- 2.1. Подгонка кривой . . . 32
- 2.2. Метод наименьших квадратов (МНК) . . . 34
- 2.3. Линейная регрессионная модель с двумя переменными . . . 38
- 2.4. Теорема Гаусса–Маркова. Оценка дисперсии ошибок σ^2 . . . 41
- 2.5. Статистические свойства МНК-оценок параметров регрессии. Проверка гипотезы $b = b_0$. Доверительные интервалы для коэффициентов регрессии 46
- 2.6. Анализ вариации зависимой переменной в регрессии. Коэффициент детерминации R^2 . . . 51

2.7. Оценка максимального правдоподобия коэффициентов регрессии . . . 55
Упражнения . . . 58

3. Модель множественной регрессии **67**
3.1. Основные гипотезы . . . 67
3.2. Метод наименьших квадратов. Теорема Гаусса–Маркова . . . 69
3.3. Статистические свойства МНК-оценок . . . 72
3.4. Анализ вариации зависимой переменной в регрессии. Коэффициенты R^2 и скорректированный R^2_{adj} 74
3.5. Проверка гипотез. Доверительные интервалы и доверительные области . . . 78
Упражнения . . . 88

4. Различные аспекты множественной регрессии **108**
4.1. Мультиколлинеарность . . . 109
4.2. Фиктивные переменные . . . 112
4.3. Частная корреляция . . . 118
4.4. Спецификация модели . . . 124
Упражнения . . . 135

5. Некоторые обобщения множественной регрессии **148**
5.1. Стохастические регрессоры . . . 149
5.2. Обобщенный метод наименьших квадратов . . . 154
5.3. Доступный обобщенный метод наименьших квадратов . . . 160
Упражнения . . . 163

6. Гетероскедастичность и корреляция по времени **167**
6.1. Гетероскедастичность . . . 168
6.2. Корреляция по времени . . . 184
Упражнения . . . 192

7. Прогнозирование в регрессионных моделях **204**
7.1. Безусловное прогнозирование . . . 205

7.2. Условное прогнозирование . . . 208
7.3. Прогнозирование при наличии авторегрессии ошибок . . . 209
Упражнения . . . 211

8. Инструментальные переменные **212**
8.1. Состоятельность оценок, полученных с помощью инструментальных переменных . . . 213
8.2. Влияние ошибок измерения . . . 214
8.3. Двухшаговый метод наименьших квадратов . . . 215
8.4. Тест Хаусмана . . . 217
Упражнения . . . 218

9. Системы регрессионных уравнений **220**
9.1. Внешне не связанные уравнения . . . 221
9.2. Системы одновременных уравнений . . . 224
Упражнения . . . 241

10. Метод максимального правдоподобия в моделях регрессии **244**
10.1. Введение . . . 245
10.2. Математический аппарат . . . 246
10.3. Оценка максимального правдоподобия параметров многомерного нормального распределения . . 248
10.4. Свойства оценок максимального правдоподобия . 249
10.5. Оценка максимального правдоподобия в линейной модели . . . 250
10.6. Проверка гипотез в линейной модели, I . . . 253
10.7. Проверка гипотез в линейной модели, II . . . 257
10.8. Нелинейные ограничения . . . 258
Упражнения . . . 260

11. Временные ряды **264**
11.1. Модели распределенных лагов . . . 266
11.2. Динамические модели . . . 268
11.3. Единичные корни и коинтеграция . . . 276

11.4. Модели Бокса–Дженкинса (ARIMA) 285
11.5. GARCH модели 311
Упражнения 315

12. Дискретные зависимые переменные и цензурированные выборки **31[illegible]**
12.1. Модели бинарного и множественного выбора . . . 321
12.2. Модели с урезанными и цензурированными выборками . . . . . 33[illegible]
Упражнения 350

13. Панельные данные **357**
13.1. Введение . . . . . 35[illegible]
13.2. Обозначения и основные модели . . . . . 36[illegible]
13.3. Модель с фиксированным эффектом 362
13.4. Модель со случайным эффектом 367
13.5. Качество подгонки . . . . . 37[illegible]
13.6. Выбор модели 375
13.7. Динамические модели . . . . . 37[illegible]
13.8. Модели бинарного выбора с панельными данными 38[illegible]
13.9. Обобщенный метод моментов . . . . . 38[illegible]
Упражнения 394

14. Предварительное тестирование: введение **39[illegible]**
14.1. Введение . . . . . 39[illegible]
14.2. Постановка задачи 400
14.3. Основной результат . . . . . 40[illegible]
14.4. *Pretest*-оценка . . . . . 40[illegible]
14.5. WALS-оценка 404
14.6. Теорема эквивалентности 405
14.7. Предварительное тестирование и эффект «занижения» 407
14.8. Эффект «занижения». Один вспомогательный параметр 412
14.9. Выбор модели: от общего к частному и от частного к общему 415

14.10. Эффект «занижения». Два вспомогательных параметра 419
14.11. Прогнозирование и предварительное тестирование 425
14.12. Обобщения 429
14.13. Другие вопросы 432
Упражнения 434

Эконометрика финансовых рынков **435**
15.1. Введение 436
15.2. Гипотеза эффективности финансового рынка . . . 438
15.3. Оптимизация портфеля ценных бумаг 446
15.4. Тест на включение новых активов в эффективный портфель 450
15.5. Оптимальный портфель при наличии безрискового актива 456
15.6. Модели оценки финансовых активов 461
Упражнения 471

Перспективы эконометрики **472**
16.1. Введение 472
16.2. Чем собственно занимается эконометрист? 473
16.3. Эконометрика и физика 474
16.4. Эконометрика и математическая статистика . . . 475
16.5. Теория и практика 476
16.6. Эконометрический метод 477
16.7. Слабое звено 480
16.8. Агрегирование 481
16.9. Как использовать другие работы 481
16.10. Заключение 482

[illegible]ложение ЛА. Линейная алгебра **484**
[illegible]1. Векторное пространство 484
[illegible]2. Векторное пространство R^n 485
[illegible]3. Линейная зависимость 485
[illegible]4. Линейное подпространство 486
[illegible]5. Базис. Размерность 486

6. Линейные операторы 487
7. Матрицы 488
8. Операции с матрицами 489
9. Инварианты матриц: след, определитель 492
10. Ранг матрицы 494
11. Обратная матрица 495
12. Системы линейных уравнений 496
13. Собственные числа и векторы 496
14. Симметричные матрицы 498
15. Положительно определенные матрицы 500
16. Идемпотентные матрицы 502
17. Блочные матрицы 503
18. Произведение Кронекера 504
19. Дифференцирование по векторному аргументу . . 505
Упражнения 507

Приложение МС. Теория вероятностей и математическая статистика **509**
1. Случайные величины, случайные векторы 509
2. Условные распределения 516
3. Некоторые специальные распределения 518
4. Многомерное нормальное распределение 524
5. Закон больших чисел. Центральная предельная теорема 528
6. Основные понятия и задачи математической статистики 531
7. Оценивание параметров 533
8. Проверка гипотез 539

Приложение ЭП. Обзор эконометрических пакетов **542**
1. Происхождение пакетов. Windows-версии. Графика 543
2. О некоторых пакетах 544
3. Опыт практической работы 546

Приложение СТ. Краткий англо-русский словарь терминов **547**

Приложение ТА. Таблицы **555**

Литература **561**

Предметный указатель **570**

Вступительное слово

Современное университетское экономическое образование держится на трех китах: макроэкономике, микроэкономике и эконометрике. После снятия идеологических барьеров с экономического образования в России — доминирования марксистской политической экономии — наши вузовские экономические программы стали перестраиваться в указанном направлении. Свидетельств этого много, в частности, появилось много учебников, в основном переводных, по макро- и микроэкономике. А вот с эконометрикой получилась промашка. Ни переводных, ни отечественных учебников нет. И почти нет преподавателей, способных читать курс эконометрики. Так что с третьим китом экономического образования пока в России сложно.

Причины в общем-то понятны. В централизованной плановой экономике эконометрика была не нужна, в отличие, скажем, от балансовых или оптимизационных методов: межотраслевого баланса и линейного программирования, по которым и книги, и специалисты у нас имелись и имеются.

Учебное пособие, которое вы, читатель, держите в руках, первое такого рода на русском языке. Оно написано по современным западным стандартам и апробировано на практике, на российских студентах. Авторы — квалифицированные специалисты в области эконометрики. Я. Магнус в течение многих лет читает курсы эконометрики в Центре экономических исследований Тилбургского университета (Голландия) и Лондонской школе экономики. П. Катышев и А. Пересецкий — математики, которые впервые столкнулись с эконометрикой, будучи ассистентами профессора Я. Магнуса в Российской экономической школе. Одновременно они являются ассистентами профессора С. А. Айвазяна, который читает там же курс прикладной статистики. В настоящее

время они читают курс лекций по эконометрике в РЭШ самостоятельно.

Почему эконометрика лежит в основе университетского экономического образования? Ответ на этот вопрос непрост. Надеюсь, читатель получит на него ответ, проработав книгу. Чем бо́льшим профессионалом становится экономист, тем яснее он понимает, что в экономике все зависит от всего. Причинно-следственными связями занимается экономическая теория, а связями вообще, без выявления их причин, — эконометрика. Поскольку экономика относится к тем областям человеческой деятельности, где интуиция, искусство не менее важны, чем наука, эконометрика нужна всем экономистам, ибо она сама есть смесь искусства и науки.

В. Л. Макаров,
директор Центрального
экономико-математического института РАН,
ректор РЭШ, академик РАН

Экономисты используют количественные данные для наблюдения за ходом развития экономики, ее анализа и прогнозов. Набор статистических методов, используемых для этих целей, называется в совокупности *эконометрикой*. Для успешного применения этих методов требуется точное (или хотя бы приблизительно верное) моделирование поведения экономических агентов, необходимо также понимание процессов, породивших имеющиеся данные, и насколько эти данные отражают те явления, которые мы пытаемся исследовать. Поскольку наши модели неполны, а данные несовершенны, значительная часть эконометрики посвящена методам, которые могли бы работать с такими моделями и данными. В конце концов, эконометрика является не более чем набором инструментов, хотя и очень полезных. Качество ингредиентов (моделей и данных) и то, как мы их используем, определяют результаты нашего анализа. Но и хорошие инструменты анализа также необходимы. Эконометрика является одновременно нашим

телескопом и нашим микроскопом для изучения окружающего экономического мира.

Книга, лежащая перед нами, является первым оригинальным учебником эконометрики, изданным на русском языке за последние десятилетия. Авторы преподавали этот материал пяти поколениям студентов Российской экономической школы и отрабатывали на них свои педагогические методики. Книга представляет собой прекрасное введение в основы эконометрики в том виде, в котором она преподается и используется во всем мире. Настало время для того, чтобы российские студенты и исследователи имели свой собственный учебник по основам этой науки. Надеюсь, что издание этой книги является только началом и в дальнейшем в России появятся как учебники, содержащие изложение дальнейших разделов эконометрики, так и высококачественные прикладные эконометрические работы, использующие методы, приведенные в данной книге.

Цви Грилихес,
профессор экономики имени Пола М. Варбурга
(Гарвардский Университет, Кембридж, США),
Президент Эконометрического общества (1975),
Президент Американской экономической ассоциации (1993)

Zvi Griliches,
Paul M. Warburg Professor of Economics,
Harvard University, Cambridge, USA,
President of Econometric Society (1975),
President of American Economic Association (1993)

От авторов

Цви Грилихес умер 4 ноября 1999 г. Будучи одним из выдающихся эконометристов нашего времени, он внес значительный вклад как в теорию, так и в приложения эконометрики. Цви стоял у истоков Российской экономической школы и до самого конца был активным и влиятельным членом Международного комитета советников РЭШ. Цви Грилихес сыграл существенную роль в развитии эконометрики в России и стал другом тех, кому выпало счастье работать с ним.

Предисловие к первому изданию

Эконометрика (наряду с микроэкономикой и макроэкономикой) входит в число базовых дисциплин современного экономического образования. Что же такое эконометрика? Когда имеешь дело с живой, развивающейся наукой, всегда возникает трудность при попытке дать краткое описание ее предмета и методов. Можно ли сказать, что эконометрика — это наука об экономических измерениях, как подсказывает само ее название? Конечно же можно, но тогда возникает вопрос, какой смысл вкладывать в термин «экономические измерения». Это аналогично тому, как если бы определить математику как науку о числах. Поэтому, не пытаясь более подробно развивать эту проблему, приведем высказывания признанных авторитетов в экономике и эконометрике.

«Эконометрика позволяет проводить количественный анализ реальных экономических явлений, основываясь на современном развитии теории и наблюдениях, связанных с методами получения выводов» (*Самуэльсон*).

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

«Цель эконометрики — эмпирический вывод экономических законов. Эконометрика дополняет теорию, используя реальные данные для проверки и уточнения постулируемых отношений» (*Маленво*).

Эта книга адресована прежде всего студентам, впервые приступающим к изучению эконометрики, и имеет две цели. Во-первых, мы хотим подготовить читателя к прикладным исследованиям в области экономики. Во-вторых, мы думаем, что она будет полезна студентам, которые собираются в дальнейшем углубленно изучать теорию эконометрики. Никаких предварительных знаний

об эконометрике не требуется. Однако предполагается знакомство с курсами линейной алгебры, теории вероятностей и математической статистики в начальном объеме (например, Гельфанд, 1971; Ильин, Позняк, 1984; Вентцель, 1964). Мы предполагаем также, что читатель владеет математическим анализом в пределах стандартного курса технического вуза.

Существует несколько прекрасных учебников по эконометрике на английском языке. Так, например, книгу (Greene, 1997) по праву можно считать «эконометрической энциклопедией» — в ней содержатся практически все разделы современной эконометрики. В учебнике (Goldberger, 1990) больше внимания уделяется формально-математической стороне эконометрики. Очень удачной, современной и сбалансированной с точки зрения теории и приложений является, на наш взгляд, книга (Johnston and DiNardo, 1997). Следует также отметить учебники (Griffits, Hill and Judge, 1993) и (Pindyck and Rubinfeld, 1991), ориентированные на читателей, не имеющих сильной математической подготовки, и снабженные большим количеством примеров и упражнений. Хорошим дополнением к стандартным учебникам может служить книга (Kennedy, 1998), где основной упор делается на содержательную сторону эконометрического анализа и которая содержит большое число интересных упражнений. Необходимо также упомянуть книгу (Hamilton, 1994), где очень подробно и на высоком математическом уровне изложена теория временны́х рядов, и книгу (Stewart, 1991), содержащую удачные и компактные разделы по теории временны́х рядов.

Поэтому, возможно, необходимо привести некоторые аргументы в пользу написания новой книги вместо простого перевода одного из существующих учебников. Наша книга основана на материале лекций, которые один из авторов (Я. Магнус) читал в качестве начального курса эконометрики по мастер-программе для студентов Российской экономической школы (РЭШ) в марте–апреле 1993 г. Два других автора (П. Катышев, А. Пересецкий) проводили практические занятия. Интенсивный 7-недельный курс включал в себя основы эконометрики. Это был первый год существования Российской экономической школы. В последующие

годы авторы сотрудничали в создании программы всех трех эконометрических курсов для студентов первого года обучения в РЭШ. В процессе работы мы, в частности, составили примеры из российской экономики, которые использовали вместо традиционно рассматриваемых примеров из экономики стран Западной Европы и США. В конце концов мы пришли к убеждению, что было бы желательно иметь учебник, написанный специально для российских студентов, и переработали программу курса в самостоятельную книгу. Настоящая книга является, таким образом, результатом пятилетнего опыта преподавания эконометрики для российских студентов.

Главы 2–4 содержат классическую теорию линейных регрессионных моделей. Этот материал является ядром эконометрики, и студенты должны хорошо освоить его перед тем, как перейти к изучению остальных частей книги. В главе 2 рассматривается простейшая модель с двумя регрессорами, глава 3 посвящена многомерным моделям. В определенном смысле глава 2 избыточна, однако с педагогической точки зрения крайне полезно изучить сначала регрессионные модели с двумя переменными. Тогда, например, можно обойтись без матричной алгебры, в двумерном случае легче также понять графическую интерпретацию регрессии. Глава 4 содержит несколько дополнительных разделов (проблема мультиколлинеарности, фиктивные переменные, спецификация модели), однако ее материал также можно отнести к стандартным основам эконометрики.

В главах 5–9[1] изучаются некоторые обобщения стандартной модели множественной регрессии, такие, как стохастические регрессоры, обобщенный метод наименьших квадратов, гетероскедастичность и автокорреляция остатков, доступный обобщенный метод наименьших квадратов, прогнозирование, метод инструментальных переменных. Удивительно в теории эконометрики то, что на этом уровне большинство теорем стандартного ядра теории (главы 2–4) остаются справедливыми, по крайней мере приближенно или асимптотически, когда условия теорем ослабляются.

[1] В данном издании материал главы 7 первого издания помещен в главу 5, и главы 8–13 первого издания имеют номера 7–12 соответственно.

Мы настоятельно рекомендуем постоянно соотносить результаты глав 5–9 с основными результатами, изложенными в главах 2–4.

Глава 10 содержит теорию систем одновременных уравнений, т. е. тот случай, когда модель содержит более одного уравнения. Рассматриваются проблемы, с которыми может встретиться эконометрист в практической работе.

В книгу включено несколько приложений, в том числе обзор эконометрических пакетов и краткий англо-русский словарь терминов.

Наш опыт показывает, что материала глав 1–7 достаточно для 7-недельного курса по 6 часов в неделю, а материала глав 1–10 — для стандартного односеместрового. Мы получали хорошие результаты со следующей структурой курса: две двухчасовые лекции в неделю и один семинар (в более малочисленных подгруппах), однако другие структуры курса также возможны.

Студентам

Решение задач — ключ к изучению математики, статистики, а также эконометрики. Об этом говорили нам наши учителя, когда мы были студентами, и мы повторяем это здесь. И это верно! Для студентов с ориентацией на практическую деятельность необходимы эксперименты с данными. Удалите несколько наблюдений из ваших данных и посмотрите, что произойдет с вашими оценками и почему. Добавьте объясняющие переменные и посмотрите, как изменятся ваши оценки и прогнозы. В общем, экспериментируйте. Студент, ориентированный на изучение теории, должен задавать себе вопрос, почему то или другое условие теоремы необходимо. Почему теорема перестает быть справедливой, если вы удаляете или изменяете одно из условий. Находите контрпримеры.

Преподавателям

Важно, чтобы все студенты обладали требуемым математическим и статистическим уровнем подготовки в начале курса. Если это не так, то курс следует начать с обзора необходимых понятий линейной алгебры и математической статистики. Главы 2–4 должны стоять в начале курса. Есть определенная свобода в выборе дальнейших тем, если время не позволяет включить в курс всю книгу. В случае недостатка времени можно отложить стохастические регрессоры (п. 5.1) и тесты на гетероскедастичность (но не саму концепцию гетероскедастичности) на следующий курс. Главы 7–10 содержат специальные, но важные разделы, которые могут быть включены в курс с той или иной степенью подробности, в зависимости от вкусов преподавателя.

Мы будем благодарны за любые замечания, сообщения об опечатках, неясных местах, ошибках в этой книге.

Благодарности

Мы в огромном долгу перед пятью поколениями студентов Российской экономической школы, которые в процессе изучения курса давали массу критических замечаний, использованных нами при работе над книгой. Без них эта книга никогда не была бы написана.

Мы благодарны выпускникам РЭШ Владиславу Каргину и Алексею Онацкому, которые подготовили для книги пример по рынку квартир в Москве, а также студенткам РЭШ Елене Пальцевой и Гаухар Турмухамбетовой, усилиями которых удалось избежать многих опечаток.

Мы также благодарны нашему коллеге Александру Сластникову, взявшему на себя труд редактирования рукописи.

В работе над рукописью П. Катышев и А. Пересецкий получали финансовую поддержку Российского гуманитарного научного фонда, проект 96-02-16011а.

Тилбург/Москва, март 1997 г.

P.S. Мы рады, что первая часть тиража (5000 экз.) быстро разошлась, следовательно, публикация была своевременной. Во второй части тиража исправлены некоторые обнаруженные неточности и опечатки. Мы благодарны С. А. Айвазяну за ряд ценных замечаний, в частности, он указал на то, что в русском языке более приняты термины «парная» и «множественная» регрессия, чем использованные в тексте «двумерная» и «многомерная» регрессия. Мы заранее благодарны всем читателям, которые сообщат нам свои замечания.

Тилбург/Москва, март 1998 г.

Предисловие к третьему изданию

Первое и второе издания нашей книги были выпущены общим тиражом 10 000 экз. и довольно быстро разошлись. Это послужило одной (но не главной) из причин написания третьего издания. Основным мотивом продолжения книги явилось желание расширить тематику начального курса эконометрики, включив те разделы, которые изучаются в магистерских программах большинства экономических вузов. Новыми являются глава 11[1] «Метод максимального правдоподобия в моделях регрессии», глава 12 «Временные ряды» и глава 13 «Дискретные зависимые переменные и цензурированные выборки».

Глава 11 содержит краткое описание общего метода максимального правдоподобия и достаточно подробно рассказывает об его использовании в моделях регрессии. Мы не ставили перед собой цель дать полное и систематическое изложение этого метода, который по традиции относится к теоретической и прикладной статистике. Более подробно о нем можно прочесть, например, в книгах (Рао, 1968), (Крамер, 1975), (Айвазян и др., 1983). В то же время мы выделили этот материал в отдельную главу, а не вынесли его в приложение по теории вероятностей и математической статистике, поскольку в двух последующих главах этот метод активно используется, и для удобства восприятия материала целесообразно прочесть о методе максимального правдоподобия непосредственно перед этими главами.

Глава 12 посвящена динамическим моделям и временны́м рядам. Эта очень обширная тематика и служит содержанием больших монографий (например, (Hamilton, 1994)). Мы рассматриваем простейшие модели с распределенными лагами и модели, в которых правые части регрессионных уравнений содержат значения зависимой переменной в предыдущие моменты времени. Значительное внимание уделяется проблемам стационарности и коинтеграции временны́х рядов. Дается изложение методологии Бокса–Дженкинса построения моделей временны́х рядов. Кратко описываются авторегрессионные условно гетероскедастичные модели (так называемые ARCH и GARCH модели), ставшие популярными в последнее время при описании финансовых рынков.

[1] В данном издании главы 11–13 третьего издания имеют номера 10–12.

В главе 13 изучаются модели, в которых есть априорные ограничения на значения зависимой переменной. Например, при изучении влияния каких-либо факторов на выбор из нескольких альтернатив зависимая переменная в соответствующей модели принимает дискретное множество значений. Ограничения на зависимые переменные возникают также при работе с цензурированными или усеченными выборками. Для подобных моделей метод наименьших квадратов не является адекватным инструментом оценивания и для построения оценок обычно используется метод максимального правдоподобия.

Мы стремились сохранить книгу как учебник по начальному курсу эконометрики. Наш опыт преподавания эконометрики в Тилбургском университете, в Российской экономической школе и в Высшей школе экономики позволяет нам рекомендовать новое издание книги в качестве учебника для годового (двухсеместрового) курса эконометрики как для бакалавров, так и для магистров (из расчета 2 часа лекций и 2 часа семинарских занятий в неделю). При этом для программы бакалавриата главы 11, 12, 13 могут быть опущены, а для магистров они могут составить основу программы второго семестра.

За время, прошедшее с момента появления первого издания нашей книги, в России были выпущены два учебника по эконометрике: С. А. Айвазян, В. М. Мхитарян «Прикладная статистика и основы эконометрики» (Айвазян, Мхитарян, 1998) и К. Доугерти «Введение в эконометрику» (Доугерти, 1997). Пер-

вая из этих книг охватывает очень широкий круг тем и является фундаментальным учебником по математической и прикладной статистике и основам эконометрики. Во второй книге изложение эконометрики дается на весьма простом уровне, часто недостаточном для программы бакалавриата.

В новом издании нашей книги мы стремились сохранить компактность изложения и в то же время его достаточно высокий математический уровень. Как и в первом издании, новые главы снабжены примерами из российской экономики. По нашему мнению, появление книг и учебников по эконометрике, ориентированных на разные группы студентов и специалистов, полезно для развития эконометрического образования в России.

Первое издание книги содержало около 50 упражнений. В дополнение к нему был выпущен сборник задач с решениями (Катышев, Пересецкий, 1999). В новом издании количество задач удвоилось, причем появились задачи, решение которых требует применения эконометрических компьютерных пакетов. В будущем мы планируем выпустить второе издание сборника задач с решениями.

Опыт преподавания эконометрики убеждает нас в необходимости освоения студентами современных эконометрических компьютерных пакетов. Решение практических задач и проведение небольших самостоятельных исследований, требующих работы с реальными данными, стимулирует интерес студентов к предмету и является, на наш взгляд, необходимой компонентой современного эконометрического образования.

Помимо добавления новых глав, были внесены некоторые изменения в содержание отдельных глав первого издания. В частности, ввиду быстрого развития эконометрического программного обеспечения информация, содержащаяся в приложении «Обзор эконометрических пакетов», в значительной мере устарела, и в новом издании мы решили его существенно сократить. Мы постарались также устранить выявленные за это время неточности и опечатки.

Обозначения

В третьем издании книги мы перешли в основном к новым обозначениям в формулах, ставшим за последние годы де-факто стандартом в международной эконометрической литературе — как в книгах, так и в научных статьях. Нам не удалось выдержать этот новый стиль во всей книге, и после долгого изучения литературы и мучительных обсуждений мы пришли к выводу, что, к сожалению, нет возможности ввести универсальную систему обозначений. Поэтому во второй главе используется «старая» система обозначений.

В сущности, единственная проблема при использовании новой системы состоит в отсутствии удобных обозначений для отклонений от средних значений.

Итак, в новой системе обозначений векторы, матрицы и их компоненты обозначаются следующим образом:

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nk} \end{bmatrix} \text{ — } n \times k \text{ матрица объясняющих переменных,}$$

а вектор отклонений от средних $\boldsymbol{y}_* = \begin{bmatrix} y_1 - \overline{y} \\ \vdots \\ y_n - \overline{y} \end{bmatrix}$.

Единственное место, где мы отступаем от этого правила, глава 2. В этой главе

$$\boldsymbol{y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} X_{11} & \cdots & X_{1k} \\ \vdots & & \vdots \\ X_{n1} & \cdots & X_{nk} \end{bmatrix}$$

и вектор отклонений от средних $\boldsymbol{y}_* = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} Y_1 - \overline{Y} \\ \vdots \\ Y_n - \overline{Y} \end{bmatrix}$.

Благодарности

Мы благодарим профессора Тилбургского университета Артура ван Суста за плодотворные обсуждения материалов третьего издания книги.

Мы благодарны профессору университета г. Левен (Бельгия) Марко Вербику за предоставленную возможность ознакомиться с рукописью его книги (Verbeek, 2000).

Выражаем благодарность профессору С. А. Айвазяну, взявшему на себя труд прочесть новые главы книги и сделавшему ряд существенных замечаний и пожеланий для улучшения качества книги.

Благодарим члена-корреспондента РАН И. И. Елисееву и профессора Б. П. Суворова за рецензирование данного издания книги.

Мы также благодарны нашему коллеге А. Д. Сластникову, осуществившему научное редактирование нового издания.

Мы искренне благодарим студентов Российской экономической школы, особенно Михаила Другова, которые своим кропотливым трудом помогли выявить неточности и опечатки в первоначальном тексте. Мы признательны также выпускнице Российской экономической школы Е. Е. Баян-оол, предоставившей пример исследования устойчивости российских банков.

Мы также благодарны преподавателям статистики и эконометрики различных университетов России — участникам региональных семинаров по преподаванию эконометрики — за полезные обсуждения новых глав книги.

Мы обязаны нашим коллегам Н. В. Третьякову, Е. В. Герасимовой, С. В. Голованю за перевод нового издания в систему TEX, причем помощь Н. В. Третьякова выходила далеко за рамки простого перенабора текста. Благодаря их усилиям новое издание книги выглядит более современно.

Мы благодарны Нидерландской организации научных исследований (NWO) за финансовую поддержку подготовки третьего издания нашей книги.

Наконец, мы считаем своим приятным долгом поблагодарить Центр экономических исследований Тилбургского университета (Нидерланды) за возможность стажировки и научных визитов, что в значительной степени способствовало появлению нового издания.

Тилбург/Москва, апрель 2000 г.

Предисловие к шестому изданию

Первые пять изданий данной книги были выпущены общим тиражом 38 000 экз. В 2001 г. эконометрика была введена в стандарт экономического образования, и возникла острая потребность в учебниках по эконометрике на русском языке.

В связи с этим в последнее время появилось несколько учебников и учебных пособий по эконометрике разного уровня, ориентированных на студентов с различной математической подготовкой (см., например, (Айвазян, 2002), (Елисеева, 2001a,b), (Бородич, 2001), (Доугерти, 2001), (Замков, 2001), (Кремер, Путко, 2002)). Наша книга содержит как достаточно строгое и полное изложение материала, удовлетворяющее студентов со склонностью к математике, так и значительное количество практических примеров и упражнений, предназначенных для тех студентов, которых больше интересуют прикладные аспекты корректного использования эконометрических методов, чем сама эконометрическая теория.

По сравнению с предыдущим изданием произошло некоторое изменение нумерации глав. Глава 7, посвященная доступному обобщенному методу наименьших квадратов, вставлена в качестве раздела 5.3 в главу 5, непосредственно после раздела, содержащего описание обобщенного метода наименьших квадратов. Соответственно, нумерация глав начиная с 8-й сдвинулась на единицу.

В связи с интенсивным расширением эконометрического образования в России изменяется понятие «начального курса эконометрики», все более приближаясь к современному состоянию эконометрики. В шестое издание книги добавлены три новые главы: 13-я «Панельные данные», 14-я «Предварительное тестирование» и 15-я «Эконометрика финансовых рынков».

Тема «Панельные данные» дополняет книгу до полного набора тем, необходимых для начального эконометрического образования. Эта тема особенно важна для экономических исследований в России, где временные ряды макроэкономических данных еще слишком короткие для применения методов анализа временных рядов.

Главы 14 и 15 (помечены *) не входят в стандартный набор тем, включаемых обычно в начальный курс эконометрики.

Материал главы 14 содержит введение в новую и интенсивно развивающуюся область эконометрики — теорию оценок, основанных на предварительном отборе моделей (*pretest-estimators*) и анализ чувствительности (*sensitivity analysis*), в которой активно работает один из авторов (Ян Магнус). Эта глава может использоваться в продвинутых курсах эконометрики для студентов с хорошей математической подготовкой, которые интересуются теоретическими и прикладными разделами эконометрики.

Глава 15 посвящена приложениям эконометрических методов к численному анализу финансовых рынков. На английском языке существует огромное количество литературы, посвященной теории и эконометрическому анализу финансовых рынков (см., например, (Cohrane, 2001), (Campbell, Lo and MacKinlay, 1997), (LeRoy and Werner, 2001), (Luenberger, 1998), (Mills, 1999), (Gourieroux, 1997)). Также издано несколько книг на русском языке, из которых можно отметить книгу (Малюгин, 2003). Данная глава может использоваться в продвинутых курсах эконометрики, ориентированных на студентов, интересующихся прикладным эконометрическим анализом финансовых рынков.

В шестое издание книги добавлены новые упражнения. Особенностью этого издания является наличие компьютерных упражнений, данные для которых доступны в Интернете по адресу `http://econometrics.nes.ru/mkp/`.

Благодарности

Мы благодарны профессорам Тилбургского университета Артуру ван Сусту (Arthur van Soest), Тео Нейману (Theo Nijman) и Бертрану Меленбергу (Bertrand Melenberg) за их лекции на выездных семинарах по эконометрике и обсуждения материалов новых 14-й и 15-й глав книги. Мы благодарны Дмитрию Данилову (Tilburg University) за помощь в подготовке главы 14. Мы также благодарны профессорам Эрику Гисельсу (Eric Ghysels, University of North Carolina), Франсу де Рону (Frans DeRoon, Tilburg University), Юрию Кабанову (Université de Besançon) и Владимиру Малюгину (Белорусский государственный университет) за ценные обсуждения материала главы 15.

Мы искренне благодарим студентов Российской экономической школы, которые своим кропотливым трудом помогли выявить неточности и опечатки в первоначальном тексте.

Мы обязаны преподавателю РЭШ Сергею Голованю за помощь в подготовке данного издания. С. В. Головань также добавил новые упражнения и подготовил все компьютерные упражнения. Он является и одним из авторов нового издания задачника по эконометрике с решениями задач из данного учебника.

Мы благодарны Нидерландской организации научных исследований (NWO) за финансовую поддержку подготовки шестого издания нашей книги.

Наконец, мы считаем своим приятным долгом поблагодарить Центр экономических исследований Тилбургского университета (Нидерланды) за возможность стажировки и научных визитов, помощь в организации выездных семинаров по эконометрике[1], что в значительной степени способствовало появлению нового издания.

Тилбург/Москва, июнь 2003 г.

[1]Подробную информацию о программе семинаров по преподаванию эконометрики, которой руководит А. Пересецкий, можно найти на сайте РЭШ `http://www.nes.ru/russian/outreach/workshops/econometrics.htm`

Глава 1

Введение

1.1. Модели

Эконометрика как наука расположена где-то между экономикой, статистикой и математикой. Один из ответов на вопрос, что такое эконометрика, может звучать так: это наука, связанная с эмпирическим выводом экономических законов. То есть мы используем данные или «наблюдения» для того, чтобы получить количественные зависимости для экономических соотношений. Данные, как правило, не являются экспериментальными, так как в экономике мы не можем проводить (многократные) эксперименты.

Но это — только малая часть работы эконометриста. Он также формулирует экономические модели, основываясь на экономической теории или на эмпирических данных, оценивает неизвестные величины (параметры) в этих моделях, делает прогнозы (и оценивает их точность) и дает рекомендации по экономической политике.

Во всей этой деятельности существенным является использование моделей. Модели должны быть «настолько простыми, насколько возможно, но не проще», сказал Эйнштейн. В большинстве случаев экономические законы выражаются в относительно простой математической форме. Рассмотрим, например, функцию

потребления

$$\ln C = \beta_0 + \beta_1 \ln Y + \beta_2 \ln P,$$

где C — потребление некоторого пищевого продукта на душу населения в некотором году, Y — реальный доход на душу населения в этом году, а P — индекс цен на этот продукт, скорректированный (дефлированный) на общий индекс стоимости жизни; β_0, β_1, β_2 — константы. Это уравнение называется уравнением поведения (*behavioural equation*). Оно описывает (в среднем) поведение потребителя по отношению к покупке данного пищевого продукта в зависимости от относительного уровня цен на продукт и реального душевого дохода. Закон поведения будет определен, как только мы найдем значения коэффициентов β_0, β_1, β_2. Соответственно задача эконометрики — определить (оценить) эти коэффициенты из подходящего набора наблюдений. Но это не единственная задача. Можно задать много других вопросов, также относящихся к эконометрике, например:

Нет ли переменных, которые следовало бы дополнительно включить в уравнение (например, цены на непродовольственные товары)?

Не следует ли исключить из уравнения некоторые переменные?

Насколько корректно измерены наши данные, представляют ли они то, что должны представлять, по нашему мнению?

Верно ли, что модель линейна? Верна ли экономическая теория?

Является ли модель полной? (В данном примере мы имеем дело с уравнением спроса и не принимаем во внимание уравнение предложения. Что произойдет, если мы будем изучать спрос и предложение одновременно?)

Достаточно ли изучать макроэкономическое уравнение, подобно приведенному выше, для ответа на интересующие нас вопросы, или необходимо изучать также индивидуальные (микро) данные?

Приведенная выше модель является статической. Возможно, более подходящей была бы динамическая модель. Например,

можно предположить, что прошлогодний доход может влиять на текущий уровень потребления. В этом случае мы должны также включить его в уравнение.

Эконометрика рассматривает все эти вопросы, и в последующих главах мы опишем способы решения поставленных проблем. Не следует предполагать, что после изучения книги читатель сразу же станет опытным эконометристом. Этому есть две причины. Во-первых, имеется много технического (теоретического) материала, не включенного в данную книгу. Во-вторых, даже если два эконометриста обладают одинаковыми теоретическими познаниями, оценки и прогнозы, полученные ими, будут различаться. Это происходит оттого, что эконометрика представляет собой нечто большее, чем только применение теоретических знаний. Она требует прочных экономических знаний и определенного скептицизма по поводу значимости «теорем» в практических приложениях.

1.2. Типы моделей

Математические модели широко применяются в бизнесе, экономике, общественных науках, исследовании экономической активности и даже в исследовании политических процессов.

Математические модели полезны для более полного понимания сущности происходящих процессов, их анализа. Модель, построенная и верифицированная на основе (уже имеющихся) наблюденных значений объясняющих переменных, может быть использована для прогноза значений зависимой переменной в будущем или для других наборов значений объясняющих переменных.

Можно выделить три основных класса моделей, которые применяются для анализа и/или прогноза.

Модели временных рядов

К этому классу относятся модели:

тренда: $y(t) = T(t) + \varepsilon_t$,

где $T(t)$ — временной тренд заданного параметрического вида (например, линейный $T(t) = a + bt$), ε_t — случайная (стохастическая) компонента;

сезонности: $y(t) = S(t) + \varepsilon_t$,

где $S(t)$ — периодическая (сезонная) компонента, ε_t — случайная (стохастическая) компонента;

тренда и сезонности: $y(t) = T(t) + S(t) + \varepsilon_t$ (аддитивная) или
$y(t) = T(t)S(t) + \varepsilon_t$ (мультипликативная),

где $T(t)$ — временной тренд заданного параметрического вида, $S(t)$ — периодическая (сезонная) компонента, ε_t — случайная (стохастическая) компонента.

К моделям временных рядов относится множество более сложных моделей, таких, как модели адаптивного прогноза, модели авторегрессии и скользящего среднего (ARIMA) и др. Их общей чертой является то, что они объясняют поведение временного ряда, исходя только из его предыдущих значений. Такие модели могут применяться, например, для изучения и прогнозирования объема продаж авиабилетов, спроса на мороженое, краткосрочного прогноза процентных ставок и т. п.

Регрессионные модели с одним уравнением

В таких моделях зависимая (объясняемая) переменная y представляется в виде функции $f(x, \beta) = f(x_1, \ldots, x_k, \beta_1, \ldots, \beta_p)$, где $x_1, \ldots, x_k$ — независимые (объясняющие) переменные, а $\beta_1, \ldots, \beta_p$ — параметры. В зависимости от вида функции $f(x, \beta)$ модели делятся на линейные и нелинейные. Например, можно исследовать спрос на мороженое как функцию от времени, температуры воздуха, среднего уровня доходов или зависимость зарплаты от возраста, пола, уровня образования, стажа работы и т. п.

Область применения таких моделей, даже линейных, значительно шире, чем моделей временных рядов. Проблемам теории оценивания, верификации, отбора значимых параметров и другим посвящен огромный объем литературы. Эта тема является, пожалуй, стержневой в эконометрике и основной в данном курсе.

Системы одновременных уравнений

Эти модели описываются системами уравнений. Системы могут состоять из тождеств и регрессионных уравнений, каждое из которых может, кроме объясняющих переменных, включать в себя также объясняемые переменные из других уравнений системы. Таким образом, мы имеем здесь набор объясняемых переменных, связанных через уравнения системы. Примером может служить модель спроса и предложения, приведенная ниже. Системы одновременных уравнений требуют относительно более сложный математический аппарат. Они могут использоваться для моделей страновой экономики и др.

Пример. ***Модель спроса и предложения.*** Пусть Q_t^D — спрос на товар в момент времени t (*demand*), Q_t^S — предложение товара в момент времени t (*supply*), P_t — цена товара в момент времени t (*price level*), Y_t — доход в момент времени t (*income*). Составим следующую систему уравнений «спрос–предложение»:

$$Q_t^S = \alpha_1 + \alpha_2 P_t + \alpha_3 P_{t-1} + \varepsilon_t \qquad \text{(предложение)},$$

$$Q_t^D = \beta_1 + \beta_2 P_t + \beta_3 Y_t + u_t \qquad \text{(спрос)},$$

$$Q_t^S = Q_t^D \qquad \text{(равновесие)}.$$

Цена товара P_t и спрос на товар $Q_t = Q_t^D = Q_t^S$ определяются из уравнений модели, т. е. являются эндогенными переменными. Предопределенными переменными в данной модели являются доход Y_t и значение цены товара в предыдущий момент времени P_{t-1}.

1.3. Типы данных

При моделировании экономических процессов мы встречаемся с двумя типами данных:

пространственные данные (*cross-sectional data*) и

временные ряды (*time-series data*).

Примером пространственных данных является, например, набор сведений (объем производства, количество работников, доход

и др.) по разным фирмам в один и тот же момент времени (пространственный срез). Другим примером могут являться данные по курсам покупки/продажи наличной валюты в какой-то день по обменным пунктам в Москве.

Примерами временных данных могут быть ежеквартальные данные по инфляции, средней заработной плате, национальному доходу, денежной эмиссии за последние годы или, например, ежедневный курс доллара США на ММВБ, цены фьючерсных контрактов на поставку доллара США (МТБ) и котировки ГКО (ММВБ) за два последних года.

Отличительной чертой временных данных является то, что они естественным образом упорядочены по времени, кроме того, наблюдения в близкие моменты времени часто бывают зависимыми.

Глава 2

Модель парной регрессии

2.1. Подгонка кривой

Пусть у нас есть набор значений двух переменных X_t, Y_t, $t = 1, \ldots, n$; можно отобразить пары (X_t, Y_t) точками на плоскости X–Y (рис. 2.1).

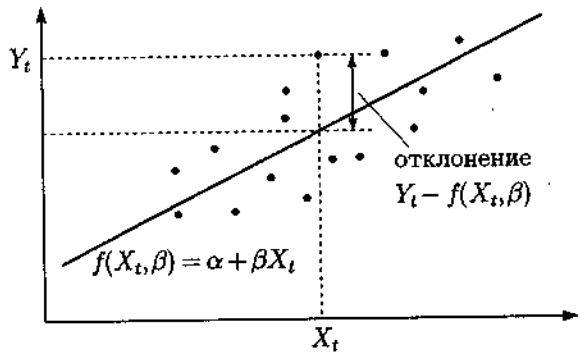


Рис. 2.1

Предположим, что нашей задачей является подобрать («подогнать») функцию $Y = f(X)$ из параметрического семейства функций $f(X, \beta)$, «наилучшим» способом описывающую зависимость Y от X. Подобрать функцию в данном случае означает

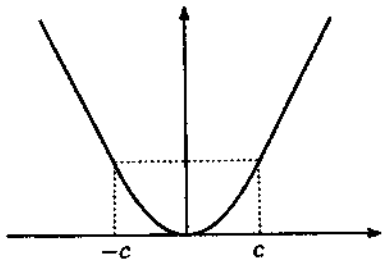
выбрать «наилучшее» значение параметра β. (Примером параметрического семейства может служить семейство линейных функций $f(X, \beta) = \alpha + \beta X$.)

В качестве меры отклонения функции $f(X, \beta)$ от набора наблюдений можно взять:

1) сумму квадратов отклонений $F = \sum\limits_{t=1}^{n} (Y_t - f(X_t, \beta))^2$,
2) сумму модулей отклонений $F = \sum\limits_{t=1}^{n} |Y_t - f(X_t, \beta)|$, или, в общем случае,
3) $F = \sum\limits_{t=1}^{n} g(Y_t - f(X_t, \beta))$, где g — «мера», с которой отклонение $Y_t - f(X_t, \beta)$ входит в функционал F.

Примером такой «меры» может служить функция Хубера, которая при малых отклонениях квадратична, а при больших линейна:

$$g(x) = \begin{cases} x^2, & |x| < c, \\ 2cx - c^2, & x \geqslant c, \\ -2cx - c^2, & x \leqslant -c. \end{cases}$$



Рассмотрим достоинства и недостатки перечисленных функционалов.

Сумма квадратов отклонений

Плюсы метода:

- легкость вычислительной процедуры;
- хорошие статистические свойства, простота математических выводов делают возможным построить развитую теорию, позволяющую провести тщательную проверку различных статистических гипотез;

минусы метода:

– чувствительность к «выбросам» (*outliers*).

Сумма модулей отклонений

Плюсы метода:

– робастность, т. е. нечувствительность к выбросам;

минусы метода:

– сложность вычислительной процедуры;

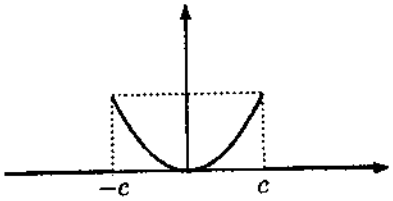
– возможно, большим отклонениям надо придавать больший вес (лучше два отклонения величиной 1, чем два отклонения величиной 0 и 2);

– неоднозначность, т. е. разным значениям параметра β могут соответствовать одинаковые суммы модулей отклонений (см. упражнение 2.11).

Функция Хубера является попыткой совместить достоинства двух первых функционалов.

Вопрос. Что будет, если взять в качестве функционала

$$F = \sum_{t=1}^{n} g(Y_t - f(X_t, \beta)),$$



где $g(X) = X^2$ для $|X| < c$, и $g(X) = 0$ для $|X| \geqslant c$? (Pindyck, Rubinfeld, 1991, п. 1.1, рис. 1.3b, стр. 6).

2.2. Метод наименьших квадратов (МНК)

Рассмотрим задачу «наилучшей» аппроксимации набора наблюдений X_t, Y_t, $t = 1, \ldots, n$, линейной функцией $f(X) = a + bX$

в смысле минимизации функционала

$$F = \sum_{t=1}^{n} \big(Y_t - (a + bX_t)\big)^2. \tag{2.1}$$

Запишем необходимые условия экстремума (*First Order Conditition, FOC*):

$$\frac{\partial F}{\partial a} = -2\sum_{t=1}^{n}(Y_t - a - bX_t) = 0, \quad \frac{\partial F}{\partial b} = -2\sum_{t=1}^{n} X_t(Y_t - a - bX_t) = 0,$$

или

$$\sum_{t=1}^{n}(Y_t - a - bX_t) = 0, \quad \sum_{t=1}^{n} X_t(Y_t - a - bX_t) = 0. \tag{2.2}$$

Раскроем скобки и получим *стандартную форму нормальных уравнений* (для краткости опустим индексы суммирования у знака суммы $\sum$):

$$an + b\sum X_t = \sum Y_t, \quad a\sum X_t + b\sum X_t^2 = \sum X_t Y_t. \tag{2.3}$$

Решения $\widehat{a}$, $\widehat{b}$ системы (2.3) можно легко найти:

$$\widehat{b} = \frac{n\sum X_t Y_t - \left(\sum X_t\right)\left(\sum Y_t\right)}{n\sum X_t^2 - \left(\sum X_t\right)^2} = \frac{\mathrm{Cov}(X, Y)}{\mathrm{Var}(X)}, \tag{2.4a}$$

$$\widehat{a} = \frac{1}{n}\sum Y_t - \frac{1}{n}\sum X_t \widehat{b} = \overline{Y} - \overline{X}\,\widehat{b}. \tag{2.4б}$$

Замечание 1. Из первого уравнения системы (2.3) следует

$$\overline{Y} = \widehat{a} + \widehat{b}\,\overline{X}, \tag{2.5}$$

т. е. уравнение прямой линии $Y = \widehat{a} + \widehat{b}X$, полученное в результате минимизации функционала (2.1), проходит через точку $(\overline{X}, \overline{Y})$. Здесь через $\overline{X}$ и $\overline{Y}$ обозначены выборочные средние значения переменных X_t и Y_t: $\overline{X} = (1/n)\sum X_t$, $\overline{Y} = (1/n)\sum Y_t$.

Замечание 2. Мы предполагаем здесь, что среди X_t, $t = 1, \ldots, n$, не все числа одинаковые, т. е. $\mathrm{Var}(X) \neq 0$ и (2.4a) имеет смысл.

Уравнения в отклонениях

Обозначим через $x_t = X_t - \overline{X}$, $y_t = Y_t - \overline{Y}$ отклонения от средних по выборке значений X_t и Y_t, $\overline{X} = (1/n)\sum X_t$, $\overline{Y} = (1/n)\sum Y_t$. (*Проверьте, что* $\overline{x} = \overline{y} = 0$.)

Решим теперь ту же задачу: подобрать линейную функцию $f(x) = a + bx$, минимизирующую функционал

$$F = \sum_{t=1}^{n}(y_t - (a + bx_t))^2.$$

Из геометрических соображений ясно, что решением задачи будет та же прямая на плоскости (x, y), что и для исходных данных X_t, Y_t. В самом деле, в силу (2.5) переход от X, Y к отклонениям x, y означает лишь перенос начала координат в точку $(\overline{X}, \overline{Y})$. Вычисления, которые необходимо проделать для решения задачи, вполне аналогичны предыдущим (с заменой X, Y на x, y). Заменив в (2.4а), (2.4б) X_t, Y_t на x_t, y_t и учитывая, что $\overline{x} = \overline{y} = (1/n)\sum x_t = (1/n)\sum y_t = 0$, получим

$$\widehat{a} = 0, \quad \widehat{b} = \frac{\sum x_t y_t}{\sum x_t^2} = \frac{\sum (X_t - \overline{X})(Y_t - \overline{Y})}{\sum (X_t - \overline{X})^2}. \tag{2.6}$$

Таким образом, мы получили другое выражение для углового коэффициента прямой $\widehat{b}$ (ср. (2.4а)).

Геометрическая интерпретация

Рассмотрим n-мерное векторное пространство R^n, снабженное стандартным евклидовым скалярным произведением: $(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}'\boldsymbol{y} = \sum X_t Y_t$, где $\boldsymbol{x}'$ — транспонированная матрица, т. е. в данном случае $1 \times n$ вектор-строка. Пусть

$$\boldsymbol{x} = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \quad \boldsymbol{\imath} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \boldsymbol{e} = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix},$$

$$\widehat{\boldsymbol{y}} = a\boldsymbol{\imath} + b\boldsymbol{x}, \quad \boldsymbol{e} = \boldsymbol{y} - \widehat{\boldsymbol{y}},$$

где a, b — числовые коэффициенты, $\widehat{\boldsymbol{y}}$ — вектор, лежащий в двумерной гиперплоскости π, натянутой на векторы $\boldsymbol{\imath}$, $\boldsymbol{x}$. (Здесь мы

снова предполагаем, что векторы $\boldsymbol{\imath}$ и $\boldsymbol{x}$ неколлинеарны; ср. *Замечание* 2, стр. 35.) Поставим задачу: найти такие a, b, чтобы вектор $\boldsymbol{e}$ имел наименьшую длину. (Другими словами, мы хотим наилучшим образом аппроксимировать вектор $\boldsymbol{y}$ вектором $\widehat{\boldsymbol{y}}$, лежащим в подпространстве π.) Очевидно, решением является такой вектор $\widehat{\boldsymbol{y}}$, для которого вектор $\boldsymbol{e}$ перпендикулярен плоскости π. Для этого необходимо и достаточно, чтобы вектор $\boldsymbol{e}$ был ортогонален векторам $\boldsymbol{\imath}$ и $\boldsymbol{x}$, порождающим плоскость π:

$$\begin{aligned} \boldsymbol{\imath}'\boldsymbol{e} = 0 &\iff \sum e_t = 0 &&\iff \sum(Y_t - a - bX_t) = 0, \\ \boldsymbol{x}'\boldsymbol{e} = 0 &\iff \sum X_t e_t = 0 &&\iff \sum X_t(Y_t - a - bX_t) = 0. \end{aligned} \tag{2.7}$$

Нетрудно заметить, что мы опять получили необходимые условия экстремума (2.2).

Матричная форма записи

Обозначим теперь через $\boldsymbol{X}$ матрицу размерности $n \times 2$

$$\boldsymbol{X} = \begin{bmatrix} 1 & X_1 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} a \\ b \end{bmatrix}$$

— 2×1 матрица (вектор) коэффициентов, $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X\beta}$, условие (2.7) ортогональности вектора $\boldsymbol{e}$ плоскости π теперь записывается как $\boldsymbol{X}'\boldsymbol{e} = \boldsymbol{0}$, или $\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X\beta}) = \boldsymbol{X}'\boldsymbol{y} - \boldsymbol{X}'\boldsymbol{X\beta} = \boldsymbol{0}$. Отсюда получаем $\boldsymbol{X}'\boldsymbol{X\beta} = \boldsymbol{X}'\boldsymbol{y}$, или

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}, \tag{2.8}$$

в предположении, конечно, что векторы $\boldsymbol{\imath}$, $\boldsymbol{x}$ линейно независимы и, следовательно, матрица $\boldsymbol{X}'\boldsymbol{X}$ обратима.

Нетрудно проверить, что (2.8) совпадает с (2.4а), (2.4б):

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = \begin{bmatrix} n & \sum X_t \\ \sum X_t & \sum X_t^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum Y_t \\ \sum X_t Y_t \end{bmatrix} = \cdots = \begin{bmatrix} \widehat{a} \\ \widehat{b} \end{bmatrix}.$$

Отметим, что матрица $\boldsymbol{X}'\boldsymbol{X}$ невырождена, так как матрица $\boldsymbol{X}$ имеет максимальный ранг 2 (см. *Замечание* 2, стр. 35, см. приложение ЛА, п. 10).

2.3. Линейная регрессионная модель с двумя переменными

В предыдущем разделе нас интересовало только качество подгонки кривой. Теперь добавим к постановке задачи некоторые статистические свойства данных.

На самом деле, для одного X мы можем наблюдать разные значения Y.

Пример 1. X — возраст индивидуума, Y — его зарплата.

Пример 2. X — доход семьи, Y — расходы на питание.

Запишем уравнение зависимости Y_t от X_t в виде

$$Y_t = a + bX_t + \varepsilon_t, \quad t = 1, \ldots, n,$$

где X_t — неслучайная (детерминированная) величина, а Y_t, ε_t — случайные величины. Y_t называется объясняемой (зависимой) переменной, а X_t — объясняющей (независимой) переменной или *регрессором*. Уравнение, приведенное выше, также называется *регрессионным уравнением*.

Какова природа ошибки ε_t?

Есть две основные возможные причины случайности:

а) Наша модель является упрощением действительности и на самом деле есть еще другие параметры (пропущенные переменные, *omitted variables*), от которых зависит Y. Зарплата, например, может зависеть от уровня образования, стажа работы, пола, типа фирмы (государственная, частная) и т. п.

б) Трудности в измерении данных (присутствуют ошибки измерений). Например, данные по расходам семьи на питание составляются на основании записей участников опросов, которые, как предполагается, тщательно фиксируют свои ежедневные расходы. Разумеется, при этом возможны ошибки.

Таким образом, можно считать, что ε_t — случайная величина с некоторой функцией распределения, которой соответствует функция распределения случайной величины Y_t.

Основные гипотезы:

1. $Y_t = a + bX_t + \varepsilon_t$, $t = 1, \ldots, n$, — спецификация модели.

2. X_t — детерминированная величина; вектор $(X_1, \ldots, X_n)'$ не коллинеарен вектору $\boldsymbol{\imath} = (1, \ldots, 1)'$.

3a. $\mathrm{E}\varepsilon_t = 0$, $\mathrm{E}(\varepsilon_t^2) = \mathrm{V}(\varepsilon_t) = \sigma^2$ — не зависит от t.

3b. $\mathrm{E}(\varepsilon_t \varepsilon_s) = 0$ при $t \neq s$, некоррелированность ошибок для разных наблюдений.

Часто добавляется условие:

3c. Ошибки ε_t, $t = 1, \ldots, n$, имеют совместное нормальное распределение: $\varepsilon_t \sim N(0, \sigma^2)$.

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

Замечание. В случае нормальной линейной регрессионной модели условие 3b эквивалентно условию статистической независимости ошибок ε_t, ε_s при $t \neq s$ (см. приложение MC, п. 4, N4).

Замечание. Позже будет показано, что многие свойства модели сохраняются при замене условий 3a,b на более слабое условие (X может быть случайной величиной):

3′a,b. $\mathrm{Cov}(X_t, \varepsilon_s) = 0$ для всех t, s,
$\mathrm{E}(\varepsilon_t \mid \boldsymbol{X}) = 0$, $\mathrm{E}(\varepsilon_t^2 \mid \boldsymbol{X}) = \sigma^2$ при всех t,
$\mathrm{E}(\varepsilon_t \varepsilon_s \mid \boldsymbol{X}) = 0$ при всех $t \neq s$.

Обсудим гипотезы, лежащие в основе линейной регрессионной модели.

1. Спецификация модели отражает наше представление о механизме зависимости Y_t от X_t и сам выбор объясняющей переменной X_t.

3a,b. Эти условия в векторной форме могут быть записаны так:

$$\mathrm{E}\boldsymbol{\varepsilon} = 0, \quad \mathrm{V}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}_n,$$

где $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)'$, $\boldsymbol{I}_n$ — $n \times n$ единичная матрица, $\mathrm{V}(\boldsymbol{\varepsilon})$ — $n \times n$ матрица ковариаций.

Условие $\mathrm{E}\boldsymbol{\varepsilon} = \mathbf{0}$ означает, что $\mathrm{E}Y_t = a + bX_t$, т.е. при фиксированном X_t среднее ожидаемое значение Y_t равно $a + bX_t$.

Условие независимости дисперсии ошибки от номера наблюдения (от регрессора X_t): $\mathrm{E}(\varepsilon_t^2) = \mathrm{V}(\varepsilon_t) = \sigma^2$, $t = 1, \ldots, n$, называется *гомоскедастичностью* (*homoscedasticity*); случай, когда условие гомоскедастичности не выполняется, называется *гетероскедастичностью* (*heteroscedasticity*). На рис. 2.2а приведен пример типичной картинки для случая гомоскедастичности ошибок; на рис. 2.2б — пример данных с гетероскедастичными ошибками (возможно, что в этом примере $\mathrm{V}(\varepsilon_t) \sim X_t^2$).

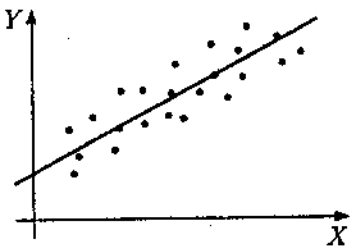


Рис. 2.2а

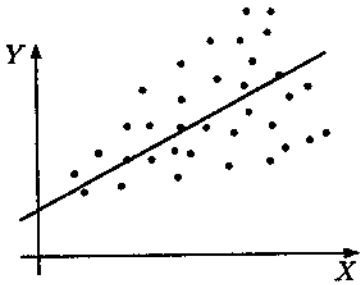


Рис. 2.2б

Условие $\mathrm{E}(\varepsilon_t \varepsilon_s) = 0$, $t \neq s$ указывает на некоррелированность ошибок для разных наблюдений. Это условие часто нарушается в случае, когда наши данные являются временными рядами. В случае, когда это условие не выполняется, говорят об *автокорреляции ошибок* (*serial correlation*).

Для простейшего случая автокорреляции ошибок, когда $\mathrm{E}(\varepsilon_t \varepsilon_{t+1}) = \rho \neq 0$, типичный вид данных представлен на рис. 2.3а ($\rho > 0$) и рис. 2.3б ($\rho < 0$).

Отметим, что условия 3a,b можно также написать в терминах зависимой переменной: $\mathrm{E}Y_t = a + bX_t$, $\mathrm{V}(Y_t) = \sigma^2$, $\mathrm{Cov}(Y_t, Y_s) = 0$, $t \neq s$.

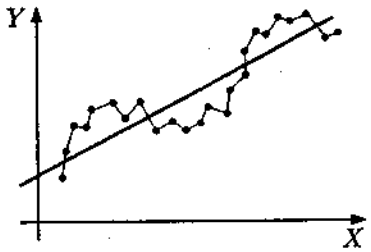


Рис. 2.3а

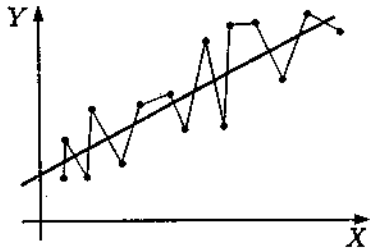


Рис. 2.3б

2.4. Теорема Гаусса–Маркова. Оценка дисперсии ошибок σ^2

Итак, мы имеем набор данных (наблюдений) (X_t, Y_t), $t = 1, \ldots, n$, и модель 1–3ab. Наша задача — оценить все три параметра модели: a, b, σ^2.

Мы хотим оценить параметры a и b «наилучшим» способом. Что значит «наилучшим»? Например, найти в классе линейных (по Y_t) несмещенных оценок наилучшую в смысле минимальной дисперсии (*Best Linear Unbiased Estimator, BLUE*).

Заметим, что когда такая оценка найдена, это вовсе не означает, что не существует нелинейной несмещенной оценки с меньшей дисперсией. Кроме того, например, можно отбросить требование несмещенности оценки и минимизировать среднеквадратичное отклонение оценки от истинного значения: $\mathrm{E}(\widehat{b} - b)^2$.

Теорема Гаусса–Маркова. *В предположениях модели* 1–3ab:

1. $Y_t = a + bX_t + \varepsilon_t$, $t = 1, \ldots, n$;
2. X_t — *детерминированная величина*;

3a. $\mathrm{E}\varepsilon_t = 0$, $\mathrm{E}(\varepsilon_t^2) = \mathrm{V}(\varepsilon_t) = \sigma^2$;

3b. $\mathrm{E}(\varepsilon_t \varepsilon_s) = 0$, *при* $t \neq s$;

оценки $\widehat{a}$, $\widehat{b}$ (2.4а), (2.4б), *полученные по методу наименьших квадратов (МНК), имеют наименьшую дисперсию в классе всех линейных несмещенных оценок.*

Доказательство. 1. Проверим, что МНК-оценки $\widehat{a}$, $\widehat{b}$ являются несмещенными оценками истинных значений a, b. Из (2.4а), (2.4б), (2.6) получаем:

$$\mathrm{E}\widehat{b} = \mathrm{E}\frac{n\sum X_tY_t - (\sum X_t)(\sum Y_t)}{n\sum X_t^2 - (\sum X_t)^2} = \mathrm{E}\frac{\sum x_ty_t}{\sum x_t^2} = \frac{\sum x_t\mathrm{E}y_t}{\sum x_t^2}$$

$$= \frac{\sum x_tbx_t}{\sum x_t^2} = b,$$

$$\mathrm{E}\widehat{a} = \mathrm{E}\left(\frac{1}{n}\sum Y_t - \frac{1}{n}\sum X_t\widehat{b}\right) = \mathrm{E}\overline{Y} - \overline{X}\mathrm{E}\widehat{b} = a + b\overline{X} - \overline{X}b = a.$$

2. Вычислим дисперсии оценок $\widehat{a}$, $\widehat{b}$.

Представим $\widehat{b}$ в виде (ср. (2.6))

$$\widehat{b} = \frac{\sum x_ty_t}{\sum x_t^2} = \sum w_ty_t, \quad \text{где } w_t = \frac{x_t}{\sum\limits_s x_s^2}. \tag{2.9}$$

Легко проверить, что w_t удовлетворяют следующим условиям:

$$\begin{aligned} &1)\quad \sum w_t = 0,\\ &2)\quad \sum w_tx_t = \sum w_tX_t = 1,\\ &3)\quad \sum w_t^2 = 1/\sum x_t^2,\\ &4)\quad \sum w_ty_t = \sum w_tY_t, \end{aligned} \tag{2.10}$$

$$\mathrm{V}(\widehat{b}) = \mathrm{V}\left(\sum w_ty_t\right) = \mathrm{V}\left(\sum w_tY_t\right) = \sum w_t^2\sigma^2 = \frac{\sigma^2}{\sum x_t^2}, \tag{2.11}$$

$$\widehat{a} = \overline{Y} - \overline{X}\widehat{b} = \overline{Y} - \overline{X}\sum w_tY_t = \sum\left(\frac{1}{n} - \overline{X}w_t\right)Y_t, \tag{2.12}$$

$$\mathrm{V}(\widehat{a}) = \mathrm{V}\left(\sum\left(\frac{1}{n} - \overline{X}w_t\right)Y_t\right) = \sigma^2\sum\left(\frac{1}{n} - \overline{X}w_t\right)^2$$

$$= \sigma^2\left(\frac{1}{n} + \frac{\overline{X}^2}{\sum x_t^2}\right) = \sigma^2\frac{\sum X_t^2}{n\sum x_t^2} \tag{2.13}$$

(при выводе (2.13) мы использовали тождество $\sum x_t^2 \equiv \sum X_t^2 - n\overline{X}^2$).

Упражнение. Используя (2.9), (2.12) и аналогичные вычисления, покажите, что

$$\mathrm{Cov}(\widehat{a}, \widehat{b}) = -\frac{\overline{X}}{\sum x_t^2}\sigma^2 = -\frac{\overline{X}}{\sum(X_t - \overline{X})^2}\sigma^2. \tag{2.14}$$

3. Покажем, что МНК-оценки являются «наилучшими» (в смысле наименьшей дисперсии) в классе всех линейных несмещенных оценок.

Пусть $\widetilde{b} = \sum c_tY_t$ — любая другая несмещенная оценка. Представим c_t в виде $c_t = w_t + d_t$, тогда

$$\mathrm{E}(\widetilde{b} - \widehat{b}) = 0 = \mathrm{E}\left(\sum d_tY_t\right) = \sum d_t(a + bX_t)$$

для всех a, b. Отсюда

$$\sum d_t = 0; \qquad \sum d_tX_t = \sum d_tx_t = 0,$$

$$\mathrm{V}(\widetilde{b}) = \mathrm{V}\left(\sum c_tY_t\right) = \sigma^2\sum c_t^2 = \sigma^2\sum(w_t + d_t)^2$$

$$= \sigma^2\left(\sum w_t^2 + 2\sum w_td_t + \sum d_t^2\right) = \sigma^2\left(\sum w_t^2 + \sum d_t^2\right)$$

$$= \mathrm{V}(\widehat{b}) + \sigma^2\sum d_t^2,$$

т. е. $\mathrm{V}(\widetilde{b}) \geqslant \mathrm{V}(\widehat{b})$, что и требовалось доказать. (Выше $\sum w_td_t = 0$ в силу определения w_t (2.9) и того, что $\sum d_tx_t = 0$.)

Аналогичные вычисления показывают, что $\mathrm{V}(\widetilde{a}) \geqslant \mathrm{V}(\widehat{a})$; мы оставим доказательство этого факта в качестве упражнения. Ниже (в главе 3) мы докажем теорему Гаусса–Маркова в общем случае.

Упражнение. Покажите, что $\mathrm{V}(\widetilde{a}) \geqslant \mathrm{V}(\widehat{a})$.

Оценка дисперсии ошибок σ^2

Итак, теперь у нас есть «наилучшие» (в смысле теоремы Гаусса–Маркова) оценки коэффициентов регрессии a, b. Однако в регрессионном уравнении есть еще один параметр — дисперсия ошибок σ^2.

Обозначим через $\widehat{Y}_t = \widehat{a} + \widehat{b}X_t$ прогноз (*fitted value*) значения Y_t в точке X_t. *Остатки* регрессии e_t определяются из уравнения

$Y_t = \widehat{Y}_t + e_t = \widehat{a} + \widehat{b}X_t + e_t$. Не следует путать остатки регрессии с ошибками регрессии в уравнении модели $Y_t = a + bX_t + \varepsilon_t$. Остатки e_t, так же как и ошибки ε_t, являются случайными величинами, однако разница состоит в том, что остатки, в отличие от ошибок, наблюдаемы.

Кажется вполне естественной гипотеза, что оценка σ^2 связана с суммой квадратов остатков регрессии $e_t = Y_t - \widehat{a} - \widehat{b}X_t$. В самом деле,

$$\begin{aligned}\sum e_t^2 &= \sum (Y_t - \widehat{a} - \widehat{b}X_t)^2 = \sum (\overline{Y} + y_t - \widehat{a} - \widehat{b}\overline{X} - \widehat{b}x_t)^2 \\ &= \sum (y_t - \widehat{b}x_t)^2 = \sum (bx_t + \varepsilon_t - \overline{\varepsilon} - \widehat{b}x_t)^2 \\ &= \sum \left((b - \widehat{b})x_t + (\varepsilon_t - \overline{\varepsilon})\right)^2 \\ &= \sum x_t^2 (b - \widehat{b})^2 + 2(b - \widehat{b}) \sum x_t(\varepsilon_t - \overline{\varepsilon}) + \sum (\varepsilon_t - \overline{\varepsilon})^2 \\ &= \mathrm{I} + \mathrm{II} + \mathrm{III}.\end{aligned}$$

Вычислим математическое ожидание $\mathrm{E}\sum e_t^2 = \mathrm{E(I)} + \mathrm{E(II)} + \mathrm{E(III)}$.

$$\mathrm{E(I)} = \mathrm{E}\Big(\sum x_t^2 (b - \widehat{b})^2\Big) = \sum x_t^2 \mathrm{V}(\widehat{b}) = \sum x_t^2 \frac{\sigma^2}{\sum x_t^2} = \sigma^2.$$

Используя соотношение $\widehat{b} = \sum w_t y_t = \sum w_t(bx_t + \varepsilon_t - \overline{\varepsilon}) = b + \sum w_t \varepsilon_t$, получаем

$$\begin{aligned}\mathrm{E(II)} &= -2\mathrm{E}\Big(\sum_t w_t \varepsilon_t \sum_s x_s(\varepsilon_s - \overline{\varepsilon})\Big) \\ &= -2\mathrm{E}\Big(\sum_{t,s} w_t x_s \varepsilon_t \varepsilon_s - \sum_t w_t \varepsilon_t \overline{\varepsilon} \sum_s x_s\Big) \\ &= -2\sum_t w_t x_t \sigma^2 = -2\sigma^2, \\ \mathrm{E(III)} &= \mathrm{E}\Big(\sum \varepsilon_t^2 - 2\overline{\varepsilon}\sum \varepsilon_t + n\overline{\varepsilon}^2\Big) \\ &= n\sigma^2 - 2n\frac{1}{n}\sigma^2 + n\frac{1}{n}\sigma^2 = (n-1)\sigma^2.\end{aligned}$$

Таким образом,

$$\mathrm{E}\sum e_t^2 = \mathrm{E(I)} + \mathrm{E(II)} + \mathrm{E(III)} = \sigma^2 - 2\sigma^2 + (n-1)\sigma^2 = (n-2)\sigma^2.$$

Отсюда следует, что

$$s^2 = \widehat{\sigma}^2 = \frac{1}{n-2}\sum e_t^2 \tag{2.15}$$

является несмещенной оценкой дисперсии ошибок σ^2.

Формулы (2.11), (2.13) дают дисперсии оценок $\widehat{a}$, $\widehat{b}$ коэффициентов регрессии в том случае, если σ^2 известно. На практике, как правило, дисперсия ошибок σ^2 неизвестна и оценивается по наблюдениям одновременно с коэффициентами регрессии a, b. В этом случае вместо дисперсий оценок $\widehat{a}$, $\widehat{b}$ мы можем получить лишь оценки дисперсий $\widehat{a}$, $\widehat{b}$, заменив σ^2 на s^2 из (2.15) в (2.11), (2.13), (2.14):

$$\begin{aligned}\widehat{\mathrm{V}}(\widehat{b}) &= s^2 \frac{1}{\sum x_t^2} = \frac{s^2}{\sum (X_t - \overline{X})^2}, \\ \widehat{\mathrm{V}}(\widehat{a}) &= s^2 \frac{\sum X_t^2}{n\sum x_t^2} = \frac{s^2 \sum X_t^2}{n\sum (X_t - \overline{X})^2}, \\ \widehat{\mathrm{Cov}}(\widehat{a}, \widehat{b}) &= -\frac{\overline{X}}{\sum x_t^2} s^2 = -\frac{\overline{X} s^2}{\sum (X_t - \overline{X})^2}.\end{aligned} \tag{2.16}$$

Стандартные отклонения оценок коэффициентов регрессии, которые приводятся в результатах регрессии в статистических пакетах, вычисляются на основе этих формул ($s_{\widehat{b}} = \sqrt{\widehat{\mathrm{V}}(\widehat{b})}$).

Замечание. Предположим, что мы изучаем зависимость Y от X и число наблюдений n задано, но мы можем выбирать набор $(X_1, X_2, \ldots, X_n)$. Как выбрать X_i так, чтобы точность оценки углового коэффициента b была наибольшей? Дисперсия оценки $\widehat{b}$ задается формулой (2.16), откуда видно, что чем больше $\sum x_t^2$, тем меньше дисперсия $\mathrm{V}(\widehat{b})$. Поэтому желательно выбирать X_t таким образом, чтобы их разброс вокруг среднего значения был большим.

Замечание. Из формулы (2.14) для ковариации оценок свободного члена a и углового коэффициента b следует, что $\mathrm{Cov}(\widehat{a}, \widehat{b}) < 0$, если $\overline{X} > 0$. Это соответствует геометрической интуиции (рис. 2.4). В самом деле, график уравнения регрессии $Y = \widehat{a} + \widehat{b}X$ в силу (2.5) проходит через точку с координатами $(\overline{X}, \overline{Y})$, поэтому

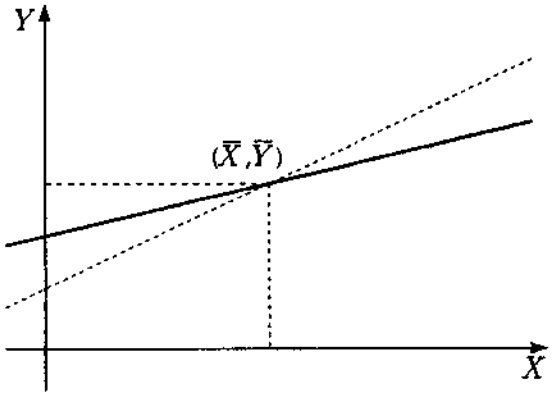


Рис. 2.4

при увеличении $\widehat{b}$ (повороте прямой вокруг точки $(\overline{X},\overline{Y})$ против часовой стрелки) величина $\widehat{a}$ уменьшается.

2.5. Статистические свойства МНК-оценок параметров регрессии. Проверка гипотезы $b = b_0$. Доверительные интервалы для коэффициентов регрессии

Пусть выполняется условие нормальной линейной регрессионной модели $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$, т. е. $\boldsymbol{\varepsilon}$ — многомерная нормально распределенная случайная величина, или, что то же самое, Y_t имеют совместное нормальное распределение. Тогда МНК-оценки коэффициентов регрессии $\widehat{a}$, $\widehat{b}$ также имеют совместное нормальное распределение, так как они являются линейными функциями (2.4а), (2.4б) от Y_t:

$$\widehat{a} \sim N\left(a, \sigma^2 \frac{\sum X_t^2}{n \sum x_t^2}\right), \quad \widehat{b} \sim N\left(b, \sigma^2 \frac{1}{\sum x_t^2}\right). \tag{2.17}$$

Если гипотеза нормальности ошибок не выполняется, то (2.17), вообще говоря, неверно, однако при некоторых условиях регуляр-

ности на поведение X_t при росте n оценки $\widehat{a}$, $\widehat{b}$ имеют асимптотически нормальное распределение, т. е. (2.17) выполняется асимптотически при $n \to \infty$.

Распределение оценки дисперсии ошибок s^2

Покажем, что в случае нормальной линейной регрессионной модели, т. е. когда $\boldsymbol{\varepsilon}$ — многомерная нормально распределенная случайная величина, выполняется

$$\frac{(n-2)s^2}{\sigma^2} \sim \chi^2(n-2).$$

Используя представление $\widehat{b}$ в виде (2.9), получаем

$$\widehat{b} = \frac{\sum x_t y_t}{\sum x_t^2} = b + \sum w_t \varepsilon_t. \tag{2.18}$$

Подставив (2.18) в уравнение регрессии $\widehat{Y} = \widehat{a} + \widehat{b}X$, получим следующую формулу для остатков регрессии:

$$\begin{aligned} e_t &= Y_t - \widehat{Y}_t = Y_t - \widehat{a} - \widehat{b}X_t = a + bX_t + \varepsilon_t - (\overline{Y} - \widehat{b}\overline{X}) - \widehat{b}X_t \\ &= a + bX_t + \varepsilon_t - a - b\overline{X} - \overline{\varepsilon} + b\overline{X} + \overline{X}\sum w_s\varepsilon_s - bX_t \\ &- X_t \sum w_s \varepsilon_s = \varepsilon_t - \overline{\varepsilon} - x_t \sum w_s \varepsilon_s. \end{aligned} \tag{2.19}$$

Представим (2.19) в векторной форме:

$$\boldsymbol{e} = \left(\boldsymbol{I} - \frac{1}{n}\boldsymbol{\imath}\boldsymbol{\imath}' - \boldsymbol{x}_*\boldsymbol{w}'\right)\boldsymbol{\varepsilon} = \boldsymbol{A}\boldsymbol{\varepsilon}; \tag{2.20}$$

здесь $\boldsymbol{\imath}$ — $n \times 1$ вектор-столбец, состоящий из одних единиц, $\boldsymbol{e} = (e_1, \ldots, e_n)'$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)'$, $\boldsymbol{A}$ — $n \times n$ матрица и $\boldsymbol{x}_* = (x_1, \ldots, x_n)'$ — вектор отклонений от среднего значения.

Далее воспользуемся свойством N8 (приложение МС, п. 4). Проверим, что матрица $\boldsymbol{A} = \boldsymbol{I} - \frac{1}{n}\boldsymbol{\imath}\boldsymbol{\imath}' - \boldsymbol{x}_*\boldsymbol{w}'$ из (2.20) является идемпотентной: $\boldsymbol{A}' = \boldsymbol{A}$, так как $(\boldsymbol{\imath}\boldsymbol{\imath}')' = \boldsymbol{\imath}''\boldsymbol{\imath}' = \boldsymbol{\imath}\boldsymbol{\imath}'$, и

$$(\boldsymbol{x}_*\boldsymbol{w}')' = \left(\frac{\boldsymbol{x}_*\boldsymbol{x}_*'}{\sum x_t^2}\right)' = \frac{(\boldsymbol{x}_*\boldsymbol{x}_*')'}{\sum x_t^2} = \frac{\boldsymbol{x}_*\boldsymbol{x}_*'}{\sum x_t^2} = \boldsymbol{x}_*\boldsymbol{w}'.$$

$$\begin{aligned}\boldsymbol{A}^2 &= \left(\boldsymbol{I} - \frac{1}{n}\boldsymbol{\imath}\boldsymbol{\imath}' - \boldsymbol{x}_*\boldsymbol{w}'\right)^2 \\ &= \boldsymbol{I} + \frac{1}{n^2}\boldsymbol{\imath}\boldsymbol{\imath}'\boldsymbol{\imath}\boldsymbol{\imath}' + \boldsymbol{x}_*\boldsymbol{w}'\boldsymbol{x}_*\boldsymbol{w}' - \frac{2}{n}\boldsymbol{\imath}\boldsymbol{\imath}' - 2\boldsymbol{x}_*\boldsymbol{w}' + \frac{1}{n}\boldsymbol{\imath}\boldsymbol{\imath}'\boldsymbol{x}_*\boldsymbol{w}' + \frac{1}{n}\boldsymbol{x}_*\boldsymbol{w}'\boldsymbol{\imath}\boldsymbol{\imath}' \\ &= \boldsymbol{I} - \frac{1}{n}\boldsymbol{\imath}\boldsymbol{\imath}' + \boldsymbol{x}_*\boldsymbol{w}' - 2\boldsymbol{x}_*\boldsymbol{w}' = \boldsymbol{I} - \frac{1}{n}\boldsymbol{\imath}\boldsymbol{\imath}' - \boldsymbol{x}_*\boldsymbol{w}' = \boldsymbol{A},\end{aligned}$$

так как $\boldsymbol{\imath}'\boldsymbol{\imath} = n$, $\boldsymbol{\imath}'\boldsymbol{x}_* = \boldsymbol{w}'\boldsymbol{\imath} = 0$ (см. (2.10)).

По условиям классической нормальной регрессионной модели $\boldsymbol{\varepsilon}/\sigma \sim N(\boldsymbol{0}, \boldsymbol{I}_n)$, таким образом, в силу свойства N8 (приложение МС, п. 4):

$$\frac{(n-2)s^2}{\sigma^2} = \frac{1}{\sigma^2}\sum e_t^2 = \left(\frac{\boldsymbol{\varepsilon}}{\sigma}\right)'\boldsymbol{A}\left(\frac{\boldsymbol{\varepsilon}}{\sigma}\right) \sim \chi^2(n-2), \tag{2.21}$$

что и требовалось доказать.

Независимость s^2 и МНК-оценок $\widehat{a}$, $\widehat{b}$

Так как оценка дисперсии ошибок s^2 является функцией от остатков регрессии e_t, то для того чтобы доказать независимость s^2 и $(\widehat{a}, \widehat{b})$, достаточно доказать независимость e_t и $(\widehat{a}, \widehat{b})$. Оценки $\widehat{a}$, $\widehat{b}$ так же, как и остатки регрессии e_t, являются линейными функциями ошибок ε_t (см. (2.4а), (2.4б), (2.20)) и поэтому имеют совместное нормальное распределение. Известно (приложение МС, п. 4, N4), что два случайных вектора, имеющие совместное нормальное распределение, независимы тогда и только тогда, когда они некоррелированы. Таким образом, чтобы доказать независимость s^2 и $(\widehat{a}, \widehat{b})$, нам достаточно доказать некоррелированность e_t и $(\widehat{a}, \widehat{b})$.

Обозначим $\xi = \sum w_t\varepsilon_t$, тогда из (2.19) $e_t = \varepsilon_t - \overline{\varepsilon} - x_t\xi$, а из (2.18) следует, что $\widehat{b} = b + \xi$.

$$\begin{aligned}\operatorname{Cov}(e_t, \widehat{b}) &= \operatorname{Cov}(\varepsilon_t - \overline{\varepsilon} - x_t\xi,\ b + \xi) = \mathrm{E}(\varepsilon_t\xi - \overline{\varepsilon}\xi - x_t\xi^2) \\ &= \sigma^2\left(w_t - \frac{1}{n}\sum w_s - x_t\sum w_s^2\right) \\ &= \sigma^2\left(w_t - x_t\frac{1}{\sum x_s^2}\right) = 0.\end{aligned} \tag{2.22}$$

Таким образом, e_t и $\widehat{b}$ независимы при всех t, что и требовалось показать.

Упражнение. Используя аналогичные выкладки, показать независимость e_t и $\widehat{a}$.

Проверка гипотезы $b = b_0$

Из (2.17) имеем $\widehat{b} - b \sim N(0, \sigma_{\widehat{b}}^2)$, где $\sigma_{\widehat{b}}^2 = \sigma^2/\sum x_t^2$. Оценка дисперсии оценки $\widehat{b}$ может быть получена по формуле $\widehat{\mathrm{V}}(\widehat{b}) = s_{\widehat{b}}^2 = \dfrac{s^2}{\sum x_t^2}$. Таким образом, $\dfrac{\widehat{b} - b}{\sigma_{\widehat{b}}} \sim N(0,1)$, а из (2.21) следует, что

$$\frac{s}{\sigma} \sim \sqrt{\frac{1}{n-2}\chi^2(n-2)},$$

т. е. по определению статистики Стьюдента (приложение МС, п. 3) имеем

$$t = \frac{(\widehat{b} - b)/\sigma_{\widehat{b}}}{s/\sigma} \sim t(n-2),$$

и, так как $\dfrac{\sigma_{\widehat{b}}}{\sigma} = \dfrac{s_{\widehat{b}}}{s}$, получаем

$$t = \frac{\widehat{b} - b}{s_{\widehat{b}}} \sim t(n-2). \tag{2.23}$$

Упражнение. Используя аналогичные выкладки, покажите, что

$$t = \frac{\widehat{a} - a}{s_{\widehat{a}}} \sim t(n-2). \tag{2.24}$$

Итак, мы показали, что в случае нормально распределенных ошибок величина (2.23) распределена по закону Стьюдента. Заметим, что при определенных условиях регулярности на x_t соотношения (2.23) и (2.24) справедливы асимптотически и без предположения о нормальности ошибок регрессии.

Статистику (2.23) можно использовать для проверки гипотезы H_0: $b = b_0$ против альтернативной гипотезы H_1: $b \neq b_0$. Предположим, что верна гипотеза H_0, тогда

$$t = \frac{\widehat{b} - b_0}{s_{\widehat{b}}} \sim t(n-2).$$

Зададимся, например, 2.5%-ной точкой t-распределения с $(n-2)$ степенями свободы $t_{0.025}$ (при 40 степенях свободы $t_{0.025} = 2.021$), т. е.

$$\mathrm{P}\{-t_{0.025} < t < t_{0.025}\} = 0.95.$$

Мы отвергаем гипотезу H_0 (и принимаем H_1) на 5%-ном уровне значимости, если $|t| > t_{0.025}$ («редкое» событие с точки зрения гипотезы H_0), в противном случае мы не можем отвергнуть H_0 (и принимаем H_0).

При тестировании нулевой гипотезы H_0: $b = b_0$ против двусторонней альтернативной гипотезы H_1: $b \neq b_0$ на доверительном уровне α нулевая гипотеза отвергается при $|t| > t_{\alpha/2}(n-2)$. Соответственно, при тестировании гипотезы H_0 против односторонней альтернативы H_1: $b > b_0$, нулевая гипотеза отвергается при $t > t_\alpha(n-2)$. (Здесь $t_\alpha(n)$ означает 100α%-ную точку распределения Стьюдента с n степенями свободы.)

Разрешив неравенство в $\mathrm{P}\{|(\widehat{b}-b)/s_{\widehat{b}}| < t_{0.025}\} = 0.95$ относительно b, получим

$$\mathrm{P}\{\widehat{b} - t_{0.025}s_{\widehat{b}} < b < \widehat{b} + t_{0.025}s_{\widehat{b}}\} = 0.95,$$

т. е. $[\widehat{b} - t_{0.025}s_{\widehat{b}},\ \widehat{b} + t_{0.025}s_{\widehat{b}}]$ — 95%-ный *доверительный интервал* для b. Доверительный интервал накрывает истинное значение параметра b с заданной вероятностью (в данном случае 95%). Соответственно, двусторонний $100(1-\alpha)$%-ный доверительный интервал для параметра b имеет следующий вид:

$$[\widehat{b} - t_{\alpha/2}(n-2)s_{\widehat{b}},\ \widehat{b} + t_{\alpha/2}(n-2)s_{\widehat{b}}].$$

Наиболее просто выглядит t-статистика при гипотезе H_0: $b = 0$, а именно, $t = \widehat{b}/s_{\widehat{b}}$. Это значение и приводится всеми компьютерными пакетами в результатах регрессии. Значение $|t| > t_{0.025}$ ($t_{0.025} \approx 2$ для больших n) позволяет сделать вывод об отличии от нуля (на соответствующем уровне значимости) коэффициента регрессии и, следовательно, о наличии влияния (связи) X на Y. Малые значения t-статистики соответствуют отсутствию достоверной статистической связи объясняющей переменной X и зависимой переменной Y.

Компьютерные пакеты приводят также двусторонние P-значения t-статистики, т. е. вероятность того, что случайная величина, распределенная по закону $t(n-2)$, принимает значение по абсолютной величине большее, чем $|t| = \left|\dfrac{\widehat{b}}{s_{\widehat{b}}}\right|$.

2.6. Анализ вариации зависимой переменной в регрессии. Коэффициент детерминации R^2

Анализ вариации зависимой переменной в регрессии

Рассмотрим вариацию (разброс) $\sum(Y_t - \overline{Y})^2$ значений Y_t вокруг среднего значения. Разобьем эту вариацию на две части: объясненную регрессионным уравнением и не объясненную (т. е. связанную с ошибками ε_t).

Обозначим через $\widehat{Y}_t = \widehat{a} + \widehat{b}X_t$ предсказанное значение Y_t, тогда $Y_t - \overline{Y} = (Y_t - \widehat{Y}_t) + (\widehat{Y}_t - \overline{Y})$ (см. рис. 2.5) и вариация Y_t представляется в виде трех слагаемых:

$$\sum(Y_t - \overline{Y})^2 = \sum(Y_t - \widehat{Y}_t)^2 + \sum(\widehat{Y}_t - \overline{Y})^2 + 2\sum(Y_t - \widehat{Y}_t)(\widehat{Y}_t - \overline{Y}). \tag{2.25}$$

Третье слагаемое в (2.25) равно нулю, так как $\boldsymbol{y} - \widehat{\boldsymbol{y}} = \boldsymbol{e}$, — вектор остатков регрессии, ортогонален константе $\boldsymbol{\imath}$ и вектору $\boldsymbol{x}$ (см. (2.7)). В самом деле, $\sum e_t(\widehat{Y}_t - \overline{Y}) = \sum e_t(\widehat{a} + \widehat{b}X_t - \overline{Y}) = (\widehat{a} + \widehat{b}\overline{X} - \overline{Y})\sum e_t + \widehat{b}\sum e_t x_t = 0$. Поэтому верно равенство

$$\underbrace{\sum(Y_t - \overline{Y})^2}_{\mathrm{TSS}} = \underbrace{\sum(Y_t - \widehat{Y}_t)^2}_{\mathrm{ESS}} + \underbrace{\sum(\widehat{Y}_t - \overline{Y})^2}_{\mathrm{RSS}}. \tag{2.26}$$

Замечание. Вектор остатков регрессии ортогонален константе, т. е. $\boldsymbol{\imath}'\boldsymbol{e} = \sum e_t = 0$, вообще говоря, *только* в том случае, когда константа *включена* в число объясняющих параметров регрессии. Поэтому (2.26) справедливо, вообще говоря, *только* в случае, когда константа *включена* в число объясняющих параметров регрессии.

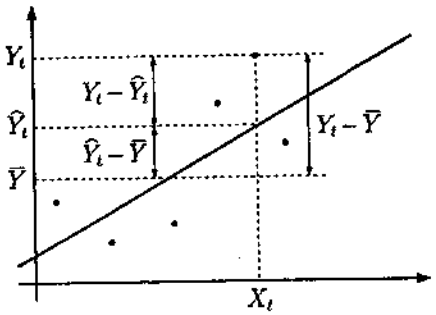


Рис. 2.5

Обозначим левую часть в (2.26) через TSS (*total sum of squares*) — вся дисперсия, первое слагаемое в правой части, соответствующее не объясненной дисперсии, через ESS (*error sum of squares*), второе слагаемое в правой части — RSS (*regression sum of squares*) — объясненная часть всей дисперсии[1].

Статистика R^2 — коэффициент детерминации

Определение. Коэффициентом детерминации, или долей объясненной дисперсии, называется

$$R^2 = 1 - \frac{\text{ESS}}{\text{TSS}} = \frac{\text{RSS}}{\text{TSS}}. \tag{2.27}$$

Заметим, что второе равенство в (2.27) верно лишь в том случае, если верно (2.26), т. е. когда константа включена в уравнение регрессии. Только в этом случае имеет смысл рассматривать статистику R^2.

В силу определения R^2 принимает значения между 0 и 1, $0 \leqslant R^2 \leqslant 1$. Если $R^2 = 0$, то это означает, что регрессия ничего не дает, т. е. X_t не улучшает качество предсказания Y_t по сравнению с тривиальным предсказанием $\widehat{Y}_t = \overline{Y}$.

[1] К сожалению, эти обозначения не унифицированы. В некоторых книгах (см., например, Johnston, DiNardo, 1997) первое слагаемое в правой части (2.26) обозначается RSS (*residual sum of squares*), а второе — ESS (*explained sum of squares*).

Другой крайний случай $R^2 = 1$ означает точную подгонку: все точки наблюдений лежат на регрессионной прямой (все $e_t = 0$).

Чем ближе к 1 значение R^2, тем лучше качество подгонки, $\widehat{\boldsymbol{y}}$ более точно аппроксимирует $\boldsymbol{y}$. Ниже в главе 3, посвященной множественной регрессии, мы более подробно остановимся на вопросе о значении R^2 при оценке качества регрессии.

Геометрическая интерпретация коэффициента R^2

Вернемся к геометрической интерпретации регрессии из раздела 2.2. Вектор $\overline{Y}\boldsymbol{\imath}$ является ортогональной проекцией вектора $\boldsymbol{y}$ на вектор $\boldsymbol{\imath}$. Вектор $\widehat{\boldsymbol{y}}$ есть ортогональная проекция вектора $\boldsymbol{y}$ на плоскость $(\boldsymbol{\imath}, \boldsymbol{x})$ (см. рис. 2.6). По теореме о трех перпендикулярах ортогональная проекция вектора $\widehat{\boldsymbol{y}}$ на вектор $\boldsymbol{\imath}$ совпадает с $\overline{Y}\boldsymbol{\imath}$. Равенство (2.26) является теоремой Пифагора для прямоугольного треугольника со сторонами $\widehat{\boldsymbol{y}} - \overline{Y}\boldsymbol{\imath}$, $\boldsymbol{y} - \overline{Y}\boldsymbol{\imath}$, $\boldsymbol{e}$, т. е. $\|\boldsymbol{y} - \overline{Y}\boldsymbol{\imath}\|^2 = \|\boldsymbol{e}\|^2 + \|\widehat{\boldsymbol{y}} - \overline{Y}\boldsymbol{\imath}\|^2$. Поэтому $R^2 = \text{RSS}/\text{TSS} = \cos^2\varphi$, где φ — угол между векторами $(\widehat{\boldsymbol{y}} - \overline{Y}\boldsymbol{\imath})$ и $(\boldsymbol{y} - \overline{Y}\boldsymbol{\imath})$.

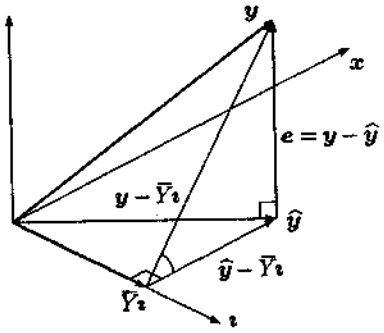


Рис. 2.6

Упражнение. Покажите, что R^2 равен квадрату выборочного коэффициента корреляции между Y_t и $\widehat{Y}_t$.

F-статистика

Снова предположим, что мы находимся в рамках нормальной линейной регрессионной модели. Из (2.17), (2.21) получаем

$$\frac{\widehat{b}-b}{\sigma_{\widehat{b}}} = \frac{\widehat{b}-b}{\sigma\Big/\sqrt{\sum x_t^2}} \sim N(0,1);$$

$$\frac{(n-2)s^2}{\sigma^2} = \frac{\sum e_t^2}{\sigma^2} \sim \chi^2(n-2).$$

Ранее мы показали, что s^2 и $\widehat{b}$ — независимые случайные величины, поэтому по определению распределений Фишера и χ-квадрат (приложение МС, п. 3 (7, 8)) получаем

$$F = \frac{\left(\dfrac{\widehat{b}-b}{\sigma_{\widehat{b}}}\right)^2 \dfrac{1}{1}}{\dfrac{\sum e_t^2}{\sigma^2}\dfrac{1}{n-2}} = \frac{(\widehat{b}-b)^2\sum x_t^2}{\sum e_t^2/(n-2)} \sim \frac{\dfrac{1}{1}\chi^2(1)}{\dfrac{1}{n-2}\chi^2(n-2)} = F(1,\ n-2). \tag{2.28}$$

Полученную F-статистику можно использовать для проверки нулевой гипотезы H_0: $b - b_0 = 0$. При этой гипотезе статистика (2.28) выглядит следующим образом:

$$F = \frac{(\widehat{b}-b_0)^2\sum x_t^2}{\sum e_t^2/(n-2)} \sim F(1,\ n-2). \tag{2.29}$$

Если нулевая гипотеза справедлива, то значение F в (2.29) мало. Таким образом, мы отвергаем нулевую гипотезу, если F превосходит критическое значение $F_\alpha(1,\ n-2)$ распределения Фишера с параметрами $(1,\ n-2)$ для выбранного уровня значимости α.

Статистика (2.29) особенно просто выглядит для гипотезы H_0: $b = 0$ (случай отсутствия линейной функциональной связи между X и Y). Преобразуя числитель следующим образом:

$$(\widehat{b})^2\sum x_t^2 = \left(\frac{\sum x_t y_t}{\sum x_t^2}\right)^2 \sum x_t^2 = \frac{(\sum x_t y_t)^2}{\sum x_t^2}$$
$$= \sum(\widehat{b}x_t)^2 = \sum(\widehat{y}_t)^2 = \widehat{\boldsymbol{y}}_*'\widehat{\boldsymbol{y}}_*,$$

получим (в векторных обозначениях для отклонений)

$$F = \frac{\widehat{\boldsymbol{y}}_*'\widehat{\boldsymbol{y}}_*}{\boldsymbol{e}'\boldsymbol{e}/(n-2)}. \tag{2.30}$$

Замечание. Сравнивая (2.23) и (2.29), мы видим что $F = t^2$, т. е. проверка гипотезы H_0, используя t- и F-статистики, дает в данном случае (для одномерной регрессионной модели) тождественные результаты.

Нетрудно заметить, что, переписывая определение R^2-статистики (2.27) в отклонениях, получаем

$$R^2 = \frac{\widehat{\boldsymbol{y}}_*'\widehat{\boldsymbol{y}}_*}{\boldsymbol{y}_*'\boldsymbol{y}_*} = \frac{\widehat{\boldsymbol{y}}_*'\widehat{\boldsymbol{y}}_*}{\boldsymbol{e}'\boldsymbol{e} + \widehat{\boldsymbol{y}}_*'\widehat{\boldsymbol{y}}_*}. \tag{2.31}$$

Используя (2.30) и (2.31), получаем следующее соотношение, связывающее R^2 и F-статистики:

$$F = (n-2)\frac{R^2}{1-R^2}. \tag{2.32}$$

Не удивительно, что малым значениям F (отсутствие значимой функциональной связи X и Y) соответствуют малые значения R^2 (плохая аппроксимация данных).

2.7. Оценка максимального правдоподобия коэффициентов регрессии

Оценка максимального правдоподобия

Наряду с методом наименьших квадратов (МНК) возможен и другой подход к оцениванию параметров линейного регрессионного уравнения по данным наблюдений — метод максимального правдоподобия. Этот метод будет рассмотрен детально в главе 10. В данном разделе мы рассмотрим его применение к оцениванию параметров парной регрессии.

Предположим, что мы ищем параметры нормальной линейной регрессионной модели

$$Y_t = a + bX_t + \varepsilon_t. \tag{2.33}$$

Ошибки регрессии ε_t независимы и распределены по нормальному закону:

$$\varepsilon_t \sim N(0, \sigma^2), \tag{2.34}$$

или, что является эквивалентной записью,

$$Y_t \sim N(a + bX_t,\ \sigma^2).$$

Имея набор наблюдений (X_t, Y_t), $t = 1, \ldots, n$, мы можем попытаться ответить на вопрос: *при каких значениях параметров a, b, σ^2 модели* (2.33)–(2.34) *вероятность получить этот набор наблюдений наибольшая*? Другими словами, каковы наиболее вероятные значения параметров модели для данного набора наблюдений?

Чтобы ответить на этот вопрос, составим *функцию правдоподобия* (приложение МС, п. 7), равную произведению плотностей вероятности отдельных наблюдений (мы считаем все ε_t независимыми):

$$L(Y_1, \ldots, Y_n, a, b, \sigma^2) = p(Y_1, \ldots, Y_n \,|\, X_1, \ldots, X_n, a, b, \sigma^2) = \prod_{t=1}^{n} p(Y_t)$$
$$= (2\pi)^{-n/2} (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum (Y_t - a - bX_t)^2\right), \tag{2.35}$$

где p обозначает плотность вероятности, зависящую от X_t, Y_t и параметров a, b, σ^2. Для того чтобы найти наиболее правдоподобные значения параметров, нам необходимо найти такие их значения, при которых функция правдоподобия L (2.35) достигает своего максимума. Так как функции L и $\ln L$ одновременно достигают своего максимума, достаточно искать максимум логарифма функции правдоподобия

$$\ln L(Y_1, \ldots, Y_n, a, b, \sigma) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2)$$
$$- \frac{1}{2\sigma^2} \sum (Y_t - a - bX_t)^2. \tag{2.36}$$

Необходимые условия экстремума функции $\ln L$ имеют вид:

$$\frac{\partial \ln L}{\partial a} = \frac{1}{\sigma^2} \sum (Y_t - a - bX_t) = 0, \tag{2.37а}$$

$$\frac{\partial \ln L}{\partial b} = \frac{1}{\sigma^2} \sum X_t (Y_t - a - bX_t) = 0, \tag{2.37б}$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4} \sum (Y_t - a - bX_t)^2 = 0. \tag{2.37в}$$

Решением системы уравнений (2.37а)–(2.37в) являются оценки *максимального правдоподобия*

$$\widehat{b}_{\mathrm{ML}} = \frac{\sum x_t y_t}{\sum x_t^2}; \quad \widehat{a}_{\mathrm{ML}} = \overline{Y} - \widehat{b}_{\mathrm{ML}}\overline{X}; \quad \widehat{\sigma}^2_{\mathrm{ML}} = \frac{1}{n}\sum e_t^2. \tag{2.38}$$

Отметим, что оценки максимального правдоподобия параметров a, b совпадают с оценками метода наименьших квадратов $\widehat{a}_{\mathrm{ML}} = \widehat{a}_{\mathrm{OLS}}$, $\widehat{b}_{\mathrm{ML}} = \widehat{b}_{\mathrm{OLS}}$. Это легко видеть из того, что уравнения (2.37а) и (2.37б) совпадают с соответствующими уравнениями метода наименьших квадратов (2.2). Оценка максимального правдоподобия для σ^2 не совпадает с $\sigma^2_{\mathrm{OLS}} = \sum e_t^2/(n-2)$, которая, как мы знаем (см. (2.15)), является несмещенной оценкой дисперсии ошибок. Таким образом, $\widehat{\sigma}^2_{\mathrm{ML}} = ((n-2)/n)\widehat{\sigma}^2_{\mathrm{OLS}}$ является смещенной, но тем не менее состоятельной оценкой σ^2.

Пример. ***Доходы семьи.*** В этом примере используются данные RLMS[2]. Пусть *Inc* обозначает реальный доход семьи, *Expend* — ее реальные расходы. Для того чтобы исследовать зависимость расходов от доходов, оценим коэффициенты регрессии *Expend* на *Inc* и константу.

Для всех типов семей (количество наблюдений 3594)

$$\mathit{Expend} = \underset{(233.6)}{4663.3} + \underset{(0.0223)}{0.686}\ \mathit{Inc}, \qquad R^2 = 0.21, \quad s = 11307.$$

В скобках приведены стандартные ошибки коэффициентов регрессии. Соответствующие t-статистики равны 19.96 и 30.81, т. е. коэффициенты статистически достоверно отличаются от нуля. Однако значение коэффициента детерминации R^2 невелико. Это объясняется, конечно, разнородностью семей как по составу, так и по другим факторам, таким, как место проживания, структура расходов,

[2]RLMS, Round VI — обследование 3781 семьи в Российской Федерации, произведенное осенью 1995 г. совместно Госкомстатом РФ, Институтом социологии РАН, Институтом питания РАМН и Университетом Северной Каролины (Чапел Хилл, США), данные можно найти по адресу: `http://www.cpc.unc.edu/projects/rlms/rlms_home.html`.

состав семьи и т. п. Таким образом, для более однородной выборки семей мы вправе ожидать увеличения значения коэффициента детерминации.

Для семей, состоящих из одного человека (количество наблюдений 509):

$$Expend = \underset{(182.0)}{3229.2} + \underset{(0.0162)}{0.355}\ Inc, \qquad R^2 = 0.39, \quad s = 4567.$$

Как и раньше, коэффициенты являются значимыми — t-статистики равны соответственно 17.74 и 20.70. Как мы и ожидали, качество подгонки улучшилось — коэффициент R^2 возрос с 0.21 до 0.39, а оценка стандартного отклонения остатков s уменьшилась с 11307 до 4567. Так как в семьях из одного человека нет расходов на содержание неработающих членов семьи (дети, престарелые), то на потребление тратится меньшая часть прироста дохода. *Склонность к потреблению*, определяемая как $\partial Expend/\partial Inc$, для семьи из одного человека равна 0.355, в то время как в среднем по всей выборке 0.686.

Обозначим через Nf количество членов в семье. Оценим регрессию среднего расхода на члена семьи на средний доход члена семьи (количество наблюдений 3594):

$$Expend/Nf = \underset{(76.8)}{2387.2} + \underset{(0.0133)}{0.447}\ Inc/Nf, \quad R^2 = 0.24, \quad s = 4202.$$

Значение R^2 увеличилось по сравнению с первой регрессией. Переход к удельным данным приводит к уменьшению дисперсии ошибок модели.

Упражнения

2.1. Наблюдения 16 пар (X, Y) дали следующие результаты:

$$\sum Y^2 = 526, \ \sum X^2 = 657, \ \sum XY = 492,$$
$$\sum Y = 64, \quad \sum X = 96.$$

Оцените регрессию $Y_t = \alpha + \beta X_t + \varepsilon_t$ и проверьте гипотезу, что коэффициент β равен 1.0.

2.2. Покажите, что $\widehat{\beta} = r_{XY}\dfrac{s_Y}{s_X}$, где r_{XY} — выборочный коэффициент корреляции между X и Y, а s_X, s_Y — стандартные отклонения X и Y, соответственно.

2.3. Пусть $\widehat{\beta}$ — оценка коэффициента наклона в регрессии Y на X, а $\widehat{\gamma}$ — оценка коэффициента наклона в регрессии X на Y. Покажите, что $\widehat{\beta} = 1/\widehat{\gamma}$ тогда и только тогда, когда $R^2 = 1$.

2.4. Рассмотрим модель $Y_t = \alpha + \beta X_t^{\gamma} + \varepsilon_t$, где ошибки являются независимыми одинаково распределенными нормальными случайными величинами. Почему для оценивания параметров нельзя применять метод наименьших квадратов? Выведите уравнение для оценок максимального правдоподобия.

2.5. Могут ли следующие уравнения быть преобразованы в уравнения, линейные по параметрам?

а) $Y_i = \alpha \cdot \exp(\beta X_i) \cdot \varepsilon_i$,

б) $Y_i = \alpha \cdot \exp(-\beta X_i) + \varepsilon_i$,

в) $Y_i = \exp(\alpha + \beta X_i + \varepsilon_i)$,

г) $Y_i = \alpha \,/\, (\beta - X_i) + \varepsilon_i$.

2.6. Зависимая переменная в регрессии $Y_i = \alpha + \beta X_i + \varepsilon_i$ разбивается на две компоненты: $Y_i = Y_{1i} + Y_{2i}$. Рассмотрим две регрессии для компонент: $Y_{1i} = \alpha_1 + \beta_1 X_i + \varepsilon_{1i}$ и $Y_{2i} = \alpha_2 + \beta_2 X_i + \varepsilon_{2i}$. Докажите следующие соотношения для МНК-оценок параметров трех регрессий: $\widehat{\alpha} = \widehat{\alpha}_1 + \widehat{\alpha}_2$; $\widehat{\beta} = \widehat{\beta}_1 + \widehat{\beta}_2$.

2.7. Уравнение $Y_i = \alpha + \beta X_i + \varepsilon_i$ оценивается методом наименьших квадратов. Остатки регрессии равны e_i, $y_i = Y_i - \overline{Y}$, $x_i = X_i - \overline{X}$, $\widehat{y}_i = \widehat{Y}_i - \overline{Y}$ — отклонения от средних. Докажите, что следующие меры качества подгонки совпадают:

а) $\left(\sum x_i y_i\right)^2 \Big/ \left(\sum x_i^2 \sum y_i^2\right)$, б) $\widehat{\beta}\left(\sum x_i y_i\right) \Big/ \left(\sum y_i^2\right)$,

в) $\left(\sum \widehat{y}_i y_i\right)^2 \Big/ \left(\sum \widehat{y}_i^2 \sum y_i^2\right)$, г) $1 - \sum e_i^2 \Big/ \sum y_i^2$.

2.8. Выведите непосредственно формулу для оценки коэффициента наклона в регрессии без свободного члена, т. е. найдите оценку параметра β в регрессии $Y_t = \beta X_t + \varepsilon_t$ минимизацией суммы квадратов отклонений $\sum (Y_t - \widehat{Y}_t)^2$.

2.9. Для наблюдений

Y	X
70	5
65	11
55	15
60	17
50	20
35	22
40	25
30	27
25	30
32	35

вычислите следующие величины:

а) коэффициент детерминации R^2 в регрессии Y_t на X_t при наличии свободного члена;

б) коэффициент детерминации R^2 в регрессии Y_t на X_t при отсутствии свободного члена;

в) коэффициент детерминации R^2 в регрессии y_t на x_t при наличии свободного члена, где y_t и x_t — отклонения переменных Y_t и X_t от их средних значений;

г) коэффициент детерминации R^2 в регрессии y_t на x_t при отсутствии свободного члена.

2.10. Предположим, что модель

$$Y_t = \alpha + \beta X_t + \varepsilon_t, \quad t = 1, \ldots, n$$

удовлетворяет условиям классической регрессии. Рассматривается следующая оценка коэффициента β:

$$\widetilde{\beta} = \frac{1}{n} \sum_{t=1}^{n} \frac{Y_t - \overline{Y}}{X_t - \overline{X}}.$$

а) Является ли оценка $\widetilde{\beta}$ несмещенной? Является ли она линейной?

б) Вычислите дисперсию оценки $\widetilde{\beta}$.

в) Проверьте теорему Гаусса–Маркова, сравнив полученную дисперсию оценки $\widetilde{\beta}$ с дисперсией МНК-оценки $\sigma^2 / \sum_{t=1}^{n} (X_t - \overline{X})^2$.

2.11. Приведите пример набора данных (X_t, Y_t), для которого решение задачи поиска параметров α, β, минимизирующих функционал

$$F = \sum_{t=1}^{n} |Y_t - (\alpha + \beta X_t)|,$$

не единственно.

2.12. Рассмотрим модель регрессии на константу

$$Y_t = \alpha + \varepsilon_t, \quad t = 1, \ldots, n.$$

а) Найдите оценки метода наименьших квадратов для α и σ^2.

б) Найдите дисперсию оценки $\widehat{\alpha}$.

в) Покажите, что статистика $\dfrac{\widehat{\alpha} - \alpha}{s_{\widehat{\alpha}}}$ имеет распределение $t(n-1)$.

г) Чему равен коэффициент детерминации R^2?

2.13. Рассмотрим модель регрессии без константы

$$Y_t = \beta X_t + \varepsilon_t, \quad t = 1, \ldots, n.$$

а) Найдите оценки метода наименьших квадратов для β и σ^2.

б) Найдите дисперсию оценки $\widehat{\beta}$.

в) Покажите, что статистика $\dfrac{\widehat{\beta} - \beta}{s_{\widehat{\beta}}}$ имеет распределение $t(n-1)$.

г) Приведите примеры данных, для которых: значение коэффициента R^2, рассчитанное по формуле $R^2 = \mathrm{RSS}/\mathrm{TSS}$, отличается от значения R^2, рассчитанного по формуле $R^2 = 1 - \mathrm{ESS}/\mathrm{TSS}$; значение коэффициента R^2, рассчитанное по формуле $R^2 = \mathrm{RSS}/\mathrm{TSS}$, больше 1; значение коэффициента R^2, рассчитанное по формуле $R^2 = 1 - \mathrm{ESS}/\mathrm{TSS}$, меньше 0.

2.14. Менеджер новой чебуречной не уверен в правильности выбранной цены на чебуреки, поэтому в течение 12 недель он варьирует цену и записывает количество проданных чебуреков. Полученные данные приведены в таблице 2.1 (t — номер недели, q_t — количество проданных чебуреков, p_t — цена одного чебурека (руб.)).

а) Оцените параметры модели

$$\ln q_t = \alpha + \beta \ln p_t + \varepsilon_t, \quad t = 1, \ldots, 12.$$

б) Используя полученные оценки коэффициентов, найдите оптимальную в смысле максимума выручки от продаж цену чебурека.

Таблица 2.1

t	p_t	q_t	t	p_t	q_t
1	12.3	795	7	12.8	714
2	11.5	915	8	9.9	1180
3	11.0	965	9	12.2	851
4	12.0	892	10	12.5	779
5	13.5	585	11	13.0	625
6	12.5	644	12	10.5	1001

2.15. Пусть $\widehat{\sigma}^2_{\mathrm{ML}} = \sum e_t^2/n$ и $\widehat{\sigma}^2_{\mathrm{OLS}} = \sum e_t^2/(n-2)$ — оценки методов максимального правдоподобия и наименьших квадратов для дисперсии ошибок σ^2 в классической модели парной регрессии $Y_t = \beta_1 + \beta_2 X_t + \varepsilon_t$, $t = 1, \ldots, n$, $\varepsilon_t \sim N(0, \sigma^2)$.

а) Найдите дисперсию и среднеквадратичное отклонение ($\mathrm{MSE}(\widehat{\theta}) = \mathrm{E}((\hat{\theta} - \theta)^2)$) каждой из двух оценок.

б) Какая из двух оценок обладает наименьшей дисперсией? Наименьшим среднеквадратичным отклонением?

2.16. Так называемая кривая Филлипса описывает связь темпа роста зарплаты и уровня безработицы. А именно,

$$\delta w_t = \beta_1 + \beta_2 \frac{1}{u_t} + \varepsilon_t,$$

где w_t — уровень заработной платы, $\delta w_t = 100(w_t - w_{t-1})/w_{t-1}$ — темп роста зарплаты (в процентах) и u_t — процент безработных в год t. Теория предполагает, что $\beta_1 < 0$ и $\beta_2 > 0$.

Используя данные для страны из таблицы 2.2, ответьте на следующие вопросы:

а) Найдите оценки коэффициентов уравнения и проверьте наличие значимой связи между δw и u.

б) Найдите «естественный уровень безработицы», т. е. такой уровень безработицы, при котором $\delta w = 0$.

в) Когда изменения в уровне безработицы оказывали наибольшее (наименьшее) влияние на темп изменения зарплаты?

г) Найдите 95%-ные доверительные интервалы для β_1 и β_2.

Таблица 2.2

Год t	w_t	u_t	Год t	w_t	u_t
1	1.62	1.0	10	2.66	1.8
2	1.65	1.4	11	2.73	1.9
3	1.79	1.1	12	2.80	1.5
4	1.94	1.5	13	2.92	1.4
5	2.03	1.5	14	3.02	1.8
6	2.12	1.2	15	3.13	1.1
7	2.26	1.0	16	3.28	1.5
8	2.44	1.1	17	3.43	1.3
9	2.57	1.3	18	3.58	1.4

2.17. В таблице 2.3 представлены расходы на агрегированное потребление Y и агрегированный располагаемый доход X в некоторой национальной экономике в течение 12 лет с 1986 по 1997 г.

а) Изобразите графически зависимость Y от X и определите, есть ли приближенная линейная зависимость Y от X.

б) Вычислите парную регрессию агрегированного потребления Y на X по данным, представленным в таблице 2.3.

в) Вычислите $s^2, s^2_{\widehat{\alpha}}, s^2_{\widehat{\beta}}$.

Таблица 2.3

Год	t	Y_t	X_t	Год	t	Y_t	X_t
1986	1	152	170	1992	7	177	200
1987	2	159	179	1993	8	179	207
1988	3	162	187	1994	9	184	215
1989	4	165	189	1995	10	186	216
1990	5	170	193	1996	11	190	220
1991	6	172	199	1997	12	191	225

2.18. Рассмотрим регрессию, построенную в упражнении 2.17.

а) Сформулируйте нулевую (основную) и альтернативную гипотезы при проверке статистической значимости коэффициентов регрессии.

б) Какое распределение имеют оценки $\widehat{\alpha}$ и $\widehat{\beta}$?

в) Какое распределение используется при проверке статистической значимости α и β?

г) Чему равно число степеней свободы?

д) Проверьте на 5%-ном уровне значимость коэффициентов α и β.

е) Постройте 95%-ный доверительный интервал для коэффициентов α и β в регрессии упражнения 2.17.

ж) Вычислите коэффициент детерминации, используя равенства $R^2 = \mathrm{RSS}/\mathrm{TSS}$ и $R^2 = 1 - \mathrm{ESS}/\mathrm{TSS}$.

2.19. Дана модель парной регрессии $Y_t = \alpha + \beta X_t + \varepsilon_t$, $t = 1, \ldots, n$, для которой выполнены стандартные условия классической линейной модели. Известно, что $n = 2m$. Все множество наблюдений (Y_t, X_t) разбито на две группы a и b по m наблюдений в каждой группе. Обозначим $\overline{X}_a, \overline{X}_b, \overline{Y}_a, \overline{Y}_b$ выборочные средние наблюдений X, Y по группам a, b, соответственно. В качестве оценки параметра β берется величина $\widetilde{\beta} = (\overline{Y}_a - \overline{Y}_b)/(\overline{X}_a - \overline{X}_b)$.

а) Найдите $\mathrm{E}(\widetilde{\beta})$ и $\mathrm{V}(\widetilde{\beta})$.

б) Каким должно быть разбиение наблюдений на группы a и b, чтобы дисперсия $\mathrm{V}(\widetilde{\beta})$ была минимальной?

2.20. Пусть $Y_t = \beta X_t + \varepsilon_t$, $t = 1, \ldots, n$, где $\mathrm{E}(\varepsilon_t) = 0$, и матрица ковариаций вектора $\boldsymbol{\varepsilon}$ известна. При каких условиях оценки

$$\frac{1}{n}\sum_{t=1}^{n}\frac{Y_t}{X_t} \quad \text{и} \quad \frac{\sum_{t=1}^{n} Y_t}{\sum_{t=1}^{n} X_t}$$

являются наилучшими среди несмещенных линейных оценок параметра β?

2.21. Проведены две регрессии:

$$Y_t = \alpha + \beta X_t + \varepsilon_t \quad \text{и} \quad Y_t = \alpha' + \beta' x_t + \varepsilon'_t, \quad t = 1, \ldots, T,$$

где $x_t = X_t - \overline{X}$.

а) По известным МНК-оценкам $\widehat{\alpha}$, $\widehat{\beta}$ параметров α, β в первой регрессии найдите МНК-оценки $\widehat{\alpha}'$, $\widehat{\beta}'$ параметров α', β' во второй регрессии.

б) Найдите $\mathrm{Cov}(\widehat{\alpha}', \widehat{\beta}')$.

2.22. В таблице 2.4 приведены ежегодные значения денежной массы и национального дохода некоторой гипотетической страны (все величины выражены в миллиардах кварков (название национальной валюты)).

Таблица 2.4

Год	Денежная масса	Нац. доход	Год	Денежная масса	Нац. доход
1981	2.0	5.0	1986	4.0	7.7
1982	2.5	5.5	1987	4.2	8.4
1983	3.2	6.0	1988	4.6	9.0
1984	3.6	7.0	1989	4.8	9.7
1985	3.3	7.2	1990	5.0	10.0

а) Проведите регрессию национального дохода (Y) на денежную массу (X) и константу.

б) Постройте 95%-ный доверительный интервал для оцениваемых параметров. Можете ли вы отвергнуть гипотезу $\beta = 0$? $\beta = 1$?

2.23. Два исследователя, работая независимо друг от друга, изучают одну и ту же регрессионную модель

$$Y_t = \alpha + \beta X_t + \varepsilon_t,$$

для которой выполнены все условия классической модели. В таблице 2.5 приведены результаты, полученные ими на основе независимых выборок:

Таблица 2.5

Выборка I	Выборка II
$n = 20$	$n = 20$
$\sum X_t = 100$	$\sum X_t = 200$
$\sum X_t^2 = 600$	$\sum X_t^2 = 2400$
$\sum Y_t = 500$	$\sum Y_t = 500$
$\widehat{\beta}_{\mathrm{I}} = 2$	$\widehat{\beta}_{\mathrm{II}} = 2.5$

Узнав о работе друг друга, они решают вывести единую оценку параметра β. Первый исследователь предлагает взять

$$\widetilde{\beta} = \frac{1}{2}(\widehat{\beta}_{\mathrm{I}} + \widehat{\beta}_{\mathrm{II}}).$$

Второй исследователь считает, что весовые коэффициенты первой и второй оценок выбраны неэффективно, и можно построить несмещенную оценку с меньшей дисперсией. Научный руководитель этих исследователей утверждает, что он знает способ еще улучшить общую оценку.

а) Какую оценку предлагает использовать второй исследователь?

б) Какую оценку предлагает использовать научный руководитель?

Оцените улучшение точности оценок пп. а), б) по сравнению с оценкой $\widetilde{\beta}$.

2.24. Предположим, что модель $Y_t = \alpha + \beta X_t + \varepsilon_t$, $t = 1, \ldots, n$, удовлетворяет условиям классической регрессии. Пусть $\widehat{\alpha}$, $\widehat{\beta}$ — оценки метода наименьших квадратов. Оценка $\widetilde{\beta}$ получена по методу наименьших квадратов при дополнительном (вообще говоря, неверном) предположении, что $\alpha = 0$.

а) Найдите МНК-оценку $\widetilde{\beta}$. При каких условиях она является несмещенной оценкой параметра β?

б) Найдите дисперсию оценки $\widetilde{\beta}$, сравните ее с дисперсией оценки $\widehat{\beta}$.

в) Обсудите, какую из двух оценок лучше использовать.

2.25. Рассмотрим модель парной регрессии $Y_t = \alpha + \beta X_t + \varepsilon_t$. Пусть $Z_t = X_t^2$. Рассмотрим следующую оценку параметра β:

$$\widetilde{\beta} = \frac{\sum_{t=1}^{n}(Z_t - \overline{Z})Y_t}{\sum_{t=1}^{n}(Z_t - \overline{Z})X_t}.$$

а) Покажите, что оценка $\widetilde{\beta}$ несмещенная.

б) Найдите дисперсию оценки $\widetilde{\beta}$.

в) Не повторяя доказательство теоремы Гаусса–Маркова, непосредственно проверьте, что $\mathrm{V}(\widetilde{\beta}) \geqslant \mathrm{V}(\widehat{\beta}_{\mathrm{OLS}})$.

Глава 3

Модель множественной регрессии

Естественным обобщением линейной регрессионной модели с двумя переменными (см. п. 2.3) является *многомерная регрессионная модель* (*multiple regression model*), или *модель множественной регрессии*:

$$y_t = \beta_1 + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + \varepsilon_t, \qquad t = 1, \ldots, n,$$

или

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + \varepsilon_t, \qquad t = 1, \ldots, n, \tag{3.1}$$

где x_{tp} — значения регрессора x_p в наблюдении t, а $x_{t1} = 1$, $t = 1, \ldots, n$. С учетом этого замечания мы не будем далее различать модели вида (3.1) со свободным членом или без свободного члена.

3.1. Основные гипотезы

Гипотезы, лежащие в основе модели множественной регрессии, являются естественным обобщением модели парной регрессии (см. п. 2.3):

1. $y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + \varepsilon_t$, $t = 1, \ldots, n$ — спецификация модели.

2. $x_{t1}, \ldots, x_{tk}$ — детерминированные величины. Векторы $\boldsymbol{x}_s = (x_{1s}, \ldots, x_{ns})'$, $s = 1, \ldots, k$ линейно независимы в R^n.

3a. $\mathrm{E}\varepsilon_t = 0$, $\mathrm{E}(\varepsilon_t^2) = \mathrm{V}(\varepsilon_t) = \sigma^2$ — не зависит от t.

3b. $\mathrm{E}(\varepsilon_t \varepsilon_s) = 0$ при $t \neq s$ — статистическая независимость (некоррелированность) ошибок для разных наблюдений.

Часто добавляется следующее условие.

3c. Ошибки ε_t, $t = 1, \ldots, n$ имеют совместное нормальное распределение: $\varepsilon_t \sim N(0, \sigma^2)$.

В этом случае модель называется *нормальной линейной регрессионной* (*classical normal linear regression model*).

Гипотезы, лежащие в основе множественной регрессии, удобно записать в матричной форме, которая главным образом и будет использоваться в дальнейшем.

Пусть $\boldsymbol{y}$ обозначает $n \times 1$ матрицу (вектор-столбец) $(y_1, \ldots, y_n)'$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$ — $k \times 1$ вектор коэффициентов; $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)'$ — $n \times 1$ вектор ошибок;

$\boldsymbol{X} = \begin{bmatrix} x_{11} & \ldots & x_{1k} \\ \ldots & \ldots & \ldots \\ x_{n1} & \ldots & x_{nk} \end{bmatrix}$ — $n \times k$ матрицу объясняющих переменных.

Столбцами матрицы $\boldsymbol{X}$ являются $n \times 1$ векторы регрессоров $\boldsymbol{x}_s = (x_{1s}, \ldots, x_{ns})'$, $s = 1, \ldots, k$. Условия 1–3 в матричной записи выглядят следующим образом:

1. $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$ — спецификация модели;

2. $\boldsymbol{X}$ — детерминированная матрица, имеет максимальный ранг k;

3a,b. $\mathrm{E}(\boldsymbol{\varepsilon}) = \boldsymbol{0}$; $\mathrm{V}(\boldsymbol{\varepsilon}) = \mathrm{E}(\boldsymbol{\varepsilon\varepsilon}') = \sigma^2 \boldsymbol{I}_n$;

дополнительное условие:

3c. $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$, т. е. $\boldsymbol{\varepsilon}$ — нормально распределенный случайный вектор со средним $\boldsymbol{0}$ и матрицей ковариаций $\sigma^2 \boldsymbol{I}_n$ (нормальная линейная регрессионная модель).

3.2. Метод наименьших квадратов. Теорема Гаусса–Маркова

Как и в случае регрессионного уравнения с одной переменной (см. п. 2.2), целью метода является выбор вектора оценок $\widehat{\boldsymbol{\beta}}$, минимизирующего сумму квадратов остатков e_t (т. е. квадрат длины вектора остатков $\boldsymbol{e}$):

$$\boldsymbol{e} = \boldsymbol{y} - \widehat{\boldsymbol{y}} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}, \quad \mathrm{ESS} = \Sigma e_t^2 = \boldsymbol{e}'\boldsymbol{e} \to \min.$$

Выразим $\boldsymbol{e}'\boldsymbol{e}$ через $\boldsymbol{X}$ и $\boldsymbol{\beta}$:

$$\begin{aligned} \boldsymbol{e}'\boldsymbol{e} &= (\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) = \boldsymbol{y}'\boldsymbol{y} - \boldsymbol{y}'\boldsymbol{X}\widehat{\boldsymbol{\beta}} - \widehat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} + \widehat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}} \\ &= \boldsymbol{y}'\boldsymbol{y} - 2\widehat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} + \widehat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}}. \end{aligned} \tag{3.2}$$

Необходимые условия минимума ESS получаются дифференцированием (3.2) по вектору $\widehat{\boldsymbol{\beta}}$ (см. (ЛА.22), (ЛА.24)):

$$\frac{\partial \mathrm{ESS}}{\partial \widehat{\boldsymbol{\beta}}} = -2\boldsymbol{X}'\boldsymbol{y} + 2\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{0}, \tag{3.3}$$

откуда, учитывая обратимость матрицы $\boldsymbol{X}'\boldsymbol{X}$ в силу условия 2 (приложение ЛА, п. 10), находим оценку метода наименьших квадратов:

$$\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}. \tag{3.4}$$

(Сравните с аналогичной формулой (2.8), полученной для регрессионного уравнения с одной независимой переменной.)

Покажем, что, как и в случае одного регрессора, (3.3) означает, что вектор остатков $\boldsymbol{e}$ ортогонален всем независимым переменным $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ (столбцам матрицы $\boldsymbol{X}$). Условие $\boldsymbol{x}_1'\boldsymbol{e} = \cdots = \boldsymbol{x}_k'\boldsymbol{e} = \boldsymbol{0}$ эквивалентно равенству $\boldsymbol{X}'\boldsymbol{e} = \boldsymbol{0}$. Действительно,

$$\begin{aligned} \boldsymbol{X}'\boldsymbol{e} &= \boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}) = \boldsymbol{X}'\boldsymbol{y} - \boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}} \\ &= \boldsymbol{X}'\boldsymbol{y} - \boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{0}. \end{aligned} \tag{3.5}$$

Получим полезную в дальнейшем формулу для суммы квадратов остатков

$$\begin{aligned} \boldsymbol{e}'\boldsymbol{e} &= \boldsymbol{y}'\boldsymbol{y} - 2\widehat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y} + \widehat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{X}\widehat{\boldsymbol{\beta}} \\ &= \boldsymbol{y}'\boldsymbol{y} - \widehat{\boldsymbol{\beta}}'(2\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}) = \boldsymbol{y}'\boldsymbol{y} - \widehat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{y}. \end{aligned} \tag{3.6}$$

Геометрическая интерпретация в основном совпадает с геометрической интерпретацией регрессионного уравнения с одной независимой переменной (см. п. 2.2). Представим $\boldsymbol{y}, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ как векторы в n-мерном евклидовом пространстве R^n. Векторы $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ порождают k-мерное подпространство π.

Вектор $\widehat{\boldsymbol{y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}$ есть ортогональная проекция вектора $\boldsymbol{y}$ на гиперплоскость π.

Вектор остатков $\boldsymbol{e} = \boldsymbol{y} - \widehat{\boldsymbol{y}}$ ортогонален подпространству π.

Как и в случае регрессионного уравнения с одной независимой переменной (см. п. 2.4), можно показать, что оценка метода наименьших квадратов является оптимальной.

Теорема Гаусса–Маркова. *Предположим, что:*

1. $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$;
2. $\boldsymbol{X}$ *— детерминированная* $n \times k$ *матрица, имеющая максимальный ранг* k;
3. $\mathrm{E}(\boldsymbol{\varepsilon}) = \boldsymbol{0}$; $\mathrm{V}(\boldsymbol{\varepsilon}) = \mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2 \boldsymbol{I}_n$.

Тогда оценка метода наименьших квадратов $\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ *является наиболее эффективной (в смысле наименьшей дисперсии) оценкой в классе линейных (по* $\boldsymbol{y}$*) несмещенных оценок (Best Linear Unbiased Estimator, BLUE).*

Доказательство. Обозначим $\boldsymbol{A} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$, $\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} = \boldsymbol{A}\boldsymbol{y}$. Любую другую линейную оценку вектора параметров $\boldsymbol{\beta}$ можно без ограничения общности представить в виде: $\boldsymbol{b} = (\boldsymbol{A} + \boldsymbol{C})\boldsymbol{y}$, где $\boldsymbol{C}$ — некоторая $k \times n$ матрица.

1. Покажем, что МНК-оценка (3.4) является несмещенной оценкой $\boldsymbol{\beta}$:

$$\begin{aligned}\mathrm{E}\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} &= \mathrm{E}\big((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}\big) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathrm{E}(\boldsymbol{y}) \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathrm{E}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathrm{E}\boldsymbol{\varepsilon} = \boldsymbol{\beta}. \qquad (3.7)\end{aligned}$$

Из условия несмещенности оценки $\boldsymbol{b}$ получаем, что для всех $\boldsymbol{\beta}$ справедливо соотношение

$$\boldsymbol{\beta} = \mathrm{E}\boldsymbol{b} = (\boldsymbol{A} + \boldsymbol{C})\mathrm{E}\boldsymbol{y} = (\boldsymbol{A} + \boldsymbol{C})\boldsymbol{X}\boldsymbol{\beta} = (\boldsymbol{I} + \boldsymbol{C}\boldsymbol{X})\boldsymbol{\beta},$$

откуда следует, что $\boldsymbol{C}\boldsymbol{X} = \boldsymbol{0}$.

2. Подсчитаем матрицу ковариаций МНК-оценки:

$$\begin{aligned}\mathrm{V}(\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}) &= \mathrm{V}(\boldsymbol{A}\boldsymbol{y}) = \boldsymbol{A}\mathrm{V}(\boldsymbol{y})\boldsymbol{A}' = \boldsymbol{A}\sigma^2\boldsymbol{I}\boldsymbol{A}' \\ &= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} \qquad (3.8)\end{aligned}$$

(здесь мы использовали симметричность матрицы $\boldsymbol{X}'\boldsymbol{X}$ и свойство матрицы ковариаций (МС.9)).

3. Используя полученное выше равенство $\boldsymbol{C}\boldsymbol{X} = \boldsymbol{0}$, получаем

$$\begin{aligned}\boldsymbol{b} - \boldsymbol{\beta} &= (\boldsymbol{A} + \boldsymbol{C})\boldsymbol{y} - \boldsymbol{\beta} = (\boldsymbol{A} + \boldsymbol{C})\boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{A} + \boldsymbol{C})\boldsymbol{\varepsilon} - \boldsymbol{\beta} \\ &= \boldsymbol{A}\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{\beta} + \boldsymbol{C}\boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{A} + \boldsymbol{C})\boldsymbol{\varepsilon} = (\boldsymbol{A} + \boldsymbol{C})\boldsymbol{\varepsilon},\end{aligned}$$

т. к. $\boldsymbol{C}\boldsymbol{X} = \boldsymbol{0}$ и $\boldsymbol{A}\boldsymbol{X} = \boldsymbol{I}$. Вычислим теперь матрицу ковариаций вектора $\boldsymbol{b}$:

$$\begin{aligned}\mathrm{V}(\boldsymbol{b}) &= \mathrm{E}\big((\boldsymbol{b} - \boldsymbol{\beta})(\boldsymbol{b} - \boldsymbol{\beta})'\big) = \mathrm{E}\big((\boldsymbol{A} + \boldsymbol{C})\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'(\boldsymbol{A} + \boldsymbol{C})'\big) \\ &= (\boldsymbol{A} + \boldsymbol{C})\sigma^2\boldsymbol{I}(\boldsymbol{A} + \boldsymbol{C})' = \sigma^2(\boldsymbol{A}\boldsymbol{A}' + \boldsymbol{C}\boldsymbol{A}' + \boldsymbol{A}\boldsymbol{C}' + \boldsymbol{C}\boldsymbol{C}') \\ &= \sigma^2\big((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} + \boldsymbol{C}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\ &\quad + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{C}' + \boldsymbol{C}\boldsymbol{C}'\big) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} + \sigma^2\boldsymbol{C}\boldsymbol{C}'.\end{aligned}$$

Таким образом (см. (3.8)),

$$\mathrm{V}(\boldsymbol{b}) = \mathrm{V}(\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}) + \sigma^2\boldsymbol{C}\boldsymbol{C}'.$$

Матрица $\boldsymbol{C}\boldsymbol{C}'$ неотрицательно определена (приложение ЛА, п. 15), поэтому

$$\mathrm{V}(\boldsymbol{b}) \geqslant \mathrm{V}(\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}). \qquad (3.9)$$

Отсюда следует утверждение теоремы. В самом деле, i-й диагональный элемент матрицы $\mathrm{V}(\boldsymbol{b})$ равен дисперсии i-й компоненты вектора коэффициентов $\mathrm{V}(b_i)$. Поэтому (приложение ЛА, п. 15) из (3.9) следует соответствующее неравенство для дисперсий оценок коэффициентов регрессии

$$\mathrm{V}(b_i) \geqslant \mathrm{V}(\widehat{\beta}_{\mathrm{OLS},i}),$$

что и требовалось доказать.

3.3. Статистические свойства МНК-оценок

Оценка дисперсии ошибок σ^2. Распределение s^2

Введем некоторые полезные в дальнейшем обозначения.

Вектор прогнозных значений

$$\widehat{y} = X\widehat{\beta} = X(X'X)^{-1}X'y = Ny, \quad N = X(X'X)^{-1}X'. \quad (3.10)$$

Вектор остатков регрессии

$$\begin{aligned} e &= y - \widehat{y} = y - X\widehat{\beta} = y - X(X'X)^{-1}X'y \\ &= (I - X(X'X)^{-1}X')y \\ &= (I - N)y = My, \quad M = I - N = I - X(X'X)^{-1}X'. \end{aligned} \quad (3.11)$$

Непосредственно из определения нетрудно проверить, что матрицы M, N *идемпотентны*, т. е. симметричны и являются проекторами:

$$N^2 = N, \qquad N' = N, \quad (3.12)$$

$$M^2 = M, \qquad M' = M. \quad (3.13)$$

В соответствии с геометрической интерпретацией регрессии из (3.10), (3.11) вытекает, что матрица N является матрицей оператора ортогонального проектирования на подпространство π, порожденное векторами x_i, а M — матрицей оператора ортогонального проектирования на $\pi^{\perp}$ — ортогональное дополнение к подпространству π в R^n. Поэтому

$$NX = X, \quad MX = 0. \quad (3.14)$$

Вычислим математическое ожидание и матрицу ковариаций вектора остатков e:

$$\begin{aligned} \mathrm{E}(e) &= \mathrm{E}(My) = M\mathrm{E}(y) = MX\beta \\ &= (I - X(X'X)^{-1}X')X\beta = X\beta - X\beta = 0, \end{aligned} \quad (3.15)$$

$$\mathrm{V}(e) = \mathrm{V}(My) = M\mathrm{V}(y)M' = M\sigma^2 IM' = \sigma^2 M. \quad (3.16)$$

Сумма квадратов остатков $\sum e_t^2 = e'e$ является естественным кандидатом на оценку дисперсии ошибок σ^2 (конечно, с некоторым поправочным коэффициентом, зависящим от числа степеней свободы):

$$\mathrm{E}(e'e) = \mathrm{tr}(\mathrm{V}(e)) = \sigma^2\,\mathrm{tr}(M) = \sigma^2\,\mathrm{tr}(I_n - N) = (n-k)\sigma^2. \quad (3.17)$$

При выводе (3.17) мы использовали (3.15), (3.16), свойства следа матрицы (приложение ЛА, п. 9), а также соотношение

$$\begin{aligned} \mathrm{tr}(N) &= \mathrm{tr}(X(X'X)^{-1}X') \\ &= \mathrm{tr}(X'X(X'X)^{-1}) = \mathrm{tr}(I_k) = k. \end{aligned} \quad (3.18)$$

При выводе последнего равенства используется свойство следа матрицы: $\mathrm{tr}(AB) = \mathrm{tr}(BA)$ (приложение ЛА, п. 9).

Из (3.17) следует, что

$$s^2 = \widehat{\sigma}^2 = \frac{e'e}{n-k} = \frac{\sum e_t^2}{n-k} \quad (3.19)$$

является несмещенной оценкой дисперсии ошибок σ^2, т. е. $\mathrm{E}s^2 = \sigma^2$.

Так как из (3.14) следует, что

$$e = My = M(X\beta + \varepsilon) = M\varepsilon \quad (3.20)$$

и $\mathrm{rank}(M) = \mathrm{rank}(I - N) = \mathrm{tr}(I - N) = n - k$ (ранг идемпотентной матрицы равен ее следу (приложение ЛА, п. 16)), то по лемме (приложение МС, п. 4, N8) распределение

$$\frac{e'e}{\sigma^2} \sim \chi^2(n-k) \quad \text{или} \quad (n-k)\frac{s^2}{\sigma^2} \sim \chi^2(n-k). \quad (3.21)$$

Независимость оценок $\widehat{\beta}$ и s^2

В предположении нормальной линейной множественной регрессионной модели удается доказать независимость оценок $\widehat{\beta}$ и s^2.

В самом деле, из (3.4) получаем

$$\widehat{\beta}_{\mathrm{OLS}} = (X'X)^{-1}X'(X\beta + \varepsilon) = \beta + (X'X)^{-1}X'\varepsilon = \beta + A\varepsilon. \quad (3.22)$$

Из (3.22) и (3.20) видно, что случайные векторы $\widehat{\boldsymbol{\beta}}$ и $\boldsymbol{e}$ имеют совместное многомерное нормальное распределение (приложение МС, п. 4). Поэтому для того чтобы доказать их независимость, достаточно показать их некоррелированность.

$$\begin{aligned}\boldsymbol{AM} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}') \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' - (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' = \boldsymbol{0},\end{aligned}$$

поэтому (т. к. $\mathrm{E}\boldsymbol{e} = \boldsymbol{0}$)

$$\mathrm{Cov}(\widehat{\boldsymbol{\beta}}, \boldsymbol{e}) = \mathrm{E}\big((\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\boldsymbol{e}'\big) = \mathrm{E}(\boldsymbol{A}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{M}) = \sigma^2\boldsymbol{AM} = \boldsymbol{0},$$

что и требовалось показать.

Так как s^2 является функцией от $\boldsymbol{e}$ (см. (3.19)), то оценки $\widehat{\boldsymbol{\beta}}$ и s^2 также независимы.

3.4. Анализ вариации зависимой переменной в регрессии. Коэффициенты R^2 и скорректированный R^2_{adj}

Как и в случае регрессионной модели с одной независимой переменной, вариацию $\sum(y_t - \overline{y})^2$ можно разбить на две части: объясненную регрессионным уравнением и необъясненную (т. е. связанную с ошибками ε) — см. (2.25):

$$\begin{aligned}\sum(y_t - \overline{y})^2 = &\sum(y_t - \widehat{y}_t)^2 \\ &+ \sum(\widehat{y}_t - \overline{y})^2 + 2\sum(y_t - \widehat{y}_t)(\widehat{y}_t - \overline{y}),\end{aligned} \tag{3.23}$$

или в векторной форме:

$$\begin{aligned}(\boldsymbol{y} - \overline{y}\boldsymbol{\imath})'(\boldsymbol{y} - \overline{y}\boldsymbol{\imath}) = &(\boldsymbol{y} - \widehat{\boldsymbol{y}})'(\boldsymbol{y} - \widehat{\boldsymbol{y}}) + (\widehat{\boldsymbol{y}} - \overline{y}\boldsymbol{\imath})'(\widehat{\boldsymbol{y}} - \overline{y}\boldsymbol{\imath}) \\ &+ 2(\boldsymbol{y} - \widehat{\boldsymbol{y}})'(\widehat{\boldsymbol{y}} - \overline{y}\boldsymbol{\imath}).\end{aligned} \tag{3.24}$$

Третье слагаемое в (3.24) равно нулю в случае, если константа, т. е. вектор $\boldsymbol{\imath} = (1, \ldots, 1)'$, принадлежит линейной оболочке векторов $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$. В самом деле,

$$(\boldsymbol{y} - \widehat{\boldsymbol{y}})'(\widehat{\boldsymbol{y}} - \overline{y}\boldsymbol{\imath}) = \boldsymbol{e}'(\boldsymbol{X}\widehat{\boldsymbol{\beta}} - \overline{y}\boldsymbol{\imath}) = \boldsymbol{e}'\boldsymbol{X}\widehat{\boldsymbol{\beta}} - \overline{y}\boldsymbol{e}'\boldsymbol{\imath} = 0,$$

т. к. в силу (3.5) $\boldsymbol{e}'\boldsymbol{X} = \boldsymbol{0}$ и $\overline{e} = \boldsymbol{e}'\boldsymbol{\imath}/n = 0$. Поэтому верно равенство

$$\underbrace{\|\boldsymbol{y} - \overline{y}\boldsymbol{\imath}\|^2}_{\text{TSS}} = \underbrace{\|\boldsymbol{y} - \widehat{\boldsymbol{y}}\|^2}_{\text{ESS}} + \underbrace{\|\widehat{\boldsymbol{y}} - \overline{y}\boldsymbol{\imath}\|^2}_{\text{RSS}}. \tag{3.25}$$

Записывая (3.25) в отклонениях $\boldsymbol{y}_* = \boldsymbol{y} - \overline{y}\boldsymbol{\imath}$; $\widehat{\boldsymbol{y}}_* = \widehat{\boldsymbol{y}} - \overline{y}\boldsymbol{\imath}$; опять получим теорему Пифагора:

$$\boldsymbol{y}_*'\boldsymbol{y}_* = \boldsymbol{e}'\boldsymbol{e} + \widehat{\boldsymbol{y}}_*'\widehat{\boldsymbol{y}}_*. \tag{3.26}$$

Как и ранее в (2.27), определим коэффициент детерминации R^2 как

$$R^2 = 1 - \frac{\text{ESS}}{\text{TSS}} = 1 - \frac{\boldsymbol{e}'\boldsymbol{e}}{\boldsymbol{y}_*'\boldsymbol{y}_*} = \frac{\widehat{\boldsymbol{y}}_*'\widehat{\boldsymbol{y}}_*}{\boldsymbol{y}_*'\boldsymbol{y}_*} = \frac{\text{RSS}}{\text{TSS}}. \tag{3.27}$$

Отметим, что коэффициент R^2 корректно определен только в том случае, если константа, т. е. вектор $\boldsymbol{\imath} = (1, \ldots, 1)'$, принадлежит линейной оболочке векторов $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$. В этом случае R^2 принимает значения из интервала $[0, 1]$.

Коэффициент R^2 показывает качество подгонки регрессионной модели к наблюденным значениям y_t.

Если $R^2 = 0$, то регрессия $\boldsymbol{y}$ на $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ не улучшает качество предсказания y_t по сравнению с тривиальным предсказанием $\widehat{y}_t = \overline{y}$.

Другой крайний случай $R^2 = 1$ означает точную подгонку: все $e_t = 0$, т. е. все точки наблюдений удовлетворяют уравнению регрессии.

В какой степени допустимо использовать критерий R^2 для выбора между несколькими регрессионными уравнениями? Следующие два замечания побуждают не полагаться *только* на значение R^2.

1. R^2, вообще говоря, возрастает при добавлении еще одного регрессора.
2. R^2 изменяется даже при простейшем преобразовании зависимой переменной (см. пример в конце раздела, стр. 76), поэтому сравнивать по значению R^2 можно только регрессии с одинаковыми зависимыми переменными.

Если взять число регрессоров равным числу наблюдений, всегда можно добиться того, что $R^2 = 1$, но это вовсе не будет означать наличие содержательной (имеющей экономический смысл) зависимости y от регрессоров.

Попыткой устранить эффект, связанный с ростом R^2 при возрастании числа регрессоров, является коррекция R^2 на число регрессоров. *Скорректированным* (*adjusted*) R^2 называется

$$R^2_{\text{adj}} = 1 - \frac{e'e/(n-k)}{y'_*y_*/(n-1)}. \qquad (3.28)$$

Заметим, что нет никакого существенного оправдания именно такого способа коррекции.

Свойства скорректированного R^2:

1. $R^2_{\text{adj}} = 1 - (1 - R^2)\dfrac{(n-1)}{(n-k)}$.
2. $R^2 \geqslant R^2_{\text{adj}}$, $k > 1$.
3. $R^2_{\text{adj}} \leqslant 1$, но может принимать значения < 0.

В определенной степени использование скорректированного коэффициента детерминации R^2_{adj} более корректно для сравнения регрессий при изменении количества регрессоров.

Например, рассмотрим две модели:

1. $y = X\beta + \varepsilon$.
2. $z = y - x_1 = X\gamma + \varepsilon$.

Строятся МНК-оценки параметров β и γ обеих моделей.

Для первой модели коэффициент детерминации R^2 равен

$$R^2_1 = 1 - \frac{e'e}{\sum(y_t - \overline{y})^2}, \quad e = My, \quad M = I - X(X'X)^{-1}X'. \qquad (3.29)$$

Подсчитаем коэффициент детерминации R^2 для второй модели. Обозначим $\delta = (1, 0, \ldots, 0)'$ — вектор-столбец; тогда $X\delta = x_1$. Матрица M одна и та же для обеих моделей, так как в них один и тот же набор регрессоров. Остатки во второй модели равны

$$\widetilde{e} = Mz = M(y - X\delta) = My - MX\delta = My = e$$

(мы использовали (3.14)). Таким образом, остатки в обеих моделях совпадают.

$$R^2_2 = 1 - \frac{e'e}{\sum(z_t - \overline{z})^2}, \qquad (3.30)$$

(3.29) и (3.30) отличаются только знаменателями y'_*y_* и z'_*z_*.

$$z'_*z_* = (y_* - x_{1*})'(y_* - x_{1*}) = y'_*y_* - 2y'_*x_{1*} + x'_{1*}x_{1*}. \qquad (3.31)$$

(Здесь $y_* = Ay$, $z_* = Az$, $x_{1*} = Ax_1$, где A — $n \times n$ матрица взятия отклонения от среднего, $A = I - \imath\imath'/n$). Из (3.31) видно, что коэффициенты детерминации, вообще говоря, не совпадают. Оценки коэффициентов двух регрессий связаны естественным соотношением:

$$\widehat{\gamma} = (X'X)^{-1}X'z = (X'X)^{-1}X'(y - X\delta) = \widehat{\beta} - \delta,$$

т. е. фактически обоим уравнениям соответствует одна и та же геометрическая картинка и экономически содержательная ситуация. Однако коэффициенты R^2 не совпадают только потому, что зависимость сформулирована в разных координатах.

Что «лучше»: y или $\widehat{y}$?

В качестве значений зависимой переменной в момент t мы можем использовать y_t или, например, прогноз $\widehat{y}_t$. Матрица ковариаций вектора y по условию модели равна $\mathrm{V}(y) = \sigma^2 I_n$. Матрица ковариаций вектора прогноза равна

$$\mathrm{V}(\widehat{y}) = \mathrm{V}(N\varepsilon) = \sigma^2 NN' = \sigma^2 N.$$

Таким образом,

$$\mathrm{V}(y) - \mathrm{V}(\widehat{y}) = \sigma^2(I - N) = \sigma^2 M.$$

Матрица M идемпотентная, поэтому, имея собственные значения только 0 или 1 (приложение ЛА, п. 16), неотрицательно определена, т. е.

$$\mathrm{V}(y) - \mathrm{V}(\widehat{y}) = \sigma^2 M \geqslant 0, \quad \text{или} \quad \mathrm{V}(y) \geqslant \mathrm{V}(\widehat{y}). \qquad (3.32)$$

Из (3.32) следует аналогичное неравенство для дисперсий наблюденных и предсказанных значений

$$\mathrm{V}(y_t) \geqslant \mathrm{V}(\widehat{y}_t). \tag{3.33}$$

Таким образом, как это ни парадоксально, в качестве значения зависимой переменной зачастую лучше брать предсказанное по модели значение, а не фактически наблюденное. При этом, естественно, предполагается, что наблюдаемые значения y_t действительно удовлетворяют соотношению $\boldsymbol{y} = \boldsymbol{X\beta}+\boldsymbol{\varepsilon}$, т.е. порождаются рассматриваемой моделью.

3.5. Проверка гипотез. Доверительные интервалы и доверительные области

Проверка гипотезы H_0: $\beta_i = \beta_{i0}$. Итак, мы доказали следующие статистические результаты.

1. Вектор оценок $\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ имеет нормальное распределение со средним $\boldsymbol{\beta}$ и матрицей ковариаций (см. (3.8)) $\mathrm{V}(\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$, т.е. (см. (3.7)) $\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} - \boldsymbol{\beta} \sim N(\boldsymbol{0}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1})$ или $\widehat{\beta}_{\mathrm{OLS},i} - \beta_i \sim N(0, \sigma^2_{\widehat{\beta}_i})$, где $\sigma^2_{\widehat{\beta}_i} = \sigma^2 q^{ii}$; q^{ii} — i-й диагональный элемент матрицы $(\boldsymbol{X}'\boldsymbol{X})^{-1}$. В качестве оценки дисперсии $\widehat{\beta}_{\mathrm{OLS},i}$ возьмем $s^2_{\widehat{\beta}_i} = \widehat{\sigma}^2_{\widehat{\beta}_i} = \widehat{\sigma}^2 q^{ii} = s^2 q^{ii}$.

2. Случайная величина $(n-k)\dfrac{s^2}{\sigma^2}$ распределена по закону хи-квадрат с $n-k$ степенями свободы $\chi^2(n-k)$ (см. (3.21)).

3. Оценки $\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ и s^2 независимы (см. п. 3.3).

Отсюда получаем (см. приложение МС, п. 3):

$$t = \frac{\widehat{\beta}_{\mathrm{OLS},i} - \beta_i}{s_{\widehat{\beta}_i}} = \frac{(\widehat{\beta}_{\mathrm{OLS},i} - \beta_i)/\sigma_{\widehat{\beta}_i}}{s_{\widehat{\beta}_i}/\sigma_{\widehat{\beta}_i}} \sim t(n-k). \tag{3.34}$$

Из (3.34) получаем, что

$$[\widehat{\beta}_{\mathrm{OLS},i} - t_c s_{\widehat{\beta}_i}; \widehat{\beta}_{\mathrm{OLS},i} + t_c s_{\widehat{\beta}_i}]$$

является $100(1-\alpha)\%$-ным *доверительным интервалом* для истинного значения коэффициента β_i, где $t_c = t_{\alpha/2}(n-k)$ — $100(\alpha/2)\%$-ная точка распределения Стьюдента с $n-k$ степенями свободы.

Для тестирования нулевой гипотезы H_0: $\beta_i = \beta_{i0}$, также можно применить статистику (3.34), а именно, нулевая гипотеза отклоняется на уровне значимости α, если

$$|t| = \left|\frac{\widehat{\beta}_{\mathrm{OLS},i} - \beta_{i0}}{s_{\widehat{\beta}_i}}\right| > t_{\alpha/2}(n-k).$$

Проверка гипотезы H_0: $\beta_2 = \beta_3 = \cdots = \beta_k = 0$. Предположим, что в число регрессоров включена константа (свободный член): $y_t = \beta_1 + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + \varepsilon_t$. Нулевая гипотеза состоит в том, что коэффициенты при всех регрессорах равны нулю.

Рассмотрим статистику

$$F = \frac{R^2}{1-R^2}\frac{n-k}{k-1} = \frac{\mathrm{RSS}}{\mathrm{ESS}}\frac{n-k}{k-1}$$
$$= \frac{\sum(\widehat{y}_t - \overline{y})^2/(k-1)}{\sum e_t^2/(n-k)} = \frac{\dfrac{\widehat{\boldsymbol{y}}'_*\widehat{\boldsymbol{y}}_*}{\sigma^2}\dfrac{1}{k-1}}{\dfrac{\boldsymbol{e}'\boldsymbol{e}}{\sigma^2}\dfrac{1}{n-k}}. \tag{3.35}$$

Как мы показали ранее, знаменатель в (3.35) имеет распределение $\dfrac{1}{n-k}\chi^2(n-k)$.

Покажем, что числитель имеет распределение $\dfrac{1}{k-1}\chi^2(k-1)$. В самом деле, $\widehat{\boldsymbol{y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{N}\boldsymbol{y}$, где $\boldsymbol{N} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ — оператор ортогонального проектирования на подпространство π, порожденное векторами $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$. Операцию взятия отклонения от среднего $y_t - \overline{y}$ можно записать в матричной форме:

$$\boldsymbol{y}_* = \boldsymbol{y} - \overline{y}\boldsymbol{\imath} = \boldsymbol{y} - \left(\frac{1}{n}\sum y_t\right)\boldsymbol{\imath} = \boldsymbol{y} - \left(\frac{1}{n}\boldsymbol{\imath}'\right)\boldsymbol{y} = (\boldsymbol{I} - \boldsymbol{P})\boldsymbol{y},$$

где $\boldsymbol{P}$ — $n \times n$ матрица, $P_{ij} = 1/n$. $\boldsymbol{P}$ есть матрица ортогонального проектирования на вектор $\boldsymbol{\imath} = (1, \ldots, 1)'$ (константа). Поскольку по нашему предположению вектор $\boldsymbol{\imath}$ принадлежит подпространству π, то $\boldsymbol{PN} = \boldsymbol{P}$. Последовательное ортогональное проектирование вектора $\boldsymbol{y}$ на π и затем на вектор $\boldsymbol{\imath}$ совпадает с ортогональным проектированием вектора $\boldsymbol{y}$ на вектор $\boldsymbol{\imath}$ (теорема о трех перпендикулярах).

Замечание. Отсюда следует, что $\overline{\widehat{y}}\boldsymbol{\imath} = \boldsymbol{P}\widehat{\boldsymbol{y}} = \boldsymbol{PNy} = \boldsymbol{Py} = \overline{y}\boldsymbol{\imath}$ и $\widehat{\boldsymbol{y}}_* = \widehat{\boldsymbol{y}} - \overline{\widehat{y}}\boldsymbol{\imath} = \widehat{\boldsymbol{y}} - \overline{y}\boldsymbol{\imath} = (\boldsymbol{N} - \boldsymbol{P})\boldsymbol{y} = (\boldsymbol{N} - \boldsymbol{P})(\boldsymbol{X\beta} + \boldsymbol{\varepsilon}) = \boldsymbol{X}_*\boldsymbol{\beta} + (\boldsymbol{N} - \boldsymbol{P})\boldsymbol{\varepsilon}$.

Здесь согласно (3.14) $\boldsymbol{NX} = \boldsymbol{X}$, поэтому $(\boldsymbol{N} - \boldsymbol{P})\boldsymbol{X} = \boldsymbol{X} - \boldsymbol{PX} = \boldsymbol{X}_*$, где $\boldsymbol{X}_*$ есть $n \times k$ матрица с нулевым первым столбцом. Поэтому при гипотезе H_0 имеем $\boldsymbol{X}_*\boldsymbol{\beta} = \boldsymbol{0}$ и $\widehat{\boldsymbol{y}}_* = (\boldsymbol{N} - \boldsymbol{P})\boldsymbol{\varepsilon}$. Матрица $\boldsymbol{N} - \boldsymbol{P}$ является идемпотентной: она, очевидно, симметричная и $(\boldsymbol{N} - \boldsymbol{P})^2 = \boldsymbol{N}^2 - \boldsymbol{PN} - \boldsymbol{NP} + \boldsymbol{P}^2 = \boldsymbol{N} - \boldsymbol{P} - \boldsymbol{N}'\boldsymbol{P}' + \boldsymbol{P} = \boldsymbol{N} - (\boldsymbol{PN})' = \boldsymbol{N} - \boldsymbol{P}$. Ранг идемпотентной матрицы равен ее следу (приложение ЛА, п. 16), поэтому $\operatorname{rank}(\boldsymbol{N} - \boldsymbol{P}) = \operatorname{tr}(\boldsymbol{N} - \boldsymbol{P}) = k - 1$ (см. (3.18)). Таким образом, из леммы (приложение МС, п. 4, N8) получаем: $\widehat{\boldsymbol{y}}_*'\widehat{\boldsymbol{y}}_* / \sigma^2 \sim \chi^2(k-1)$, что и требовалось показать.

Как мы установили ранее в п. 3.3, $\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ и $\boldsymbol{e}$ независимы, поэтому статистика F из (3.35) имеет распределение Фишера

$$F = \frac{R^2}{1 - R^2}\frac{n-k}{k-1} = \frac{\mathrm{RSS}}{\mathrm{ESS}}\frac{n-k}{k-1} = \frac{\widehat{\boldsymbol{y}}_*'\widehat{\boldsymbol{y}}_*/(k-1)}{\boldsymbol{e}'\boldsymbol{e}/(n-k)} \sim F(k-1, n-k) \quad (3.36)$$

и ее можно использовать для проверки гипотезы H_0: $\beta_2 = \beta_3 = \cdots = \beta_k = 0$. А именно, гипотеза H_0 отвергается, например, на 5%-ном уровне значимости, если $F > F_c$, где $F_c = F_{0.05}(k-1, n-k)$ — 5%-ная точка распределения Фишера $F(k-1, n-k)$.

Линейное ограничение общего вида H_0: $\boldsymbol{H\beta} = \boldsymbol{r}$. Пусть $\boldsymbol{H}$ — $q \times k$ матрица, $\boldsymbol{\beta}$ — $k \times 1$ вектор коэффициентов, $\boldsymbol{r}$ — $q \times 1$ вектор.

Естественно считать, что число ограничений не превосходит числа параметров и ограничения линейно независимы, т. е. $q \leqslant k$ и матрица $\boldsymbol{H}$ имеет максимальный ранг: $\operatorname{rank}(\boldsymbol{H}) = q$.

В качестве примера рассмотрим следующие матрицы $\boldsymbol{H}, \boldsymbol{r}$ для $k = 3$, $q = 2$:

$$\boldsymbol{H\beta} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix} = \boldsymbol{r}.$$

Это условие соответствует системе двух линейных ограничений:

$$\begin{cases} \beta_1 = 2, \\ \beta_2 - \beta_3 = 0. \end{cases}$$

Из (3.4), (3.7), (3.8) видно, что вектор $\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ имеет нормальное распределение со средним $\boldsymbol{\beta}$ и матрицей ковариаций $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$.

$$\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} \sim N(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}).$$

Отсюда получаем, что $\boldsymbol{H}\widehat{\boldsymbol{\beta}} - \boldsymbol{r} \sim N(\boldsymbol{H\beta} - \boldsymbol{r}, \boldsymbol{\Sigma})$, где $\boldsymbol{\Sigma}$ — $q \times q$ матрица и $\boldsymbol{\Sigma} = \mathrm{V}(\boldsymbol{H}\widehat{\boldsymbol{\beta}} - \boldsymbol{r}) = \mathrm{V}(\boldsymbol{H}\widehat{\boldsymbol{\beta}}) = \boldsymbol{H}\mathrm{V}(\widehat{\boldsymbol{\beta}})\boldsymbol{H}' = \sigma^2\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}'$. Итак,

$$\boldsymbol{H}\widehat{\boldsymbol{\beta}} - \boldsymbol{r} \sim N(\boldsymbol{H\beta} - \boldsymbol{r}, \sigma^2\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}'). \quad (3.37)$$

По лемме (приложение МС, п. 4, N9) из (3.37) при условии справедливости гипотезы H_0: $\boldsymbol{H\beta} = \boldsymbol{r}$, получаем:

$$\frac{1}{\sigma^2}(\boldsymbol{H}\widehat{\boldsymbol{\beta}} - \boldsymbol{r})'\left(\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}'\right)^{-1}(\boldsymbol{H}\widehat{\boldsymbol{\beta}} - \boldsymbol{r}) \sim \chi^2(q). \quad (3.38)$$

Из (3.21) и (3.38), используя независимость $\widehat{\boldsymbol{\beta}}$ и $\boldsymbol{e}$, получаем:

$$F = \frac{(\boldsymbol{H}\widehat{\boldsymbol{\beta}} - \boldsymbol{r})'\left(\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}'\right)^{-1}(\boldsymbol{H}\widehat{\boldsymbol{\beta}} - \boldsymbol{r})/q}{\boldsymbol{e}'\boldsymbol{e}/(n-k)} \sim F(q, n-k). \quad (3.39)$$

Если справедлива гипотеза H_0: $\boldsymbol{H\beta} - \boldsymbol{r} = \boldsymbol{0}$, то статистика F в (3.39) не должна принимать слишком больших значений, а именно, с вероятностью $1-\alpha$ имеем $F < F_\alpha(q, n-k)$, где $F_\alpha(q, n-k)$ есть 100α%-ная точка распределения Фишера $F(q, n-k)$.

Из (3.21), (3.37), снова используя независимость $\widehat{\boldsymbol{\beta}}$ и $\boldsymbol{e}$, получаем

$$F = \frac{(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{H}'\left(\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}'\right)^{-1}\boldsymbol{H}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})/q}{\boldsymbol{e}'\boldsymbol{e}/(n-k)} \sim F(q, n-k). \quad (3.40)$$

Условие $F < F_\alpha(q, n-k)$ задает $100(1-\alpha)\%$-ную *доверительную область* для коэффициентов $\boldsymbol{\beta}$.

Так как в числителе (3.40) стоит неотрицательно определенная квадратичная форма от β_i, то эта доверительная область является выпуклым множеством.

В случае $\boldsymbol{H} = \boldsymbol{I}$ статистика F в (3.40) выглядит следующим образом:

$$F = \frac{(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\boldsymbol{X}'\boldsymbol{X})(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})/k}{\boldsymbol{e}'\boldsymbol{e}/(n-k)} \sim F(k, n-k).$$

В этом случае доверительная область является эллипсоидом в k-мерном пространстве коэффициентов $\boldsymbol{\beta}$.

H_0: $\beta_{k-q+1} = \beta_{k-q+2} = \dots = \beta_k = 0$. Гипотеза является, конечно, частным случаем общей линейной гипотезы $\boldsymbol{H\beta} = \boldsymbol{r}$. Эту гипотезу приходится проверять наиболее часто, и она представляет самостоятельный интерес.

Разобьем все матрицы с размером по одной из сторон k на блоки со сторонами $(k-q)$ и q:

$$\boldsymbol{H} = [\boldsymbol{0} \;\; \boldsymbol{I}_q], \quad \widehat{\boldsymbol{\beta}} = \begin{bmatrix} \widehat{\boldsymbol{\beta}}_1 \\ \widehat{\boldsymbol{\beta}}_2 \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix}, \quad \boldsymbol{X} = [\boldsymbol{X}_1 \;\; \boldsymbol{X}_2],$$

$$\boldsymbol{H}\widehat{\boldsymbol{\beta}} = [\boldsymbol{0} \;\; \boldsymbol{I}_q] \begin{bmatrix} \widehat{\boldsymbol{\beta}}_1 \\ \widehat{\boldsymbol{\beta}}_2 \end{bmatrix} = \widehat{\boldsymbol{\beta}}_2, \quad \boldsymbol{X}\widehat{\boldsymbol{\beta}} = \boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 + \boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2,$$

здесь $\boldsymbol{X}_1 - n \times (k-q)$, $\boldsymbol{X}_2 - n \times q$, $\boldsymbol{\beta}_1, \widehat{\boldsymbol{\beta}}_1 - (k-q) \times 1$, $\boldsymbol{\beta}_2$, $\widehat{\boldsymbol{\beta}}_2 - q \times 1$ матрицы. Введем обозначения:

$$\boldsymbol{Q} = \boldsymbol{X}'\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1' \\ \boldsymbol{X}_2' \end{bmatrix} [\boldsymbol{X}_1 \;\; \boldsymbol{X}_2] = \begin{bmatrix} \boldsymbol{X}_1'\boldsymbol{X}_1 & \boldsymbol{X}_1'\boldsymbol{X}_2 \\ \boldsymbol{X}_2'\boldsymbol{X}_1 & \boldsymbol{X}_2'\boldsymbol{X}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{Q}_{11} & \boldsymbol{Q}_{12} \\ \boldsymbol{Q}_{21} & \boldsymbol{Q}_{22} \end{bmatrix},$$

$$\boldsymbol{Q}^{-1} = \begin{bmatrix} \boldsymbol{Q}^{11} & \boldsymbol{Q}^{12} \\ \boldsymbol{Q}^{21} & \boldsymbol{Q}^{22} \end{bmatrix}.$$

В этих обозначениях числитель дроби в выражении (3.40) для F при условии, что верна нулевая гипотеза H_0: $\boldsymbol{\beta}_2 = \boldsymbol{0}$, имеет вид

(с точностью до множителя $1/q$)

$$\begin{aligned}
&(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{I}_q \end{bmatrix} \left([\boldsymbol{0} \;\; \boldsymbol{I}_q] \; \boldsymbol{Q}^{-1} \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{I}_q \end{bmatrix} \right)^{-1} [\boldsymbol{0} \;\; \boldsymbol{I}_q] \; (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \\
&= \widehat{\boldsymbol{\beta}}_2'(\boldsymbol{Q}^{22})^{-1}\widehat{\boldsymbol{\beta}}_2 = \widehat{\boldsymbol{\beta}}_2'(\boldsymbol{Q}_{22} - \boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12})\widehat{\boldsymbol{\beta}}_2 \\
&= \widehat{\boldsymbol{\beta}}_2'(\boldsymbol{X}_2'\boldsymbol{X}_2 - \boldsymbol{X}_2'\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1'\boldsymbol{X}_2)\widehat{\boldsymbol{\beta}}_2 \\
&= \widehat{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'(\boldsymbol{I} - \boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1')\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2 = \widehat{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2.
\end{aligned} \tag{3.41}$$

Здесь через $\boldsymbol{M}_1$ обозначена матрица ортогонального проектирования на $\pi_1^\perp$ — ортогональное дополнение к подпространству π_1 в R^n (порожденному $k-q$ столбцами матрицы $\boldsymbol{X}_1$). При выводе (3.41) мы использовали формулу (ЛА.18): $(\boldsymbol{Q}^{22})^{-1} = \boldsymbol{Q}_{22} - \boldsymbol{Q}_{21}\boldsymbol{Q}_{11}^{-1}\boldsymbol{Q}_{12}$.

Докажем теперь, что (3.41) равно $\boldsymbol{e}^{*\prime}\boldsymbol{e}^* - \boldsymbol{e}'\boldsymbol{e}$, где $\boldsymbol{e}^*$ — остатки «короткой» регрессии (только на $\boldsymbol{X}_1$), а $\boldsymbol{e}$ — остатки «длинной» регрессии (на $\boldsymbol{X} = [\boldsymbol{X}_1 \, \boldsymbol{X}_2]$). В самом деле,

$$\begin{aligned}
\boldsymbol{e}^* &= \boldsymbol{M}_1\boldsymbol{y} = \boldsymbol{M}_1(\boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 + \boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2 + \boldsymbol{e}) \\
&= \boldsymbol{M}_1\boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1 + \boldsymbol{M}_1\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2 + \boldsymbol{M}_1\boldsymbol{e} = \boldsymbol{M}_1\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2 + \boldsymbol{e}
\end{aligned}$$

(здесь слагаемое $\boldsymbol{M}_1\boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1$ равно нулю, т.к. $\boldsymbol{M}_1$ — проектор на подпространство, ортогональное $\boldsymbol{X}_1$, и $\boldsymbol{M}_1\boldsymbol{e} = \boldsymbol{e}$, т.к. $\boldsymbol{e}$ ортогонально $\boldsymbol{X}_1$ и $\boldsymbol{X}_2$). Отсюда

$$\begin{aligned}
\boldsymbol{e}^{*\prime}\boldsymbol{e}^* &= (\boldsymbol{e} + \boldsymbol{M}_1\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2)'(\boldsymbol{e} + \boldsymbol{M}_1\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2) \\
&= \boldsymbol{e}'\boldsymbol{e} + \widehat{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2 + \boldsymbol{e}'\boldsymbol{M}_1\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2 + \widehat{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{e},
\end{aligned} \tag{3.42}$$

два последних слагаемых в (3.42) равны нулю, т.к. $\boldsymbol{M}_1\boldsymbol{e} = \boldsymbol{e}$, а $\boldsymbol{X}_2'\boldsymbol{e} = \boldsymbol{0}$ (остатки ортогональны регрессорам). Таким образом,

$$\boldsymbol{e}^{*\prime}\boldsymbol{e}^* - \boldsymbol{e}'\boldsymbol{e} = \widehat{\boldsymbol{\beta}}_2'\boldsymbol{X}_2'\boldsymbol{M}_1\boldsymbol{X}_2\widehat{\boldsymbol{\beta}}_2. \tag{3.43}$$

Полученное выражение совпадает с (3.41), поэтому статистику F из (3.40) в нашем случае можно записать как

$$F = \frac{(\boldsymbol{e}^{*\prime}\boldsymbol{e}^* - \boldsymbol{e}'\boldsymbol{e})/q}{\boldsymbol{e}'\boldsymbol{e}/(n-k)} = \frac{(\mathrm{ESS_R} - \mathrm{ESS_{UR}})/q}{\mathrm{ESS_{UR}}/(n-k)} \sim F(q, n-k). \tag{3.44}$$

Здесь $\mathrm{ESS_R}$ — сумма квадратов остатков «короткой» (*restricted*) регрессии; $\mathrm{ESS_{UR}}$ — сумма квадратов остатков «длинной» (*unrestricted*) регрессии.

Как и ранее (см. (2.32)), F-статистику (3.44) можно выразить через коэффициенты детерминации R^2 для «короткой» и «длинной» регрессий:

$$F = \frac{(R^2_{\mathrm{UR}} - R^2_{\mathrm{R}})/q}{(1 - R^2_{\mathrm{UR}})/(n-k)} \sim F(q, n-k) \tag{3.45}$$

(при выводе (3.45) используется определение R^2 (3.27) и то, что $\mathrm{TSS_{UR}} \equiv \mathrm{TSS_R}$).

Важное замечание. *Можно доказать, что представления F-статистики в формах* (3.44), (3.45) *справедливы и в общем случае произвольного линейного ограничения $\boldsymbol{H\beta} = \boldsymbol{r}$.*

В этом случае «длинной» регрессией является регрессия без ограничений на параметры $\boldsymbol{\beta}$, а «короткой» — регрессия с ограничениями $\boldsymbol{H\beta} = \boldsymbol{r}$. МНК при этом состоит в минимизации функции ESS (3.2) при условии $\boldsymbol{H\beta} = \boldsymbol{r}$.

Упражнение. Докажите, что представления (3.44), (3.45) для F-статистики справедливы в общем случае произвольного линейного ограничения $\boldsymbol{H\beta} = \boldsymbol{r}$. (*Указание*: линейной заменой регрессоров сведите общий случай к рассмотренному выше случаю ограничения вида $\beta_{k-q+1} = \beta_{k-q+2} = \cdots = \beta_k = 0$.)

$\mathrm{H_0}$: $\boldsymbol{c}'\boldsymbol{\beta} = \theta$, $\boldsymbol{c}$ — $k \times 1$ *вектор*. Эта гипотеза также является частным случаем общей линейной гипотезы $\boldsymbol{H\beta} = \boldsymbol{r}$ для $1 \times k$ матрицы $\boldsymbol{H} = \boldsymbol{c}'$. Соответственно для проверки гипотезы можно использовать F-статистику (3.39), которая в данном случае распределена по закону Фишера $F(1, n-k)$.

Эту гипотезу можно проверять, используя также t-статистику, аналогичную (3.34). Заметим, что $\boldsymbol{c}'\widehat{\boldsymbol{\beta}} \sim N(\boldsymbol{c}'\boldsymbol{\beta}, \sigma^2_{\boldsymbol{c}'\widehat{\boldsymbol{\beta}}})$, как линейная комбинация совместно нормально распределенных случайных величин. Дисперсия $\boldsymbol{c}'\widehat{\boldsymbol{\beta}}$ равна $\sigma^2_{\boldsymbol{c}'\widehat{\boldsymbol{\beta}}} = \mathrm{V}(\boldsymbol{c}'\widehat{\boldsymbol{\beta}}) = \boldsymbol{c}'\mathrm{V}(\widehat{\boldsymbol{\beta}})\boldsymbol{c} = \sigma^2\boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{c}$, оценка дисперсии $\widehat{\sigma}^2_{\boldsymbol{c}'\widehat{\boldsymbol{\beta}}} = \widehat{\sigma}^2\boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{c}$.

Из рассуждений, совершенно аналогичных рассуждениям при выводе t-статистики для гипотезы $\beta_i = \beta_{i0}$, получаем

$$t = \frac{\boldsymbol{c}'\widehat{\boldsymbol{\beta}} - \boldsymbol{c}'\boldsymbol{\beta}}{\widehat{\sigma}_{\boldsymbol{c}'\widehat{\boldsymbol{\beta}}}} = \frac{\boldsymbol{c}'\widehat{\boldsymbol{\beta}} - \boldsymbol{c}'\boldsymbol{\beta}}{\sqrt{s^2\boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{c}}} \sim t(n-k) \tag{3.46}$$

или, если справедлива гипотеза $\mathrm{H_0}$: $\boldsymbol{c}'\boldsymbol{\beta} = \theta$,

$$t = \frac{\boldsymbol{c}'\widehat{\boldsymbol{\beta}} - \theta}{\sqrt{s^2\boldsymbol{c}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{c}}} \sim t(n-k). \tag{3.47}$$

Конечно, в данном случае t-статистика эквивалентна F-статистике.

$\mathrm{H_0}$: $\boldsymbol{\beta}' = \boldsymbol{\beta}''$; $\sigma' = \sigma''$ (***тест Чоу (Chow)***). Предположим, у нас есть две выборки данных. По каждой выборке мы строим регрессионную модель. Вопрос, который нас интересует: верно ли, что эти две модели совпадают? Рассмотрим модели:

$$y_t = \beta'_1 x_{t1} + \beta'_2 x_{t2} + \cdots + \beta'_k x_{tk} + \varepsilon'_t, \quad t = 1, \ldots, n, \tag{3.48а}$$

$$y_t = \beta''_1 x_{t1} + \beta''_2 x_{t2} + \cdots + \beta''_k x_{tk} + \varepsilon''_t, \quad t = n+1, \ldots, n+m, \tag{3.48б}$$

в первой выборке n наблюдений и m наблюдений во второй. Например, y — заработная плата, x_i — регрессоры (возраст, стаж, уровень образования и т. п.), и пусть первая выборка относится к женщинам, вторая — к мужчинам. Вопрос: следует ли из оценки моделей (3.48а), (3.48б), что модель зависимости зарплаты от регрессоров одна и та же для мужчин и женщин?

Сведем эту ситуацию к общей схеме проверки линейных ограничений на параметры модели. Регрессией без ограничений здесь является объединение двух регрессий (3.48а), (3.48б), т. е. $\mathrm{ESS_{UR}} = \mathrm{ESS_1} + \mathrm{ESS_2}$, число степеней свободы при этом равно $(n-k) + (m-k) = n + m - 2k$. Предположим теперь, что верна нулевая гипотеза. Тогда регрессия с ограничениями записывается одним уравнением

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + \varepsilon_t, \qquad t = 1, \ldots, n+m. \tag{3.49}$$

Оценивая (3.49), получаем $\mathrm{ESS_R}$. Тогда, учитывая, что наложено k ограничений на параметры модели, получаем (ср. (3.44))

$$F = \frac{(\mathrm{ESS_R} - \mathrm{ESS_{UR}})/k}{\mathrm{ESS_{UR}}/(n+m-2k)} \sim F(k, n+m-2k). \qquad (3.50)$$

Если F-статистика (3.50) больше критического значения $F_c = F_\alpha(k, n+m-2k)$, то нулевая гипотеза отвергается на уровне значимости α. В этом случае мы не можем объединить две выборки в одну.

Рассмотрим пример исследования, использующего модель множественной регрессии.

Пример. ***Рынок квартир в Москве*** (см. Каргин, Онацкий, 1996). Данные для этого исследования собраны студентами РЭШ в 1994 и 1996 гг.

После проведенного анализа была выбрана логарифмическая форма модели, как более соответствующая данным:

$$\begin{aligned}\mathrm{LOGPRICE} = {} & \beta_0 + \beta_1\mathrm{LOGLIVSP} + \beta_2\mathrm{LOGPLAN} \\ & + \beta_3\mathrm{LOGKITSP} + \beta_4\mathrm{LOGDIST} + \beta_5\mathrm{FLOOR} + \beta_6\mathrm{BRICK} \\ & + \beta_7\mathrm{BAL} + \beta_8\mathrm{LIFT} + \beta_9\mathrm{R1} + \beta_{10}\mathrm{R2} + \beta_{11}\mathrm{R3} + \beta_{12}\mathrm{R4} + \varepsilon. \quad (*)\end{aligned}$$

Здесь LOGPRICE — логарифм цены квартиры (в долл. США), LOGLIVSP — логарифм жилой площади (в кв. м), LOGPLAN — логарифм площади нежилых помещений (в кв. м), LOGKITSP — логарифм площади кухни (в кв. м), LOGDIST — логарифм расстояния от центра Москвы (в км). Включены также бинарные, «фиктивные» переменные (см. п. 4.2), принимающие значения 0 или 1: FLOOR — принимает значение 1, если квартира расположена на первом или на последнем этаже, BRICK — принимает значение 1, если квартира находится в кирпичном доме, BAL — принимает значение 1, если в квартире есть балкон, LIFT — принимает значение 1, если в доме есть лифт, R1 — принимает значение 1 для однокомнатных квартир и 0 для всех остальных, R2, R3, R4 — аналогичные переменные для двух-, трех- и четырехкомнатных квартир.

Результаты оценивания уравнения (*) для 464 наблюдений, относящихся к 1996 г., приведены в таблице 3.1.

Таблица 3.1

Переменная	Коэффициент	Стандартная ошибка	t-статистика	P-значение
CONST	7.106	0.290	24.5	0.0000
LOGLIVSP	0.670	0.069	9.65	0.0000
LOGPLAN	0.431	0.049	8.71	0.0000
LOGKITSP	0.147	0.060	2.45	0.0148
LOGDIST	−0.114	0.016	−7.11	0.0000
BRICK	0.134	0.024	5.67	0.0000
FLOOR	−0.0686	0.021	−3.21	0.0014
LIFT	0.114	0.024	4.79	0.0000
BAL	0.042	0.020	2.08	0.0385
R1	0.214	0.109	1.957	0.0510
R2	0.140	0.080	1.75	0.0809
R3	0.164	0.060	2.74	0.0065
R4	0.169	0.054	3.11	0.0020

R^2=0.8921, R^2_{adj}=0.8892, стандартная ошибка регрессии 0.2013.

Из анализа t-статистик видно, что все коэффициенты, кроме коэффициентов при R1 и R2, значимы на 5%-ном уровне (см. (3.34)).

Коэффициент при LOGLIVSP, равный 0.67, означает, что увеличение жилой площади квартиры на 1% увеличивает ее цену на 0.67%. Иначе говоря, эластичность цены квартиры по жилой площади равна 0.67.

Несколько сложнее объяснить значение коэффициентов при LOGPLAN и LOGKITSP. Для их объяснения мы решили использовать следующий пример. Предположим, что есть две квартиры с одинаковой кухней, скажем 9 кв. м, но разными по площади остальными вспомогательными помещениями. Например, в первой квартире эта площадь равна 11 кв. м, а во второй 12 кв. м. Таким образом, во второй квартире общая площадь вспомогательных помещений (21 кв. м) на 5% больше, чем в первой. Такое увеличение площади, с фиксированной площадью кухни, в соответствии с нашей моделью должно привести к увеличению цены второй квартиры по сравнению с первой на $5 \cdot 0.431 = 2.15\%$. Теперь представим себе, что имеется третья квартира с кухней 10 кв. м и площадью остальных вспомогательных помещений 11 кв. м. Общая площадь вспомогательных помещений в такой квартире, как и в предыдущем случае, 21 кв. м. Однако теперь мы ожидаем увеличение цены третьей квартиры по сравнению с первой квартирой на

$5 \cdot 0.431 + 5 \cdot 0.147 = 2.89\%$, то есть увеличение площади вспомогательных помещений за счет кухни приводит к большему увеличению цены квартиры, чем такое же увеличение за счет, скажем, коридора.

Отрицательное значение коэффициента при LOGDIST (-0.114) означает, что увеличение расстояния от центра города на 1% уменьшает цену квартиры на 0.11%. Эксперты считают, что в действительности цена квартиры зависит также от «качества» района, в котором она расположена, а не только от ее расстояния от центра, однако влияние фактора «качества» не рассматривалось в данном исследовании.

Ниже (п. 4.2) мы вернемся к этому примеру для интерпретации коэффициентов при «фиктивных» переменных.

Существует мнение экспертов, что рынок квартир достаточно отчетливо делится на три сектора: рынок однокомнатных квартир, рынок квартир среднего размера (от 2 до 4 комнат) и рынок больших квартир. Для проверки этого утверждения тестируем с помощью F-статистики (3.39) гипотезу H_0, что коэффициенты при R2, R3, R4 равны: $\beta_{10} = \beta_{11}$; $\beta_{11} = \beta_{12}$. Получаем следующий результат:

F-статистика 0.22315 P-значение 0.8001,

который показывает, что мы не можем отвергнуть гипотезу, что для квартир с числом комнат 2–4 формулы (*) расчета цены совпадают. Однако тестирование гипотезы H_0: $\beta_9 = \beta_{10}$ о совпадении формул для одно- и двухкомнатных квартир дает следующее значение F-статистики:

F-статистика 3.03188 P-значение 0.0823,

С вероятностью ошибиться, меньшей 10%, можно отвергнуть гипотезу о совпадении формул (*) для одно- и двухкомнатных квартир.

Упражнения

3.1. Рассмотрим уравнения:

$$\ln y_t = \beta_1 + \beta_2 \ln w_t + \beta_3 s_t + \varepsilon_t,$$
$$\ln(y_t/w_t) = \gamma_1 + \gamma_2 \ln w_t + \gamma_3 s_t + \varepsilon_t, \qquad t = 1, \ldots, n,$$

где y_t — годовой доход t-го индивидуума, w_t — число его рабочих недель в году, s_t — полное число лет, потраченных им на образование.

а) Покажите, что для соответствующих МНК-оценок выполнены соотношения: $\widehat{\gamma}_1 = \widehat{\beta}_1$, $\widehat{\gamma}_3 = \widehat{\beta}_3$, $\widehat{\gamma}_2 = \widehat{\beta}_2 - 1$.

б) Покажите, что остатки этих регрессий совпадают.

в) При каких условиях коэффициент детерминации R^2 в первой регрессии будет больше коэффициента детерминации второй регрессии? Что при этом можно сказать о качестве подгонки?

3.2. Покажите, что в регрессии y_i на прогнозные значения $\widehat{y}_i$ и константу свободный член равен 0, а угловой коэффициент равен 1.

3.3. Дано регрессионное уравнение $y_t = \beta x_t + \varepsilon_t$, $t = 1, \ldots, T$. Ошибки ε_t — независимые одинаково распределенные нормальные величины. Мы хотим проверить гипотезу, что после наблюдения с номером n значение параметра β изменилось. Сумма квадратов остатков с ограничением $\mathrm{ESS_R}$ получается из регрессии y на x по всем T наблюдениям. Для нахождения $\mathrm{ESS_{UR}}$ используются две разные процедуры: 1) оцениваем суммы квадратов остатков регрессий по двум подпериодам $\mathrm{ESS_1}$ и $\mathrm{ESS_2}$, затем их складываем: $\mathrm{ESS_{UR}} = \mathrm{ESS_1} + \mathrm{ESS_2}$; 2) переписываем уравнение в виде

$$y_t = \beta_1 x_t d_{t1} + \beta_2 x_t d_{t2} + \varepsilon_t,$$

где

$$d_{t1} = \begin{cases} 1, & t = 1, \ldots, n, \\ 0, & t = n+1, \ldots, T, \end{cases} \qquad d_{t2} = \begin{cases} 0, & t = 1, \ldots, n, \\ 1, & t = n+1, \ldots, T. \end{cases}$$

Далее мы получаем $\mathrm{ESS_{UR}}$ как сумму квадратов остатков этой регрессии по всем T наблюдениям.

Докажите, что эти две процедуры дают одинаковые значения F-статистик.

Покажите также, что тот же результат может быть получен, если регрессия без ограничений записана в виде $y_t = \beta_1 x_t + \delta x_t d_{t2} + \varepsilon_t$, где $\delta = \beta_2 - \beta_1$.

3.4. Регрессия зависимой переменной y на три независимые переменные на основе $n = 30$ наблюдений дала следующие результаты:

y	$= 25.1$	$+ 1.2x_1$	$+ 1.0x_2$	$- 0.50x_3$
Стандартные ошибки	(2.1)	(1.5)	(1.3)	(0.060)
t-значения	(11.9)	( )	( )	( )
95%-ные доверительные границы	(±4.3)	( )	( )	( )

а) Заполните пропуски.

б) Истинны или ложны следующие утверждения (если ложны, исправьте их):

1) Оценка коэффициента при x_1 есть 1.2. Другие исследователи могут собрать другие данные и построить другие оценки этого коэффициента. Распределение этих оценок сосредоточено вокруг истинного значения 1.2. Поэтому оценка называется несмещенной.

2) Если есть априорная уверенность в том, что x_1 не влияет на y, то представляется разумным отвергнуть нулевую гипотезу H_0: $\beta_1 = 0$ на 5%-ном уровне значимости.

3) Если есть априорная уверенность в том, что x_2 влияет на y, то представляется более разумным использовать оценку 1.0, чем принимать нулевую гипотезу H_0: $\beta_2 = 0$.

3.5. Бюджетное обследование пяти случайно выбранных семей дало следующие результаты (в тыс. руб.):

Семья	Накопления, S	Доход, Y	Имущество, W
1	3.0	40	60
2	6.0	55	36
3	5.0	45	36
4	3.5	30	15
5	1.5	30	90

а) Оцените регрессию S на Y и W.

б) Спрогнозируйте накопления семьи, имеющей доход 40 тыс. руб. и имущество стоимостью 25 тыс. руб.

в) Предположим, что доход семьи возрос на 10 тыс. руб., в то время как стоимость имущества не изменилась. Оцените, как возрастут ее накопления.

г) Оцените, как возрастут накопления семьи, если ее доход вырос на 5 тыс. руб., а стоимость имущества увеличилась на 15 тыс. руб.

д) Найдите сумму квадратов остатков и постройте оценку дисперсии регрессии.

3.6. Рассмотрим регрессию $S = \beta_1 + \beta_2 Y + \beta_3 W + \varepsilon$ из предыдущего упражнения 3.5.

а) Постройте 95%-ное доверительное множество для

1) β_2 и β_3;

2) β_2;

3) β_3;

4) β_1 и β_2.

б) Проверьте с 5%-ным уровнем значимости следующие гипотезы:

1) $\beta_2 = 0$ и $\beta_3 = 0$;

2) $\beta_3 = 0$ (стоимость имущества несущественна);

3) $\beta_2 = 0$ (величина дохода несущественна);

4) $\beta_2 = 1$ (таким мог быть ответ вашего коллеги на вопрос о зависимости накопления от дохода);

5) $\beta_2 = 1.57$ (такое значение коэффициента β_2 могло быть с высокой степенью надежности установлено для другой страны и вас интересует вопрос, верно ли это для вашей страны);

6) $\beta_2 = -5\beta_3$ (т. е. эффект дохода противоположен эффекту богатства в фиксированной пропорции).

в) Пусть некоторая семья имеет доход $Y = 30$ тыс. руб. и имущество стоимостью $W = 52.5$ тыс. руб.

1) Чему равна прогнозная величина ее накоплений?

2) В каком смысле эта семья может рассматриваться как средняя между семьями 4 и 5 (упражнение 3.5)? Почему прогнозная величина ее накоплений не есть среднее между 3.5 и 1.5 тыс. руб.?

3) Постройте 95%-ный доверительный интервал для прогнозной величины накоплений этой семьи.

3.7. Всегда ли доверительный интервал для $\beta_1 + \beta_2$ шире каждого из доверительных интервалов для β_1 и β_2? Если да, то почему?

3.8. В этом упражнении изучается влияние преобразований зависимой и независимых переменных на МНК-оценки.

а) Что произойдет с МНК-оценками в парной регрессии y на x, если добавить константу к каждому наблюдению y? к каждому наблюдению x? Что произойдет с МНК-оценками в множественной

регрессии y на x_1 и x_2, если добавить константу c_1 к каждому наблюдению x_1 и другую константу c_2 к каждому наблюдению x_2?

б) Что произойдет с МНК-оценками в множественной регрессии y на x_1 и x_2, если переменные x_1 и x_2 заменить их отклонениями от средних значений?

в) Что произойдет с МНК-оценками в множественной регрессии, если умножить зависимую переменную Y на константу? если на константу умножить какой-либо регрессор?

3.9. Рассмотрим оценку вида $\widetilde{\boldsymbol{\beta}} = \left((\boldsymbol{X}'\boldsymbol{X})^{-1} + \gamma \boldsymbol{I}\right)\boldsymbol{X}'\boldsymbol{y}$ для вектора коэффициентов регрессионного уравнения $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. ($\boldsymbol{I}$ — единичная $k \times k$ матрица.)

а) Найдите математическое ожидание, матрицу ковариаций и матрицу среднеквадратичных отклонений оценки $\widetilde{\boldsymbol{\beta}}$ ($\mathrm{MSE}(\widetilde{\boldsymbol{\theta}}) = \mathrm{E}((\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta})')$).

б) Можно ли найти γ такое, что оценка $\widetilde{\boldsymbol{\beta}}$ более эффективна, чем оценка метода наименьших квадратов $\widehat{\boldsymbol{\beta}}$ (т.е. для всех $i = 1, \ldots, k$, $\mathrm{MSE}(\widetilde{\beta}_i) < \mathrm{MSE}(\widehat{\beta}_i)$)?

3.10. Рассмотрим оценку вида $\widetilde{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X} + \tau \boldsymbol{D})^{-1}\boldsymbol{X}'\boldsymbol{y}$ (ридж-регрессия (*ridge regression*)) для вектора коэффициентов регрессионного уравнения $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, где $\boldsymbol{D}$ — диагональная $k \times k$ матрица, состоящая из диагональных элементов матрицы $\boldsymbol{X}'\boldsymbol{X}$.

а) Найдите математическое ожидание, матрицу ковариаций и матрицу среднеквадратичных отклонений оценки $\widetilde{\boldsymbol{\beta}}$ ($\mathrm{MSE}(\widetilde{\boldsymbol{\theta}}) = \mathrm{E}((\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta})(\widetilde{\boldsymbol{\theta}} - \boldsymbol{\theta})')$).

б) Покажите, что существует $\tau > 0$ такое, что $\mathrm{V}(\widetilde{\boldsymbol{\beta}}) < \mathrm{V}(\widehat{\boldsymbol{\beta}})$, где $\widehat{\boldsymbol{\beta}}$ — оценка метода наименьших квадратов.

в) Можно ли найти такое $\tau > 0$, что оценка $\widetilde{\boldsymbol{\beta}}$ более эффективна, чем оценка метода наименьших квадратов $\widehat{\boldsymbol{\beta}}$ (т.е. для всех $i = 1, \ldots, k$, $\mathrm{MSE}(\widetilde{\beta}_i) < \mathrm{MSE}(\widehat{\beta}_i)$)?

3.11. После финансового кризиса спрос на чебуреки (см. упражнение 2.14) упал, и менеджер был вынужден тратить часть средств на рекламу. Для изучения зависимости объема продаж от цены и расходов на

рекламу менеджер использует следующую модель:

$$q_t = \beta_1 + \beta_2 p_t + \beta_3 a_t + \beta_4 a_t^2 + \varepsilon_t.$$

В таблице 3.2 приведены данные наблюдений за 20 недель (t — номер недели, q_t — количество проданных чебуреков, p_t — цена одного чебурека (руб.), a_t — затраты на рекламу (100 руб.)).

Таблица 3.2

t	q_t	p_t	a_t	t	q_t	p_t	a_t
1	525	5.92	4.79	11	407	6.67	5.19
2	567	6.50	3.61	12	608	6.92	3.27
3	396	6.54	5.49	13	399	6.97	4.69
4	726	6.11	2.78	14	631	6.59	3.79
5	265	6.62	5.74	15	545	6.50	4.29
6	615	5.15	1.34	16	512	6.86	2.71
7	370	5.02	5.81	17	845	5.09	2.21
8	789	5.02	3.39	18	571	6.08	3.09
9	513	6.77	3.74	19	539	6.36	4.65
10	661	5.57	3.59	20	620	6.22	1.97

Используя данные таблицы 3.2, ответьте на следующие вопросы:

а) Отклик количества проданных чебуреков на изменение цены измеряется коэффициентом $\beta_2 = \partial q/\partial p$. Аналогично, $\partial q/\partial a = \beta_3 + 2\beta_4 a$. Какие знаки β_2, β_3, β_4, вы ожидаете получить?

б) Найдите оценки коэффициентов регрессии и их стандартные ошибки. Соответствуют ли знаки оценок вашим ожиданиям?

в) Пусть себестоимость производства одного чебурека равна 2 рубля. Тогда чистый доход за неделю задается формулой $profit = pq - 2q - 100a$.

г) Найдите оптимальную цену при расходах на рекламу, равных 280 руб.

д) Найдите оптимальный уровень расходов на рекламу при цене чебурека, равной 6 руб.

е) Помогите менеджеру найти оптимальное решение (максимизирующее чистый доход).

ж) Найдите 95%-ные доверительные интервалы для β_2, β_3, β_4. Проверьте значимость влияния цены, а также расходов на рекламу на количество проданных чебуреков.

3.12. В кейнсианской теории спрос на деньги зависит от доходов и процентных ставок. Рассмотрим следующую модель:

$$m_t = \beta_1 + \beta_2 y_t + \beta_3 i_t + \varepsilon_t, \qquad (*)$$

где m_t — агрегат денежной массы M1 (млрд. долл.), y_t — валовой внутренний продукт (ВВП) (млрд. долл.), i_t — процентные ставки по 6-месячным государственным облигациям США (*6-month US Treasury Bills*, %). В таблице 3.3 представлены данные по этим переменным за период 1960–1983 гг. по экономике США.

Таблица 3.3

Год	y_t	m_t	i_t	Год	y_t	m_t	i_t
1960	506.5	141.8	3.247	1961	524.6	146.5	2.605
1962	565.0	149.2	2.908	1963	596.7	154.7	3.253
1964	637.7	161.8	3.686	1965	691.1	169.5	4.055
1966	756.0	173.7	5.082	1967	799.6	185.1	4.630
1968	873.4	199.4	5.470	1969	944.0	205.8	6.853
1970	992.7	216.5	6.562	1971	1077.6	230.7	4.511
1972	1185.9	251.9	4.466	1973	1326.4	265.8	7.178
1974	1434.2	277.5	7.926	1975	1549.2	291.1	6.122
1976	1718.0	310.4	5.266	1977	1918.3	335.5	5.510
1978	2163.9	363.2	7.572	1979	2417.8	389.0	10.017
1980	2631.7	414.1	11.374	1981	2954.1	440.6	13.776
1982	3073.0	478.2	11.084	1983	3309.5	521.1	8.750

Источник: *Economic Report of the President*, Department of Commerce, Bureau of Economic Analysis.

а) Найдите оценки коэффициентов регрессии (*). Интерпретируйте знаки коэффициентов.

б) Рассчитайте прогноз спроса на деньги при значениях: (1) $y = 1000$, $i = 10$ и (2) $y = 2500$, $i = 5$.

в) Рассчитайте эластичность спроса на деньги m по доходам y и по процентным ставкам ($\partial \ln m/\partial \ln y$, $\partial \ln m/\partial \ln i$) в двух точках (1) и (2) из б). Сравните результаты.

г) Рассмотрим модель

$$\ln m_t = \beta_1 + \beta_2 \ln y_t + \beta_3 \ln i_t + \varepsilon_t. \qquad (**)$$

Повторите б) и в) и сравните результаты, полученные по разным моделям. Сравните модели (*) и (**). Какая из них вам представляется более предпочтительной?

3.13. Рассмотрим классическую модель линейной регрессии $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$ с ограничением $\boldsymbol{H\beta} = \boldsymbol{r}$ на вектор коэффициентов.

а) Покажите, что оценка метода наименьших квадратов при наличии ограничения $\widehat{\boldsymbol{\beta}}_{\mathrm{R}}$, получающаяся из решения соответствующей задачи минимизации, следующим образом выражается через обычную оценку метода наименьших квадратов без учета ограничения $\widehat{\boldsymbol{\beta}}_{\mathrm{UR}}$:

$$\widehat{\boldsymbol{\beta}}_{\mathrm{R}} = \widehat{\boldsymbol{\beta}}_{\mathrm{UR}} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}'\left(\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}'\right)^{-1}(\boldsymbol{r} - \boldsymbol{H}\widehat{\boldsymbol{\beta}}_{\mathrm{UR}}).$$

б) Покажите, что

$$(\boldsymbol{H}\widehat{\boldsymbol{\beta}}_{\mathrm{UR}} - r)'\left(\boldsymbol{H}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{H}'\right)^{-1}(\boldsymbol{H}\widehat{\boldsymbol{\beta}}_{\mathrm{UR}} - r) = \boldsymbol{e}_{\mathrm{R}}'\boldsymbol{e}_{\mathrm{R}} - \boldsymbol{e}_{\mathrm{UR}}'\boldsymbol{e}_{\mathrm{UR}},$$

где $\boldsymbol{e}_{\mathrm{R}} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\mathrm{R}}$, $\boldsymbol{e}_{\mathrm{UR}} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\mathrm{UR}}$ — векторы остатков в регрессиях с ограничениями и без ограничений, соответственно.

3.14. Оценивание модели $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4} + \varepsilon_t$ методом наименьших квадратов по 26 наблюдениям дало следующие результаты:

$$y_t = 2 + \underset{(1.9)}{3.5}\, x_{t2} - \underset{(2.2)}{0.7}\, x_{t3} + \underset{(1.5)}{2.0}\, x_{t4} + e_t, \qquad R^2 = 0.882$$

(в скобках даны значения t-статистик).

Оценивание той же модели при ограничении $\beta_2 = \beta_4$ дало следующие результаты:

$$y_t = 1.5 + \underset{(2.7)}{3.0}\,(x_{t2} + x_{t4}) - \underset{(2.4)}{0.6}\, x_{t3} + u_t, \qquad R^2 = 0.876.$$

а) Проверьте значимость вектора $\beta' = (\beta_2, \beta_3, \beta_4)$ в регрессии без ограничений.

б) Проверьте ограничение $\beta_2 = \beta_4$.

3.15. В таблице 3.4 представлены реальный доход на душу населения y (тыс. долл.), процент рабочей силы, занятой в сельском хозяйстве, x_1 и средний уровень образования населения в возрасте после 25 лет x_2 (число лет, проведенных в учебных заведениях) для 15 развитых стран в 1983 г.

а) Проведите множественную регрессию y на константу, x_1 и x_2 и проинтерпретируйте полученные результаты.

Таблица 3.4

Страна	y	x_1	x_2	Страна	y	x_1	x_2
1	7	8	9	9	10	6	12
2	9	9	13	10	11	7	14
3	9	7	11	11	11	6	11
4	8	6	11	12	12	4	15
5	8	10	12	13	9	8	15
6	14	4	16	14	10	5	10
7	9	5	11	15	12	8	13
8	8	5	11				

б) Определите s^2, $s^2_{\hat\beta_1}$ и $s^2_{\hat\beta_2}$.

в) Почему, как правило, константа β_0 не играет существенной роли при рассмотрении регрессии?

г) Постройте 95%-ные доверительные интервалы для коэффициентов β_1, β_2 и вычислите коэффициент детерминации R^2 и скорректированный коэффициент детерминации R^2_{adj}.

д) Проверьте на 5%-ном уровне значимость коэффициентов β_1, β_2.

3.16. Вместо того, чтобы оценивать параметры $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ в модели

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}, \qquad (*)$$

($\boldsymbol{X}_1, \boldsymbol{X}_2$ — $n \times k_1$, $n \times k_2$ матрицы, соответственно, $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2$ — векторы размерности k_1, k_2, соответственно), строятся МНК-оценки этих параметров исходя из модели

$$\boldsymbol{y} = \boldsymbol{X}_1^*\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}^*, \qquad (**)$$

где $\boldsymbol{X}_1^*$ — матрица остатков, полученных в результате регрессии каждого столбца матрицы $\boldsymbol{X}_1$ на $\boldsymbol{X}_2$.

а) Покажите, что полученная таким образом оценка вектора $\boldsymbol{\beta}_2$ совпадает с оценкой, полученной в результате регрессии $\boldsymbol{y}$ только на $\boldsymbol{X}_2$.

б) Найдите смещение оценки вектора $\boldsymbol{\beta}_2$.

в) Покажите, что МНК-оценки вектора $\boldsymbol{\beta}_1$, построенные по моделям (*) и (**), совпадают.

3.17. Строится регрессия $n \times 1$ вектора $\boldsymbol{y}$ на $n \times k$ матрицу регрессоров $\boldsymbol{X}$ и вычисляется коэффициент детерминации R_1^2. Затем к матрице $\boldsymbol{X}$ добавляется дополнительный $(k+1)$-й столбец, проводится регрессия $\boldsymbol{y}$ на новую матрицу и вычисляется коэффициент детерминации R_2^2. При каких условиях $R_1^2 = R_2^2$?

3.18. Рассматривается стандартная линейная модель множественной регрессии $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, где $\boldsymbol{X}$ — $n \times k$ матрица ранга k.

а) Пусть $\boldsymbol{G}$ — $k \times m$ матрица, имеющая ранг $m < k$, и пусть $L = \{\boldsymbol{\beta} : \boldsymbol{\beta} = \boldsymbol{G}\boldsymbol{\gamma}$ для некоторого $\boldsymbol{\gamma}\}$. Постройте тест для проверки гипотезы $\text{H}_0 : \boldsymbol{\beta} \in L$ против альтернативы $\text{H}_1 : \boldsymbol{\beta} \notin L$.

б) Пусть матрица $\boldsymbol{X}$ разбита на две матрицы $\boldsymbol{X} = [\boldsymbol{X}_1\ \boldsymbol{X}_2]$, где $\boldsymbol{X}_1$ — $n \times k_1$ матрица, $\boldsymbol{X}_2$ — $n \times k_2$ матрица, и пусть $\boldsymbol{q}_1 = \boldsymbol{X}_1\boldsymbol{r}_1$, $\boldsymbol{q}_2 = \boldsymbol{X}_2\boldsymbol{r}_2$, где $\boldsymbol{r}_1, \boldsymbol{r}_2$ — известные векторы. Рассматривается новая модель $\boldsymbol{y} = \alpha_1\boldsymbol{q}_1 + \alpha_2\boldsymbol{q}_2$, где α_1, α_2 — скалярные параметры.

Каким образом, используя результаты а), можно проверить, является ли новая модель приемлемой?

3.19. Покажите, что при добавлении в модель регрессора скорректированный коэффициент детерминации R^2_{adj} увеличивается тогда и только тогда, когда t-статистика оценки коэффициента при этом регрессоре по модулю превосходит единицу.

3.20. Оценивание четырех регрессионных моделей на основании 40 наблюдений дало следующие результаты:

$$W = \underset{(5.0)}{20} + \underset{(0.09)}{0.8}\,AGE + \underset{(1.31)}{3.7}\,EDU, \qquad R^2 = 0.40,$$

$$\ln W = \underset{(3.0)}{3.2} + \underset{(0.009)}{0.10}\,\ln AGE + \underset{(0.03)}{0.19}\,\ln EDU, \qquad R^2 = 0.71,$$

$$W = \underset{(0.3)}{20} + \underset{(0.09)}{0.6}\,AGE + \underset{(0.12)}{0.4}\,EXP, \qquad R^2 = 0.59,$$

$$W = \underset{(0.4)}{2.05} + \underset{(0.19)}{0.5}\,AGE + \underset{(0.35)}{0.6}\,EDU + \underset{(0.13)}{0.2}\,EXP, \qquad R^2 = 0.63$$

(в скобках указаны стандартные ошибки), где W — зарплата работника, AGE — его возраст (в годах), EDU — уровень образования (число лет, проведенных в учебных заведениях), EXP — стаж работы.

а) Сравните эти четыре регрессии с точки зрения их качества и прогностической силы.

б) Дайте интерпретацию коэффициентов при переменных AGE и $\ln AGE$ в первом и втором уравнениях соответственно.

3.21. Рассмотрим 3 модели

$$y_t = \frac{\alpha}{t} + \varepsilon_t,$$
$$y_t = \alpha g^t + \varepsilon_t,$$
$$y_t = \alpha + \frac{\beta}{t^2} + \varepsilon_t.$$

где $t = 1, \ldots, T$, $\varepsilon \sim N(0, \sigma^2 I)$. Во втором уравнении g — известная константа.

а) Покажите, что МНК-оценка параметра α в первом уравнении не может быть состоятельна. Верно ли то же самое для второго уравнения?

б) Являются ли состоятельными МНК-оценки параметров α и β в третьем уравнении?

Указание. $\sum_{t=1}^{\infty} t^{-2} = \pi^2/6$, $\sum_{t=1}^{\infty} t^{-4} = \pi^4/90$.

3.22. Пусть истинная модель, $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \beta_4 x_{t4} + \varepsilon_t$, удовлетворяет условиям теоремы Гаусса–Маркова. Оценки $\widetilde{\beta}_1, \widetilde{\beta}_2, \widetilde{\beta}_3$ являются МНК-оценками в регрессии y на x_2 и x_3. Покажите, что

$$\mathrm{E}\widetilde{\beta}_2 = \beta_2 + \beta_4 \frac{\sum_{t=1}^{n} r_{t2} x_{t4}}{\sum_{t=1}^{n} r_{t2}^2},$$

где r_{t2} — МНК-остатки в регрессии x_2 на x_3.
Указание. Покажите сначала, что МНК-оценка коэффициента β_2 в уравнении $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \varepsilon_t$ представляется в виде

$$\widehat{\beta}_2 = \frac{\sum_{t=1}^{n} r_{t2} y_t}{\sum_{t=1}^{n} r_{t2}^2}.$$

3.23. Рассматривается классическая линейная нормальная модель $y = X\beta + \varepsilon$, $\mathrm{V}(\varepsilon) = 2I$, причем известно, что

$$X'X = \begin{bmatrix} 5 & 2 \\ 2 & 4 \end{bmatrix} \quad \text{и} \quad \widehat{\beta}_1 = 3,\ \widehat{\beta}_2 = 2.$$

а) Постройте 95%-ный доверительный интервал для $\theta = \beta_1 + \beta_2$.

б) Постройте 95%-ную доверительную область для вектора $\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$.

3.24. Для города и для деревни рассматриваются две модели парной регрессии. 20 наблюдений для города дали следующие результаты:

$$X'X = \begin{bmatrix} 20 & 20 \\ 20 & 25 \end{bmatrix}, \quad X'y = \begin{bmatrix} 10 \\ 20 \end{bmatrix}, \quad y'y = 30,$$

а 10 наблюдений для деревни —

$$X'X = \begin{bmatrix} 10 & 10 \\ 10 & 20 \end{bmatrix}, \quad X'y = \begin{bmatrix} 8 \\ 20 \end{bmatrix}, \quad y'y = 24.$$

На 95%-ном доверительном уровне проверьте гипотезу о том, что эти две модели совпадают.

3.25. Проведены две регрессии ежеквартальных данных со второго квартала 1990 г. по третий квартал 2001 г. Они имеют следующий вид:

$$\widehat{y} = 40 + 0.3x_2 + 0.8x_3 - 1.8x_4, \qquad R^2 = 0.82,$$
$$\widehat{y} = 60 + 0.5x_2 + 0.6x_3, \qquad R^2 = 0.75.$$

Для первой регрессии проверьте (на 95%-ном уровне значимости) гипотезу H_0: $\beta_4 = 1$.

3.26. Известно, что процесс, порождающий данные (истинная модель), описывается классической линейной моделью регрессии $y = X\beta + \varepsilon$. Оценка $\widehat{\beta}_R$ получается регрессией y на X (МНК-оценка) при ограничении $H\beta = r$. Найдите матрицу ковариаций $\mathrm{V}(\widehat{\beta}_R)$ и сравните ее с матрицей ковариаций $\mathrm{V}(\widehat{\beta})$ — МНК-оценки в регрессии без ограничений. Как полученный вами результат соотносится с теоремой Гаусса–Маркова?

3.27. При каких условиях добавление в уравнение еще одного регрессора не изменяет коэффициент детерминации?

3.28. Оценивание производственной функции по методу наименьших квадратов дало следующие результаты:

$$\ln Q = 1.37 + \underset{(0.257)}{0.632} \ln K + \underset{(0.219)}{0.452} \ln L, \quad R^2 = 0.98,\ \mathrm{Cov}(\widehat{\beta}_K, \widehat{\beta}_L) = 0.055$$

(в скобках даны стандартные ошибки). Проверьте гипотезы:

а) эластичности по труду и капиталу совпадают;

б) выполнено свойство постоянства отдачи на масштаб.

Замечание. В задаче не указано число наблюдений. Будут ли ваши выводы зависеть от этого числа?

3.29. Рассматривается стандартная линейная регрессионная модель

$$y_t = \alpha + \beta x_t + \delta w_t + \theta z_t + \varepsilon_t.$$

а) Какую регрессию следует осуществить, чтобы учесть (истинную) информацию, что $\beta = 2\delta$?

б) Будет ли коэффициент детерминации R^2 этой регрессии (п. а)) больше, меньше или равен R^2 исходной регрессии?

в) Будут ли оценки параметра θ в исходной модели и в п. а несмещенными?

г) Будет ли дисперсия этой оценки (п. а)) больше, меньше или равна дисперсии оценки θ в исходной регрессии. Объясните на содержательном уровне.

3.30. В файле `gnovgorod.xls` содержатся данные по стоимости квартир в Новгороде.

а) Постройте и оцените минимальную модель, с помощью которой вы сможете оценить параметр r, равный относительному приросту стоимости квартиры при добавлении к ней комнаты площадью 18 кв.м.

б) Найдите 95%-ный доверительный интервал для r.

в) Помогает ли включение в модель дополнительных параметров более точно оценить параметр r?

г) Можете ли вы предложить модель, в которой параметр r был бы одним из коэффициентов? Изменяется ли при этом способе оценивания доверительный интервал?

д) Зависит ли параметр r от количества комнат в квартире? Почему?

3.31. В примере рассматриваются данные по стоимости квартир в Москве, собранные студентами первого курса РЭШ осенью 1997 г. Описание переменных содержится в таблице 3.5.

Данные находятся в файле `flat98s.xls`

Таблица 3.5

Переменная	Описание
TOTSP	Общая площадь, кв.м
PRICE	Цена квартиры, тыс. долл.
ROOMS	Количество жилых комнат
LIVSP	Жилая площадь, кв.м
KITSP	Площадь кухни, кв.м
DIST	Расстояние до центра, км
METRDIST	Расстояние до ближайшей станции метро, мин
WALK	1, если пешком от метро, 0 — иначе
BRICK	1, если дом кирпичный, 0 — иначе
TEL	1, если есть телефон, 0 — иначе
BAL	1, если есть балкон или лоджия, 0 — иначе
FLOOR	0, если квртира находится на первом или последнем этаже, 1 — иначе

а) Постройте модель стоимости квартиры (или стоимости квадратного метра жилой площади квартиры) в зависимости от имеющихся параметров.

б) Проверьте гипотезу, что модели для 1, 2, 3–4-комнатных квартир различаются между собой, т. е. гипотезу, что рынок распадается на рынки однокомнатных, двухкомнатных и трех-четырехкомнатных квартир.

3.32. (автор — Arthur van Soest, Tilburg University)
Введение. Рассматриваемые здесь упражнения в значительной мере опираются на статью (Mankiw et al., 1992) и направлены на проверку полученных там результатов (в первую очередь, с точки зрения здравого экономического смысла). В цитированной статье изучается расширенный вариант модели экономического роста Солоу. Основным объектом, изучаемым в модели Солоу, является удельная величина валового внутреннего продукта (ВВП) в стационарном состоянии. Таким образом, модель объясняет различие в уровне благосостояния разных стран в долговременном плане. Обобщение модели Солоу в работе (Mankiw et al., 1992) состоит в том, что в отличие от первоначальной модели здесь допускаются инвестиции не только в физический, но и в человеческий капитал. Приведенный там эмпирический анализ основан на межстрановых данных, взятых из работы (Summers, Heston, 1988). Мы также будем использовать эти данные.

Обобщенная модель Солоу. Дадим краткое описание обобщенной модели Солоу, предложенной в работе (Mankiw et al., 1992). Исходная модель Солоу изложена во многих учебниках по макроэкономике (см., например, (Romer, 2001)). Предполагается, что в каждый момент времени t производство задается производственной функцией Кобба–Дугласа с постоянной отдачей на масштаб:

$$Y_t = K_t^{\alpha} H_t^{\beta} (A_t L_t)^{1-\alpha-\beta},$$

где Y — выпуск, K и H — объем физического и человеческого капитала, соответственно, L — труд, а переменная A описывает уровень технологии. Предположение о постоянстве отдачи на масштаб позволяет оперировать с удельными величинами (на единицу эффективного труда):

$$y = \frac{Y}{AL},\ k = \frac{K}{AL},\ h = \frac{H}{AL}.$$

Будем также считать, что выполнены следующие условия:

- фиксированные доли s_k, s_h суммарного выпуска Y инвестируются в физический и человеческий капитал, соответственно;
- $L_t = L_0 e^{nt}$, где n — скорость роста населения;
- $A_t = A_0 e^{gt}$, где g — скорость роста технологического уровня;
- интенсивность амортизации δ одинакова для физического и человеческого капитала.

Из этих предположений вытекает, что эволюция капитала описывается следующей системой дифференциальных уравнений:

$$\dot{k}_t = s_k y_t - (n + g + \delta) k_t,$$
$$\dot{h}_t = s_h y_t - (n + g + \delta) h_t.$$

Стационарное состояние характеризуется условиями $\dot{k}_t = \dot{h}_t = 0$.

3.32.1. Покажите, что в стационарном состоянии выполнено равенство

$$\ln y_t = \frac{1}{1-\alpha-\beta}(\alpha \ln s_k + \beta \ln s_h - (\alpha+\beta)\ln(n+g+\delta)). \qquad (3.51)$$

Равенство (3.51) устанавливает соотношение (в стационарном состоянии) между благосостоянием страны, скоростью роста ее населения и интенсивностью инвестиций в физический и человеческий капитал.

Следствием этого соотношения является то, что и в долговременном плане можно ожидать сохранение различия в уровне благосостояния разных стран.

Модель также позволяет описать траекторию сходимости к стационарному состоянию. Пусть y^* — значение y_t в стационарном состоянии. Тогда можно показать, что имеет место следующее приближенное соотношение:

$$\frac{d \ln y_t}{dt} = \lambda(\ln y^* - \ln y_t),$$

где $\lambda = (n + g + \delta)(1 - \alpha - \beta)$. Решая это уравнение, получаем

$$\ln y_t = (1 - e^{-\lambda t} \ln y^* + e^{-\lambda t} \ln y_0). \qquad (3.52)$$

3.32.2. Покажите, что из (3.51) и (3.52) вытекает следующее уравнение для траектории сходимости:

$$\ln \frac{y_t}{y_0} = (1 - e^{-\lambda t})$$
$$\times \left\{ \frac{1}{1-\alpha-\beta}(\alpha \ln s_k + \beta \ln s_h - (\alpha+\beta)\ln(n+g+\delta)) - \ln y_0 \right\}. \qquad (3.53)$$

Данные. Используются данные, извлеченные из архива журнала Journal of Applied Econometrics. Они соответствуют работе (Durlauf, Johnson 1995). Начало этим исследованиям положила работа (Summers, Heston 1988). Единицей наблюдения является страна, даны результаты наблюдений 121 страны. Используются переменные, перечисленные в таблице 3.6.

Все данные, за исключением LIT, взяты из работы (Mankiw et al., 1992); переменная LIT взята из доклада Всемирного банка. Данные содержатся в файле `growth.xls`. Список стран приведен в приложении к работе (Mankiw et al., 1992).

3.32.3. а) Вычислите суммарные статистики всех переменных. Проверьте, имеют ли смысл ваши результаты.

б) Вычислите корреляционную матрицу всех переменных. Дайте интерпретацию наиболее важных результатов. Соответствуют ли они тому, что вы ожидали?

Таблица 3.6

Переменная	Описание
NUM	номер страны в базе данных Summers, Heston (1988)
$NOIL$	1 для страны, не добывающей нефть, 0 — для добывающей
$INTER$	1 для страны с хорошим качеством данных, 0 — в противном случае
$OECD$	1 для страны, входящей в Организацию экономического сотрудничества и развития, 0 — в противном случае
$GDP60$	ВВП на душу населения в 1960 г. (долл.)
$GDP85$	ВВП на душу населения в 1985 г. (долл.)
$GDPGRO$	средний рост ВВП на душу населения с 1960 по 1985 г. (%)
$POPGRO$	средний рост работоспособного населения с 1960 по 1985 г. (%)
$IONY$	средняя доля инвестиций (включая государственные) в общем объеме ВВП с 1960 по 1985 г. (%)
SCH	средняя доля населения, продолжающего получать образование одновременно с работой с 1960 по 1985 г. (%)
LIT	доля людей среди населения старше 15 лет, умеющих читать и писать в 1960 г.

Анализ стационарного состояния. Если предположить, что в 1985 г. страны достигли стационарного состояния, то мы можем использовать достигнутый в 1985 г. уровень ВВП для оценивания уравнения (3.51). Поскольку мы используем данные, относящиеся к одному и тому же году, то индекс t можно опустить. Уравнение (3.51) переписывается в следующем виде:

$$\ln GDP85 = \pi_0 + \pi_1 \ln s_k + \pi_2 \ln s_h + \pi_3 \ln(n + g + \delta), \qquad (3.54)$$

где $\pi_0 = \ln A_0 + gt$ — постоянный член. При оценивании уравнения (3.54) представляется разумным в качестве s_k использовать переменную $IONY$, а в качестве s_h переменную SCH. Мы не наблюдаем величины g и δ, поэтому будем считать, как в работе (Mankiw et al., 1992), что $g = 2\%$ и $\delta = 3\%$. В качестве n берется переменная $POPGRO$.

3.32.4. а) Оцените уравнение (3.54), используя данные по всем странам, за исключением тех, для которых пропущены наблюдения какой-либо переменной.

б) Исходная модель Солоу не включает человеческий капитал. Оцените уравнение (3.54), удалив переменную $\ln SCH$. Сравните с результатом, полученным в п. а). В чем состоит основное различие? Объясните это различие, используя также результаты упражнения 3.32.3.

3.32.5. а) Структурная форма (3.51) накладывает некоторое линейное ограничение на параметры π_1, π_2, π_3 приведенной формы. Что это за ограничение?

б) Протестируйте (на 5%-ном уровне значимости) выполнимость этого ограничения.

в) Оцените вновь уравнение (3.54), используя это ограничение. Сравните ваш результат с результатом, полученным в упражнении 3.32.4 а).

г) Выразите структурные параметры α и β через π_1, π_2 и постройте, таким образом, их оценки.

3.32.6. а) Добавьте в регрессионное уравнение упражнения 3.32.4 а) фиктивные переменные $NOIL$ и $OECD$ и проверьте их значимость.

б) Проверьте, является линейная спецификация (3.54) разумной, добавляя квадраты независимых переменных и перекрестные члены.

3.32.7. Согласно «золотому правилу накопления капитала», доли инвестиций s_k, s_h должны выбираться таким образом, чтобы в стационарном состоянии величина $c = (1 - s_k - s_h)y$ была максимальна.

а) Найдите теоретические оптимальные значения величин s_k, s_h.

б) Используя оценки, полученные в упражнении 3.32.5, проверьте, удовлетворяют ли в среднем инвестиции в физический капитал «золотому правилу».

Рост ВВП. Уравнение (3.53) служит основой эмпирического анализа роста ВВП в период с 1960 ($t = 0$) по 1985 г. ($t = 25$). Заметим, что уравнение (3.53) можно переписать следующим образом:

$$\ln \frac{GDP85}{GDP60} = \pi_0 + \pi_1 \ln s_k + \pi_2 \ln s_h + \pi_3 \ln(n + g + \delta) + \pi_4 \ln GDP60. \qquad (3.55)$$

При оценивании этого уравнения будем использовать те же предположения, что и в предыдущих разделах. Так, например, $g = 2\%$, $\delta = 3\%$ и т. д.

3.32.8. а) Оцените уравнение (3.55), интерпретируйте результат.

б) Исходная модель Солоу не включает человеческий капитал. Оцените уравнение (3.55), удалив переменную $\ln SCH$. Сравните результат с тем, что получен в п. а). Объясните разницу.

3.32.9. а) Структурная форма (3.53) накладывает некоторое линейное ограничение на параметры π_1, π_2, π_3 приведенной формы. Что это за ограничение?

б) Протестируйте (на 5%-ном уровне значимости) выполнимость этого ограничения.

в) Оцените вновь уравнение (3.55), используя это ограничение. Сравните ваш результат с результатом, полученным в упражнении 3.32.8 а).

г) Используя результат п. в), постройте оценки структурных параметров λ, α, β. Проинтерпретируйте результаты. Сравните ваши оценки параметров α, β, с оценками, полученными в предыдущих упражнениях.

3.32.10. а) Добавьте фиктивные переменные $NOIL$ и $OECD$ в уравнение упражнения 3.32.8 а) и проверьте их значимость.

б) Проверьте, является линейная спецификация (3.55) разумной, добавляя квадраты независимых переменных и перекрестные члены.

3.32.11. а) Оцените уравнение (3.55) отдельно для стран — членов OECD и для стран — нечленов OECD и проинтерпретируйте результаты.

б) Проверьте, совпадают ли коэффициенты уравнения (3.55) (за исключением свободного члена) для стран — членов OECD и для стран — нечленов OECD.

3.32.12. Выберите наилучшее, с вашей точки зрения, уравнение и постройте 95%-ный доверительный интервал для скорости сходимости λ. Проинтерпретируйте результат.

3.33. Рассматривается информация о стоимости коттеджей в Московской области по Киевскому направлению (по данным строительной компании «Стройсервис», осень 1997 г.).

Данные находятся в файле `villa.xls`. Переменные описаны в таблице 3.7.

Таблица 3.7

Переменная	Описание
N	Номер по порядку
$Price$	Цена в тыс. долл.
$Dist$	Расстояние от кольцевой автодороги в км
$House$	Площадь дома в кв.м
$Area$	Площадь участка в сотках

Подберите функциональную форму зависимости цены коттеджа от его параметров, учитывая такие факторы, как t-статистики и коэффициент детерминации R^2.

Глава 4

Различные аспекты множественной регрессии

В предыдущих главах были изучены основные теоретико-статистические вопросы регрессионных моделей. В данной главе рассматриваются некоторые проблемы, часто возникающие при их практическом использовании.

На практике исследователю нередко приходится сталкиваться с ситуацией, когда полученная им регрессия является «плохой», т. е. t-статистики большинства оценок малы, что свидетельствует о незначимости соответствующих независимых переменных (регрессоров). В то же время F-статистика (3.36) может быть достаточно большой, что говорит о значимости регрессии в целом. Одна из возможных причин такого явления носит название *мультиколлинеарности* и возникает при наличии высокой корреляции между регрессорами. Проблеме мультиколлинеарности посвящено начало этой главы.

Регрессионные модели являются достаточно гибким инструментом, позволяющим, в частности, оценивать влияние качественных признаков (пол, профессия, наличие детей и т. п.) на изучаемую переменную. Это достигается введением в число регрессоров так называемых *фиктивных переменных*, принимающих, как правило, значения 1 или 0 в зависимости от наличия или отсутствия

соответствующего признака в очередном наблюдении. С формальной точки зрения фиктивные переменные ничем не отличаются от других регрессоров. Однако следует обратить особое внимание на правильное их использование и интерпретацию оценок.

В этой главе мы также рассмотрим задачу нахождения частной корреляции между переменными и так называемую проблему спецификации модели.

4.1. Мультиколлинеарность

Одним из условий классической регрессионной модели является предположение о линейной независимости объясняющих переменных, что означает линейную независимость столбцов матрицы регрессоров $\boldsymbol{X}$ или (эквивалентно) что матрица $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ имеет полный ранг k. При нарушении этого условия, т. е. когда один из столбцов матрицы $\boldsymbol{X}$ есть линейная комбинация остальных столбцов, говорят, что имеет место *полная коллинеарность*. В этой ситуации нельзя построить МНК-оценку параметра $\boldsymbol{\beta}$, что формально следует из сингулярности матрицы $\boldsymbol{X}'\boldsymbol{X}$ и невозможности решить нормальные уравнения. Нетрудно также понять и содержательный смысл этого явления. Рассмотрим следующий простой пример регрессии (Greene, 1997): $C = \beta_1 + \beta_2 S + \beta_3 N + \beta_4 T + \varepsilon$, где C — потребление, S — зарплата, N — доход, получаемый вне работы, T — полный доход. Поскольку выполнено равенство $T = S + N$, то для произвольного числа h исходную регрессию можно переписать в следующем виде: $C = \beta_1 + \beta_2' S + \beta_3' N + \beta_4' T + \varepsilon$, где $\beta_2' = \beta_2 + h$, $\beta_3' = \beta_3 + h$, $\beta_4' = \beta_4 - h$. Таким образом, одни и те же наблюдения могут быть объяснены различными наборами коэффициентов $\boldsymbol{\beta}$. Эта ситуация тесно связана с проблемой *идентифицируемости* системы, о чем более подробно будет говориться позднее. Кроме того, если с учетом равенства $T = S + N$ переписать исходную систему в виде $C = \beta_1 + (\beta_2 + \beta_4)S + (\beta_3 + \beta_4)N + \varepsilon$, то становится ясно, что оценить можно лишь три параметра β_1, $(\beta_2 + \beta_4)$ и $(\beta_3 + \beta_4)$, а не четыре исходных. В общем случае можно показать, что если $\operatorname{rank}(\boldsymbol{X}'\boldsymbol{X}) = l < k$, то оценить можно только l

линейных комбинаций исходных коэффициентов. Если есть полная коллинеарность, то можно выделить в матрице X максимальную линейно независимую систему столбцов и, удалив остальные столбцы, провести новую регрессию.

На практике полная коллинеарность встречается исключительно редко. Гораздо чаще приходится сталкиваться с ситуацией, когда матрица X имеет полный ранг, но между регрессорами имеется высокая степень корреляции, т. е. когда матрица $X'X$, говоря нестрого, близка к вырожденной. Тогда говорят о наличии мультиколлинеарности. В этом случае МНК-оценка формально существует, но обладает «плохими» свойствами. Это нетрудно объяснить, используя геометрическую интерпретацию метода наименьших квадратов. Как уже отмечалось, регрессию можно рассматривать как проекцию в пространстве R^n вектора y на подпространство, порожденное столбцами матрицы X. Если между этими векторами существует приблизительная линейная зависимость, то операция проектирования становится неустойчивой: небольшое изменение в исходных данных может привести к существенному изменению оценок. Рисунок 4.1 наглядно это демонстрирует. Векторы y и y' мало отличаются друг от друга, но в

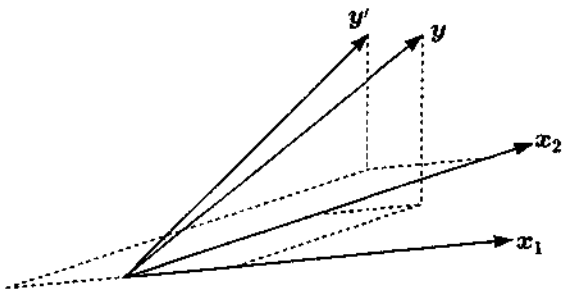


Рис. 4.1

силу того, что угол между векторами (регрессорами) x_1 и x_2 мал, разложения проекций этих векторов по x_1 и x_2 отличаются значительно. У проекции вектора y оба коэффициента разложения по x_1 и x_2 (напомним, что это и есть МНК-оценки) положительны и относительно невелики. У проекции вектора y' коэффициент

при x_1 принимает большое отрицательное значение. В силу этого обстоятельства интерпретация коэффициентов регрессии становится весьма проблематичной.

Мультиколлинеарность может возникать в силу разных причин. Например, несколько независимых переменных могут иметь общий временной тренд, относительно которого они совершают малые колебания. В частности, так может случиться, когда значения одной независимой переменной являются лагированными значениями другой.

Выделим некоторые наиболее характерные признаки мультиколлинеарности.

1. Небольшое изменение исходных данных (например, добавление новых наблюдений) приводит к существенному изменению оценок коэффициентов модели.

2. Оценки имеют большие стандартные ошибки, малую значимость, в то время как модель в целом является значимой (высокое значение коэффициента детерминации R^2 и соответствующей F-статистики).

3. Оценки коэффициентов имеют неправильные с точки зрения теории знаки или неоправданно большие значения.

Что же делать, если по всем признакам имеется мультиколлинеарность? Однозначного ответа на этот вопрос нет, и среди эконометристов есть разные мнения на этот счет. Существует даже такая школа, представители которой считают, что и не нужно ничего делать, поскольку «так устроен мир» (см. Kennedy, 1992). Мы здесь не ставим цель дать достаточно полное описание методов борьбы с мультиколлинеарностью. Более подробно об этом можно прочесть, например, в (Greene, 1997, глава 9).

У неискушенного исследователя при столкновении с проблемой мультиколлинеарности может возникнуть естественное желание отбросить «лишние» независимые переменные, которые, возможно, служат ее причиной. Однако следует помнить, что при этом могут возникнуть новые трудности. Во-первых, далеко не

всегда ясно, какие переменные являются лишними в указанном смысле. Мультиколлинеарность означает лишь приблизительную линейную зависимость между столбцами матрицы X, но это не всегда выделяет «лишние» переменные. Во-вторых, во многих ситуациях удаление каких-либо независимых переменных может значительно отразиться на содержательном смысле модели. Наконец, как будет показано в п. 4.4, отбрасывание так называемых существенных переменных, т. е. независимых переменных, которые реально влияют на изучаемую зависимую переменную, приводит к смещенности МНК-оценок.

4.2. Фиктивные переменные

Как правило, независимые переменные в регрессионных моделях имеют «непрерывные» области изменения (национальный доход, уровень безработицы, размер зарплаты и т. п.). Однако теория не накладывает никаких ограничений на характер регрессоров, в частности, некоторые переменные могут принимать всего два значения или, в более общей ситуации, дискретное множество значений. Необходимость рассматривать такие переменные возникает довольно часто в тех случаях, когда требуется принимать во внимание какой-либо качественный признак. Например, при исследовании зависимости зарплаты от различных факторов может возникнуть вопрос, влияет ли на ее размер, и если да, то в какой степени, наличие у работника высшего образования. Также можно задать вопрос, существует ли дискриминация в оплате труда между мужчинами и женщинами. В принципе можно оценивать соответствующие уравнения внутри каждой категории, а затем изучать различия между ними, но введение дискретных переменных позволяет оценивать одно уравнение сразу по всем категориям.

Покажем, как это можно сделать в примере с зарплатой. Пусть $x_t = (x_{t1}, \dots, x_{tk})'$ — набор объясняющих (независимых) переменных, т. е. первоначальная модель описывается уравнениями

$$y_t = x_{t1}\beta_1 + \dots + x_{tk}\beta_k + \varepsilon_t = x_t'\beta + \varepsilon_t, \quad t = 1, \dots, n, \tag{4.1}$$

где y_t — размер зарплаты t-го работника. Теперь мы хотим включить в рассмотрение такой фактор, как наличие или отсутствие высшего образования. Введем новую, бинарную, переменную d, полагая $d_t = 1$, если в t-м наблюдении индивидуум имеет высшее образование, и $d_t = 0$ в противном случае, и рассмотрим новую систему

$$y_t = x_{t1}\beta_1 + \dots + x_{tk}\beta_k + d_t\delta + \varepsilon_t = z_t'\gamma + \varepsilon_t, \quad t = 1, \dots, n, \tag{4.2}$$

где $z = (x_1, \dots, x_k, d)' = (x', d)'$, $\gamma = (\beta_1, \dots, \beta_k, \delta)'$. Иными словами, принимая модель (4.2), мы считаем, что средняя зарплата есть $x'\beta$ при отсутствии высшего образования и $x'\beta + \delta$ — при его наличии. Таким образом, величина δ интерпретируется как среднее изменение зарплаты при переходе из одной категории (без высшего образования) в другую (с высшим образованием) при неизменных значениях остальных параметров. К системе (4.2) можно применить метод наименьших квадратов и получить оценки соответствующих коэффициентов. Легко понять, что, тестируя гипотезу $\delta = 0$, мы проверяем предположение о несущественном различии в зарплате между категориями.

Замечание. В англоязычной литературе по эконометрике переменные указанного выше типа называются *dummy variables*, что на русский язык часто переводится как «фиктивные переменные» (см., например, Джонстон, 1980). Следует, однако, ясно понимать, что d такая же «равноправная» переменная, как и любой из регрессоров x_j, $j = 1, \dots, k$. Ее «фиктивность» состоит только в том, что она количественным образом описывает качественный признак.

Качественное различие можно формализовать с помощью любой переменной, принимающей два значения, а не обязательно значения 0 или 1. Однако в эконометрической практике почти всегда используют лишь фиктивные переменные типа «0–1», поскольку в этом случае интерпретация выглядит наиболее просто. Если бы в рассмотренном выше примере переменная d принимала значение, скажем, 5 для индивидуума с высшим образованием и 2 для индивидуума без высшего образования, то коэффициент при

этом регрессоре равнялся бы трети среднего изменения зарплаты при получении высшего образования.

Если включаемый в рассмотрение качественный признак имеет не два, а несколько значений, то в принципе можно было бы ввести дискретную переменную, принимающую такое же количество значений. Но этого фактически никогда не делают, так как тогда трудно дать содержательную интерпретацию соответствующему коэффициенту. В этих случаях целесообразнее использовать несколько бинарных переменных. Типичным примером подобной ситуации является исследование сезонных колебаний. Пусть, например, y_t — объем потребления некоторого продукта в месяц t, и есть все основания считать, что потребление зависит от времени года. Для выявления влияния сезонности можно ввести три бинарные переменные d_1, d_2, d_3:

$d_{t1} = 1$, если месяц t является зимним, $d_{t1} = 0$ в остальных случаях;

$d_{t2} = 1$, если месяц t является весенним, $d_{t2} = 0$ в остальных случаях;

$d_{t3} = 1$, если месяц t является летним, $d_{t3} = 0$ в остальных случаях,

и оценивать уравнение

$$y_t = \beta_0 + \beta_1 d_{t1} + \beta_2 d_{t2} + \beta_3 d_{t3} + \varepsilon_t. \tag{4.3}$$

Отметим, что мы не вводим четвертую бинарную переменную d_4, относящуюся к осени, иначе тогда для любого месяца t выполнялось бы тождество $d_{t1} + d_{t2} + d_{t3} + d_{t4} = 1$, что означало бы линейную зависимость регрессоров в (4.3) и, как следствие, невозможность получения МНК-оценок. (Такая ситуация, когда сумма фиктивных переменных тождественно равна константе, также включенной в регрессию, называется «*dummy trap*».) Иными словами, среднемесячный объем потребления есть β_0 для осенних месяцев, $\beta_0+\beta_1$ — для зимних, $\beta_0+\beta_2$ — для весенних и $\beta_0+\beta_3$ — для летних. Таким образом, оценки коэффициентов β_i, $i = 1, 2, 3$, показывают средние сезонные отклонения в объеме потребления по

отношению к осенним месяцам. Тестируя, например, стандартную гипотезу $\beta_3 = 0$, мы проверяем предположение о несущественном различии в объеме потребления между летним и осенним сезонами, гипотеза $\beta_1 = \beta_2$ эквивалентна предположению об отсутствии различия в потреблении между зимой и весной и т. д.

Фиктивные переменные, несмотря на свою внешнюю простоту, являются весьма гибким инструментом при исследовании влияния качественных признаков. Рассмотрим еще один пример. В предыдущей модели мы интересовались сезонными различиями лишь для среднемесячного объема потребления. Модифицируем ее, введя новую независимую переменную i — доход, используемый на потребление. Как известно, в регрессии

$$y_t = \beta_0 + \beta_1 i_t + \varepsilon_t \tag{4.4}$$

коэффициент β_1 носит название «склонность к потреблению». Поэтому естественно поставить задачу исследовать влияние сезона на склонность к потреблению. Для этого можно рассмотреть модель

$$\begin{aligned} y_t = {} & \beta_0 + \beta_1 d_{t1} + \beta_2 d_{t2} + \beta_3 d_{t3} \\ & + \beta_4 d_{t1} i_t + \beta_5 d_{t2} i_t + \beta_6 d_{t3} i_t + \beta_7 i_t + \varepsilon_t, \end{aligned} \tag{4.5}$$

согласно которой склонность к потреблению зимой, весной, летом и осенью есть $\beta_4+\beta_7$, $\beta_5+\beta_7$, $\beta_6+\beta_7$ и β_7 соответственно. Как и в предыдущей модели, можно тестировать гипотезы об отсутствии сезонных влияний на склонность к потреблению.

Фиктивные переменные позволяют строить и оценивать так называемые кусочно-линейные модели, которые можно применять для исследования структурных изменений. Как и раньше, проще всего это продемонстрировать на примере.

Пусть y — зависимая переменная и пусть для простоты есть только две независимые переменные: x и постоянный член. Предположим, что x и y представлены в виде временных рядов $\{(x_t, y_t), t = 1, \ldots, n\}$ (например, x_t — размер основного фонда некоторого предприятия в период t, y_t — объем продукции, выпущенной в этот же период). Из некоторых априорных соображений иссле-

дователь считает, что в момент t_0 произошла структурная перестройка и линия регрессии будет отличаться от той, что была до момента t_0, но общая линия остается непрерывной (рис. 4.2).

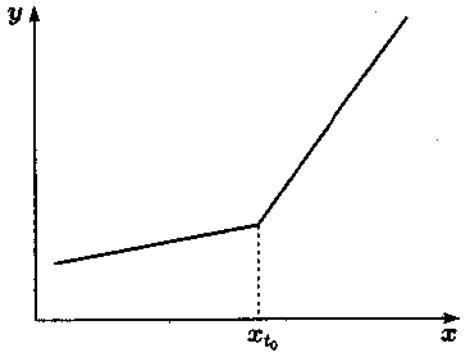


Рис. 4.2

Чтобы оценить такую модель, введем бинарную переменную r, полагая $r_t = 0$, если $t \leqslant t_0$ и $r_t = 1$, если $t > t_0$, и запишем следующее регрессионное уравнение:

$$y_t = \beta_1 + \beta_2 x_t + \beta_3 (x_t - x_{t_0}) r_t + \varepsilon_t. \qquad (4.6)$$

Нетрудно проверить, что регрессионная линия, соответствующая (4.6), имеет коэффициент наклона β_2 для $t \leqslant t_0$ и $\beta_2 + \beta_3$ для $t > t_0$, и разрыва в точке x_{t_0} не происходит. Таким образом, тестируя гипотезу $\beta_3 = 0$, мы проверяем предположение о том, что фактически структурного изменения не произошло.

Этот подход легко обобщается на случай нескольких структурных изменений в пределах одного временного интервала. Мы предлагаем читателю сделать это самостоятельно.

В заключение этого раздела отметим, что с помощью фиктивных переменных можно исследовать влияние разных качественных признаков (например, уровень образования и наличие или отсутствие детей), а также их взаимное влияние. Следует только быть внимательным, чтобы при включении нескольких бинарных переменных не нарушить линейную независимость регрессоров (см. выше пример с сезонными колебаниями).

Выводы:

1) для исследования влияния качественных признаков в модель можно вводить бинарные (фиктивные) переменные, которые, как правило, принимают значение 1, если данный качественный признак присутствует в наблюдении, и значение 0 при его отсутствии;

2) способ включения фиктивных переменных зависит от априорной информации относительно влияния соответствующих качественных признаков на зависимую переменную и от гипотез, которые проверяются с помощью модели;

3) от способа включения фиктивной переменной зависит и интерпретация оценки коэффициента при ней.

Пример. ***Рынок квартир в Москве*** (Каргин, Онацкий, 1996). Продолжение 1 (см. начало — п. 3.5).

Напомним, что рассматривается стоимость квартир в Москве. Построена оценка методом наименьших квадратов (см. п. 3.5, таблица 3.1, стр. 87):

$$\begin{aligned}\mathrm{LOGPRICE} = {} & \underset{(24.5)}{7.106} + \underset{(9.65)}{0.670}\,\mathrm{LOGLIVSP} + \underset{(8.71)}{0.431}\,\mathrm{LOGPLAN} \\ & + \underset{(2.45)}{0.147}\,\mathrm{LOGKITSP} - \underset{(-7.11)}{0.114}\,\mathrm{LOGDIST} \\ & - \underset{(-3.21)}{0.0686}\,\mathrm{FLOOR} + \underset{(5.67)}{0.134}\,\mathrm{BRICK} \\ & + \underset{(2.08)}{0.042}\,\mathrm{BAL} + \underset{(4.79)}{0.114}\,\mathrm{LIFT} + \underset{(1.957)}{0.214}\,\mathrm{R1} \\ & + \underset{(1.75)}{0.140}\,\mathrm{R2} + \underset{(2.74)}{0.164}\,\mathrm{R3} + \underset{(3.11)}{0.169}\,\mathrm{R4}. \qquad (*)\end{aligned}$$

Рассмотрим интерпретацию фиктивных переменных FLOOR, BRICK, BAL, LIFT, R1, R2, R3, R4.

Отрицательный коэффициент при FLOOR означает, что квартира на первом или последнем этаже стоит на 6.9% дешевле аналогичной квартиры на средних этажах. Квартира в кирпичном доме (BRICK = 1) стоит на 13.4% дороже аналогичной квартиры в панельном доме. Присутствие лифта (LIFT = 1) увеличивает стоимость квартиры на 11.4%, а наличие балкона (BAL = 1) — на 4.2%.

Умеренная величина t-статистики (2.08), соответствующей коэффициенту при BAL, может означать, что связь между ценой квартиры и наличием в ней балкона сомнительна. Возможно, что 5%-я значимость коэффициента при BAL получилась за счет существования косвенной зависимости цены квартиры от балкона. Например, балкона на первом этаже, скорее всего, нет, и, следовательно, значимый коэффициент при BAL может быть результатом некоего сложного комплекса влияний факторов этажности на цену.

Переменные R1, R2, R3 и R4 были включены в регрессию, чтобы учесть возможные различия в структуре рынка жилья для квартир с разным количеством комнат. Заметим, что в выборке есть пяти-, шести- и даже восьмикомнатные квартиры, поэтому переменные R1, R2, R3 и R4 в сумме не дают константу.

Выше (Пример. *Рынок квартир в Москве*, начало, п. 3.5, стр. 86) было показано, что коэффициенты при R2, R3 и R4 можно считать равными. Из уравнения (*), таким образом, видно, что квартиры с числом комнат от 2 до 4 стоят дороже многокомнатных, а однокомнатные — еще дороже (при прочих равных условиях).

4.3. Частная корреляция

В том случае, когда имеются одна независимая и одна зависимая переменные, естественной мерой зависимости (в рамках линейного подхода) является (выборочный) коэффициент корреляции между ними. Использование множественной регрессии позволяет обобщить это понятие на случай, когда имеется несколько независимых переменных. Корректировка здесь необходима по следующим очевидным соображениям. Высокое значение коэффициента корреляции между исследуемой зависимой и какой-либо независимой переменной может, как и раньше, означать высокую степень зависимости, но может быть обусловлено и другой причиной. А именно, есть третья переменная, которая оказывает сильное влияние на две первые, что и служит в конечном счете причиной их высокой коррелированности. Поэтому возникает естественная задача найти «чистую» корреляцию между двумя переменными, исключив (линейное) влияние других факторов. Это можно сделать с помощью коэффициента частной корреляции. Для простоты предположим, что имеется регрессионная модель

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

где, как обычно, y — $n \times 1$ вектор наблюдений зависимой переменной, x_1, x_2 — $n \times 1$ векторы независимых переменных, α, β_1, β_2 — (скалярные) параметры, ε — $n \times 1$ вектор ошибок. Наша цель — определить корреляцию между y и, например, первым регрессором x_1 после исключения влияния x_2.

Соответствующая процедура устроена следующим образом.

1. Осуществим регрессию y на x_2 и константу и получим прогнозные значения $\widehat{y} = \widehat{\alpha}_1 + \widehat{\alpha}_2 x_2$.
2. Осуществим регрессию x_1 на x_2 и константу и получим прогнозные значения $\widehat{x}_1 = \widehat{\gamma}_1 + \widehat{\gamma}_2 x_2$.
3. Удалим влияние x_2, взяв остатки $e_y = y - \widehat{y}$ и $e_{x_1} = x_1 - \widehat{x}_1$.
4. Определим (выборочный) *коэффициент частной корреляции* между y и x_1 при исключении влияния x_2 как (выборочный) коэффициент корреляции между e_y и e_{x_1}:

$$r(y,\, x_1|x_2) = r(e_y,\, e_{x_1}). \tag{4.7}$$

Напомним, что из свойств метода наименьших квадратов следует (см. (3.5)), что e_y и e_{x_1} не коррелированы с x_2. Именно в этом смысле указанная процедура соответствует интуитивному представлению об «исключении (линейного) влияния переменной x_2».

Прямыми вычислениями (см. упражнение 4.2) можно показать, что справедлива следующая формула, связывающая коэффициенты частной и обычной корреляции:

$$r(y,\, x_1|x_2) = \frac{r(y,x_1) - r(y,x_2)r(x_1,x_2)}{\sqrt{1 - r^2(x_1,x_2)}\sqrt{1 - r^2(y,x_2)}}. \tag{4.8}$$

Значения $r(y,\, x_1|x_2)$ лежат в интервале $[-1, 1]$, как у обычного коэффициента корреляции. Равенство коэффициента $r(y,\, x_1|x_2)$

нулю означает, говоря нестрого, отсутствие *прямого* (линейного) *влияния* переменной x_1 на y.

Существует тесная связь между коэффициентом частной корреляции $r(y, x_1|x_2)$ и коэффициентом детерминации R^2, а именно

$$r^2(y, x_1|x_2) = \frac{R^2 - r^2(y, x_2)}{1 - r^2(y, x_2)}, \qquad (4.9)$$

или

$$1 - R^2 = \big(1 - r^2(y, x_2)\big)\big(1 - r^2(y, x_1|x_2)\big).$$

Описанная выше процедура очевидным образом обобщается на случай, когда исключается влияние не одной, а нескольких переменных: достаточно переменную x_2 заменить на набор переменных X_2, сохраняя определение (4.7). Формула (4.8), естественно, усложнится. Подробнее об этом можно прочесть в книге (Айвазян и др., 1985).

Проиллюстрируем приведенное выше понятие частных коэффициентов корреляции и их отличие от обычных коэффициентов корреляции на следующем примере.

Пример. ***Рынки валютных фьючерсов.*** Рассмотрим вопрос о связи российского и западного рынков валютных фьючерсов.

В настоящее время несколько российских бирж ведут торговлю срочными контрактами на поставку доллара США: МТБ, МЦФБ, РТСБ и др. Однако (см. Яковлев, Бессонов, 1995а, 1995б) в течение периода наблюдений (ноябрь 1992 г. – сентябрь 1995 г.) на МТБ приходилось от 75 до 85% общего объема торговли. Поэтому в качестве цен фьючерсных контрактов на поставку доллара США мы выбрали котировки контрактов на МТБ.

Динамика цен валютных фьючерсов на Западе не сильно зависит от биржи. Для анализа мы взяли биржу с наибольшим объемом торговли — IMM (International Monetary Market, Chicago).

Мы используем ежедневные данные — цена закрытия для IMM и котировочная цена для МТБ — показатели, которые используют торговые палаты этих бирж для ежедневного перерасчета позиций инвесторов (вариационной маржи).

В качестве параметров для сравнения мы взяли не сами цены контрактов, а «доходности», приведенные к годичному базису,

определяемые как

$$y_{t,T} = (\ln F_t^T - \ln S_t)/(T - t) \cdot 365, \qquad (*)$$

где F_t^T — цена контракта в момент времени t на поставку 1 доллара в момент времени T (т. е. со сроком до поставки $T - t$); S_t — спот-курс доллара в момент t. (Для рубля — данные ММВБ, для немецкой марки DM, британского фунта BP, японской иены JY — данные IMM.) $y_{t,T}^{\mathrm{RU}}$, $y_{t,T}^{\mathrm{DM}}$, $y_{t,T}^{\mathrm{BP}}$, $y_{t,T}^{\mathrm{JY}}$ обозначают доходности (*) контрактов на поставку 1 доллара в рублях, DM, BP, JY. На наш взгляд, этот показатель в меньшей мере зависит от темпа инфляции, чем сама цена контракта. Время t измеряется в днях.

Рассмотрим таблицу коэффициентов корреляции доходностей $y_{t,T}^{\mathrm{RU}}$, $y_{t,T}^{\mathrm{DM}}$, $y_{t,T}^{\mathrm{BP}}$, $y_{t,T}^{\mathrm{JY}}$:

Таблица 4.1

	RU	DM	BP	JY
RU	1			
DM	0.626	1		
BP	0.380	0.775	1	
JY	0.615	0.919	0.602	1

Из таблицы 4.1 видны высокие (0.602, 0.775, 0.919) значения коэффициентов корреляции показателей для западных валют, что неудивительно ввиду высокой степени интегрированности западных финансовых рынков. Удивление вызывают высокие 0.615 (0.626) значения коэффициентов корреляции показателей для рубля и японской иены (немецкой марки).

Рассмотрим теперь таблицу коэффициентов частной корреляции между доходностями $y_{t,T}^{\mathrm{XX}}$ для XX = RU, DM, BP, JY (устранено влияние *временного тренда* t).

Таблица 4.2

	RU	DM	BP	JY
RU	1			
DM	0.024	1		
BP	0.008	0.807	1	
JY	-0.003	0.488	0.276	1

Теперь мы видим картину более реалистичную! Наиболее тесно связаны между собой европейские валюты (BP, DM), слабее связь европейских валют и японской иены и практически отсутствует связь российской валюты с западными.

Таким образом, высокие коэффициенты корреляции в первой таблице, например 0.626 для RU–DM, были лишь следствием того, что на интервале наблюдений (ноябрь 1992 г. – сентябрь 1995 г.) отмечалось падение курса рубля по отношению к доллару и падение курса доллара по отношению к немецкой марке, т. е. эта корреляция является следствием наличия временного тренда в $y_{t,T}^{\mathrm{RU}}$ и $y_{t,T}^{\mathrm{DM}}$. Наш вывод подтверждается также тем, что коэффициенты корреляции $y_{t,T}^{\mathrm{RU}}$ и $y_{t,T}^{\mathrm{DM}}$ с t достаточно высоки (−0.673; −0.920).

Процедура пошагового отбора переменных

Коэффициент частной корреляции часто используется при решении проблемы *спецификации модели* (см. далее п. 4.4). Остановимся на этом аспекте более подробно.

Иногда исследователь заранее знает характер зависимости исследуемых величин, опираясь, например, на экономическую теорию, предыдущие результаты, априорные знания и т. п., и задача состоит лишь в оценивании неизвестных параметров. (По существу, во всех наших предыдущих рассуждениях мы неявно предполагали, что имеется именно такая ситуация.) Классический пример — оценивание параметров производственной функции Кобба–Дугласа $Y = AK^{\alpha}L^{\beta}$, где Y — совокупный выпуск, K — капиталовложения и L — трудозатраты. Логарифмируя это равенство, получаем линейное относительно $\ln A$, α, β уравнение, из которого, например, с помощью метода наименьших квадратов можно получить оценки этих параметров, проверять те или иные гипотезы и т. д.

Однако на практике довольно часто приходится сталкиваться с ситуацией, когда имеется большое число наблюдений различных параметров (независимых переменных), но нет априорной модели изучаемого явления. Возникает естественная проблема, какие переменные включить в регрессионную схему. Теоретические вопросы, связанные с этой проблемой, будут изложены далее, в п. 4.4.

В компьютерные пакеты включены различные эвристические процедуры *пошагового отбора регрессоров*. Основными пошаговыми процедурами являются *процедура последовательного присоединения*, *процедура присоединения-удаления* и *процедура последовательного удаления*. Опишем кратко одну из таких процедур, использующую понятие коэффициента частной корреляции.

Процедура присоединения-удаления

На первом шаге из исходного набора объясняющих переменных выбирается (включается в число регрессоров) переменная, имеющая наибольший по модулю коэффициент корреляции с зависимой переменной y.

Второй шаг состоит из двух подшагов. На первом из них, который выполняется, если число регрессоров уже больше двух, делается попытка исключить один из регрессоров. Ищется тот регрессор x_s, удаление которого приводит к наименьшему уменьшению коэффициента детерминации. Затем сравнивается значение F-статистики (3.44) для проверки гипотезы H_0 о незначимости этого регрессора с некоторым заранее заданным пороговым значением $F_{\text{искл}}$. Если $F < F_{\text{искл}}$, то x_s удаляется из списка регрессоров. Заметим, что гипотеза H_0 о равенстве коэффициента при x_s нулю эквивалентна гипотезе о равенстве коэффициентов детерминации до и после удаления регрессора (см. (3.45)), а также гипотезе о том, что коэффициент частной корреляции x_s и y равен 0. Второй подшаг состоит в попытке включения нового регрессора из исходного набора предсказывающих переменных. Ищем переменную x_p с наибольшим по модулю частным коэффициентом корреляции (исключается влияние ранее включенных в уравнение регрессоров) и сравниваем значение F-статистики (3.44) для проверки гипотезы H_0 о незначимости этого регрессора с некоторым заранее заданным пороговым значением $F_{\text{вкл}}$. Если $F > F_{\text{вкл}}$, то x_p включается в список регрессоров. Обычно выбирают $F_{\text{искл}} < F_{\text{вкл}}$. Второй шаг повторяется до тех пор, пока происходит изменение списка регрессоров. Конечно, ни одна из пошаговых процедур не гарантирует получение оптимального по какому-либо критерию набора регрессоров.

Подробное описание пошаговых процедур содержится в книге (Айвазян и др., 1985).

Следует отметить, что пошаговый отбор является формально-аналитической процедурой, и его надо рассматривать как вспомогательный метод. Основным критерием является содержательный экономический смысл модели.

4.4. Спецификация модели

Все наши предыдущие рассуждения и выводы, касающиеся схемы классической множественной регрессии, основывались явно или неявно на предположении, что мы имеем дело с правильной *спецификацией модели*. Иными словами, мы считали, что зависимая переменная y, регрессоры X и оцениваемые параметры β связаны соотношением

$$y = X\beta + \varepsilon, \tag{4.10}$$

и выполнены условия 1–3 п. 3.1. При этом часто говорят, что соотношение (4.10) описывает «процесс, порождающий данные» или что (4.10) является «истинной моделью». Как правило, на практике истинная модель неизвестна, исследователь оценивает модель, которая лишь приближенно соответствует процессу, порождающему данные. (Напомним, что сам выбор регрессоров называется *спецификацией модели*.) Поэтому возникает естественный вопрос соотношения между МНК-оценками параметров в истинной и выбранной моделях. Мы рассмотрим две, в определенном смысле противоположные, ситуации: в оцениваемой модели отсутствует часть независимых переменных, имеющихся в истинной модели (исключение существенных переменных); в оцениваемой модели присутствуют независимые переменные, которых нет в истинной модели (включение несущественных переменных).

Итак, будем изучать два основных случая.

Случай 1. *Исключены существенные переменные.*

Процесс, порождающий данные: $y = X\beta + Z\gamma + \varepsilon.$ (4.11а)

Модель: $y = X\beta + \varepsilon.$ (4.11б)

Случай 2. *Включены несущественные переменные.*

Процесс, порождающий данные: $y = X\beta + \varepsilon.$ (4.12а)

Модель: $y = X\beta + Z\gamma + \varepsilon.$ (4.12б)

Здесь X — $n \times k$ матрица; Z — $n \times l$ матрица; y — $n \times 1$ вектор наблюдений зависимой переменной; β — $k \times 1$, γ — $l \times 1$ векторы коэффициентов. Часто регрессию (4.11а) называют длинной, а регрессию (4.11б) — короткой.

Исключение существенных переменных

Напомним, что МНК-оценка вектора параметров β для модели (4.11б) (в короткой регрессии) имеет вид (см. (3.4)):

$$\widehat{\beta}_{\text{OLS}} = \widehat{\beta} = (X'X)^{-1}X'y. \tag{4.13}$$

Обозначим также $\widehat{\delta}^* = \begin{bmatrix} \widehat{\beta}^* \\ \widehat{\gamma}^* \end{bmatrix}$ МНК-оценку вектора коэффициентов $\delta = \begin{bmatrix} \beta \\ \gamma \end{bmatrix}$ в длинной регрессии (в истинной модели) (4.11а).

Учитывая (4.11а), получаем (ср. (3.7), (3.8)):

$$\begin{aligned} \mathrm{E}\widehat{\beta} &= (X'X)^{-1}X'\mathrm{E}y = \beta + (X'X)^{-1}X'Z\gamma, \\ \mathrm{V}(\widehat{\beta}) &= \sigma^2(X'X)^{-1}. \end{aligned} \tag{4.14}$$

Из (4.14) видно, что оценка $\widehat{\beta}$, вообще говоря, смещена за исключением двух случаев:

а) $\gamma = 0$ (очевидный случай) и

б) $X'Z = 0$ (ортогональность регрессоров X и Z).

Рассмотрим эту ситуацию более подробно. Имеем:

$$[X \quad Z]'\,[X \quad Z] = \begin{bmatrix} X' \\ Z' \end{bmatrix}[X \quad Z] = \begin{bmatrix} X'X & 0 \\ 0 & Z'Z \end{bmatrix},$$

$$\Big[[X \quad Z]'\,[X \quad Z]\Big]^{-1} = \begin{bmatrix} (X'X)^{-1} & 0 \\ 0 & (Z'Z)^{-1} \end{bmatrix},$$

$$\widehat{\delta}^* = \left[[X \;\; Z]' \, [X \;\; Z]\right]^{-1} [X \;\; Z]' \, y = \begin{bmatrix} (X'X)^{-1}X'y \\ (Z'Z)^{-1}Z'y \end{bmatrix}.$$

Таким образом, $\widehat{\beta} = \widehat{\beta}^*$, т. е. МНК-оценки вектора β, полученные в длинной и короткой регрессиях, совпадают. (Если пользоваться геометрической интерпретацией, то содержательно полученный результат выражает хорошо известную теорему о трех перпендикулярах.)

Сумма квадратов остатков вычисляется по формуле (см. (3.11), (3.20))

$$\mathrm{ESS} = y'My, \quad M = I - X(X'X)^{-1}X'.$$

Так как $MX = 0$, то, согласно известным формулам (см. приложение МС, п. 4, N8), получаем:

$$\begin{aligned} \mathrm{E(ESS)} &= \mathrm{E}(y'My) = \mathrm{E}((X\beta + Z\gamma + \varepsilon)'M(X\beta + Z\gamma + \varepsilon)) \\ &= \mathrm{E}(\varepsilon'M\varepsilon + 2\gamma'Z'M\varepsilon + \gamma'Z'MZ\gamma) \\ &= \sigma^2(n-k) + \gamma'Z'MZ\gamma. \end{aligned} \tag{4.15}$$

Последнее слагаемое в (4.15) неотрицательно при любом γ, поскольку $Z'MZ = Z'M^2Z = (MZ)'(MZ)$, т. е. матрица $Z'MZ$ неотрицательно определена. Таким образом, оценка $s^2 = \widehat{\sigma}^2 = e'e/(n-k) = \sum e_t^2/(n-k)$ является, вообще говоря, смещенной оценкой дисперсии σ^2: $\mathrm{E}(\widehat{\sigma}^2) \geqslant \sigma^2$.

Заметим, что если $X'Z = 0$, то оценка $\widehat{\beta}$ несмещена, но оценка $\widehat{\sigma}^2$ по-прежнему смещена:

$$\mathrm{E}(\widehat{\sigma}^2) = \sigma^2 + \frac{1}{n-k}\gamma'Z'Z\gamma > \sigma^2.$$

Отметим, что так как $\widehat{\sigma}^2$ участвует во многих статистических тестах, то в этом случае можно получить ложные выводы.

Сравним теперь ковариационные матрицы оценок $\widehat{\beta}$ и $\widehat{\beta}^*$. Из (4.13) следует, что

$$\mathrm{V}(\widehat{\beta}) = \sigma^2(X'X)^{-1}, \tag{4.16}$$

а ковариационная матрица вектора $\widehat{\beta}^*$ есть левый верхний квадратный блок размера k матрицы $\mathrm{V}(\widehat{\delta}^*) = \sigma^2([X \;\; Z]' \, [X \;\; Z])^{-1}$. Вспоминая правило обращения блочных матриц (см. приложение

ЛА, п. 17), получаем

$$\mathrm{V}(\widehat{\beta}^*) = \sigma^2(X'X - X'Z(Z'Z)^{-1}Z'X)^{-1}. \tag{4.17}$$

Поэтому в силу (4.16) и (4.17)

$$(\mathrm{V}(\widehat{\beta}))^{-1} - (\mathrm{V}(\widehat{\beta}^*))^{-1} = \frac{1}{\sigma^2}X'Z(Z'Z)^{-1}Z'X \geqslant 0.$$

Это, в свою очередь, означает, что $\mathrm{V}(\widehat{\beta}^*) \geqslant \mathrm{V}(\widehat{\beta})$. Иными словами, оценка, полученная в короткой регрессии, в общем случае смещена, но обладает меньшей вариацией.

Для наглядности рассмотрим простейший случай $k = l = 1$, т. е. предположим, что длинная регрессия есть

$$y = x\beta + z\gamma + \varepsilon$$

(мы здесь предполагаем, что $\overline{y} = \overline{x} = \overline{z} = 0$), а короткая —

$$y = x\beta + \varepsilon,$$

где y, x, z, ε — $n \times 1$ векторы, β, γ — скаляры. Тогда (см. (2.6))

$$\widehat{\beta} = \frac{x'y}{x'x} = \frac{x'(x\beta + z\gamma + \varepsilon)}{x'x} = \beta + \frac{x'z}{x'x}\gamma + \frac{x'\varepsilon}{x'x}$$

и

$$\mathrm{E}\widehat{\beta} = \beta + \frac{x'z}{x'x}\gamma.$$

При этом

$$\mathrm{V}(\widehat{\beta}) = \frac{\sigma^2}{x'x}, \qquad \mathrm{V}(\widehat{\beta}^*) = \frac{\sigma^2}{x'x(1-r^2)}, \tag{4.18}$$

где r — выборочный коэффициент корреляции между x и z.

Включение несущественных переменных

Можно получить два эквивалентных выражения для оценки по модели (4.12б) (упражнение 4.3):

$$\widehat{\beta} = (X'X)^{-1}X'y - (X'X)^{-1}X'Z(Z'MZ)^{-1}Z'My, \tag{4.19}$$

$$\widehat{\beta} = (X'M_ZX)^{-1}X'M_Zy, \quad M_Z = I - Z(Z'Z)^{-1}Z'. \tag{4.20}$$

Из (4.19) или (4.20) получаем

$$\mathrm{E}\widehat{\beta} = \beta, \qquad \mathrm{V}(\widehat{\beta}) = \sigma^2(X'M_ZX)^{-1},$$

$$\mathrm{V}(\widehat{\beta}) = \sigma^2\big((X'X)^{-1} + (X'X)^{-1}X'Z(Z'MZ)^{-1}Z'X(X'X)^{-1}\big),$$
$$\mathrm{V}(\widehat{\beta}) > \sigma^2(X'X)^{-1}.$$

Таким образом, как и следовало ожидать, в этом случае оценка $\widehat{\beta}$ несмещенная, однако дисперсия оценки увеличивается от включения в модель несущественных переменных.

Рассматривая оценку для σ^2, получаем

$$e^{*\prime}e^* = e^{*\prime}M^*e^* = (y - X\widehat{\beta} - Z\widehat{\gamma})'M^*(y - X\widehat{\beta} - Z\widehat{\gamma}) = y'M^*y$$

($X^* = [X\ Z]$, $M^* = I - X^*(X^{*\prime}X^*)^{-1}X^{*\prime}$), так как $M^*X^* = [M^*X\ \ M^*Z] = 0$ и, значит, $M^*X = 0$, $M^*Z = 0$. Аналогично,

$$y'M^*y = (X\beta + \varepsilon)'M^*(X\beta + \varepsilon) = \varepsilon'M^*\varepsilon,$$

$$\mathrm{E}\frac{\varepsilon'M^*\varepsilon}{\sigma^2} = \operatorname{rank} M^* = n - k - l, \qquad \mathrm{E}\widehat{\sigma}^2 = \mathrm{E}\frac{e^{*\prime}e^*}{n-k-l} = \sigma^2.$$

Итак, обе оценки $\widehat{\beta}$ и $\widehat{\sigma}^2$ являются несмещенными.

Поскольку включение несущественных переменных сохраняет несмещенность, у исследователя может возникнуть соблазн включать в модель как можно больше объясняющих переменных с целью получить лучшую подгонку. Однако следует помнить, что точность оценок при этом уменьшается. Кроме того, увеличение числа регрессоров часто приводит к неустойчивости модели из-за наличия сильной корреляции между ними (см. рассмотренную выше (п. 4.1) проблему мультиколлинеарности). В простейшем случае указанное обстоятельство наглядно демонстрирует формула (4.18): при стремлении r^2 к 1 дисперсия оценки $\widehat{\beta}^*$ стремится к бесконечности.

Короткая или длинная регрессия?

До сих пор мы рассматривали общий случай. С этого момента и до конца главы будем предполагать, что $l = 1$.

Рассмотрим теперь другую, более близкую к практике, постановку задачи, когда нам неизвестен процесс, порождающий данные («истинная модель»).

Таким образом, мы сравниваем две модели:

I. $y = X\beta + z\gamma + \varepsilon$ (модель без ограничений),

II. $y = X\beta + \varepsilon$ (модель с ограничением),

где z — *один* дополнительный регрессор (X — $n \times k$, z — $n \times 1$ матрицы). (Использование терминов «без ограничений» и «с ограничением» совершенно понятно: если в модели I наложить ограничение $\gamma = 0$, то получается модель II. Далее мы будем использовать индексы u (*unrestricted*) и r (*restricted*) для моделей без ограничения и с ограничением.)

Как выбрать одну из этих моделей? Рассмотрим разные возможности сравнения.

***Способ* 1** (основан на R^2).

Нам известно, что всегда $R_u^2 \geqslant R_r^2$ (см. (3.45)), поскольку $e_u'e_u \leqslant e_r'e_r$. Поэтому такой способ плох.

***Способ* 2** (основан на R^2_{adj}).

По определению (см. главу 3)

$$R^2 = 1 - \frac{e'e}{y'Ay}, \quad A = I_n - \frac{1}{n}\imath\imath', \quad R^2_{\mathrm{adj}} = 1 - \frac{e'e/(n-k)}{y'Ay/(n-1)}.$$

Для моделей с ограничением и без ограничения получаем

$$R^2_{u,\,\mathrm{adj}} = 1 - \frac{e_u'e_u/(n-k-1)}{y'Ay/(n-1)}, \quad R^2_{r,\,\mathrm{adj}} = 1 - \frac{e_r'e_r/(n-k)}{y'Ay/(n-1)}$$

и

$$R^2_{r,\,\mathrm{adj}} - R^2_{u,\,\mathrm{adj}} = \frac{e_u'e_u/(n-k-1) - e_r'e_r/(n-k)}{y'Ay/(n-1)}.$$

Мы ранее показали (см. (3.44)), что гипотеза H_0: $\gamma = 0$ проверяется с помощью F-статистики (или t-статистики):

$$F = \frac{(e_r'e_r - e_u'e_u)/1}{e_u'e_u/(n-k-1)} \sim F(1, n-k-1) \sim t^2(n-k-1).$$

Следовательно,

$$R^2_{r,\,\mathrm{adj}} - R^2_{u,\,\mathrm{adj}} = \frac{e_u'e_u/(n-k-1)}{y'Ay/(n-1)} \cdot \frac{1-t^2}{n-k}.$$

Таким образом, если $|t| > 1$, то $R^2_{r,\mathrm{adj}} < R^2_{u,\mathrm{adj}}$, и наоборот. Если мы верим, что скорректированный коэффициент детерминации R^2_{adj} является правильным критерием, то мы должны выбрать модель без ограничения тогда и только тогда, когда $|t| > 1$. Ниже мы увидим, что $t = 1$ является естественной границей.

Вернемся к исходной задаче. У нас есть модель с ограничением и модель без ограничения, и наша цель — оценить $\boldsymbol{\beta}$. Взяв модель с ограничением, мы можем получить смещенную оценку, взяв модель без ограничений, можем получить неэффективную оценку. Естественным компромиссом является

***Способ* 3** (основан на наименьшем среднеквадратичном отклонении MSE, Mean Squared Error).

Будем сравнивать модели I и II по критерию $\mathrm{MSE}(\widehat{\boldsymbol{\beta}}) = \mathrm{E}((\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}))$. Как и ранее, обозначим $\boldsymbol{M} = \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. Введем также обозначения

$$\boldsymbol{q} = \frac{\sigma}{\sqrt{\boldsymbol{z}'\boldsymbol{M}\boldsymbol{z}}}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{z}, \qquad \theta = \frac{\gamma}{\sigma/\sqrt{\boldsymbol{z}'\boldsymbol{M}\boldsymbol{z}}}.$$

Предположим, что вектор ошибок имеет стандартное многомерное нормальное распределение. Тогда из предыдущих результатов (см. (4.13), (4.14), (4.19), (4.20)) следует, что

$$\widehat{\boldsymbol{\beta}}_r \sim N(\boldsymbol{\beta} + \theta\boldsymbol{q}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}), \qquad \widehat{\boldsymbol{\beta}}_u \sim N(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} + \boldsymbol{q}\boldsymbol{q}').$$

Отсюда, пользуясь свойствами нормального распределения, получаем (см. приложение МС, п. 4)

$$\mathrm{MSE}(\widehat{\boldsymbol{\beta}}_r) - \mathrm{MSE}(\widehat{\boldsymbol{\beta}}_u) = (\theta^2 - 1)\boldsymbol{q}'\boldsymbol{q}.$$

Мы снова видим важность условия $|\theta| > 1$. Но на этот раз θ — «теоретическое t-отношение», а не то, которое получено из наблюдений.

Полученные результаты дают пищу для размышления. В том случае, когда мы хотим оценить коэффициент $\boldsymbol{\beta}$ и не уверены, должна ли переменная z присутствовать в модели, то вопрос «Верна ли гипотеза H_0: $\gamma = 0$?» не является подходящим! Ответ на него покажет, равно γ нулю или нет, но это не то, что мы хотели бы знать, а именно, «что лучше — $\widehat{\boldsymbol{\beta}}_r$ или $\widehat{\boldsymbol{\beta}}_u$?» Правильным

будет вопрос: «Верна ли гипотеза H_0: $|\theta| > 1$?», так как именно это условие различает, какое из двух чисел больше: $\mathrm{MSE}(\widehat{\boldsymbol{\beta}}_r)$ или $\mathrm{MSE}(\widehat{\boldsymbol{\beta}}_u)$.

Аспект выбора модели и аспект оценки параметров модели должны быть, строго говоря, совмещены.

Предыдущая оценка, основанная на предварительном тесте, может быть записана в виде

$$\widehat{\boldsymbol{\beta}} = \lambda(\widehat{\theta})\widehat{\boldsymbol{\beta}}_u + (1 - \lambda(\widehat{\theta}))\widehat{\boldsymbol{\beta}}_r, \quad \text{где } \lambda(\widehat{\theta}) = \begin{cases} 0, & |\widehat{\theta}| \leqslant 1, \\ 1, & |\widehat{\theta}| > 1. \end{cases}$$

Такую оценку естественно назвать «крайней». Можно ее улучшать, выбирая $\lambda(\widehat{\theta})$, например, в виде

$$\lambda(\widehat{\theta}) = \frac{\widehat{\theta}^2}{c^2 + \widehat{\theta}^2}, \quad 0 \leqslant c \leqslant \infty.$$

Выбор оценки также может быть основан на байесовском подходе, но эти аспекты выходят за пределы нашей книги.

Более подробно свойства оценок, полученных в результате предварительного тестирования, будут рассмотрены в главе 14.

Выводы:

1) при исключении существенных переменных МНК-оценка, получаемая в короткой регрессии, в общем случае смещена и обладает меньшей ковариационной матрицей, чем оценка, полученная в длинной регрессии (в истинной модели);

2) если $\gamma = 0$, то смещение отсутствует;

3) если исключаемые независимые переменные ортогональны оставшимся ($\boldsymbol{X}'\boldsymbol{Z} = \boldsymbol{0}$), то оценки, полученные в длинной и короткой регрессиях, совпадают, но остатки регрессий разные и оценка дисперсии смещена;

4) оценка дисперсии в короткой регрессии имеет неотрицательное смещение;

5) при включении несущественных переменных оценка параметров $\boldsymbol{\beta}$ является несмещенной;

6) ее ковариационная матрица больше, чем у оценки, получаемой в истинной модели;

7) оценка дисперсии также является несмещенной.

Сравнение не вложенных моделей

Рассмотрим простой случай, когда надо выбрать между двумя не вложенными друг в друга линейными моделями. Пусть есть две модели:

$$\text{модель A}: \quad y_t = x_t'\beta + \varepsilon_t \tag{4.21}$$

и

$$\text{модель B}: \quad y_t = z_t'\gamma + \nu_t. \tag{4.22}$$

В работе (Davidson and MacKinnon, 1981) предложена следующая процедура, названная J-тестом. Построим модель, включающую как частный случай модели A и B:

$$y_t = (1-\delta)x_t'\beta + \delta z_t'\gamma + u_t. \tag{4.23}$$

При $\delta = 0$ модель (4.23) совпадает с моделью A, а при $\delta = 1$ — с моделью B. Однако уравнение (4.23) невозможно оценить, поскольку параметры β, γ, δ не могут быть идентифицированы одновременно. Оценим из уравнения (4.22) параметр γ и заменим в уравнении (4.23) γ на полученную оценку $\widehat{\gamma}$:

$$y_t = (1-\delta)x_t'\beta + \delta z_t'\widehat{\gamma} + u_t = x_t'\beta^* + \delta\widehat{y}_{tB} + u_t. \tag{4.24}$$

Здесь $\widehat{y}_{tB}$ — прогнозные значения, полученные по модели B, а $\beta^* = (1-\delta)\beta$. Из уравнения (4.24) можно получить оценку $\widehat{\delta}$. Как показали Davidson и MacKinnon, если верна нулевая гипотеза (модель A), то $p\lim\widehat{\delta} = 0$ и обычная t-статистика коэффициента δ имеет асимптотически стандартное нормальное распределение. Если полученное значение t-статистики больше критического, то нулевая гипотеза отвергается. Аналогичную процедуру можно проделать, взяв за нулевую гипотезу модель B. В двух из четырех возможных исходов теста, когда обе модели отвергаются или обе модели не отвергаются, ситуация остается неопределенной.

Другая ситуация необходимости сравнения двух не вложенных моделей возникает, когда, скажем, надо сделать выбор между линейной и лог-линейной моделями. Поскольку в этих моделях разные зависимые переменные (y_t и $\ln y_t$), то критерии качества подгонки модели, такие как коэффициент детерминации R^2 или критерии Акаике или Шварца (см. главу 11, (11.97), (11.98)), неприменимы. В работе (MacKinnon et al., 1983) предложен PE-тест, который состоит в следующем. Оценим обе модели, линейную и лог-линейную, методом наименьших квадратов и получим соответствующие прогнозные значения $\widehat{y}_t$ и $\widehat{\ln y_t}$. Тогда мы можем тестировать гипотезу H_0: линейная модель против альтернативной гипотезы H_1: лог-линейная модель, проверяя гипотезу $\delta_{LIN} = 0$ (с помощью обычной t-статистики, которая имеет при нулевой гипотезе приблизительно стандартное нормальное распределение) в уравнении

$$y_t = x_t'\beta + \delta_{LIN}(\ln\widehat{y}_t - \widehat{\ln y_t}) + \varepsilon_t. \tag{4.25}$$

Аналогично, проверяя гипотезу $\delta_{LOG} = 0$ в уравнении

$$y_t = (\ln x_t)'\beta + \delta_{LOG}(\widehat{y}_t - \exp(\widehat{\ln y_t}) + \varepsilon_t, \tag{4.26}$$

можно тестировать лог-линейную модель относительно линейной модели. Как и ранее, возможны четыре исхода теста.

Заметим, что PE-тест может применяться в значительно более общей ситуации, подробнее см. (MacKinnon et al., 1983) или (Greene, 1997, п. 10.2.4).

Тест на функциональную форму

Самый простой способ тестировать справедливость линейной спецификации модели (4.27)

$$y_t = x_t'\beta + \varepsilon_t \tag{4.27}$$

— это добавить в правую часть нелинейные члены и тестировать их значимость с помощью обычного F-теста. Недостаток этого метода состоит в том, что мы должны задавать альтернативную гипотезу. В работе (Ramsey, 1969) предложен $RESET$-тест (*Regression Equation Specification Error Test*), основанный на следующей идее. Если модель (4.27) верна, то добавление нелинейных

функций $\widehat{y}_t = x_t'\widehat{\beta}$ не должно помогать объяснять y_t. В частности, можно добавлять степени:

$$y_t = x_t'\beta + \alpha_2\widehat{y}_t^2 + \alpha_3\widehat{y}_t^3 + \dots + \alpha_m\widehat{y}_t^m + \varepsilon_t. \qquad (4.28)$$

Гипотезу H_0: $\alpha_2 = \ldots = \alpha_m$ можно тестировать с помощью обычного F-теста или теста Вальда (см. главу 10, п. 10.6). Обычно тест применяется при небольших значениях $m = 2, 3, 4$. Заметим, что тест может отвергать нулевую гипотезу не потому, что в истинной модели есть нелинейные члены, а в силу того, что в уравнении (4.27) пропущена переменная, влияние которой частично учтено нелинейными членами в (4.28).

Пример. ***Зарплата в Нидерландах*** (Arthur van Soest). Данные содержатся в файле `wages.wf1`. Имеется 150 наблюдений, 75 мужчин и 75 женщин, работавших на полную ставку (не менее 4 дней в неделю в 1987 г.). Переменные: W — зарплата (гульденов/час) до вычета налогов; AGE — возраст (лет); SEX — 1 (мужчины), 2 (женщины); EDU — уровень образования: 1 (начальная школа или менее), 2 (низшее ремесленное), 3 (среднее), 4 (высшее ремесленное), 5 (университет).

Рассмотрим две простейшие модели для зависимости зарплаты от образования, пола и возраста: линейную и полулогарифмическую. Получаем следующие оценки (в скобках указаны P-значения):

$$W = \underset{(0.258)}{3.515} - \underset{(0.004)}{3.551}SEX + \underset{(0.000)}{3.245}EDU + \underset{(0.000)}{0.441}AGE, \qquad (4.29)$$

$$\ln W = \underset{(0.000)}{2.816} - \underset{(0.001)}{0.167}SEX + \underset{(0.000)}{0.142}EDU + \underset{(0.000)}{0.020}AGE. \qquad (4.30)$$

Коэффициенты детерминации равны соответственно 0.516 и 0.580, однако, поскольку зависимые переменные в уравнениях разные, мы не можем использовать их для выбора модели.

Применим описанную выше процедуру PE-теста:

$$W = \underset{(0.307)}{3.151} - \underset{(0.002)}{3.831}SEX + \underset{(0.000)}{3.545}EDU + \underset{(0.000)}{0.453}AGE - \underset{(0.045)}{13.96}(\ln\widehat{W} - \widehat{\ln W}), \qquad (4.31)$$

$$\ln W = \underset{(0.000)}{2.173} - \underset{(0.001)}{0.164}SEX + \underset{(0.000)}{0.139}EDU + \underset{(0.000)}{0.020}AGE - \underset{(0.234)}{0.017}(\widehat{W} - \exp(\widehat{\ln W})). \qquad (4.32)$$

Получаем, что на 5%-ном уровне значимости нулевая гипотеза линейной модели отвергается (P-значение 0.045 в уравнении (4.31)), а нулевая гипотеза полулогарифмической модели не отвергается (P-значение 0.234 в уравнении (4.32)).

$RESET$-тест для уравнения (4.30) при $m = 3$ дает следующий результат:

$$\ln W = \underset{(0.001)}{41.94} - \underset{(0.001)}{5.898}SEX + \underset{(0.001)}{5.001}EDU + \underset{(0.001)}{0.705}AGE - \underset{(0.001)}{11.01}(\widehat{\ln W})^2 + \underset{(0.0.002)}{1.164}(\widehat{\ln W})^3.$$

Величина F-статистики для тестирования гипотезы о равенстве двух последних коэффициентов во вспомогательном уравнении равна 6.81 (соответствующее P-значение равно 0.0015), поэтому тест показывает на ошибочную спецификацию уравнения (4.30). В уравнение надо попробовать включить нелинейные члены (например, AGE^2) или другие переменные.

Упражнения

4.1. С помощью бинарных переменных напишите уравнение, соответствующее наличию двух структурных изменений в моменты времени t_0 и t_1 (предполагается, что $t_0 < t_1$).

4.2. Докажите равенство (4.8).

4.3. Докажите эквивалентность (4.19) и (4.20).
(*Указание.* Первое выражение получается из блочного представления матриц; второе получается после замены X и y на их остатки при регрессии на Z.)

4.4. Предположим, что вы оцениваете линейную функцию потребления $c_t = \alpha + \beta y_t + \varepsilon_t$ среди n индивидуумов. Как учесть возможный сдвиг этой функции при переходе от городского к сельскому потребителю, если вы считаете, что маргинальная склонность к потреблению постоянна, в то время как средняя склонность к потреблению может меняться? Как проверить гипотезу о том, что маргинальные склонности к потреблению индивидуумов с доходом выше и ниже уровня y^* отличаются?

4.5. Рассмотрим регрессию

$$y_t = \beta_1 + \beta_2 d_t + \varepsilon_t, \quad t = 1, \ldots, n,$$

где d — некоторая фиктивная переменная. Пусть $\overline{y}_0$ — среднее значение переменной y по n_0 наблюдениям, для которых $d = 0$ и $\overline{y}_1$ — среднее значение переменной y по n_1 наблюдениям, для которых $d = 1$ $(n_0 + n_1 = n)$. Найдите $\mathrm{V}(\widehat{\beta}_1)$, $\mathrm{V}(\widehat{\beta}_2)$.

4.6. На основе квартальных данных с 1971 по 1976 г. с помощью метода наименьших квадратов получено следующее уравнение:

$$y_t = \underset{(2.14)}{1.12} - \underset{(0.0034)}{0.0098} x_{t1} - \underset{(3.42)}{5.62} x_{t2} + \underset{(0.009)}{0.044} x_{t3},$$

в скобках указаны стандартные ошибки, RSS = 110.32, ESS = 21.43.

а) Проверьте значимость каждого из коэффициентов.

б) Найдите коэффициент детерминации.

в) Протестируйте значимость регрессии в целом.

г) Когда в уравнение были добавлены три фиктивные переменные, соответствующие трем первым кварталам года, величина RSS выросла до 118.20. Проверьте гипотезу о наличии сезонности, сформулировав необходимые предположения о виде этой сезонности.

д) Для той же исходной модели были раздельно проведены две регрессии на основе данных: 1-й квартал 1971 г. – 1-й квартал 1975 г. и 2-й квартал 1975 г. – 4-й квартал 1976 г., соответственно, и получены следующие значения сумм квадратов остатков: $\mathrm{ESS}_1 = 12.25$, $\mathrm{ESS}_2 = 2.32$. Проверьте гипотезу о том, что между 1-м и 2-м кварталами 1975 г. произошло структурное изменение.

4.7. Процесс, порождающий данные (истинная модель), описывается соотношениями

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \varepsilon_t,$$
$$\mathrm{E}(\varepsilon_t) = 0, \quad \mathrm{E}(\varepsilon_t^2) = \sigma^2, \quad \mathrm{E}(\varepsilon_t \varepsilon_s) = 0, \quad t \neq s, \quad t = 1, \ldots, n.$$

Обозначим через $\widehat{\beta}_1$ МНК-оценку параметра β_1 в этой регрессии, а через $\widehat{\beta}_1^*$ — МНК-оценку параметра β_1 в регрессии y только на x_1.

а) Покажите, что

$$\frac{\mathrm{MSE}(\widehat{\beta}_1^*)}{\mathrm{MSE}(\widehat{\beta}_1)} = 1 + r_{12}^2 (t_2^2 - 1),$$

где

$$r_{12}^2 = \frac{(\sum x_1 x_2)^2}{\sum x_1^2 \sum x_2^2}, \quad t_2^2 = \frac{\beta_2^2}{\mathrm{V}(\widehat{\beta}_2)}.$$

б) Рассмотрим смесь оценок $\widetilde{\beta}_1 = \lambda \widehat{\beta}_1 + (1-\lambda)\widehat{\beta}_1^*$. При каком значении λ величина $\mathrm{MSE}(\widetilde{\beta}_1)$ минимальна?

4.8. Процесс, порождающий данные (истинная модель), описывается соотношениями:

$$y_t = \beta x_t + \gamma z_t + \varepsilon_t,$$
$$\mathrm{E}(\varepsilon_t) = 0, \quad \mathrm{E}(\varepsilon_t^2) = \sigma^2, \quad \mathrm{E}(\varepsilon_t \varepsilon_s) = 0, \quad t \neq s, \quad t = 1, \ldots, n.$$

Переменная x наблюдается с ошибками, т. е. в регрессии могут быть использованы лишь величины $w_t = x_t + u_t$, при этом предполагается, что ошибки u удовлетворяют условиям $\mathrm{E}(u_t) = 0$, $\mathrm{E}(u_t^2) = \sigma^2$, $\mathrm{E}(u_t u_s) = 0$, $t \neq s$, $\mathrm{E}(u_t \varepsilon_s) = 0$, $\forall s, t$. Проводятся две регрессии: первая — y на z; вторая — y на z и w.

Покажите, что смещение оценки параметра γ во второй регрессии меньше, чем в первой.

4.9. Процесс, порождающий данные (истинная модель), описывается соотношениями:

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \varepsilon_t,$$
$$\mathrm{E}(\varepsilon_t) = 0, \quad \mathrm{E}(\varepsilon_t^2) = \sigma^2, \quad \mathrm{E}(\varepsilon_t \varepsilon_s) = 0, \quad t \neq s, \quad t = 1, \ldots, n.$$

Проводится регрессия y на x_1 и стандартным образом через остатки этой регрессии оценивается дисперсия σ^2. Покажите, что полученная оценка смещена вверх.

4.10. Предположим, что некоторые ежегодные данные удовлетворяют соотношениям:

$$y_t = \alpha + \beta_1 x_t + \beta_2 t + \varepsilon_t \quad \text{(истинная модель)},$$

причем выполнены все условия классической регрессии. Однако оценивается «неправильная» модель без временного тренда

$$y_t = a + b_1 x_t + \nu_t.$$

а) Какие из условий классической регрессии не выполнены для модели без временного тренда?

б) Будет ли равна нулю сумма остатков для этой регрессии? Как это связано с ошибочным предположением, что $\mathrm{E}(\nu_t) = 0$?

в) Предположим, что коэффициент β_2 положителен и нарисован график остатков регрессии $y_t = a + b_1 x_t + \nu_t$ как функция времени. Как должен выглядеть этот график?

4.11. Дана стандартная модель парной регрессии

$$y_t = \alpha + \beta x_t + \varepsilon_t, \quad t = 1, \ldots, n.$$

а) Чему равна МНК-оценка коэффициента β при ограничении $\alpha = 0$?

б) Чему равна дисперсия оценки в а)? Покажите, что она меньше, чем $\sigma^2 / \sum_{t=1}^n (x_t - \bar{x})^2$ — дисперсия МНК-оценки β в регрессии без ограничения. Противоречит ли это теореме Гаусса–Маркова?

4.12. Рассмотрим регрессионную модель

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \varepsilon_t, \quad t = 1, \ldots, n,$$

в которой переменные представлены в виде отклонений от выборочных средних (т.е. $\bar{y} = 0$, $\bar{x}_1 = 0$, $\bar{x}_2 = 0$).

а) Покажите, что дисперсии и ковариация оценок метода наименьших квадратов $\widehat{\beta}_1$ и $\widehat{\beta}_2$ равны:

$$\mathrm{V}(\widehat{\beta}_1) = \frac{\sigma^2}{\sum_{t=1}^n x_{t1}^2 (1 - r_{12}^2)}, \quad \mathrm{V}(\widehat{\beta}_2) = \frac{\sigma^2}{\sum_{t=1}^n x_{t2}^2 (1 - r_{12}^2)},$$

$$\mathrm{Cov}(\widehat{\beta}_1, \widehat{\beta}_2) = \frac{-\sigma^2 r_{12}^2}{(1 - r_{12}^2)\sqrt{\sum_{t=1}^n x_{t1}^2 \sum_{t=1}^n x_{t2}^2}},$$

где

$$r_{12} = \frac{\sum_{t=1}^n x_{t1} x_{t2}}{\sqrt{\sum_{t=1}^n x_{t1}^2 \sum_{t=1}^n x_{t2}^2}}$$

— выборочный коэффициент корреляции между x_1 и x_2.

б) Чему равны дисперсии и ковариация в случае $r_{12} = 0$? Как это связано с проблемой мультиколлинеарности?

в) Постройте график отношения $\mathrm{V}(\widehat{\beta}_1)$ к значению $\mathrm{V}(\widehat{\beta}_1)$, полученному в б), в диапазоне $0 < r_{12} < 1$. Как этот график связан с проблемой мультиколлинеарности?

г) Что происходит с 95%-ными доверительными интервалами для β_1 и β_2 и ковариацией $\mathrm{Cov}(\widehat{\beta}_1, \widehat{\beta}_2)$ при возрастании r_{12} в диапазоне $0 < r_{12} < 1$?

4.13. Некоторая фирма занимается продажей молока. В таблице 4.3 представлены объемы ежемесячных продаж Q (тыс. литров) по различным ценам P (руб. за литр). Во время пятого, шестого и седьмого месяцев на одном из предприятий фирмы происходила забастовка.

Таблица 4.3

Месяцы	Q	P	Месяцы	Q	P	Месяцы	Q	P
1	98	10.0	6	87	14.6	11	121	14.2
2	100	11.0	7	94	14.9	12	123	14.4
3	103	12.5	8	113	13.0	13	126	15.0
4	105	12.5	9	116	13.0	14	128	16.1
5	80	14.6	10	118	13.8			

С помощью регрессий Q на P определите:

а) произошел ли сдвиг свободного члена (константы) во время забастовки по сравнению с обычным режимом;

б) произошел ли сдвиг как константы, так и коэффициента наклона при P.

4.14. В таблице 4.4 представлены совокупный объем внутренних инвестиций y и валовой внутренний продукт США (млрд. долл.) за период с 1939 по 1954 г.

Таблица 4.4. Инвестиции и ВВП США

Годы	y	x	Годы	y	x
1939	9.3	90.8	1947	34.0	232.8
1940	13.1	100.0	1948	45.9	259.1
1941	17.9	124.9	1949	35.3	258.0
1942	9.9	158.3	1950	53.8	286.2
1943	5.8	192.0	1951	59.2	330.2
1944	7.2	210.5	1952	52.1	347.2
1945	10.6	212.3	1953	53.3	366.1
1946	30.7	209.3	1954	52.7	366.3

Источник: *D.Salvatore. Statistics and Econometrics*, McGraw-Hill, 1982.

Напишите и оцените уравнения, позволяющие ответить на вопрос, изменилась ли зависимость инвестиций от валового внутреннего продукта во время войны (1942–1945 гг.) по сравнению с мирным временем.

4.15. В таблице 4.5 представлены квартальные данные об объемах продаж и доходах текстильных корпораций США с первого квартала 1974 г. по третий квартал 1979 г. Введите сезонные фиктивные переменные и с помощью регрессии дохода на объем продаж исследуйте наличие или отсутствие сезонных колебаний.

Таблица 4.5

Годы	Кварталы	Объем продаж	Доход
1974	I	242.0	13.5
	II	269.4	16.3
	III	272.1	15.5
	IV	277.0	13.4
1975	I	247.1	9.3
	II	265.8	12.4
	III	271.0	13.2
	IV	281.3	14.2
1976	I	284.2	14.8
	II	307.6	18.1
	III	301.6	16.0
	IV	309.8	15.6
1977	I	311.5	15.6
	II	338.6	19.7
	III	331.7	16.7
	IV	346.2	18.4
1978	I	340.2	16.0
	II	377.5	22.1
	III	376.9	20.4
	IV	401.8	22.6
1979	I	406.2	22.6
	II	436.4	26.8
	III	437.5	24.8

Источник: *D.Salvatore. Statistics and Econometrics*, McGraw-Hill, 1982.

4.16. Таблица 4.6 содержит данные об объеме импорта y (млрд. долл.), валовом национальном продукте x_1 (млрд. долл.) и индексе потребительских цен x_2 в США за период с 1964 по 1979 г.

Таблица 4.6

Годы	y	x_1	x_2	Годы	y	x_1	x_2
1964	28.4	635.7	92.9	1972	75.9	1171.1	125.3
1965	32.0	688.1	94.5	1973	94.4	1306.6	133.1
1966	37.7	753.0	97.2	1974	131.9	1412.9	147.7
1967	40.6	796.3	100.0	1975	126.9	1528.8	161.2
1968	47.7	868.5	104.2	1976	155.4	1702.2	170.5
1969	52.9	935.5	109.8	1977	185.8	1899.5	181.5
1970	58.5	982.4	116.3	1978	217.5	2127.6	195.4
1971	64.0	1063.4	121.3	1979	260.9	2368.5	217.4

Источник: *D.Salvatore. Statistics and Econometrics*, McGraw-Hill, 1982.

а) Вычислите выборочный коэффициент корреляции между x_1 и x_2.

б) Оцените регрессию y на константу и x_1.

в) Оцените регрессию y на константу и x_2.

г) Оцените регрессию y на константу, x_1 и x_2.

Как можно проинтерпретировать полученные результаты? Можно ли ограничиться только одной из регрессий б) или в)?

4.17. В таблице 4.7 представлены выпуск Q, трудозатраты L и капиталовложения K 15 фирм некоторой отрасли.

Таблица 4.7

Фирма	Q	L	K	Фирма	Q	L	K
1	2350	2334	1570	9	2550	2446	1880
2	2470	2425	1850	10	2450	2403	1790
3	2110	2230	1150	11	2290	2301	1480
4	2560	2463	1940	12	2160	2253	1240
5	2650	2565	2450	13	2400	2367	1660
6	2240	2278	1340	14	2490	2430	1850
7	2430	2380	1700	15	2590	2470	2000
8	2530	2437	1860				

а) Оцените по этим данным производственную функцию Кобба–Дугласа $Q = \alpha L^{\beta_1} K^{\beta_2}$, вычислите коэффициент детерминации, скорректированный коэффициент детерминации и выборочный коэффициент корреляции между $\ln L$ и $\ln K$.

б) Проведите регрессию $\ln Q$ только на $\ln K$.

Как можно проинтерпретировать полученные результаты?

4.18. Можно ли преодолеть проблему мультиколлинеарности, возникающую в упражнении 4.17, если известно, что производственная функция обладает постоянной отдачей на масштаб ($\beta_1 + \beta_2 = 1$)?

4.19. Пусть $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$ — стандартная модель множественной регрессии и $\widehat{\boldsymbol{\beta}}$ — МНК-оценка вектора коэффициентов $\boldsymbol{\beta}$.

а) Покажите, каким образом можно использовать $\widehat{\boldsymbol{\beta}}$ для получения более эффективной оценки параметров $\boldsymbol{\beta}$, если известно, что $\boldsymbol{\beta}$ удовлетворяет линейному ограничению $\boldsymbol{R\beta} = \boldsymbol{r}$.

б) Для модели $y_t = \alpha + \beta x_{t1} + \gamma x_{t2} + \varepsilon_t$ по $n = 100$ наблюдениям получены следующие данные (матрица сумм произведений соответствующих переменных):

	$y - \overline{y}$	$x_1 - \overline{x}_1$	$x_2 - \overline{x}_2$
$y - \overline{y}$	2000	100	90
$x_1 - \overline{x}_1$	100	10	5
$x_2 - \overline{x}_2$	90	5	5

Проверьте гипотезу H_0: $5\beta = \gamma$ против альтернативы H_1: $5\beta \neq \gamma$

4.20. С помощью обычного метода наименьших квадратов получены две спецификации модели: $\boldsymbol{y} = \widehat{\alpha} + \widehat{\beta}\boldsymbol{x} + \boldsymbol{e}$ и $\boldsymbol{y} = \widehat{\alpha}^* + \widehat{\beta}^*\boldsymbol{x} + \widehat{\gamma}^*\boldsymbol{z} + \boldsymbol{u}$, где $\boldsymbol{e}$, $\boldsymbol{u}$ — остатки соответствующих регрессий. Объясните, при каких обстоятельствах выполнены следующие условия:

а) $\widehat{\beta} = \widehat{\beta}^*$;

б) $\sum u_t^2 \leqslant \sum e_t^2$;

в) оценка $\widehat{\beta}$ статистически значима на 5%-ном уровне, а оценка $\widehat{\beta}^*$ незначима;

г) оценка $\widehat{\beta}^*$ статистически значима на 5%-ном уровне, а оценка $\widehat{\beta}$ незначима.

4.21. Дана стандартная модель множественной регрессии $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$.

а) Выразите матрицу ковариаций МНК-оценки вектора $\boldsymbol{\beta}$ в терминах собственных значений и собственных векторов матрицы $\boldsymbol{X}'\boldsymbol{X}$.

б) Объясните, как соотносится результат а) с проблемой мультиколлинеарности.

4.22. Дана модель множественной регрессии $\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$, где $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$ — векторы размерности k_1, k_2, соответственно. Предположим, что у вектора $\boldsymbol{\beta}_1$ есть несмещенная оценка $\widehat{\boldsymbol{\beta}}_1$, некоррелированная с $\boldsymbol{\varepsilon}$, с известной ковариационной матрицей V_1.

а) Вычислите ковариационную матрицу оценки вектора $\boldsymbol{\beta}_2$, получаемой регрессией $\boldsymbol{y} - \boldsymbol{X}_1\widehat{\boldsymbol{\beta}}_1$ на $\boldsymbol{X}_2$.

б) Сравните ковариационную матрицу, полученную в а), с ковариационной матрицей МНК-оценки вектора $\boldsymbol{\beta}_2$ в исходной модели.

в) Можете ли вы предложить более эффективную оценку вектора $\boldsymbol{\beta}_2$, чем те, что получены в а) и б)?

г) Как полученные результаты связаны с проблемой мультиколлинеарности?

4.23. Для проверки гипотезы о том, что удельный выпуск Q/L в металлургической промышленности зависит от уровня зарплаты W, на основе межстрановых наблюдений была получена регрессия

$$\ln\frac{Q}{L} = 0.374 + \underset{(0.049)}{0.805}W + e, \qquad R^2 = 0.929$$

(в скобках указана стандартная ошибка).

а) Проверьте гипотезу.

б) Было высказано предположение, что приведенное выше уравнение содержит ошибки спецификации, поскольку оно не учитывает разницу в эффективности между странами, которая оказывает влияние на удельный выпуск и положительно коррелирована с зарплатой. Как это предположение повлияет на ваш вывод?

4.24. Модель

$$y_t = \beta_1 + \beta_2 d_{t2} + \beta_3 d_{t3} + \beta_4 d_{t4} + \beta_5 x_t + \varepsilon_t$$

оценивается с помощью обычного метода наименьших квадратов на основе ежеквартальных наблюдений, где d_{ti}, $i = 2, 3, 4$ — фиктивные переменные для соответствующих кварталов, т. е.

$d_{t2} = 1$, если t — второй квартал, $= 0$ в остальных случаях,
$d_{t3} = 1$, если t — третий квартал, $= 0$ в остальных случаях,
$d_{t4} = 1$, если t — четвертый квартал, $= 0$ в остальных случаях.

а) Почему в модель не включена переменная d_{t1}?

б) Покажите, что оценка $\widehat{\beta}_5$ совпадает с МНК-оценкой коэффициента β в регрессии $y_t^* = \alpha + \beta x_t^* + u_t$, где y_t^* — остатки регрессии y_t на d_{t2}, d_{t3}, d_{t4} и константу, а x_t^* — остатки регрессии x_t на d_{t2}, d_{t3}, d_{t4} и константу.

4.25. В программе исследований k разных удобрений, предназначенных для повышения урожайности лимонных бананов, использованы в опытах на $n = n_1 + \cdots + n_k$ опытных участках. Удобрение номер s ($s = 1, \ldots, k$) использовалось на n_s опытных участках. Для изучения влияния удобрений использовалась регрессионная модель

$$y_t = \beta_1 d_{1t} + \cdots + \beta_k d_{kt}, \quad t = 1, \ldots, n.$$

Здесь y — урожайность, d_s — фиктивная переменная, равная 1 для участка номер s и 0 в других случаях. Известны выборочные средние $\overline{y}_s$ и стандартные отклонения s_s для $s = 1, \ldots, k$.

$$\overline{y}_s = \frac{1}{n_s} \sum_{d_{st}=1} y_t, \quad s_s = \frac{1}{n_s - 1} \sum_{d_{st}=1} (y_t - \overline{y}_s).$$

Выразить через известные величины F-статистику для тестирования нулевой гипотезы о равном влиянии всех удобрений ($\beta_1 = \cdots = \beta_k$).

4.26. Вы в скором времени планируете поступить на должность политического аналитика на некую телевизионную станцию. Все телевизионные компании уделяют большое внимание освещению выборов в Конгресс (США), а ваших знаний по этому вопросу недостаточно, чтобы получить эту работу. Поэтому вы решили рассмотреть несколько регрессионных моделей, чтобы подкрепить свое мнение относительно выборов в Конгресс 1996 года.

Вам понадобится три модели. Каждая из них пытается объяснить различия в проценте голосов, отданных Республиканской партии, среди всех 50 штатов. Т. е. все модели включают 50 наблюдений, каждое соответствует одному штату. У вас также есть четыре типа объясняющих переменных:

1) уровень безработицы для каждого штата;

2) региональные фиктивные переменные, показывающие, что штат находится на северо-востоке, юге, среднем западе или на западе;

3) фиктивная переменная, показывающая, что Альберт Гор (вице-президент, демократ) появлялся в этом штате, агитируя за кандидатов в Конгресс;

4) перекрестные произведения региональных фиктивных переменных и фиктивной переменной Гора.

Три ваших модели отличаются только набором объясняющих переменных:

Модель I содержит переменные 1) и 2).
Модель II содержит переменные 1), 2) и 3).
Модель III содержит переменные 1) и 4).

а) Запишите уравнение регрессии для каждой из моделей. Это можно сделать разными способами, используйте формулировку, которая вам больше нравится.

б) Укажите, как бы вы тестировали с помощью этих моделей следующие гипотезы (если вы хотите предложить F-тест, укажите регрессию с ограничениями и без ограничений):

1) появление Гора не оказывает влияния на процент голосов, отданных республиканцам;

2) вся страна голосует одинаково, без различий по региональному признаку;

3) северо-восток и средний запад («пояс холода») голосуют одинаково;

4) «пояс холода» голосует одинаково, «солнечный пояс» (юг и запад) голосует одинаково, но между этими поясами может быть разница;

5) появление Гора приводит к одному и тому же эффекту для всех регионов.

4.27. На основе квартальных данных с 1971 по 1976 г. с помощью метода наименьших квадратов получено следующее уравнение:

$$y_t = \underset{(2.14)}{1.12} - \underset{(0.0034)}{0.0098} x_{t1} - \underset{(3.42)}{5.62} x_{t2} + \underset{(0.009)}{0.044} x_{t3},$$

в скобках указаны стандартные ошибки, RSS = 101.32, ESS = 21.43.

а) Проверьте значимость каждого из коэффициентов.

б) Найдите коэффициент детерминации.

в) Протестируйте значимость регрессии в целом.

г) Когда в уравнение были добавлены три фиктивные переменные, соответствующие трем первым кварталам года, величина RSS выросла до 118.20. Проверьте гипотезу о наличии сезонности, сформулировав необходимые предположения о виде этой сезонности.

д) Для той же исходной модели были раздельно проведены две регрессии на основе данных: 1-й квартал 1971 г.–1-й квартал 1975 г. и 2-й квартал 1975 г.–4-й квартал 1976 г., соответственно, и получены следующие значения сумм квадратов остатков: ESS1 = 11.25, ESS2 = 2.32. Проверьте гипотезу о том, что между первым и вторым кварталами 1975 г. произошло структурное изменение.

4.28. Ниже приведены результаты регрессии W — зарплаты менеджера фирмы, на объем ее продаж S и доход P (число наблюдений $n = 102$, в скобках даны стандартные ошибки):

$$W = \underset{(0.83)}{0.50}\,S + \underset{(0.83)}{0.4}\,P, \quad e'e = 250, \quad X'X = \begin{bmatrix} 10 & 8 \\ 8 & 10 \end{bmatrix}$$

(для удобства все переменные представлены в отклонениях от средних). Ввиду большой зависимости между объемами продаж и доходом возникает проблема мультиколлинеарности, что не позволяет точно оценить соответствующие параметры. Для решения этой проблемы было предложено действовать следующим образом.

1) Провести регрессию P на S и получить остатки r.

2) Провести регрессию W на S и r.

Обозначим результат последней регрессии $W = c_1 S + c_2 r$.

а) Вычислите c_1, c_2.

б) Дайте оценку предложенному методу как способу борьбы с мультиколлинеарностью.

в) Дайте оценку предложенному методу как способу получения более точных оценок исходных параметров.

4.29. Файл `usa_import.xls` содержит данные об объеме импорта Imp (млрд. долл.), валовом национальном продукте GDP (млрд. долл.) и индексе потребительских цен CPI в США за период с 1964 по 1979 г.

а) Вычислите выборочный коэффициент корреляции между GDP и CPI.

б) Оцените регрессию Imp на константу и GDP.

в) Оцените регрессию Imp на константу и CPI.

г) Оцените регрессию Imp на константу, GDP и CPI.

Как можно интерпретировать полученные результаты? Можно ли ограничиться только одной из регрессий б) или в)?

4.30. Построение модели цены колготок в московских оптовых торговых фирмах, осень 1997 г. Данные содержатся в файле `tights.xls` (всего 74 наблюдения). Описание переменных содержится в таблице 4.8.

Таблица 4.8

Переменная	Описание
N	номер по порядку
$Price$	цена колготок в рублях 1997 г.
DEN	плотность в DEN
$polyamid$	содержание полиамида
$lykra$	содержание лайкры
$cotton$	% хлопка
$wool$	% шерсти
$firm$	фирма-производитель: 0 — Levante, 1 — Golden Lady

а) Постройте уравнения зависимости цены колготок от их плотности, состава и производителя. Подберите наиболее подходящую форму модели. Какие проблемы с данными вы при этом встретили?

б) С помощью построенной модели ответьте на вопрос: верно ли, что цены колготок двух фирм-производителей различаются статистически достоверно? Какая из фирм устанавливает более высокие цены?

Глава 5

Некоторые обобщения множественной регрессии

В этой главе классическая регрессионная схема обобщается в двух направлениях. Первое связано с отказом от предположения, что независимые переменные являются неслучайными величинами. Оказывается, что при выполнении некоторых естественных условий МНК-оценка вектора неизвестных параметров сохраняет основные свойства МНК-оценки в стандартной модели. Главным условием, гарантирующим наличие этих свойств, является некоррелированность (независимость) матрицы регрессоров X и вектора ошибок ε.

Второе направление — это изучение линейной модели, в которой ковариационная матрица Ω вектора ошибок ε не обязательно является скалярной (т. е. имеет вид $\sigma^2 I_n$), а может быть произвольной симметричной положительно определенной матрицей. С помощью линейного преобразования исходную систему можно свести к обычному регрессионному уравнению и построить для него МНК-оценку вектора коэффициентов. Эта оценка зависит от матрицы ковариаций ошибки ε, а способ оценивания носит название *обобщенный метод наименьших квадратов* (ОМНК). (Соответствующий термин на английском языке — *Generalized Least Squares*, GLS.) Для ОМНК-оценки устанавливается аналог теоре-

мы Гаусса–Маркова, а именно, доказывается, что в классе всех несмещенных линейных оценок она обладает наименьшей матрицей ковариаций. Обобщенный метод наименьших квадратов позволяет с единых позиций изучать некоторые важные классы регрессионных моделей: так называемые модели с *гетероскедастичностью*, когда матрица Ω является диагональной, но имеет разные элементы на главной диагонали, и модели, в которых наблюдения имеют смысл временных рядов, а ошибки коррелированы по времени. Эти вопросы изучаются в главе 6. Следует подчеркнуть, что практическое использование обобщенного метода наименьших квадратов усложнено тем обстоятельством, что для построения ОМНК-оценки требуется знать матрицу Ω, которая реально почти всегда неизвестна. В связи с этим возникает проблема построения так называемого *доступного обобщенного метода наименьших квадратов* (*Feasible Generalized Least Squares*), который обсуждается в разделе 5.3.

5.1. Стохастические регрессоры

В предыдущих разделах предполагалось, что независимые переменные (матрица X) являются неслучайными. Ясно, что такое условие выполнено не всегда, например, во многих ситуациях при измерении независимых переменных могут возникать случайные ошибки. Кроме того, при анализе временных рядов значение исследуемой величины в момент t может зависеть от ее значений в предыдущие моменты времени, т. е. в некоторых уравнениях эти значения выступают в качестве независимых, а в других — в качестве зависимых переменных (модели с лагированными переменными). Поэтому возникает необходимость рассматривать модели со стохастическими регрессорами.

Предлагаемый здесь подход позволяет сохранить, по существу, все основные свойства МНК-оценок в классической регрессии. Условия, накладываемые на систему со стохастическими регрессорами, почти дословно повторяют ограничения стандартной модели, но только теперь их следует понимать, говоря не совсем стро-

го, в условном смысле (при фиксированной матрице $\boldsymbol{X}$). Мы предполагаем, что читателю известны простейшие свойства условных распределений и условных математических ожиданий, используемые ниже (см. приложение МС, п. 2).

Пусть, как и раньше,

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

где $\boldsymbol{y}$ — $n \times 1$ вектор зависимых переменных, $\boldsymbol{X}$ — $n \times k$ матрица независимых переменных, $\boldsymbol{\varepsilon}$ — $n \times 1$ вектор случайных ошибок. Но теперь будем считать, что элементы матрицы $\boldsymbol{X}$ также являются случайными величинами. Предположим, что выполнены следующие условия:

1) $\mathrm{E}(\boldsymbol{\varepsilon} \mid \boldsymbol{X}) = \boldsymbol{0}$,
2) $\mathrm{V}(\boldsymbol{\varepsilon} \mid \boldsymbol{X}) = \sigma^2 \boldsymbol{I}$,
3) при любой реализации (т. е. с вероятностью 1) матрица $\boldsymbol{X}$ имеет ранг k.

Здесь $\mathrm{E}(\boldsymbol{\varepsilon} \mid \boldsymbol{X})$ — условное математическое ожидание случайного вектора $\boldsymbol{\varepsilon}$ при фиксированной матрице $\boldsymbol{X}$, $\mathrm{V}(\boldsymbol{\varepsilon} \mid \boldsymbol{X}) = \mathrm{E}(\boldsymbol{\varepsilon\varepsilon}' \mid \boldsymbol{X})$ — его условная ковариационная матрица. Заметим, что условия 1), 2) эквивалентны условиям:

1′) $\mathrm{E}(\boldsymbol{y} \mid \boldsymbol{X}) = \boldsymbol{X\beta}$,
2′) $\mathrm{V}(\boldsymbol{y} \mid \boldsymbol{X}) = \sigma^2 \boldsymbol{I}$.

Пусть $\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} = \widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ — МНК-оценка вектора $\boldsymbol{\beta}$ (которая существует при любой реализации $\boldsymbol{X}$ в силу условия 3)), $\boldsymbol{e} = \boldsymbol{My}$ — вектор остатков, $\widehat{\sigma}^2 = \boldsymbol{e}'\boldsymbol{e}/(n-k)$ — оценка дисперсии, $\widehat{\mathrm{V}}(\widehat{\boldsymbol{\beta}}) = \widehat{\sigma}^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$ — оценка ковариационной матрицы $\widehat{\boldsymbol{\beta}}$. Тогда (ср. (3.7), (3.8))

$$\begin{aligned}
\mathrm{E}(\widehat{\boldsymbol{\beta}} \mid \boldsymbol{X}) &= \mathrm{E}\big(\boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon} \mid \boldsymbol{X}\big) = \boldsymbol{\beta} + \mathrm{E}\big((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon} \mid \boldsymbol{X}\big) \\
&= \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathrm{E}(\boldsymbol{\varepsilon} \mid \boldsymbol{X}) = \boldsymbol{\beta}; \\
\mathrm{V}(\widehat{\boldsymbol{\beta}} \mid \boldsymbol{X}) &= \mathrm{V}\big((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \mid \boldsymbol{X}\big) \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\mathrm{V}(\boldsymbol{y} \mid \boldsymbol{X})\,\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1};
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}(\boldsymbol{e} \mid \boldsymbol{X}) &= \mathrm{E}(\boldsymbol{My} \mid \boldsymbol{X}) = \boldsymbol{M}\mathrm{E}(\boldsymbol{y} \mid \boldsymbol{X}) = \boldsymbol{MX\beta} = \boldsymbol{0}; \\
\mathrm{V}(\boldsymbol{e} \mid \boldsymbol{X}) &= \mathrm{V}(\boldsymbol{My} \mid \boldsymbol{X}) = \boldsymbol{M}\mathrm{V}(\boldsymbol{y} \mid \boldsymbol{X})\boldsymbol{M}' = \sigma^2\boldsymbol{M}.
\end{aligned}$$

Отсюда следует, что

$$\begin{aligned}
\mathrm{E}(\boldsymbol{e}'\boldsymbol{e} \mid \boldsymbol{X}) &= \sigma^2\,\mathrm{tr}(\boldsymbol{M}); \\
\mathrm{E}(\widehat{\sigma}^2 \mid \boldsymbol{X}) &= \sigma^2; \\
\mathrm{E}(\widehat{\mathrm{V}}(\widehat{\boldsymbol{\beta}}) \mid \boldsymbol{X}) &= \mathrm{E}(\widehat{\sigma}^2(\boldsymbol{X}'\boldsymbol{X})^{-1} \mid \boldsymbol{X}) = \mathrm{E}(\widehat{\sigma}^2 \mid \boldsymbol{X})(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}.
\end{aligned}$$

При выводе этих равенств мы постоянно используем тот факт, что сомножитель, функционально зависящий от условия (матрицы $\boldsymbol{X}$), например, $\boldsymbol{M}$, можно выносить из-под знака условного математического ожидания. Таким образом, оценки $\widehat{\boldsymbol{\beta}}$, $\widehat{\sigma}^2$ и $\widehat{\mathrm{V}}(\widehat{\boldsymbol{\beta}})$ являются условно (относительно $\boldsymbol{X}$) несмещенными. Используя еще одно свойство условного математического ожидания — правило повторного ожидания, нетрудно установить безусловную несмещенность этих оценок:

$$\begin{aligned}
\mathrm{E}(\widehat{\boldsymbol{\beta}}) &= \mathrm{E}(\mathrm{E}(\widehat{\boldsymbol{\beta}} \mid \boldsymbol{X})) = \mathrm{E}(\boldsymbol{\beta}) = \boldsymbol{\beta}; \\
\mathrm{E}\widehat{\sigma}^2 &= \mathrm{E}(\mathrm{E}(\widehat{\sigma}^2 \mid \boldsymbol{X})) = \sigma^2; \\
\mathrm{E}(\widehat{\mathrm{V}}(\widehat{\boldsymbol{\beta}})) &= \mathrm{E}(\mathrm{E}(\widehat{\mathrm{V}}(\widehat{\boldsymbol{\beta}}) \mid \boldsymbol{X})) = \sigma^2\mathrm{E}\big((\boldsymbol{X}'\boldsymbol{X})^{-1}\big) = \mathrm{V}(\widehat{\boldsymbol{\beta}}).
\end{aligned} \tag{5.1}$$

Нетрудно также доказать соответствующий вариант теоремы Гаусса-Маркова, а именно, что среди всех линейных условно несмещенных оценок вектора $\boldsymbol{\beta}$ его МНК-оценка обладает наименьшей условной ковариационной матрицей. Итак, при выполнении условий 1), 2), 3) МНК-оценка в модели со стохастическими регрессорами обладает свойствами, аналогичными свойствам МНК-оценки в классической модели.

Следует понимать, что условия 1), 2) касаются совместного распределения $\boldsymbol{X}$ и $\boldsymbol{\varepsilon}$. Из 1), в частности, вытекает некоррелированность $\boldsymbol{X}$ и $\boldsymbol{\varepsilon}$. Действительно, поскольку $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathrm{E}(\mathrm{E}(\boldsymbol{\varepsilon} \mid \boldsymbol{X})) = \boldsymbol{0}$, то $\mathrm{Cov}(x_{ij}, \varepsilon_m) = \mathrm{E}(x_{ij}\varepsilon_m) = \mathrm{E}(\mathrm{E}(x_{ij}\varepsilon_m \mid \boldsymbol{X})) = \mathrm{E}(x_{ij}\mathrm{E}(\varepsilon_m \mid \boldsymbol{X})) = 0$. Обратное, вообще говоря, неверно. Однако

если $\boldsymbol{X}$ и ε независимы и $\mathrm{E}(\varepsilon) = \mathbf{0}$, $\mathrm{E}(\varepsilon\varepsilon') = \sigma^2\boldsymbol{I}$, то выполнены 1) и 2).

Остановимся, наконец, на проблеме состоятельности МНК-оценки в этой модели. Напомним, что оценка параметра называется состоятельной, если ее предел по вероятности при увеличении числа наблюдений стремится к истинному значению параметра. В данном случае требуется сформулировать условия, при выполнении которых $p\lim_{n\to\infty}\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}$. Итак, пусть $n \to \infty$ (напомним, что n представляет количество наблюдений и, следовательно, при увеличении n возрастают размерности векторов $\boldsymbol{y}$ и ε, увеличивается количество строк матрицы $\boldsymbol{X}$, в то время как число ее столбцов и размерность вектора $\widehat{\boldsymbol{\beta}}$ остаются равными k). Сделаем элементарное преобразование:

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\varepsilon \\ &= \boldsymbol{\beta} + \left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\frac{1}{n}\boldsymbol{X}'\varepsilon\right) \qquad (5.2)\end{aligned}$$

и предположим, что выполнены следующие условия:

4) существует $p\lim_{n\to\infty}(1/n)\,\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{A}$, причем матрица $\boldsymbol{A}$ положительно определена (и, следовательно, существует $\boldsymbol{A}^{-1}$);

5) $p\lim_{n\to\infty}(1/n)\,\boldsymbol{X}'\varepsilon = \mathbf{0}$.

Тогда из теоремы Слуцкого (см. приложение МС, п. 5) и (5.2) следует, что $p\lim_{n\to\infty}\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}$, т. е. оценка $\widehat{\boldsymbol{\beta}}$ состоятельна.

В некоторых случаях условия 4), 5) достаточно легко проверяются. Пусть, например, строки матрицы $\boldsymbol{X}$ независимы и одинаково распределены (как случайные k-мерные векторы), вектор ошибок ε состоит из независимых и одинаково распределенных компонент, $\mathrm{E}\varepsilon = \mathbf{0}$, $\boldsymbol{X}$ и ε независимы. Иными словами, значения объясняющих переменных в каждом наблюдении выбираются из одной и той же генеральной совокупности, причем наблюдения между собой независимы и не зависят от случайных ошибок. Обозначим $a_{ij} = \mathrm{E}(x_{ti}x_{tj})$, $i, j = 1, \ldots, k$ (эти числа не зависят от t, поскольку строки матрицы X одинаково распределены), и пусть $\boldsymbol{A} = (a_{ij})$. Тогда по закону больших чисел $p\lim_{n\to\infty}(1/n)\boldsymbol{X}'\boldsymbol{X} = \boldsymbol{A}$, и если распределение каждой

строки не сосредоточено на какой-либо гиперплоскости пространства R^k, то матрица $\boldsymbol{A}$ положительно определена. Аналогично, $p\lim_{n\to\infty}\Big((1/n)\boldsymbol{X}'\varepsilon\Big)_i = \mathrm{E}(x_{ti}\varepsilon_t) = 0$ в силу независимости $\boldsymbol{X}$ и ε.

Подчеркнем, что из представления (5.2) следует, что при наличии корреляции между $\boldsymbol{X}$ и ε МНК-оценка будет, вообще говоря, смещенной и несостоятельной.

Замечание. В рамках ограничений 1), 2), 3) для состоятельности МНК-оценки достаточно требовать выполнения условия 4) (с точностью до некоторых математических тонкостей), так как в силу (5.1)

$$\mathrm{V}(\widehat{\boldsymbol{\beta}}) = \sigma^2\mathrm{E}((\boldsymbol{X}'\boldsymbol{X})^{-1}) = \frac{1}{n}\sigma^2\mathrm{E}\left(\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right) \to \mathbf{0} \quad \text{при } n \to \infty$$

и, следовательно, $p\lim_{n\to\infty}\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}$. Мы, однако, привели условия 4) и 5) ввиду их большей универсальности: как легко понять, для доказательства состоятельности МНК-оценки при выполнении 4), 5) требуются только равенство $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \varepsilon$ и вид МНК-оценки, а условия 1), 2), 3) явно не используются.

Можно сделать следующие выводы:

1) если в регрессионной модели объясняющие переменные случайны и выполнены условия 1)–3) (в частности, регрессоры и ошибки должны быть некоррелированы), то МНК-оценка и связанные с ней статистики (оценка дисперсии и ковариационной матрицы) являются как условно (при фиксированной матрице $\boldsymbol{X}$), так и безусловно несмещенными;

2) имеет место условный вариант теоремы Гаусса–Маркова;

3) при выполнении условий 4), 5) МНК-оценка состоятельна, в частности, это справедливо, если в каждом наблюдении значения объясняющих переменных выбираются из одной и той же генеральной совокупности, а ошибки независимы, одинаково распределены и не зависят от регрессоров;

4) если регрессоры и ошибки коррелированы, то МНК-оценка будет в общем случае смещенной и несостоятельной.

5.2. Обобщенный метод наименьших квадратов

Одно из предположений классической регрессионной модели состоит в том, что случайные ошибки некоррелированы между собой и имеют постоянную дисперсию. В тех случаях, когда наблюдаемые объекты достаточно однородны, не сильно отличаются друг от друга, такое допущение оправдано. Однако во многих ситуациях такое предположение нереалистично. Например, если исследуется зависимость расходов на питание в семье от ее общего дохода, то естественно ожидать, что разброс в данных будет выше для семей с более высоким доходом. Это означает, что дисперсии зависимых величин (а следовательно, и случайных ошибок) не постоянны. Это явление в эконометрике называется *гетероскедастичностью* (в отличие от *гомоскедастичности* — равенства дисперсий). Кроме того, при анализе временных рядов в довольно редких случаях можно считать, что наблюдения некоррелированы во времени. Как правило, значение исследуемой величины в текущий момент времени статистически зависит от ее значений в прошлом, что означает наличие корреляции между ошибками. Поэтому естественно изучать модели регрессии без предположения, что $\mathrm{V}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}$.

В данном разделе мы будем рассматривать так называемую обобщенную регрессионную модель

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{5.3}$$

где $\boldsymbol{y}$ — $n \times 1$ вектор зависимой переменной, $\boldsymbol{X}$ — $n \times k$ матрица независимых переменных, $\boldsymbol{\beta}$ — $k \times 1$ вектор неизвестных параметров, $\boldsymbol{\varepsilon}$ — $n \times 1$ вектор случайных ошибок, причем:

1) матрица $\boldsymbol{X}$ неслучайна и имеет полный ранг;

2) $\mathrm{E}\boldsymbol{\varepsilon} = \boldsymbol{0}$;

3) $\mathrm{V}(\boldsymbol{\varepsilon}) = \boldsymbol{\Omega}$, и матрица $\boldsymbol{\Omega}$ положительно определена.

Иными словами, обобщенная модель отличается от классической только условием 3).

1. *Обычный метод наименьших квадратов.* К системе (5.3) можно применить обычный метод наименьших квадратов. Пусть $\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} = \widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ — МНК-оценка вектора $\boldsymbol{\beta}$, $\boldsymbol{e} = \boldsymbol{M}\boldsymbol{y} = (\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{y}$ — вектор остатков. Тогда нетрудно проверить, что

$$\mathrm{E}(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathrm{E}(\boldsymbol{\varepsilon}) = \boldsymbol{\beta},$$

т. е. $\widehat{\boldsymbol{\beta}}$ является несмещенной оценкой,

$$\mathrm{V}(\widehat{\boldsymbol{\beta}}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathrm{V}(\boldsymbol{y})\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1},$$
$$\mathrm{E}(\boldsymbol{e}) = \boldsymbol{M}\mathrm{E}(\boldsymbol{y}) = \boldsymbol{0}, \quad \mathrm{V}(\boldsymbol{e}) = \boldsymbol{M}\boldsymbol{\Omega}\boldsymbol{M}' = \boldsymbol{M}\boldsymbol{\Omega}\boldsymbol{M}$$

(напомним, что $\boldsymbol{M}^2 = \boldsymbol{M}$, $\boldsymbol{M}' = \boldsymbol{M}$ и $\boldsymbol{M}\boldsymbol{X} = \boldsymbol{0}$). Поэтому для математического ожидания суммы квадратов остатков получаем следующее выражение (ср. (3.17)):

$$\mathrm{E}(\boldsymbol{e}'\boldsymbol{e}) = \mathrm{tr}\,(\mathrm{V}(\boldsymbol{e})) = \mathrm{tr}\,(\boldsymbol{M}\boldsymbol{\Omega}\boldsymbol{M}) = \mathrm{tr}\,(\boldsymbol{M}^2\boldsymbol{\Omega}) = \mathrm{tr}\,(\boldsymbol{M}\boldsymbol{\Omega}).$$

Следовательно,

$$\mathrm{E}(\widehat{\sigma}^2) = \mathrm{E}\,\frac{\boldsymbol{e}'\boldsymbol{e}}{n-k} = \frac{\mathrm{tr}\,(\boldsymbol{M}\boldsymbol{\Omega})}{n-k}.$$

Таким образом, если в качестве оценки матрицы ковариаций $\mathrm{V}(\widehat{\boldsymbol{\beta}})$ взять стандартную оценку $\widehat{V}(\widehat{\boldsymbol{\beta}}) = \widehat{\sigma}^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$, то $\mathrm{E}(\widehat{V}(\widehat{\boldsymbol{\beta}})) = (1/(n-k))\,\mathrm{tr}\,(\boldsymbol{M}\boldsymbol{\Omega})(\boldsymbol{X}'\boldsymbol{X})^{-1}$, что, в общем случае, не совпадает с $\mathrm{V}(\widehat{\boldsymbol{\beta}})$. Значит, оценка матрицы ковариаций вектора $\widehat{\boldsymbol{\beta}}$, получаемая при использовании обычного метода наименьших квадратов, является смещенной.

Заметим, что с очевидными изменениями, подобно тому, как это сделано в п. 5.1, можно получить аналогичные результаты для стохастических регрессоров $\boldsymbol{X}$. В частности,

$$\begin{aligned}\mathrm{V}(\widehat{\boldsymbol{\beta}}) &= \mathrm{E}((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}) \\ &= \frac{1}{n}\mathrm{E}\left(\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{\Omega}\boldsymbol{X}\right)\left(\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\right),\end{aligned}$$

откуда следует, что если при $n \to \infty$ матрицы $(1/n)\boldsymbol{X}'\boldsymbol{X}$ и $(1/n)\boldsymbol{X}'\boldsymbol{\Omega}\boldsymbol{X}$ стремятся к положительно определенным матрицам,

то $\mathrm{V}(\widehat{\beta}) \to 0$ и, значит, оценка $\widehat{\beta}$ будет состоятельной. Однако в отличие от классической модели, она не будет оптимальной в смысле теоремы Гаусса–Маркова. Для получения эффективной оценки надо воспользоваться так называемым *обобщенным методом наименьших квадратов* (ОМНК).

2. *Обобщенный метод наименьших квадратов.* Ответ на вопрос об эффективной линейной несмещенной оценке вектора β для модели (5.3) дает следующая теорема.

Теорема Айткена. *В классе линейных несмещенных оценок вектора β для обобщенной регрессионной модели оценка*

$$\widehat{\beta}^* = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \tag{5.4}$$

имеет наименьшую матрицу ковариаций.

Доказательство. Нетрудно проверить, что оценка (5.4) действительно несмещена. Далее, в силу условия 3) матрица Ω^{-1} положительно определена и симметрична, поэтому существует такая невырожденная $n \times n$ матрица P, что

$$P'P = \Omega^{-1}. \tag{5.5}$$

В самом деле, так как Ω^{-1} симметрична, то существует ортогональная матрица S, такая что $\Omega^{-1} = S'\Lambda S$, где Λ — диагональная матрица, на главной диагонали которой стоят собственные числа λ_i, $i = 1, \ldots, n$, матрицы Ω^{-1}. В силу положительной определенности Ω^{-1} все они положительны, поэтому можно определить диагональную матрицу $\Lambda^{1/2}$, на главной диагонали которой стоят числа $\lambda_i^{1/2}$, $i = 1, \ldots, n$. Теперь достаточно взять $P = \Lambda^{1/2}S$. Заметим, что представление (5.5) не единственно, но для наших рассуждений это несущественно. Умножим равенство (5.3) слева на P и обозначим $y^* = Py$, $X^* = PX$, $\varepsilon^* = P\varepsilon$. Таким образом,

$$y^* = X^*\beta + \varepsilon^*, \tag{5.6}$$

причем $\mathrm{E}(\varepsilon^*) = 0$ и $\mathrm{V}(\varepsilon^*) = P\Omega P' = I$, поскольку в силу (5.5) $P' = \Omega^{-1}P^{-1}$. Кроме того, $\mathrm{rank}(X^*) = k$, так как P невырождена. Это означает, что для модели (5.6) выполнены условия

теоремы Гаусса–Маркова и, следовательно, оптимальной в классе несмещенных и линейных по y^* оценок вектора β является оценка (см. (3.4))

$$\begin{aligned}\widehat{\beta}^* = (X^{*\prime}X^*)^{-1}X^{*\prime}y^* &= (X'P'PX)^{-1}X'P'Py \\ &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y,\end{aligned} \tag{5.7}$$

что совпадает с (5.4). Остается заметить, что поскольку матрица P невырождена, то класс оценок, линейных по y^*, совпадает с классом оценок, линейных по y. Доказательство закончено.

Так как $\mathrm{V}(y) = \Omega$, то из (5.7) непосредственно следует, что

$$\mathrm{V}(\widehat{\beta}^*) = (X'\Omega^{-1}X)^{-1}. \tag{5.8}$$

Оценку $\widehat{\beta}^*$ часто будем обозначать $\widehat{\beta}_{\mathrm{GLS}}$ (GLS, *Generalized Least Squares*). Нетрудно проверить, что если $\Omega = \sigma^2 I$, т. е. модель является классической, то $\widehat{\beta}_{\mathrm{GLS}} = \widehat{\beta}_{\mathrm{OLS}}$, как и следовало ожидать. Использование термина «обобщенный метод наименьших квадратов» объясняется следующим соображением. Как показано при доказательстве теоремы Айткена, оценка $\widehat{\beta}_{\mathrm{GLS}}$ получается минимизацией по b сумм квадратов отклонений $f(b) = e^{*\prime}e^* = (y^* - X^*b)'(y^* - X^*b)$ для системы (5.6). Но $f(b) = (y - Xb)'P'P(y - Xb) = (y - Xb)'\Omega^{-1}(y - Xb) = e'\Omega^{-1}e$, т. е. для построения оптимальной оценки в модели (5.3) надо минимизировать «обобщенную» сумму квадратов отклонений $e'\Omega^{-1}e$.

Проверять гипотезы о наличии линейных ограничений можно как непосредственно, используя (5.8), так и с помощью вспомогательной регрессии (5.6). Например, если в предположении нормальности ошибок ε требуется проверить гипотезу H_0: $R\beta = r$ против альтернативной, то можно воспользоваться тем фактом, что статистика

$$\begin{aligned}&(R\widehat{\beta}^* - r)'(\mathrm{V}(R\widehat{\beta}^*))^{-1}(R\widehat{\beta}^* - r) \\ &\qquad = (R\widehat{\beta}^* - r)'\big(R(X'\Omega^{-1}X)^{-1}R'\big)^{-1}(R\widehat{\beta}^* - r)\end{aligned} \tag{5.9}$$

при гипотезе H_0 имеет распределение $\chi^2(q)$, где $q = \mathrm{rank}(R)$ (приложение МС, п. 4, N9). А можно точно так же, как и в п. (3.5),

для системы (5.6) провести две регрессии — без ограничения и с ограничением, получить соответствующие остатки e^*_{UR} и e^*_{R} и составить статистику

$$F = \frac{(e^{*\prime}_{\mathrm{R}} e^*_{\mathrm{R}} - e^{*\prime}_{\mathrm{UR}} e^*_{\mathrm{UR}})/q}{(e^{*\prime}_{\mathrm{UR}} e^*_{\mathrm{UR}})/(n-k)},$$

которая при гипотезе H_0 имеет распределение Фишера $F(q,\, n-k)$.

Если вектор ошибок $\boldsymbol{\varepsilon}$ имеет многомерное нормальное распределение, то можно проверить, что оценка вектора $\boldsymbol{\beta}$, получаемая с помощью обобщенного метода наименьших квадратов, совпадает с оценкой максимального правдоподобия (естественно, при известной матрице $\boldsymbol{\Omega}$): $\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}} = \widehat{\boldsymbol{\beta}}_{\mathrm{ML}}$.

Для обобщенной регрессионной модели, в отличие от классической, коэффициент детерминации

$$R^2 = 1 - \frac{(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}})'(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}})}{\sum(y_t - \overline{y})^2}$$

не может служить удовлетворительной мерой качества подгонки. В общем случае он даже не обязан лежать в интервале [0, 1], а добавление или удаление независимой переменной не обязательно приводит к его увеличению или уменьшению. Также нет особого смысла ориентироваться на коэффициент детерминации для регрессии (5.6): во-первых, даже если среди исходных регрессоров $\boldsymbol{X}$ содержался постоянный член, в преобразованных регрессорах $\boldsymbol{X}^*$ его может не оказаться, а во-вторых, в общем случае трудно установить связь между качеством подгонки вспомогательной модели (5.6) и исходной модели.

Подчеркнем еще раз, что для применения ОМНК необходимо знать матрицу $\boldsymbol{\Omega}$, что на практике бывает крайне редко. Поэтому вполне естественным кажется такой способ: оценить (каким-нибудь образом) матрицу $\boldsymbol{\Omega}$, а затем использовать эту оценку в формуле (5.4) вместо $\boldsymbol{\Omega}$. Этот подход составляет суть так называемого *доступного обобщенного метода наименьших квадратов* (*Feasible Generalized Least Squares*), о котором подробнее говорится в разделе 5.3. Построенную с его помощью оценку обозначим

$\widehat{\boldsymbol{\beta}}_{\mathrm{FGLS}}$. Следует понимать, что в общем случае матрица $\boldsymbol{\Omega}$ содержит $n(n+1)/2$ неизвестных параметров (в силу ее симметричности) и, имея только n наблюдений, нет никакой надежды получить для нее «хорошую» оценку. Поэтому для получения содержательных результатов приходится вводить дополнительные условия на структуру матрицы $\boldsymbol{\Omega}$.

Выводы:

1) для обобщенной регрессионной модели обычная МНК-оценка $\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ вектора $\boldsymbol{\beta}$ является несмещенной, состоятельной, но в отличие от классического случая не эффективной (в смысле минимума ковариационной матрицы);

2) оценка матрицы ковариаций вектора $\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ является смещенной;

3) эффективной в классе линейных несмещенных оценок является оценка (5.4), получаемая обобщенным методом наименьших квадратов (ОМНК);

4) для нахождения ОМНК-оценки $\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}$ необходимо знать ковариационную матрицу $\boldsymbol{\Omega}$ вектора ошибок;

5) ОМНК-оценка может быть получена применением обычного метода наименьших квадратов к вспомогательной системе (5.6), получаемой линейным преобразованием исходной модели (5.3);

6) проверка гипотез о наличии линейных ограничений проводится так же, как и в классическом случае либо непосредственно, либо с помощью остатков регрессий без ограничений и с ограничением для вспомогательной модели (5.6);

7) в случае нормального распределения вектора ошибок ОМНК-оценка совпадает с оценкой максимального правдоподобия;

8) коэффициент детерминации не может служить удовлетворительной мерой качества подгонки при использовании обобщенного метода наименьших квадратов.

5.3. Доступный обобщенный метод наименьших квадратов

Рассмотрим стандартную линейную модель

$$y = X\beta + \varepsilon, \tag{5.10}$$

где

$$\varepsilon \sim N(0, \sigma^2\Omega). \tag{5.11}$$

В случае когда $n \times n$ нормированная[1] матрица ковариаций Ω полностью известна, то, как было показано в разделе 5.2, наилучшая линейная несмещенная оценка (а также оценка максимального правдоподобия для β) задается формулой (см. (5.4))

$$\widehat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \tag{5.12}$$

и распределена по нормальному закону:

$$\widehat{\beta} \sim N(\beta, \sigma^2(X'\Omega^{-1}X)^{-1}). \tag{5.13}$$

Напомним (п. 5.2), что оценка (5.12) называется оценкой обобщенного метода наименьших квадратов и может быть получена из решения оптимизационной задачи:

$$\min_{\beta} (y - X\beta)'\Omega^{-1}(y - X\beta). \tag{5.14}$$

На практике матрица Ω почти никогда неизвестна. Мы предположим, что нам задана *структура* матрицы Ω (т. е. форма ее функциональной зависимости от сравнительно небольшого количества параметров), но не сами значения параметров. Например, мы можем знать (или допустить), что ошибки в (5.10) порождаются авторегрессионным процессом первого порядка, так что

$$\Omega(\rho) = \begin{bmatrix} 1 & \rho & \dots & \rho^{n-1} \\ \rho & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho^{n-1} & \dots & \rho & 1 \end{bmatrix}, \tag{5.15}$$

[1]Здесь мы используем нормировку $\mathrm{tr}(\Omega) = 1$.

где ρ — неизвестный параметр, который следует оценить. (Конечно, на практике мы обычно не знаем структуры Ω, но отказ от этого предположения выходит за рамки данной книги.)

Итак, предположим, что Ω зависит от конечного числа параметров $\theta_1, \dots, \theta_m$. Обозначим через θ вектор параметров $\theta = (\theta_1, \dots, \theta_m)'$ и будем считать, что β и (σ^2, θ) *функционально независимы*. Этим мы исключаем случаи, когда параметры ковариационной матрицы являются функциями от β.

Пусть $\widehat{\theta}$ — *состоятельная* оценка параметра θ. Обозначим $\widehat{\Omega} = \Omega(\widehat{\theta})$. Тогда оценкой *доступного обобщенного метода наименьших квадратов* (*Feasible Generalised Least Squares, FGLS*) называется величина

$$\widehat{\widehat{\beta}} = (X'\widehat{\Omega}^{-1}X)^{-1}X'\widehat{\Omega}^{-1}y. \tag{5.16}$$

Если $\widehat{\theta}$ — состоятельная оценка параметра θ, то можно было бы предположить, что $\widehat{\widehat{\beta}}$ тоже является состоятельной оценкой для β. Однако в общем случае это неверно.

Тем не менее можно показать, что если выполнены условия

$$\lim \frac{X'\Omega^{-1}X}{n} = Q \tag{5.17}$$

(Q — конечная, невырожденная матрица) и

$$p\lim \frac{X'\Omega^{-1}\varepsilon}{n} = 0, \tag{5.18}$$

то оценка доступного обобщенного метода наименьших квадратов $\widehat{\widehat{\beta}}$ состоятельна.

Оценка максимального правдоподобия[2]. Напомним, что функция правдоподобия для системы (5.10) при условии (5.11) есть (см. главу 10)

$$L = (2\pi)^{-n/2}(\sigma^2)^{-n/2}|\Omega|^{-1/2}\exp\left(-\frac{(y - X\beta)'\Omega^{-1}(y - X\beta)}{2\sigma^2}\right),$$

[2]Подробно применение оценок максимального правдоподобия в регрессии рассматривается в главе 10.

а ее логарифм равен

$$\ln L = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) + \frac{1}{2}\ln|\boldsymbol{\Omega}^{-1}| - \frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{\Omega}^{-1}(\boldsymbol{y}-\boldsymbol{X\beta}).$$

Приравнивая к нулю производные $\ln L$ по $\boldsymbol{\beta}$ и по σ^2, получаем

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\widehat{\boldsymbol{\Omega}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\Omega}}^{-1}\boldsymbol{y}, \tag{5.19}$$

$$\widehat{\sigma}^2 = \frac{1}{n}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})'\widehat{\boldsymbol{\Omega}}^{-1}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}), \tag{5.20}$$

дифференцирование по θ_j $(j = 1, \ldots, m)$ дает

$$\frac{\boldsymbol{e}'\boldsymbol{C}_j\boldsymbol{e}}{\boldsymbol{e}'\widehat{\boldsymbol{\Omega}}^{-1}\boldsymbol{e}} = \frac{1}{n}\operatorname{tr}(\boldsymbol{C}_j\widehat{\boldsymbol{\Omega}}), \tag{5.21}$$

где

$$\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}, \quad \boldsymbol{C}_j = \frac{\partial\boldsymbol{\Omega}^{-1}(\widehat{\boldsymbol{\theta}})}{\partial\theta_j}.$$

Решением системы уравнений (5.19)–(5.21) являются оценки максимального правдоподобия $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\theta}}$, $\widehat{\sigma}^2$. В этой системе только решение уравнений (5.21) может представлять трудность. В некоторых случаях (5.21) удается решить явно, но в большинстве случаев необходимо применять численные итерационные процедуры.

Интересно заметить, что из (5.21) и симметричности распределения $\boldsymbol{\varepsilon}$ следует, что $(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ и $-(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ имеют одинаковую плотность. Отсюда вытекает, что $\widehat{\boldsymbol{\beta}}$ симметрично распределено вокруг $\boldsymbol{\beta}$ и, следовательно, является несмещенной оценкой, если существует ее математическое ожидание.

Есть несколько способов работы с системой (5.19)–(5.21). Можно искать точное решение — оценку максимального правдоподобия, можно также использовать следующую весьма популярную двухшаговую процедуру.

1) Вычисляем оценку метода наименьших квадратов $\widehat{\boldsymbol{\beta}}_{(1)} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$.

   Вычисляем остатки метода наименьших квадратов $\boldsymbol{e}_{(1)}$.

   Решаем систему (5.21) при заданных остатках. Получаем $m \times 1$ вектор $\widehat{\boldsymbol{\theta}}_{(1)}$.

   Вычисляем $\widehat{\boldsymbol{\Omega}}_{(1)} = \boldsymbol{\Omega}(\widehat{\boldsymbol{\theta}}_{(1)})$.

2) $\widehat{\boldsymbol{\beta}}_{(2)} = (\boldsymbol{X}'\widehat{\boldsymbol{\Omega}}_{(1)}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\widehat{\boldsymbol{\Omega}}_{(1)}^{-1}\boldsymbol{y}$.

При некоторых слабых предположениях (таких, например, как состоятельность $\widehat{\boldsymbol{\theta}}_{(1)}$) $\widehat{\boldsymbol{\beta}}_{(2)}$ будет асимптотически эквивалентна оценке максимального правдоподобия. А как известно, в широком числе случаев оценка максимального правдоподобия асимптотически эффективна.

Большинство двухшаговых процедур (например, процедура Кохрейна–Оркатта) могут быть интерпретированы как итеративные процедуры в рамках метода максимального правдоподобия и, таким образом, при слабых предположениях, асимптотически эквивалентны оценке максимального правдоподобия.

Упражнения

5.1. Проверьте несмещенность оценки (5.4).

5.2. Проверьте равенство (5.8).

5.3. Докажите, что $\operatorname{Cov}(\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}, \widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}) = \mathrm{V}(\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}})$.

5.4. Согласно результатам п. 3.2 для классической регрессионной модели $\operatorname{Cov}(\widehat{y}_t, e_t) = 0$, $t = 1, \ldots, n$, где $\widehat{\boldsymbol{y}} = (\widehat{y}_1, \ldots, \widehat{y}_n)' = \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ — прогнозное значение $\boldsymbol{y}$, $\boldsymbol{e} = (e_1, \ldots, e_n)' = \boldsymbol{y} - \widehat{\boldsymbol{y}}$ — вектор остатков. Сохраняется ли это свойство для обобщенной регрессионной модели (5.3), т. е. верно ли, что $\operatorname{Cov}(\widehat{\boldsymbol{y}}, \boldsymbol{e}) = 0$, где $\widehat{\boldsymbol{y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ и $\boldsymbol{e} = \boldsymbol{y} - \widehat{\boldsymbol{y}}$?

5.5. Докажите, что если в (5.3) вектор ошибок $\boldsymbol{\varepsilon}$ имеет многомерное нормальное распределение, то $\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}} = \widehat{\boldsymbol{\beta}}_{\mathrm{ML}}$.

5.6. Рассмотрим уравнение регрессии:

$$y_t = \beta + \varepsilon_t, \quad t = 1, \ldots, n.$$

Пусть ошибки регрессии удовлетворяют следующим условиям:

$$\mathrm{E}(\varepsilon_t) = 0; \quad \operatorname{Cov}(\varepsilon_t, \varepsilon_s) = 0,\ t \neq s; \quad \mathrm{V}(\varepsilon_t) = \sigma^2 x_t,\ x_t > 0.$$

а) Найдите оценку метода наименьших квадратов $\widehat{\beta}$ и ее дисперсию.

б) Предложите несмещенную оценку, обладающую меньшей дисперсией, чем оценка метода наименьших квадратов. Получите дисперсию этой оценки и сравните ее с дисперсией оценки метода наименьших квадратов. Интерпретируйте результат.

5.7. Рассмотрим следующую регрессионную модель, в которой $2n$ наблюдений разбиты на две равные группы по n наблюдений в каждой:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad \mathrm{E}(\boldsymbol{\varepsilon}) = \boldsymbol{0}; \quad \mathrm{Cov}(\varepsilon_t, \varepsilon_s) = 0,\ t \neq s;$$
$$\mathrm{V}(\varepsilon_t) = \sigma_1^2,\ t = 1, \ldots, n; \quad \mathrm{V}(\varepsilon_t) = \sigma_2^2,\ t = n+1, \ldots, 2n.$$

Введем естественное разбиение матриц на блоки:

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{bmatrix}.$$

(Здесь $\boldsymbol{y}_1$, $\boldsymbol{y}_2$, $\boldsymbol{\varepsilon}_1$, $\boldsymbol{\varepsilon}_2$ — $n \times 1$ векторы, $\boldsymbol{X}_1$, $\boldsymbol{X}_2$ — $n \times k$ матрицы.)

а) Пусть $\widehat{\boldsymbol{\beta}}_1$, $\widehat{\boldsymbol{\beta}}_2$ и $\widehat{\boldsymbol{\beta}}$ — оценки метода наименьших квадратов вектора коэффициентов $\boldsymbol{\beta}$ по первой группе наблюдений, по второй группе наблюдений и по всем $2n$ наблюдениям, соответственно. Покажите, что $\widehat{\boldsymbol{\beta}}$ есть «взвешенное среднее» оценок $\widehat{\boldsymbol{\beta}}_1$ и $\widehat{\boldsymbol{\beta}}_2$, в том смысле, что $\widehat{\boldsymbol{\beta}} = \boldsymbol{L}_1\widehat{\boldsymbol{\beta}}_1 + \boldsymbol{L}_2\widehat{\boldsymbol{\beta}}_2$, где $\boldsymbol{L}_1$ и $\boldsymbol{L}_1$ — $k \times k$ матрицы такие, что $\boldsymbol{L}_1 + \boldsymbol{L}_2 = \boldsymbol{I}_k$.

б) Выведите следующие формулы для оценки обобщенного метода наименьших квадратов:

$$\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}} = \left(\frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{\sigma_1^2} + \frac{\boldsymbol{X}_2'\boldsymbol{X}_2}{\sigma_2^2}\right)^{-1} \left(\frac{\boldsymbol{X}_1'\boldsymbol{y}_1}{\sigma_1^2} + \frac{\boldsymbol{X}_2'\boldsymbol{y}_2}{\sigma_2^2}\right),$$
$$\mathrm{V}(\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}) = \left(\frac{\boldsymbol{X}_1'\boldsymbol{X}_1}{\sigma_1^2} + \frac{\boldsymbol{X}_2'\boldsymbol{X}_2}{\sigma_2^2}\right)^{-1}.$$

в) Покажите, что $\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}$ также является «взвешенным средним» оценок $\widehat{\boldsymbol{\beta}}_1$ и $\widehat{\boldsymbol{\beta}}_2$, в том смысле, что существуют $k \times k$ матрицы $\boldsymbol{\Lambda}_1$ и $\boldsymbol{\Lambda}_2$ такие, что $\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}} = \boldsymbol{\Lambda}_1\widehat{\boldsymbol{\beta}}_1 + \boldsymbol{\Lambda}_2\widehat{\boldsymbol{\beta}}_2$ и $\boldsymbol{\Lambda}_1 + \boldsymbol{\Lambda}_2 = \boldsymbol{I}_k$.

5.8. Рассмотрим модель из упражнения 5.7. Опишите процедуру доступного обобщенного метода наименьших квадратов в применении к этой модели.

5.9. В этом упражнении мы покажем, что если $\boldsymbol{\Omega} = \boldsymbol{\Omega}(\boldsymbol{\theta})$ и $\widehat{\boldsymbol{\theta}}$ — состоятельная оценка $\boldsymbol{\theta}$, то, вообще говоря, оценка доступного обобщенного метода наименьших квадратов не будет иметь то же асимптотическое распределение, что и оценка обобщенного метода наименьших квадратов.

Пусть

$$\boldsymbol{x} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \boldsymbol{\Omega}^{-1}(\theta) = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & \theta & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \theta^{n-1} \end{bmatrix} \quad \text{и} \quad \widehat{\theta} = \theta + \frac{1}{n}.$$

а) Покажите, что $p\lim(\widehat{\theta}) = \theta$.

б) При $\theta = 1$ покажите, что

$$\frac{1}{n}\boldsymbol{x}'\boldsymbol{\Omega}^{-1}\boldsymbol{x} = 1$$

и

$$\frac{\boldsymbol{x}'\boldsymbol{\Omega}^{-1}\boldsymbol{\varepsilon}}{\sqrt{n}} \sim N(0, \sigma^2).$$

в) Пусть $\widehat{\beta}(\theta) = (\boldsymbol{x}'\boldsymbol{\Omega}^{-1}(\theta)\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{\Omega}^{-1}(\theta)\boldsymbol{y}$. Покажите, что при $\theta = 1$

$$\sqrt{n}(\widehat{\beta}(\theta) - \beta) \sim N(0, \sigma^2).$$

г) С другой стороны, покажите, что при $\theta = 1$

$$\frac{\boldsymbol{x}'\boldsymbol{\Omega}^{-1}(\widehat{\theta})\boldsymbol{x}}{n} \to e - 1$$

и

$$\frac{\boldsymbol{x}'\boldsymbol{\Omega}^{-1}(\widehat{\theta})\boldsymbol{\varepsilon}}{\sqrt{n}} \xrightarrow{d} N\left(0, \sigma^2\frac{e^2-1}{2}\right).$$

д) Следовательно, при $\theta = 1$

$$\sqrt{n}(\widehat{\beta}(\widehat{\theta}) - \beta) \xrightarrow{d} N\left(0, \sigma^2\frac{e+1}{2(e-1)}\right)$$

(напомним, что символом $\xrightarrow{d}$ обозначается сходимость по распределению (см. приложение МС, п. 6)).

е) Выведите отсюда, что когда значение θ равно 1, асимптотическое распределение $\widehat{\beta}(\widehat{\theta})$ (оценки доступного обобщенного метода наименьших квадратов) *не совпадает* с асимптотическим распределением $\widehat{\beta}(\theta)$ (оценки обобщенного метода наименьших квадратов) вопреки тому, что $\widehat{\theta}$ является состоятельной оценкой θ.

5.10. Дана обобщенная линейная регрессионная модель $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$, $\mathrm{E}\boldsymbol{\varepsilon} = 0$, $\mathrm{V}(\boldsymbol{\varepsilon}) = \boldsymbol{\Omega}$. Пусть $\widehat{\boldsymbol{\beta}}$ — оценка вектора $\boldsymbol{\beta}$ с помощью обычного метода наименьших квадратов, и пусть $\widehat{\boldsymbol{y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}$.

а) Вычислите $\mathrm{V}(\widehat{\boldsymbol{y}})$.

б) Вычислите $\mathrm{V}(\boldsymbol{e}) = \mathrm{V}(\boldsymbol{y} - \widehat{\boldsymbol{y}})$.

в) Покажите, что в общем случае $\boldsymbol{e}$ и $\widehat{\boldsymbol{y}}$ коррелированы.

5.11. Как известно (см. задачу 3.26), для классической линейной модели $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$, $\mathrm{E}\boldsymbol{\varepsilon} = 0$, $\mathrm{V}(\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{I}$ выполнено неравенство $\mathrm{V}(\widehat{\boldsymbol{\beta}}_R) \leqslant \mathrm{V}(\widehat{\boldsymbol{\beta}})$, где $\widehat{\boldsymbol{\beta}}$ — МНК-оценка вектора $\boldsymbol{\beta}$, а $\widehat{\boldsymbol{\beta}}_R$ — оценка, получаемая регрессией $\boldsymbol{y}$ на $\boldsymbol{X}$ при линейном ограничении $\boldsymbol{H\beta} = \boldsymbol{r}$. Сохраняется ли это неравенство (для тех же оценок), если модель обобщенная, т. е. $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$, $\mathrm{E}\boldsymbol{\varepsilon} = 0$, $\mathrm{V}(\boldsymbol{\varepsilon}) = \boldsymbol{\Omega}$?

5.12. Пусть $\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}} = (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{y}$ — оценка, полученная с помощью обобщенного метода наименьших квадратов в обобщенной модели $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$, $\mathrm{E}\boldsymbol{\varepsilon} = 0$, $\mathrm{V}(\boldsymbol{\varepsilon}) = \boldsymbol{\Omega}$. Определим коэффициент детерминации

$$R^2_{\mathrm{GLS}} = 1 - \frac{\boldsymbol{e}'\boldsymbol{e}}{\sum_{t=1}^{n}(y_t - \overline{y})^2}, \quad \text{где} \quad \boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}.$$

Обладает ли этот коэффициент привычными свойствами коэффициента детерминации в классической линейной модели? В частности, верно ли, что R^2_{GLS} лежит в интервале $[0, 1]$?

5.13. Пусть в уравнении $y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + \varepsilon_t$, $t = 1, \ldots, n$, ошибки удовлетворяют уравнению авторегрессии первого порядка $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$, $u_t \sim iid(0, \sigma_u^2)$. Пусть $\boldsymbol{\Omega} = \mathrm{V}(\boldsymbol{\varepsilon})$. Найти матрицу $\boldsymbol{P}$ такую, что $\boldsymbol{\Omega}^{-1} = \boldsymbol{P}'\boldsymbol{P}$. Покажите, как выглядит преобразованное уравнение $\boldsymbol{Py} = \boldsymbol{PX\beta} + \boldsymbol{P\varepsilon}$, которое используется для вычисления оценок обобщенного метода наименьших квадратов.

Глава 6

Гетероскедастичность и корреляция по времени

Эта глава посвящена изучению двух важных классов обобщенных регрессионных моделей. Первый составляют модели с гетероскедастичностью. Этот термин применяется в ситуации, когда матрица ковариаций вектора ошибок является диагональной, но элементы главной диагонали, вообще говоря, различны. Иными словами, ошибки в разных наблюдениях некоррелированы, но их дисперсии — разные. Модели второго класса, как правило, используются при анализе данных, имеющих характер временных рядов. В этих случаях часто приходится принимать во внимание то обстоятельство, что наблюдения в разные моменты времени статистически зависимы (типичный пример — ежедневный обменный курс доллара по отношению к рублю). Следовательно, ошибки, относящиеся к разным наблюдениям (разным моментам времени), могут быть коррелированы, и ковариационная матрица вектора ошибок не является диагональной. Формально проблему оценивания неизвестных параметров решает обобщенный метод наименьших квадратов, рассмотренный в предыдущей главе. Однако, как там отмечалось, его применение требует знания матрицы ковариаций $\boldsymbol{\Omega}$ вектора ошибок, что бывает крайне редко. Поэтому, помимо те-

оретических вопросов, в данной главе будут затронуты некоторые аспекты практического использования ОМНК.

6.1. Гетероскедастичность

В этом разделе мы рассмотрим частный случай обобщенной регрессионной модели, а именно, модель с *гетероскедастичностью*. Это означает, что ошибки некоррелированы, но имеют непостоянные дисперсии. (Классическая модель с постоянными дисперсиями ошибок называется *гомоскедастичной*.) Как уже отмечалось, гетероскедастичность довольно часто возникает, если анализируемые объекты, говоря нестрого, неоднородны. Например, если исследуется зависимость прибыли предприятия от каких-либо факторов, скажем, от размера основного фонда, то естественно ожидать, что для больших предприятий колебание прибыли будет выше, чем для малых.

Метод взвешенных наименьших квадратов

Итак, пусть

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}, \tag{6.1}$$

и предположим, что ковариационная матрица $\boldsymbol{\Omega}$ вектора ошибок $\boldsymbol{\varepsilon}$ диагональна, $\mathrm{V}(\varepsilon_t) = \sigma_t^2$, $t = 1, \ldots, n$. Иногда удобно использовать представление $\sigma_t^2 = \sigma^2\omega_t$, где числа ω_t нормированы таким образом, что $\sum \omega_t = n$. Тогда при $\omega_t = 1$, $t = 1, \ldots, n$, модель сводится к классической.

Обобщенный метод наименьших квадратов в данном случае выглядит очень просто — вспомогательная система (5.6) получается делением каждого уравнения в (6.1) на соответствующее σ_t (здесь нам удобнее выписать каждое уравнение):

$$\frac{y_t}{\sigma_t} = \sum_{j=1}^{k} \beta_j \frac{x_{tj}}{\sigma_t} + u_t, \qquad t = 1, \ldots, n, \tag{6.2}$$

где $u_t = \varepsilon_t/\sigma_t$, причем $\mathrm{V}(u_t) = 1$, $\mathrm{Cov}\,(u_t, u_s) = 0$ при $t \neq s$. Применяя к (6.2) стандартный метод наименьших квадратов, ОМНК-

оценку получаем минимизацией по $\boldsymbol{b} = (b_1, \ldots, b_k)'$ суммы

$$f(\boldsymbol{b}) = \sum_{t=1}^{n} \left(\frac{1}{\sigma_t} \left(y_t - \sum_{j=1}^{k} b_j x_{tj} \right) \right)^2.$$

Нетрудно понять содержательный смысл этого преобразования. Используя обычный метод наименьших квадратов, мы минимизируем сумму квадратов отклонений $\varphi(\boldsymbol{b}) = \sum_{t=1}^{n}(y_t - \sum_{j=1}^{k} b_j x_{tj})^2$, в которую, говоря нестрого, разные слагаемые дают разный статистический вклад из-за различных дисперсий, что в конечном итоге и приводит к неэффективности МНК-оценки. «Взвешивая» каждое наблюдение с помощью коэффициента $1/\sigma_t$, мы устраняем такую неоднородность (заметим, что это означает, что мы придаем больший «вес» наблюдениям с меньшей дисперсией, т.е. более «точным»). Поэтому часто обобщенный метод наименьших квадратов для системы с гетероскедастичностью называют методом *взвешенных наименьших квадратов*. Можно непосредственно проверить (упражнение 6.1), что применение метода взвешенных наименьших квадратов приводит к уменьшению дисперсий оценок по сравнению с обычным методом наименьших квадратов.

Коррекция на гетероскедастичность

Если числа σ_t неизвестны (что, как правило, и бывает на практике), необходимо использовать доступный обобщенный метод наименьших квадратов, который требует оценивания дисперсий σ_t^2. Так как число этих параметров равно n, то без дополнительных ограничений на структуру матрицы $\boldsymbol{\Omega}$ нет надежды получить приемлемые оценки дисперсий. Ниже мы рассмотрим несколько классов моделей с гетероскедастичностью, где такие ограничения накладываются и благодаря этому удается построить удовлетворительные оценки матрицы $\boldsymbol{\Omega}$, а следовательно, используя доступный обобщенный метод наименьших квадратов, и оценку $\widehat{\boldsymbol{\beta}}_{\mathrm{FGLS}}$.

1. *Стандартное отклонение ошибки пропорционально независимой переменной.* В некоторых ситуациях априорно можно счи-

тать, что стандартное отклонение ошибки прямо пропорционально одной из независимых переменных, например, x_k: $\sigma_t^2 = \sigma^2 x_{tk}^2$. Тогда, разделив t-е уравнение на x_{tk}, $t = 1, \ldots, n$, и вводя новые независимые переменные $x_{tj}^* = x_{tj}/x_{tk}$ и новую зависимую переменную $y_t^* = y_t/x_{tk}$, $t = 1, \ldots, n$, $j = 1, \ldots, k$, получим классическую регрессионную модель. МНК-оценки коэффициентов этой модели дают непосредственно оценки исходной модели. Следует только помнить, что если первый регрессор в $\boldsymbol{X}$ есть набор единиц, то оценки свободного члена и коэффициента при $x_{t1}^* = 1/x_{tk}$ в новой модели являются оценками соответственно коэффициента при x_{tk} и свободного члена в исходной модели.

Возникает естественный вопрос, при каких обстоятельствах можно пользоваться описанным выше методом. Ниже будут описаны некоторые процедуры, позволяющие выявлять гетероскедастичность того или иного рода (тесты на гетероскедастичность). Здесь мы ограничимся лишь практическими рекомендациями. Если есть предположение о зависимости ошибок от одной из независимых переменных, то целесообразно расположить наблюдения в порядке возрастания значений этой переменной, а затем провести обычную регрессию и получить остатки. Если размах их колебаний тоже возрастает (это хорошо заметно при обычном визуальном исследовании), то это говорит в пользу исходного предположения. Тогда надо сделать описанное выше преобразование, вновь провести регрессию и исследовать остатки. Если теперь их колебание имеет неупорядоченный характер, то это может служить показателем того, что коррекция на гетероскедастичность прошла успешно. Естественно, следует сравнивать и другие параметры регрессии (значимость оценок, сумму квадратов отклонений и т. п.) и только тогда принимать окончательное решение, какая из моделей более приемлема.

Пример. ***Рынок квартир в Москве*** (см. Каргин, Онацкий, 1996). Продолжение 2 (см. начало — п. 3.5, продолжение 1 — п. 4.2).

Как мы увидим далее, при более тщательном изучении данных примера (см. продолжение 3, п. 6.1), гипотеза гомоскедастичности ошибок должна быть отвергнута. Это ставит под сомнение выводы о значимости регрессоров (и результаты тестирования гипотез на

коэффициенты), сделанные в начале и продолжении 1 этого примера.

Для учета гетероскедастичности использовалась так называемая двухшаговая процедура оценки (*two-step estimation* (см. Greene, 1997), не путать с двухшаговым методом наименьших квадратов, *two stage least squares*, п. 8.2), которая является обобщением описанной выше ситуации, когда дисперсия пропорциональна одному из регрессоров. Такая двухшаговая процедура дает асимптотически несмещенные оценки стандартных ошибок коэффициентов регрессии.

Предполагается, что дисперсия ошибки есть линейная функция от нескольких регрессоров, в данном случае

$$\sigma_i^2 = \beta_0 + \beta_1 \mathrm{LOGLIVSP}_i + \beta_2 \mathrm{BRICK}_i. \qquad (*)$$

На первом шаге процедуры мы оцениваем регрессионное уравнение нашей модели:

$$\begin{aligned}\mathrm{LOGPRICE} = {} & \beta_0 + \beta_1 \mathrm{LOGLIVSP} + \beta_2 \mathrm{LOGPLAN} + \beta_3 \mathrm{LOGKITSP} \\ & + \beta_4 \mathrm{LOGDIST} + \beta_5 \mathrm{FLOOR} + \beta_6 \mathrm{BRICK} + \beta_7 \mathrm{BAL} \\ & + \beta_8 \mathrm{LIFT} + \beta_9 \mathrm{R1} + \beta_{10} \mathrm{R2} + \beta_{11} \mathrm{R3} + \beta_{12} \mathrm{R4} + \varepsilon, \quad (**)\end{aligned}$$

и из регрессии (*) (подставив вместо σ_i остатки e_i регрессии (**)) находим состоятельную оценку вектора дисперсий $\widehat{\sigma}_i^2$.

На втором шаге полученные оценки $\widehat{\sigma}_i$ используются в качестве весовых коэффициентов для *взвешенного метода наименьших квадратов* (i-е уравнение делится на $\widehat{\sigma}_i$) (см. стр. 168).

В таблице 6.1 приведены результаты описанной выше двухшаговой процедуры.

Из сравнения таблиц 3.1 и 6.1 видно, что существенных изменений в оценках не произошло.

Наибольшему изменению подверглись оценки коэффициентов при фиктивных переменных, отражающих количество комнат в квартире. Коэффициенты при R1 и R2 оказались значимыми в противоположность ранее полученному результату.

Больше других изменилась также оценка эластичности цены по жилой площади. Новая оценка эластичности равна 0.76 по сравнению с 0.67 ранее.

Пользуясь полученными результатами, мы сможем провести статистически осмысленную проверку гипотезы о равенстве коэффициентов при R2, R3 и R4 и неравенстве коэффициентов при R1

и R2 (см. п. 3.5, пример *Рынок квартир в Москве*, начало). Как и ранее, применяем F-тест для проверки гипотез:

1. H_0: $\beta_{10} = \beta_{11}$; $\beta_{11} = \beta_{12}$;

   F-статистика 1.548415 P-значение 0.213713.

2. H_0: $\beta_9 = \beta_{10}$;

   F-статистика 10.41677 P-значение 0.001340.

Таблица 6.1

Переменная	Коэффициент	Стандартная ошибка	t-статистика	P-значение
CONST	6.693	0.251	26.69	0.0000
LOGLIVSP	0.756	0.0536	14.11	0.0000
LOGPLAN	0.438	0.0468	9.36	0.0000
LOGKITSP	0.129	0.0479	2.69	0.0073
LOGDIST	−0.110	0.0135	−8.08	0.0000
BRICK	0.130	0.0198	6.58	0.0000
FLOOR	−0.0658	0.0170	−3.87	0.0001
LIFT	0.131	0.0187	7.01	0.0000
BAL	0.0350	0.0170	2.05	0.0400
R1	0.365	0.0892	4.08	0.0001
R2	0.249	0.0639	3.90	0.0001
R3	0.257	0.0473	5.43	0.0000
R4	0.205	0.0376	5.44	0.0000

Статистика R^2 (невзвешенная): 0.891.

Таким образом, как и ранее, гипотеза о равенстве коэффициентов при R2, R3 и R4 не может быть отвергнута; гипотеза о равенстве коэффициентов при R1 и R2 отвергается еще более уверенно — на 0.5%-ном уровне значимости.

2. *Дисперсия ошибки принимает только два значения.* Пусть известно, что $\sigma_t^2 = \omega_1^2$ для $t = 1, \ldots, n_1$ и $\sigma_t^2 = \omega_2^2$ для $t = n_1+1, \ldots, n_1+n_2$ $(n_1+n_2 = n)$, но числа ω_1^2 и ω_2^2 неизвестны. Иными словами, в первых n_1 наблюдениях дисперсия ошибки имеет одно значение, в последующих n_2 — другое. В этом случае естественным является следующий вариант доступного обобщенного метода наименьших квадратов:

1) провести обычную регрессию (6.1), получить вектор остатков e и разбить его на два подвектора e_1, e_2 размерности n_1 и n_2 соответственно;

2) построить оценки $\widehat{\omega}_1^2 = e_1'e_1/n_1$ и $\widehat{\omega}_2^2 = e_2'e_2/n_2$ дисперсий ω_1^2 и ω_2^2;

3) преобразовать переменные, разделив первые n_1 уравнений на $\widehat{\omega}_1$, а последующие n_2 — на $\widehat{\omega}_2$;

4) провести обычную регрессию для преобразованной модели.

Хотя, как было установлено ранее (п. 5.2), оценки $\widehat{\omega}_1^2$ и $\widehat{\omega}_2^2$ являются смещенными, можно показать (Goldberger, 1990, глава 30.5) их состоятельность.

Ясно, что эта модель допускает обобщение на случай, когда дисперсия принимает не два, а несколько значений.

3. *Состоятельное оценивание дисперсий.* Предположим теперь, что в модели (6.1) с гетероскедастичностью для оценки вектора параметра $\boldsymbol{\beta}$ используется обычный метод наименьших квадратов. Как установлено в главе 5, эта оценка является состоятельной и несмещенной, однако стандартная оценка ее матрицы ковариаций ((3.8), (3.19)) $\widehat{V}(\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}) = \widehat{\sigma}^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$ смещена и несостоятельна. Отметим, что компьютерные пакеты при оценивании коэффициентов регрессии вычисляют стандартные ошибки коэффициентов регрессии именно по этой формуле. Можно ли сделать поправку на гетероскедастичность и «улучшить» оценку матрицы ковариаций? Положительный ответ дают приводимые ниже два способа оценивания.

Стандартные ошибки в форме Уайта. Предположим, что $\boldsymbol{\Omega}$, матрица ковариаций вектора ошибок ε, диагональна, $\mathrm{V}(\varepsilon_t) = \sigma_t^2$, $t = 1, \ldots, n$. Тогда поскольку $\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}} = \widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}$, то

$$\mathrm{V}(\widehat{\boldsymbol{\beta}}) = \mathrm{E}\left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{X}\,(\boldsymbol{X}'\boldsymbol{X})^{-1}\right) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}\,\boldsymbol{X}\,(\boldsymbol{X}'\boldsymbol{X})^{-1}$$
$$= n(\boldsymbol{X}'\boldsymbol{X})^{-1}\Big(\frac{1}{n}\,\boldsymbol{X}'\boldsymbol{\Omega}\,\boldsymbol{X}\Big)(\boldsymbol{X}'\boldsymbol{X})^{-1}.$$

Рассмотрим матрицу $\boldsymbol{X}'\boldsymbol{\Omega}\,\boldsymbol{X}$. Имеем $(\boldsymbol{X}'\boldsymbol{\Omega}\,\boldsymbol{X})_{ij} = \sum\limits_{s=1}^{n} x_{si}\sigma_s^2 x_{sj}$. Обозначим через $\boldsymbol{x}_s'$, $s = 1, \ldots, n$, $1 \times k$ векторы-строки матрицы

регрессоров $\boldsymbol{X}$. Тогда

$$\boldsymbol{X}'\boldsymbol{\Omega}\,\boldsymbol{X} = \sum_{s=1}^{n} \sigma_s^2 \boldsymbol{x}_s \boldsymbol{x}_s'.$$

Уайт (White, 1980) показал, что

$$\widehat{\mathrm{V}}(\widehat{\beta}) = n(\boldsymbol{X}'\boldsymbol{X})^{-1}\left(\frac{1}{n}\sum_{s=1}^{n} e_s^2 \boldsymbol{x}_s \boldsymbol{x}_s'\right)(\boldsymbol{X}'\boldsymbol{X})^{-1} \qquad (6.3)$$

является состоятельной оценкой матрицы ковариаций оценок коэффициентов регрессии.

Стандартные отклонения, рассчитанные по формуле (6.3), называются *стандартными ошибками в форме Уайта* (*White standard errors*) или *состоятельными стандартными ошибками при наличии гетероскедастичности* (*Heteroscedasticity Consistent standard errors, HC s. e.*).

Стандартные ошибки в форме Ньюи–Веста. Для более сложного случая, когда в матрице ковариаций ошибок $\mathrm{V}(\boldsymbol{\varepsilon}) = \boldsymbol{\Omega} = (\omega_{ij})$ ненулевые элементы стоят не только на главной диагонали, но и на соседних диагоналях, отстоящих от главной не более чем на L (т. е. $\omega_{ij} = 0$, $|i-j| > L$), Ньюи и Вест (Newey, West, 1987) показали, что оценка

$$\widehat{\mathrm{V}}(\widehat{\boldsymbol{\beta}}) = n(\boldsymbol{X}'\boldsymbol{X})^{-1}\left(\frac{1}{n}\sum_{s=1}^{n} e_s^2 \boldsymbol{x}_s \boldsymbol{x}_s' + \frac{1}{n}\sum_{j=1}^{L}\sum_{t=j+1}^{n} w_j e_t e_{t-j}(\boldsymbol{x}_t \boldsymbol{x}_{t-j}' + \boldsymbol{x}_{t-j}\boldsymbol{x}_t')\right)(\boldsymbol{X}'\boldsymbol{X})^{-1} \qquad (6.4)$$

является состоятельной оценкой матрицы ковариаций оценок коэффициентов регрессии.

Существует несколько способов выбора весовых коэффициентов w_j.

1. Наиболее простым кажется взять $w_j = 1$. Однако при таком выборе матрица (6.4) может оказаться не неотрицательно определенной.

2. $w_j = 1 - \dfrac{j}{L+1}$ (*Бартлетт*).

3. $w_j = \begin{cases} 1 - 6\left(\frac{j}{L+1}\right)^2 + 6\left(\frac{j}{L+1}\right)^3, & 1 \leqslant j \leqslant \frac{L+1}{2}, \\ 2\left(1 - \frac{j}{L+1}\right)^2, & \frac{L+1}{2} < j \leqslant L \end{cases}$ (*Парзен*).

В большинстве случаев использование весов Парзена предпочтительнее.

Стандартные отклонения, рассчитанные по формуле (6.4), называются *стандартными ошибками в форме Ньюи-Веста* (*Newey-West standard errors*) или *стандартными ошибками с учетом гетероскедастичности и автокорреляции* (*Heteroscedasticity and Autocorrelation Consistent standard errors, HAC s. e.*).

Рассмотрим пример регрессии, в котором матрица ковариаций ошибок заведомо недиагональная.

Пример. ***Премия за риск*** (см. Peresetsky, de Roon, 1997).

Пусть S_t — спот-курс доллара (руб./долл.) в день t. $F_t^{(n)}$ — цена фьючерсного контракта в день t на поставку доллара через n дней (в день $t+n$).

Представляет интерес вопрос: является ли $F_t^{(n)}$ несмещенной оценкой будущего спот-курса S_{t+n}? А именно, верно ли равенство:

$$\mathrm{E}\,(S_{t+n} \mid I_t) = F_t^{(n)},$$

где I_t — информационное множество, содержащее всю информацию, доступную в момент t, $\mathrm{E}\,(S_{t+n} \mid I_t)$ — условное математическое ожидание. На самом деле, в присутствии на рынке агентов, избегающих риска, это равенство не обязано выполняться и цена фьючерса отличается от ожидаемого спот-курса на величину *премии за риск* $\widehat{\pi}_t^{(n)}$:

$$\mathrm{E}\,(S_{t+n} \mid I_t) = F_t^{(n)}(1 + \widehat{\pi}_t^{(n)}).$$

Удобно перейти к модели, в которой используются логарифмы переменных:

$$\mathrm{E}\,(s_{t+n} \mid I_t) = f_t^{(n)} + \pi_t^{(n)},$$

здесь $f_t^{(n)} = \ln F_t^{(n)}$; $s_t^{(n)} = \ln S_t^{(n)}$ и $\pi_t^{(n)} = \ln(1 + \widehat{\pi}_t^{(n)})$ — переопределение премии за риск ($\pi_t^{(n)} \approx \widehat{\pi}_t^{(n)}$ при малых $\widehat{\pi}_t^{(n)}$).

Для тестирования гипотезы о несмещенности оценки рассмотрим следующую регрессию:

$$s_{t+n} - f_t^{(n)} = (\beta_0 + \beta_1 n) + (\beta_2 + \beta_3 n) f_t^{(n)} + (\beta_4 + \beta_5 n)(f_t^{(n)} - s_t) + \varepsilon_{t+n}^{(n)}. \qquad (*)$$

Если в результате оценивания этого уравнения мы получим, что хотя бы один коэффициент β_i статистически достоверно отличается от 0, это будет означать наличие отличной от нуля премии за риск. (Отметим, что все регрессоры в уравнении (*) принадлежат информационному множеству I_t.) Более того, (*) предлагает определенную *временную структуру* премии за риск:

$$\pi_t^{(n)} = (\beta_0 + \beta_1 n) + (\beta_2 + \beta_3 n) f_t^{(n)} + (\beta_4 + \beta_5 n)(f_t^{(n)} - s_t). \qquad (**)$$

Рассматривается период с ноября 1992 г. по октябрь 1995 г. Данные о спот-курсе взяты с ММВБ (Московская межбанковская валютная биржа), а данные о котировках фьючерсных контрактов — с МТБ (Московская товарная биржа), на которой в этот период были самые значительные (по сравнению с другими биржами) объемы торговли валютными фьючерсами.

Упорядочим наблюдения по возрастанию n, а для одинаковых n — по возрастанию t. Отметим, что для уравнения (*) условия гомоскедастичности, безусловно, нарушены по двум причинам: 1) ошибки $\varepsilon_{t+n}^{(n)}$, $\varepsilon_{t+m}^{(m)}$ коррелированы (цены контрактов с разными сроками поставок в один и тот же день) и 2) $\varepsilon_{t+1+n}^{(n-1)}$, $\varepsilon_{t+n}^{(n)}$ также коррелированы (цены контракта со сроком поставки $t+n$ в последовательные дни).

Однако для наблюдений, отстоящих достаточно далеко друг от друга в нашем упорядочении, ошибки можно считать независимыми. Таким образом, для оценивания уравнения (*) нельзя использовать МНК-оценки стандартных отклонений коэффициентов и поэтому мы используем стандартные ошибки в форме Ньюи–Веста.

В статье (Яковлев, Бессонов, 1995б) рассматриваемый период развития фьючерсного рынка разбивается на три подпериода по институциональным признакам (состав участников рынка, объем торговли и т. п.). Следуя этой разбивке, мы оценивали уравнение (*) отдельно для каждого из подпериодов.

В таблице 6.2 приведены результаты оценивания. В квадратных скобках указаны стандартные ошибки в форме Ньюи–Веста с лагом 150.

Таблица 6.2

Период	const	n	$(f_t^{(n)} - s_t)$	$(f_t^{(n)} - s_t)n$	$f_t^{(n)}$	$f_t^{(n)}n$
11.92 – 10.93	1.54* [0.64]	0.016* [0.0076]	−0.99** [0.35]	0.0056 [0.0031]	−0.22* [0.091]	−0.0023* [0.0011]
10.93 – 03.94	−1.08 [0.94]	0.041** [0.0035]	−1.33** [0.25]	0.0080** [0.00054]	0.15 [0.13]	−0.0056** [0.00047]
03.94 – 10.95	−0.53 [0.41]	0.035** [0.0016]	−0.69 [0.38]	0.0015 [0.00088]	0.060 [0.052]	−0.0041** [0.00023]

* Отличается от нуля на 5%-ном уровне значимости.

** Отличается от нуля на 1%-ном уровне значимости.

Из таблицы 6.2 мы видим, что:

1) во всех трех периодах премия за риск $\pi_t^{(n)}$, моделируемая уравнением (**), статистически достоверно отличается от нуля;
2) премия за риск статистически достоверно различается от периода к периоду.

Таким образом, подтверждается разбиение истории развития рынка валютных фьючерсов на три периода, полученное А. Яковлевым и В. Бессоновым на основе анализа институциональных признаков.

Тесты на гетероскедастичность

Опишем несколько общеупотребительных статистических тестов на гетероскедастичность, не проводя их детального исследования. Как правило, из определения тестов будет ясно, какова их значимость. Проблему мощности тестов мы рассматривать не будем. Во всех этих тестах проверяется основная гипотеза H_0: $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_n^2$ против альтернативной гипотезы H_1: не H_0.

Большинство тестов ориентированы на те или иные ситуации, когда относительно характера гетероскедастичности есть априорные структурные ограничения. Исключение составляет тест Уайта.

Тест Уайта (White). Содержательный смысл этого теста состоит в следующем. Если в модели присутствует гетероскедастичность, то очень часто это связано с тем, что дисперсии ошибок

некоторым образом (возможно, довольно сложно) зависят от регрессоров, а гетероскедастичность должна как-то отражаться в остатках обычной регрессии исходной модели. Реализуя эти идеи, Уайт (White, 1980) предложил метод тестирования гипотезы H_0 без каких-либо предположений относительно структуры гетероскедастичности. Сначала к исходной модели (6.1) применяется обычный метод наименьших квадратов и находятся остатки регрессии e_t, $t = 1, ..., n$. Затем осуществляется регрессия квадратов этих остатков e_t^2 на все регрессоры $\boldsymbol{X}$, их квадраты, попарные произведения и константу, если ее не было в составе исходных регрессоров. Тогда при гипотезе H_0 величина nR^2 асимптотически имеет распределение $\chi^2(N-1)$, где R^2 — коэффициент детерминации, а N — число регрессоров второй регрессии.

Привлекательной чертой теста Уайта является его универсальность. Однако если гипотеза H_0 отвергается, этот тест не дает указания на функциональную форму гетероскедастичности, и единственным способом коррекции на гетероскедастичность является применение стандартных ошибок в форме Уайта.

Тест Голдфелда–Куандта (Goldfeld–Quandt). Этот тест применяется, как правило, когда есть предположение о прямой зависимости дисперсии ошибки от величины некоторой независимой переменной (ср. *коррекция на гетероскедастичность*, стр. 169, п. 1). Кратко тест можно описать следующим образом:

1) упорядочить данные по убыванию той независимой переменной, относительно которой есть подозрение на гетероскедастичность;

2) исключить d средних (в этом упорядочении) наблюдений (d должно быть примерно равно четверти общего количества наблюдений);

3) провести две независимые регрессии первых $n/2 - d/2$ наблюдений и последних $n/2 - d/2$ наблюдений и построить соответствующие остатки e_1 и e_2;

4) составить статистику $F = e_1'e_1/e_2'e_2$. Если верна гипотеза H_0, то F имеет распределение Фишера с $(n/2-d/2-k, n/2-d/2-k)$ степенями свободы (числитель и знаменатель в выражении для F следует разделить на соответствующее число степеней свободы, но в данном случае они одинаковы).

Большая величина этой статистики означает, что гипотезу H_0 следует отвергнуть. Количество исключаемых наблюдений не должно быть ни слишком мало, ни слишком велико. Формально тест работает и без исключения наблюдений, но, как показывает опыт, при этом его мощность уменьшается.

Аналогично этот тест используется, если есть предположение о межгрупповой гетероскедастичности, когда дисперсия ошибки принимает, например, только два возможных значения.

Тест Бреуша–Пагана (Breusch–Pagan). Этот тест применяется в тех случаях, когда априорно предполагается, что дисперсии σ_t^2 зависят от некоторых дополнительных переменных:

$$\sigma_t^2 = \gamma_0 + \boldsymbol{z}_t'\gamma, \quad t = 1, \ldots, n,$$

где $\boldsymbol{z}_t = (z_{t1}, \ldots, z_{tp})'$ — вектор (наблюдаемых) независимых переменных, γ_0, $\gamma = (\gamma_1, \ldots, \gamma_p)'$ — неизвестные параметры. В соответствии с тестом Бреуша–Пагана следует действовать так:

1) провести обычную регрессию (6.1) и получить вектор остатков $\boldsymbol{e} = (e_1, \ldots, e_n)'$;

2) построить оценку $\widehat{\sigma}^2 = (1/n)\sum e_t^2$;

3) провести регрессию $\dfrac{e_t^2}{\widehat{\sigma}^2} = \gamma_0 + \boldsymbol{z}_t'\gamma + \nu_t$ и найти для нее объясненную часть вариации RSS;

4) построить статистику RSS/2. В работе (Breusch, Pagan, 1979) установлено, что если верна гипотеза H_0 (отсутствие гетероскедастичности), то величина RSS/2 асимптотически имеет распределение $\chi^2(p)$.

При выявлении гетероскедастичности с помощью этого теста можно попытаться осуществить коррекцию с помощью метода взвешенных наименьших квадратов, выбирая в качестве весов величины $(\widehat{\gamma}_0 + \boldsymbol{z}_t'\widehat{\gamma})^{-1/2}$, где $\widehat{\gamma}_0$, $\widehat{\gamma}$ — оценки, полученные в п. 3).

При этом может оказаться, что $\widehat{\gamma}_0 + z_t'\widehat{\gamma} < 0$ для некоторых t. Если число таких наблюдений невелико, то их можно просто выбросить. В противном случае можно попытаться использовать мультипликативную форму гетероскедастичности:

$$\sigma_t^2 = e^{\gamma_0 + z_t'\gamma}, \quad t = 1, \ldots, n.$$

Процедура теста Бреуша–Пагана тогда выглядит совершенно аналогично изложенной выше в п. 3). Точно так же можно действовать для произвольной формы гетероскедастичности $\sigma_t^2 = f(\gamma_0 + z_t'\gamma)$.

Выводы:

1) применение обобщенного метода наименьших квадратов при наличии гетероскедастичности сводится к минимизации суммы взвешенных квадратов отклонений;
2) использование доступного обобщенного метода наименьших квадратов в общем случае требует оценивания n параметров по n наблюдениям, что не позволяет получать состоятельные оценки;
3) в некоторых ситуациях (ошибка пропорциональна одной из независимых переменных, дисперсии ошибок принимают два значения) можно применять доступный обобщенный метод наименьших квадратов и получать состоятельные оценки коэффициентов регрессии;
4) если в модели с гетероскедастичностью использовать обычный метод наименьших квадратов, то для получения состоятельной оценки соответствующей матрицы ковариаций можно применять оценки ошибок в форме Уайта (6.3) или Ньюи–Веста (6.4).

Пример. ***Рынок квартир в Москве*** (см. Каргин, Онацкий, 1996). Продолжение 3 (см. начало — п. 3.5, продолжение 1 — п. 4.2, продолжение 2 — п. 6.1).

Для тестирования ошибок модели (*) примера о ценах на квартиры в Москве на гетероскедастичность применяем тест Голдфелда–Куандта (см. выше) по переменной LOGLIVSP. Данные (464 наблюдения) делятся на три группы, примерно равные по объему. В первую группу попадают наблюдения с LOGLIVSP > 3.8 (155 наблюдений), во вторую — с LOGLIVSP < 3.35 (149 наблюдений).

Из-за возникновения *dummy trap* проблемы в первом случае пришлось «выбросить» переменную R1, а во втором — переменные R2, R3 и R4, таким образом, количество регрессоров в обоих случаях отличалось от первоначального числа 13. Соответственно число степеней свободы равнялось $143 = 155 - 12$ и $139 = 149 - 10$ (12 и 10 — это количество регрессоров соответственно в первой и во второй регрессии). Здесь использовано очевидное обобщение теста Голдфелда–Куандта на случай разного количества регрессоров.

После прогонки регрессий в каждой из групп получены следующие значения сумм квадратов остатков: $\mathbf{e}_1'\mathbf{e}_1 = 6.80$ и $\mathbf{e}_2'\mathbf{e}_2 = 3.76$. Таким образом, $F = \frac{\mathbf{e}_1'\mathbf{e}_1/143}{\mathbf{e}_2'\mathbf{e}_2/139} \approx 1.7$. Вероятность того, что случайная величина с распределением Фишера $F(143, 139)$ принимает значение меньше единицы, равна 95%. Полученная величина $F = 1.7$ превышает $F_{0.05}(143, 139)$, и гипотеза гомоскедастичности остатков должна быть отвергнута.

Замечание

Отметим отдельно, что надо внимательно относиться к интерпретации результатов тестов на гетероскедастичность. Дело в том, что неверная спецификация функциональной формы модели может привести к тому, что тест отвергает гипотезу гомоскедастичности. Поясним это на простейшем примере. Пусть истинная модель имеет вид $\exp(y_t) = \alpha + \beta x_t + \varepsilon_t$ с гомоскедастичными ошибками, т. е. $\mathrm{V}(\varepsilon_t) = \sigma^2$, а мы оцениваем линейную модель $y_t = \alpha + \beta x_t + \varepsilon_t$. В результате мы получим картину, похожую на приведенную на рис. 6.1.

Мы видим, что квадраты остатков регрессии e_t^2, которыми оперируют тесты на гетероскедастичность, зависят от значения переменной x_t, и, соответственно, тесты отвергают гипотезу гомоскедастичности, что в данном случае является следствием ошибки спецификации модели.

Рассмотрим пример, в котором мы встречаемся с данной ситуацией.

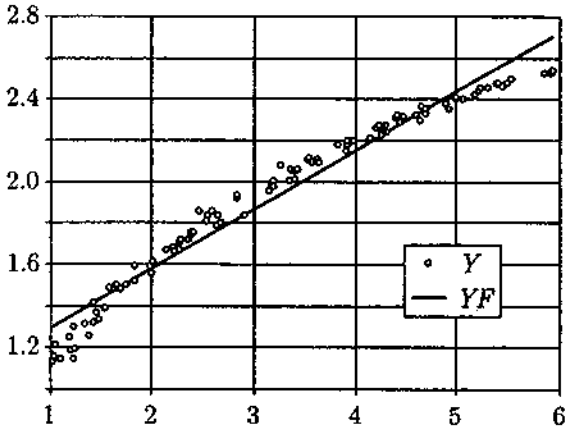


Рис. 6.1

Пример. ***Зарплата в Нидерландах*** (Arthur van Soest). Продолжение 1 (см. начало — п. 4.4, стр. 134).

Попробуем исследовать на этих данных зависимость зарплаты от возраста. Мы ожидаем, что до некоторого возраста зарплата растет (идет накопление опыта), а далее — убывает. Простейший способ учесть этот эффект — включить в уравнение как AGE так и AGE^2. Мы ожидаем получить положительный коэффициент при AGE и отрицательный при AGE^2. Результаты регрессии W на остальные переменные приведены в таблице 6.3.

Таблица 6.3

Dependent Variable: W

Variable	*Coefficient*	*Std.Error*	*t-Statistic*	*Probability*
const	8.0110	6.7978	1.1785	0.2405
SEX	−3.6826	1.2275	−3.0000	0.0032
EDU	3.3468	0.5511	6.0726	0.0000
AGE	0.1707	0.3687	0.4631	0.6400
AGE^2	0.0036	0.0048	0.7433	0.4585
R-squared	0.5173			

White Heteroscedasticity Test:

F-statistic	6.4422	*Probability*	0.0000
*Obs*R-squared*	31.9177	*Probability*	0.0000

Из таблицы видно, что коэффициенты при интересующих нас переменных AGE и AGE^2 не значимы. Тест Уайта показывает наличие гетероскедастичности. Прежде чем начать коррекцию гетероскедастичности, вспомним, что тест может давать такой результат при ошибке спецификации функциональной формы. В самом деле, поскольку, как правило, все надбавки к зарплате формулируются в мультипликативной форме («увеличение на 5%»), то более естественно взять в качестве зависимой переменной логарифм зарплаты $\ln W$. Результаты регрессии $\ln W$ на остальные переменные приведены в таблице 6.4.

Таблица 6.4

Dependent Variable: $\ln W$

Variable	*Coefficient*	*Std.Error*	*t-Statistic*	*Probability*
const	1.4217	0.2585	5.5001	0.0000
SEX	−0.1447	0.0467	−3.0005	0.0023
EDU	0.1244	0.0210	5.9367	0.0000
AGE	0.0660	0.0140	4.7045	0.0000
AGE^2	−0.00061	0.0002	−3.3225	0.0011
R-squared	0.6098			

White Heteroscedasticity Test:

F-statistic	1.5619	*Probability*	0.1625
*Obs*R-squared*	9.2254	*Probability*	0.1613

Теперь оба коэффициента значимо отличаются от нуля и имеют «правильные знаки». Тест Уайта показывает отсутствие гетероскедастичности. Из последнего уравнения можно также получить, что возраст, при котором достигается максимальная зарплата, равен примерно 54 годам, что согласуется со здравым смыслом. По-видимому следует заключить, что в первом уравнении результат теста указывал на ошибку спецификации. Пример показывает, что при эконометрическом анализе полезна любая дополнительная информация (в нашем случае — механизм формирования зарплаты).

6.2. Корреляция по времени

Авторегрессионный процесс первого порядка

При анализе временных рядов часто приходится учитывать статистическую зависимость наблюдений в разные моменты времени. Иными словами, для многих временных рядов предположение о некоррелированности ошибок не выполняется. В этом разделе мы рассмотрим наиболее простую модель, в которой ошибки образуют так называемый авторегрессионный процесс первого порядка (точное определение будет дано ниже). Как было показано ранее (глава 5), применение обычного метода наименьших квадратов к этой системе дает несмещенные и состоятельные оценки параметров, однако можно показать (см., например, Johnston and DiNardo, 1997), что получаемая при этом оценка дисперсии оказывается смещенной вниз, что может отрицательно сказаться при проверке гипотез о значимости коэффициентов. Образно говоря, МНК рисует более оптимистичную картину регрессии, чем есть на самом деле.

Как и раньше, рассмотрим модель

$$y = X\beta + \varepsilon, \tag{6.5}$$

где t-я компонента вектора y представляет значение зависимой переменной в момент времени t, $t = 1, \ldots, n$. Будем для определенности считать, что первым регрессором в X является константа. Запишем подробнее уравнение для наблюдения в момент времени t:

$$y_t = \beta_1 + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + \varepsilon_t = x_t'\beta + \varepsilon_t, \tag{6.6}$$

где $x_t' = (1, x_{t2}, \ldots, x_{tk})$ — t-я строка матрицы X.

Один из наиболее простых способов учета коррелированности ошибок (в разные моменты времени) состоит в предположении, что случайная последовательность $\{\varepsilon_t,\ t = 1, \ldots, n\}$ образует *авторегрессионный процесс первого порядка*. Это означает, что ошибки удовлетворяют рекуррентному соотношению

$$\varepsilon_t = \rho\varepsilon_{t-1} + \nu_t, \tag{6.7}$$

где $\{\nu_t,\ t = 1, \ldots, n\}$ — последовательность независимых нормально распределенных случайных величин с нулевым средним и постоянной дисперсией σ_ν^2, а ρ — некоторый параметр, называемый *коэффициентом авторегрессии* ($|\rho| < 1$). Строго говоря, для полного описания модели надо определить ε_0. Будем считать, что ε_0 — нормальная случайная величина с нулевым средним и дисперсией $\sigma_\varepsilon^2 = \sigma_\nu^2/(1-\rho^2)$, не зависящая от $\{\nu_t,\ t = 1, \ldots, n\}$. Из дальнейшего станет ясно, почему у ε_0 именно такие параметры. Взяв математическое ожидание от обеих частей (6.7), получим $\mathrm{E}\,\varepsilon_t = \rho\,\mathrm{E}\,\varepsilon_{t-1}$, откуда следует, что $\mathrm{E}\,\varepsilon_t = 0$, $t = 1, \ldots, n$. Поскольку ε_{t-1} выражается через $\nu_1, \ldots, \nu_{t-1}$ (см. (6.7)), то ε_{t-1} и ν_t независимы. Поэтому

$$\mathrm{E}\,(\varepsilon_t^2) = \mathrm{E}\,(\rho\varepsilon_{t-1} + \nu_t)^2 = \rho^2\,\mathrm{E}\,(\varepsilon_{t-1}^2) + \mathrm{E}\,\nu_t^2 = \rho^2\,\mathrm{E}\,(\varepsilon_{t-1}^2) + \sigma_\nu^2.$$

Легко проверяется, что если $\mathrm{E}\,(\varepsilon_0^2) = \sigma_\nu^2/(1-\rho^2)$, то

$$\sigma_\varepsilon^2 = \mathrm{E}\,(\varepsilon_t^2) = \mathrm{V}\,(\varepsilon_t) = \sigma_\nu^2/(1-\rho^2), \qquad t = 1, \ldots, n. \tag{6.8}$$

Умножая (6.7) на ε_{t-1} и вновь пользуясь независимостью ε_{t-1} и ν_t, получим

$$\mathrm{E}\,(\varepsilon_t\,\varepsilon_{t-1}) = \mathrm{Cov}\,(\varepsilon_t, \varepsilon_{t-1}) = \rho\,\mathrm{V}\,(\varepsilon_{t-1}) = \rho\sigma_\varepsilon^2. \tag{6.9}$$

Аналогично $\mathrm{Cov}\,(\varepsilon_t, \varepsilon_{t-2}) = \rho^2\sigma_\varepsilon^2$ и вообще

$$\mathrm{Cov}\,(\varepsilon_t, \varepsilon_{t-m}) = \rho^m\sigma_\varepsilon^2. \tag{6.10}$$

Таким образом, последовательность $\{\varepsilon_t\}$ образует стационарный[1] случайный процесс. Именно этим обстоятельством диктовался выбор параметров начальной величины ε_0. На самом деле, с течением времени зависимость ε_t от ε_0 быстро уменьшается, поэтому в большинстве книг по эконометрике проблему начальных условий для $\{\varepsilon_t\}$ просто не рассматривают, неявно подразумевая, что процесс (6.7) при любом начальном значении быстро сходится к стационарному. Отметим также, что условие $|\rho| < 1$ является необходимым для стационарности.

[1] Более подробно понятие стационарности будет рассмотрено в главе 11.

Из (6.9) следует, что

$$\rho = \mathrm{Cov}\,(\varepsilon_t, \varepsilon_{t-1})/\sigma_\varepsilon^2 = \mathrm{Cov}\,(\varepsilon_t, \varepsilon_{t-1})/\left(\mathrm{V}\,(\varepsilon_t)^{1/2}\mathrm{V}\,(\varepsilon_{t-1})^{1/2}\right),$$

т. е. ρ есть в точности коэффициент корреляции между двумя соседними ошибками. Пользуясь (6.10), можно выписать ковариационную матрицу случайного вектора $\boldsymbol{\varepsilon}$:

$$\boldsymbol{\Omega} = \frac{\sigma_\nu^2}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \dots & \dots & \dots & \dots & \dots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{bmatrix}.$$

Оценивание в модели с авторегрессией

Проблему оценивания системы (6.5) рассмотрим отдельно для случая, когда коэффициент ρ известен, и отдельно — когда неизвестен.

1. ***Значение ρ известно.*** В этом случае для оценивания системы (6.5) можно применить обобщенный метод наименьших квадратов. В данном случае нетрудно найти матрицу $\boldsymbol{P}$, для которой $\boldsymbol{P}'\boldsymbol{P} = \boldsymbol{\Omega}^{-1}$ (см. (5.5)). Здесь весьма просто догадаться, какое линейное преобразование исходной системы (6.5) надо провести, чтобы получить классическую модель. Напишем (6.6) для момента времени $t-1$ ($t \geqslant 2$)

$$y_{t-1} = \boldsymbol{x}_{t-1}'\boldsymbol{\beta} + \varepsilon_{t-1},$$

умножим обе части на ρ и вычтем почленно из (6.6). Тогда с учетом (6.7) получим

$$y_t - \rho y_{t-1} = (\boldsymbol{x}_t - \rho\boldsymbol{x}_{t-1})'\boldsymbol{\beta} + \nu_t. \tag{6.11}$$

При $t = 1$ достаточно обе части уравнения (6.6) умножить на $\sqrt{1-\rho^2}$:

$$\sqrt{1-\rho^2}\,y_1 = \sqrt{1-\rho^2}\,\boldsymbol{x}_1'\boldsymbol{\beta} + \sqrt{1-\rho^2}\,\varepsilon_1. \tag{6.12}$$

В системе (6.11), (6.12) ошибки удовлетворяют условиям уже обычной регрессионной модели. Действительно, в (6.11) случайные величины $\{\nu_t,\ t = 2, \dots, n\}$ независимы и имеют постоянную дисперсию σ_ν^2, а в (6.12) ошибка $\sqrt{1-\rho^2}\varepsilon_1$ не зависит от $\{\nu_t,\ t = 2, \dots, n\}$ и, согласно (6.8), также имеет дисперсию σ_ν^2.

На практике часто опускают преобразование (6.12), игнорируя тем самым первое наблюдение. С одной стороны, благодаря этому, преобразование исходной модели (6.5) становится единообразным. В частности, для получения оценки параметра β_1 достаточно оценку свободного члена в (6.11) разделить на $(1-\rho)$. С другой стороны, отбрасывание первого наблюдения может привести к потере важной информации, особенно в выборках небольшого размера.

2. ***Значение ρ неизвестно.*** Ситуации, когда параметр авторегрессии ρ известен, встречаются крайне редко. Поэтому возникает необходимость в процедурах оценивания при неизвестном ρ. Как правило, они имеют итеративный характер. Опишем три наиболее употребительные. Мы не будем устанавливать сходимость этих процедур, практика их применения показала, что они достаточно эффективны.

Процедура Кохрейна–Оркатта (Cochrane–Orcutt). Начальным шагом этой процедуры является применение обычного метода наименьших квадратов к исходной системе (6.5) и получение соответствующих остатков $\boldsymbol{e} = (e_1, \dots, e_n)'$. Далее,

1) в качестве приближенного значения ρ берется его МНК-оценка r в регрессии $e_t = \rho e_{t-1} + \nu_t$;

2) проводится преобразование (6.11) (или (6.11), (6.12)) при $\rho = r$, и находятся МНК-оценки $\widehat{\boldsymbol{\beta}}$ вектора параметров $\boldsymbol{\beta}$;

3) строится новый вектор остатков $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X}\,\widehat{\boldsymbol{\beta}}$;

4) процедура повторяется, начиная с п. 1).

Процесс обычно заканчивается, когда очередное приближение ρ мало отличается от предыдущего. Иногда просто фиксируется

количество итераций. Процедура Кохрейна–Оркатта реализована в большинстве эконометрических компьютерных программ.

Процедура Хилдрета–Лу (Hildreth–Lu). Суть процедуры достаточно проста. Из интервала $(-1,1)$ возможного изменения коэффициента ρ берутся последовательно некоторые значения (например, числа с постоянным шагом 0.1 или 0.05) и для каждого из них проводится оценивание преобразованной системы (6.11). Определяется то значение этого параметра, для которого сумма квадратов отклонений в (6.11) минимальна. Затем в некоторой окрестности этого значения устраивается более мелкая сетка и процесс повторяется. Итерации заканчиваются, когда будет достигнута желаемая точность. Время работы процедуры, очевидно, сокращается, если есть априорная информация об области изменения параметра ρ.

Процедура Дарбина (Durbin). Преобразованная система (6.11) переписывается в следующем виде:

$$y_t = \beta_1(1-\rho)+\rho y_{t-1}+\beta_2 x_{t2}-\rho\beta_2 x_{t-1\,2}+\cdots+\beta_k x_{tk}-\rho\beta_k x_{t-1\,k}+\nu_t,$$

т. е. y_{t-1} включается в число регрессоров, а ρ — в число оцениваемых параметров. Для этой системы строятся обычные МНК-оценки r и $\widehat{\theta}_j$ параметров ρ и $\rho\beta_j$ соответственно. В качестве оценки $\widehat{\beta}_j$ берут $\widehat{\theta}_j/r$. Можно улучшить качество оценок $\widehat{\boldsymbol{\beta}}$, подставив полученное значение r в систему (6.11), и найти новые МНК-оценки параметров $\boldsymbol{\beta}$.

Тест Дарбина–Уотсона на наличие или отсутствие корреляции по времени

Большинство тестов на наличие корреляции по времени в ошибках системы (6.5) используют следующую идею: если корреляция есть у ошибок ε, то она присутствует и в остатках $\boldsymbol{e}$, получаемых после применения к (6.5) обычного метода наименьших квадратов. Здесь мы рассмотрим только одну реализацию этого подхода. Пусть нулевая гипотеза состоит в отсутствии корреляции, т. е. H_0: $\rho = 0$. В качестве альтернативной может выступать либо просто H_1: «не H_0», либо односторонняя гипотеза, например, H_1: $\rho > 0$.

Наиболее широко используется тест Дарбина–Уотсона (*Durbin–Watson*). Он основан на статистике

$$\mathrm{DW} = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}. \tag{6.13}$$

Будем считать, что постоянный член включен в число регрессоров. Тогда нетрудно проверить, что эта статистика тесно связана с выборочным коэффициентом корреляции между e_t и e_{t-1}. Действительно, проводя элементарные выкладки, имеем

$$\begin{aligned}\mathrm{DW} &= \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2} = \frac{\sum_{t=2}^{n} e_t^2 + \sum_{t=2}^{n} e_{t-1}^2 - 2\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2}\\ &= \frac{\sum_{t=1}^{n} e_t^2 - e_1^2 + \sum_{t=1}^{n} e_t^2 - e_n^2 - 2\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2}\\ &= 2\left(1 - \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2}\right) - \frac{e_1^2 + e_n^2}{\sum_{t=1}^{n} e_t^2}.\end{aligned} \tag{6.14}$$

Предполагая число наблюдений достаточно большим, можно считать, что приближенно выполнены следующие равенства: $\frac{1}{n-1}\sum_{t=2}^{n} e_t = -e_1/(n-1) \approx 0$ и $\frac{1}{n-1}\sum_{t=2}^{n} e_{t-1} = -e_n/(n-1) \approx 0$ (поскольку выполнено точное равенство $\sum_{t=1}^{n} e_t = 0$ в силу наличия постоянного регрессора). Поэтому выборочный коэффициент корреляции r между e_t и e_{t-1} можно приближенно представить в виде

$$r \approx \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sqrt{\sum_{t=2}^{n} e_t^2 \ \sum_{t=1}^{n-1} e_t^2}} \approx \frac{\sum_{t=2}^{n} e_t e_{t-1}}{\sum_{t=1}^{n} e_t^2}.$$

Наконец, пренебрегая в (6.14) слагаемыми e_1^2 и e_n^2 по сравнению с общей суммой $\sum_{t=1}^{n} e_t^2$, окончательно получим

$$\mathrm{DW} \approx 2(1-r). \tag{6.15}$$

Понятен и содержательный смысл статистики DW: если между e_t и e_{t-1} имеется достаточно высокая положительная корреляция, то в определенном смысле e_t и e_{t-1} близки друг к другу и величина статистики DW мала. Это согласуется с (6.15): если коэффициент r близок к единице, то величина DW близка к нулю. Отсутствие корреляции означает, что DW близка к 2. Таким образом, если бы распределение статистики DW было известно, то для проверки гипотезы H_0: $\rho = 0$ против альтернативы H_1: $\rho > 0$ можно было бы для заданного уровня значимости (например, для 5%-ного уровня) найти такое критическое значение d^*, что если $DW > d^*$, то гипотеза H_0 не отвергается, в противном случае она отвергается в пользу H_1. Проблема, однако, состоит в том, что распределение DW зависит не только от числа наблюдений n и количества регрессоров k, но и от всей матрицы X, и, значит, практическое применение этой процедуры невозможно, поскольку нельзя же составить таблицу критических значений d^* для всех матриц X! Тем не менее, Дарбин и Уотсон доказали (Durbin, Watson, 1951), что существуют две границы, обычно обозначаемые d_u и d_l, $d_u > d_l$ (u = upper — верхняя, l = low — нижняя), которые зависят лишь от n, k и уровня значимости (а следовательно, могут быть затабулированы) и обладают следующим свойством: если $DW > d_u$, то $DW > d^*$ и, значит, гипотеза H_0 не отвергается, а если $DW < d_l$, то $DW < d^*$, и гипотеза H_0 отвергается в пользу H_1. В случае $d_l < DW < d_u$ ситуация неопределенна, т. е. нельзя высказаться в пользу той или иной гипотезы. Если альтернативной является гипотеза об отрицательной корреляции H_1: $\rho < 0$, то соответствующими верхними и нижними границами будут $4 - d_l$ и $4 - d_u$. Целесообразно представить эти результаты в виде таблицы (см. таблицу 6.5).

Наличие зоны неопределенности, конечно, представляет определенные трудности при использовании теста Дарбина–Уотсона. Ее ширина может быть довольно значительной. К примеру, при $n = 19$, $k = 3$ она образует интервал (0.97, 1.68). Отметим, что некоторые компьютерные пакеты, например *SHAZAM*, численно вычисляют точные критические значения (зависящие от X).

Таблица 6.5

Значение статистики DW	Вывод
$4 - d_l < DW < 4$	Гипотеза H_0 отвергается, есть отрицательная корреляция
$4 - d_u < DW < 4 - d_l$	Неопределенность
$d_u < DW < 4 - d_u$	Гипотеза H_0 не отвергается
$d_l < DW < d_u$	Неопределенность
$0 < DW < d_l$	Гипотеза H_0 отвергается, есть положительная корреляция

Сделаем еще одно важное замечание. Тест Дарбина–Уотсона построен в предположении, что регрессоры X и ошибки ε не коррелированы. Поэтому его нельзя применять, в частности, когда среди регрессоров содержатся лагированные значения зависимой переменной y.

Выводы:

1) при анализе временных рядов следует учитывать, что, как правило, ошибки коррелированы во времени, что требует коррекции обычного метода наименьших квадратов;

2) во многих случаях можно считать, что ошибки образуют стационарный авторегрессионный процесс первого порядка (6.7);

3) МНК-оценки в случае авторегрессии первого порядка несмещены, состоятельны, но неэффективны;

4) оценка дисперсии при использовании МНК является заниженной;

5) если коэффициент авторегрессии известен, то обобщенный метод наименьших квадратов сводится к преобразованию (6.11), (6.12) исходной системы и дальнейшему применению МНК;

6) при неизвестном коэффициенте авторегрессии существует несколько процедур доступного обобщенного метода наименьших квадратов, суть которых состоит в оценивании этого коэффициента, а затем в применении преобразования (6.11), (6.12);

7) одним из наиболее распространенных тестов проверки гипотезы об отсутствии корреляции является тест Дарбина–Уотсона, основанный на статистике DW (6.13). Его особенность заключается в наличии зоны неопределенности для DW, когда нет оснований ни принимать, ни отвергать гипотезу об отсутствии корреляции.

Упражнения

6.1. Проверьте непосредственно, что для парной регрессии (п. 2.3) с гетероскедастичностью дисперсия оценки параметра b, полученная с помощью метода взвешенных наименьших квадратов, меньше дисперсии МНК-оценки.

6.2. Процесс, порождающий данные, описывается уравнением

$$y_t = \beta x_t + \varepsilon_t,$$
$$\mathrm{E}\,\varepsilon_t = 0, \quad \mathrm{E}\,(\varepsilon_t^2) = \sigma^2, \quad \mathrm{E}\,(\varepsilon_t \varepsilon_s) = 0, \quad t \neq s, \quad t = 1, \ldots, n.$$

Экспериментатор не имеет доступа к исходным данным, а может использовать лишь «групповые» данные. А именно, значения независимой переменной упорядочиваются по величине ($x_1 < x_2 < \cdots < x_n$), вычисляются средние значения в первой группе из n_1 наблюдений

$$\overline{x}_1 = \frac{1}{n_1}\sum_{t=1}^{n_1} x_t, \quad \overline{y}_1 = \frac{1}{n_1}\sum_{t=1}^{n_1} y_t,$$

во второй группе из n_2 наблюдений

$$\overline{x}_2 = \frac{1}{n_2}\sum_{t=n_1+1}^{n_1+n_2} x_t, \quad \overline{y}_2 = \frac{1}{n_2}\sum_{t=n_1+1}^{n_1+n_2} y_t,$$

и т. д. Всего есть J групп наблюдений, j-я группа имеет объем n_j. Параметр β оценивается с помощью регрессии $\overline{y}_j$ на $\overline{x}_j$, $j = 1, \ldots, J$. Вычислите среднее значение и дисперсию оценки. Оцените потерю эффективности в результате такой группировки данных.

6.3. Рассмотрим уравнение $y_t = \alpha + \beta x_t + \varepsilon_t$, где ошибки ε_t порождаются авторегрессионным процессом второго порядка:

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + u_t.$$

Предложите обобщение итеративной процедуры Кохрейна–Оркатта для оценивания параметров этой модели.

6.4. Рассмотрим модель $y_t = \beta x_t + \varepsilon_t$, где

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t, \quad \mathrm{E}\,u_t = 0, \quad \mathrm{E}\,u_t^2 = \sigma^2, \quad \mathrm{E}\,u_t u_s = 0, \qquad t \neq s.$$

Предлагается оценивать параметр β с помощью регрессии первой разности $\Delta y_t = y_t - y_{t-1}$ на Δx_t.

а) Покажите, что эта оценка является линейной и несмещенной.

б) Вычислите дисперсию оценки и покажите, что стандартная оценка этой дисперсии смещена.

6.5. Предположим, что для системы $y_t = \alpha + \beta x_t + \varepsilon_t$, $t = 1, \ldots, n$, выполнены все предположения классической нормальной модели за одним исключением: дисперсии ошибок удовлетворяют соотношениям

$$\sigma_t^2 = \mu + \delta x_t.$$

Предложите двухшаговую процедуру оценивания параметров α и β.

6.6. Рассмотрим модель, связывающую количество вакансий w_t и уровень безработицы u_t:

$$\ln w_t = \beta_1 + \beta_2 \ln u_t + \varepsilon_t.$$

Ошибки ε_t независимы и нормально распределены $N(0, \sigma_\varepsilon^2)$.

а) Используя (искусственные) данные из таблицы 6.6, найдите МНК-оценки параметров β_1 и β_2, а также 95%-ный доверительный интервал для β_2.

б) Вычислите статистику Дарбина–Уотсона. Что ее значение говорит об исходном предположении об ошибках ε_t? Что можно сказать о доверительном интервале, найденном в а)?

в) Оцените модель заново, используя модель автокорреляции первого порядка для ошибок регрессии. Найдите 95%-ный доверительный интервал для β_2. Сравните результат с интервалом, полученным в а).

Таблица 6.6

t	w_t	u_t	t	w_t	u_t	t	w_t	u_t
1	1.73	8.65	9	5.06	2.87	17	3.15	4.72
2	1.94	4.82	10	2.81	5.29	18	1.92	7.45
3	3.05	2.67	11	4.43	3.31	19	2.26	6.21
4	4.17	2.67	12	3.19	5.44	20	6.18	2.64
5	2.52	2.58	13	2.23	6.80	21	2.07	8.55
6	1.71	8.07	14	2.06	8.25	22	8.39	2.60
7	1.95	8.83	15	3.33	3.44	23	2.75	6.25
8	2.57	5.54	16	2.12	7.80	24	6.10	2.70

6.7. В таблице 6.7 представлены данные о потребительских расходах C и располагаемом доходе Y_d тридцати семей (долл.).

Таблица 6.7

Потребление			Доход
10700	10900	11200	12000
11400	11700	12100	13000
12300	12600	13200	14000
13000	13300	13600	15000
13800	14000	14200	16000
14400	14900	15300	17000
15000	15700	16400	18000
15900	16500	16900	19000
16900	17500	18100	20000
17200	17800	18500	21000

а) Проведите регрессию C на Y_d и проверьте наличие или отсутствие гетероскедастичности.

б) Если в а) выявлена гетероскедастичность, осуществите коррекцию на гетероскедастичность.

6.8. Таблица 6.8 содержит данные об уровнях запасов I, объемов продаж S (млн. долл.) и процентные ставки по кредитам R в 35 фирмах некоторой отрасли. Экономическая интуиция подсказывает, что I должно быть положительно связано с S и отрицательно с R.

а) Проведите регрессию I на S и R и тест на гетероскедастичность.

б) Если в а) выявлена гетероскедастичность, осуществите коррекцию на гетероскедастичность, предполагая, что дисперсия ошибки пропорциональна S^2.

Таблица 6.8

Фирма	I	S	R	Фирма	I	S	R
1	10	100	17	19	15	122	11
2	10	101	17	20	15	123	11
3	10	103	17	21	15	125	11
4	11	105	16	22	16	128	10
5	11	106	16	23	16	128	10
6	11	106	16	24	16	131	10
7	12	108	15	25	17	133	10
8	12	109	15	26	17	134	9
9	12	111	14	27	17	135	9
10	12	111	14	28	17	136	9
11	12	112	14	29	18	139	8
12	13	113	14	30	18	143	8
13	13	114	13	31	19	147	8
14	13	114	13	32	19	151	8
15	14	116	12	33	19	157	8
16	14	117	12	34	20	163	7
17	14	118	12	35	20	171	7
18	15	120	11				

6.9. В таблице 6.9 приведены данные об объеме импорта M и ВНП США (млрд. долл.) за период с 1960 по 1979 г.

Таблица 6.9

Год	M	ВНП	Год	M	ВНП
1960	23.2	506.0	1970	58.5	982.4
1961	23.1	523.3	1971	64.0	1063.4
1962	25.2	563.8	1972	75.9	1171.1
1963	26.4	594.7	1973	94.4	1306.6
1964	28.4	635.7	1974	131.9	1424.9
1965	32.0	688.1	1975	126.9	1528.1
1966	37.7	753.0	1976	155.4	1702.2
1967	40.6	796.3	1977	185.8	1899.5
1968	47.7	868.5	1978	217.5	2127.6
1969	52.9	935.5	1979	260.9	2368.5

Источник: *D.Salvatore. Statistics and Econometrics*, McGraw-Hill, 1982.

а) Проведите регрессию M на ВНП и на 5%-ном уровне значимости протестируйте гипотезу об отсутствии автокорреляции ошибок.

б) Если в а) гипотеза отвергается, проведите коррекцию на автокорреляцию.

6.10. Рассмотрим модель $y_t = \beta x_t + \varepsilon_t$, $t = 1, 2, ..., n$, где β, x_t — скаляры, и предположим, что $\lim_{n\to\infty}(1/n)\sum_{t=1}^{n} x_t^2 = \sigma_x^2$, $\lim_{n\to\infty}(1/n)\sum_{t=s+1}^{n} x_t x_{t-s} = \alpha^s \sigma_x^2$, где $|\alpha| \le 1$ и $\mathrm{E}(\varepsilon_t) = 0$, $\mathrm{E}(\varepsilon_t \varepsilon_{t-s}) = \rho^s \sigma_\varepsilon^2$ при всех s, где $|\rho| < 1$.

а) Вычислите асимптотическую дисперсию МНК-оценки параметра β, т.е. $\lim_{n\to\infty} n\mathrm{V}(\widehat{\beta}_{\mathrm{OLS}})$.

б) Вычислите асимптотическую дисперсию оценки параметра β, полученной с помощью обобщенного метода наименьших квадратов, и покажите что при $\alpha = \rho$ асимптотическая эффективность МНК-оценки равна $(1-\rho^2)/(1+\rho^2)$.

6.11. Рассмотрим модель $y_t = \beta x_t + \varepsilon_t$, $t = 1, \ldots, n$, где $\mathrm{E}(\varepsilon_t) = 0$, $\mathrm{E}(\varepsilon_t^2) = a x_t^2$, $\mathrm{E}(\varepsilon_t \varepsilon_s) = 0$ при $t \ne s$ и $\sum_{t=1}^{n} x_t^2 = n$.

а) Покажите, что МНК-оценка $\widehat{\beta}$ параметра β является несмещенной, но неэффективной.

б) Покажите, что стандартная оценка дисперсии $\widehat{\beta}$ смещена вниз по отношению к истинной дисперсии $\widehat{\beta}$.

6.12. Уравнение $y_t = \beta_1 + \beta_2 x_t + \varepsilon_t$ оценивается по следующим наблюдениям (см. таблицу 6.10):

Таблица 6.10

x	4	1	5	8	2
y	6	3	12	15	4

Известна функциональная форма гетероскедастичности: $\mathrm{V}(\varepsilon_t) = \sigma_t^2 = \sigma^2 x_t^2$. Вычислить оценки обобщенного метода наименьших квадратов параметров β_1, β_2 и их стандартные отклонения.

6.13. В модели $y_t = \beta x_t + \varepsilon_t$ (β, x_t — скаляры и $x_t > 0$) ошибки ε_t образуют авторегрессию первого порядка: $\varepsilon_t = \rho\varepsilon_{t-1} + u_t$, $0 < \rho < 1$. Покажите, что стандартная оценка дисперсии ε_t, полученная с помощью обычного метода наименьших квадратов, смещена вниз.

6.14. Расходы домашних хозяйств в Нидерландах (см. продолжение в упражнении 12.15).

Введение. Традиционной задачей эмпирических исследований в микроэкономике является оценивание кривых Энгеля. Эрнст Энгель установил, что при увеличении дохода семьи доля расходов на питание уменьшается (закон Энгеля). В современных микроэкономических терминах это означает, что эластичность расходов на питание по доходу меньше единицы. (При этом говорят также, что еда является необходимым

товаром, а не предметом роскоши.) Зависимость расходов на приобретение некоторого вида товара от доходов называется кривой Энгеля. В настоящее время принято, как правило, вместо дохода рассматривать полные расходы.

В этой серии упражнений мы будем изучать ежегодные расходы домашних хозяйств (*household*) на питание в зависимости от полных ежегодных расходов и некоторых других переменных на основании данных по расходам семей в Нидерландах.

Данные. Используются данные, полученные из архива журнала *Journal of Applied Econometrics* (`expend.xls`). Для наших целей мы взяли данные о годовых расходах на питание, отдых и другие товары за период с октября 1986 г. по сентябрь 1987 г. (427 наблюдений). Список переменных содержится в таблице 6.11.

Таблица 6.11

Переменная	Описание
f3	Расходы на питание одной семьи с октября 1986 г. по сентябрь 1987 г. в голландских гульденах
v3	Расходы на отдых с октября 1986 г. по сентябрь 1987 г. в голландских гульденах
tot3	Полные расходы с октября 1986 г. по сентябрь 1987 г. в голландских гульденах
prov	Провинция
reg	Регион
scl	Социальный класс (1 — нижний класс, ..., 5 — верхний класс)
nahm	Число членов семьи старше 11 лет
durb	Степень урбанизации (1 — маленькая деревня, ..., 13 — большой город)
nch06	Число детей младше 6 лет
nch711	Число детей от 7 до 11 лет
nch1217	Число детей от 12 до 17 лет
nch18	Число детей старше 18 лет

6.14.1. а) Вычислите суммарные статистики всех переменных. Проверьте, имеют ли смысл ваши результаты.

б) Вычислите корреляционную матрицу переменных *f3*, *v3*, *tot3* и *nahm*. Проинтерпретируйте результат. Соответствует ли он вашим ожиданиям?

Расходы на питание. Здесь мы изучим линейную модель для объяснения логарифма расходов на питание: $lf3 = \ln(f3)$. Мы рассмотрим также некоторую модель для объяснения доли расходов на питание.

6.14.2. а) Проведите парную регрессию $lf3$ на $ltot3 = \ln(tot3)$. С ее помощью постройте 95%-ный двусторонний доверительный интервал для эластичности по доходу.

б) В регрессии п. а) не принимается во внимание размер семьи. Объясните, почему это может привести к смещенности оценки. Объясните, почему таким образом вы, возможно, переоцениваете эластичность по доходу.

6.14.3. а) Проведите регрессию $lf3$ на $ltot3$, $nahm$, $nch06$, $nch711$ и константу.

б) С ее помощью постройте 95%-ный двусторонний доверительный интервал для эластичности по доходу. Сравните ваш результат с результатом упражнения 6.14.2 а).

в) Проинтерпретируйте эффект включения в регрессию состава семьи.

г) Проверьте, отличается ли влияние детей в возрасте до 6 лет от влияния детей в возрасте от 7 до 11 лет.

д) Проверьте, зависит ли каким-нибудь образом влияние наличия детей от их возраста.

6.14.4. а) Добавьте переменную $ltot3s = ltot3 \cdot ltot3$ в правую часть регрессионного уравнения упражнения 6.14.3 и проведите новую регрессию. Является ли переменная $ltot3s$ значимой? Что она означает?

б) Воспользуйтесь полученными результатами для оценивания эластичности при различных уровнях дохода $tot3$.

в) Воспользуйтесь полученными результатами и/или дополнительными вычислениями для построения доверительного интервала для эластичности при $tot3 = 36000$.

6.14.5. а) Постройте фиктивные переменные для каждой из 13 провинций. Проведите регрессию $lf3$ на переменные, включенные в упражнении 6.14.2, и на двенадцать из тринадцати фиктивных переменных. Почему не следует включать все 13 фиктивных переменных? Проверьте (на 95%-ном доверительном уровне) совместную значимость эффекта провинции.

б) Проведите аналогичную процедуру с заменой провинций на социальные классы.

6.14.6. а) Воспользуйтесь стратегией «от общего к частному» для построения наиболее подходящей модели, объясняющей величину $lf3$. Проинтерпретируйте ваш результат. Используйте выбранную вами модель в последующих упражнениях.

б) Оцените эластичность по доходу для различных уровней переменной $tot3$. Является ли еда необходимым товаром или предметом роскоши?

в) Постройте доверительный интервал для эластичности по доходу при $tot3 = 36000$.

г) Постройте двусторонний 95%-ный доверительный интервал для прогнозного значения расходов на питание бездетной семьи из двух человек, принадлежащей среднему классу и проживающей в Амстердаме, если ее общие расходы составляют 50000 гульденов.

д) Нарисуйте график зависимости остатков регрессии от $ltot3$. Какой вывод вы можете сделать относительно предположения о независимости ошибок и регрессора $ltot3$?

6.14.7. а) Постройте переменную $sf3 = 100 \cdot f3/tot3$ (доля расходов на питание в общем бюджете, в %). Следуя процедуре упражнения 6.14.6, постройте наиболее подходящую модель для объяснения $sf3$.

б) Используя результаты п. а), оцените эластичность по доходу для различных значений $tot3$. Сравните с результатами, полученными в упражнении 6.14.6.

в) Нарисуйте график зависимости остатков регрессии от $ltot3$. Какой вывод вы можете сделать?

6.14.8. Долю расходов на питание в общем бюджете можно рассматривать как отрицательный показатель благосостояния: более высокое значение этой доли соответствует более низкому благосостоянию семьи. Тогда «стоимость детей» можно измерить, ответив на следующий вопрос: какой дополнительный доход требуется семье с каждым новым ребенком, чтобы остаться на том же уровне благосостояния, т. е. чтобы иметь ту же долю расходов на питание в общем бюджете?

а) Воспользуйтесь результатами, полученными в упражнении 6.14.7, чтобы оценить стоимость одного ребенка в возрасте 12 лет, беря в качестве отправной точки бездетную семью из двух человек.

б) Вычислите стоимость (первого) ребенка в каждой возрастной группе. Сравните с п. а). Проинтерпретируйте результаты.

Расходы на питание и гетероскедастичность. В предыдущих упражнениях мы предполагали, что ошибки *гомоскедастичны*. Сейчас мы попытаемся ответить на вопрос, является ли эта гипотеза приемлемой. В тех случаях, когда ее целесообразно отвергнуть, мы будем исследовать модель с учетом *гетероскедастичности*.

6.14.9. В качестве отправной точки используется модель упражнения 6.14.6.

а) Вычислите остатки e в этой модели и постройте переменную $\ln(e)$.

б) Проведите регрессию $\ln(e)$ на все независимые переменные модели упражнения 6.14.6.

в) Проверьте совместную значимость всех переменных в б).

г) Объясните, почему этот тест можно рассматривать как тест на наличие экспоненциальной гетероскедастичности ($\sigma^2 = \exp(\boldsymbol{x}'\boldsymbol{\alpha})$), и проинтерпретируйте результат.

6.14.10. В качестве отправной точки используется модель упражнения 6.14.7.

а) С помощью теста Бреуша–Пагана проверьте гипотезу о гетероскедастичности вида $\sigma = f(\boldsymbol{x}'\boldsymbol{\alpha})$ с неизвестной функцией f.

б) С помощью теста Голдфилда–Квандта проверьте гипотезу о наличии гетероскедастичности типа «σ возрастает с ростом *ltot3*».

в) Проверьте гипотезу о наличии экспоненциальной гетероскедастичности, следуя той же схеме, что и в упражнении 6.14.9.

6.14.11. В качестве отправной точки используется модель упражнения 6.14.7, но теперь допускается наличие экспоненциальной гетероскедастичности ($\sigma = \exp(\boldsymbol{z}'\boldsymbol{\alpha})$)

а) Основываясь на результатах упражнения 6.14.10 в), выберите в качестве $\boldsymbol{z}$ подходящий подвектор вектора $\boldsymbol{x}$.

б) Оцените $\boldsymbol{\alpha}$.

в) Примените метод взвешенных наименьших квадратов с весом $\exp(-\boldsymbol{z}'\widehat{\boldsymbol{\alpha}})$.

г) Сравните результаты с результатами упражнения 6.14.7.

6.15. Рассматриваются следующие данные из газеты «Из рук в руки» за период с декабря 1996 г. по сентябрь 1997 г., касающиеся стоимости однокомнатных квартир в юго-западной части Москвы. Данные содержатся в файле `room1.xls`. В таблице 6.12 приведено описание переменных.

Таблица 6.12

Переменная	Описание
n	Номер по порядку
distc	Удаленность от центра, км
distm	Удаленность от метро, мин
totsq	Общая площадь квартиры, кв.м
kitsq	Площадь кухни, кв.м
livsq	Площадь комнаты, кв.м
floor	Этаж, 0 — первый или последний, 1 — нет
cat	Категория дома, 1 — кирпичный, 0 — нет
date	Дата рекламного объявления
price	Цена квартиры, тыс. долл.

а) Найдите среднее, стандартное отклонение и другие выборочные статистики переменных. Найдите коэффициенты корреляции переменных с ценой квартиры. Соответствуют ли полученные значения экономической интуиции?

б) Исследуйте значимость влияния различных факторов на цену квартиры. (Вы можете брать в качестве зависимой переменной цену квартиры, цену квадратного метра общей площади или их логарифмы.)

в) Есть ли существенная зависимость цены квартиры от расстояния до центра? От расстояния до метро? Как интерпретировать результаты?

г) Подберите модель, которая наилучшим способом прогнозирует цену квартиры по имеющимся данным. Проверьте наличие гетероскедастичности.

6.16. Выборка состоит из 70 объявлений о продаже двухкомнатных квартир из газеты «Недвижимость» за сентябрь 1997 г. Были отобраны квартиры в окраинных районах Москвы (новостройки). Данные находятся в файле `room2.xls`, таблица 6.13 содержит описание переменных.

а) Найдите среднее, стандартное отклонение и другие выборочные статистики переменных. Найдите коэффициенты корреляции переменных с ценой квартиры. Соответствуют ли полученные значения экономической интуиции?

б) Исследуйте значимость влияния различных факторов на цену квартиры. (Вы можете брать в качестве зависимой переменной

Таблица 6.13

Переменная	Описание
n	Номер по порядку
$price$	Цена квартиры, тыс. долл.
$totsq$	Общая площадь квартиры, кв.м
$livsq$	Жилая площадь квартиры, кв.м
$kitsq$	Площадь кухни, кв.м
$distm$	Расстояние пешком до метро, мин
$floor$	Этаж, 0 — первый или последний, 1 — нет
cat	Категория дома, 1 — кирпичный, 0 — нет
tel	Телефон, 1 — есть, 0 — нет
$lift$	Лифт, 1 — есть, 0 — нет
$balc$	Балкон, 1 — есть, 0 — нет

цену квартиры, цену квадратного метра общей площади или их логарифмы.)

в) Есть ли существенная зависимость цены квартиры от расстояния до метро? От наличия телефона? лифта? Как интерпретировать результаты?

г) Что «стоит дороже»: квадратный метр кухни, коридора или комнаты?

д) Подберите модель, которая наилучшим способом прогнозирует цену квартиры по имеющимся данным. Проверьте наличие гетероскедастичности.

6.17. (Arthur van Soest, Tilburg University) Файл `wages.xls` содержит данные о 75 мужчинах и 75 женщинах, работавших на полную ставку (не менее 4 дней в неделю в 1987 г.). Данные получены на основании опроса. В таблице 6.14 приведено описание переменных.

Вопросы для обсуждения

- Верно ли, что зарплата мужчин выше, чем зарплата женщин? Если да, то может ли это быть объяснено разницей в возрасте или образовании?
- Какова отдача от образования?
- Одинакова ли зависимость зарплаты от возраста для мужчин и женщин?

Таблица 6.14

Переменная	Описание
W	Зарплата, гульденов/час до вычета налогов (1987 г.)
AGE	Возраст, лет
SEX	Пол, 1 — для мужчин, 2 — для женщин
EDU	Уровень образования,
	1 — начальная школа или менее;
	2 — низшее ремесленное;
	3 — среднее;
	4 — высшее ремесленное;
	5 — университет

а) Вычислите описательные статистики. Постройте матрицу корреляций.

б) Создайте переменную $S = SEX - 1$. Обсудите регрессию $W = \beta_0 + \beta_1 S + \beta_2 AGE$. Проделайте тест на гетероскедастичность. Получите оценку Уайта стандартных отклонений коэффициентов в МНК-оценивании. Проделайте двухшаговую процедуру коррекции на гетероскедастичность.

в) Обсудите регрессию $W = \beta_0 + \beta_1 S + \beta_2 AGE + \beta_3 EDU$. Что можно сказать о коэффициенте при S в этой и предыдущей регрессиях? Насколько реалистична эта модель?

г) Обсудите регрессию $W = \beta_0 + \beta_1 S + \beta_2 AGE + \beta_3 EDU + \beta_5 AGE^2$. Что будет, если взять полулогарифмическую модель? При каком возрасте зарплата наибольшая? Зависит ли этот возраст от уровня образования? Как интерпретировать коэффициент при S в предыдущих регрессиях?

Глава 7

Прогнозирование в регрессионных моделях

Одна из важнейших целей моделирования заключается в прогнозировании поведения исследуемого объекта. Обычно термин «прогнозирование» используется в тех ситуациях, когда требуется предсказать состояние системы в будущем. Для регрессионных моделей он имеет, однако, более широкое значение. Как уже отмечалось, данные могут не иметь временной структуры, но и в этих случаях вполне может возникнуть задача оценить значение зависимой переменной для некоторого набора независимых, объясняющих переменных, которых нет в исходных наблюдениях. Именно в этом смысле — как построение оценки зависимой переменной — и следует понимать прогнозирование в эконометрике.

Проблема прогнозирования имеет много различных аспектов. Можно различать точечное и интервальное прогнозирование. В первом случае оценка — это конкретное число, во втором — интервал, в котором истинное значение переменной находится с заданным уровнем доверия. Выделяют также безусловное и условное прогнозирование в зависимости от того, известны ли интересующие нас объясняющие переменные точно или приближенно. Кроме того, для временных рядов при нахождении прогноза существенно наличие или отсутствие корреляции по времени между ошибками.

Рассмотрим вначале классическую регрессионную модель

$$y = X\beta + \varepsilon, \tag{7.1}$$

где, как и раньше, y — $n \times 1$ вектор зависимых переменных, X — $n \times k$ матрица независимых переменных, ε — $n \times 1$ вектор ошибок, β — $k \times 1$ вектор параметров, $\mathrm{E}\varepsilon = 0$, $\mathrm{V}(\varepsilon) = \sigma^2 I$. Предположим теперь, что есть еще один набор $x_{n+1} = (x_{n+1,1}, \ldots, x_{n+1,k})'$ объясняющих переменных и известно, что соответствующая зависимая переменная удовлетворяет модели (7.1), т. е.

$$y_{n+1} = x'_{n+1}\beta + \varepsilon_{n+1}, \tag{7.2}$$

где $\mathrm{E}\varepsilon_{n+1} = 0$, $\mathrm{V}(\varepsilon_{n+1}) = \sigma^2$, и случайная величина ε_{n+1} не коррелирована с ε. Требуется по (y, X, x_{n+1}) оценить y_{n+1}. Подчеркнем, что в данном случае надо построить оценку не параметра, а случайной величины.

7.1. Безусловное прогнозирование

Термин *безусловное прогнозирование* означает, что вектор независимых переменных x_{n+1} известен точно.

Предположим, что мы знаем значения параметров β и σ^2. Тогда естественно в качестве оценки $\widehat{y}_{n+1} = \widehat{y}$ величины y_{n+1} взять $\mathrm{E}(y_{n+1}) = x'_{n+1}\beta$. Среднеквадратичная ошибка такого прогноза есть $\mathrm{E}(y_{n+1} - \widehat{y})^2 = \mathrm{E}(\varepsilon^2_{n+1}) = \sigma^2$. Если дополнительно предположить, что ε_{n+1} имеет нормальное распределение, то $(\widehat{y} - \sigma z_{\alpha/2}, \widehat{y} + \sigma z_{\alpha/2})$ есть интервал, в котором y_{n+1} находится с вероятностью $1-\alpha$, где $z_{\alpha/2}$ есть $100(\alpha/2)$%-ная точка стандартного нормального распределения.

Предположим теперь, что параметры β и σ^2 неизвестны, что, как правило, и бывает на практике. Обозначим $\widehat{\beta}$ и s^2 их МНК-оценки на основании модели (7.1): $\widehat{\beta} = (X'X)^{-1}X'y$, $s^2 = e'e/(n-k)$. Возьмем в качестве оценки y_{n+1} величину

$$\widehat{y} = x'_{n+1}\widehat{\beta}. \tag{7.3}$$

Нетрудно проверить, что поскольку $\mathrm{E}\widehat{\beta} = \beta$, то $\mathrm{E}\widehat{y} = \mathrm{E}y_{n+1}$, т. е. оценка $\widehat{y}$ является несмещенной. Оказывается, в классе линейных (по y) несмещенных оценок она обладает наименьшей среднеквадратичной ошибкой.

Теорема. *Пусть $\widetilde{y} = c'y$ — оценка величины y_{n+1}, где $c = (c_1, \ldots, c_n)'$ — некоторый вектор, и пусть оценка $\widetilde{y}$ несмещенная,*

$$\mathrm{E}\widetilde{y} = \mathrm{E}y_{n+1} = x'_{n+1}\beta. \tag{7.4}$$

Тогда

$$\mathrm{E}(\widetilde{y} - y_{n+1})^2 \geqslant \mathrm{E}(\widehat{y} - y_{n+1})^2.$$

Доказательство. Так как в силу (7.4) $\mathrm{E}\widetilde{y} = c'X\beta = x'_{n+1}\beta$ при любом β, то

$$c'X = x'_{n+1}. \tag{7.5}$$

Далее,

$$\begin{aligned}\mathrm{E}(\widetilde{y} - y_{n+1})^2 &= \mathrm{E}(\widetilde{y} - \widehat{y} + \widehat{y} - y_{n+1})^2 = \mathrm{E}(\widetilde{y} - \widehat{y})^2 \\ &\quad + \mathrm{E}(\widehat{y} - y_{n+1})^2 + 2\mathrm{E}\big((\widetilde{y} - \widehat{y})(\widehat{y} - y_{n+1})\big).\end{aligned} \tag{7.6}$$

Покажем, что

$$\mathrm{E}\big((\widetilde{y} - \widehat{y})(\widehat{y} - y_{n+1})\big) = 0. \tag{7.7}$$

Имеем

$$\begin{aligned}\mathrm{E}\big((\widetilde{y} - \widehat{y})(\widehat{y} - y_{n+1})\big) &= \mathrm{E}(c'yx'_{n+1}\widehat{\beta}) - \mathrm{E}(x'_{n+1}\widehat{\beta}x'_{n+1}\widehat{\beta}) \\ &\quad - \mathrm{E}\big(c'y(x'_{n+1}\beta + \varepsilon_{n+1})\big) \\ &\quad + \mathrm{E}\big(x'_{n+1}\widehat{\beta}(x'_{n+1}\beta + \varepsilon_{n+1})\big).\end{aligned}$$

Первое слагаемое:

$$\begin{aligned}\mathrm{E}(c'yx'_{n+1}\widehat{\beta}) &= \mathrm{E}(c'y\widehat{\beta}'x_{n+1}) \\ &= \mathrm{E}\big(c'yy'X(X'X)^{-1}x_{n+1}\big) \\ &= c'\mathrm{E}(yy')X(X'X)^{-1}x_{n+1} \\ &= c'(\sigma^2 I + X\beta\beta'X')X(X'X)^{-1}x_{n+1} \\ &= \sigma^2 c'X(X'X)^{-1}x_{n+1} \\ &\quad + c'X\beta\beta'X'X(X'X)^{-1}x_{n+1} \\ \text{(в силу (7.5))} \qquad &= \sigma^2 x'_{n+1}(X'X)^{-1}x_{n+1} + x'_{n+1}\beta\beta'x_{n+1}.\end{aligned}$$

Второе слагаемое:

$$\begin{aligned}\mathrm{E}(x'_{n+1}\widehat{\beta}x'_{n+1}\widehat{\beta}) &= x'_{n+1}\mathrm{E}(\widehat{\beta}\widehat{\beta}')x_{n+1} \\ &= x'_{n+1}(\sigma^2(X'X)^{-1} + \beta\beta')x_{n+1} \\ &= \sigma^2 x'_{n+1}(X'X)^{-1}x_{n+1} + x'_{n+1}\beta\beta'x_{n+1}.\end{aligned}$$

Третье слагаемое:

$$\begin{aligned}\mathrm{E}\big(c'y(x'_{n+1}\beta + \varepsilon_{n+1})\big) &= c'\mathrm{E}(y)\beta'x_{n+1} \\ &= c'X\beta\beta'x_{n+1} \\ \text{(в силу (7.5))} \qquad &= x'_{n+1}\beta\beta'x_{n+1}.\end{aligned}$$

Четвертое слагаемое:

$$\mathrm{E}\big(x'_{n+1}\widehat{\beta}(x'_{n+1}\beta + \varepsilon_{n+1})\big) = x'_{n+1}\beta\beta'x_{n+1}.$$

(Мы постоянно пользуемся тем, что для векторов x и y одинаковой размерности $x'y = y'x$.)
Таким образом, выполнено (7.7), и теорема доказана.

Нетрудно проверить, что среднеквадратичная ошибка прогноза есть

$$\mathrm{E}(\widehat{y} - y_{n+1})^2 = \sigma^2(1 + x'_{n+1}(X'X)^{-1}x_{n+1}). \tag{7.8}$$

Заменим σ^2 на ее оценку s^2 и обозначим

$$\delta = \sqrt{s^2(1 + x'_{n+1}(X'X)^{-1}x_{n+1})}.$$

Используя те же аргументы, что и в п. 3.5, получаем, что если ошибки $(\varepsilon, \varepsilon_{n+1})$ имеют совместное нормальное распределение, то случайная величина $(\widehat{y} - y_{n+1})/\delta$ имеет распределение Стьюдента с $n-k$ степенями свободы. Поэтому доверительным интервалом для y_{n+1} с уровнем доверия $1-\alpha$ будет интервал $(\widehat{y} - \delta t_{\alpha/2}, \widehat{y} + \delta t_{\alpha/2})$, где $t_{\alpha/2}$ есть $100(\alpha/2)$%-ная точка распределения Стьюдента с $n-k$ степенями свободы.

Можно показать, что в случае парной регрессии, т. е. когда система (7.1) имеет вид

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t, \qquad t = 1, \ldots, n,$$

формула (7.8) выглядит так:

$$\mathrm{E}(\widehat{y} - y_{n+1})^2 = \sigma^2\left(1 + \frac{1}{n} + \frac{(x_{n+1} - \overline{x})^2}{\sum(x_t - \overline{x})^2}\right), \tag{7.9}$$

где $\overline{x} = \frac{1}{n}\sum x_t$. Из (7.9) следует, что среднеквадратичная ошибка прогноза минимальна при $x_{n+1} = \overline{x}$, и чем дальше x_{n+1} от $\overline{x}$, тем шире соответствующий доверительный интервал (см. рис. 7.1).

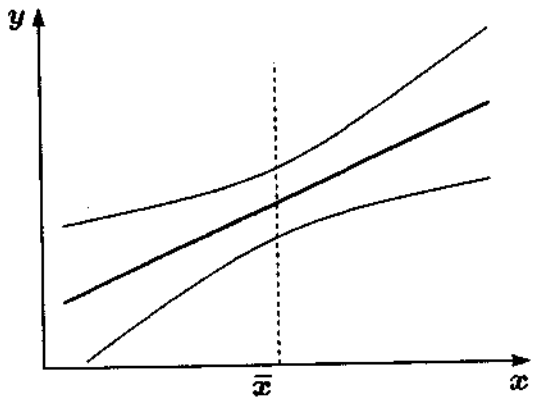


Рис. 7.1

7.2. Условное прогнозирование

В предыдущих рассуждениях мы предполагали, что независимая переменная x_{n+1} известна точно. Однако на практике встречаются ситуации, когда в x_{n+1} содержатся ошибки. Так, при прогнозировании временных рядов часто приходится прогнозировать значения независимых переменных, что неизбежно приводит к отклонениям от истинных значений. Поэтому рассмотрим теперь задачу *условного прогнозирования*. Пусть выполнены соотношения (7.1) и (7.2), но вектор $\boldsymbol{x}_{n+1}$ наблюдается с ошибкой

$$\boldsymbol{z} = \boldsymbol{x}_{n+1} + \boldsymbol{u}, \tag{7.10}$$

где $\boldsymbol{u}$ — $k\times 1$ случайный вектор, не зависящий от $(\boldsymbol{\varepsilon}, \varepsilon_{n+1})$, $\mathrm{E}\boldsymbol{u} = \boldsymbol{0}$, $\mathrm{V}(\boldsymbol{u}) = \sigma_u^2\boldsymbol{I}$. Прогноз (7.3) заменяется теперь на

$$\widehat{y} = \boldsymbol{z}'\widehat{\boldsymbol{\beta}}. \tag{7.11}$$

Пусть $e = \widehat{y} - y_{n+1}$ — ошибка прогнозирования. Тогда

$$\begin{aligned}\mathrm{E}e &= \mathrm{E}(\boldsymbol{z}'\widehat{\boldsymbol{\beta}}) - \boldsymbol{x}'_{n+1}\boldsymbol{\beta} = \mathrm{E}\big((\boldsymbol{x}_{n+1} + \boldsymbol{u})'\widehat{\boldsymbol{\beta}}\big) - \boldsymbol{x}'_{n+1}\boldsymbol{\beta}\\ &= \mathrm{E}(\boldsymbol{x}'_{n+1}\widehat{\boldsymbol{\beta}}) + \mathrm{E}(\boldsymbol{u}'\widehat{\boldsymbol{\beta}}) - \boldsymbol{x}'_{n+1}\boldsymbol{\beta} = 0,\end{aligned}$$

так как $\boldsymbol{u}$ и $\widehat{\boldsymbol{\beta}}$ независимы и $\mathrm{E}\boldsymbol{u} = \boldsymbol{0}$. Иными словами, оценка (7.11) является несмещенной. Можно проверить (мы оставляем это читателю в виде упражнения), что

$$\mathrm{E}e^2 = \sigma^2\big(1 + \boldsymbol{x}'_{n+1}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}_{n+1} + \sigma_u^2\,\mathrm{tr}((\boldsymbol{X}'\boldsymbol{X})^{-1})\big) + \sigma_u^2\boldsymbol{\beta}'\boldsymbol{\beta}. \tag{7.12}$$

Таким образом, при наличии ошибок в независимой переменной к ошибке прогнозирования (7.8) добавляются два новых положительных слагаемых, пропорциональных дисперсии σ_u^2.

В случае условного прогнозирования нельзя так же просто, как при безусловном прогнозировании, построить доверительный интервал для y_{n+1}. Это связано с тем, что при нормально распределенных ошибках $(\boldsymbol{\varepsilon}, \varepsilon_{n+1}, \boldsymbol{u})$ оценка $\widehat{y}$ есть скалярное произведение двух независимых нормальных векторов. Поэтому доверительный интервал нельзя найти аналитически, однако существуют численные процедуры, позволяющие строить его приближенно.

7.3. Прогнозирование при наличии авторегрессии ошибок

В заключение остановимся на задаче прогнозирования, когда ошибки в исходной модели (7.1), (7.2) коррелированы по времени, а именно, образуют авторегрессионный процесс первого порядка:

$$\varepsilon_t = \rho\varepsilon_{t-1} + \nu_t, \qquad t = 1, \ldots, n, n+1, \tag{7.13}$$

где $\{\nu_t,\ t = 1, \ldots, n, n+1\}$ — последовательность независимых нормально распределенных случайных величин с нулевым средним и постоянной дисперсией σ_ν^2, $|\rho| < 1$. Покажем, как можно использовать информацию об ошибках (7.13) для улучшения прогнозирования. Предположим, как и в начале этого раздела, что все параметры $(\boldsymbol{\beta}, \rho)$ известны. Но теперь в качестве оценки $\widehat{y}$ величины y_{n+1} возьмем не $\boldsymbol{x}'_{n+1}\boldsymbol{\beta}$, как раньше, а

$$\widehat{y} = \boldsymbol{x}'_{n+1}\boldsymbol{\beta} + \rho\varepsilon_n = \boldsymbol{x}'_{n+1}\boldsymbol{\beta} + \rho(y_n - \boldsymbol{x}'_n\boldsymbol{\beta}). \tag{7.14}$$

Нетрудно проверить, что

$$e = y_{n+1} - \widehat{y} = \nu_{n+1},$$

откуда сразу следует, что $\mathrm{E}e = 0$ и

$$\mathrm{E}e^2 = \sigma_\nu^2 = (1-\rho^2)\sigma_\varepsilon^2. \qquad (7.15)$$

Таким образом, удается уменьшить ошибку прогноза по сравнению со случаем некоррелированных ошибок.

Реально параметры регрессии неизвестны, поэтому при прогнозировании величины y_{n+1} в формуле (7.14) значения $\boldsymbol{\beta}$ и ρ заменяют их оценками, полученными с помощью, например, одной из процедур, описанных в п. 6.2:

$$\widehat{y} = \boldsymbol{x}'_{n+1}\widehat{\boldsymbol{\beta}} + r(y_n - \boldsymbol{x}'_n\widehat{\boldsymbol{\beta}}). \qquad (7.16)$$

Мы не можем дать аналитическое выражение для среднеквадратичной ошибки прогноза. На практике используют формулу (7.15) с заменой величины σ_ν^2 на ее оценку, получаемую из регрессии (6.11), (6.12).

Выводы:

1) прогноз $\widehat{y}$ величины y_{n+1} в модели (7.1), (7.2) задается равенством (7.3), где $\widehat{\boldsymbol{\beta}}$ — МНК-оценка вектора $\boldsymbol{\beta}$, полученная в регрессии (7.1);
2) эта оценка обладает минимальной среднеквадратичной ошибкой в классе линейных несмещенных оценок величины y_{n+1};
3) среднеквадратичная ошибка прогнозирования определяется равенством (7.8);
4) при наличии ошибок в независимых переменных ошибка прогнозирования возрастает в соответствии с формулой (7.12);
5) если ошибки в модели (7.1), (7.2) образуют авторегрессионный процесс первого порядка, то можно уменьшить ошибку прогнозирования, воспользовавшись формулой (7.16).

Упражнения

7.1. Проверьте формулу (7.8).

7.2. Докажите равенство (7.9).

7.3. Имеется $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ — классическая регрессионная модель ($\boldsymbol{y}$ — $n \times 1$ вектор, $\boldsymbol{X}$ — $n \times k$ матрица, $\boldsymbol{\varepsilon}$ — $n \times 1$ вектор ошибок, $\boldsymbol{\beta}$ — $k \times 1$ вектор коэффициентов, $\mathrm{E}\boldsymbol{\varepsilon} = \boldsymbol{0}$, $\mathrm{V}(\boldsymbol{\varepsilon}) = \sigma^2\boldsymbol{I}$). Пусть $\boldsymbol{x}_{n+1} = (x_{n+1,1}, \ldots, x_{n+1,k})'$ — дополнительное наблюдение независимых переменных и $y_{n+1} = \boldsymbol{x}'_{n+1}\boldsymbol{\beta} + \varepsilon_{n+1}$.

Покажите, что если матрица $\boldsymbol{X}$ содержит константу, то ошибка прогноза минимальна, если каждое $x_{n+1,j}$ равно среднему j-го столбца матрицы $\boldsymbol{X}$.

7.4. Для модели парной регрессии $y_t = \alpha + \beta x_t + \varepsilon_t$, $t = 1, \ldots, 10$ известно, что

$$\sum y_t = 8, \ \sum x_t = 40, \ \sum y_t^2 = 26, \ \sum x_t^2 = 200, \ \sum y_t x_t = 20$$

(всюду суммирование от 1 до 10). Для некоторого наблюдения s дано $x_s = 10$. Предполагая, что наблюдение s удовлетворяет исходной модели,

а) вычислите наилучший линейный несмещенный прогноз величины y_s;

б) оцените стандартную ошибку прогноза.

7.5. Стандартная линейная модель $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, где $\boldsymbol{y}$ — $n \times 1$ вектор, $\boldsymbol{X}$ — $n \times k$ матрица, оценивается обычным методом наименьших квадратов. Имеется дополнительное наблюдение y_0, $\boldsymbol{x}'_0 = (x_{0\,1}, \ldots, x_{0\,k})$.

С помощью какой статистики можно ответить на вопрос, удовлетворяет ли это наблюдение исходной модели?

7.6. Дана регрессионная модель $y_t = \alpha + \beta x_t + \varepsilon_t$, $t = 1, \ldots, n$. Предположим, что параметр β известен. Предложите способ прогноза величины y_{n+1} (для заданного x_{n+1}) и найдите дисперсию ошибки прогноза.

Глава 8

Инструментальные переменные

В п. 5.1 при рассмотрении модели со стохастическими регрессорами отмечалось, что при наличии корреляции между независимыми переменными и ошибками МНК-оценки могут быть смещенными и несостоятельными. Один из путей преодоления этой трудности — использование других независимых переменных, которые носят название *инструментальные переменные*. Как будет показано ниже, для получения состоятельных оценок надо, чтобы они обладали двумя свойствами:

1) новые независимые переменные должны быть «хорошо коррелированы» с исходными независимыми переменными;

2) новые переменные не должны быть коррелированы с ошибками.

Точный смысл этих условий проясняется при более детальном изучении модели.

8.1. Состоятельность оценок, полученных с помощью инструментальных переменных

Пусть исходная модель описывается равенством

$$y = X\beta + \varepsilon, \tag{8.1}$$

где, как и раньше, y — $n \times 1$ вектор зависимых переменных, X — $n \times k$ матрица независимых переменных, ε — $n \times 1$ вектор ошибок, β — $k \times 1$ вектор параметров. Предположим также, что задана матрица Z размера $n \times k$ — матрица инструментальных переменных, причем $k \times k$ матрица $Z'X$ обратима. Тогда по определению оценкой параметров β с помощью *инструментальных переменных* (*Instrumental Variables, IV*) называется вектор

$$\widehat{\beta}_{\mathrm{IV}} = (Z'X)^{-1}Z'y. \tag{8.2}$$

Нетрудно понять, почему именно в таком виде следует брать оценку, а не в виде $\widehat{\beta}_{\mathrm{IV}} = (Z'Z)^{-1}Z'y$, как могло бы показаться на первый взгляд. Действительно, подставляя в (8.2) выражение для y из (8.1), получаем

$$\begin{aligned}\widehat{\beta}_{\mathrm{IV}} &= (Z'X)^{-1}Z'(X\beta + \varepsilon) = \beta + (Z'X)^{-1}Z'\varepsilon \\ &= \beta + \left(\frac{1}{n}Z'X\right)^{-1}\frac{1}{n}Z'\varepsilon.\end{aligned} \tag{8.3}$$

Предположим, что выполнено следующее условие, формально выражающее требование «хорошей корреляции» между X и Z:

последовательность матриц $\frac{1}{n}Z'X$ *сходится по вероятности при* $n \to \infty$ *к некоторой невырожденной матрице.* (8.4)

В силу отсутствия корреляции между Z и ε, член $(1/n)Z'\varepsilon$ стремится по вероятности к $\mathbf{0}$.[1]

[1] Заметим, что условие некоррелированности Z и ε можно заменить более слабым условием: $p\lim \frac{1}{n}Z'\varepsilon = \mathbf{0}$.

Таким образом, из (8.3) и (8.4) следует, что оценка $\widehat{\beta}_{\mathrm{IV}}$ является состоятельной. В то же время следует подчеркнуть, что в общем случае $\widehat{\beta}_{\mathrm{IV}}$ является смещенной и не обладает минимальной матрицей ковариаций, т. е., вообще говоря, не является эффективной (что естественно, поскольку оценка $\widehat{\beta}_{\mathrm{IV}}$ явно зависит от Z).

Возникает естественный вопрос, как находить нужные инструментальные переменные. Дать строгий однозначный ответ на него невозможно. Все зависит от конкретной ситуации. Может так случиться, что инструментальных переменных нельзя найти, а может быть, что существует и несколько таких инструментов.

8.2. Влияние ошибок измерения

Типичной ситуацией, когда требуется использовать инструментальные переменные, является наличие ошибок в измерениях. Рассмотрим этот случай подробнее.

Ошибки в измерениях зависимой переменной. Предположим, что истинной является модель (8.1), но вектор y измеряется с ошибкой, т. е. наблюдается вектор $y^* = y + u$, где u — ошибки, имеющие нулевое математическое ожидание и не зависящие от ε и X. Тогда нетрудно понять, что построение МНК-оценок на основании y^* эквивалентно регрессии

$$y^* = X\beta + (\varepsilon + u),$$

откуда видно, что оценка параметра β будет несмещенной и состоятельной, так как $\mathrm{E}(\varepsilon + u) = 0$ и $\mathrm{Cov}(X, \varepsilon + u) = 0$. Наличие же ошибки приводит лишь к увеличению дисперсии регрессии: $\mathrm{V}(\varepsilon + u) = \sigma_\varepsilon^2 + \sigma_u^2$.

Ошибки в измерении независимой переменной. Ситуация радикальным образом меняется в худшую сторону, когда есть ошибки в измерении регрессоров. Пусть, как и раньше, истинной является модель (8.1), но наблюдается матрица $X^* = X + V$, где матрица ошибок V имеет нулевое математическое ожидание и не зависит

от ε. Тогда реально будет осуществляться регрессия

$$y = X^*\beta + (\varepsilon - V\beta) = X^*\beta + \varepsilon^*,$$

в которой регрессоры и ошибки уже являются коррелированными: $\mathrm{E}(X^{*\prime}\varepsilon^*) = \mathrm{E}[(X' + V')(\varepsilon - V\beta)] = -\mathrm{E}(V'V)\beta$. Это означает, что в общем случае МНК-оценки будут смещенными и несостоятельными. Степень смещения определяется как истинным значением параметра, так и матрицей ковариаций ошибок. Здесь можно применять метод инструментальных переменных.

Понятно, что можно рассматривать общий случай, когда есть ошибки в измерениях независимых и зависимых переменных. Ясно, что, как и в предыдущем случае, применение метода наименьших квадратов будет приводить к смещенным и несостоятельным оценкам.

8.3. Двухшаговый метод наименьших квадратов

Нетрудно понять, что метод оценивания с помощью инструментальных переменных является обобщением обычного метода наименьших квадратов. Подчеркнем еще раз, что нахождение нужных инструментальных переменных является нелегкой задачей, решение которой зависит от конкретной ситуации.

Совпадение числа инструментальных переменных с числом исходных регрессоров не является обязательным условием. На самом деле, достаточно требовать, чтобы число инструментов было не меньше, чем число независимых переменных. Выведем формулу для $\widehat{\beta}_{\mathrm{IV}}$ в этом случае. Итак, пусть Z — $n \times m$ матрица, столбцы которой линейно независимы (т. е. $\mathrm{rank}(Z) = m$), причем $m \geqslant k$. Если воспользоваться геометрической интерпретацией метода наименьших квадратов, то достаточно легко понять смысл оценки $\widehat{\beta}_{\mathrm{IV}}$ и получить требуемую формулу. Рассмотрим каждый столбец x_j, $j = 1, \ldots, k$, матрицы X как n-мерный вектор и спроектируем его на m-мерное подпространство, порожденное столбцами матрицы Z, получив векторы $\widehat{x}_j$, $j = 1, \ldots, k$. Это эквива-

лентно тому, что мы осуществляем регрессию x_j на Z и находим прогнозные значения $\widehat{x}_j$. Их мы теперь будем считать новыми независимыми переменными, и оценка $\widehat{\beta}_{\mathrm{IV}}$ вектора параметров β строится с помощью обычной регрессии y на $\widehat{x}_j$, $j = 1, \ldots, k$. Таким образом, метод наименьших квадратов применяется здесь дважды — сначала для построения регрессоров $\widehat{x}_j$, $j = 1, \ldots, k$, а затем для нахождения оценки $\widehat{\beta}_{\mathrm{IV}}$. Эта процедура носит название *двухшагового метода наименьших квадратов*, о котором мы будем более подробно говорить, изучая системы регрессионных уравнений. Вспоминая, что проекция на пространство, порожденное векторами-столбцами матрицы Z, осуществляется с помощью матрицы $Z(Z'Z)^{-1}Z'$ (см. п. 3.3), имеем

$$\widehat{x}_j = Z(Z'Z)^{-1}Z'x_j, \qquad j = 1, \ldots, k,$$

и, следовательно,

$$\widehat{X} = [\widehat{x}_1 \ \ldots \ \widehat{x}_k] = Z(Z'Z)^{-1}Z'X.$$

Наконец,

$$\begin{aligned}\widehat{\beta}_{\mathrm{IV}} &= (\widehat{X}'\widehat{X})^{-1}\widehat{X}'y \\ &= \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1}X'Z(Z'Z)^{-1}Z'y. \qquad (8.5)\end{aligned}$$

При этом мы, конечно, предполагаем, что матрица $\widehat{X}$ имеет полный ранг k. Отсюда следует, что условие $m \geqslant k$ является необходимым для использования инструментальных переменных.

Как и раньше, нетрудно показать, что если выполнены условия, аналогичные условию (8.4), а именно:

1) *последовательность матриц* $\frac{1}{n}Z'X$ *сходится по вероятности при* $n \to \infty$ *к некоторой матрице полного ранга* k;

2) *последовательность матриц* $\frac{1}{n}Z'Z$ *сходится по вероятности при* $n \to \infty$ *к некоторой невырожденной матрице,*

то оценка (8.5) будет состоятельной.

8.4. Тест Хаусмана

Как определить, следует ли использовать инструментальные переменные или достаточно применять обычный метод наименьших квадратов? Ответ на него равносилен тестированию гипотезы H_0: $p\lim(1/n)X'\varepsilon = 0$ против альтернативы H_1: $p\lim(1/n)X'\varepsilon \neq 0$. Довольно очевидно, что при наличии только наблюдений y, X проверить эту гипотезу нельзя, поэтому предположим, что наряду с обычной МНК-оценкой $\widehat{\beta}_1 = \widehat{\beta}_{\mathrm{OLS}}$ есть оценка $\widehat{\beta}_2 = \widehat{\beta}_{\mathrm{IV}}$, полученная с помощью некоторых инструментальных переменных. При гипотезе H_0 оценка $\widehat{\beta}_1$ является состоятельной и эффективной, а при альтернативной гипотезе H_1 — несостоятельной. Оценка $\widehat{\beta}_2$ состоятельна как при нулевой, так и при альтернативной гипотезах. Таким образом, при нулевой гипотезе разность $\widehat{\beta}_2 - \widehat{\beta}_1$ стремится к нулю, и естественно ожидать, что при соответствующей нормировке распределение этой разности будет асимптотически совпадать с каким-нибудь известным распределением.

Хаусман (Hausman, 1978) доказал, что асимптотически $\mathrm{V}(\widehat{\beta}_2 - \widehat{\beta}_1) = \mathrm{V}(\widehat{\beta}_2) - \mathrm{V}(\widehat{\beta}_1)$ и величина

$$(\widehat{\beta}_2 - \widehat{\beta}_1)'\left(\mathrm{V}(\widehat{\beta}_2) - \mathrm{V}(\widehat{\beta}_1)\right)^{-1}(\widehat{\beta}_2 - \widehat{\beta}_1)$$

асимптотически имеет хи-квадрат распределение с k степенями свободы.

Более подробно о тесте Хаусмана можно прочесть, например, в (Johnston and DiNardo, 1997).

Выводы:

1) при наличии корреляции между независимыми переменными и ошибками МНК-оценки смещены и несостоятельны; для получения состоятельных оценок можно воспользоваться инструментальными переменными;

2) число инструментальных переменных должно быть не меньше числа исходных независимых переменных;

3) инструментальные переменные должны быть некоррелированы с ошибками и коррелированы с исходными независи-

мыми переменными (условие (8.4); в этом случае $\widehat{\boldsymbol{\beta}}_{\mathrm{IV}}$ состоятельна, но, вообще говоря, неэффективна;

4) при наличии ошибок в измерениях зависимой переменной МНК-оценка остается несмещенной и состоятельной, если же с ошибками измеряются независимые переменные, то это приводит к возникновению корреляции между регрессорами и ошибками и, как следствие, к несостоятельности МНК-оценки.

Упражнения

8.1. Проверьте формулу (8.5).

8.2. Докажите, что при $m = k$ оценка (8.5) совпадает с (8.2).

8.3. Найдите $\mathrm{V}(\widehat{\boldsymbol{\beta}}_{\mathrm{IV}})$ для оценок (8.2) и (8.5).

8.4. Пусть мы оцениваем регрессионное уравнение

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t, \quad t = 1, \ldots, n$$

с помощью метода инструментальных переменных, используя переменную z_t как инструмент для x_t.

Покажите, что оценки коэффициентов имеют вид

$$\widehat{\beta}_{1\mathrm{IV}} = \overline{y} - \widehat{\beta}_{2\mathrm{IV}}\overline{x}, \qquad \widehat{\beta}_{2\mathrm{IV}} = \frac{\sum(z_t - \overline{z})(y_t - \overline{y})}{\sum(z_t - \overline{z})(x_t - \overline{x})},$$

и являются решениями системы уравнений

$$n\widehat{\beta}_{1\mathrm{IV}} + \left(\sum x_t\right)\widehat{\beta}_{2\mathrm{IV}} = \sum y_t,$$
$$\left(\sum z_t\right)\widehat{\beta}_{1\mathrm{IV}} + \left(\sum z_t x_t\right)\widehat{\beta}_{2\mathrm{IV}} = \sum z_t y_t.$$

8.5. Рассмотрим модель (8.3)

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}, \quad \mathrm{V}(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I},$$

в которой регрессоры x_{tp} коррелированы с ошибками ε_t. Пусть $\boldsymbol{Z}$ — некоторая матрица. Преобразуем исходное уравнение, умножив его слева на $\boldsymbol{Z}'$:

$$\boldsymbol{Z}'\boldsymbol{y} = \boldsymbol{Z}'\boldsymbol{X\beta} + \boldsymbol{Z}'\boldsymbol{\varepsilon}. \qquad (*)$$

Покажите, что оценка обобщенного метода наименьших квадратов (5.4) для вектора коэффициентов уравнения (*) равна

$$\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}} = \left(\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{Z}(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{Z}'\boldsymbol{y}.$$

Сравните результат с формулой (8.5) для оценки метода инструментальных переменных.

8.6. Пусть переменные y_t^*, z_t^* связаны (точным) уравнением

$$y_t^* = \beta_1 + \beta_2 z_t^*.$$

Однако вместо точных значений мы наблюдаем измеренные (с ошибками измерений) значения $y_t = y_t^* + u_t$ и $z_t = z_t^* + v_t$, где $u_t \sim iid(0, \sigma_u^2)$, $v_t \sim iid(0, \sigma_v^2)$, ошибки u_t и v_s независимы при всех t и s. Мы оцениваем методом наименьших квадратов уравнение

$$y_t = \beta_1 + \beta_2 z_t + \varepsilon_t.$$

а) Удовлетворяют ли ошибки в данном уравнении условиям стандартной линейной модели?

б) Найти $\mathrm{Cov}(z_t, \varepsilon_t)$.

в) Найти $p\lim \widehat{\beta}_2$.

Глава 9

Системы регрессионных уравнений

При моделировании достаточно сложных экономических объектов часто приходится вводить не одно, а несколько связанных между собой уравнений, т. е. описывать модель системой уравнений. А значит, при проведении регрессионного анализа модели может возникнуть необходимость оценивать систему уравнений. Например, простейшая макроэкономическая кейнсианская модель потребления может быть представлена в следующем виде:

$$C_t = \beta_1 + \beta_2 Y_t + \varepsilon_t,$$
$$Y_t = C_t + I_t,$$

где C_t — агрегированное потребление, Y_t — национальный доход, I_t — инвестиции в период времени t. Коэффициент β_2 носит название *склонность к потреблению*.

Как будет показано ниже, наличие связи между переменными C_t и Y_t, определяемой вторым уравнением, требует корректировки метода наименьших квадратов для оценивания параметров модели β_1 и β_2. Вообще, оценивание систем уравнений требует введения новых понятий и разработки новых методов, чему и посвящена данная глава. Сначала мы рассмотрим более простую задачу оценивания системы, в которой уравнения связаны лишь благодаря наличию корреляции между ошибками в разных уравнениях, — это так называемая система *внешне не связанных между собой уравнений* (*Seemingly Unrelated Regression, SUR*). Затем мы исследуем общие системы регрессионных уравнений, которые в эконометрике называются *системами одновременных уравнений* (*Simultaneous equations*).

9.1. Внешне не связанные уравнения

Чтобы понять постановку задачи и суть проблемы, рассмотрим следующий пример. Предположим, что исследуется зависимость инвестиций y, осуществляемых некоторым предприятием (например, компанией «Газпром»), от его дохода x_1 и размера основного фонда x_2:

$$y_t = \beta_1 + \beta_2 x_{t1} + \beta_3 x_{t2} + \varepsilon_t, \quad t = 1, \ldots, n. \tag{9.1}$$

Представим теперь, что имеется ряд наблюдений другого аналогичного предприятия (например, компании «ЛУКОЙЛ»):

$$z_t = \gamma_1 + \gamma_2 p_{t1} + \gamma_3 p_{t2} + u_t, \quad t = 1, \ldots, n. \tag{9.2}$$

Конечно, можно оценивать уравнения (9.1), (9.2) по отдельности. Внешне они выглядят как не связанные друг с другом. Но ясно, что в данной ситуации естественно считать ошибки ε_t и u_t коррелированными, поскольку предприятия в каждый период t действуют в «одной экономической среде». Поэтому целесообразно объединить уравнения (9.1), (9.2) и оценивать их совместно, используя доступный обобщенный метод наименьших квадратов.

Общая задача формулируется следующим образом. Даны M регрессионных уравнений:

$$\begin{aligned} \boldsymbol{y}_1 &= \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1, \\ \boldsymbol{y}_2 &= \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_2, \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ \boldsymbol{y}_M &= \boldsymbol{X}_M\boldsymbol{\beta}_M + \boldsymbol{\varepsilon}_M, \end{aligned} \tag{9.3}$$

где y_i — $n \times 1$ вектор зависимых переменных, X_i — $n \times k_i$ матрица независимых переменных, β_i — $k_i \times 1$ вектор неизвестных параметров, ε_i — $n \times 1$ вектор ошибок, $i = 1, \ldots, M$. Будем предполагать, что $\mathrm{E}\varepsilon_i = \mathbf{0}$ и $\mathrm{E}(\varepsilon_{is}\varepsilon_{jt}) = \sigma_{ij}$ при $s = t$ и 0 в противном случае. Последнее условие можно представить так:

$$\mathrm{E}(\varepsilon_i\varepsilon_j') = \sigma_{ij}I_n, \quad i, j = 1, \ldots, M, \tag{9.4}$$

где I_n — единичная матрица размера $n \times n$. Иными словами, заданы M регрессионных уравнений, по каждому из которых имеется n наблюдений. Если данные имеют структуру временных рядов, то считается, что ошибки во всех уравнениях коррелированы в один и тот же момент времени и некоррелированы для разных моментов. Равенство (9.4) определяет связь между этими уравнениями. Каждое отдельное уравнение в (9.3) удовлетворяет условиям классической регрессионной модели и может быть оценено обычным методом наименьших квадратов. Однако, если объединить эти уравнения и применить обобщенный метод наименьших квадратов, то можно повысить эффективность оценивания.

Обозначим

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix}, \quad X = \begin{bmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X_M \end{bmatrix},$$

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_M \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_M \end{bmatrix}, \quad \Sigma = (\sigma_{ij}), \quad i, j = 1, \ldots, M.$$

Тогда система (9.3) переписывается в виде

$$y = X\beta + \varepsilon.$$

Используя понятие произведения Кронекера двух матриц, ковариационную матрицу вектора ошибок можно представить так:

$$\mathrm{E}(\varepsilon\varepsilon') = \Omega = \Sigma \otimes I_n$$

(приложение ЛА, п. 18). Предположим, что матрица Σ не вырождена. Для построения оценки вектора β применим обобщенный метод наименьших квадратов (п. 5.2, формула (5.4)):

$$\begin{aligned} \widehat{\beta}_{\mathrm{GLS}} &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \\ &= (X'(\Sigma^{-1} \otimes I_n)X)^{-1}X'(\Sigma^{-1} \otimes I_n)y \end{aligned} \tag{9.5}$$

(здесь мы воспользовались известным свойством произведения Кронекера: для двух квадратных невырожденных матриц A и B выполнено равенство $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ (приложение ЛА, п. 18)).

Нетрудно понять, что в общем случае оценка (9.5) отличается от оценки, полученной в результате применения обычного метода наименьших квадратов к каждому уравнению в системе (9.3). Есть, однако, две ситуации, когда эти оценки совпадают.

1. Уравнения в (9.3) действительно не связаны друг с другом, т. е. $\sigma_{ij} = 0$ при $i \neq j$.

2. Все уравнения в (9.3) имеют один и тот же набор независимых переменных, т. е. $X_1 = X_2 = \ldots = X_M$.

Первое утверждение почти очевидно, поскольку матрица Ω в этом случае является диагональной. Доказательство второго утверждения требует некоторых вычислений, мы его оставляем читателю в качестве упражнения.

Для использования доступного обобщенного метода наименьших квадратов нужно оценить матрицу Σ. Это можно сделать, применяя к каждому уравнению системы (9.3) обычный метод наименьших квадратов, получая векторы остатков e_i, $i = 1, \ldots, M$, и беря в качестве оценок ковариаций σ_{ij} величины $s_{ij} = (e_i'e_j)/n$. Можно проверить, что эти оценки являются состоятельными.

Отметим в заключение, что эффективность оценки $\widehat{\beta}_{\mathrm{GLS}}$ (или ее доступного варианта) по сравнению с МНК-оценками тем выше, чем сильнее корреляция между ошибками.

9.2. Системы одновременных уравнений

Примеры: *кривые спроса и предложения*

Пример 1. Рассмотрим вначале простой пример системы одновременных уравнений, который демонстрирует основные проблемы, возникающие при попытке оценить неизвестные параметры. (Этот пример входит практически во все учебники по эконометрике.) Предположим, что исследуется зависимость спроса и предложения некоторого товара от его цены и дохода — так называемые кривые спроса и предложения:

$$Q_t^S = \alpha_1 + \alpha_2 P_t + \varepsilon_t \qquad \text{(предложение)},$$
$$Q_t^D = \beta_1 + \beta_2 P_t + \beta_3 Y_t + u_t \qquad \text{(спрос)},$$

где P_t — цена товара, Y_t — доход в момент времени t. Предполагается, что на рынке существует равновесие, т. е. в каждый момент времени наблюдается одна величина

$$Q_t^S = Q_t^D = Q_t \qquad \text{(равновесие)}.$$

Записывая каждое уравнение, для простоты в отклонениях от средних значений (см. п. 2.2), получаем следующую систему:

$$q_t = \alpha_2 p_t + \varepsilon_t \qquad \text{(предложение)}, \tag{9.6}$$
$$q_t = \beta_2 p_t + \beta_3 y_t + u_t \qquad \text{(спрос)}. \tag{9.7}$$

Отметим, что в соответствии с этой моделью цена и величина спроса-предложения определяются одновременно (отсюда и термин «одновременные уравнения») и поэтому обе эти переменные должны считаться *эндогенными*. В отличие от них доход y_t является *экзогенной* переменной. Подчеркнем, что деление переменных на экзогенные и эндогенные определяется содержательной стороной модели. *Предполагается, что в каждом уравнении экзогенные переменные некоррелированы с ошибкой*. В то же время эндогенные переменные, стоящие в правых частях уравнений, как

правило, имеют ненулевую корреляцию с ошибкой в соответствующем уравнении. Действительно, разрешим систему (9.6), (9.7) относительно q_t и p_t:

$$q_t = \frac{\alpha_2\beta_3 y_t}{\alpha_2 - \beta_2} + \frac{\alpha_2 u_t - \beta_2\varepsilon_t}{\alpha_2 - \beta_2}, \tag{9.8}$$
$$p_t = \frac{\beta_3 y_t}{\alpha_2 - \beta_2} + \frac{u_t - \varepsilon_t}{\alpha_2 - \beta_2}. \tag{9.9}$$

Тогда, учитывая некоррелированность y_t с u_t и ε_t, из (9.9) получаем

$$\operatorname{Cov}(p_t, \varepsilon_t) = \frac{\operatorname{Cov}(u_t, \varepsilon_t) - \operatorname{V}(\varepsilon_t)}{\alpha_2 - \beta_2}, \tag{9.10}$$

что, в общем случае, не равно 0. В п. 5.1 при рассмотрении модели со стохастическими регрессорами отмечалось, что наличие корреляции между регрессорами и ошибками приводит к смещенности и несостоятельности МНК-оценок. В нашем простом примере величину асимптотического смещения можно получить в явном виде. Как известно (см. (2.6)), МНК-оценка коэффициента α_2 в уравнении (9.6) имеет следующий вид:

$$\widehat{\alpha}_2 = \sum_{t=1}^n q_t p_t \Big/ \sum_{t=1}^n p_t^2 = \alpha_2 + \sum_{t=1}^n \varepsilon_t p_t \Big/ \sum_{t=1}^n p_t^2. \tag{9.11}$$

Так как в последнем слагаемом в (9.11) числитель состоит из коррелированных величин, а числитель и знаменатель зависимы, то нет никакой надежды, что в общем случае

$$\operatorname{E}\left(\sum_{t=1}^n \varepsilon_t p_t \Big/ \sum_{t=1}^n p_t^2\right) = 0,$$

т. е. оценка $\widehat{\alpha}_2$ является смещенной. Кроме того, она и несостоятельна. Действительно, предположим для простоты, что ошибки u, ε независимы и распределения u_t и ε_t не зависят от t. Будем считать также, что существует $p\lim_{n\to\infty}(1/n)\sum_{t=1}^n y_t^2 = \operatorname{V}(y)$. Перепишем (9.11) так:

$$\widehat{\alpha}_2 = \alpha_2 + \left(\frac{1}{n}\sum_{t=1}^n \varepsilon_t p_t\right) \Big/ \left(\frac{1}{n}\sum_{t=1}^n p_t^2\right).$$

Тогда в силу закона больших чисел и формулы (9.10) имеем

$$\operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}\varepsilon_t p_t = -\frac{\mathrm{V}(\varepsilon_t)}{\alpha_2-\beta_2}.$$

(Напомним, что $\mathrm{Cov}(u_t,\varepsilon_t)=0$.) Далее, из (9.9) в силу некоррелированности y с u и ε получаем

$$\operatorname*{plim}_{n\to\infty}\frac{1}{n}\sum_{t=1}^{n}p_t^2 = \frac{1}{(\alpha_2-\beta_2)^2}(\beta_3^2\mathrm{V}(y)+\mathrm{V}(\varepsilon)+\mathrm{V}(u)).$$

Окончательно,

$$\operatorname*{plim}_{n\to\infty}\widehat{\alpha}_2 = \alpha_2 - \frac{\mathrm{V}(\varepsilon)(\alpha_2-\beta_2)}{\beta_3^2\mathrm{V}(y)+\mathrm{V}(u)+\mathrm{V}(\varepsilon)} = \lambda\alpha_2+(1-\lambda)\beta_2,$$

где

$$\lambda = \frac{\beta_3^2\mathrm{V}(y)+\mathrm{V}(u)}{\beta_3^2\mathrm{V}(y)+\mathrm{V}(u)+\mathrm{V}(\varepsilon)},\quad 1-\lambda = \frac{\mathrm{V}(\varepsilon)}{\beta_3^2\mathrm{V}(y)+\mathrm{V}(u)+\mathrm{V}(\varepsilon)}.$$

Таким образом, $\operatorname{plim}_{n\to\infty}\widehat{\alpha}_2=\alpha_2$, только если $\mathrm{V}(\varepsilon)=0$.

Система (9.6), (9.7) называется *структурной формой модели*, соответственно коэффициенты этих уравнений называются *структурными коэффициентами*. Система (9.8), (9.9) называется *приведенной формой модели*. Обозначая

$$\pi_1=(\alpha_2\beta_3)/(\alpha_2-\beta_2),\qquad \nu_{1t}=(\alpha_2u_t-\beta_2\varepsilon_t)/(\alpha_2-\beta_2), \tag{9.12}$$

$$\pi_2=\beta_3/(\alpha_2-\beta_2),\qquad \nu_{2t}=(u_t-\varepsilon_t)/(\alpha_2-\beta_2), \tag{9.13}$$

перепишем (9.8) и (9.9):

$$q_t=\pi_1y_t+\nu_{1t},$$
$$p_t=\pi_2y_t+\nu_{2t}.$$

Здесь уже в каждом уравнении экзогенная переменная некоррелирована с ошибкой, поэтому метод наименьших квадратов даст состоятельные оценки $\widehat{\pi}_1$ и $\widehat{\pi}_2$ коэффициентов π_1 и π_2. Заметим, что $\alpha_2=\pi_1/\pi_2$, поэтому (в силу теоремы Слуцкого) величина $\widehat{\alpha}_{2\,\mathrm{ILS}}=\widehat{\pi}_1/\widehat{\pi}_2$ будет состоятельной оценкой структурного параметра α_2. Такой способ оценивания структурных коэффициентов с помощью оценок коэффициентов приведенной формы

называется *косвенным методом наименьших квадратов* (*Indirect Least Squares, ILS*). Следовательно, для структурного коэффициента первого уравнения можно построить состоятельную оценку, используя косвенный метод наименьших квадратов.

В главе 8 отмечалось, что при наличии корреляции между регрессорами и ошибками для получения состоятельных оценок можно воспользоваться методом инструментальных переменных. В нашей модели для оценивания α_2 в качестве инструмента естественно использовать y — эта переменная некоррелирована с ε по условию и в силу (9.9) коррелирована с p. Тогда согласно (8.2)

$$\widehat{\alpha}_{2\,\mathrm{IV}}=\sum_{t=1}^{n}q_ty_t\Big/\sum_{t=1}^{n}p_ty_t=\widehat{\pi}_1/\widehat{\pi}_2=\widehat{\alpha}_{\mathrm{ILS}},$$

поскольку

$$\widehat{\pi}_1=\sum_{t=1}^{n}q_ty_t\Big/\sum_{t=1}^{n}y_t^2,\qquad \widehat{\pi}_2=\sum_{t=1}^{n}p_ty_t\Big/\sum_{t=1}^{n}y_t^2.$$

Таким образом, в данном случае оценки, полученные косвенным методом наименьших квадратов и с помощью инструментальных переменных, совпадают.

Пример 2. Усложним нашу исходную модель, включив в уравнение (9.7) для спроса процентную ставку r_t:

$$q_t=\beta_2p_t+\beta_3y_t+\beta_4r_t+u_t\qquad\text{(спрос)}, \tag{9.14}$$

считая эту переменную экзогенной. Проводя непосредственные вычисления, для системы (9.6), (9.14) получаем следующую приведенную форму:

$$q_t=\pi_{11}y_t+\pi_{12}r_t+\nu_{1t},$$
$$p_t=\pi_{21}y_t+\pi_{22}r_t+\nu_{2t},$$

где

$$\pi_{11}=(\alpha_2\beta_3)/(\alpha_2-\beta_2),\qquad \pi_{12}=(\alpha_2\beta_4)/(\alpha_2-\beta_2),$$
$$\pi_{21}=\beta_3/(\alpha_2-\beta_2),\qquad \pi_{22}=\beta_4/(\alpha_2-\beta_2),$$

а ν_{1t} и ν_{2t} — такие же, как и в (9.12), (9.13). Очевидно, что $\alpha_2=\pi_{11}/\pi_{21}=\pi_{12}/\pi_{22}$. Поэтому при использовании косвенного метода

наименьших квадратов можно в качестве оценки структурного параметра α_2 брать либо $\widehat{\pi}_{11}/\widehat{\pi}_{21}$, либо $\widehat{\pi}_{12}/\widehat{\pi}_{22}$, причем, в общем случае, это будут *разные* оценки. Точно так же можно в качестве инструмента использовать как y_t, так и r_t, тоже получая разные оценки. При этом, естественно, возникает вопрос, какая из них лучше. Ответ на него будет дан ниже при рассмотрении общей задачи.

В то же время, как нетрудно проверить, даже знание точных значений коэффициентов приведенной формы и для исходной, и для усложненной моделей не позволяет сделать никаких выводов относительно структурных параметров второго уравнения. Для этого уравнения также невозможно использовать y или r в качестве инструментальной переменной из-за возникающей при этом линейной зависимости между регрессорами. Это явление тесно связано с так называемой проблемой *идентификации*, о которой подробно будет говориться ниже. В данном случае нетрудно понять, почему уравнение (9.7) для спроса неидентифицируемо. Действительно, возьмем произвольное число λ и составим линейную комбинацию уравнений (9.6) и (9.7), умножая первое на λ, второе — на $(1-\lambda)$ и складывая их:

$$q_t = \gamma_2 p_t + \gamma_3 y_t + \eta_t, \tag{9.15}$$

где

$$\gamma_2 = \lambda\alpha_2 + (1-\lambda)\beta_2, \quad \gamma_3 = (1-\lambda)\beta_3, \quad \eta_t = \lambda\varepsilon_t + (1-\lambda)u_t.$$

Уравнение (9.15) имеет точно такой же вид, что и уравнение (9.7). Иными словами, существует бесконечно много структурных форм, совместимых с имеющимися данными q_t, p_t, y_t. Поэтому какой бы метод оценивания структурных коэффициентов уравнения спроса ни был выбран, нельзя сказать, какое отношение полученные оценки имеют к исходным параметрам β_2 и β_3. Подчеркнем, что это не статистическая проблема, не проблема количества наблюдений: даже имея бесконечное число наблюдений, невозможно «правильно» оценить уравнение спроса (9.7).

Сформулируем выводы, которые мы получили, и проблемы, с которыми столкнулись, исследуя этот простой пример системы одновременных уравнений.

1) Переменные в системах одновременных уравнений делятся на экзогенные и эндогенные. Первые отличаются от вторых тем, что в каждом уравнении они некоррелированы с соответствующей ошибкой.
2) Из-за наличия корреляции между эндогенными переменными и ошибками непосредственное применение метода наименьших квадратов к структурной форме модели приводит к смещенным и несостоятельным оценкам структурных коэффициентов.
3) Коэффициенты приведенной формы модели могут быть состоятельно оценены методом наименьших квадратов. Эти оценки могут быть использованы для оценивания структурных параметров (косвенный метод наименьших квадратов). При этом возможны три ситуации: структурный коэффициент однозначно выражается через коэффициенты приведенной системы, структурный коэффициент допускает несколько разных оценок косвенного метода наименьших квадратов, структурный коэффициент не может быть выражен через коэффициенты приведенной системы. В последнем случае соответствующее структурное уравнение является неидентифицируемым. Неидентифицируемость уравнения не связана с числом наблюдений.
4) Экзогенные переменные можно использовать в качестве инструментальных. В том случае, когда оценка косвенного метода единственна, она совпадает с оценкой, полученной с помощью инструментальных переменных.

Прежде чем перейти к общей теории, рассмотрим две модификации исходной модели (9.6), (9.7) с тем, чтобы дать более наглядное представление о понятии идентифицируемости.

Пример 3. Спрос и предложение зависят только от цены:

$$q_t = \alpha_2 p_t + \varepsilon_t \qquad \text{(предложение)},$$
$$q_t = \beta_2 p_t + u_t \qquad \text{(спрос)}.$$

На плоскости (Q, P) равновесие представляется как пересечение кривых (в данном случае прямых) спроса и предложения.

В этой модели имеются только одна кривая спроса и одна кривая предложения, а различие в наблюдаемых значениях обусловлено только случайными ошибками ε и u (см. рис. 9.1).

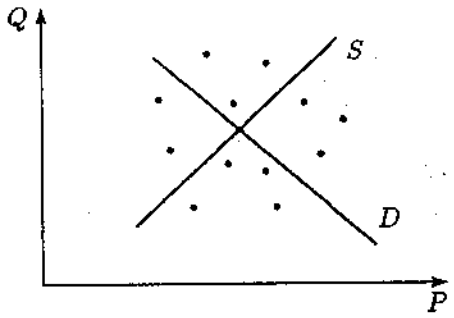


Рис. 9.1

Понятно, что, имея только «облако» наблюдений (Q_t, P_t), $t = 1, \ldots, n$, ничего нельзя сказать об «истинных» прямых D и S, поскольку каждая точка (Q_t, P_t) может быть реализована как пересечение двух прямых, имеющих произвольный наклон. Заметим, что этот вывод подтверждается и приведенной формой модели

$$q_t = (\alpha_2 u_t - \beta_2 \varepsilon_t)/(\alpha_2 - \beta_2),$$
$$p_t = (u_t - \varepsilon_t)/(\alpha_2 - \beta_2),$$

которая в правых частях уравнений содержит только случайные ошибки.

Пример 4. Исходная модель (9.6), (9.7). Здесь имеются одна кривая предложения и несколько кривых спроса, благодаря наличию экзогенной переменной y, а разброс в наблюдениях обусловлен не только случайными ошибками, но и сдвигом кривой спроса вдоль единственной кривой предложения. Это обстоятельство и позволяет оценить параметры последней (см. рис. 9.2). В то же время о положении прямых D_t ничего сказать нельзя, поскольку, как и в предыдущем примере, любой их наклон совместим с имеющимися наблюдениями.

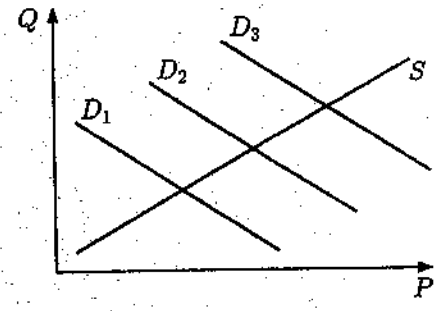


Рис. 9.2

Пример 5. Предложение также зависит от дохода:

$$q_t = \alpha_2 p_t + \alpha_3 y_t + \varepsilon_t \qquad \text{(предложение)},$$
$$q_t = \beta_2 p_t + \beta_3 y_t + u_t \qquad \text{(спрос)}.$$

В этом случае разброс наблюдений объясняется не только наличием случайных ошибок, но и одновременным сдвигом обеих кривых (см. рис. 9.3).

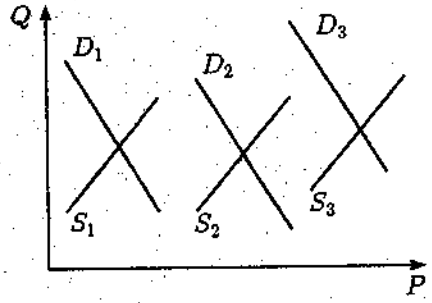


Рис. 9.3

Так же, как и в первом примере, ни одна из кривых не может быть идентифицирована.

Системы одновременных уравнений в матричной форме. Проблема идентифицируемости

Предположим теперь, что имеется следующая система уравнений, называемая *структурной формой модели* (с этим понятием мы

встречались ранее в п. 9.2, пример 1):

$$\begin{array}{l} \beta_{11}y_{1t} + \beta_{12}y_{2t} + \ldots + \beta_{1m}y_{mt} + \gamma_{11}x_{1t} + \ldots + \gamma_{1k}x_{kt} = \varepsilon_{1t} \\ \beta_{21}y_{1t} + \beta_{22}y_{2t} + \ldots + \beta_{2m}y_{mt} + \gamma_{21}x_{1t} + \ldots + \gamma_{2k}x_{kt} = \varepsilon_{2t} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \beta_{m1}y_{1t} + \beta_{m2}y_{2t} + \ldots + \beta_{mm}y_{mt} + \gamma_{m1}x_{1t} + \ldots + \gamma_{mk}x_{kt} = \varepsilon_{mt} \end{array} \qquad (9.16)$$

Переменные $y_1, \ldots, y_m$, определяемые внутри системы, называются *эндогенными*, в переменные $x_1, \ldots, x_k$ могут быть включены как внешние по отношению к системе (*экзогенные*) переменные, так и лагированные значения эндогенных переменных, которые называются *предопределенными* переменными.

Индекс t, как и раньше, означает номер наблюдения, $t = 1, \ldots, n$, а $\varepsilon_{1t}, \ldots, \varepsilon_{mt}$ — случайные ошибки. Будем считать, что в каждом уравнении один из коэффициентов β при какой-либо эндогенной переменной равен 1 — это естественное *условие нормировки*. Оно позволяет представить каждое уравнение в привычном виде, когда в левой части стоит одна эндогенная переменная, а в правой части — остальные переменные с неизвестными коэффициентами плюс случайная ошибка. Обозначим

$$\boldsymbol{y}_t = \begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{mt} \end{bmatrix}, \quad \boldsymbol{x}_t = \begin{bmatrix} x_{1t} \\ x_{2t} \\ \vdots \\ x_{kt} \end{bmatrix}, \quad \boldsymbol{\varepsilon}_t = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{mt} \end{bmatrix},$$

$$\mathbf{B} = \begin{bmatrix} \beta_{11} & \beta_{12} & \ldots & \beta_{1m} \\ \beta_{21} & \beta_{22} & \ldots & \beta_{2m} \\ \ldots & \ldots & \ldots & \ldots \\ \beta_{m1} & \beta_{m2} & \ldots & \beta_{mm} \end{bmatrix}, \quad \mathbf{\Gamma} = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \ldots & \gamma_{1k} \\ \gamma_{21} & \gamma_{22} & \ldots & \gamma_{2k} \\ \ldots & \ldots & \ldots & \ldots \\ \gamma_{m1} & \gamma_{m2} & \ldots & \gamma_{mk} \end{bmatrix}.$$

Тогда систему (9.16) можно переписать в следующем виде:

$$\mathbf{B}\boldsymbol{y}_t + \mathbf{\Gamma}\boldsymbol{x}_t = \boldsymbol{\varepsilon}_t. \qquad (9.17)$$

Подчеркнем, что деление переменных на экзогенные и эндогенные должно быть проведено вне модели. Одно из основных

требований к экзогенным переменным — некоррелированность векторов $\boldsymbol{x}_t$ и $\boldsymbol{\varepsilon}_t$ в каждом наблюдении t. Будем предполагать, что

1) $\mathrm{E}(\boldsymbol{\varepsilon}_t) = \mathbf{0}$;
2) $\mathrm{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \mathbf{\Sigma}$, причем матрица $\mathbf{\Sigma}$ не зависит от t и положительно определена;
3) при $t \neq s$ векторы $\boldsymbol{\varepsilon}_t$ и $\boldsymbol{\varepsilon}_s$ некоррелированы;
4) матрица $\mathbf{B}$ невырождена.

Используя условие 4), умножим обе части равенства (9.17) слева на $\mathbf{B}^{-1}$:

$$\boldsymbol{y}_t = -\mathbf{B}^{-1}\mathbf{\Gamma}\boldsymbol{x}_t + \mathbf{B}^{-1}\boldsymbol{\varepsilon}_t = \mathbf{\Pi}\boldsymbol{x}_t + \boldsymbol{\nu}_t, \qquad (9.18)$$

где $\mathbf{\Pi} = -\mathbf{B}^{-1}\mathbf{\Gamma}$, $\boldsymbol{\nu}_t = \mathbf{B}^{-1}\boldsymbol{\varepsilon}_t$. Система (9.18) называется *приведенной формой модели* (ср. п. 9.2, пример 1). Элементы матриц $\mathbf{B}$ и $\mathbf{\Gamma}$ в (9.17) иногда для краткости называют *структурными коэффициентами*, а элементы матрицы $\mathbf{\Pi}$ в (9.18) — *коэффициентами приведенной формы*.

Нетрудно понять, что в общем случае эндогенные переменные и ошибки в структурной системе коррелированы (пример 1 данной главы), поэтому, как уже неоднократно отмечалось, применение к какому-либо из уравнений метода наименьших квадратов даст смещенные и несостоятельные оценки структурных коэффициентов. В то же время коэффициенты приведенной формы могут быть состоятельно оценены, поскольку переменные $\boldsymbol{x}_t$ некоррелированы со структурными ошибками $\boldsymbol{\varepsilon}_t$ и, следовательно, с ошибками приведенной формы модели $\boldsymbol{\nu}_t$.

Мы не будем давать формального определения идентифицируемости структурной модели, а для более подробного ознакомления с этой проблемой можем рекомендовать, например (Greene, 1997, глава 16). Говоря же нестрого, тот или иной структурный коэффициент идентифицируем, если он может быть вычислен на основе коэффициентов приведенной формы. Соответственно какое-либо уравнение в структурной форме модели будет называться идентифицируемым, если идентифицируемы все его коэффициенты. Подчеркнем еще раз, что

проблема идентифицируемости логически предшествует задаче оценивания, отсутствие идентифицируемости означает, что существует бесконечно много моделей, совместимых с имеющимися наблюдениями, и это никак не связано с количеством наблюдений.

Приведенная форма (9.18) позволяет состоятельно оценить mk элементов матрицы $\mathbf{\Pi}$ и $m(m+1)/2$ элементов матрицы ковариаций вектора ошибок $\boldsymbol{\nu}$. В то же время в структурной форме неизвестными являются $m^2 - m$ элементов матрицы $\mathbf{B}$ (условие нормировки), mk элементов матрицы $\mathbf{\Gamma}$ и $m(m+1)/2$ элементов матрицы ковариаций вектора ошибок $\boldsymbol{\varepsilon}$. Таким образом, превышение числа структурных коэффициентов над числом коэффициентов приведенной формы есть $m^2 - m$ и, следовательно, в общем случае система неидентифицируема. Однако, как было показано ранее (пример 1 данной главы), некоторые структурные коэффициенты или структурные уравнения могут быть идентифицированы. Основная причина этого — наличие априорных ограничений на структурные коэффициенты.

Мы будем изучать лишь задачу идентифицируемости отдельного уравнения в том случае, когда ограничения имеют наиболее простой вид, а именно, часть структурных коэффициентов равна 0. Более полное изложение проблем, связанных с идентифицируемостью модели, содержится, например, в (Greene, 1997, глава 16).

Для определенности рассмотрим задачу идентифицируемости первого уравнения системы (9.16) при условии, что какие-то структурные коэффициенты равны 0, т.е. из уравнения исключены некоторые переменные, и идентифицируемость будем понимать как возможность вычисления структурных коэффициентов уравнения по коэффициентам приведенной формы.

Для удобства изложения далее мы объединим в одну группу экзогенные и предопределенные переменные и будем называть их просто экзогенными переменными.

Без ограничения общности можно считать, что первые q коэффициентов при эндогенных переменных и первые p коэффициентов при экзогенных переменных не равны 0, а остальные коэффициенты — нулевые. Тогда первое уравнение в (9.16) (опуская индекс t) можно переписать следующим образом:

$$[\beta_{11}, \ldots, \beta_{1q}, 0, \ldots, 0] \begin{bmatrix} y_1 \\ \vdots \\ y_q \\ \vdots \\ y_m \end{bmatrix} + [\gamma_{11}, \ldots, \gamma_{1p}, 0, \ldots, 0] \begin{bmatrix} x_1 \\ \vdots \\ x_p \\ \vdots \\ x_k \end{bmatrix} = \varepsilon_1. \tag{9.19}$$

Обозначим

$$\boldsymbol{y}_* = \begin{bmatrix} y_1 \\ \vdots \\ y_q \end{bmatrix}, \quad \boldsymbol{y}_{**} = \begin{bmatrix} y_{q+1} \\ \vdots \\ y_m \end{bmatrix}, \quad \boldsymbol{x}_\times = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}, \quad \boldsymbol{x}_{\times\times} = \begin{bmatrix} x_{p+1} \\ \vdots \\ x_k \end{bmatrix},$$
$$\boldsymbol{\beta}_* = [\beta_{11}, \ldots, \beta_{1q}]', \quad \boldsymbol{\gamma}_\times = [\gamma_{11}, \ldots, \gamma_{1p}]'. \tag{9.20}$$

Тогда уравнение (9.19) записывается кратко так:

$$\boldsymbol{\beta}_*' \boldsymbol{y}_* + \boldsymbol{\gamma}_\times' \boldsymbol{x}_\times = \varepsilon_1.$$

В соответствии с разбиением (9.20) представим $m \times k$ матрицу $\mathbf{\Pi}$ в (9.18) в блочном виде

$$\mathbf{\Pi} = \begin{bmatrix} \mathbf{\Pi}_{*,\times} & \mathbf{\Pi}_{*,\times\times} \\ \mathbf{\Pi}_{**,\times} & \mathbf{\Pi}_{**,\times\times} \end{bmatrix}$$

и заметим, что приведенная форма модели (9.18) в наших обозначениях имеет вид

$$\begin{bmatrix} \boldsymbol{y}_* \\ \boldsymbol{y}_{**} \end{bmatrix} = \begin{bmatrix} \mathbf{\Pi}_{*,\times} & \mathbf{\Pi}_{*,\times\times} \\ \mathbf{\Pi}_{**,\times} & \mathbf{\Pi}_{**,\times\times} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_\times \\ \boldsymbol{x}_{\times\times} \end{bmatrix} + \boldsymbol{\nu}.$$

Напомним, что $\mathbf{B\Pi} = -\mathbf{\Gamma}$, и тогда для первых строк этих матриц получаем равенство

$$\begin{bmatrix} \boldsymbol{\beta}_* \\ \mathbf{0}_{m-q} \end{bmatrix}' \begin{bmatrix} \mathbf{\Pi}_{*,\times} & \mathbf{\Pi}_{*,\times\times} \\ \mathbf{\Pi}_{**,\times} & \mathbf{\Pi}_{**,\times\times} \end{bmatrix} = - \begin{bmatrix} \boldsymbol{\gamma}_\times \\ \mathbf{0}_{k-p} \end{bmatrix}',$$

где $\mathbf{0}_r$ — нулевой вектор-столбец размерности r. По правилу действия с блочными матрицами имеем

$$\boldsymbol{\beta}_*'\mathbf{\Pi}_{*,\times} = -\boldsymbol{\gamma}_\times', \tag{9.21}$$

$$\boldsymbol{\beta}_*'\mathbf{\Pi}_{*,\times\times} = \mathbf{0}_{k-p}'. \tag{9.22}$$

Соотношение (9.22) является (линейной) системой $k-p$ уравнений (относительно $\boldsymbol{\beta}_*$) с $q-1$ неизвестными, поскольку в силу условия нормировки один из элементов вектора $\boldsymbol{\beta}_*$ равен 1. Очевидно, что если коэффициенты $\boldsymbol{\beta}_*$ найдены, то коэффициенты $\boldsymbol{\gamma}_\times$ определяются равенством (9.21). Ясно также, что для того чтобы параметры $\boldsymbol{\beta}_*$ в системе (9.22) можно было бы как-нибудь выразить через элементы матрицы $\mathbf{\Pi}_{*,\times\times}$, необходимо, чтобы число уравнений в (9.22) было не меньше числа неизвестных, т. е. выполнялось неравенство

$$k-p \geqslant q-1. \tag{9.23}$$

Иными словами, число исключенных из уравнения экзогенных переменных должно быть не меньше числа включенных эндогенных переменных минус единица. Неравенство (9.23) носит название *порядковое условие* (*order condition*) и является лишь *необходимым* условием идентифицируемости уравнения, поскольку даже при его выполнении уравнения в (9.22) могут оказаться линейно зависимыми. Из общей теории систем линейных уравнений известно, что для разрешимости системы (9.22) необходимо и достаточно, чтобы матрица $\mathbf{\Pi}_{*,\times\times}$ имела ранг $q-1$:

$$\operatorname{rank}(\mathbf{\Pi}_{*,\times\times}) = q-1. \tag{9.24}$$

Это равенство называется *ранговым условием* (*rank condition*), и оно является необходимым и достаточным для идентифицируемости уравнения.

Если условие (9.23) выполняется со знаком равенства, то говорят, что уравнение *точно идентифицируемо* (*exactly identified*), если со знаком строгого неравенства, то *сверхидентифицируемо* (*overidentified*). В последнем случае число уравнений превышает число неизвестных, и некоторые из структурных коэффициентов $\boldsymbol{\beta}_*$ могут быть выражены разным способом через коэффициенты матрицы $\mathbf{\Pi}_{*,\times\times}$, как в примере 2 данной главы.

Оценивание систем одновременных уравнений. Двухшаговый метод наименьших квадратов

Как мы выяснили в предыдущем разделе, приступать к оцениванию того или иного уравнения в системе (9.16) имеет смысл лишь после того, как установлена его идентифицируемость. Как и раньше, будем рассматривать для определенности первое уравнение и предположим, что оно содержит q эндогенных и p экзогенных переменных и идентифицируемо (в частности, выполнено порядковое условие (9.23)), при этом без ограничения общности можно считать, что коэффициент при y_1 равен 1.

1. *Косвенный метод наименьших квадратов.* При идентифицируемости уравнения оценки структурных коэффициентов можно найти, оценив методом наименьших квадратов приведенную форму модели (9.18), а затем решив систему (9.21), (9.22), заменяя элементы матрицы $\mathbf{\Pi}$ их оценками. Этот способ носит название *косвенный метод наименьших квадратов* (*Indirect Least Squares*) (см. п. 9.2, пример 1). В силу теоремы Слуцкого полученные оценки являются состоятельными, поскольку состоятельны оценки коэффициентов приведенной формы модели. Однако у этого метода есть серьезный недостаток: если уравнение сверхидентифицируемо, то в системе (9.22) число уравнений превышает число неизвестных, а это значит, что один и тот же структурный коэффициент допускает разные выражения через коэффициенты приведенной формы. Это сужает область его применения как с теоретической (не ясно, какую же оценку следует предпочесть), так и с практической точки зрения (трудность алгоритмизации). Поэтому, как правило, используется

2. *Двухшаговый метод наименьших квадратов.* Представим первое уравнение в следующем виде:

$$\begin{aligned} y_{1t} = &-\beta_{12}y_{2t} - \ldots - \beta_{1q}y_{qt} \\ &- \gamma_{11}x_{1t} - \gamma_{12}x_{2t} - \ldots - \gamma_{1p}x_{pt} + \varepsilon_{1t}, \quad t = 1, \ldots, n. \end{aligned} \tag{9.25}$$

Обозначим

$$\boldsymbol{y}_1 = \begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n} \end{bmatrix}, \quad \boldsymbol{Y}_1 = \begin{bmatrix} y_{21} & \dots & y_{q1} \\ y_{22} & \dots & y_{q2} \\ \dots & \dots & \dots \\ y_{2n} & \dots & y_{qn} \end{bmatrix}, \quad \boldsymbol{X}_1 = \begin{bmatrix} x_{11} & \dots & x_{p1} \\ x_{12} & \dots & x_{p2} \\ \dots & \dots & \dots \\ x_{1n} & \dots & x_{pn} \end{bmatrix},$$

$$\boldsymbol{\beta}_1 = \begin{bmatrix} -\beta_{12} \\ \vdots \\ -\beta_{1q} \end{bmatrix}, \ \boldsymbol{\gamma}_1 = \begin{bmatrix} -\gamma_{11} \\ \vdots \\ -\gamma_{1p} \end{bmatrix}, \ \boldsymbol{\varepsilon}_1 = \begin{bmatrix} \varepsilon_{11} \\ \vdots \\ \varepsilon_{1n} \end{bmatrix}, \ \boldsymbol{X} = \begin{bmatrix} x_{11} & \dots & x_{k1} \\ x_{12} & \dots & x_{k2} \\ \dots & \dots & \dots \\ x_{1n} & \dots & x_{kn} \end{bmatrix},$$

и перепишем (9.25) в виде

$$\boldsymbol{y}_1 = \boldsymbol{Y}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_1\boldsymbol{\gamma}_1 + \boldsymbol{\varepsilon}_1. \tag{9.26}$$

Поскольку элементы матрицы $\boldsymbol{Y}_1$ коррелированы с вектором ошибок $\boldsymbol{\varepsilon}_1$, непосредственное применение метода наименьших квадратов приведет к смещенным и несостоятельным оценкам. В таких ситуациях, как мы знаем, целесообразно воспользоваться инструментальными переменными. В качестве инструментов можно взять, например, экзогенные переменные, не содержащиеся в $\boldsymbol{X}_1$ (очевидно, что использовать в качестве инструментов только переменные, содержащиеся в $\boldsymbol{X}_1$, нельзя, поскольку в системе возникнет полная коллинеарность). Известно (глава 8), что число инструментальных переменных должно быть не меньше, чем число переменных в $\boldsymbol{Y}_1$, т. е. $q-1$. В нашем случае это требование выполнено благодаря порядковому условию (9.23). Очевидно также, что в случае сверхидентификации уравнения возможны разные наборы инструментальных переменных. Наиболее распространенным способом выбора инструментальных переменных является *двухшаговый метод наименьших квадратов* (*Two Stage Least Squares*, 2SLS), который устроен следующим образом:

1) проводится регрессия каждого столбца матрицы $\boldsymbol{Y}_1$ на *все* экзогенные переменные, т. е. рассматривается регрессия

$$\boldsymbol{Y}_1 = \boldsymbol{X}\boldsymbol{\Pi}_1 + \boldsymbol{V},$$

где $\boldsymbol{\Pi}_1$ — $k \times (q-1)$ матрица коэффициентов приведенной формы;

2) строится прогнозное значение $\widehat{\boldsymbol{Y}}_1 = \boldsymbol{X}\widehat{\boldsymbol{\Pi}}_1$, где $\widehat{\boldsymbol{\Pi}}_1 = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}_1$;

3) осуществляется регрессия (9.26) с заменой в правой части $\boldsymbol{Y}_1$ на $\widehat{\boldsymbol{Y}}_1$, т. е. строятся МНК-оценки структурных параметров $\boldsymbol{\beta}_1$ и $\boldsymbol{\gamma}_1$ в регрессии

$$\boldsymbol{y}_1 = \widehat{\boldsymbol{Y}}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_1\boldsymbol{\gamma}_1 + \boldsymbol{\varepsilon}_1. \tag{9.27}$$

Мы не ставим целью в данной книге дать подробное описание свойств двухшагового метода наименьших квадратов. Перечислим без доказательства лишь основные результаты, касающиеся этого способа оценивания.

1) Если для уравнения выполнено ранговое условие идентификации и порядковое условие (9.23) выполнено со знаком равенства (точная идентификация), то 2SLS-оценка совпадает с оценкой, полученной косвенным методом наименьших квадратов.

2) 2SLS-оценка совпадает с оценкой, полученной методом инструментальных переменных, когда в качестве последних используются $\widehat{\boldsymbol{Y}}_1$ и $\boldsymbol{X}_1$.

3) Пусть в качестве инструментальных переменных для замены $\boldsymbol{Y}_1$ выбраны любые линейные комбинации столбцов матрицы $\boldsymbol{X}$. Тогда матрица ковариаций этой оценки не меньше, чем матрица ковариаций 2SLS-оценки.

Последнее свойство означает эффективность 2SLS-оценки в соответствующем классе оценок. Отметим, что в большинстве эконометрических компьютерных пакетов для оценивания систем одновременных уравнений реализован именно двухшаговый метод наименьших квадратов. В заключение подчеркнем, что при использовании двухшагового метода наименьших квадратов фактически каждое уравнение оценивается независимо от других. Су-

ществует так называемый *трехшаговый метод наименьших квадратов*, который учитывает взаимодействие уравнений в системе, что приводит к повышению эффективности оценки, однако его описание выходит за рамки данной книги.

Завершим эту главу описанием классической макроэкономической модели Клейна и результатов ее оценивания с помощью обычного и двухшагового метода наименьших квадратов.

Пример 6. *Модель Клейна 1.* В 1950 г. Л. Клейн предложил динамическую модель макроэкономики, получившую название *модель Клейна 1*. Она описывается следующей системой уравнений:

$$C_t = \alpha_0 + \alpha_1 P_t + \alpha_2 P_{t-1} + \alpha_3 (W_t^p + W_t^g) + \varepsilon_{1t} \qquad \text{(потребление)},$$

$$I_t = \beta_0 + \beta_1 P_t + \beta_2 P_{t-1} + \beta_3 K_{t-1} + \varepsilon_{2t} \qquad \text{(инвестиции)},$$

$$W_t^p = \gamma_0 + \gamma_1 X_t + \gamma_2 X_{t-1} + \gamma_3 A_t + \varepsilon_{3t} \qquad \text{(зарплата в частном секторе)},$$

$$X_t = C_t + I_t + G_t \qquad \text{(совокупный спрос в равновесии)},$$

$$P_t = X_t - T_t - W_t^p \qquad \text{(доход частного сектора)},$$

$$K_t = K_{t-1} + I_t \qquad \text{(капитал)}.$$

Переменные, стоящие в левых частях уравнений, являются эндогенными. Экзогенными переменными в данной модели являются: G — государственные расходы, не включающие зарплату, T — непрямые налоги плюс чистый доход от экспорта, W^g — зарплата в государственном секторе, A_t — временной тренд (в годах, начиная с 1931 г.). Кроме того, включены три предопределенные (лагированные) переменные. Таким образом, модель содержит три поведенческих уравнения, одно уравнение равновесия и два тождества.

Приведем результаты оценивания первых трех уравнений на основе ежегодных данных для экономики США за период с 1921 по 1941 г. с помощью обычного метода наименьших квадратов и двухшагового метода наименьших квадратов (в скобках указаны оценки стандартных ошибок).

Обычный метод наименьших квадратов:

$$C_t = \underset{(1.30)}{16.2} + \underset{(0.091)}{0.193} P_t + \underset{(0.091)}{0.090} P_{t-1} + \underset{(0.040)}{0.796} (W_t^p + W_t^g),$$

$$I_t = \underset{(5.47)}{10.1} + \underset{(0.097)}{0.480} P_t + \underset{(0.101)}{0.333} P_{t-1} - \underset{(0.027)}{0.112} K_{t-1},$$

$$W_t^p = \underset{(1.27)}{1.48} + \underset{(0.032)}{0.439} X_t + \underset{(0.037)}{0.146} X_{t-1} + \underset{(0.032)}{0.130} A_t.$$

Двухшаговый метод наименьших квадратов:

$$C_t = \underset{(1.32)}{16.6} + \underset{(0.118)}{0.017} P_t + \underset{(0.107)}{0.216} P_{t-1} + \underset{(0.040)}{0.810} (W_t^p + W_t^g),$$

$$I_t = \underset{(7.54)}{20.3} + \underset{(0.173)}{0.150} P_t + \underset{(0.162)}{0.616} P_{t-1} - \underset{(0.036)}{0.158} K_{t-1},$$

$$W_t^p = \underset{(1.15)}{1.50} + \underset{(0.036)}{0.439} X_t + \underset{(0.039)}{0.147} X_{t-1} + \underset{(0.029)}{0.130} A_t.$$

Упражнения

9.1. Рассмотрим следующую модель:

$$\begin{cases} C_t = \alpha + \beta Y_t + \varepsilon_{1t}, \\ Y_t = C_t + I_t + G_t, \\ I_t = \gamma + \delta Y_t + \varepsilon_{2t}. \end{cases}$$

Эндогенные переменные — C_t, Y_t, I_t, экзогенная переменная — G_t. Напишите эту модель в матричной форме и найдите ее приведенную форму. Сколько ограничений накладывается на шесть коэффициентов приведенной формы модели и каковы эти ограничения? Покажите, что при заданных значениях коэффициентов приведенной формы можно единственным образом получить значения коэффициентов α, β, γ и δ, т. е. при заданной матрице $\mathbf{\Pi}$ уравнение $\mathbf{B\Pi} + \mathbf{\Gamma}$ имеет единственное решение относительно $\mathbf{B}$ и $\mathbf{\Gamma}$.

9.2. Рассмотрим проблему идентифицируемости каждого из уравнений в следующей модели:

$$\begin{cases} P_t + \beta_{12} W_t + \gamma_{11} Q_t + \gamma_{13} P_{t-1} = \varepsilon_{1t}, \\ \beta_{21} P_t + W_t + \beta_{23} N_t + \gamma_{22} S_t + \gamma_{24} W_{t-1} = \varepsilon_{2t}, \\ \beta_{32} W_t + N_t + \gamma_{32} S_t + \gamma_{33} P_{t-1} + \gamma_{34} W_{t-1} = \varepsilon_{3t}, \end{cases}$$

где P_t, W_t, N_t — индекс цен, зарплата, профсоюзный взнос соответственно (эндогенные переменные), а Q_t и S_t — производительность труда и количество забастовок (экзогенные переменные). Как выглядят порядковое и ранговое условия, если известно, что:

а) $\gamma_{11} = 0$,

б) $\beta_{21} = \gamma_{22} = 0$,

в) $\gamma_{33} = 0$?

9.3. Опишите процедуру оценивания каждого из уравнений следующей системы:

$$\begin{cases} y_{1t} + \beta_{12}y_{2t} + \gamma_{11} + \gamma_{12}x_{2t} = \varepsilon_{1t}, \\ y_{2t} + \gamma_{21} + \gamma_{23}x_{3t} = \varepsilon_{2t}, \\ \beta_{32}y_{2t} + y_{3t} + \gamma_{31} + \gamma_{33}x_{3t} = \varepsilon_{3t}. \end{cases}$$

9.4. Рассматривается следующая система уравнений:

$$\begin{cases} y_{1t} = \gamma_{10} + \beta_{12}y_{2t} + \beta_{13}y_{3t} + \gamma_{11}x_{1t} + \gamma_{12}x_{2t} + \varepsilon_{1t}, \\ y_{2t} = \gamma_{20} + \beta_{21}y_{1t} + \gamma_{21}x_{1t} + \varepsilon_{2t}, \\ y_{3t} = \gamma_{30} + \beta_{31}y_{1t} + \beta_{32}y_{2t} + \gamma_{31}x_{1t} + \gamma_{33}x_{3t} + \varepsilon_{3t}. \end{cases}$$

Идентифицируемо ли каждое из уравнений системы? Что получится, если применить к первому уравнению двухшаговый метод наименьших квадратов?

9.5. Задана система одновременных уравнений (y_1, y_2, y_3 — эндогенные переменные).

$$\begin{cases} y_{1t} = \gamma_{10} + \beta_{12}y_{2t} + \gamma_{11}x_{1t} + \varepsilon_{1t}, \\ y_{2t} = \gamma_{20} + \beta_{23}y_{3t} + \gamma_{21}x_{1t} + \gamma_{23}x_{3t} + \varepsilon_{2t}, \\ y_{3t} = \beta_{31}y_{1t} + \beta_{32}y_{2t} + \gamma_{31}x_{1t} + \gamma_{32}x_{2t} + \gamma_{33}x_{3t} + \varepsilon_{3t}, \end{cases}$$

а) Для каждого из трех уравнений определите, выполняются ли порядковые и ранговые условия идентифицируемости.

б) Повторите а) при дополнительном ограничении: $\gamma_{32} = 0$.

в) Повторите а) при дополнительном ограничении: $\gamma_{32} = 1$.

г) Повторите а) при дополнительном ограничении: $\gamma_{32} = \gamma_{33}$.

9.6. Рассматривается модель, состоящая из двух внешне не связанных уравнений (SUR):

$$\begin{cases} y_{1t} = \beta_1 + \varepsilon_1, \\ y_{2t} = \beta_2 x_t + \varepsilon_2. \end{cases}$$

По 50 наблюдениям (по каждому уравнению) получены следующие результаты: $\sum x_t = 100$, $\sum x_t^2 = 600$, $\sum x_t y_{t1} = 60$, $\sum x_t y_{t2} = 50$, $\sum y_{t1} = 150$, $\sum y_{t1}^2 = 500$, $\sum y_{t1}y_{t2} = 40$, $\sum y_{t2} = 50$, $\sum y_{t2}^2 = 90$.

а) Напишите формулу для GLS-оценки параметров β_1, β_2.

б) Найдите OLS-оценку этих параметров.

в) Найдите SUR (FGLS)-оценку этих параметров и оцените матрицы ковариаций этих оценок.

Глава 10

Метод максимального правдоподобия в моделях регрессии

Данная глава несколько отличается от других глав. Разделы 10.1–10.4 фактически содержат справочный материал по методу максимального правдоподобия, широко применяемому в математической статистике. Подробное изложение этого материала можно найти, например, в (Айвазян (1983), Крамер (1975), Рао (1968)). Раздел 10.5 во многом повторяет описанные кратко в разделах 2.7, 5.3 и приложении МС (п. 7) способы применения этого метода к моделям парной и множественной регрессии. Причина, по которой мы поместили этот материал не в приложении МС, а здесь, состоит в следующем. Первое, метод максимального правдоподобия является традиционно трудным для студентов разделом курса математической статистики, и его, по нашему мнению, следует повторить в курсе эконометрики, включающем в себя темы временных рядов и дискретных зависимых переменных, в которых этот метод интенсивно используется. Второе, удобство читателя, для которого все необходимые факты по методу максимального правдоподобия собраны в одном месте книги.

10.1. Введение

Принцип *максимального правдоподобия* (*maximum likelihood, ML*) уже использовался в нашей книге в главе 2 (п. 2.7) для случая парной регрессии и в главе 5 (п. 5.3) для случая множественной регрессии с нормальным распределением вектора ошибок. Краткое описание метода также можно найти в приложении (см. МС, п. 7). В данной главе мы дадим более подробное описание метода максимального правдоподобия. Начнем с простого примера.

Предположим, у нас есть выборка из биномиального распределения $B(n,p)$, где $n = 10$, а вероятность p неизвестна. Выборка состоит из 7 единиц и 3 нулей, не обязательно в этом порядке. Вероятность 7 успехов в 10 испытаниях равна (см. МС, п. 3)

$$h(p) = C_{10}^{7}\, p^{7}(1-p)^{3}. \tag{10.1}$$

Для того чтобы найти значение p, максимизирующее (10.1), мы вычисляем производную логарифма (10.1)

$$\frac{d\ln h(p)}{dp} = \frac{7}{p} - \frac{3}{1-p}, \tag{10.2}$$

и приравниваем ее нулю. Получаем $\widehat{p} = 0.7$, значение параметра p, при котором вероятность получения такой выборки максимальна.

Рассмотрим более общий случай. Пусть у нас есть n наблюдений, $(y_1, \ldots, y_n)$, где все y_i равны 1 или 0. Вероятность получения в точности x успехов в n испытаниях равна

$$C_n^x\, p^x(1-p)^{n-x}. \tag{10.3}$$

Это выражение рассматривается обычно как функция x при заданных значениях параметров n, p и называется *распределением* (*distribution*). В отличие от этого, в методе максимального правдоподобия мы рассматриваем (10.3) как функцию p (предполагаем сейчас, что n известно), при данном x (из наблюденной выборки), и называем (10.3) *функцией правдоподобия* (*likelihood function*).

Метод максимального правдоподобия является *конструктивным* методом. В простых случаях, подобно приведенному выше,

удается получить явную формулу для оценки. В более сложных случаях получить явную формулу не удается, однако можно определить численное значение оценки, максимизирующее функцию правдоподобия. Но и в этой ситуации можно многое сказать о статистических свойствах оценки.

10.2. Математический аппарат

Рассмотрим последовательность случайных величин $\{y_1, y_2, \ldots\}$, не обязательно независимых или одинаково распределенных. Пусть $h_n(\cdot, \boldsymbol{\theta}_0)$ — совместная плотность распределения случайных величин $\boldsymbol{y} = (y_1, \ldots, y_n)$. Предположим, что вид этой функции известен, за исключением вектора параметров $\boldsymbol{\theta}_0$, который мы хотим оценить. Мы предполагаем, что $\boldsymbol{\theta}_0 \in \Theta$, где множество возможных значений параметра Θ принадлежит конечномерному евклидову пространству.

Для каждого (фиксированного) $\boldsymbol{y}$ вещественная функция

$$L_n(\boldsymbol{\theta}) = L_n(\boldsymbol{\theta}, \boldsymbol{y}) = h_n(\boldsymbol{y}, \boldsymbol{\theta}), \qquad \boldsymbol{\theta} \in \Theta, \tag{10.4}$$

называется *функцией правдоподобия* (*likelihood function*), и ее логарифм $\ln L_n(\boldsymbol{\theta})$ называется *логарифмической функцией правдоподобия* (*loglikelihood function*).

Для фиксированного $\boldsymbol{y}$ любое значение $\widehat{\boldsymbol{\theta}}_n(\boldsymbol{y}) \in \Theta$, такое, что

$$L_n(\widehat{\boldsymbol{\theta}}_n(\boldsymbol{y}), \boldsymbol{y}) = \sup_{\boldsymbol{\theta} \in \Theta} L_n(\boldsymbol{\theta}, \boldsymbol{y}), \tag{10.5}$$

называется *оценкой максимального правдоподобия* (*maximum likelihood, ML, estimate*) параметра $\boldsymbol{\theta}_0$. В общем случае нет гарантии, что ML-оценка параметра $\boldsymbol{\theta}_0$ существует для (почти) всех значений $\boldsymbol{y}$, но если это верно, функция $\widehat{\boldsymbol{\theta}}_n$ называется *оценкой максимального правдоподобия* (*ML estimator*) неизвестного параметра $\boldsymbol{\theta}_0$.

Когда максимум в (10.5) достигается во внутренней точке пространства параметров Θ, а $L_n(\boldsymbol{\theta})$ является дифференцируемой (по $\boldsymbol{\theta}$) функцией, то вектор частных производных $\partial \ln L_n(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$

в этой точке равен нулю. Тогда $\widehat{\boldsymbol{\theta}}_n$ является решением векторного уравнения

$$\frac{\partial \ln L_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}.$$

В случае, когда $L_n(\boldsymbol{\theta})$ дважды дифференцируема по $\boldsymbol{\theta}$, гессиан (*Hessian matrix*) определяется как

$$\boldsymbol{H}_n(\boldsymbol{\theta}) = \frac{\partial^2 \ln L_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}'} \tag{10.6}$$

и *информационная матрица* в точке $\boldsymbol{\theta}_0$ равна

$$\boldsymbol{\mathcal{F}}_n(\boldsymbol{\theta}_0) = -\mathrm{E}\big(\boldsymbol{H}_n(\boldsymbol{\theta}_0)\big). \tag{10.7}$$

Заметим, что информационная матрица вычислена в точке истинного значения параметра $\boldsymbol{\theta}_0$. *Асимптотическая информационная матрица* для параметра $\boldsymbol{\theta}_0$ определяется как

$$\boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0) = \lim_{n \to \infty} (1/n) \boldsymbol{\mathcal{F}}_n(\boldsymbol{\theta}_0), \tag{10.8}$$

если предел существует. Если матрица $\boldsymbol{\mathcal{F}}(\boldsymbol{\theta}_0)$ положительно определена, то обратная к ней матрица $\boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{\theta}_0)$ является нижней границей для асимптотической матрицы ковариаций любой состоятельной оценки параметра $\boldsymbol{\theta}_0$ (асимптотическое неравенство Рао–Крамера) (ср. МС, п. 7). При некоторых условиях регулярности матрица ковариаций оценки максимального правдоподобия асимптотически приближается к этой нижней границе. Вследствие этого $\boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{\theta}_0)$ называется *асимптотической матрицей ковариаций* оценки максимального правдоподобия $\widehat{\boldsymbol{\theta}}$. Точный смысл последнего утверждения состоит в том, что при некоторых условиях регулярности последовательность случайных векторов

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \tag{10.9}$$

сходится по распределению к нормально распределенному случайному вектору с нулевым средним и матрицей ковариаций $\boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{\theta}_0)$.

10.3. Оценка максимального правдоподобия параметров многомерного нормального распределения

Рассмотрим теперь выборку из многомерного нормального распределения с математическим ожиданием $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)'$ и матрицей ковариаций $\sigma^2 \boldsymbol{I}_m$. У нас есть n наблюдений (случайных векторов) $(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$. Функция плотности для каждого наблюдения равна

$$h(\boldsymbol{y}_i) = (2\pi)^{-m/2}(\sigma^2)^{-m/2} \exp\left\{-\frac{1}{2}\left(\frac{(\boldsymbol{y}_i - \boldsymbol{\mu})'(\boldsymbol{y}_i - \boldsymbol{\mu})}{\sigma^2}\right)\right\}. \quad (10.10)$$

Взяв сумму логарифмов этих выражений по всем n наблюдениям, получаем логарифмическую функцию правдоподобия:

$$\ln L = -\frac{nm}{2}\ln 2\pi - \frac{nm}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_i (\boldsymbol{y}_i - \boldsymbol{\mu})'(\boldsymbol{y}_i - \boldsymbol{\mu}). \quad (10.11)$$

Дифференцируя эту функцию по параметрам распределения $\boldsymbol{\mu}$ и σ^2, получаем следующие уравнения (необходимые условия экстремума):

$$\frac{\partial \ln L}{\partial \boldsymbol{\mu}} = \frac{1}{\sigma^2}\sum_i (\boldsymbol{y}_i - \boldsymbol{\mu}) = 0, \quad (10.12)$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{nm}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_i (\boldsymbol{y}_i - \boldsymbol{\mu})'(\boldsymbol{y}_i - \boldsymbol{\mu}). \quad (10.13)$$

Решая систему (10.12)–(10.13), получаем оценки максимального правдоподобия

$$\widehat{\mu}_j = \overline{y}_j \quad \text{и} \quad \widehat{\sigma}^2 = \frac{\sum_j \sum_i (y_{ji} - \overline{y}_j)^2}{nm}. \quad (10.14)$$

Здесь y_{ji} обозначает j-ю компоненту вектора $\boldsymbol{y}_i$, а $\overline{y}_j = (1/n)\sum_{i=1}^n y_{ji}$.

10.4. Свойства оценок максимального правдоподобия

Оценки максимального правдоподобия привлекательны благодаря своим *асимптотическим* свойствам. При выполнении весьма общих предположений оценки максимального правдоподобия обладают следующими четырьмя свойствами.

1. *Инвариантность.* Пусть $\widehat{\boldsymbol{\theta}}$ — оценка максимального правдоподобия параметра $\boldsymbol{\theta}$ и $\boldsymbol{g}(\boldsymbol{\theta})$ — непрерывная функция. Тогда $\boldsymbol{g}(\widehat{\boldsymbol{\theta}})$ является оценкой максимального правдоподобия параметра $\boldsymbol{g}(\boldsymbol{\theta})$.

Например, из того, что оценка $\widehat{\sigma}^2$ является оценкой максимального правдоподобия параметра σ^2 (см. (10.13)), сразу вытекает, что $\widehat{\sigma}$ и $1/\widehat{\sigma}$ являются оценками максимального правдоподобия для σ и $1/\sigma$ соответственно. Из свойства инвариантности вытекает, в частности, что оценки максимального правдоподобия, в общем случае не являются несмещенными (почему?). Из этого свойства следует также, что мы можем параметризовать функцию правдоподобия любым способом, что часто существенно облегчает вычисление оценки.

Свойство инвариантности выполняется для конечных выборок, в то время как следующие три свойства оценок максимального правдоподобия являются асимптотическими.

2. *Состоятельность.* Выполнено равенство: $p\lim \widehat{\boldsymbol{\theta}} = \boldsymbol{\theta}$, т. е. $\mathrm{P}(|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}| > \varepsilon) \to 0$ при $n \to \infty$ для всякого $\varepsilon > 0$. Состоятельность рассматривается обычно как самое важное свойство оценки. Все наиболее часто встречающиеся оценки состоятельны. Другими словами, состоятельность — минимальное требование, предъявляемое к любой оценке.

Из того, что оценка $\widehat{\boldsymbol{\theta}}$ состоятельна, следует (см. МС, п. 5), что она сходится к истинному значению параметра по распределению: $\widehat{\boldsymbol{\theta}} \xrightarrow{d} \boldsymbol{\theta}$. Это означает, что предельное распределение сосредоточено в точке $\boldsymbol{\theta}$, или, другими словами, предельное распределение $(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ сосредоточено в точке $\mathbf{0}$. К сожалению, это не очень информативно. В этом случае для детального рассмотрения ситуа-

ции используют подходящую нормировку оценки так, чтобы нормированная оценка имела предельное невырожденное распределение. Рассмотрим частный случай. Пусть $(y_1, \ldots, y_n)$ — выборка из нормального распределения $N(\mu, \sigma^2)$. В этом случае $\overline{y}$ является состоятельной оценкой μ. При этом среднее $\overline{y} \sim N(\mu, \sigma^2/n)$ и, следовательно, $\sqrt{n}(\overline{y} - \mu) \sim N(0, \sigma^2)$, поэтому $\sqrt{n}(\overline{y} - \mu) \xrightarrow{d} N(0, \sigma^2)$. В этом частном случае распределение нормированной оценки не зависит от n и равно предельному, а в общем случае мы имеем следующее свойство асимптотической нормальности.

3. ***Асимптотическая нормальность.*** Имеет место следующее асимптотическое поведение оценки максимального правдоподобия:

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{\theta})), \tag{10.15}$$

где $\boldsymbol{\mathcal{F}}(\boldsymbol{\theta})$ обозначает асимптотическую информационную матрицу. Следует отметить, что $\boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{\theta})$ является матрицей ковариаций асимптотического распределения $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$; оценка матрицы ковариаций собственно оценки $\widehat{\boldsymbol{\theta}}$ равна $\widehat{\mathrm{V}}(\widehat{\boldsymbol{\theta}}) = \boldsymbol{\mathcal{F}}_n^{-1}(\widehat{\boldsymbol{\theta}}) = (1/n)\boldsymbol{\mathcal{F}}^{-1}(\widehat{\boldsymbol{\theta}})$.

4. ***Асимптотическая эффективность.*** Оценка максимального правдоподобия асимптотически эффективна. Это означает, что если мы сравним оценку максимального правдоподобия $\widehat{\boldsymbol{\theta}}$ с любой другой оценкой $\widetilde{\boldsymbol{\theta}}$, также состоятельной и асимптотически нормальной, то $\mathrm{V}(\widehat{\boldsymbol{\theta}}) \leqslant \mathrm{V}(\widetilde{\boldsymbol{\theta}})$, т.е. разность $\mathrm{V}(\widetilde{\boldsymbol{\theta}}) - \mathrm{V}(\widehat{\boldsymbol{\theta}})$ является неотрицательно определенной матрицей. В частности, это означает, что дисперсия каждой из компонент вектора $\widetilde{\boldsymbol{\theta}}$ не меньше дисперсии соответствующей компоненты вектора $\widehat{\boldsymbol{\theta}}$, т.е. оценка максимального правдоподобия $\widehat{\boldsymbol{\theta}}$ «лучше» оценки $\widetilde{\boldsymbol{\theta}}$.

10.5. Оценка максимального правдоподобия в линейной модели

Рассмотрим стандартную линейную модель (см. главу 3) с нормально распределенными ошибками

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}, \qquad \boldsymbol{u} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I}_n), \tag{10.16}$$

или ее эквивалентную форму записи:

$$\boldsymbol{y} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n). \tag{10.17}$$

Плотность распределения случайного вектора $\boldsymbol{y}$ равна

$$h(\boldsymbol{y}) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2}\frac{(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{\sigma^2}\right\}. \tag{10.18}$$

Отсюда получаем логарифмическую функцию правдоподобия

$$\ln L(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 - \frac{1}{2}\frac{(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{\sigma^2}, \tag{10.19}$$

которая является функцией параметров $\boldsymbol{\beta}$ и σ^2. Частные производные первого порядка этой функции равны

$$\begin{aligned} \frac{\partial \ln L}{\partial \boldsymbol{\beta}'} &= \frac{1}{\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{X}, \\ \frac{\partial \ln L}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{2\sigma^4}. \end{aligned} \tag{10.20}$$

Приравнивая первые производные (10.20) нулю, находим оценки максимального правдоподобия

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}, \qquad \widehat{\sigma}^2 = \frac{\boldsymbol{e}'\boldsymbol{e}}{n}, \tag{10.21}$$

где $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}$ обозначает вектор остатков. Заметим, что оценка максимального правдоподобия $\widehat{\boldsymbol{\beta}}$ совпадает с оценкой метода наименьших квадратов (3.4) для $\boldsymbol{\beta}$, в то время как оценка максимального правдоподобия $\widehat{\sigma}^2$ *не равна* оценке (3.19) $s^2 = \boldsymbol{e}'\boldsymbol{e}/(n-k)$ для σ^2, обычно используемой в методе наименьших квадратов.

Частные производные второго порядка от (10.19) равны

$$\frac{\partial^2 \ln L}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} = -\frac{\boldsymbol{X}'\boldsymbol{X}}{\sigma^2}, \qquad \frac{\partial^2 \ln L}{\partial \sigma^2 \partial \boldsymbol{\beta}'} = -\frac{(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{X}}{\sigma^4}, \tag{10.22}$$

и

$$\frac{\partial^2 \ln L}{(\partial \sigma^2)^2} = \frac{n}{2\sigma^4} - \frac{(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})}{\sigma^6}. \tag{10.23}$$

Взяв математическое ожидание (с обратным знаком) от производных второго порядка, получаем

$$-\mathrm{E}\left(\frac{\partial^2 \ln L}{\partial\beta\partial\beta'}\right) = \frac{X'X}{\sigma^2}, \qquad -\mathrm{E}\left(\frac{\partial^2 \ln L}{(\partial\sigma^2)^2}\right) = \frac{n}{2\sigma^4}, \tag{10.24}$$

поскольку $\mathrm{E}\big((y-X\beta)'(y-X\beta)\big) = n\sigma^2$. Кроме того, математическое ожидание от смешанной производной равно нулю в силу того, что $\mathrm{E}(y - X\beta) = 0$. Поэтому информационная матрица равна

$$\mathcal{F}_n(\beta,\sigma^2) = \begin{bmatrix} (1/\sigma^2)X'X & 0 \\ 0 & n/(2\sigma^4) \end{bmatrix}. \tag{10.25}$$

Таким образом, асимптотическая информационная матрица имеет вид

$$\mathcal{F}(\beta,\sigma^2) = \lim_{n\to\infty}(1/n)\mathcal{F}_n = \begin{bmatrix} (1/\sigma^2)Q & 0 \\ 0 & 1/(2\sigma^4) \end{bmatrix}, \tag{10.26}$$

где предполагается, что существует предельная, положительно определенная матрица $Q = \lim_{n\to\infty}(1/n)X'X$. Легко вычисляется обратная матрица

$$\mathcal{F}^{-1} = \begin{bmatrix} \sigma^2 Q^{-1} & 0 \\ 0 & 2\sigma^4 \end{bmatrix} \tag{10.27}$$

и из свойств оценок максимального правдоподобия следует, что $(\widehat{\sigma}^2/n)Q^{-1}$ является оценкой асимптотической матрицы ковариаций вектора оценок $\widehat{\beta}$. На практике асимптотическая матрица ковариаций аппроксимируется $\widehat{\sigma}^2(X'X)^{-1}$. В самом деле, мы знаем, что *точная* матрица ковариаций вектора $\widehat{\beta}$ равна $\sigma^2(X'X)^{-1}$.

Асимптотическую дисперсию оценки $\widehat{\sigma}^2$ можно оценить величиной $2\widehat{\sigma}^4/n$. Так как внедиагональные блоки матрицы $\mathcal{F}$ равны нулю, оценки $\widehat{\beta}$ и $\widehat{\sigma}^2$ асимптотически независимы. Такая (асимптотическая) независимость оценок средних значений и дисперсии является общей чертой теории регрессии и имеет важные последствия для оценивания и тестирования.

10.6. Проверка гипотез в линейной модели, I

Рассмотрим теперь обобщенную линейную модель $y = X\beta + u$ (см. (5.3)) с матрицей ковариаций ошибок $\mathrm{V}(u) = \Omega$, где Ω — *известная* положительно определенная симметричная матрица. Мы ослабим это требование в следующем разделе. Пусть мы хотим проверить гипотезу о том, что выполнена система q ($q < k$) независимых линейных ограничений $R\beta = r$. Здесь R — известная $q \times k$ матрица ранга q, а r — известный $q \times 1$ вектор. В данном разделе мы расмотрим три различных теста для проверки этой гипотезы, основанные на разных подходах.

Выпишем логарифмическую функцию правдоподобия

$$\ln L(\beta) = \mathrm{const} + \frac{1}{2}\ln|\Omega^{-1}| - \frac{1}{2}(y - X\beta)'\Omega^{-1}(y - X\beta), \tag{10.28}$$

ее частные производные первого порядка

$$\frac{\partial \ln L(\beta)}{\partial\beta} = X'\Omega^{-1}(y - X\beta) \tag{10.29}$$

и информационную матрицу

$$\mathcal{F}(\beta) = -\mathrm{E}\left(\frac{\partial^2 \ln L(\beta)}{\partial\beta\partial\beta'}\right) = X'\Omega^{-1}X. \tag{10.30}$$

ML-оценка для β в регрессии без ограничения задается уравнением $\partial \ln L(\beta)/\partial\beta = 0$, откуда получается GLS-оценка (см. (5.4))

$$\widehat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y. \tag{10.31}$$

Вектор остатков равен $\widehat{u} = y - X\widehat{\beta}$, а соответствующее максимальное значение логарифмической функции правдоподобия равно

$$\ln L(\widehat{\beta}) = \mathrm{const} + \frac{1}{2}\ln|\Omega^{-1}| - \frac{1}{2}\widehat{u}'\Omega^{-1}\widehat{u}. \tag{10.32}$$

ML-оценка для β в регрессии с ограничением ($R\beta = r$) получается максимизацией функции $\ln L(\beta)$ при условии $R\beta = r$.

Чтобы найти эту оценку, запишем функцию Лагранжа

$$\psi(\boldsymbol{\beta}) = \ln L(\boldsymbol{\beta}) - \boldsymbol{l}'(\boldsymbol{R\beta} - \boldsymbol{r}), \tag{10.33}$$

где через $\boldsymbol{l} = (l_1, \ldots, l_q)'$ обозначен вектор q множителей Лагранжа. Условия экстремума имеют вид

$$\partial\psi(\boldsymbol{\beta})/\partial\boldsymbol{\beta} = \mathbf{0} \qquad \text{и} \qquad \boldsymbol{R\beta} = \boldsymbol{r} \tag{10.34}$$

или

$$\boldsymbol{X}'\boldsymbol{\Omega}^{-1}(\boldsymbol{y} - \boldsymbol{X\beta}) - \boldsymbol{R}'\boldsymbol{l} = \mathbf{0} \qquad \text{и} \qquad \boldsymbol{R\beta} = \boldsymbol{r}. \tag{10.35}$$

Обозначим через $\widetilde{\boldsymbol{\beta}}$ и $\widetilde{\boldsymbol{l}}$ решение системы (10.35). Получаем

$$\widetilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}} - (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{R}'\widetilde{\boldsymbol{l}} \tag{10.36}$$

и, следовательно,

$$\boldsymbol{r} = \boldsymbol{R}\widetilde{\boldsymbol{\beta}} = \boldsymbol{R}\widehat{\boldsymbol{\beta}} - \big(\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{R}'\big)\widetilde{\boldsymbol{l}}. \tag{10.37}$$

Выразив $\widetilde{\boldsymbol{l}}$ из (10.37):

$$\widetilde{\boldsymbol{l}} = \big(\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{R}'\big)^{-1}(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r}), \tag{10.38}$$

и, подставив это выражение в (10.36), найдем оценку $\widetilde{\boldsymbol{\beta}}$ в регрессии с ограничением:

$$\widetilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}} - (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{R}'\big(\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{R}'\big)^{-1}(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r}). \tag{10.39}$$

Обозначим через $\widetilde{\boldsymbol{u}} = \boldsymbol{y} - \boldsymbol{X}\widetilde{\boldsymbol{\beta}}$ вектор остатков в регрессии с ограничением. Тогда максимальное значение логарифмической функции правдоподобия $\ln L(\boldsymbol{\beta})$ равно

$$\ln L(\widetilde{\boldsymbol{\beta}}) = \text{const} + \frac{1}{2}\ln|\boldsymbol{\Omega}^{-1}| - \frac{1}{2}\widetilde{\boldsymbol{u}}'\boldsymbol{\Omega}^{-1}\widetilde{\boldsymbol{u}}. \tag{10.40}$$

Теперь мы готовы сформулировать три теста, о которых шла речь в начале раздела. Во всех этих тестах нулевой гипотезой H_0 является наличие ограничения $\boldsymbol{R\beta} = \boldsymbol{r}$.

Тест Вальда (W) (*Wald test*). Тест Вальда основан на идее, что при выполнении нулевой гипотезы вектор $\boldsymbol{R}\widehat{\boldsymbol{\beta}}$ должен быть близок к $\boldsymbol{r}$. Из (10.31) получаем

$$\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r} \sim N(\boldsymbol{R\beta} - \boldsymbol{r}, \boldsymbol{R}(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{R}'). \tag{10.41}$$

Следовательно, если имеет место нулевая гипотеза, то

$$\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r} \sim N(\mathbf{0}, \boldsymbol{R}(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{R}'). \tag{10.42}$$

Используя свойство нормального распределения (приложение МС, п. 4, N9), получаем

$$\mathrm{W} \equiv (\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r})'\big(\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{R}'\big)^{-1}(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r}) \sim \chi^2(q). \tag{10.43}$$

Отметим, что тест Вальда использует только оценки в модели без ограничения на параметры.

Тест множителей Лагранжа (LM) (*Lagrange multiplier test*). Тест множителей Лагранжа основан на идее, что при выполнении нулевой гипотезы все множители Лагранжа должны быть равны нулю, поэтому и вектор $\widetilde{\boldsymbol{l}}$ должен быть близок к нулю. Из (10.38) и (10.41) получаем, что в том случае, когда выполняется нулевая гипотеза,

$$\widetilde{\boldsymbol{l}} \sim N\big(\mathbf{0}, \big(\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{R}'\big)^{-1}\big) \tag{10.44}$$

и, следовательно,

$$\mathrm{LM} \equiv \widetilde{\boldsymbol{l}}'\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{R}'\widetilde{\boldsymbol{l}} \sim \chi^2(q). \tag{10.45}$$

Поскольку в силу (10.35) $\boldsymbol{X}'\widetilde{\boldsymbol{\Omega}}^{-1}\widetilde{\boldsymbol{u}} = \boldsymbol{R}'\widetilde{\boldsymbol{l}}$, получаем эквивалентное представление LM статистики

$$\mathrm{LM} = \widetilde{\boldsymbol{u}}'\boldsymbol{\Omega}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\widetilde{\boldsymbol{u}}.$$

В отличие от теста Вальда тест множителей Лагранжа использует только оценки в модели с ограничением на параметры.

Тест отношения правдоподобия (LR) (*Likelihood ratio test*). Тест отношения правдоподобия использует как регрессию с ограничением, так и регрессию без ограничения. Он основан на том,

что если ограничение справедливо, то отношение максимальных значений функций правдоподобия для регрессии *с ограничением* и *без ограничения* должно быть близко к 1. Таким образом, в качестве критической статистики теста берется разница максимумов логарифмических функций правдоподобия (10.32) и (10.40):

$$\mathrm{LR} \equiv -2(\ln L(\widetilde{\beta}) - \ln L(\widehat{\beta})) = \widetilde{u}'\Omega^{-1}\widetilde{u} - \widehat{u}'\Omega^{-1}\widehat{u}. \qquad (10.46)$$

Можно показать, что при выполнении нулевой гипотезы (10.46) имеет $\chi^2(q)$ распределение. Действительно, поскольку $X\widetilde{\beta} = X\widehat{\beta} - X(\widehat{\beta} - \widetilde{\beta})$, то $\widetilde{u} = \widehat{u} + X(\widehat{\beta} - \widetilde{\beta})$. Домножив на $\Omega^{-1/2}$, получаем

$$\Omega^{-1/2}\widetilde{u} = \Omega^{-1/2}\widehat{u} + \Omega^{-1/2}X(\widehat{\beta} - \widetilde{\beta}). \qquad (10.47)$$

Взяв скалярный квадрат обеих частей (10.47) и учитывая, что $X'\Omega^{-1}\widehat{u} = 0$ (см. (10.31)), имеем

$$\widetilde{u}'\Omega^{-1}\widetilde{u} = \widehat{u}'\Omega^{-1}\widehat{u} + (\widehat{\beta} - \widetilde{\beta})'X'\Omega^{-1}X(\widehat{\beta} - \widetilde{\beta}). \qquad (10.48)$$

Из (10.36) вытекает, что

$$\widehat{\beta} - \widetilde{\beta} = (X'\Omega^{-1}X)^{-1}R'\widetilde{l}, \qquad (10.49)$$

поэтому из (10.45) следует:

$$\mathrm{LR} = \widetilde{u}'\Omega^{-1}\widetilde{u} - \widehat{u}'\Omega^{-1}\widehat{u} = \widetilde{l}'R(X'\Omega^{-1}X)^{-1}R'\widetilde{l} \sim \chi^2(q). \qquad (10.50)$$

Мы видим, что все три критические статистики имеют одно и то же $\chi^2(q)$ распределение. Более того, из (10.50)) следует LR = LM. Из равенства (10.38) вытекает, что W = LM. Следовательно,

$$\mathrm{LM} = \mathrm{LR} = \mathrm{W}. \qquad (10.51)$$

Однако, это справедливо только в простейшем случае, когда матрица ковариаций ошибок Ω полностью известна. Ситуация усложняется в том (более реальном) случае, когда матрица Ω неизвестна.

10.7. Проверка гипотез в линейной модели, II

Предположим теперь, что матрица Ω неизвестна, однако известна ее *структура*, т. е. $\Omega = \Omega(\theta)$ является функцией неизвестного $p \times 1$ вектора параметров θ, который необходимо оценить. (Эту ситуацию мы рассматривали ранее в главе 5, п. 5.3.) Три критические статистики в этом случае принимают вид:

$$\mathrm{W} = (R\widehat{\beta} - r)'(R(X'\widehat{\Omega}^{-1}X)^{-1}R')^{-1}(R\widehat{\beta} - r), \qquad (10.52)$$

$$\mathrm{LM} = \widetilde{u}'\widetilde{\Omega}^{-1}X(X'\widetilde{\Omega}^{-1}X)^{-1}X'\widetilde{\Omega}^{-1}\widetilde{u}, \qquad (10.53)$$

$$\mathrm{LR} = -2(\ln L(\widetilde{\beta}, \widetilde{\theta}) - \ln L(\widehat{\beta}, \widehat{\theta})), \qquad (10.54)$$

где $\widehat{\Omega} \equiv \Omega(\widehat{\theta})$ и $\widetilde{\Omega} \equiv \Omega(\widetilde{\theta})$. Теперь мы покажем, что верно знаменитое неравенство

$$\mathrm{LM} \leqslant \mathrm{LR} \leqslant \mathrm{W}. \qquad (10.55)$$

Доказательство основано на довольно тонком замечании, принадлежащем Бреушу (Breusch, 1979). Отметим сначала, что $\widehat{\beta}$ является точкой максимума функции $\ln L(\beta, \widehat{\theta})$, а $\widetilde{\beta}$ является точкой условного максимума функции $\ln L(\beta, \widehat{\theta})$ при ограничении $R\beta = r$. Введем теперь два новых вектора: $\widetilde{\beta}_u$, являющийся точкой максимума функции $\ln L(\beta, \widetilde{\theta})$, и $\widehat{\beta}_r$, который является точкой условного максимума функции $\ln L(\beta, \widehat{\theta})$ при ограничении $R\beta = r$.

Равенства, выведенные в предыдущем разделе (для случая, когда θ известно), позволяют получить следующие выражения для статистик W и LM, представив их в виде, аналогичном статистике LR:

$$\mathrm{W} = -2\left(\ln L(\widehat{\beta}_r, \widehat{\theta}) - \ln L(\widehat{\beta}, \widehat{\theta})\right), \qquad (10.56)$$

$$\mathrm{LM} = -2\left(\ln L(\widetilde{\beta}, \widetilde{\theta}) - \ln L(\widetilde{\beta}_u, \widetilde{\theta})\right). \qquad (10.57)$$

Кроме того, критическая статистика теста отношения правдоподобий равна

$$\mathrm{LR} = -2\left(\ln L(\widetilde{\beta},\widetilde{\theta}) - \ln L(\widehat{\beta},\widehat{\theta})\right). \tag{10.58}$$

Отсюда мы получаем, что

$$\mathrm{LR} - \mathrm{LM} = 2\left(\ln L(\widehat{\beta},\widehat{\theta}) - \ln L(\widetilde{\beta}_u,\widetilde{\theta})\right) \geqslant 0, \tag{10.59}$$

$$\mathrm{W} - \mathrm{LR} = 2\left(\ln L(\widetilde{\beta},\widetilde{\theta}) - \ln L(\widehat{\beta}_r,\widehat{\theta})\right) \geqslant 0. \tag{10.60}$$

Следовательно, $\mathrm{LM} \leqslant \mathrm{LR} \leqslant \mathrm{W}$.

Это неравенство справедливо при нулевой гипотезе. Для конечных выборок уровень значимости тестов, если применять одно и то же критическое значение, будет различен. Неравенство само по себе не дает никакой информации о сравнительной мощности W, LR и LM тестов. Подробное обсуждение этого неравенства и его следствий содержится в работах (Evans and Savin, 1982) и (Godfrey, 1988).

10.8. Нелинейные ограничения

Предположим, что нулевая гипотеза состоит из системы q нелинейных ограничений на вектор коэффициентов $\boldsymbol{\beta}$. Пусть дана линейная модель $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}$ при стандартном условии на распределение вектора ошибок $\boldsymbol{u} \sim N(\boldsymbol{0}, \boldsymbol{\Omega}(\boldsymbol{\theta}))$. Запишем ограничения в виде

$$\boldsymbol{g}(\boldsymbol{\beta}) = \boldsymbol{0}. \tag{10.61}$$

Таким образом, мы собираемся тестировать нулевую гипотезу H_0: $\boldsymbol{g}(\boldsymbol{\beta}) = \boldsymbol{0}$ против альтернативной гипотезы H_1: $\boldsymbol{g}(\boldsymbol{\beta}) \neq \boldsymbol{0}$.

Предположим, что все q компонент вектора $\boldsymbol{g}(\boldsymbol{\beta})$ являются непрерывными дважды дифференцируемыми функциями. Обозначим через $\boldsymbol{G}(\boldsymbol{\beta})$ $q \times k$ матрицу первых производных и предположим, что она имеет полный ранг в некоторой окрестности истинного значения $\boldsymbol{\beta}$.

Логарифмическая функция правдоподобия равна

$$\ln L(\boldsymbol{\beta},\boldsymbol{\theta}) = \mathrm{const} + \frac{1}{2}\ln|\boldsymbol{\Omega}^{-1}(\boldsymbol{\theta})| - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X\beta})'\boldsymbol{\Omega}^{-1}(\boldsymbol{\theta})(\boldsymbol{y} - \boldsymbol{X\beta}). \tag{10.62}$$

Как и ранее, обозначим оценку максимального правдоподобия без ограничения через $(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\theta}})$, а с ограничением — через $(\widetilde{\boldsymbol{\beta}},\widetilde{\boldsymbol{\theta}})$. После несложных вычислений три критические статистики записываются в виде

$$\mathrm{W} = \boldsymbol{g}(\widehat{\boldsymbol{\beta}})'\left(\boldsymbol{G}(\widehat{\boldsymbol{\beta}})(\boldsymbol{X}'\widehat{\boldsymbol{\Omega}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{G}(\widehat{\boldsymbol{\beta}})'\right)^{-1}\boldsymbol{g}(\widehat{\boldsymbol{\beta}}), \tag{10.63}$$

$$\mathrm{LM} = \widetilde{\boldsymbol{u}}'\widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{X}(\boldsymbol{X}'\widetilde{\boldsymbol{\Omega}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\widetilde{\boldsymbol{\Omega}}^{-1}\widetilde{\boldsymbol{u}}, \tag{10.64}$$

$$\mathrm{LR} = -2\left(\ln L(\widetilde{\boldsymbol{\beta}},\widetilde{\boldsymbol{\theta}}) - \ln L(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\theta}})\right). \tag{10.65}$$

(Сравните эти выражения с (10.52)–(10.54).) Бреуш (Breusch, 1979) показал, что и в нелинейном случае по-прежнему верно, что $\mathrm{LR} \geqslant \mathrm{LM}$, однако второе неравенство $\mathrm{W} \geqslant \mathrm{LR}$, вообще говоря, не выполняется.

В заключение обратим внимание на некоторую проблему, связанную со статистикой Вальда. Дело в том, что всегда существует множество способов записи одного и того же ограничения. Например, в простой модели

$$\boldsymbol{y} = \beta_1\boldsymbol{x}_1 + \beta_2\boldsymbol{x}_2 + \boldsymbol{u}, \qquad \boldsymbol{u} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I}), \tag{10.66}$$

мы собираемся тестировать условие $\beta_1 = \beta_2$. Функцию, задающую это ограничение, можно записать в виде $g_1(\boldsymbol{\beta}) = \beta_1 - \beta_2$ или эквивалентным способом $g_2(\boldsymbol{\beta}) = (\beta_1/\beta_2) - 1$. Конечно, можно записать это же условие и многими другими способами. Функции g_1 и g_2 совпадают при наличии ограничения, но отличаются в других значениях аргументов. Производные функций g_1 и g_2 равны

$$\boldsymbol{G}_1(\boldsymbol{\beta}) = (1, -1), \qquad \boldsymbol{G}_2(\boldsymbol{\beta}) = \left(\frac{1}{\beta_2}, \frac{-\beta_1}{\beta_2^2}\right). \tag{10.67}$$

Отсюда критические статистики тестов Вальда имеют вид

$$\mathrm{W}_i = \frac{(\widehat{\beta}_1 - \widehat{\beta}_2)^2}{\widehat{\sigma}^2 \boldsymbol{d}_i'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{d}_i} \quad (i = 1, 2), \tag{10.68}$$

где

$$\boldsymbol{d}_1' = (1, -1), \qquad \boldsymbol{d}_2' = (1, -(\widehat{\beta}_1/\widehat{\beta}_2)). \tag{10.69}$$

Мы можем заключить отсюда, что статистика теста Вальда — в отличие от LR и LM статистик — неинвариантна по отношению к тривиальным преобразованиям в нулевой гипотезе. Причина, по которой это происходит, состоит в том, что критическая статистика теста Вальда выводится из линейной аппроксимации вектора ограничений в точке $\widehat{\boldsymbol{\beta}}$, а различные способы выражения ограничения приводят к различным линеаризациям. Эти различия асимптотически исчезают, однако могут быть существенны в конечных выборках.

Упражнения

10.1. Рассмотрим выборку размера 10 из пуассоновского распределения с параметром θ: 1, 4, 3, 2, 3, 0, 1, 1, 0, 5.

а) Вычислите оценку максимального правдоподобия для θ.

б) Покажите графически, что функция правдоподобия и ее логарифм достигают максимума в одной и той же точке θ.

10.2. Дана выборка размера n из нормального распределения $N(\mu, \sigma^2)$. Запишите логарифмическую функцию правдоподобия и найдите ML-оценки параметров μ и σ^2. Найдите смещения этих оценок.

10.3. Пусть $y_1, \ldots, y_n$ — выборка из распределения с плотностью $h(y, \theta) = 1/\theta$, если $0 < x \leqslant \theta$, и $h(y, \theta) = 0$ — в остальных случаях $(0 < \theta < \infty)$. Покажите, что $\widehat{\theta} = \max y_i$ является оценкой максимального правдоподобия, и найдите ее смещение.

10.4. Выведите оценки максимального правдоподобия для параметров $\boldsymbol{\mu}$ и $\boldsymbol{\Omega}$ многомерного нормального распределения по выборке размера n.

10.5. Пусть $L_n(\boldsymbol{\theta})$ — функция правдоподобия. Докажите, что

$$\mathrm{E}\left(\frac{\partial \ln L_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right) = \boldsymbol{0}, \qquad \mathrm{V}\left(\frac{\partial \ln L_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right) = \boldsymbol{\mathcal{F}}_n(\boldsymbol{\theta}).$$

10.6. Пусть $y_1, \ldots, y_n$ — независимые, одинаково распределенные случайные величины, равномерно распределенные на интервале $(\theta, 2\theta)$. Покажите, что:

а) оценка максимального правдоподобия есть $\widehat{\theta} = \max y_i/2$;

б) $\widehat{\theta}$ является смещенной, но асимптотически несмещенной;

в) $\mathrm{V}(\widehat{\theta})$ асимптотически равна $\theta^2/(4n^2)$.

10.7. (Продолжение упражнения 10.6) Рассмотрим альтернативную оценку:

$$\widetilde{\theta} = (\min y_i + 2\max y_i)/5.$$

а) Покажите, что $\mathrm{V}(\widetilde{\theta})$ асимптотически равна $\theta^2/(5n^2)$.

б) Покажите, что $\widetilde{\theta}$ более эффективна, чем $\widehat{\theta}$.

в) Противоречит ли это асимптотической эффективности оценки максимального правдоподобия?

10.8. Дана линейная модель $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$, $\boldsymbol{u} \sim N(0, \sigma^2 \boldsymbol{I}_n)$. Покажите, что

$$\mathrm{LM} = n(\mathrm{ESS_R} - \mathrm{ESS_{UR}})/\mathrm{ESS_R}, \qquad \mathrm{LR} = n\ln(\mathrm{ESS_R}/\mathrm{ESS_{UR}}),$$

и

$$\mathrm{W} = n(\mathrm{ESS_R} - \mathrm{ESS_{UR}})/\mathrm{ESS_{UR}}.$$

Покажите, что выполняются неравенства $\mathrm{LM} \leqslant \mathrm{LR} \leqslant \mathrm{W}$.

10.9. Представим стандартную линейную модель в следующем виде:

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \boldsymbol{y}_3 \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \boldsymbol{X}_3 \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \boldsymbol{\varepsilon}_3 \end{bmatrix},$$

где $\boldsymbol{y}_i, \boldsymbol{\varepsilon}_i$ — $n_i \times 1$ векторы, $\boldsymbol{X}_i$ — $n_i \times k$ матрицы, $\boldsymbol{\beta}$ — $k \times 1$ вектор, векторы $\boldsymbol{\varepsilon}_i$ имеют нормальное распределение $N\left(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_{n_i}\right)$ и независимы. Вектор $\boldsymbol{y}_3$ представляет собой пропущенные наблюдения зависимой переменной, а матрица $\boldsymbol{X}_2$ — пропущенные наблюдения независимых переменных.

Вычислите следующие оценки вектора β и сравните их свойства:

а) МНК-оценка только по полным наблюдениям y_1, X_1;

б) МНК-оценка при замене матрицы X_2 на нулевую и исключении наблюдений y_3, X_3;

в) МНК-оценка по всей модели при замене y_3, X_2 на соответственно нулевой вектор и нулевую матрицу;

г) оценка максимального правдоподобия, предполагая y_3, X_2 неизвестными параметрами наряду с β.

10.10. Известно, что в модели множественной регрессии $y = X\beta + \varepsilon$ имеется гетероскедастичность, причем

$$\begin{aligned} \mathrm{V}(\varepsilon_t) &= \sigma_1^2, \quad t = 1, \ldots, n_1, \\ \mathrm{V}(\varepsilon_t) &= \sigma_2^2, \quad t = n_1 + 1, \ldots, n_1 + n_2 \ (n = n_1 + n_2), \\ \mathrm{E}(\varepsilon_t \varepsilon_s) &= 0, \quad t \neq s. \end{aligned}$$

В предположении нормальности вектора ошибок постройте тест отношения правдоподобия (LR-test) для проверки гипотезы H_0: $\sigma_1^2 = \sigma_2^2$.

10.11. Дана линейная модель $y = X\beta + \epsilon$, где y — $n \times 1$ вектор, β — $k \times 1$ вектор, ϵ — $n \times 1$ вектор, X — $n \times k$ матрица, $\epsilon \sim N(0, \Omega)$ и

$$\Omega = \begin{bmatrix} \sigma_1^2 I_{n_1} & 0 & \ldots & 0 \\ 0 & \sigma_2^2 I_{n_2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_r^2 I_{n_r} \end{bmatrix}, \quad n_1 + n_2 + \cdots + n_r = n$$

(групповая гетероскедастичность). Как выглядит LR-тест (тест отношения правдоподобия) для проверки гипотезы H_0: $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_r^2$? *Указание.* Рекомендуется получить ответ в терминах ML-оценок дисперсий σ^2.

10.12. Пусть p — вероятность выпадения орла при бросании монеты. Из $n = 100$ испытаний $x = 42$ раза выпал орел и 58 — решка. Тестируйте на 5%-ном уровне значимости гипотезу $\mathrm{H}_0 : p = 0.5$:

а) при помощи теста Вальда (W);

б) при помощи теста множителей Лагранжа (LM);

в) при помощи теста отношения правдоподобия (LR).

10.13. Имеется 80 наблюдений пуассоновской случайной величины X. Их среднее значение равно $\overline{x} = 1.7$. Тестируйте на 5%-ном уровне значимости гипотезу $\mathrm{H}_0 : \lambda = 2.0$:

а) при помощи теста Вальда (W);

б) при помощи теста множителей Лагранжа (LM);

в) при помощи теста отношения правдоподобия (LR).

Глава 11

Временные ряды

Во многих экономических задачах встречаются лагированные (взятые в предыдущий момент времени) переменные. Например, Y_t — выпуск предприятия в год t, может зависеть не только от инвестиций I_t в этот год, но и от инвестиций в предыдущие годы:

$$Y_t = \alpha + \beta_0 I_t + \beta_1 I_{t-1} + \cdots + \beta_k I_{t-k}.$$

Такие модели встречаются всякий раз, когда эндогенная переменная с запаздыванием реагирует на изменения экзогенной переменной. При этом в модели могут использоваться лагированные значения экзогенной или эндогенной переменной или одновременно и те, и другие. Для статистического моделирования полезно различать два случая. Обе модели

$$y_t = \beta_1 + \beta_2 x_t + \beta_3 x_{t-1} + \varepsilon_t \tag{11.1}$$

и

$$y_t = \beta_1 + \beta_2 x_t + \beta_3 y_{t-1} + \varepsilon_t \tag{11.2}$$

включают в себя лагированные значения переменных, но существенно различаются с точки зрения статистического оценивания параметров. Действительно, в (11.1) регрессоры некоррелированы с ошибками (мы здесь предполагаем, что экзогенная переменная x_t детерминированная). Поэтому (11.1) можно оценивать с помощью МНК. В модели (11.2) y_{t-1} включает в себя ε_{t-1}, поэтому

вектор ошибок $\boldsymbol{\varepsilon}$ и матрица регрессоров $\boldsymbol{X}$ коррелированы. В этом случае оценки МНК, вообще говоря, не являются несмещенными (п. 5.1).

Уравнение (11.1) является примером модели *распределенных лагов* (*distributed lags*), DL(1). В скобках указан порядок модели — максимальный лаг. Уравнение (11.2) является *авторегрессионной моделью распределенных лагов*, ADL(1,0). В скобках указаны максимальные лаги эндогенной и экзогенной переменных.

Рассмотрим отклик зависимой переменной y на единичное приращение экзогенной переменной x. Отклик за один период (*short run*) равен β_2 в обеих моделях. Суммарное влияние (*long run*) равно $\beta_2 + \beta_3$ в модели (11.1) и $\beta_2 + \beta_2\beta_3 + \beta_2\beta_3^2 + \cdots$ в модели (11.2). В самом деле, если y_{t-1} изменяется на β_2 единиц, то y_t изменится на $\beta_2\beta_3$ единиц, и т. д. Если выполнено неравенство $|\beta_3| < 1$, то ряд сходится к $\beta_2/(1-\beta_3)$. Условие $|\beta_3| < 1$ является условием *устойчивости* и встречается в том или ином виде во всех моделях с авторегрессионными членами.

Оператор сдвига

Для аналитических вычислений с моделями, включающими лаговые переменные, удобно использовать оператор сдвига (*lag operator*) $Lx_t = x_{t-1}$. Так, например, модель ADL(p,q)

$$\begin{aligned} &y_t - \alpha_1 y_{t-1} - \cdots - \alpha_p y_{t-p} \\ &\quad = \delta + \beta_0 x_t + \beta_1 x_{t-1} + \cdots + \beta_q x_{t-q} + \varepsilon_t; \qquad t = 1, \ldots, n, \end{aligned} \tag{11.3}$$

обобщающая модели (11.1) и (11.2), может быть записана в более компактном виде:

$$A(L)y_t = \delta + B(L)x_t + \varepsilon_t; \qquad t = 1, \ldots, n, \tag{11.4}$$

где $A(L)$ и $B(L)$ — полиномы от оператора сдвига:

$$\begin{aligned} A(L) &= 1 - \alpha_1 L - \cdots - \alpha_p L^p, \\ B(L) &= \beta_0 + \beta_1 L + \cdots + \beta_q L^q. \end{aligned}$$

11.1. Модели распределенных лагов

Уравнение (11.4) в случае отсутствия авторегрессионных членов принимает вид

$$y_t = \delta + \beta_0 x_t + \beta_1 x_{t-1} + \cdots + \beta_q x_{t-q} + \varepsilon_t, \qquad t = 1, \ldots, n, \quad (11.5)$$

и называется *моделью распределенных лагов*. Как и раньше, суммарное влияние равно $\beta = \beta_0 + \beta_1 + \cdots + \beta_q$. Вклад отдельного лага s равен $w_s = \beta_s/\beta$, $\sum_{k=0}^{q} w_k = 1$. Функция целого аргумента w_s называется распределением лагов. Для того чтобы измерить скорость реакции y на изменение x, можно ввести понятие среднего лага, равного $\sum_{k=0}^{q} k w_k$. Малые значения среднего лага соответствуют быстрой реакции y на изменения x, и, наоборот, большим значениям среднего лага соответствует замедленная реакция.

Оценивание

В случае, когда x детерминированы, а ошибки $\varepsilon_t \sim iid(0, \sigma^2)$ независимые, одинаково распределенные с нулевым средним и дисперсией σ^2 (*independent identically distributed*), модель (11.5) удовлетворяет условиям классической модели линейной регрессии (п. 3.1), однако на практике при ее оценивании могут встретиться трудности. Во-первых, может оказаться, что количество коэффициентов $q + 2$ слишком велико, если по смыслу задачи ожидается влияние с большим запаздыванием. Во-вторых, в том случае, если ряд x_t имеет некоторую структуру, например, автокорреляцию или сезонность, матрица $X'X$ может оказаться близкой к вырожденной, и мы оказываемся в ситуации мультиколлинеарности (см. п. 4.1).

Для преодоления этих трудностей обычно предполагается та или иная форма «гладкости» распределения лагов w_s. Это приводит к уменьшению числа оцениваемых параметров. Рассмотрим две популярные модели такого рода: *полиномиальных лагов* (метод Алмон (Almon)) и *геометрических лагов* (модель Койка (Koyck)).

Модель полиномиальных лагов

В этой модели зависимость β_i от i аппроксимируется полиномом некоторой степени r:

$$\beta_i = \gamma_0 + \gamma_1 i + \cdots + \gamma_r i^r, \qquad r \leqslant q. \quad (11.6)$$

Таким образом, после подстановки (11.6) в (11.5) получаем модель, содержащую только $r + 2$ неизвестных параметров и имеющую вид:

$$y_t = \delta + \gamma_0 \widetilde{x}_{0t} + \cdots + \gamma_r \widetilde{x}_{rt} + \varepsilon_t, \qquad t = 1, \ldots, n, \quad (11.7)$$

где переменные $\widetilde{x}_0, \ldots, \widetilde{x}_r$ являются линейными комбинациями переменных $x_t, \ldots, x_{t-q}$.

Как определить порядок r полинома (11.6)? Для проверки адекватности модели (11.7) можно применить обычный F-тест (см. (3.44))

$$F = \frac{(\mathrm{ESS_R} - \mathrm{ESS_{UR}})/(q - r)}{\mathrm{ESS_{UR}}/(n - q - 2)}.$$

(Здесь (11.5) — регрессия без ограничений, а (11.7) — регрессия с ограничениями.) В том случае, если значение F «достаточно мало» (меньше критического значения F-статистики), модель полиномиальных лагов адекватна данным. Как обычно, при прочих равных условиях, надо выбирать модель с наименьшим количеством параметров.

Модель геометрических лагов

В этой модели предполагается, что влияние переменной x не заканчивается через время q, а продолжается бесконечно, убывая на один и тот же процент с каждым шагом по времени. Такая модель представляется достаточно правдоподобной в примере с выпуском и инвестициями в оборудование, приведенным в начале главы. Модель имеет вид:

$$y_t = \delta + \beta x_t + \beta\lambda x_{t-1} + \beta\lambda^2 x_{t-2} + \cdots + \varepsilon_t, \qquad t = 1, \ldots, n. \quad (11.8)$$

Параметр λ $(0 < \lambda < 1)$ связан обратной зависимостью со скоростью реакции; $\lambda = 0$ означает мгновенную полную реакцию y на изменение x. Суммарное влияние равно $\sum_{k=0}^{\infty} \beta\lambda^k = \beta/(1-\lambda)$. Распределение лагов имеет вид $w_s = (1-\lambda)\lambda^s$.

Модель (11.8) содержит только три параметра (δ, β, λ), однако ее оценивание затруднено тем, что она является нелинейной. Можно предложить эвристическую процедуру оценивания этой модели, подобную процедуре Хилдреда-Лу (п. 6.2). Перебираем с некоторым шагом значения λ из интервала $(0,1)$ и для каждого находим МНК-оценку уравнения (11.8). Затем выбирается значение λ, соответствующее наименьшей сумме квадратов остатков.

Другой способ оценивания состоит в следующем: из уравнения (11.8) вычитается то же самое уравнение, сдвинутое по времени на один шаг назад, и умноженное на λ:

$$y_t - \lambda y_{t-1} = \delta(1-\lambda) + \beta x_t + u_t, \qquad t = 1, \ldots, n$$

или

$$y_t = \delta(1-\lambda) + \lambda y_{t-1} + \beta x_t + u_t, \qquad t = 1, \ldots, n. \tag{11.9}$$

Здесь $u_t = \varepsilon_t - \lambda\varepsilon_{t-1}$. Уравнение (11.9) линейно по комбинациям параметров, через которые эти параметры можно выразить. Однако (11.9) содержит лагированную эндогенную переменную и ошибки, не удовлетворяющие условиям классической модели линейной регрессии. Поэтому можно показать, что МНК-оценки коэффициентов уравнения являются несостоятельными. Для получения состоятельных оценок можно применить метод инструментальных переменных (п. 8.1), взяв, например, x_{t-1} в качестве инструмента для y_{t-1}, или воспользоваться методом максимального правдоподобия (глава 10).

11.2. Динамические модели

Рассмотрим особенности динамических моделей (содержащих лагированные эндогенные переменные в правой части) на простейшем примере (11.2):

$$y_t = \beta_1 + \beta_2 x_t + \beta_3 y_{t-1} + \varepsilon_t, \qquad t = 1, \ldots, n.$$

Запишем это уравнение в обозначениях (11.4):

$$A(L)y_t = \beta_1 + B(L)x_t + \varepsilon_t, \qquad t = 1, \ldots, n, \tag{11.10}$$

где $A(L) = 1 - \beta_3 L$ и $B(L) = \beta_2$. Начнем с простейшего случая: $B(L) \equiv \beta_1 \equiv 0$ (опустим индекс у β_3):

$$y_t = \beta y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim iid(0, \sigma^2), \qquad t = 1, \ldots, n. \tag{11.11}$$

Такой процесс называется авторегрессионным процессом первого порядка, AR(1). В главе 6 (п. 6.2) мы рассматривали подобную модель для ошибок регрессии. Как и ранее, мы предполагаем, что $|\beta| < 1$, тогда

$$y_t = \sum_{s=0}^{\infty} \beta^s \varepsilon_{t-s}, \quad \mathrm{E}(y_t) = 0; \quad \mathrm{V}(y_t) = \sigma^2/(1-\beta^2). \tag{11.12}$$

МНК-оценка параметра β равна:

$$\widehat{\beta} = \sum y_t y_{t-1} / \sum y_{s-1}^2 = \beta + \sum \left(y_{t-1} / \sum y_{s-1}^2 \right) \varepsilon_t. \tag{11.13}$$

Поскольку $y_{t-1} / \sum y_{s-1}^2$ и ε_t зависимы при всех $t < n$, то оценка (11.13) является, вообще говоря, смещенной. Однако можно показать, что если $|\beta| < 1$ и существуют необходимые моменты распределения ε, то оценка $\widehat{\beta}$ (11.13) является состоятельной и асимптотически нормальной:

$$\sqrt{n}(\widehat{\beta} - \beta) \xrightarrow{d} N(0, 1-\beta^2). \tag{11.14}$$

Итак, предыдущие аргументы показывают, что уравнение с авторегрессионными членами может быть оценено при помощи МНК. Существенными тут являются два условия. 1) Устойчивость. Для уравнения (11.11) это означает $|\beta| < 1$, лучше, если значения параметров будут отстоять на некоторое расстояние от границы критической области. 2) Отсутствует автокорреляция ошибок ε_t.

Авторегрессионная модель при наличии автокорреляции ошибок

Усложним модель (11.11), добавив в нее автокорреляцию ошибок:

$$\begin{aligned} y_t &= \beta y_{t-1} + u_t, \qquad t = 1, \ldots, n, \\ u_t &= \rho u_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim iid(0, \sigma^2). \end{aligned} \tag{11.15}$$

Теперь мы имеем другую ситуацию: y_{t-1} и u_t коррелированы, так как обе эти случайные величины зависят от u_{t-1}. При выполнении условий устойчивости $|\beta| < 1$, $|\rho| < 1$ можно вычислить предел по вероятности МНК-оценки (11.13) параметра β:

$$p\lim \widehat{\beta} = \frac{\beta + \rho}{1 + \beta\rho} \neq \beta. \tag{11.16}$$

Таким образом, МНК-оценка коэффициентов регрессии оказывается несостоятельной в моделях с авторегрессионными членами и автокорреляцией ошибок. Можно показать, что оценка $\widehat{\rho}$, полученная из остатков МНК, также не является состоятельной:

$$p\lim \widehat{\rho} = \frac{\beta\rho(\beta + \rho)}{1 + \beta\rho} \neq \rho. \tag{11.17}$$

Заметим, что именно для модели (11.15) не существует состоятельного метода оценивания. В самом деле, вычитая умноженное на ρ лагированное уравнение из исходного, получаем:

$$y_t = (\beta + \rho)y_{t-1} - \beta\rho y_{t-2} + \varepsilon_t, \qquad t = 1, \ldots, n, \tag{11.18}$$

т. е. параметры β и ρ неразличимы ($(\beta, \rho) = (0.1, 0.2)$ и $(\beta, \rho) = (0.2, 0.1)$ порождают то же уравнение (11.15)) и уравнение неидентифицируемо.

Оценивание. Метод инструментальных переменных

Так же как и в главе 8 (п. 8.1), в случае корреляции регрессоров с ошибкой можно применить метод инструментальных переменных

для оценивания моделей с авторегрессионными членами и автокорреляцией ошибок. Рассмотрим, например, модель

$$\begin{aligned} y_t &= \beta_1 + \beta_2 x_t + \beta_3 y_{t-1} + u_t, \qquad t = 1, \ldots, n, \\ u_t &= \rho u_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim iid(0, \sigma^2). \end{aligned} \tag{11.19}$$

Переменная x является экзогенной, y_{t-1} коррелирована с x_{t-1}, поэтому x_{t-1} можно взять в качестве инструмента для y_{t-1}. Оценка, полученная по методу инструментальных переменных, будет состоятельной. Однако вследствие автокорреляции ошибок оценки дисперсий оценок коэффициентов не будут состоятельными.

Оценивание. Метод максимального правдоподобия

Используя обычное преобразование, приведем модель (11.19) к виду

$$y_t = \beta_1(1 - \rho) + \beta_2 x_t - \beta_2\rho x_{t-1} + (\beta_3 + \rho)y_{t-1} - \beta_3\rho y_{t-2} + \varepsilon_t. \tag{11.20}$$

Коэффициенты уравнения (11.20) можно оценить при помощи метода максимального правдоподобия (п. 10.5), который легко применяется здесь, так как ошибки некоррелированы.

Замечание. Нелинейный метод наименьших квадратов состоит в данном случае в минимизации функции

$$\sum \left(y_t - \beta_1(1-\rho) - \beta_2 x_t + \beta_2\rho x_{t-1} - (\beta_3 + \rho)y_{t-1} + \beta_3\rho y_{t-2}\right)^2 \tag{11.21}$$

по параметрам $\beta_1, \beta_2, \beta_3$ и ρ. Отметим, что нелинейный метод наименьших квадратов является аппроксимацией метода максимального правдоподобия и отличается от него только отсутствием оптимизации по начальным значениям y.

Из предыдущего следует, что, перед тем как оценивать модель с авторегрессионными членами, необходимо проверить наличие автокорреляции ошибок.

Тест на автокорреляцию ошибок

Заметим прежде всего, что тест Дарбина–Уотсона (DW) (п. 6.2) в данном случае неприменим, так как не выполнены условия, лежащие в его обосновании. При наличии лагированных эндогенных

переменных результаты теста DW смещены в сторону принятия гипотезы отсутствия автокорреляции ошибок.

Тест множителей Лагранжа (*Lagrange Multiplier*, LM) (п. 10.6) тем не менее применим и в данной ситуации. Можно использовать также h тест Дарбина, который реализован во многих компьютерных пакетах. Этот тест, в отличие от LM теста, предназначен только для проверки на присутствие автокорреляции первого порядка. Например, для уравнения (11.19) критическая статистика имеет вид:

$$h = \left(1 - \frac{1}{2}\mathrm{DW}\right)\sqrt{\frac{n}{1 - n\widehat{\mathrm{V}}(\widehat{\beta}_3)}}, \qquad (11.22)$$

где DW — значение статистики Дарбина–Уотсона, $\widehat{\beta}_3$ — оценка коэффициента при y_{t-1}, полученная применением МНК непосредственно к исходному уравнению (11.19). Если верна нулевая гипотеза H_0 отсутствия автокорреляции ошибок первого порядка, то статистика h имеет асимптотически стандартное нормальное распределение. Гипотеза H_0 отвергается на 5%-ном уровне значимости в пользу гипотезы наличия положительной автокорреляции, если $h > 1.645$.

Тест Дарбина не работает, если $n\widehat{\mathrm{V}}(\widehat{\beta}_3) > 1$. Дарбин показал, что следующая процедура асимптотически эквивалентна h-тесту. 1) Вычислим остатки МНК-регрессии уравнения (11.19) e_t. 2) Оценим вспомогательную регрессию e_t на e_{t-1}, y_{t-1}, x_t. 3) Проверим гипотезу H_0 с помощью обычного t-теста на значимость коэффициента при e_{t-1} во вспомогательной регрессии.

Некоторые примеры моделей с лагированными переменными

В этом разделе мы рассмотрим частные случаи модели ADL(1,1)

$$y_t = \beta_1 + \beta_2 x_t + \beta_3 x_{t-1} + \beta_4 y_{t-1} + \varepsilon_t, \qquad t = 1, \ldots, n. \qquad (11.23)$$

Модель частичного приспособления

В модели частичного приспособления (*partial adjustment*) предполагается, что желаемое (или оптимальное, целевое) значение переменной y определяется уравнением:

$$y_t^* = \alpha + \beta x_t + \varepsilon_t, \qquad \varepsilon_t \sim iid(0, \sigma^2), \qquad t = 1, \ldots, n. \qquad (11.24)$$

Наблюдаемое значение переменной y, однако, не выходит мгновенно на желаемое значение, изменяясь только на долю δ в нужном направлении:

$$(y_t - y_{t-1}) = \delta(y_t^* - y_{t-1}), \qquad 0 < \delta < 1. \qquad (11.25)$$

Пусть, например, (11.24) определяет оптимальный размер запасов y_t^* в зависимости от уровня продаж x_t. Уравнение (11.25) можно переписать в виде:

$$y_t = \delta y_t^* + (1 - \delta) y_{t-1},$$

т. е. размер запасов равен взвешенному среднему оптимального размера запасов и размера запасов в предыдущем периоде.

Подставив (11.25) в (11.24), получаем:

$$y_t = \delta\alpha + \delta\beta x_t + (1 - \delta) y_{t-1} + \delta\varepsilon_t, \qquad t = 1, \ldots, n. \qquad (11.26)$$

Поскольку ошибки не коррелированы, состоятельные оценки параметров можно получить, применяя МНК к оцениванию составных параметров $\delta\alpha, \delta\beta$ и $(1 - \delta)$ в уравнении (11.26).

Модель (11.26) получается из модели (11.23) путем введения в последнюю ограничения $\beta_3 = 0$.

Модель адаптивных ожиданий

Обозначим через x_{t+1}^* ожидаемое (в момент t) будущее значение переменной x_t. Предположим, что значение величины y_t определяется этим ожидаемым значением:

$$y_t = \alpha + \beta x_{t+1}^* + \varepsilon_t, \qquad t = 1, \ldots, n. \qquad (11.27)$$

Гипотеза *адаптивных ожиданий* (*adaptive expectations*) предполагает, что ожидания пересматриваются в некоторой пропорции

от разницы (расхождения) между наблюденным значением и прогнозом переменной x на предыдущем шаге:

$$(x^*_{t+1} - x^*_t) = (1-\lambda)(x_t - x^*_t), \qquad 0 \leqslant \lambda < 1. \tag{11.28}$$

Такая модель возникает, например, в случае, когда фирма принимает решение об объеме производимой в период t продукции y_t до того, как известна цена x_{t+1}, по которой эта продукция может быть продана в следующий период. Поскольку цена x_{t+1} не известна в период t, то решение принимается на основе ожидаемого значения x^*_{t+1}. Гипотезу адаптивных ожиданий (11.28) можно записать в виде

$$x^*_{t+1} = \lambda x^*_t + (1-\lambda)x_t,$$

т. е. ожидаемое значение цены x^*_{t+1} является взвешенным средним наблюдаемой цены x и ожидаемой цены x^* в период t.

Итерируя (11.28) и подставляя затем результат в (11.27), получаем:

$$y_t = \alpha + \beta(1-\lambda)(x_t + \lambda x_{t-1} + \lambda^2 x_{t-2} + \cdots) + \varepsilon_t, \quad t = 1, \ldots, n. \tag{11.29}$$

Заметим, что полученное уравнение совпадает с моделью геометрических лагов (11.8) и может быть преобразовано к виду (11.9), который является частным случаем модели (11.23), если в последнем положить $\beta_3 = 0$ и ввести автокорреляцию ошибок.

Модель коррекции ошибок

Предположим, что модель (11.23) имеет стационарное состояние (y^*, x^*) (разумеется, для этого должно быть выполнено условие устойчивости $|\beta_4| < 1$). Записав (11.23) в стационарном состоянии, получаем:

$$(1-\beta_4)y^* = \beta_1 + (\beta_2 + \beta_3)x^*$$

или

$$y^* = \frac{\beta_1}{1-\beta_4} + \frac{\beta_2+\beta_3}{1-\beta_4}x^*. \tag{11.30}$$

Сделаем в уравнении (11.23) замену переменных $y_t = y_{t-1} + \Delta y_t$, $x_t = x_{t-1} + \Delta x_t$. Получим:

$$\Delta y_t = \beta_2 \Delta x_t - (1-\beta_4)\left(y_{t-1} - \frac{\beta_1}{1-\beta_4} - \frac{\beta_2+\beta_3}{1-\beta_4}x_{t-1}\right) + \varepsilon_t. \tag{11.31}$$

Уравнение (11.31) называется *моделью коррекции ошибок* (*error correction*). Изменение y на текущем шаге состоит из двух компонент. Первая пропорциональна текущему изменению x, вторая является частичной коррекцией отклонения y на предыдущем шаге от равновесного состояния, определяемого значением x в соответствии с уравнением (11.30).

Коэффициенты уравнения (11.31) могут быть оценены при помощи МНК. Результат оценивания полностью идентичен оцениванию этих же параметров в уравнении (11.23), так как уравнения получаются одно из другого невырожденной линейной заменой переменных.

Тест Гранжера на причинно-следственную зависимость

В экономике часто возникает вопрос о причинно-следственной связи между переменными. Например, верно ли, что увеличение денежной массы влечет за собой инфляцию?

Идея теста, предложенного в работе (Granger, 1969), проста: если x влияет на y, то изменения x должны предшествовать изменениям y, но не наоборот. Иначе говоря, должны выполняться два условия: во-первых, x должен вносить вклад в прогноз y; во-вторых, y не должен вносить значимый вклад в прогноз x. Если же каждая из этих двух переменных дает значимый вклад в прогноз другой, то, скорее всего, существует третья переменная z, влияющая на обе переменные.

Для того чтобы тестировать нулевую гипотезу «x не влияет на y», мы оцениваем регрессию y на лагированные значения y и лагированные значения x:

$$y_t = \alpha_0 + \sum_{j=1}^{m} \alpha_j y_{t-j} + \sum_{j=1}^{m} \beta_j x_{t-j} + \varepsilon_t. \tag{11.32}$$

На языке этой модели гипотеза «x не влияет на y» формулируется как H_0: $\beta_1 = \ldots = \beta_m = 0$. Для ее тестирования применяется обычный F-тест (3.44)–(3.45). Гипотеза «y не влияет на x» тестируется аналогично, надо только поменять местами x и y в уравнении (11.32). Для того чтобы прийти к заключению, что «x влияет на y», надо, чтобы гипотеза «x не влияет на y» была отвергнута, а гипотеза «y не влияет на x» была принята.

Подчеркнем, что «x влияет на y» не означает наличие причинно-следственной связи между x и y, а означает то, что предшествующие значения x объясняют последующие значения y, т. е. означает *возможность* наличия причинно-следственной связи. Если же гипотеза «x не влияет на y» не отвергается, то это означает, что x не является причиной y.

Описанный выше тест называется *тестом Гранжера на причинно-следственную зависимость* (*Granger causality test*). Заметим, что выбор m, вообще говоря, может повлиять на результат теста. Как правило, лучше проделать тест для нескольких разных значений m и выяснить, насколько результат теста чувствителен к выбору m. В том случае, если имеются основания предполагать наличие автокорреляции ошибок в модели (11.32), для тестирования гипотезы H_0 рекомендуется применять тест множителей Лагранжа (п. 10.6).

11.3. Единичные корни и коинтеграция

До сих пор мы говорили об устойчивости (стабильности) временного ряда. Дадим теперь более точное понятие стационарности.

Стационарность

Ряд y_t называется *строго стационарным* (*strictly stationary*) или *стационарным в узком смысле*, если совместное распределение m наблюдений $y_{t_1}, y_{t_2}, \ldots, y_{t_m}$ не зависит от сдвига по времени, то есть совпадает с распределением $y_{t_1+t}, y_{t_2+t}, \ldots, y_{t_m+t}$ для любых m, t, $t_1, \ldots, t_m$. Обычно нас интересуют средние значения и ковариации, а не все распределение. Поэтому часто используется

понятие *слабой стационарности* (*weak stationarity*) или *стационарности в широком смысле*, которое состоит в том, что среднее, дисперсия и ковариации y_t не зависят от момента времени t:

$$\mathrm{E}(y_t) = \mu < \infty, \quad \mathrm{V}(y_t) = \gamma_0, \quad \mathrm{Cov}(y_t, y_{t-k}) = \gamma_k. \tag{11.33}$$

Конечно, из строгой стационарности следует слабая стационарность (при условии конечности первого и второго моментов распределения). В дальнейшем мы будем везде под «стационарностью» понимать слабую стационарность.

Введем понятие автокорреляционной функции (*autocorrelation function*), ACF:

$$\rho_k = \frac{\mathrm{Cov}(y_t, y_{t-k})}{\mathrm{V}(y_t)} = \frac{\gamma_k}{\gamma_0}. \tag{11.34}$$

Заметим, что $\rho_0 = 1$, а $|\rho_k| \leqslant 1$. ACF играет важную роль в задаче идентификации моделей временных рядов.

Рассмотрим примеры временных рядов. Самым простым является ряд с независимыми одинаково распределенными наблюдениями:

$$y_t = \varepsilon_t, \qquad \varepsilon_t \sim iid(0, \sigma^2), \quad t = 1, \ldots, n. \tag{11.35}$$

Этот процесс называется «*белым шумом*» (*white noise*), у него $\mu = 0$, $\gamma_0 = \sigma^2$, $\gamma_k = 0$, $k > 0$.

Другим примером является AR(1) процесс:

$$y_t = m + \phi y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim iid(0, \sigma^2), \quad t = 1, \ldots, n. \tag{11.36}$$

Предполагается, что $|\phi| < 1$. Этот процесс уже рассматривался ранее в главе 6 (п. 6.2). Используя оператор сдвига (11.4), запишем (11.36) в виде:

$$(1 - \phi L) y_t = m + \varepsilon_t, \tag{11.37}$$

или

$$\begin{aligned} y_t &= (1 - \phi L)^{-1}(m + \varepsilon_t) = (1 + \phi L + \phi^2 L^2 + \cdots)(m + \varepsilon_t) \\ &= (1 + \phi + \phi^2 + \cdots) m + \varepsilon_t + \phi \varepsilon_{t-1} + \phi^2 \varepsilon_{t-2} + \cdots . \end{aligned} \tag{11.38}$$

Поскольку мы предполагаем, что $|\phi| < 1$, то из (11.38) получаем

$$\mathrm{E}(y_t) = \frac{m}{1-\phi} = \mu, \tag{11.39}$$

т. е. среднее не зависит от времени. При таком же условии на ϕ получаем

$$\mathrm{V}(y_t) = \gamma_0 = \sigma_y^2 = \frac{\sigma^2}{1-\phi^2}. \tag{11.40}$$

Аналогично (см. п. (6.2)) можно показать, что

$$\mathrm{Cov}(y_t, y_{t-k}) = \gamma_k = \phi^k \sigma_y^2 = \frac{\phi^k \sigma^2}{1-\phi^2}. \tag{11.41}$$

Таким образом, AR(1) процесс является стационарным при условии $|\phi| < 1$ и его автокорреляционная функция равна

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi^k, \qquad k = 1, 2, \ldots \tag{11.42}$$

Важным примером является процесс

$$y_t = y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim iid(0, \sigma^2), \quad t = 1, \ldots, n, \tag{11.43}$$

называемый *случайным блужданием* (*random walk*). Этот процесс по виду похож на AR(1) (11.36) с $\phi = 1$, однако существенно отличается от стационарного процесса AR(1) (с $|\phi| < 1$) по своим свойствам. Из (11.43), учитывая, что ошибка ε_t некоррелирована с y_{t-1}, можно получить:

$$\mathrm{E}(y_t) = \mathrm{E}(y_{t-1}) + 0; \qquad \mathrm{V}(y_t) = \mathrm{V}(y_{t-1}) + \sigma^2. \tag{11.44}$$

Отсюда ясно, что случайное блуждание нестационарно, так как $\mathrm{V}(y_t) \neq \mathrm{V}(y_{t-1})$. Если положить, что процесс начинается с момента $t = 1$ и $\mathrm{E}(y_1) = \mu$, $\mathrm{V}(y_1) = \sigma^2$, то $\mathrm{E}(y_t) = \mu$, $\mathrm{V}(y_t) = \sigma^2 t$, при $t = 1, 2, \ldots$, т. е. дисперсия неограниченно возрастает со временем.

Процесс случайного блуждания отличается от стационарного AR(1) процесса тем, что в (11.43) влияние возмущений ε_t не затухает: $y_t = \varepsilon_t + \varepsilon_{t-1} + \ldots$, в то время как в (11.36) влияние возмущений затухает со временем: $y_t = \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \ldots$ (при $m = 0$).

Легко показать, что процесс вида (11.36) с $|\phi| > 1$ тем более не является стационарным (и не встречается в реальных экономических примерах).

Единичные корни

Рассмотрим AR(1) процесс (11.36) в форме (11.37) с нулевым средним

$$A(L)y_t = (1 - \phi L)y_t = \varepsilon_t, \qquad \varepsilon_t \sim iid(0, \sigma^2). \tag{11.45}$$

В предыдущем параграфе мы видели, что для того чтобы процесс (11.45) был стационарным, необходимо условие $|\phi| < 1$, т. е. существование обратного оператора $A(L)^{-1} = (1 - \phi L)^{-1}$.

Возьмем другой пример — AR(2) процесс:

$$A(L)y_t = (1 - \phi_1 L - \phi_2 L^2)y_t = \varepsilon_t, \quad \varepsilon_t \sim iid(0, \sigma^2). \tag{11.46}$$

Как и всякий многочлен, $A(L)$ может быть разложен на множители над полем комплексных чисел:

$$A(L) = (1 - \phi_1 L - \phi_2 L^2) = (1 - \lambda_1 L)(1 - \lambda_2 L). \tag{11.47}$$

Нетрудно понять, что для существования обратного оператора $A(L)^{-1}$ необходима обратимость каждого сомножителя в (11.47), а это означает, что все λ_i по модулю меньше единицы. Часто это условие формулируется следующим образом: все корни $\mu_i = 1/\lambda_i$ многочлена $A(x)$ должны лежать вне единичного круга.

В предыдущем разделе мы видели, что наличие единичного корня в (11.45) существенно влияет на свойства процесса. Как определить по имеющимся наблюдениям верно ли, что в (11.45) $\phi = 1$? Из п. 3.5 мы знаем, как тестировать гипотезу подобного рода с помощью t-статистики $t = (\widehat{\phi} - \phi)/s_{\widehat{\phi}}$, которая имеет распределение Стьюдента и асимптотически стандартное нормальное распределение. Однако, как показали Дики и Фуллер (*D. A. Dickey, W. A. Fuller*) (см. Fuller, 1976), в случае, если истинное значение $\phi = 1$, то t-статистика не распределена по закону Стьюдента и ее распределение не стремится к стандартному нормальному при увеличении количества наблюдений.

Распределение t-статистики при условии $\phi = 1$ в (истинной) модели (11.45), описано Дики и Фуллером для уравнения (11.45) и двух его модификаций:

$$y_t = b_1 y_{t-1} + \varepsilon_{1t}, \tag{11.48}$$

$$y_t = a_2 + b_2 y_{t-1} + \varepsilon_{2t}, \tag{11.49}$$

$$y_t = a_2 + b_3 y_{t-1} + c_3 t + \varepsilon_{3t}. \tag{11.50}$$

Уравнение (11.49) соответствует (ошибочно) включенному свободному члену, в уравнение (11.50) кроме свободного члена включен также и временной тренд. В таблице 11.1 приведены односторонние критические значения статистики Дики–Фуллера (DF).

Таблица 11.1

Доверительный уровень	Размер выборки			
	25	50	100	∞
AR модель (11.48)				
0.010	−2.66	−2.62	−2.60	−2.58
0.025	−2.26	−2.25	−2.24	−2.23
0.050	−1.95	−1.95	−1.95	−1.95
AR модель с константой (11.49)				
0.010	−3.75	−3.58	−3.51	−3.43
0.025	−3.33	−3.22	−3.17	−3.12
0.050	−3.00	−2.93	−2.89	−2.86
AR модель с константой и трендом (11.50)				
0.010	−4.38	−4.15	−4.04	−3.96
0.025	−3.95	−3.80	−3.69	−3.66
0.050	−3.60	−3.50	−3.45	−3.41

Источник: (Fuller, 1976).

Предположим, мы тестируем гипотезу H_0: $\phi = 1$ против альтернативной гипотезы H_1: $\phi < 1$ для уравнения (11.49) при 100 наблюдениях и 5%-ном уровне значимости. В том случае, если мы используем стандартную процедуру тестирования гипотезы (п. 3.5), мы должны отвергнуть H_0 при значении t-статистики меньшем, чем −1.66. Однако, если мы используем таблицу 11.1, то мы должны отвергать H_0 при значении t-статистики меньшем, чем −2.89.

Таким образом, используя стандартную процедуру, мы часто (ошибочно) отвергаем верную гипотезу наличия единичного корня.

Удивительным фактом является то, что критические значения, указанные в таблице 11.1, остаются справедливыми, если в правые части регрессий (11.48)–(11.50) добавить слагаемые вида $\Delta y_{t-1}, \Delta y_{t-2}, \ldots$. Это позволяет тестировать наличие единичного корня в AR моделях порядка больше первого. Тест, соответствующий уравнению с лагированными значениями приращений $\Delta y_{t-1}, \Delta y_{t-2}, \ldots$ в правой части, называется *расширенным* тестом Дики–Фуллера (*augmented DF test, ADF*).

Если, например, в уравнении (11.46) есть один единичный корень, то $\lambda_1 = 1$ и $|\lambda_2| < 1$, а из $1 + \lambda_2 = \phi_1$ и $1 \cdot \lambda_2 = -\phi_2$ следует $\phi_1 + \phi_2 = 1$ и $|\phi_2| < 1$. Уравнение (11.46) может быть переписано в виде:

$$y_t = (\phi_1 + \phi_2) y_{t-1} - \phi_2 (y_{t-1} - y_{t-2}) + \varepsilon_t, \quad \varepsilon_t \sim iid(0, \sigma^2), \tag{11.51}$$

или:

$$\Delta y_t = (\phi_1 + \phi_2 - 1) y_{t-1} - \phi_2 \Delta y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim iid(0, \sigma^2). \tag{11.52}$$

Поэтому гипотеза наличия единичного корня может быть тестирована в духе процедуры Дики–Фуллера, описанной выше.

Если порядок процесса AR(p) заранее неизвестен, то рекомендуется включать возможно большее количество лагов, чтобы устранить возможную автокорреляцию ошибок. Дело в том, что в ADF тесте предполагается, что ошибки являются белым шумом и критические значения, указанные в таблице 11.1, справедливы только при этом условии. Однако включение чрезмерного количества лагов снижает мощность теста. Чтобы определить количество лагов, которое надо включить в уравнение, можно использовать критерии выбора порядка ARMA модели, описанные ниже, или статистическую значимость дополнительной лаговой переменной. Заметим, что тест Дики–Фуллера включен во все современные эконометрические пакеты.

Мнимая регрессия

В предыдущем разделе мы рассматривали проблемы, возникающие в авторегрессионных процессах с единичными корнями. Рассмотрим еще один пример регрессии, в которой участвуют нестационарные временные ряды. Возьмем два независимых случайных блуждания:

$$\begin{aligned} x_t &= x_{t-1} + \varepsilon_{1t}, \qquad \varepsilon_{1t} \sim iid(0, \sigma_1^2), \qquad t = 1, \ldots, n, \\ y_t &= y_{t-1} + \varepsilon_{2t}, \qquad \varepsilon_{2t} \sim iid(0, \sigma_2^2), \qquad t = 1, \ldots, n. \end{aligned} \tag{11.53}$$

Так как ε_1 и ε_2 независимы, то между x и y нет ничего общего. Предположим, что исследователь не знает механизмов, порождающих x и y, и оценивает регрессию:

$$y_t = \alpha + \beta x_t + \varepsilon_t. \tag{11.54}$$

В работе (Granger, Newbold, 1974) методами имитационного моделирования показано, что если тестировать значимость зависимости (11.54) при помощи t-статистики, то очень вероятно, что будет получен ложный вывод о наличии значимой связи.

Причиной этого является то, что ошибка ε_t является нестационарным процессом и поэтому (11.54) не удовлетворяет условиям классической регрессионной модели (постоянство дисперсии ошибок). В работе (Philips, 1986) показано, что асимптотическая теория для МНК-оценок уравнения (11.54) в этом случае совершенно другая. Например, t-статистика не имеет предельного распределения и расходится при $n \to \infty$. Поэтому, чем больше выборка, тем больше шансов прийти к ложному заключению. Такая ситуация называется «мнимая регрессия» (*spurious regression*). На практике признаками мнимой регрессии являются высокое значение R^2 и малое значение статистики Дарбина–Уотсона DW.

Пример. ***Мнимая регрессия.***

Сгенерируем два ряда наблюдений в соответствии с (11.53), где $y_0 = x_0 = 0$ и $\sigma_1 = \sigma_2 = 1$. Ниже приведены результаты регрессии для 300 наблюдений.

$$y_t = -\underset{(-5.77)}{2.79} - \underset{(-21.5)}{0.52}\, x_t; \quad R^2 = 0.607, \quad \mathrm{DW} = 0.057.$$

В скобках указаны t-статистики. В этом примере мы видим довольно типичное поведение МНК-оценок в случае мнимой регрессии.

Коинтеграция

Предыдущий пример показывает опасность, которая может встретиться в случае регрессии нестационарных рядов. Однако не всегда дело обстоит столь безнадежно. Одними из первых подход к регрессии нестационарных рядов предложили Энгель и Гранжер (Engel and Granger, 1987).

Предположим, у нас есть нестационарный ряд x_t. Возьмем его первые разности $\Delta x_t = x_t - x_{t-1}$. Если ряд Δx_t является стационарным, то x_t называется *интегрируемым порядка* 1 (*integrated order* 1), $I(1)$. Соответственно, стационарный ряд Δx_t называется $I(0)$. Вообще, ряд называется *интегрируемым порядка* k, $I(k)$, если он и его разности до порядка $k-1$ включительно нестационарны, а k-я разность стационарна.

Пусть теперь у нас есть два $I(1)$ ряда, x_t и y_t. Пусть, кроме того, их линейная комбинация $y_t - \beta x_t$ является стационарной, $I(0)$. В этом случае ряды x_t и y_t называются *коинтегрированными* (*cointegrated*), а вектор $(1, -\beta)'$ называется *коинтегрирующим вектором*. Оказывается, в этом случае можно получить состоятельную оценку $\widehat{\beta}$, применяя МНК к уравнению

$$y_t = \alpha + \beta x_t + \varepsilon_t, \qquad t = 1, \ldots, n. \tag{11.55}$$

Асимптотические свойства оценки будут при этом другие. Если обычно $\sqrt{n}(\widehat{\beta} - \beta)$ имеет предельное нормальное распределение, то в данном случае $n(\widehat{\beta} - \beta)$ имеет некоторое предельное распределение. Такая оценка называется *супер*состоятельной, так как сходится к истинному значению быстрее, чем в случае классической регрессии.

Таким образом, чтобы проверить наличие коинтеграции, надо рассмотреть остатки e_t, полученные при МНК-оценивании коинтегрирующего уравнения (11.55). Нулевой гипотезой является отсутствие коинтеграции, т. е. наличие единичного корня в ряде остатков e_t. Однако, как показано в работе (Philips and Ouliaris,

1990), к проверке ряда e_t нельзя применять DF или ADF тесты. Дело в том, что МНК «выбирает» остатки так, чтобы они имели наименьшую возможную вариацию, поэтому, даже если переменные не коинтегрированы, МНК делает остатки «похожими» на стационарные. Поэтому при использовании ADF теста гипотеза нестационарности отвергается слишком часто (и соответственно ошибочно принимается гипотеза наличия коинтеграции). В работах (MacKinnon, 1991; Davidson and MacKinnon, 1993) имитационным методом получены уточненные по сравнению с работой (Philips and Ouliaris, 1990) асимптотические критические значения t-статистики, подходящие для данного случая. Эти значения приведены в таблице 11.2.

Таблица 11.2

Число переменных	Тип теста	Доверительный уровень		
		0.01	0.05	0.10
2	константа	−3.90	−3.34	−3.04
	константа и тренд	−4.32	−3.78	−3.50
3	константа	−4.29	−3.74	−3.45
	константа и тренд	−4.66	−4.12	−3.84
4	константа	−4.64	−4.10	−3.81
	константа и тренд	−4.97	−4.43	−4.15
5	константа	−4.96	−4.42	−4.13
	константа и тренд	−5.25	−4.72	−4.43
6	константа	−5.25	−4.71	−4.42
	константа и тренд	−5.52	−4.98	−4.70

Источник: (Davidson and MacKinnon, 1993).

В первой графе таблицы приведено количество экзогенных переменных в уравнении коинтеграции (в случае уравнения (11.55) число переменных равно двум). Во второй графе приведен тип теста: в уравнение, тестирующее наличие единичных корней в остатках, могут быть включены константа и временной тренд (аналог уравнений (11.49) и (11.50)).

Понятие коинтегрируемости связано с концепцией долгосрочного динамического равновесия (*long-run equilibrium*). Если x_t и y_t коинтегрированы, то y_t и βx_t содержат общую нестационарную компоненту (долговременную тенденцию), а разность $y_t - \alpha - \beta x_t$ стационарна и совершает флуктуации около нуля.

11.4. Модели Бокса–Дженкинса (ARIMA)

В предыдущих разделах этой главы мы обсуждали применение методов регрессионного анализа к различным моделям, включающим в себя несколько временных рядов. В данном разделе мы рассмотрим модели временных рядов в узком смысле, т. е. модели, объясняющие поведение временного ряда, исходя исключительно из его значений в предыдущие моменты времени.

Как мы видели выше, статистические свойства стационарных и нестационарных временных рядов существенно отличаются, и для их моделирования должны применяться различные методы. В данном разделе мы в основном рассмотрим частный случай модели Бокса–Дженкинса, ARMA модели для стационарных временных рядов. Почему большое внимание уделяется именно моделям стационарных временных рядов? Дело в том, что многие временные ряды могут быть приведены к стационарному ряду после операций выделения тренда, сезонной компоненты или взятия разности.

Тренд, сезонность и взятие разности

Рассмотрим различные примеры нестационарных временных рядов.

Тренд

Рассмотрим следующий временной ряд:

$$y_t = \alpha + \beta t + \varepsilon_t. \tag{11.56}$$

Здесь ряд y_t представлен в виде композиции детерминированной составляющей $\alpha + \beta t$ (*линейный тренд*) и случайной составляющей ε_t, являющейся стационарным временным рядом с нулевым средним. Часто встречаются другие примеры тренда: квадратичный, $\alpha + \beta t + \gamma t^2$; экспоненциальный $\alpha e^{\beta t}$ и т. п.

Для того чтобы выделить тренд в модели (11.56) (и ей подобных), мы можем применить обычную технику оценивания параметров регрессионных уравнений, считая t независимой переменной. После этого мы получим ряд остатков, для описания которого можно будет применить модели стационарных временных рядов.

Сезонность

В экономических данных часто встречается сезонная компонента. Например, в квартальных данных может наблюдаться сезонная компонента с периодом 4:

$$y_t = S(t) + \varepsilon_t, \qquad S(t+4) \equiv S(t). \tag{11.57}$$

Здесь ряд y_t представлен в виде композиции периодической детерминированной составляющей $S(t)$ (*сезонная компонента*) и случайной составляющей ε_t, являющейся стационарным временным рядом с нулевым средним. Сезонную компоненту $S(t)$ можно представить в виде $S(t) = \beta_1 d_{1t} + \beta_2 d_{2t} + \beta_3 d_{3t} + \beta_4 d_{4t}$, где d_i — фиктивные (бинарные) переменные для кварталов (п. 4.2). Для выделения сезонной компоненты мы можем применить методы оценивания параметров регрессий к уравнению:

$$y_t = \beta_1 d_{1t} + \beta_2 d_{2t} + \beta_3 d_{3t} + \beta_4 d_{4t} + \varepsilon_t. \tag{11.58}$$

Часто модель (11.58) представляют в виде регрессии с ограничением, включая в нее константу:

$$y_t = \alpha + \beta_1 d_{1t} + \beta_2 d_{2t} + \beta_3 d_{3t} + \beta_4 d_{4t} + \varepsilon_t, \quad \sum \beta_i = 0. \tag{11.59}$$

В (11.59) коэффициенты β_i представляют отклонение от среднего за год уровня в квартале i.

Как и в случае выделения тренда, методы моделирования стационарных временных рядов применяются далее к ряду остатков регрессии (11.58).

Типичные примеры графиков нестационарных временных рядов приведены на рисунках 11.1–11.4.

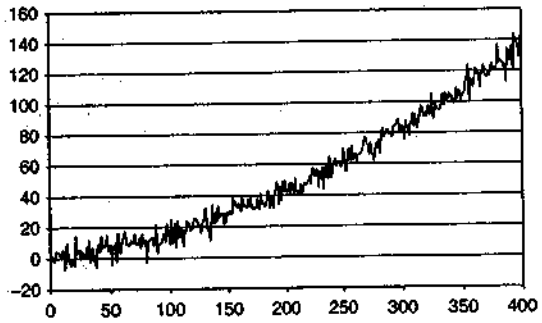


Рис. 11.1. Тренд (модельный пример)

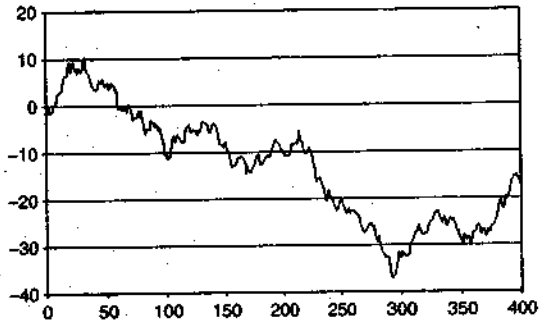


Рис. 11.2. Случайное блуждание (модельный пример)

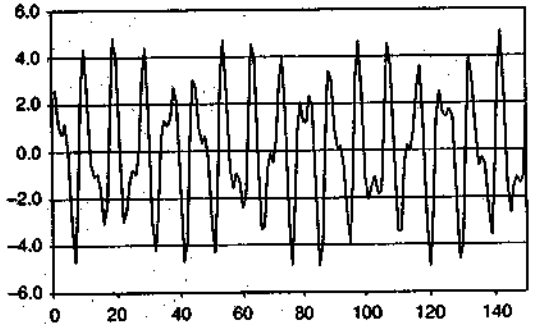


Рис. 11.3. Сезонность (солнечные пятна, числа Вольфа)

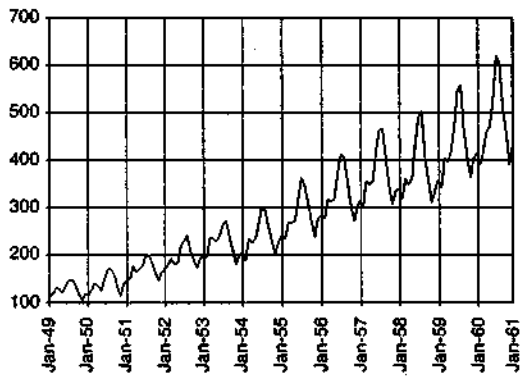


Рис. 11.4. Тренд и сезонность (объем авиаперевозок)

Взятие последовательной разности

Случайное блуждание (11.43) является примером нестационарного временного ряда. Однако, если к нему применить операцию взятия последовательной разности, получим стационарный временной ряд:

$$z_t = \Delta y_t = (1 - L)y_t = y_t - y_{t-1}, \qquad z_t = \varepsilon_t.$$

Для более общего процесса $A(L)y_t = \varepsilon_t$ (ср. (11.45)), где $A(L)$ имеет один единичный корень, т. е. $A(L) = B(L)(1 - L)$, а все корни $B(L)$ лежат вне единичного круга, это же преобразование $z_t = \Delta y_t$ приводит к стационарному процессу $B(L)z_t = \varepsilon_t$.

Взятие разности также приводит к стационарному процессу ряд (11.56) с линейным трендом:

$$\Delta y_t = \beta + u_t, \qquad u_t = \Delta\varepsilon_t = \varepsilon_t - \varepsilon_{t-1}. \tag{11.60}$$

В случае квадратичного тренда $\alpha+\beta t+\gamma t^2$ взятие первой разности не приводит к стационарному ряду, но если взять вторую разность $\Delta^2 y_t = \Delta(\Delta y_t) = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$, то

$$\Delta y_t = \beta + \gamma(2t - 1) + \Delta\varepsilon_t,$$
$$\Delta^2 y_t = 2\gamma + \Delta^2\varepsilon_t,$$

и $\Delta^2 y_t$ уже является стационарным временным рядом.

В случае наличия сезонной компоненты (11.57) устранить последнюю можно при помощи оператора взятия *сезонной* последовательной разности $\Delta_4 y_t = (1 - L^4)y_t = y_t - y_{t-4}$. (Конечно, если период сезонной компоненты равен 12, что может случиться для месячных данных, то надо применять оператор Δ_{12}.)

Заметим, что применение оператора последовательной разности не обязательно приводит нестационарный ряд к стационарному. Например, процесс

$$y_t = \beta y_{t-1} + \varepsilon_t, \quad \beta > 1 \tag{11.61}$$

не является стационарным: вычисляя дисперсию от обеих частей (11.61), получаем $V(y_t) = \beta^2 V(y_{t-1}) + \sigma_\varepsilon^2$, и если бы было выполнено условие стационарности $V(y_t) = V(y_{t-1})$, то $V(y_t) = \sigma_\varepsilon^2/(1 - \beta^2) < 0$. Полученная отрицательная дисперсия приводит нас к выводу, что $V(y_t) \neq V(y_{t-1})$, т. е. ряд (11.61) не является стационарным. Применив к (11.61) оператор разности, получим

$$\Delta y_t = \beta\Delta y_{t-1} + \Delta\varepsilon_t, \quad \beta > 1, \tag{11.62}$$

то есть процесс по-прежнему остался нестационарным. Ситуация даже осложнилась наличием в (11.62) корреляции ошибок. Повторное применение оператора разности также не приводит к стационарному ряду.

Таким образом, применяя выделение тренда, сезонности и/или оператор последовательной (и сезонной) разности, часто можно получить из исходного временного ряда стационарный.

Остается вопрос, как по имеющимся наблюдениям определить, является ли ряд стационарным.

Проверка на стационарность

Первое, что следует сделать, — посмотреть на график полученных наблюдений. Возможно, он содержит очевидный на глаз тренд или периодичную компоненту (сезонность). Также возможно, что разброс наблюдений возрастает или убывает со временем. Это может служить указанием на зависимость среднего или соответ-

ственно дисперсии от времени. В обоих случаях ряд будет, скорее всего, нестационарный.

Второе — построить график выборочной автокорреляционной функции (ACF) (ср. (11.42)), или *коррелограммы* (*correlogram*)

$$r_k = \widehat{\rho}_k = \frac{\sum_{t=k+1}^{n}(y_t - \overline{y})(y_{t-k} - \overline{y})}{\sum_{t=1}^{n}(y_t - \overline{y})^2}, \quad k = 1, 2, \ldots . \qquad (11.63)$$

Коррелограмма стационарного временного ряда «быстро убывает» с ростом k после нескольких первых значений. Если же график убывает достаточно медленно, то есть основания предположить нестационарность ряда. Кроме ACF, можно также построить график частной автокорреляционной функции, PACF, которая также должна быстро убывать для стационарного процесса.

Частная автокорреляционная функция (PACF)

В главе 4 (п. 4.3) было введено понятие частного коэффициента корреляции. Содержательно *частная автокорреляционная функция* PACF(k) (*partial autocorrelation function*) есть «чистая корреляция» между y_t и y_{t-k} при исключении влияния промежуточных значений $y_{t-1}, y_{t-2}, \ldots, y_{t-k+1}$.

Если применить процедуру вычисления выборочного частного коэффициента корреляции (см. п. 4.3), то оказывается, что в случае стационарного ряда y_t значение выборочной частной автокорреляционной функции PACF(k) вычисляется как МНК-оценка последнего коэффициента β_k в AR(k) регрессионном уравнении:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_k y_{t-k} + \varepsilon_t. \qquad (11.64)$$

На рисунках 11.5–11.8 приведены автокорреляционные и частные автокорреляционные функции нестационарных временных рядов, представленных на рисунках 11.1–11.4.

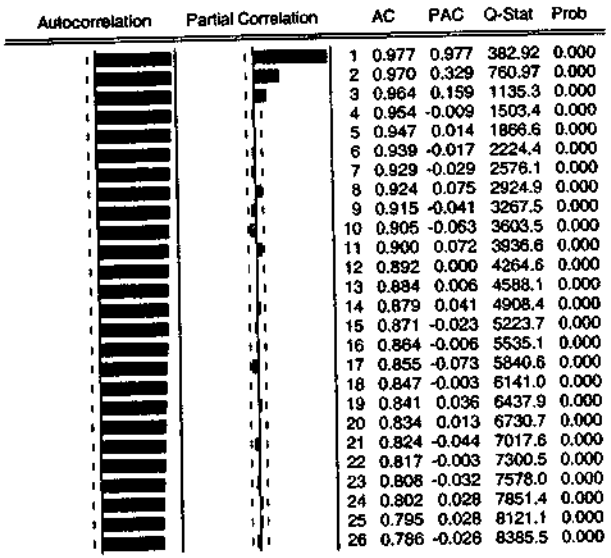


Рис. 11.5. Тренд (модельный пример)

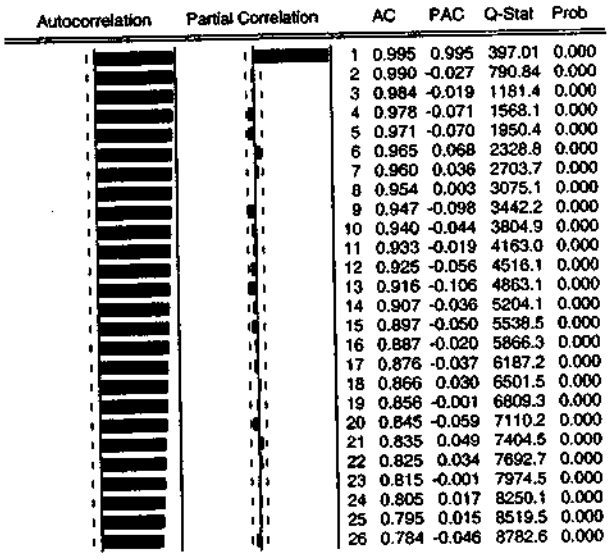
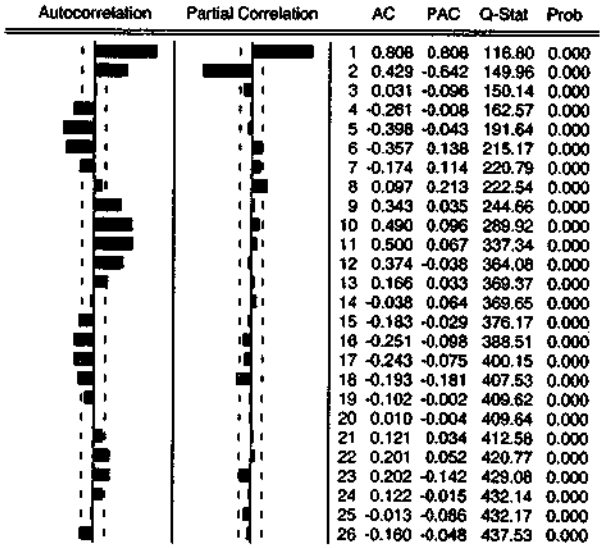
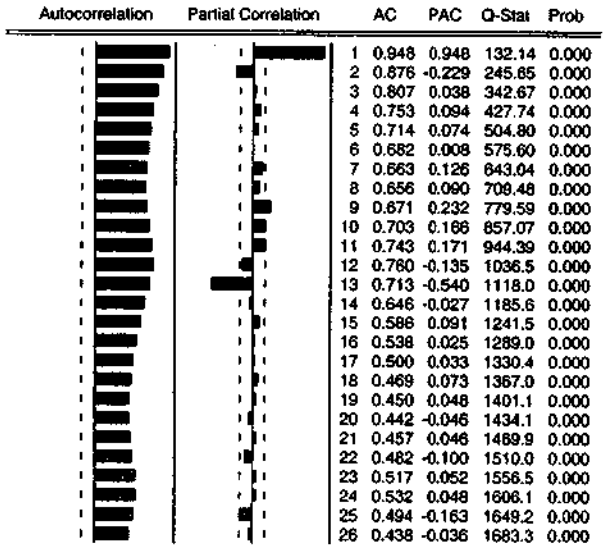


Рис. 11.6. Случайное блуждание (модельный пример)



	AC	PAC	Q-Stat	Prob
1	0.808	0.808	116.80	0.000
2	0.429	-0.642	149.96	0.000
3	0.031	-0.096	150.14	0.000
4	-0.261	-0.008	162.57	0.000
5	-0.398	-0.043	191.64	0.000
6	-0.357	0.138	215.17	0.000
7	-0.174	0.114	220.79	0.000
8	0.097	0.213	222.54	0.000
9	0.343	0.035	244.66	0.000
10	0.490	0.096	289.92	0.000
11	0.500	0.067	337.34	0.000
12	0.374	-0.038	364.08	0.000
13	0.166	0.033	369.37	0.000
14	-0.038	0.064	369.65	0.000
15	-0.183	-0.029	376.17	0.000
16	-0.251	-0.098	388.51	0.000
17	-0.243	-0.075	400.15	0.000
18	-0.193	-0.181	407.53	0.000
19	-0.102	-0.002	409.62	0.000
20	0.010	-0.004	409.64	0.000
21	0.121	0.034	412.58	0.000
22	0.201	0.052	420.77	0.000
23	0.202	-0.142	429.08	0.000
24	0.122	-0.015	432.14	0.000
25	-0.013	-0.086	432.17	0.000
26	-0.160	-0.048	437.53	0.000

Рис. 11.7. Сезонность (солнечные пятна, числа Вольфа)



	AC	PAC	Q-Stat	Prob
1	0.948	0.948	132.14	0.000
2	0.876	-0.229	245.65	0.000
3	0.807	0.038	342.67	0.000
4	0.753	0.094	427.74	0.000
5	0.714	0.074	504.80	0.000
6	0.682	0.008	575.60	0.000
7	0.663	0.126	643.04	0.000
8	0.656	0.090	709.48	0.000
9	0.671	0.232	779.59	0.000
10	0.703	0.166	857.07	0.000
11	0.743	0.171	944.39	0.000
12	0.760	-0.135	1036.5	0.000
13	0.713	-0.540	1118.0	0.000
14	0.646	-0.027	1185.6	0.000
15	0.586	0.091	1241.5	0.000
16	0.538	0.025	1289.0	0.000
17	0.500	0.033	1330.4	0.000
18	0.469	0.073	1367.0	0.000
19	0.450	0.048	1401.1	0.000
20	0.442	-0.046	1434.1	0.000
21	0.457	0.046	1469.9	0.000
22	0.482	-0.100	1510.0	0.000
23	0.517	0.052	1556.5	0.000
24	0.532	0.048	1606.1	0.000
25	0.494	-0.163	1649.2	0.000
26	0.438	-0.036	1683.3	0.000

Рис. 11.8. Тренд и сезонность (объем авиаперевозок)

Третье — можно использовать формальные тесты на наличие единичного корня (тест Дики–Фуллера DF, расширенный тест Дики–Фуллера ADF, тест МакКинли) и др., часть из которых рассмотрена выше в разделе 11.3.

Модели авторегрессии и скользящего среднего (ARMA)

Рассмотрим следующий класс моделей стационарных временных рядов:

$$y_t - \phi_1 y_{t-1} - \dots - \phi_p y_{t-p} = \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q},$$
$$\varepsilon_t \sim iid(0, \sigma^2), \tag{11.65}$$

или в более короткой записи

$$\Phi(L) y_t = \delta + \Theta(L)\varepsilon_t, \qquad \varepsilon_t \sim iid(0, \sigma^2), \tag{11.66}$$

где $\Phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$ и $\Theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q$ — полиномы от оператора сдвига. Такая модель называется *моделью авторегрессии и скользящего среднего* (*autoregressive moving average*) или ARMA(p, q).

Рассмотрим сначала простые примеры ARMA моделей.

AR(1)

Процесс ARMA(1,0)

$$y_t = \delta + \phi_1 y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim iid(0, \sigma^2) \tag{11.67}$$

является AR(1) процессом и подробно рассмотрен ранее (п. 11.3), (11.35)–(11.42). Перечислим кратко его свойства:

$$\mathrm{E}(y_t) = \frac{\delta}{1 - \phi_1}, \quad \mathrm{V}(y_t) = \gamma_0 = \frac{\sigma^2}{1 - \phi_1^2},$$
$$\gamma_k = \phi_1^k \gamma_0, \quad \rho_k = \gamma_k / \gamma_0 = \phi_1^k. \tag{11.68}$$

Неравенство $|\phi_1| < 1$ является необходимым условием стационарности процесса y_t.

Частная автокорреляционная функция процесса AR(1) равна нулю для значений $k > 1$. (По определению ACF(1)=PACF(1).)

AR(2)

Возьмем в качестве примера авторегрессионного процесса высокого порядка процесс AR(2) (для простоты положим свободный член равным нулю):

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t, \qquad \varepsilon_t \sim iid(0, \sigma^2). \tag{11.69}$$

Для $k > 0$ вычислим ковариацию обеих частей (11.69) с y_{t-k}:

$$\begin{aligned}\gamma_k &= \mathrm{Cov}(y_t, y_{t-k}) = \mathrm{Cov}(\phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t, y_{t-k}) \\ &= \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2},\end{aligned} \tag{11.70}$$

разделив на γ_0, получим

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, \quad k = 1, 2, \ldots . \tag{11.71}$$

Взяв (11.71) при $k = 1, 2$ и учитывая, что $\rho_0 = 1$, $\rho_{-1} = \rho_1$, получаем систему уравнений с неизвестными ρ_1 и ρ_2:

$$\begin{aligned}\rho_1 &= \phi_1 + \phi_2 \rho_1, \\ \rho_2 &= \phi_1 \rho_1 + \phi_2.\end{aligned} \tag{11.72}$$

Система (11.72) называется системой уравнений Юла–Уолкера (*Yule–Walker*) для AR(2) процесса. Решая эту систему, найдем два первых значения автокорреляционной функции

$$\rho_1 = \frac{\phi_1}{1 - \phi_2}, \qquad \rho_2 = \frac{\phi_1^2}{1 - \phi_2} + \phi_2. \tag{11.73}$$

Следующие значения автокорреляционной функции вычисляются по формуле (11.71). Если умножить обе части (11.69) на y_t и взять математическое ожидание, получим следующее выражение для дисперсии y_t:

$$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \sigma^2. \tag{11.74}$$

Решая это уравнение совместно с двумя уравнениями (11.70) для $k = 1, 2$, получаем:

$$\gamma_0 = \frac{(1 - \phi_2)\sigma^2}{(1 + \phi_2)(1 - \phi_1 - \phi_2)(1 + \phi_1 - \phi_2)}. \tag{11.75}$$

Отсюда, учитывая, что дисперсия должна быть положительна, получаем условия стационарности AR(2) процесса:

$$|\phi_2| < 1, \quad \phi_2 + \phi_1 < 1, \quad \phi_2 - \phi_1 < 1. \tag{11.76}$$

Можно показать, что при выполнении условий стационарности автокорреляционная функция процесса убывает экспоненциально в случае, когда корни характеристического полинома $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2$ действительны, или изменяется по синусоиде с экспоненциально убывающей амплитудой, если корни комплексные.

Опишем схему вычисления частной автокорреляционной функции для процесса AR(2). Запишем три уравнения Юла–Уолкера (типа (11.72)) для AR(3) процесса. Коэффициент ϕ_3 равен коэффициенту частной корреляции между y_t и y_{t-3}. Для AR(2) процесса из (11.71) получаем $\rho_3 = \phi_1 \rho_2 + \phi_2 \rho_1$. Подставляя это выражение в третье уравнение Юла–Уолкера, получаем $\phi_3 = 0$. Таким образом, PACF(k)=0 для $k > 2$. Аналогично, можно показать, что для AR(p) процесса частная автокорреляционная функция PACF(k) равна нулю, начиная с $k = p+1$. Следует иметь в виду, что этот результат верен для теоретической частной автокорреляционной функции и может не выполняться для выборочной частной автокорреляционной функции. Однако на практике следует ожидать резкое убывание PACF до значений, близких к нулю, за порогом, равным порядку авторегрессионного процесса.

Процессы скользящего среднего (MA)

Моделью *скользящего среднего* (*moving average*) порядка q называется модель ARMA$(0, q)$

$$\begin{aligned}y_t &= \delta + \Theta(L)\varepsilon_t \\ &= \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}, \quad \varepsilon_t \sim iid(0, \sigma^2),\end{aligned} \tag{11.77}$$

которая обозначается MA(q). Из (11.77) видно, что процесс MA(q) стационарен при любом q и любых θ_i.

Сформулируем условие *обратимости* процесса, т.е. возможности его представления в виде AR процесса.

Рассмотрим в качестве примера модель скользящего среднего первого порядка MA(1)

$$y_t = \delta + \Theta(L)\varepsilon_t = \delta + \varepsilon_t - \theta_1\varepsilon_{t-1}, \quad \varepsilon_t \sim iid(0, \sigma^2). \tag{11.78}$$

Представим MA(1) процесс в виде авторегрессионного процесса:

$$\Theta(L)^{-1}y_t = \Theta(L)^{-1}\delta + \varepsilon_t, \tag{11.79}$$

или

$$y_t = \frac{\delta}{1-\theta_1} - \theta_1 y_{t-1} - \theta_1^2 y_{t-2} - \cdots + \varepsilon_t. \tag{11.80}$$

Ясно, что такое AR(∞) представление MA(1) процесса (11.78) возможно только в случае обратимости оператора $\Theta(L) = 1 - \theta_1 L$, т. е. когда выполняется *условие обратимости* $|\theta_1| < 1$.

Нетрудно вычислить среднее и дисперсию процесса MA(1):

$$\mathrm{E}(y_t) = \delta, \qquad \mathrm{V}(y_t) = \sigma^2(1+\theta_1^2). \tag{11.81}$$

Найдем автокорреляционную функцию MA(1) процесса:

$$\gamma_1 = \mathrm{Cov}(y_t, y_{t-1}) = \mathrm{E}((\varepsilon_t - \theta_1\varepsilon_{t-1})(\varepsilon_{t-1} - \theta_1\varepsilon_{t-2})). \tag{11.82}$$

Если раскрыть скобки, то только одно слагаемое из четырех будет отлично от нуля: $\mathrm{E}(-\theta_1\varepsilon_{t-1}^2) = -\theta_1\sigma^2$. Поэтому

$$\gamma_1 = \mathrm{Cov}(y_t, y_{t-1}) = -\theta_1\sigma^2. \tag{11.83}$$

Аналогичные вычисления показывают, что $\gamma_k = 0$ при $k > 1$. Получаем:

$$\rho_1 = \gamma_1/\gamma_0 = -\theta_1/(1+\theta_1^2), \qquad \rho_k = 0, \quad k > 1. \tag{11.84}$$

Проделав аналогичные вычисления для MA(q) процесса, получим, что его автокорреляционная функция ACF(k) равна 0 для $k > q$, т. е. ее вид аналогичен виду PACF для AR(q) процесса.

Частная автокорреляционная функция PACF(k) для MA(q) процесса, аналогично ACF(k) для AR(q) процесса, экспоненциально убывает. Таким образом, имеет место некоторая симметрия: пара графиков (ACF, PACF) для MA(q) процесса имеет такой же вид, как пара графиков (PACF, ACF) для AR(q).

Отметим, что подобно AR(∞) представлению (11.80) для MA(1) процесса (11.78) существует MA(∞) представление для AR(1) процесса (11.67):

$$y_t = (1-\phi_1 L)^{-1}(\delta+\varepsilon_t) = \frac{\delta}{1-\phi_1} + \varepsilon_t + \phi_1\varepsilon_{t-1} + \phi_1^2\varepsilon_{t-2} + \cdots. \tag{11.85}$$

Смешанные процессы

Рассмотрим простейший смешанный ARMA(1,1) процесс (11.66) с $\Phi(L) = 1 - \phi_1 L$ и $\Theta(L) = 1 - \theta_1 L$:

$$y_t - \phi_1 y_{t-1} = \delta + \varepsilon_t - \theta_1\varepsilon_{t-1}, \quad \varepsilon_t \sim iid(0, \sigma^2). \tag{11.86}$$

Будем считать, что $|\phi_1| < 1$ и $|\theta_1| < 1$. Как и в случае AR(1) и MA(1) процессов, можно показать, что тогда процесс ARMA(1,1) является стационарным и обратимым.

Применяя те же методы, что и ранее, получим следующие выражения для среднего, дисперсии и ковариации ARMA(1,1) процесса:

$$\mathrm{E}(y_t) = \frac{\delta}{1-\phi_1}, \tag{11.87а}$$

$$\gamma_0 = \mathrm{V}(y_t) = \sigma^2\frac{1+\theta_1^2 - 2\phi_1\theta_1}{1-\phi_1^2}, \tag{11.87б}$$

$$\gamma_1 = \mathrm{Cov}(y_t, y_{t-1}) = \phi_1\gamma_0 - \theta_1\sigma^2. \tag{11.87в}$$

Для автокорреляций порядка больше 1 получаем рекуррентное соотношение

$$\gamma_k = \mathrm{Cov}(y_t, y_{t-1}) = \phi_1\gamma_{k-1}, \quad k > 1.$$

Применяя рекуррентно это соотношение, получаем:

$$\rho_k = \phi_1^{k-1}\rho_1, \quad k > 1, \qquad \rho_1 = \frac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1+\theta_1^2 - 2\phi_1\theta_1}. \tag{11.88}$$

Из (11.88) видно, что ACF для ARMA(1,1) процесса ведет себя так же, как ACF для AR(1) процесса (ср.(11.68)). Хотя значение ρ_1 другое, но соотношение между ρ_1 и последующими значениями ACF точно такое же.

Этот вывод можно обобщить на случай ARMA(p,q) процесса. Первые q значений ACF определяются взаимодействием AR и MA компонент, а дальнейшее поведение автокорреляционной функции такое же, как в AR(p) процессе.

Аналогичный вывод справедлив для частичной автокорреляционной функции ARMA(p,q) процесса. Она убывает подобно PACF для MA(q) процесса.

Методология Бокса–Дженкинса (ARIMA)

ARIMA модели

Как мы видели выше, некоторые нестационарные временные ряды могут быть приведены к стационарным при помощи оператора последовательной разности. Предположим, что временной ряд y_t после того, как к нему применили d раз оператор последовательной разности, стал стационарным рядом $\Delta^d y_t$, удовлетворяющим ARMA(p,q) модели (11.65). Тогда процесс y_t называется *интегрированным процессом авторегрессии и скользящего среднего* (*integrated autoregression and moving average*), ARIMA(p,d,q). (Заметим, из модели, например, для ряда Δy_t легко получить модель для исходного ряда y_t, используя соотношение $y_t = y_{t-1} + \Delta y_t$.)

Методология Бокса–Дженкинса (Box, Jenkins, 1976) подбора ARIMA модели для данного ряда наблюдений состоит из трех этапов.

I. *Идентификация модели*

I.1. Первый шаг — получение стационарного ряда. Мы тестируем ряд на стационарность, используя описанные выше методы: визуальный анализ графика, визуальный анализ ACF и PACF, тесты на единичные корни. Если получается стационарный ряд, то переходим к следующему пункту, если нет, то применяем оператор взятия последовательной разности и повторяем тестирование. На практике последовательная разность берется, как правило, не более двух раз.

I.2. После того как получен стационарный временной ряд, строятся его выборочные ACF и PACF, которые, как было показано выше, являются своеобразными «отпечатками пальцев» ARMA(p,q) процесса и позволяют сформулировать несколько гипотез о возможных порядках авторегрессии (p) и скользящего среднего (q). Обычно рекомендуется использовать модели возможно более низкого порядка, как правило, с $p+q \leqslant 3$ (если нет сезонной компоненты).

Выборочные ACF и PACF, конечно, не обязаны в точности следовать теоретическим аналогам, но должны быть «достаточно близки» к ним. На рисунках 11.9–11.21 представлены выборочные ACF и PACF, построенные по модельным примерам.

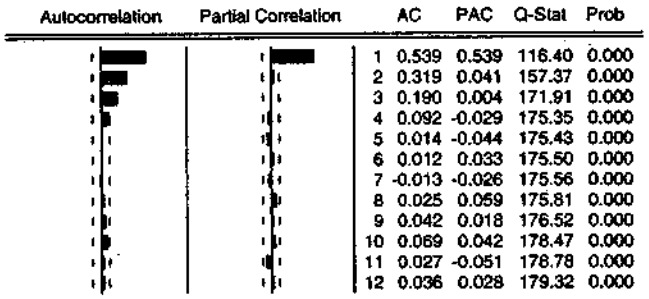


Рис. 11.9. AR(1). $Y_t = 0.5Y_{t-1} + \varepsilon_t$. Корень $\mu = 2$

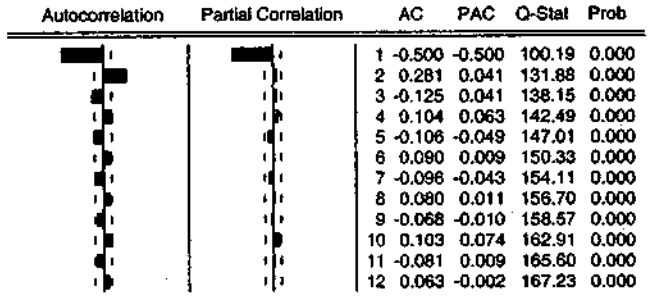
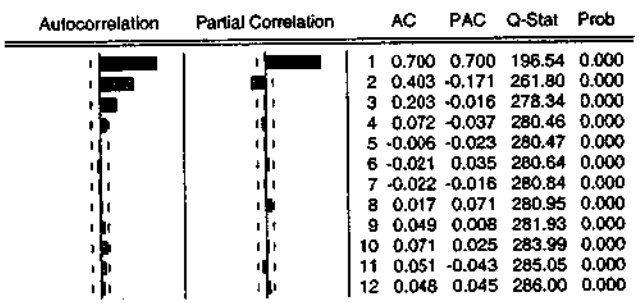
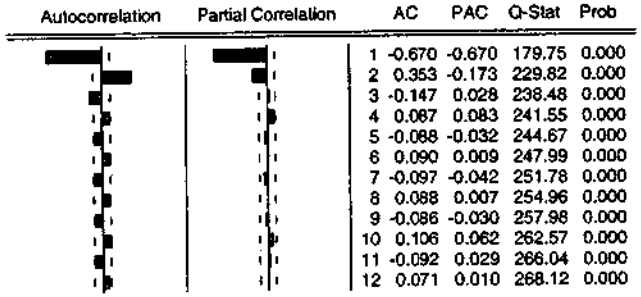


Рис. 11.10. AR(1). $Y_t = -0.5Y_{t-1} + \varepsilon_t$. Корень $\mu = -2$



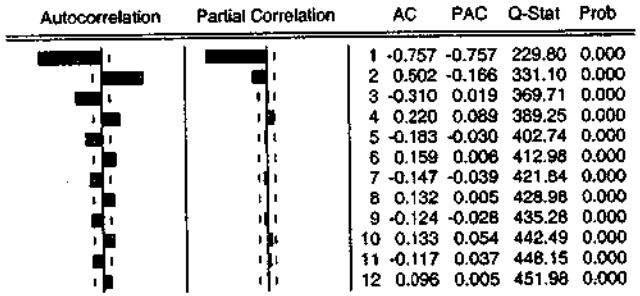
	AC	PAC	Q-Stat	Prob
1	0.700	0.700	196.54	0.000
2	0.403	-0.171	261.80	0.000
3	0.203	-0.016	278.34	0.000
4	0.072	-0.037	280.46	0.000
5	-0.006	-0.023	280.47	0.000
6	-0.021	0.035	280.64	0.000
7	-0.022	-0.016	280.84	0.000
8	0.017	0.071	280.95	0.000
9	0.049	0.008	281.93	0.000
10	0.071	0.025	283.99	0.000
11	0.051	-0.043	285.05	0.000
12	0.048	0.045	286.00	0.000

Рис. 11.11. AR(2). $Y_t = 0.8Y_{t-1} - 0.2Y_{t-2} + \varepsilon_t$. Корни $\mu_1 = 2 + i, \mu_2 = 2 - i$



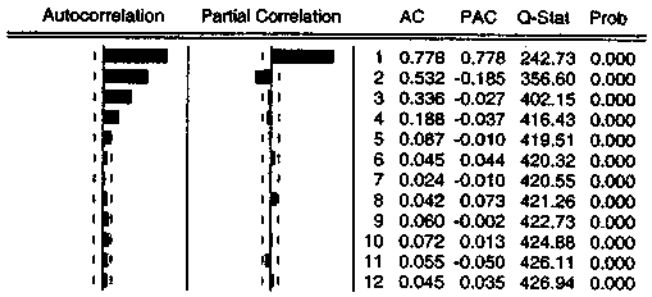
	AC	PAC	Q-Stat	Prob
1	-0.670	-0.670	179.75	0.000
2	0.353	-0.173	229.82	0.000
3	-0.147	0.028	238.48	0.000
4	0.087	0.083	241.55	0.000
5	-0.088	-0.032	244.67	0.000
6	0.090	0.009	247.99	0.000
7	-0.097	-0.042	251.78	0.000
8	0.088	0.007	254.96	0.000
9	-0.086	-0.030	257.98	0.000
10	0.106	0.062	262.57	0.000
11	-0.092	0.029	266.04	0.000
12	0.071	0.010	268.12	0.000

Рис. 11.12. AR(2). $Y_t = -0.8Y_{t-1} - 0.2Y_{t-2} + \varepsilon_t$. Корни $\mu_1 = -2 + i, \mu_2 = -2 - i$



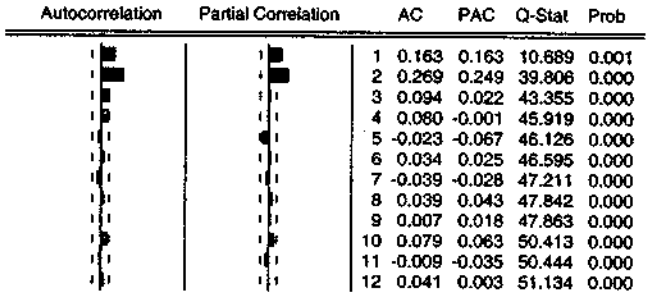
	AC	PAC	Q-Stat	Prob
1	-0.757	-0.757	229.80	0.000
2	0.502	-0.166	331.10	0.000
3	-0.310	0.019	369.71	0.000
4	0.220	0.089	389.25	0.000
5	-0.183	-0.030	402.74	0.000
6	0.159	0.006	412.98	0.000
7	-0.147	-0.039	421.84	0.000
8	0.132	0.005	428.98	0.000
9	-0.124	-0.028	435.28	0.000
10	0.133	0.054	442.49	0.000
11	-0.117	0.037	448.15	0.000
12	0.096	0.005	451.98	0.000

Рис. 11.13. AR(2). $Y_t = -0.9Y_{t-1} - 0.2Y_{t-2} + \varepsilon_t$. Корни $\mu_1 = -2.5$, $\mu_2 = -2$



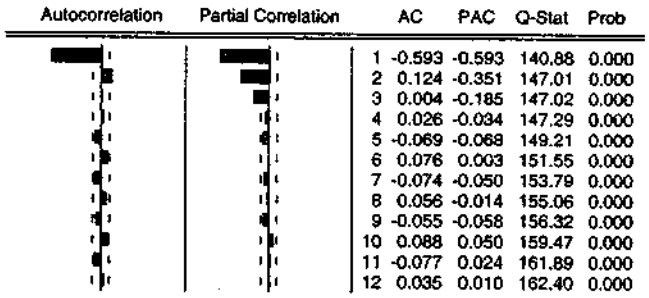
	AC	PAC	Q-Stat	Prob
1	0.778	0.778	242.73	0.000
2	0.532	-0.185	356.60	0.000
3	0.336	-0.027	402.15	0.000
4	0.188	-0.037	416.43	0.000
5	0.087	-0.010	419.51	0.000
6	0.045	0.044	420.32	0.000
7	0.024	-0.010	420.55	0.000
8	0.042	0.073	421.26	0.000
9	0.060	-0.002	422.73	0.000
10	0.072	0.013	424.88	0.000
11	0.055	-0.050	426.11	0.000
12	0.045	0.035	426.94	0.000

Рис. 11.14. AR(2). $Y_t = 0.9Y_{t-1} - 0.2Y_{t-2} + \varepsilon_t$. Корни $\mu_1 = 2.5$, $\mu_2 = 2$



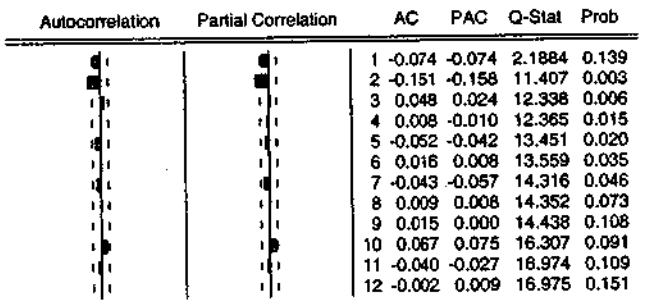
	AC	PAC	Q-Stat	Prob
1	0.163	0.163	10.689	0.001
2	0.269	0.249	39.806	0.000
3	0.094	0.022	43.355	0.000
4	0.080	-0.001	45.919	0.000
5	-0.023	-0.067	46.126	0.000
6	0.034	0.025	46.595	0.000
7	-0.039	-0.028	47.211	0.000
8	0.039	0.043	47.842	0.000
9	0.007	0.018	47.863	0.000
10	0.079	0.063	50.413	0.000
11	-0.009	-0.035	50.444	0.000
12	0.041	0.003	51.134	0.000

Рис. 11.15. AR(2). $Y_t = 0.1Y_{t-1} + 0.2Y_{t-2} + \varepsilon_t$. Корни $\mu_1 = -2.5$, $\mu_2 = 2$



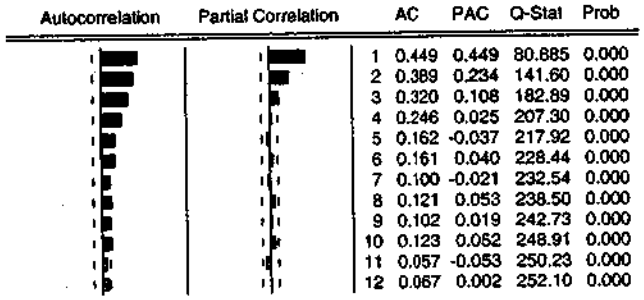
	AC	PAC	Q-Stat	Prob
1	-0.593	-0.593	140.88	0.000
2	0.124	-0.351	147.01	0.000
3	0.004	-0.185	147.02	0.000
4	0.026	-0.034	147.29	0.000
5	-0.069	-0.068	149.21	0.000
6	0.076	0.003	151.55	0.000
7	-0.074	-0.050	153.79	0.000
8	0.056	-0.014	155.06	0.000
9	-0.055	-0.058	156.32	0.000
10	0.088	0.050	159.47	0.000
11	-0.077	0.024	161.89	0.000
12	0.035	0.010	162.40	0.000

Рис. 11.16. MA(2). $Y_t = \varepsilon_t - 0.9\varepsilon_{t-1} + 0.2\varepsilon_{t-2}$. Корни $\mu_1 = 2.5$, $\mu_2 = 2$



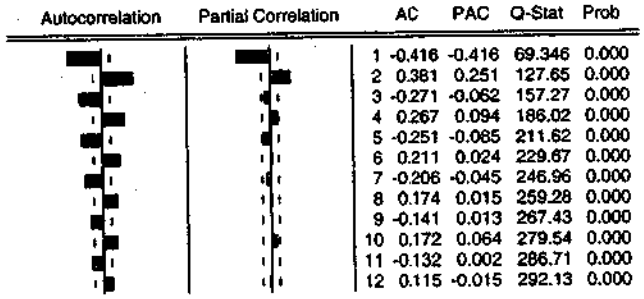
	AC	PAC	Q-Stat	Prob
1	-0.074	-0.074	2.1884	0.139
2	-0.151	-0.158	11.407	0.003
3	0.048	0.024	12.338	0.006
4	0.008	-0.010	12.365	0.015
5	-0.052	-0.042	13.451	0.020
6	0.016	0.008	13.559	0.035
7	-0.043	-0.057	14.316	0.046
8	0.009	0.008	14.352	0.073
9	0.015	0.000	14.438	0.108
10	0.067	0.075	16.307	0.091
11	-0.040	-0.027	16.974	0.109
12	-0.002	0.009	16.975	0.151

Рис. 11.17. MA(2). $Y_t = \varepsilon_t - 0.1\varepsilon_{t-1} - 0.2\varepsilon_{t-2}$. Корни $\mu_1 = -2.5, \mu_2 = 2$



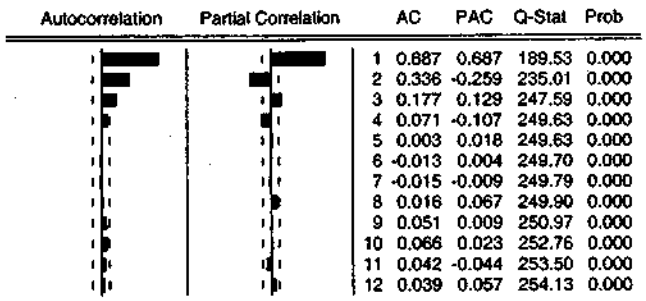
	AC	PAC	Q-Stat	Prob
1	0.449	0.449	80.885	0.000
2	0.389	0.234	141.60	0.000
3	0.320	0.108	182.89	0.000
4	0.246	0.025	207.30	0.000
5	0.162	-0.037	217.92	0.000
6	0.161	0.040	228.44	0.000
7	0.100	-0.021	232.54	0.000
8	0.121	0.053	238.50	0.000
9	0.102	0.019	242.73	0.000
10	0.123	0.052	248.91	0.000
11	0.057	-0.053	250.23	0.000
12	0.067	0.002	252.10	0.000

Рис. 11.18. ARMA(1,1). $Y_t = 0.8Y_{t-1} + \varepsilon_t - 0.5\varepsilon_{t-1}$. Корни $\mu_{AR} = 1.125, \mu_{MA} = 2$



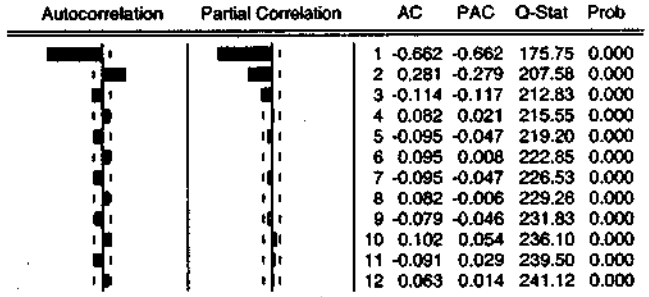
	AC	PAC	Q-Stat	Prob
1	-0.416	-0.416	69.346	0.000
2	0.381	0.251	127.65	0.000
3	-0.271	-0.062	157.27	0.000
4	0.267	0.094	186.02	0.000
5	-0.251	-0.065	211.62	0.000
6	0.211	0.024	229.67	0.000
7	-0.206	-0.045	246.96	0.000
8	0.174	0.015	259.28	0.000
9	-0.141	0.013	267.43	0.000
10	0.172	0.064	279.54	0.000
11	-0.132	0.002	286.71	0.000
12	0.115	-0.015	292.13	0.000

Рис. 11.19. ARMA(1,1). $Y_t = -0.8Y_{t-1} + \varepsilon_t + 0.5\varepsilon_{t-1}$. Корни $\mu_{AR} = 1.125, \mu_{MA} = -2$



	AC	PAC	Q-Stat	Prob
1	0.687	0.687	189.53	0.000
2	0.336	-0.259	235.01	0.000
3	0.177	0.129	247.59	0.000
4	0.071	-0.107	249.63	0.000
5	0.003	0.018	249.63	0.000
6	-0.013	0.004	249.70	0.000
7	-0.015	-0.009	249.79	0.000
8	0.016	0.067	249.90	0.000
9	0.051	0.009	250.97	0.000
10	0.066	0.023	252.76	0.000
11	0.042	-0.044	253.50	0.000
12	0.039	0.057	254.13	0.000

Рис. 11.20. ARMA(1,1). $Y_t = 0.4Y_{t-1} + \varepsilon_t + 0.5\varepsilon_{t-1}$. Корни $\mu_{AR} = 2, \mu_{MA} = -2$



	AC	PAC	Q-Stat	Prob
1	-0.662	-0.662	175.75	0.000
2	0.281	-0.279	207.58	0.000
3	-0.114	-0.117	212.83	0.000
4	0.082	0.021	215.55	0.000
5	-0.095	-0.047	219.20	0.000
6	0.095	0.008	222.85	0.000
7	-0.095	-0.047	226.53	0.000
8	0.082	-0.006	229.26	0.000
9	-0.079	-0.046	231.83	0.000
10	0.102	0.054	236.10	0.000
11	-0.091	0.029	239.50	0.000
12	0.063	0.014	241.12	0.000

Рис. 11.21. ARMA(1,1). $Y_t = -0.4Y_{t-1} + \varepsilon_t - 0.5\varepsilon_{t-1}$. Корни $\mu_{AR} = -2, \mu_{MA} = 2$

II. *Оценивание модели и проверка адекватности модели*

II.1. Для каждой из выбранных на первом этапе моделей оцениваются их параметры и вычисляются остатки.

II.2. Каждая из моделей проверяется, насколько она соответствует данным. Из моделей, адекватных данным, выбирается самая простая модель, т. е. модель с наименьшим количеством параметров.

III. *Прогнозирование*

После того как на втором этапе выбрана модель, можно строить прогноз на один или несколько шагов по времени и оценивать доверительные границы прогнозных значений.

Остановимся подробнее на втором и третьем этапах методики Бокса–Дженкинса.

Оценивание ARMA моделей

В современные компьютерные пакеты включены различные методы оценивания ARMA моделей, такие, как линейный или нелинейный МНК, полный или условный метод максимального правдоподобия.

Рассмотрим пример ARMA(1,1) модели (11.86). Запишем ее в виде:

$$\Theta(L)^{-1}y_t = \Theta(L)^{-1}(\delta + \phi_1 y_{t-1}) + \varepsilon_t, \qquad (11.89)$$

где $\Theta(L) = 1 - \theta_1 L$ и $\Theta(L)^{-1} = 1 + \theta_1 L + \theta_1^2 L^2 + \cdots$. В (11.89) надо каким-то образом интерпретировать переменную $y_t^* = \Theta(L)^{-1}y_t$, которая является бесконечной взвешенной суммой предыдущих значений y_t. Одним из возможных решений является следующее. Приравняем нулю все значения, предшествующие началу наблюдений: $y_0 = y_{-1} = \ldots = 0$. При этом получим:

$$y_1^* = y_1, \quad y_2^* = y_2 + \theta_1 y_1, \quad \ldots \quad y_t^* = y_t + \theta_1 y_{t-1} + \ldots + \theta_1^{t-1} y_1.$$

В этих обозначениях уравнение (11.89) принимает вид

$$y_t^* = \delta^* + \phi_1 y_{t-1}^* + \varepsilon_t, \quad \delta^* = \frac{\delta}{1-\theta_1}. \qquad (11.90)$$

В том случае, если θ_1 известно, это уравнение является линейным по δ^*, ϕ_1, однако в общем случае оно нелинейно по параметрам.

Для оценивания уравнения (11.89) применим *условный метод максимального правдоподобия* (*conditional ML*), когда y_1 предполагается заданным, считая, что ошибки $\varepsilon_t \sim iidN(0, \sigma^2)$. Условная

функция правдоподобия равна

$$L^* = p(y_2^*, y_3^*, \ldots, y_n^* | y_1^*) = \prod_{t=2}^{n} p(y_t^* | y_{t-1}^*)$$
$$= \prod_{t=2}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y_t^* - \delta^* - \phi_1 y_{t-1}^*)^2\right). \qquad (11.91)$$

Логарифм условной функции максимального правдоподобия равен

$$l^* = \ln L^* = \text{const} - \frac{n-1}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{t=2}^{n}(y_t^* - \delta^* - \phi_1 y_{t-1}^*)^2. \qquad (11.92)$$

Из вида функции l^* видно, что оценка коэффициентов δ, ϕ_1 по условному методу максимального правдоподобия совпадает с оценкой нелинейного метода наименьших квадратов. (Заметим, что сумма в правой части (11.92) является нелинейной функцией параметров δ, ϕ_1.)

Полный метод максимального правдоподобия (*full ML*) состоит в максимизации функции правдоподобия

$$L = p(y_1)L^*.$$

Известно (см.(11.39),(11.40)), что при гипотезе нормальности ошибок $y_1^* \sim N(\delta^*/(1-\phi_1), \sigma^2/(1-\phi_1^2))$. Поэтому логарифм функции правдоподобия равен

$$l(\delta, \phi_1, \theta_1, \sigma^2) = \ln p(y_1) + l^*$$
$$= \text{const} - \frac{n}{2}\ln\sigma^2 + \frac{1}{2}\ln(1-\phi_1^2) - \frac{1-\phi_1^2}{2\sigma^2}\left(y_1^* - \frac{\delta^*}{1-\phi_1}\right)^2$$
$$- \frac{1}{2\sigma^2}\sum_{t=2}^{n}(y_t^* - \delta^* - \phi_1 y_{t-1}^*)^2. \qquad (11.93)$$

(Конечно, в (11.93) следует подставить выражения для δ^* и y_t^* через δ и θ_1.)

Проверка адекватности ARMA моделей

Есть несколько критериев оценки того, насколько ARMA модель, которую мы оцениваем, соответствует нашим данным.

Во-первых, оценки коэффициентов модели должны статистически достоверно отличаться от нуля, т. е. соответствующие P-значения t-статистик должны быть меньше выбранного порогового значения.

Во-вторых, согласно модели ошибки ε_t являются белым шумом. Соответственно их оценки, т. е. остатки регрессии e_t, должны быть также похожи на белый шум. Поэтому остатки должны иметь нулевую автокорреляцию.

В модели, включающей константу, среднее остатков равно 0. Поэтому выборочная автокорреляционная функция остатков вычисляется по формуле:

$$r_k = \frac{\sum_{t=k+1}^{n} e_t e_{t-k}}{\sum_{t=1}^{n} e_t^2}, \quad k = 1, 2, \ldots . \tag{11.94}$$

Если модель адекватна данным, ошибки являются белым шумом, и при больших значениях n и k величина r_k имеет распределение, близкое к нормальному $N(0, \frac{1}{n})$. Причем на практике хорошая аппроксимация начинается с $k = 5 \div 6$. Поэтому значение r_k вне интервала $0 \pm \frac{2}{\sqrt{n}}$ позволяет на 5%-ном уровне значимости отвергнуть гипотезу равенства нулю коэффициента корреляции ρ_k.

Другие тесты проверяют гипотезу равенства нулю сразу K первых значений автокорреляционной функции остатков.

Q-статистика Бокса-Пирса (Box, Pierce, 1970) определяется как

$$Q = n \sum_{k=1}^{K} r_k^2. \tag{11.95}$$

При нулевой гипотезе отсутствия автокорреляции Q имеет распределение $\chi^2(K - p - q)$, где p, q — параметры ARMA модели. Нулевая гипотеза отвергается, если полученное значение Q больше соответствующего критического значения.

Тест Льюнга-Бокса (Ljung, Box, 1978) является модификацией теста Бокса–Пирса. Соответствующая статистика

$$\widetilde{Q} = n(n+2) \sum_{k=1}^{K} \frac{r_k^2}{n-k} \tag{11.96}$$

имеет такое же асимптотическое распределение, как и Q, однако ее распределение ближе к χ^2 для конечных выборок.

Если тесты показывают наличие автокорреляции остатков, это означает, что рассматриваемая ARMA модель не подходит, и ее надо модифицировать. Например, если в автокорреляционной функции отличны от нуля значения с номерами, кратными 4, то стоит попробовать ввести сезонную авторегрессию четвертого порядка. Если единственное отличающееся от нуля значение соответствует лагу, равному 4, можно попробовать ввести сезонный MA-член порядка 4.

Если мы имеем ситуацию, когда несколько ARMA моделей оказываются адекватными данным, то, руководствуясь принципом «экономии мышления», следует выбрать модель с наименьшим количеством параметров.

В компьютерных пакетах среди результатов оценивания приводится *информационный критерий Акаике* AIC (*Akaike information criterion*) (Akaike, 1973), определяемый формулой

$$AIC = 2\frac{p+q}{n} + \ln\left(\frac{\sum_{t=1}^{n} e_t^2}{n}\right). \tag{11.97}$$

Критерий Акаике является эвристической попыткой свести в один показатель два требования: уменьшение числа параметров модели и качество подгонки модели. Согласно этому критерию, из двух моделей следует выбрать модель с меньшим значением AIC.

Обычно также приводится значение *критерия Шварца* (*Schwarz criterion*) (Schwarz, 1978)

$$\frac{(p+q)\ln n}{n} + \ln\left(\frac{\sum_{t=1}^{n} e_t^2}{n}\right), \tag{11.98}$$

отличие которого от AIC состоит в большем штрафе за количество параметров.

Заметим, что по своей идеологии критерии Акаике и Шварца близки к скорректированному R^2 (3.28).

Прогнозирование с ARIMA моделями

Главная цель использования ARIMA моделей — построение прогноза за пределы выборки. Есть два источника неточности прогноза: первый — игнорирование будущих ошибок ε_t, второй — отклонение оценок коэффициентов модели от их истинных значений. В данном разделе мы будем рассматривать только первый источник ошибок прогноза или, другими словами, прогнозирование в рамках теоретических моделей.

Рассмотрим проблему прогнозирования на примере ARMA(1,1) и ARIMA(1,1,0) моделей (несколько более простых примеров вынесено в упражнения).

ARMA(1,1) модель. Прогнозирование

Из (11.86) получаем значение y в момент $n+1$:

$$y_{n+1} = \delta + \phi_1 y_n + \varepsilon_{n+1} - \theta_1 \varepsilon_n.$$

Используя обозначение $\mu = \mathrm{E}(y_t) = \delta/(1-\phi_1)$, получаем:

$$(y_{n+1} - \mu) = \phi_1(y_n - \mu) + \varepsilon_{n+1} - \theta_1\varepsilon_n. \tag{11.99}$$

Прогноз на один шаг, минимизирующий среднеквадратичное отклонение, равен $\widehat{y}_{n+1} = \mathrm{E}(y_{n+1}|I_n)$ (см. МС, п.2), где I_n — информация, доступная в момент n. Из (11.99) получаем:

$$(\widehat{y}_{n+1} - \mu) = \phi_1(y_n - \mu) - \theta_1\varepsilon_n. \tag{11.100}$$

Ошибка прогноза и ее дисперсия равны

$$e_{n+1} = y_{n+1} - \widehat{y}_{n+1} = \varepsilon_{n+1}, \qquad \mathrm{V}(e_{n+1}) = \sigma^2. \tag{11.101}$$

Используя две итерации уравнения (11.99), получаем

$$(y_{n+2} - \mu) = \phi_1^2(y_n - \mu) + \varepsilon_{n+2} + (\phi_1 - \theta_1)\varepsilon_{n+1} - \phi_1\theta_1\varepsilon_n. \tag{11.102}$$

Отсюда аналогично (11.100) вычисляется прогноз на два шага:

$$\begin{aligned} (\widehat{y}_{n+2} - \mu) &= \phi_1^2(y_n - \mu) - \phi_1\theta_1\varepsilon_n, \\ \mathrm{V}(e_{n+2}) &= \sigma^2\left(1 + (\phi_1 - \theta_1)^2\right). \end{aligned} \tag{11.103}$$

Продолжая итерации, можно получить

$$(\widehat{y}_{n+s} - \mu) = \phi_1^s(y_n - \mu) - \phi_1^{s-1}\theta_1\varepsilon_n, \tag{11.104}$$

откуда видно, что прогноз стремится к среднему μ, когда горизонт прогноза возрастает. Можно показать, что

$$\lim_{s\to\infty} \mathrm{V}(e_{n+s}) = \sigma^2\left(\frac{1 - 2\phi_1\theta_1 + \theta_1^2}{1 - \phi_1^2}\right). \tag{11.105}$$

Заметим, что это выражение совпадает с дисперсией ряда y, полученной в (11.87б).

ARIMA(1,1,0) модель. Прогнозирование

Прогноз нестационарного временного ряда несколько отличается от выше разобранного случая. Рассмотрим временной ряд y_t, первые разности которого z_t являются AR(1) процессом (см. (11.67)):

$$z_t = y_t - y_{t-1}, \qquad z_t - \mu = \phi_1(z_{t-1} - \mu) + \varepsilon_t. \tag{11.106}$$

Многократное применение (11.106) дает

$$\begin{aligned} y_{n+s} &= y_n + z_{n+1} + z_{n+2} + \cdots + z_{n+s} \\ &= (y_n + s\mu) + (z_{n+1} - \mu) + \cdots + (z_{n+s} - \mu). \end{aligned} \tag{11.107}$$

Подставляя $z_t - \mu$ из (11.106) в (11.107), получаем:

$$y_{n+s} = y_n + s\mu + \frac{\phi_1(1-\phi_1^s)}{1-\phi_1}(y_n - y_{n-1} - \mu) + e_{n+s}, \tag{11.108}$$

где

$$\begin{aligned} e_{n+s} &= \varepsilon_{n+s} + (1+\phi_1)\varepsilon_{n+s-1} + \cdots \\ &\quad + (1 + \phi_1 + \phi_1^2 + \cdots + \phi_1^{s-1})\varepsilon_{n+1}. \end{aligned} \tag{11.109}$$

Очевидно, что прогноз, минимизирующий среднеквадратичное отклонение, равен сумме первых трех слагаемых в (11.108). Заметим, что второе и третье слагаемые растут с ростом s. Ошибка прогноза на s шагов равна e_{n+s}. В силу формулы (11.109) дисперсия ошибки равна

$$\begin{aligned} \mathrm{V}(e_{n+s}) &= \sigma^2\big(1 + (1+\phi_1)^2 + \cdots \\ &\quad + (1 + \phi_1 + \phi_1^2 + \cdots + \phi_1^{s-1})^2\big). \end{aligned} \tag{11.110}$$

Мы видим, что в случае нестационарного временного ряда дисперсия ошибки прогноза монотонно растет с ростом горизонта прогноза s.

Еще раз отметим, что все вычисления в этом разделе были проведены для теоретической модели, т. е. в предположении, что коэффициенты модели известны точно. Обычно на практике мы имеем дело с оценками коэффициентов, что добавляет дополнительную неопределенность в прогноз. Поэтому полученные оценки точности прогноза являются излишне «оптимистическими». Заметим, что некоторые компьютерные пакеты (например, EViews) корректно рассчитывают дисперсии ошибок прогноза, учитывая и неопределенность в коэффициентах.

Сезонность в ARIMA моделях

В этом разделе мы лишь кратко упомянем обобщение ARIMA моделей на случай наличия сезонной компоненты. Рассмотрим следующий пример. Пусть ряд y_t имеет сезонную (квартальную) компоненту. Тогда можно написать простейшую модель, связывающую значение переменной в текущем квартале с ее значением в том же квартале предыдущего года:

$$y_t = \phi_1^{(s)} y_{t-4} + u_t.$$

Так как временные ряды обладают статистической связью соседних значений, то можно предположить, что ошибки u_t удовлетворяют AR(1) процессу

$$u_t = \phi_1 u_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim iid(0, \sigma^2).$$

Из двух последних уравнений получаем:

$$(1 - \phi_1 L)(1 - \phi_1^{(s)} L^4) y_t = \varepsilon_t,$$

или

$$y_t = \phi_1 y_{t-1} + \phi_1^{(s)} y_{t-4} - \phi_1 \phi_1^{(s)} y_{t-5} + \varepsilon_t.$$

Такая модель обозначается AR(1)×SAR(1). Она похожа на AR(5) модель с тремя (нелинейными) ограничениями на коэффициенты.

Подробнее о свойствах подобных моделей можно прочитать, например, в (Johnston and DiNardo, 1997), (Box and Jenkins, 1976).

11.5. GARCH модели

В данном разделе мы дадим лишь краткое описание ARCH и GARCH моделей, ставших весьма популярными, особенно в литературе по финансовым рынкам, во второй половине 80-х и в 90-х годах.

Суть модели состоит в следующем. Предположим, мы имеем регрессию временного ряда y_t на другие временные ряды (все ряды предполагаются стационарными):

$$y_t = x_t' \beta + u_t. \tag{11.111}$$

Из эмпирических наблюдений за поведением таких рядов, как процентные ставки, обменные курсы и т. п., было замечено, что наблюдения с большими и малыми отклонениями от средних имеют тенденцию к образованию кластеров (см. рисунок 11.22). То есть периоды «спокойного» и «возмущенного» состояний рынка чередуются.

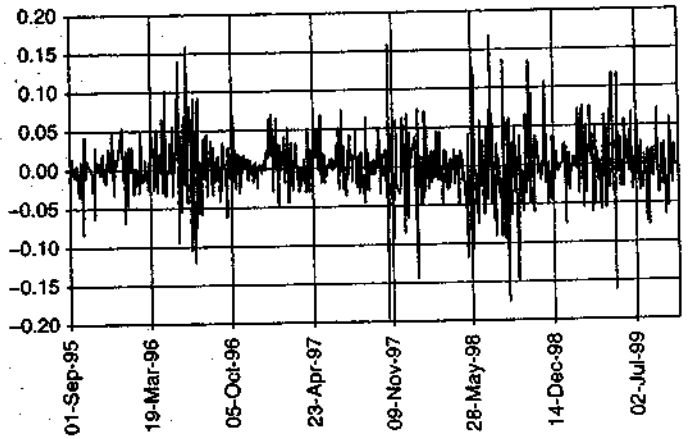


Рис. 11.22. Однодневные приращения индекса РТС

В работе (Engle, 1982) был предложен следующий способ моделирования этого явления. Пусть $\sigma_t^2 = \mathrm{V}(u_t | u_{t-1}, \ldots, u_{t-p}) = \mathrm{E}(u_t^2 | u_{t-1}, \ldots, u_{t-p})$ — условная дисперсия ошибок u_t (как обычно, $\mathrm{E}(u_t | u_{t-1}, \ldots, u_{t-p}) = 0$). Эффект «кластеризации» возмущений

можно объяснить следующей моделью зависимости условной дисперсии ошибок u_t от предыстории:

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \dots + \alpha_p u_{t-p}^2. \qquad (11.112)$$

Процесс (11.111)–(11.112) называется *авторегрессионной условно гетероскедастичной* моделью порядка p (*AutoRegressive, Conditional Heteroscedastic*), ARCH(p).

Простейшая модель такого рода, ARCH(1), имеет вид:

$$\begin{aligned} y_t &= x_t'\beta + u_t. \\ u_t &= \varepsilon_t \left(\alpha_0 + \alpha_1 u_{t-1}^2\right)^{1/2}, \quad \varepsilon_t \sim iidN(0,1). \end{aligned} \qquad (11.113)$$

В этой модели *условная* дисперсия ошибок зависит от времени: $\mathrm{V}(u_t|u_{t-1}) = \mathrm{E}(u_t^2|u_{t-1}) = \alpha_0 + \alpha_1 u_{t-1}^2$, в то время как *безусловная* дисперсия ошибок не зависит от времени: $\mathrm{V}(u_t) = \mathrm{V}(u_{t-1}) = \alpha_0/(1-\alpha_1)$. Таким образом, модель (11.113) удовлетворяет всем условиям классической линейной регрессионной модели и МНК-оценки являются наиболее эффективными *линейными* оценками.

Замечание. Существуют более эффективные *нелинейные* оценки, получающиеся из метода максимального правдоподобия (п. 10.5). Можно показать, что логарифм функции правдоподобия для (11.113) с точностью до константы равен

$$\ln L = -\frac{1}{2}\sum_{t=1}^{n} \ln\left(\alpha_0 + \alpha_1 u_{t-1}^2\right) - \frac{1}{2}\sum_{t=1}^{n} \frac{u_t^2}{\alpha_0 + \alpha_1 u_{t-1}^2}, \qquad (11.114)$$

где $u_t = y_t - \beta' x_t$.

Еще раз отметим, что ошибки u_t в ARCH(p) модели (11.111)–(11.112) являются стационарным процессом.

Как же определить, являются ли ошибки в уравнении (11.111) условно гетероскедастичными? Естественная процедура тестирования состоит из трех шагов:

1. Применяем МНК к уравнению (11.111) и вычисляем остатки e_t.
2. Оцениваем по МНК регрессию $e_t^2 = \widehat{\alpha}_0 + \widehat{\alpha}_1 e_{t-1}^2 + \dots + \widehat{\alpha}_p e_{t-p}^2 +$ *остаток*.

3. Тестируем гипотезу H_0: $\alpha_1 = \ldots = \alpha_p = 0$. Для тестирования можно применить F-тест или тест множителей Лагранжа LM (п. 10.6).

В работе (Bollerslev, 1986) была предложена более общая спецификация модели для уравнения условной дисперсии ошибок (11.112):

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \dots + \alpha_p u_{t-p}^2 + \gamma_1 \sigma_{t-1}^2 + \dots + \gamma_q \sigma_{t-q}^2. \qquad (11.115)$$

Такая модель (11.111)–(11.115) называется *обобщенной авторегрессионной условно гетероскедастичной* порядка p, q (*Generalized Auto-Regressive, Conditional Heteroscedastic*), GARCH(p,q). В этой модели ряд u_t^2 удовлетворяет ARMA($\max(p,q), q$) модели (11.56).

На практике наиболее часто применяется GARCH(1,1) модель. Существуют различные варианты и обобщения GARCH моделей, например, ARCH-M, EGARCH. Подробнее о моделях типа GARCH можно прочитать, например, в (Greene, 1997), (Hamilton, 1994).

Пример. ***GARCH модель.*** Рассмотрим связь рынка государственных облигаций (ГКО) и рынка корпоративных ценных бумаг (Российская торговая система, РТС) (Peresetsky, Ivanter, 2000). $BLCP_t$ и GKO_t — ежедневные значения индексов «голубых фишек» (наиболее ликвидных акций) и государственных облигаций. Обозначим через X_t разность однодневных доходностей двух рынков:

$$X_t = \ln\frac{BLCP_t}{BLCP_{t-1}} - \ln\frac{GKO_t}{GKO_{t-1}}.$$

Рассмотрим модель[1] выравнивания доходностей двух рынков:

$$\Delta X_t = \mathrm{const} + \mu X_{t-1} + \varepsilon_t.$$

Здесь $\Delta X_t = X_t - X_{t-1}$. Параметр μ имеет смысл скорости выравнивания доходностей и показывает степень интеграции рынков. Оценивание GARCH(1,1) модели на интервале 10.01.96–10.10.97 дает следующие результаты (использовалась программа EViews)

[1] Аналогичная модель рассматривалась в (Peresetsky, Turmuhambetova and Urga, 2001) для анализа рынка фьючерсов на ГКО.

Variable	*Coefficient*	*Std.Error*	*t-Statistic*	*Prob*
C	0.000735	0.001554	0.4728	0.6368
X_{t-1}	−0.768658	0.067768	−11.34244	0.0000
Variance Equation				
C	$5.14 \cdot 10^{-5}$	$2.14 \cdot 10^{-5}$	2.401	0.0171
$ARCH(1)$	0.1851	0.0792	2.336	0.0203
$GARCH(1)$	0.7296	0.0875	8.337	0.0000

R-squared	0.379	Mean Dependent var	−0.000341
Adjusted R-squared	0.369	S.D. dependent var	0.0288
S.E. of regression	0.0228	Akaike info criterion	−7.538
Sum squared resid	0.128	Schwarz criterion	−7.538
Log likelihood	603.5	F-statistic	37.57
Durbin–Watson stat	1.916	Prob(F-statistic)	0.0000

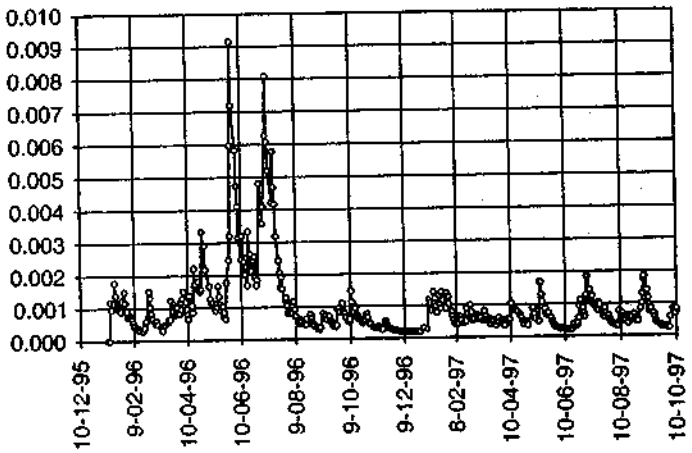


Рис. 11.23. График условного стандартного отклонения

Условное стандартное отклонение в этой модели можно интерпретировать как волатильность[2] рынка. На графике видны пики волатильности, соответствующие 30 мая 1996 года и 9 июля 1996 года, связанные с президентскими выборами. Видна стабилизация соотношения двух рынков после президентских выборов летом 1996 года.

[2]В литературе по финансовым рынкам волатильностью (*volatility*) называют меру нестабильности рынка.

Упражнения

11.1. Покажите, что выражение для суммарного влияния в модели (11.2) может быть получено как изменение (отклик) эндогенной переменной при единичном приращении экзогенной переменной в стационарном состоянии (т. е. когда $y_t = y_{t+1} = \cdots = \overline{y}$ и $x_t = x_{t+1} = \cdots = \overline{x}$).

11.2. Покажите, что суммарное влияние x на y в модели (11.3) равно $B(1)/A(1)$.

11.3. Покажите, что уравнение (11.3) устойчиво, если выполнено условие: все корни многочлена $A(x) = 1 - \alpha_1 x - \cdots - \alpha_p x^p$ лежат вне единичной окружности.

11.4. Выведите формулы, выражающие переменные $\widetilde{x}_{0t}, \ldots, \widetilde{x}_{rt}$ в уравнении (11.7), через переменные $x_t, x_{t-1}, \ldots, x_{t-q}$ из уравнения (11.5).

11.5. Выведите формулу для дисперсии (11.12).

11.6. Покажите, что для случайных величин, удовлетворяющих уравнению (11.11), при условии существовании момента $\mathrm{E}(\varepsilon_t^4)$ выполнено:

а) $p\lim (1/n) \sum y_{t-1}\varepsilon_t = 0$.

б) $p\lim (1/n) \sum y_{t-1}^2 = \sigma^2/(1-\beta^2)$.

11.7. Докажите формулы (11.16) и (11.17).

11.8. Докажите, что многочлен

$$A(L) = (1 - \phi_1 L - \phi_2 L^2) = (1 - \lambda_1 L)(1 - \lambda_2 L)$$

обратим, тогда и только тогда, когда $|\lambda_1| < 1$ и $|\lambda_2| < 1$.

11.9. Покажите, что тестирование наличия одного единичного корня в процессе AR(p) можно свести к тесту ADF, приведя уравнение к виду, аналогичному (11.52) для AR(2) модели.

11.10. Вычислите частную автокорреляционную функцию PACF(k) для AR(1) процесса.

11.11. Покажите, что автокорреляционная функция стационарного процесса AR(2) убывает экспоненциально в случае, когда характеристические корни $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2$ действительны, или изменяется по синусоиде с экспоненциально убывающей амплитудой, когда корни комплексные.

11.12. Покажите, что условия (11.76) стационарности процесса AR(2) эквивалентны тому, что оба корня характеристического уравнения $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 = 0$ лежат вне единичной окружности.

11.13. Вычислите частную автокорреляционную функцию PACF(k) для AR(2) процесса.

11.14. Примените процедуру вычисления выборочного частного коэффициента корреляции (см. п. 4.3) для стационарного ряда Y_t. Покажите, что k-е значение выборочной частной автокорреляционной функции PACF(k) вычисляется как МНК-оценка последнего коэффициента β_k в AR(k) регрессионном уравнении:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_k y_{t-k} + \varepsilon_t.$$

11.15. Покажите, что для MA(q) процесса ACF(k)=0 при $k > q$.

11.16. Сформулируйте условия обратимости MA(2) процесса.

11.17. Покажите, что для AR(1) процесса (11.67) в виде $y_t - \mu = \phi_1(y_{t-1} - \mu) + \varepsilon_t$ прогноз на s шагов вперед вычисляется по формуле $\widehat{y}_{n+s} = \mu + \phi_1^s(y_n - \mu)$, а дисперсия ошибки прогноза равна $\mathrm{V}(e_{n+s}) = \left(1 + \phi_1^2 + \phi_1^4 + \cdots + \phi_1^{2s-2}\right)\sigma^2$.

11.18. Покажите, что для MA(1) процесса (11.78) прогноз на s шагов вперед вычисляется по формулам $\widehat{y}_{n+1} = \delta - \theta_1\varepsilon_n$, $\widehat{y}_{n+s} = \delta$, $s \geqslant 2$, а дисперсия ошибки прогноза равна $\mathrm{V}(e_{n+s}) = (1 + \theta_1^2)\,\sigma^2 = \mathrm{V}(y_t)$.

11.19. Выведите формулы (11.104), (11.105).

11.20. Вычислите автокорреляционную функцию ACF(k) и частную автокорреляционную функцию PACF(k) для MA(2) процесса.

11.21. Покажите, что остатки при оценивании методом наименьших квадратов уравнений

$$y_t = \alpha y_{t-1} + \beta x_t + \varepsilon_t,$$
$$\Delta y_t = \gamma y_{t-1} + \beta x_t + \varepsilon_t,$$

совпадают. (Здесь $\Delta y_t = y_t - y_{t-1}$.)

11.22. Пусть $y_t = 1 + 0.4y_{t-1} + 0.3y_{t-2} + u_t$ — AR(2) процесс, где u_t — независимые $N(0,1)$ случайные величины. Вычислите прогнозные значения

$$\mathrm{E}(y_t \mid y_{t-1}), \quad \mathrm{E}(y_t \mid y_{t-1}, y_{t-2}), \quad \mathrm{E}(y_t \mid y_{t-1}, y_{t-2}, y_{t-3}).$$

11.23. Дана линейная модель

$$y_t = \beta_1 x_t + \beta_2 y_{t-1} + u_t, \quad u_t = \rho u_{t-1} + \varepsilon_t,$$

где $0 < \rho < 1$ и ε — гауссовский белый шум. Проводятся две регрессии: для исходных величин и их разностей, т. е.

$$y_t = \beta_1 x_t + \beta_2 y_{t-1} + u_t, \qquad (*)$$
$$\Delta y_t = \beta_1 \Delta x_t + \beta_2 \Delta y_{t-1} + v_t, \qquad (**)$$

где $v_t = (\rho - 1)u_{t-1} + \varepsilon_t$.

а) Покажите, что в обеих регрессиях (*) и (**) МНК-оценки вектора $\boldsymbol{\beta} = [\beta_1\ \beta_2]'$ будут смещенными и несостоятельными.

б) Покажите, что смещение в регрессии (*) не снижается до нуля, когда $\rho \to 1$.

в) Предложите оценку вектора $\boldsymbol{\beta}$ с помощью инструментальных переменных и покажите, что она состоятельна.

Глава 12

Дискретные зависимые переменные и цензурированные выборки

Ранее мы рассмотрели модели, в которых какие-либо *независимые* переменные принимают дискретные значения, например, 0 или 1, выражая некоторые качественные признаки (фиктивные переменные). Относительно *зависимой* переменной явно или неявно предполагалось, что она выражает количественный признак, принимая «непрерывное» множество значений. В частности, в нормальной линейной регрессионной модели (п. 2.3) предполагается, что ошибка имеет гауссовское распределение, откуда следует, что зависимая переменная y может принимать любые значения. В то же время довольно часто интересующая нас величина по своей природе является дискретной. Выделим несколько типичных ситуаций.

1. Выбор из двух или нескольких альтернатив. Примеры:
- голосование;
- решение работать или не работать;
- решение покупать или не покупать какой-либо товар длительного пользования (автомобиль, дом и т. п.);
- форма собственности (государственная, смешанная, частная);
- выбор профессии (научный работник, преподаватель, консультант, менеджер);
- способ попадания из дома на работу (пешком, автобус, метро, метро и автобус, автомобиль); и т. д.

Если есть только две возможности (бинарный выбор), то результат наблюдения обычно описывается переменной, принимающей значения 0 или 1, называемой *бинарной*. В общем случае при наличии k альтернатив результат выбора можно представить переменной, принимающей, например, значения $1, \ldots, k$. Если альтернативы нельзя естественным образом упорядочить (как в двух последних примерах), то их нумерация может быть произвольной. В этих случаях соответствующую переменную называют *номинальной* (*qualitative*).

2. Ранжированный выбор. Как и в первом случае, есть несколько альтернатив, но они некоторым образом упорядочены. Примеры:
- доход семьи (низкий, средний, высокий, очень высокий);
- уровень образования (незаконченное среднее, среднее, среднее техническое, высшее);
- состояние больного (плохое, удовлетворительное, хорошее); и т. д.

Соответствующая переменная называется *порядковой, ординальной* или *ранговой* (*ranking*).

3. Количественная целочисленная характеристика. Примеры:
- количество прибыльных предприятий;
- количество частных университетов;
- число патентов, зарегистрированных в течение года; и т. д.

Для моделей с дискретными зависимыми переменными конечно же возможно формальное применение метода наименьших квадратов, однако достаточно удовлетворительные с содержательной точки зрения результаты можно при этом получить,

как правило, лишь для моделей третьей группы с количественными целочисленными переменными. В случае порядковых переменных интерпретация оценок коэффициентов при объясняющих переменных значительно затруднена: увеличение на единицу порядковой переменной означает переход к следующей по рангу альтернативе, однако далеко не всегда переход от первой альтернативы ко второй численно эквивалентен переходу от второй к третьей. Если же зависимая переменная является номинальной и количество альтернатив больше двух, то результаты оценивания вообще теряют смысл в силу произвольности нумерации альтернатив. Таким образом, стандартная регрессионная схема, которую мы использовали ранее для анализа зависимости интересующей нас переменной от экзогенных факторов, в случае номинальных эндогенных переменных нуждается в существенной коррекции.

Сначала мы рассмотрим модели бинарного выбора, затем будет показано, что модели с несколькими альтернативами могут быть либо непосредственно сведены к моделям бинарного выбора, либо могут быть исследованы аналогичными методами.

Другой класс моделей, рассматриваемых в данной главе, связан с *цензурированными* (*censored*) и *урезанными* (*truncated*) выборками. Классический пример цензурирования дает изучение расходов семей на покупку товаров длительного пользования (автомобиля, дома и т. п.). Ясно, что эти расходы не могут быть отрицательными и в то же время при проведении обследования будут встречаться наблюдения с нулевым значением этих расходов, что просто означает отказ от покупки соответствующего товара. Здесь осуществляется цензурирование выборки на уровне 0 значения зависимой переменной. Другой пример дает определение «времени жизни» технического изделия с помощью испытания в одинаковых условиях в течение определенного периода нескольких экземпляров изделия. Для тех образцов, которые в процессе испытаний вышли из строя, время жизни будет зафиксировано точно, для остальных временем жизни будет считаться длительность испытаний, а истинное его значение останется неизвестным. В этом случае уровнем цензурирования является период испытаний. Можно показать, что в подобных ситуациях непосредственное применение метода наименьших квадратов дает смещенные оценки параметров.

Пример урезания выборки дает исследование распределения семей по объему выплачиваемых налогов или изучение зависимости выплачиваемых налогов от размера семьи, возраста ее членов и т. п. Здесь из рассмотрения могут исключаться семьи, имеющие доход ниже официального уровня бедности. В этом случае уровень бедности определяет урезание выборки. Метод наименьших квадратов здесь также приводит к смещенным оценкам.

Подчеркнем разницу между цензурированием и урезанием. В первой ситуации даже для цензурированного наблюдения известны значения независимых переменных, в то время как во втором случае известен лишь уровень урезания, а значения независимых переменных для исключенных из рассмотрения объектов неизвестны.

12.1. Модели бинарного и множественного выбора

Для наглядности будем изучать модели бинарного выбора на примере покупки семьей автомобиля. Обозначая, как и раньше, зависимую переменную y, будем считать, что $y = 1$, если в течение исследуемого периода времени семья купила автомобиль, и $y = 0$ в противном случае. Ясно, что на решение о покупке автомобиля влияют самые различные факторы: доход семьи, количество ее членов, их возраст, место проживания семьи и т. п. Набор этих характеристик можно представить вектором $x = (x_1, \ldots, x_k)'$ (независимые переменные). Сохраняя основные идеи регрессионного подхода, будем предполагать, что на решение семьи влияют также неучтенные случайные факторы (ошибки). Выдвигая различные предположения о характере зависимости y от x, будем получать разные модели. Здесь мы рассмотрим три модели: линейную модель вероятности и так называемые *probit*- и *logit*-модели.

Линейная модель вероятности

Воспользуемся обычной линейной моделью регрессии:

$$y_t = x_t'\beta + \varepsilon_t, \qquad t = 1, \ldots, n, \tag{12.1}$$

где t — номер наблюдения (семьи), $\beta = (\beta_1, \ldots, \beta_k)'$ — набор неизвестных параметров (коэффициентов), ε_t — случайная ошибка. Так как y_t принимает значения 0 или 1 и $\mathrm{E}(\varepsilon_t) = 0$, то

$$\mathrm{E}(y_t) = 1 \cdot \mathrm{P}(y_t = 1) + 0 \cdot \mathrm{P}(y_t = 0) = \mathrm{P}(y_t = 1) = x_t'\beta.$$

Таким образом, модель (12.1) может быть записана в виде

$$\mathrm{P}(y_t = 1) = x_t'\beta, \tag{12.2}$$

поэтому ее называют *линейной моделью вероятности* (*linear probability model*).

Отметим некоторые особенности этой модели, наличие которых не позволяет успешно применять метод наименьших квадратов для оценивания коэффициентов β и прогнозирования.

Из соотношения (12.1) следует, что ошибка ε в каждом наблюдении может принимать только два значения: $\varepsilon_t = 1 - x_t'\beta$ с вероятностью $\mathrm{P}(y_t = 1)$ и $\varepsilon_t = -x_t'\beta$ с вероятностью $1 - \mathrm{P}(y_t = 1)$. Это, в частности, не позволяет считать ошибку нормально распределенной или имеющей распределение, близкое к нормальному. Далее, непосредственным вычислением получаем, что дисперсия ошибки $\mathrm{V}(\varepsilon_t) = x_t'\beta(1 - x_t'\beta)$ зависит от x_t, т.е. модель (12.1) гетероскедастична (п. 6.1). Как известно, оценки коэффициентов β, полученные обычным методом наименьших квадратов, в этом случае не являются эффективными, и желательно пользоваться доступным обобщенным методом наименьших квадратов (п. 5.3).

Самым серьезным недостатком линейной модели вероятности является тот факт, что прогнозные значения $\widehat{y}_t = x_t'\widehat{\beta}$, которые по смыслу модели есть прогнозные значения вероятности $\mathrm{P}(y_t = 1)$, могут лежать вне отрезка $[0, 1]$ ($\widehat{\beta}$ — оценка коэффициентов β, полученная с помощью обычного или обобщенного метода наименьших квадратов), что, конечно же, не поддается разумной интерпретации. Это обстоятельство существенно ограничивает область

применимости линейной модели вероятности. Ее целесообразно использовать при большом числе наблюдений и при достаточно точной спецификации модели, а также как инструмент первичной обработки данных для сравнения с результатами, получаемыми более тонкими методами.

Probit- и *logit*-модели

Описание моделей

Основной недостаток линейной модели вероятности есть следствие предположения о линейной зависимости вероятности $\mathrm{P}(y_t = 1)$ от β (см. (12.2)). Его можно преодолеть, если считать, что

$$\mathrm{P}(y_t = 1) = F(x_t'\beta), \tag{12.3}$$

где $F(\cdot)$ — некоторая функция, область значений которой лежит в отрезке $[0, 1]$. В частности, в качестве $F(\cdot)$ можно взять функцию распределения некоторой случайной величины. Одна из возможных интерпретаций модели (12.3) выглядит следующим образом. Предположим, что существует некоторая количественная переменная y_t^*, связанная с независимыми переменными x_t обычным регрессионным уравнением

$$y_t^* = x_t'\beta + \varepsilon_t, \tag{12.4}$$

где ошибки ε_t независимы и одинаково распределены с нулевым средним и дисперсией σ^2. Пусть также $F(\cdot)$ — функция распределения нормированной случайной ошибки ε_t/σ. Величина y_t^* является ненаблюдаемой (латентной), а решение, соответствующее значению $y_t = 1$, принимается тогда, когда y_t^* превосходит некоторое пороговое значение. Так, в примере с покупкой автомобиля можно считать, что y_t^* представляет накопления семьи с номером t. Без ограничения общности, если константа включена в число регрессоров, можно считать это пороговое значение равным нулю. Величину y_t^* можно также интерпретировать как разность полезностей альтернативы 1 и альтернативы 0.

Таким образом,

$$\begin{aligned} y_t &= 1, \quad \text{если } y_t^* \geqslant 0, \\ y_t &= 0, \quad \text{если } y_t^* < 0. \end{aligned} \tag{12.5}$$

Тогда, предполагая, что случайные ошибки ε_t имеют одно и то же симметричное распределение $F(\cdot)$ (т. е. $F(-x) = 1 - F(x)$), получаем:

$$\begin{aligned} \mathrm{P}(y_t = 1) &= \mathrm{P}(y_t^* \geqslant 0) = \mathrm{P}(\boldsymbol{x}_t'\boldsymbol{\beta} + \varepsilon_t \geqslant 0) \\ &= \mathrm{P}(\varepsilon_t \geqslant -\boldsymbol{x}_t'\boldsymbol{\beta}) = \mathrm{P}(\varepsilon_t \leqslant \boldsymbol{x}_t'\boldsymbol{\beta}) = F\left(\frac{\boldsymbol{x}_t'\boldsymbol{\beta}}{\sigma}\right), \end{aligned} \tag{12.6}$$

что с точностью до нормировки совпадает с (12.3).

Замечание. В модели (12.4)–(12.6) параметры $\boldsymbol{\beta}$ и σ участвуют только в виде отношения и не могут быть по отдельности идентифицированы (т. е. оценить можно лишь $\boldsymbol{\beta}/\sigma$). Поэтому в данном случае без ограничения общности можно считать, что $\sigma = 1$.

Наиболее часто в качестве функции $F(\cdot)$ используют:

— функцию стандартного нормального распределения:

$$F(u) = \Phi(u) = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{u} e^{-\frac{z^2}{2}}\, dz$$

и соответствующую модель называют *probit-моделью*;

— функцию логистического распределения:

$$F(u) = \Lambda(u) = \frac{e^u}{1 + e^u}$$

и соответствующую модель называют *logit-моделью*.

В свете рассмотренной выше интерпретации модели (12.3) использование функции нормального распределения представляется достаточно естественным. Применение функции логистического распределения во многом объясняется простотой численной реализации процедуры оценивания параметров. Вопрос о том, какую из моделей (*probit* или *logit*) следует использовать в том или ином случае, является достаточно сложным. Можно, например, выбрать ту модель, для которой больше значение соответствующей функции правдоподобия. Отметим также, что для

значений u, достаточно близких по модулю к нулю (напри при $u \in [-1.2, 1.2]$), функции $\Phi(u)$ и $\Lambda(u)$ ведут себя прим одинаково, в то же время «хвосты» логистического распре ния значительно «тяжелее» «хвостов» нормального распре ния. Практический опыт показывает, что для выборок с не шим разбросом объясняющих переменных и при отсутстви щественного преобладания одной альтернативы над другой ственные выводы, получаемые с помощью *probit*- и *logit*-мо будут, как правило, совпадать.

Поскольку модель (12.3) нелинейна по параметрам $\boldsymbol{\beta}$, интерпретация отличается от привычной интерпретации коэ циентов линейных регрессионных моделей. Предположим, функции распределения $F(\cdot)$ есть плотность $p(\cdot)$. Дифферен руя по векторному аргументу $\boldsymbol{x}$ (приложение ЛА, п. 19) и опу нижний индекс t (номер наблюдения), получаем:

$$\frac{\partial \mathrm{P}(y = 1)}{\partial \boldsymbol{x}} = F'(\boldsymbol{x}'\boldsymbol{\beta})\,\boldsymbol{\beta} = p(\boldsymbol{x}'\boldsymbol{\beta})\,\boldsymbol{\beta}. \quad ($$

Таким образом, предельный эффект каждого объясняющего тора x_j, $j = 1, \ldots, k$ является переменным и зависит от знач всех остальных факторов $\boldsymbol{x} = (x_1, \ldots, x_k)'$.

При использовании этой модели для получения представл о «среднем» предельном эффекте рекомендуется вычислять изводные (12.7) для средних по выборке значений независ переменных $\boldsymbol{x}$.

Оценивание модели

Для оценивания параметров $\boldsymbol{\beta}$ модели (12.3) обычно исполь метод максимального правдоподобия (глава 10). Предполо что наблюдения $y_1, \ldots, y_n$ независимы. Поскольку y_t может нимать значения только 0 или 1, то функция правдоподобия следующий вид:

$$L = L(y_1, \ldots, y_n) = \prod_{y_t=0} (1 - F(\boldsymbol{x}_t'\boldsymbol{\beta})) \prod_{y_t=1} F(\boldsymbol{x}_t'\boldsymbol{\beta}). \quad ($$

Отсюда легко вытекает, что

$$L = \prod_t [F(x_t'\beta)]^{y_t}[1 - F(x_t'\beta)]^{1-y_t}.$$

Логарифмируя, получаем:

$$l = \ln L = \sum_t [y_t \ln F(x_t'\beta) + (1 - y_t)\ln(1 - F(x_t'\beta))]. \qquad (12.9)$$

Дифференцируя равенство (12.9) по вектору β (приложение ЛА, п. 19), получаем векторное уравнение правдоподобия

$$\frac{\partial l}{\partial \beta} = \sum_t \left(\frac{y_t p(x_t'\beta)}{F(x_t'\beta)} - \frac{(1 - y_t)p(x_t'\beta)}{1 - F(x_t'\beta)} \right) x_t = 0. \qquad (12.10)$$

Для *logit*-модели оно существенно упрощается. Действительно, пользуясь легко проверяемым тождеством $\Lambda'(u) = \Lambda(u)(1 - \Lambda(u))$, имеем

$$\sum_t (y_t - \Lambda(x_t'\beta))x_t = 0.$$

Уравнение правдоподобия (12.10) является системой нелинейных (относительно β) уравнений, в общем случае нельзя найти ее аналитическое решение и приходится прибегать к численным методам. Отметим также, что уравнение правдоподобия есть лишь необходимое условие локального экстремума. Можно показать (см., например, (Greene, 1997)), что для *probit*- и *logit*-моделей логарифмическая функция правдоподобия l является вогнутой по β функцией и, значит, решение уравнения (12.10) дает оценку максимального правдоподобия набора параметров β. Процедуры оценивания *probit*- и *logit*-моделей реализованы в большинстве современных эконометрических компьютерных пакетов.

Пример. ***Факторы некредитоспособности российских банков.*** Этот пример основан на результатах дипломной работы выпускницы РЭШ 1999 г. Е. Е. Баян-оол.

Как те или иные характеристики банка влияют на его жизнеспособность? Для исследования этой проблемы была рассмотрена

logit-модель с бинарной переменной, принимающей значения 1 или 0, в зависимости от того, находится ли банк в критическом состоянии или нет. Решение о том, является ли банк проблемным или нет, принималось на основании рейтинга банков, опубликованного в журнале «Профиль» от 21 июня 1999 г. Значение 1 приписывалось банкам с отрицательным капиталом; банкам, имеющим 4-ю группу проблемности; банкам, у которых отозвана лицензия или принято решение об отзыве. Остальным банкам присвоено значение 0. Из многочисленных характеристик банков (возраст, капитал, ликвидные активы, работающие рисковые активы, обязательства до востребования, суммарные обязательства, уставный фонд, чистые активы и т. п.) в окончательную модель после анализа и многочисленных попыток были включены следующие переменные:

TOTLIAB — суммарные обязательства (тыс. руб.);
CURRENCY — валютная составляющая (%);
EQUITY/ASS, где EQUITY — недвижимость (тыс. руб.), ASS — чистые активы (тыс. руб.);
PROFIT/ASS, где PROFIT — прибыль (убыток) (тыс. руб.);
RETAIL/TOTLIAB, где RETAIL — средства частных лиц (тыс. руб.);
TOTLIAB/PREF, где PREF — работающие рисковые активы (тыс. руб.).

Модель включает 182 наблюдения. Результаты оценивания с помощью *logit*-модели приведены в таблице 12.1. Полученные результаты согласуются с экономической интуицией. В частности, в результате кризиса наименее устойчивыми оказались крупные и мельчайшие банки. Поскольку в выборке представлены средние и крупные банки, то положительность коэффициентов при переменных, характеризующих величину банка, согласуется с реальностью.

Таблица 12.1

Переменная	Коэфф.	Станд. откл.	t-статист.	P-знач.
C	-3.19	0.87	-3.68	0.000
TOTLIAB	2.14E-07	7.21E-08	2.97	0.003
CURRENCY	-0.096	0.031	-3.13	0.021
EQUITY/ASS	-8.58	3.71	-2.31	0.022
PROFIT/ASS	-26.99	9.49	-2.84	0.005
RETAIL/TOTLIAB	3.53	1.23	2.87	0.005
TOTLIAB/PREF	2.00	0.90	2.23	0.027

Проверка гипотез

Для *probit*- или *logit*-моделей проверка гипотез о наличии ограничений на коэффициенты, в частности, гипотез о значимости одного или группы коэффициентов, может проводиться с помощью любого из трех тестов — Вальда, отношения правдоподобия, множителей Лагранжа, рассмотренных в главе 10 (п. 10.6). Большинство эконометрических пакетов, в которых реализованы *probit*- или *logit*-модели, имеют встроенные процедуры проверки ограничений с указанием метода тестирования.

Ошибки спецификации

Кратко рассмотрим проблемы, возникающие при нарушении некоторых предположений, лежащих в основе модели (12.3). Подробное изложение этого материала требует привлечения довольно сложных методов и выходит за рамки данной книги. Более детально эти вопросы изложены, например, в (Greene, 1997) или (Johnston and DiNardo, 1997).

Рассматривая реализацию (12.4), (12.5) модели (12.3) с помощью ненаблюдаемой переменной y^*, мы предполагали, что ошибки ε_t одинаково распределены, в частности, гомоскедастичны. Известно (п. 6.1), что при нарушении этого условия, т. е. при наличии гетероскедастичности, оценки метода наименьших квадратов в линейных регрессионных моделях перестают быть эффективными, но остаются несмещенными и состоятельными. В нашем случае гетероскедастичность, вообще говоря, приводит к нарушению состоятельности и асимптотической несмещенности. На содержательном уровне это нетрудно понять, исходя из следующих соображений. Пусть ошибки ε_t, $t = 1, \ldots, n$ распределены нормально с нулевым средним и дисперсиями σ_t, $t = 1, \ldots, n$ (гетероскедастичность) и предположим, что выполнено (12.5). Тогда, повторяя выкладки (12.6), получим:

$$\mathrm{P}(y_t = 1) = \Phi\left(\frac{x_t'\beta}{\sigma_t}\right),$$

где $\Phi(\cdot)$ — функция стандартного нормального распределения.

Соответствующим образом изменится логарифмическая функция правдоподобия (12.9):

$$l = \ln L = \sum_t \left[y_t \ln \Phi\left(\frac{x_t'\beta}{\sigma_t}\right) + (1 - y_t) \ln\left(1 - \Phi\left(\frac{x_t'\beta}{\sigma_t}\right)\right)\right].$$

Это означает, что теперь необходимо оценивать $n + k - 1$ неизвестных параметров (без ограничения общности одну из дисперсий можно считать равной 1), что без дополнительных предположений невозможно сделать состоятельно на основе n наблюдений.

Аналогично тому, как это делается в тесте Бреуша–Пагана (глава 6), можно предполагать ту или иную форму зависимости дисперсий от экзогенных факторов и тестировать гипотезы об отсутствии гетероскедастичности (подробнее см. (Greene, 1997)).

В п. 4.4 мы рассмотрели проблемы исключения существенных и включения несущественных переменных для линейных регрессионных моделей. Можно поставить аналогичный вопрос: какое влияние оказывает пропуск существенных переменных в уравнении (12.4) на оценивание модели бинарного выбора (12.3)? Исчерпывающий ответ на него выходит за рамки нашей книги. Отметим лишь, что в данном случае, даже если исключенные существенные переменные ортогональны включенным, оценки параметров будут, в отличие от линейной схемы, смещенными и несостоятельными (подробнее см. (Greene, 1997) и (Johnston and DiNardo, 1997)).

Модели множественного выбора

Модели множественного выбора, когда имеется не две, а несколько альтернатив, можно строить и изучать, обобщая подходы и методы, используемые для моделей бинарного выбора.

Номинальные зависимые переменные

Если соответствующая переменная является номинальной (качественной), то множественный выбор может быть представлен как последовательность бинарных выборов. Поясним это простым примером. Предположим, что изучается выбор одной из трех

профессий: инженер, научный работник, преподаватель. Введем три бинарных переменных, соответствующих каждой профессии: $y^i = 1$ для инженеров, $y^i = 0$ для всех остальных; $y^s = 1$ для научных работников, $y^s = 0$ для всех остальных; $y^t = 1$ для преподавателей, $y^t = 0$ для всех остальных.

Тогда выбор одной из трех альтернатив можно описать в виде «дерева» последовательных решений, в узлах которого происходит бинарный выбор.

В каждом узле, применяя технику оценивания для бинарных моделей, можно оценить условную вероятность выбора соответствующей альтернативы. Безусловная вероятность вычисляется по формуле умножения вероятностей. Так, например,

$$\mathrm{P}(y^t = 1) = \mathrm{P}(y^i = 0,\, y^s = 0) = \mathrm{P}(y^i = 0)\mathrm{P}(y^s = 0 \mid y^i = 0).$$

В последнем произведении первый сомножитель оценивается в первом узле (стрелка вниз), второй — во втором (стрелка вниз). Обобщение этого метода на случай любого числа альтернатив не представляет труда.

Однако у данного способа построения моделей множественного выбора есть очевидный недостаток: «дерево» последовательных решений можно строить по-разному, и результаты оценивания будут, вообще говоря, разными.

Другой подход к моделям множественного выбора с качественной зависимой переменной основан на понятии случайной полезности (как уже отмечалось выше, в *probit*- или *logit*-моделях скрытую переменную можно интерпретировать как разность полезностей альтернатив 1 и 0).

Итак, предположим, что имеется m альтернатив. Будем считать, что для индивидуума t альтернатива j имеет полезность $U_{tj} = u_{tj} + \varepsilon_{tj}$, где u_{tj} — неслучайная составляющая, а ε_{tj} — случайная составляющая полезности. Тогда индивидуум t выберет альтернативу j, если $U_{tj} > U_{tk}$ для любого $k \neq j$. Иными словами,

$$\mathrm{P}(y_t = j) = \mathrm{P}(u_{tj} + \varepsilon_{tj} > u_{tk} + \varepsilon_{tk}\ \forall\, k \neq j,\ k = 1, \dots m). \quad (12.11)$$

В общем случае для нахождения этой вероятности требуется вычислять многомерные интегралы по соответствующим областям от плотности совместного распределения ошибок ε_{tj}. Как правило (в частности, для нормально распределенных ошибок ε_{tj}), эти интегралы невозможно выразить аналитически, а можно лишь найти численно, что, в конечном итоге, делает модель не применимой на практике. Есть, однако, некоторое специальное распределение, для которого вероятность $\mathrm{P}(y_t = j)$ в (12.11) допускает достаточно простое представление. Предположим, что ошибки ε_{tj} независимы и имеют функцию распределения $F(x) = \exp(-e^{-x})$ (такое распределение возникает при изучении максимума независимых случайных величин, поэтому его часто называют распределением экстремальных значений). Тогда можно доказать, что

$$\mathrm{P}(y_t = j) = \frac{\exp(u_{tj})}{\exp(u_{t1}) + \dots + \exp(u_{tm})}. \quad (12.12)$$

Предполагая, что полезность u_{tj} зависит от наблюдаемых экзогенных характеристик $\boldsymbol{x}_{tj}$ и неизвестных параметров $\boldsymbol{\beta}$:

$$u_{tj} = \boldsymbol{x}'_{tj}\boldsymbol{\beta},$$

получаем модель

$$\mathrm{P}(y_t = j) = \frac{\exp(\boldsymbol{x}'_{tj}\boldsymbol{\beta})}{\exp(\boldsymbol{x}'_{t1}\boldsymbol{\beta}) + \dots + \exp(\boldsymbol{x}'_{tm}\boldsymbol{\beta})}, \quad (12.13)$$

которая называется *logit*-моделью множественного выбора (*multinomial logit model*)[1].

Среди экзогенных переменных x_{tj} могут быть характеристики, зависящие только от индивидуума и не зависящие от альтернативы. Если, например, анализируется проблема выбора профессии, то естественно включить в x_{tj} такие факторы, как возраст, уровень образования, социальный статус и т. п., которые не зависят от профессии. Выделим такие переменные: $x'_{tj} = [y'_{tj}, z'_t]$, и соответствующим образом разобьем вектор неизвестных параметров на две компоненты: $\beta' = [\gamma', \delta']$. Тогда числитель и знаменатель правой части формулы (12.13) будут содержать общий сомножитель $\exp(z'_t\delta)$, а это означает, что вектор параметров δ оценить невозможно (неидентифицируемость). Следовательно, если необходимо учесть индивидуальные эффекты, *logit*-модель множественного выбора должна быть модифицирована. Например, можно считать, что коэффициенты δ могут зависеть от альтернативы, т. е. $u_{tj} = y'_{tj}\gamma + z'_t\delta_j$. В примере с выбором профессии подобное предположение выглядит реалистичным: при одном и том же уровне образования полезность разных профессий разная (при прочих равных).

Часто в литературе рассматривается модель, когда

$$u_{tj} = x'_t\beta_j,$$

т. е. когда экзогенные переменные не зависят от альтернативы, а коэффициенты могут от нее зависеть. В этом случае

$$\mathrm{P}(y_t = j) = \frac{\exp(x'_t\beta_j)}{\exp(x'_t\beta_1) + \cdots + \exp(x'_t\beta_m)}, \tag{12.14}$$

и эту модель также называют *logit*-моделью множественного выбора. Заметим, что модель (12.14) неидентифицируема, поскольку правая часть формулы (12.14) зависит только от разностей

[1]Иногда (см., например, (Greene, 1997)) эту модель называют условной *logit*-моделью (*conditional logit model*).

$\beta_2 - \beta_1, \ldots, \beta_m - \beta_1$. Поэтому для идентифицируемости модели (12.14) обычно используют нормировку $\beta_1 = 0$:

$$\begin{aligned} \mathrm{P}(y_t = 1) &= \frac{1}{1 + \exp(x'_t\beta_2) + \cdots + \exp(x'_t\beta_m)}, \\ \mathrm{P}(y_t = j) &= \frac{\exp(x'_t\beta_j)}{1 + \exp(x'_t\beta_2) + \cdots + \exp(x'_t\beta_m)}, \quad j \geqslant 2. \end{aligned} \tag{12.15}$$

Нетрудно видеть, что при $m = 2$ модель (12.15) — это обычная *logit*-модель бинарного выбора (12.3). Модель (12.13) при $m = 2$ тоже сводится к обычной *logit*-модели, если в качестве независимых переменных рассматривать $x_{t2} - x_{t1}$.

Существенным ограничением, лежащим в основе *logit*-модели множественного выбора, является предположение о статистической независимости полезностей u_{tj} по j. Оно выглядит нереалистичным, если среди альтернатив есть достаточно близкие. Классический пример такой ситуации, содержащийся во многих книгах по эконометрике, дает анализ того, каким образом индивидуум попадает из дома на работу (пешком, метро, автобус, личный автомобиль). Предположим, что в городе существуют две конкурирующие транспортные компании, предоставляющие примерно одинаковые по качеству услуги («красный автобус» и «синий автобус»). Следует ожидать, что полезности этих двух альтернатив достаточно близки, что вступает в противоречие с их независимостью.

На эту же проблему можно посмотреть немного иначе. Из (12.12) следует, что

$$\frac{\mathrm{P}(y_t = j)}{\mathrm{P}(y_t = k)} = \frac{\exp(u_{tj})}{\exp(u_{tk})}$$

при любых $j, k = 1, \ldots, m$, т. е. отношение вероятностей двух альтернатив не зависит от остальных возможностей. Это свойство получило название «независимость от посторонних альтернатив» (*independence of irrelevant alternatives*). Если в нашем примере считать, что первая альтернатива — это личный автомобиль, а

вторая — «красный автобус», то отношение $\mathrm{P}(y_t = 1)/\mathrm{P}(y_t = 2)$ должно быть одно и то же, независимо от того, является третьей альтернативой «синий автобус» или метро, что выглядит весьма нереалистично.

Более подробно о *logit*-модели множественного выбора можно прочесть в книге (Greene, 1997).

Порядковые зависимые переменные

Если альтернативы упорядочены, то, используя скрытую (латентную) переменную, можно построить естественное обобщение модели (12.4), (12.5). Поясним на примере. Предположим, что у семьи есть три возможности провести отпуск:

1 — отдыхать на даче;
2 — отдыхать в Крыму;
3 — отдыхать в Испании.

Выбор места отдыха, описываемый переменной y, зависит от текущих накоплений y^* следующим образом:

$$\begin{aligned} y &= 1, \quad \text{если } y^* \leqslant c_1, \\ y &= 2, \quad \text{если } c_1 < y^* \leqslant c_2, \\ y &= 3, \quad \text{если } y^* > c_2, \end{aligned}$$

где c_1, c_2 — некоторые фиксированные уровни. Предполагая, что величина y^* удовлетворяет уравнению (12.4), и считая для простоты, что дисперсия ошибок $\sigma = 1$, имеем:

$$\begin{aligned} \mathrm{P}(y_t = 1) &= F(c_1 - x_t'\beta), \\ \mathrm{P}(y_t = 2) &= F(c_2 - x_t'\beta) - F(c_1 - x_t'\beta), \\ \mathrm{P}(y_t = 3) &= 1 - F(c_2 - x_t'\beta). \end{aligned} \tag{12.16}$$

Выбирая в качестве функции $F(\cdot)$ функцию нормального или логистического распределения, будем получать порядковые *probit*- или *logit*-модели.

Функция правдоподобия имеет следующий вид (ср. (12.8)):

$$L = \prod_{y_t=1} F(c_1 - x_t'\beta) \prod_{y_t=2} (F(c_2 - x_t'\beta) - F(c_1 - x_t'\beta)) \prod_{y_t=3} (1 - F(c_2 - x_t'\beta)).$$

Уровни c_1, c_2 могут быть априорно заданы, а могут быть неизвестны. В любом случае на основании этой формулы для порядковых *probit*- или *logit*-моделей можно строить оценки максимального правдоподобия параметров β и, если необходимо, c_1, c_2.

В общем случае модель упорядоченного множественного выбора с m альтернативами кратко описывается следующим образом. Пусть $-\infty = c_0 < c_1 < \cdots < c_{m-1} < c_m = \infty$ — точки на числовой прямой, и (скрытая) переменная y^* удовлетворяет уравнению (12.4):

$$y_t^* = x_t'\beta + \varepsilon_t.$$

Тогда

$$\mathrm{P}(y_t = j) = \mathrm{P}(c_{j-1} < y_t^* \leq c_j), \; j = 1, \ldots, m. \tag{12.17}$$

Таким образом,

$$\mathrm{P}(y_t = j) = F(c_j - x_t'\beta) - F(c_{j-1} - x_t'\beta), \; j = 1, \ldots, m, \tag{12.18}$$

где $F(\cdot)$ — функция распределения ошибки ε_t. Предполагая независимость ошибок, получаем следующее выражение для функции правдоподобия:

$$L = \prod_{j=1}^{m} \prod_{\{t:\, y_t = j\}} (F(c_j - x_t'\beta) - F(c_{j-1} - x_t'\beta)).$$

Максимизируя эту функцию, получаем оценки параметров β и c_j, $j = 1, \ldots, m-1$.

Пример. ***Анализ рейтингов российских банков.*** Этот пример основан на работе (Пересецкий, Карминский, ван Суст, 2003). Одной из важнейших характеристик банка является его надежность. Различные организации (журналы, аналитические центры и т. п.) регулярно публикуют рейтинги надежности банков. Часто методики, по которым эти рейтинги строятся, являются закрытыми и, как правило, используют экспертные оценки. Возникает естественный вопрос, можно ли связать рейтинг банка с показателями его деятельности. В работе (Пересецкий, Карминский, ван Суст,

2003), в частности, построена модель упорядоченного множественного выбора (12.17), (12.18), использующая данные Информационного центра (ИЦ) «Рейтинг». В этих данных содержится информация о рейтингах 115 российских банков, ранжированных по 6 группам надежности ($n = 115$, $m = 6$). В качестве объясняющих переменных x были выбраны следующие показатели:

BP/SK — прибыльность капитала;

DOSTKAP — достаточность капитала (Норматив Н1 ЦБ РФ);

DKE/VB — доля долгосрочных кредитов экономике в валюте банка;

MGLIK — мгновенная ликвидность (Норматив Н2 ЦБ РФ);

SK — собственный капитал.

Оценивание модели (12.17), (12.18) дало результаты, представленные в таблице 12.2.

Таблица 12.2

Переменная	Коэффициент
BP/SK	−0.132
DOSTKAP	0.054
DKE/VB	2.904
MGLIK	−0.019
SK	$-7.63 \cdot 10^{-07}$

При этом все коэффициенты, за исключением коэффициента при DKE/VB, оказались значимыми на 1%-ном уровне, а коэффициент при DKE/VB значим на 5%-ном уровне. Знаки коэффициентов согласуются со здравым смыслом и экономической интуицией. Отметим, что в этой модели уменьшение зависимой переменной соответствует повышению рейтинга, поэтому отрицательный знак коэффициента означает, что увеличение соответствующего фактора (при прочих равных) способствует повышению надежности.

Для оценки качества модели для некоторой группы банков были построены рейтинги, предписываемые моделью (модельные рейтинги) и проведено сравнение с рейтингами ИЦ «Рейтинг» (реальные рейтинги). Коэффициент Спирмэна между этими рейтингами оказался равным 0.76, для 80% банков модельные рейтинги совпали с реальными, и для всех банков отклонение модельного рейтинга от реального не превышало 1 (в ту или другую сторону). Эти факты свидетельствуют о достаточно высоком качестве построенной модели.

12.2. Модели с урезанными и цензурированными выборками

Урезанные выборки

Если выборка производится не из всей возможной совокупности наблюдений, а лишь из тех, что удовлетворяют каким-то априорным ограничениям, то такую выборку называют *урезанной*. Как правило, урезание приводит к смещенности МНК-оценок, поэтому для урезанных выборок используют в основном метод максимального правдоподобия (глава 10). В этом разделе мы рассмотрим случай, когда урезание осуществляется пороговым значением для зависимой переменной, т. е. исключаются все те наблюдения, у которых значение зависимой переменной меньше некоторой заданной величины.

Пусть случайная величина Z имеет плотность распределения $p(z)$. Нетрудно проверить, что для любого числа a условная плотность $p(z \mid Z > a)$ задается равенством

$$p(z \mid Z > a) = \frac{p(z)}{1 - F(a)}, \qquad z \geqslant a, \tag{12.19}$$

где $F(\cdot)$ — функция распределения Z.

Условное среднее случайной величины Z при условии $Z > a$ есть

$$\mathrm{E}(Z \mid Z > a) = \int_a^\infty z p(z \mid Z > a)\, dz.$$

Аналогично определяется условная дисперсия $\mathrm{V}(Z \mid Z > a)$.

Можно показать, что если $Z \sim N(m, \sigma^2)$, то

$$\mathrm{E}(Z \mid Z > a) = m + \sigma\lambda(c), \quad \mathrm{V}(Z \mid Z > a) = \sigma^2(1 - \gamma(c)), \tag{12.20}$$

где

$$c = (a - m)/\sigma, \quad \lambda(c) = \frac{\phi(c)}{1 - \Phi(c)}, \quad \gamma(c) = \lambda(c)(\lambda(c) - c),$$

а $\Phi(\cdot)$ и $\phi(\cdot)$ — соответственно функция и плотность стандартного нормального распределения. Заметим, что $\gamma(c) < 1$ при любом c.

Предположим, что имеется нормальная линейная регрессионная модель (см. п. 3.1)

$$y_t = x_t'\beta + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma^2), \tag{12.21}$$

и пусть выбираются только те наблюдения, для которых $y_t > a$. Тогда согласно (12.20) имеем:

$$\mathrm{E}(y_t \mid y_t > a) = x_t'\beta + \sigma\lambda(c_t), \tag{12.22}$$

и

$$\mathrm{V}(y_t \mid y_t > a) = \sigma^2(1 - \gamma(c_t)), \tag{12.23}$$

где $c_t = (a - x_t'\beta)/\sigma$. Найдем предельный эффект объясняющих факторов. Дифференцируя (12.22), опуская нижний индекс t и пользуясь легко проверяемым равенством $\phi'(u) = -u\phi(u)$, получаем:

$$\frac{\partial \mathrm{E}(y \mid y > a)}{\partial x} = \beta + \sigma\frac{d\lambda}{dc}\frac{\partial c}{\partial x} = \beta(1 - \gamma(c)). \tag{12.24}$$

Поскольку $0 < \gamma(c) < 1$ при любом c, то предельный эффект каждого фактора меньше соответствующего коэффициента.

Если для урезанной выборки записать регрессионное уравнение

$$y_t = x_t'\beta + u_t, \tag{12.25}$$

то в силу (12.22) ошибка u_t имеет среднее значение $\sigma\lambda(c_t)$ и дисперсию (12.23). Это означает, что, применяя к модели метод наименьших квадратов, мы будем получать смещенные и несостоятельные оценки параметров β.

Как и раньше, для построения состоятельных оценок можно воспользоваться методом максимального правдоподобия. Согласно (12.19) плотность распределения случайной величины y_t в урезанной выборке есть

$$p(y) = \frac{1}{\sigma}\frac{\phi((y - x_t'\beta)/\sigma)}{1 - \Phi((a - x_t'\beta)/\sigma)},$$

откуда следует, что логарифмическая функция правдоподобия задается равенством

$$l = -\frac{n}{2}(\ln(2\pi) + \ln\sigma^2) - \frac{1}{2\sigma^2}\sum_t (y_t - x_t'\beta)^2 - \sum_t \ln\left[1 - \Phi\left(\frac{a - x_t'\beta}{\sigma}\right)\right]. \tag{12.26}$$

Проделав необходимые вычисления, можно показать, что (векторное) уравнение правдоподобия выглядит так:

$$\frac{\partial l}{\partial \beta} = \sum_t \left(\frac{y_t - x_t'\beta}{\sigma^2} - \frac{\lambda(c_t)}{\sigma}\right) x_t = \mathbf{0}, \tag{12.27}$$

$$\frac{\partial l}{\partial \sigma^2} = \sum_t \left[-\frac{1}{2\sigma^2} + \frac{(y_t - x_t'\beta)^2}{2\sigma^4} - \frac{c_t\lambda(c_t)}{2\sigma^2}\right] = 0, \tag{12.28}$$

где, как и ранее,

$$c_t = \frac{a - x_t'\beta}{\sigma}, \quad \lambda(c_t) = \frac{\phi(c_t)}{1 - \Phi(c_t)}.$$

Исследование достаточности условий (12.27), (12.28) для максимизации функции правдоподобия (12.26) является весьма сложной задачей и выходит за рамки нашей книги.

Совершенно аналогично можно рассматривать модели, в которых урезание происходит с помощью условий $y_t < a$ или $a < y_t < b$.

Метод максимального правдоподобия для оценивания моделей с урезанными выборками реализован во многих современных эконометрических компьютерных пакетах.

Цензурированные выборки. Tobit-модель

Начало систематическому изучению в эконометрике моделей с цензурированными выборками положила работа Дж. Тобина (Tobin, 1958), в которой исследовались расходы семей на автомобили. Для некоторых семей эти расходы равнялись нулю (отказ от

покупки). Дж. Тобин заметил, что если в такой ситуации осуществить регрессию логарифма расходов на логарифм доходов, то оценка эластичности спроса на автомобили по доходам окажется смещенной и несостоятельной, и предложил методы состоятельного оценивания. Суть моделей с цензурированием состоит в том, что для части наблюдений известно не «истинное» значение зависимой переменной, а ее усеченное значение, определяемое уровнем цензурирования.

Модель с цензурированием может быть получена с помощью небольшой модификации модели (12.4), (12.5). Пусть ненаблюдаемая величина y^* удовлетворяет регрессионному уравнению

$$y_t^* = x_t'\beta + \varepsilon_t, \tag{12.29}$$

а наблюдается величина

$$y_t = \begin{cases} y_t^*, & \text{если } y_t^* > 0, \\ 0, & \text{если } y_t^* \leqslant 0. \end{cases} \tag{12.30}$$

Модель (12.29), (12.30) называют *tobit*-моделью, соединяя в ее названии имя Tobin и термины *probit*, *logit*.

Найдем $\mathrm{E}(y_t)$, предполагая, что ошибка ε_t в (12.29) имеет нормальное распределение с нулевым средним и дисперсией σ^2. Имеем:

$$\begin{aligned} \mathrm{E}(y_t) &= \mathrm{E}(y_t \mid y_t^* \leqslant 0)\mathrm{P}(y_t^* \leqslant 0) + \mathrm{E}(y_t \mid y_t^* > 0)\mathrm{P}(y_t^* > 0) \\ &= 0 \cdot \mathrm{P}(y_t^* \leqslant 0) + \mathrm{P}(\varepsilon_t > -x_t'\beta)(x_t'\beta + \mathrm{E}(\varepsilon_t \mid \varepsilon_t > -x_t'\beta)) \\ &= \Phi\left(\frac{x_t'\beta}{\sigma}\right) \cdot x_t'\beta + \sigma\phi\left(\frac{x_t'\beta}{\sigma}\right), \end{aligned} \tag{12.31}$$

что в общем случае не равно $x_t'\beta$. В последнем равенстве мы воспользовались первым равенством из (12.20). Если теперь провести регрессию y на x

$$y_t = x_t'\beta + u_t,$$

то метод наименьших квадратов даст смещенные и несостоятельные оценки параметров β, так как $\mathrm{E}(u_t) \neq 0$ в силу (12.31), откуда следует смещенность и несостоятельность МНК-оценок (п. 3.2).

Для получения состоятельных и асимптотически несмещенных оценок параметров β можно вновь, как и в предыдущем разделе, воспользоваться методом максимального правдоподобия. Поскольку в данном случае наблюдения (12.30) имеют смешанное распределение, то функция правдоподобия имеет следующий вид:

$$L = \prod_{y_t=0}\left(1-\Phi\left(\frac{x_t'\beta}{\sigma}\right)\right)\prod_{y_t>0}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2\sigma^2}(y_t-x_t'\beta)^2\right). \tag{12.32}$$

Первая группа сомножителей соответствует цензурированным наблюдениям, вторая — всем остальным. Отметим, что в отличие от *probit*- или *logit*-моделей, где параметры β и σ не идентифицируются по отдельности, здесь они в функции L «разделены» и каждый из них может быть оценен.

Процедура построения оценок максимального правдоподобия путем максимизации функции (12.32) реализована в большинстве современных эконометрических пакетов.

Кратко остановимся на вопросе интерпретации коэффициентов *tobit*-модели. Дифференцируя (12.31) по x и опуская нижний индекс t, получаем:

$$\frac{\partial \mathrm{E}(y)}{\partial x} = \Phi\left(\frac{x'\beta}{\sigma}\right)\beta, \tag{12.33}$$

т. е. предельный эффект объясняющих факторов является переменным, как и должно быть для нелинейной модели. Как и в случае *probit*- или *logit*-модели, для определения «среднего» предельного эффекта рекомендуется вычислять производные (12.33) для средних по выборке значений независимых переменных x.

В работе (McDonald and Moffit, 1980) было предложено следующее разложение для предельного эффекта (12.33):

$$\frac{\partial \mathrm{E}(y)}{\partial x} = [\Phi(z)(1 - r(z+r)) + \phi(z)(z+r)]\,\beta, \tag{12.34}$$

где

$$z = \frac{x'\beta}{\sigma}, \qquad r = \frac{\phi(z)}{\Phi(z)}.$$

Можно проверить, что соотношение (12.34) представимо в виде

$$\frac{\partial \mathrm{E}(y)}{\partial x} = \mathrm{P}(y^* > 0)\frac{\partial \mathrm{E}(y^* \mid y^* > 0)}{\partial x} + \frac{\partial \mathrm{P}(y^* > 0)}{\partial x}\mathrm{E}(y^* \mid y^* > 0). \qquad (12.35)$$

Равенство (12.35) показывает, что изменение экзогенных факторов приводит к изменению как условного среднего величины y^* в положительной области ее распределения, так и вероятности попадания в эту часть распределения.

Для *tobit*-модели, так же как для *probit*- и *logit*-моделей, существенную роль играет правильная спецификация модели; рассмотрение проблем, возникающих в связи с ошибками спецификации, можно найти, например, в (Greene, 1997), (Johnston and DiNardo, 1997).

Модель Хекмана

Анализируя рассмотренную в предыдущем разделе *tobit*-модель, нетрудно обнаружить одно ее существенное ограничение. Наблюдение $y_t = 0$ можно интерпретировать как отказ индивидуума t от участия в «мероприятии» (например, покупать — не покупать автомобиль, работать — не работать и т. п.). Если же $y_t > 0$, то величину y_t можно назвать интенсивностью участия (расходы на покупку, число рабочих часов в неделю и т. п.). Нетрудно понять, что в *tobit*-модели одни и те же факторы влияют как на вероятность, так и на интенсивность участия.

Предположим, что для некоторого фактора x_j соответствующий коэффициент больше нуля: $\beta_j > 0$. Тогда увеличение этого фактора будет повышать вероятность участия и в среднем увеличивать интенсивность участия. Иными словами, с точки зрения изменения вероятности и интенсивности участия каждый фактор действует «в одном направлении». Для многих ситуаций такое предположение оправдано. Например, при покупке автомобиля полный доход семьи может оказывать подобное воздействие. Однако так бывает далеко не всегда. Противоположная ситуация возникает, например, при анализе расходов на отдых. Тогда количество детей в семье отрицательно влияет на принятие решения о поездке в отпуск, но если такое решение принято, то этот же фактор оказывает положительное воздействие на увеличение расходов.

Чтобы преодолеть эти ограничения, рассмотрим модель, в которой принятие решения «участвовать — не участвовать» и определение степени участия разделены и могут зависеть от разных факторов.

Предположим, что величина y_t^* удовлетворяет уравнению линейной регрессии (степень участия)

$$y_t^* = x_t'\beta + \varepsilon_t. \qquad (12.36)$$

Решение «участвовать — не участвовать» описывается обычной моделью бинарного выбора:

$$g_t^* = z_t'\gamma + u_t, \qquad (12.37)$$

$$\begin{aligned} g_t &= 1, \quad \text{если } g_t^* \geqslant 0, \\ g_t &= 0, \quad \text{если } g_t^* < 0, \end{aligned} \qquad (12.38)$$

где z_t — экозогенные переменные, которые могут иметь общие компоненты с x_t, u_t — случайная ошибка. Наблюдения задаются следующим образом:

$$\begin{aligned} &y_t = y_t^*,\ g_t = 1, && \text{если } g_t^* \geqslant 0, \\ &y_t \text{ не наблюдается},\ g_t = 0, && \text{если } g_t^* < 0. \end{aligned} \qquad (12.39)$$

Переменные x_t, z_t наблюдаются при каждом t. Наконец, предположим, что $[\varepsilon_t, u_t]'$ — двумерный нормальный вектор, $\mathrm{E}(\varepsilon_t) = \mathrm{E}(u_t) = 0, \mathrm{V}(\varepsilon_t) = \sigma_\varepsilon^2, \mathrm{V}(u_t) = \sigma_u^2 = 1, \mathrm{Cov}(\varepsilon_t, u_t) = \sigma_{\varepsilon u}$. По t ошибки независимы. Напомним, что равенство $\sigma_u^2 = 1$ — это обычное условие нормировки в *probit*-моделях.

Модель (12.36)–(12.39) называется моделью Хекмана (Heckman model, см. (Heckman, 1979))[2]. Нетрудно проверить, что при $x_t = z_t$, $\beta = \gamma$, $\varepsilon_t = u_t$ мы получаем обычную *tobit*-модель.

[2]В англоязычной литературе эта модель называется также *sample selection model*. Кроме того, используют термины модель *tobit* II и, продолжая традиции, *heckit*-модель.

Вычислим среднее значение величины y_t (при условии, что она наблюдаема). Предварительно заметим (см. упражнение 12.15), что

$$\varepsilon_t = \frac{\sigma_{\varepsilon u}}{\sigma_u^2} u_t + \eta_t = \sigma_{\varepsilon u} u_t + \eta_t, \tag{12.40}$$

причем $\mathrm{E}(\eta_t) = 0$, и величины u_t и η_t независимы (на самом деле, первое слагаемое в правой части (12.40) — это условное среднее $\mathrm{E}(\varepsilon_t \mid u_t)$). Имеем далее

$$\begin{aligned}
\mathrm{E}(y_t \mid g_t = 1) &= \boldsymbol{x}_t'\boldsymbol{\beta} + \mathrm{E}(\varepsilon_t \mid g_t = 1) \\
&= \boldsymbol{x}_t'\boldsymbol{\beta} + \mathrm{E}(\varepsilon_t \mid u_t \geq -\boldsymbol{z}_t'\boldsymbol{\gamma}) \\
&= \boldsymbol{x}_t'\boldsymbol{\beta} + \sigma_{\varepsilon u}\mathrm{E}(u_t \mid u_t \geq -\boldsymbol{z}_t'\boldsymbol{\gamma}) \\
&= \boldsymbol{x}_t'\boldsymbol{\beta} + \sigma_{\varepsilon u}\frac{\varphi(\boldsymbol{z}_t'\boldsymbol{\gamma})}{\Phi(\boldsymbol{z}_t'\boldsymbol{\gamma})}.
\end{aligned} \tag{12.41}$$

В последнем равенстве мы воспользовались соотношениями (12.20).

Из равенства (12.41) следует, что применение обычного метода наименьших квадратов к наблюдениям y_t приведет, в общем случае, к смещенным оценкам параметров $\boldsymbol{\beta}$. Если же $\sigma_{\varepsilon u} = 0$, т. е. когда механизм выбора и степень участия независимы, смещение отсутствует. Величину $\varphi(\boldsymbol{z}_t'\boldsymbol{\gamma})/\Phi(\boldsymbol{z}_t'\boldsymbol{\gamma})$ в (12.41) обозначают $\lambda(\boldsymbol{z}_t'\boldsymbol{\gamma})$ и называют «лямбда Хекмана» (Heckman lambda).

Формально экзогенные переменные $\boldsymbol{x}_t, \boldsymbol{z}_t$ в (12.36), (12.37) могут быть произвольными, однако в реальных моделях, как правило, набор факторов, составляющих $\boldsymbol{x}_t$, включается и в $\boldsymbol{z}_t$.

Рассмотрим типичный пример. Пусть нас интересует, от чего зависит и как устанавливается зарплата. Предположим, что зарплата w_t^*, предлагаемая на рынке труда, удовлетворяет обычному уравнению линейной регрессии

$$w_t^* = \boldsymbol{x}_t'\boldsymbol{\beta} + \varepsilon_t, \tag{12.42}$$

где $\boldsymbol{x}_t$ — экзогенные переменные (возраст, стаж работы, уровень образования и т. п.), и это уравнение представляет главный интерес для исследователя. При оценивании уравнения (12.42) наблюдаться будет только зарплата работающих индивидуумов. Естественно, что для тех, кто не работает, величина w_t^* не наблюдается. Одна из возможных моделей, описывающая механизм принятия индивидуумом решения «работать — не работать», основана на понятии резервной зарплаты (*reservation wage*) (см. (Gronau, 1974)). Содержательно, резервная зарплата — это тот минимальный уровень зарплаты, который индивидуум считает для себя приемлемым. Если предлагаемая ему зарплата ниже этого уровня, он просто отказывается работать. Ясно, что резервная зарплата зависит от других переменных по сравнению с переменными, определяющими зарплату на рынке труда. Можно отметить такие факторы, как социальный статус, семейное положение, общий семейный доход, возраст и количество детей, которые формируют внутреннее представление индивидуума о минимальном уровне зарплаты, но, как правило, не учитываются работодателями при оплате труда.

Пусть w_t^r — резервная зарплата индивидуума t, и предположим, что

$$w_t^r = \boldsymbol{x}_{1t}'\boldsymbol{\beta}_1 + \varepsilon_{1t},$$

где $\boldsymbol{x}_{1t}$ — набор индивидульных характеристик. Тогда, обозначая $g_t^* = w_t^* - w_t^r$, получим

$$g_t^* = \boldsymbol{x}_t'\boldsymbol{\beta} - \boldsymbol{x}_{1t}'\boldsymbol{\beta}_1 + \varepsilon_t - \varepsilon_{1t},$$

что после очевидных переобозначений может быть представлено в виде (12.37):

$$g_t^* = \boldsymbol{z}_t'\boldsymbol{\gamma} + u_t.$$

Индивидуум t работает и, следовательно, величина w_t^* наблюдается, если $g_t^* \geq 0$, в противном случае зарплата w_t^* неизвестна. Следовательно, эта модель укладывается в схему модели Хекмана (12.36)–(12.39). При этом переменные $\boldsymbol{z}_t$, входящие в уравнение выбора, содержат переменные $\boldsymbol{x}_t$ и еще некоторые дополнительные переменные.

Можно считать, что ошибки ε_t и ε_{1t} независимы, поэтому в данном случае $\sigma_{\varepsilon u} = \mathrm{Cov}(\varepsilon_t, \varepsilon_t - \varepsilon_{1t}) = \sigma_\varepsilon^2 > 0$. Таким образом,

этот пример показывает, что в общем случае ошибки в уравнениях выбора и интенсивности участия коррелированы.

Оценивать модель Хекмана можно с помощью метода максимального правдоподобия. Обозначим M_0 множество тех t, для которых $g_t = 0$, и M_1 — множество тех t, для которых $g_t = 1$. Тогда логарифмическая функция правдоподобия имеет следующий вид:

$$l(\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma_\varepsilon^2, \sigma_{\varepsilon u}) = \sum_{M_0} \ln \mathrm{P}(g_t = 0) + \sum_{M_1} \ln f(y_t, g_t = 1), \qquad (12.43)$$

где $f(y_t, g_t = 1)$ — совместное (смешанное) распределение y_t и g_t. Используя условные распределения, получаем

$$f(y_t, g_t = 1) = \mathrm{P}(g_t = 1 \mid y_t) f(y_t),$$

где $f(y_t)$ — плотность распределения y_t. Поэтому (12.43) можно переписать так:

$$l(\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma_\varepsilon^2, \sigma_{\varepsilon u}) = \sum_{M_0} \ln \mathrm{P}(g_t = 0) + \sum_{M_1} [\ln \mathrm{P}(g_t = 1 \mid y_t) + \ln f(y_t)]. \qquad (12.44)$$

Здесь

$$\mathrm{P}(g_t = 0) = 1 - \Phi(\boldsymbol{z}_t'\boldsymbol{\gamma}), \qquad (12.45)$$

и

$$f(y_t) = \frac{1}{\sqrt{2\pi\sigma_\varepsilon^2}} \exp\left(-\frac{(y_t - \boldsymbol{x}_t'\boldsymbol{\beta})^2}{2\sigma_\varepsilon^2}\right). \qquad (12.46)$$

Наконец, для вычисления $\mathrm{P}(g_t = 1 \mid y_t)$ воспользуемся результатом (приложение МС, п. 4, N13): условное распределение $g_t^* \mid y_t$ является нормальным со средним

$$\mathrm{E}(g_t^* \mid y_t) = \boldsymbol{z}_t'\boldsymbol{\gamma} + \frac{\sigma_{\varepsilon u}}{\sigma_\varepsilon^2}(y_t - \boldsymbol{x}_t'\boldsymbol{\beta})$$

и дисперсией

$$\mathrm{V}(g_t^* \mid y_t) = 1 - \frac{\sigma_{\varepsilon u}^2}{\sigma_\varepsilon^2}.$$

Следовательно,

$$\mathrm{P}(g_t = 1 \mid y_t) = \Phi\left(\frac{\boldsymbol{z}_t'\boldsymbol{\gamma} + (\sigma_{\varepsilon u}/\sigma_\varepsilon^2)(y_t - \boldsymbol{x}_t'\boldsymbol{\beta})}{\sqrt{1 - \sigma_{\varepsilon u}^2/\sigma_\varepsilon^2}}\right). \qquad (12.47)$$

Таким образом, формулы (12.44)–(12.47) решают задачу вычисления логарифмической функции правдоподобия для модели Хекмана.

Применение метода максимального правдоподобия для оценивания модели Хекмана требует, как правило, создания программы, реализующей формулы (12.44)–(12.47) и последующую максимизацию функции (12.44). Чтобы избежать этого, в эмпирических исследованиях часто ограничиваются двухшаговым методом оценивания, который основан на формуле (12.41). Действительно, равенство (12.41) можно переписать в следующем виде:

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + \sigma_{\varepsilon u}\lambda(\boldsymbol{z}_t'\boldsymbol{\gamma}) + \nu_t, \qquad (12.48)$$

где, как нетрудно проверить, ошибка ν_t имеет нулевое математическое ожидание и некоррелирована с $\boldsymbol{x}_t, \boldsymbol{\gamma}_t$. Поэтому если бы величина $\lambda_t = \lambda(\boldsymbol{z}_t'\boldsymbol{\gamma})$ была известна, то применяя к уравнению (12.48) обычный метод наименьших квадратов, можно было бы получить несмещенные и состоятельные оценки параметров $\boldsymbol{\beta}$. Параметры $\boldsymbol{\gamma}$ неизвестны, но их можно состоятельно оценить по модели бинарного выбора (12.37)–(12.38) (первый шаг). Подставляя эти оценки вместо $\boldsymbol{\gamma}$ в (12.48) и применяя метод наименьших квадратов, можно получить состоятельные оценки параметров $\boldsymbol{\beta}$ (второй шаг). Конечно, эти оценки не будут эффективными. Одной из причин этого является, например, гетероскедастичность ошибок ν_t в (12.48). Поэтому на практике либо корректируют обычный метод наименьших квадратов (подробнее об этом можно прочесть, например, в книге (Greene, 1997)), либо используют МНК-оценки в качестве начального приближения в методе максимального правдоподобия.

Модели «времени жизни»

Цензурирование или урезание наблюдений естественным образом возникает при исследовании длительности какого-либо процесса, времени нахождения в каком-либо состоянии: период безотказной работы прибора, время жизни пациента после трансплантации сердца, промежуток времени между двумя арестами рецидивиста, период безработицы, длительность забастовки и т. п. Модели подобных явлений называют моделями *«времени жизни»* (*duration models*). Если к моменту наблюдения процесс не завершился, то точное значение его длительности неизвестно и она цензурируется моментом наблюдения или же это наблюдение исключается из рассмотрения (урезание).

Модели «времени жизни» уже в течение нескольких десятилетий изучаются в технике, медицине, демографии. В экономике их исследование началось сравнительно недавно ((Lancaster, 1974), (Lancaster, 1985), (Heckman and Singer, 1984), (Kiefer, 1988)).

Будем считать «время жизни» изучаемого объекта случайной величиной и обозначать τ. Одной из основных вероятностных характеристик τ является *интенсивность отказов* или *коэффициент смертности* (*hazard rate*):

$$\lambda(t) = \frac{p(t)}{1 - F(t)}, \tag{12.49}$$

где $F(\cdot)$ — функция, а $p(\cdot)$ — плотность распределения τ. Функцию (12.49) называют также *обратным отношением Миллса* (*inverse Mills ratio*). Эта функция уже встречалась ранее (12.19). Она имеет следующий смысл: если к моменту времени t процесс еще не завершился, т. е. $\tau > t$, то вероятность его окончания в течение следующего малого промежутка времени Δt есть $\lambda(t)\Delta t$, что оправдывает название функции $\lambda(\cdot)$. В моделях «времени жизни» наибольший интерес представляет именно эта функция. Вот типичные вопросы, ответы на которые зависят от поведения функции интенсивности отказов:

— Один пациент с пересаженным сердцем прожил после операции три года, второй — пять лет. Для кого из них опасность умереть в течение ближайшего года выше?

— Два человека ищут работу, один в течение трех месяцев, другой — в течение полугода. Для кого из них шанс найти работу в течение ближайшей недели выше?

— Какая забастовка более вероятно закончится завтра: та, что длится уже месяц, или та, что длится всего неделю?

Заметим, что из (12.49) легко следует, что

$$F(t) = 1 - \exp\left(-\int_0^t \lambda(s)\,ds\right),$$

т. е. распределение «времени жизни» τ однозначно восстанавливается по $\lambda(\cdot)$.

Наиболее часто в моделях «времени жизни» для случайной величины τ рассматривают распределения

показательное	$1 - F(t) = e^{-\lambda t}$	$\lambda(t) = \lambda$
Вейбулла	$1 - F(t) = e^{-\lambda t^\alpha}$	$\lambda(t) = \lambda\,\alpha\,t^{\alpha-1}$
лог-логистическое	$1 - F(t) = \dfrac{1}{1 + (\lambda t)^\alpha}$	$\lambda(t) = \dfrac{\lambda\,\alpha(\lambda t)^{\alpha-1}}{1 + (\lambda t)^\alpha}$

Здесь $\lambda > 0$, $\alpha > 0$ — параметры распределений. Показательное распределение является частным случаем распределения Вейбулла и характеризуется постоянной интенсивностью отказа; для распределения Вейбулла функция $\lambda(t)$ возрастает при $\alpha > 1$ и убывает при $\alpha < 1$; у лог-логистического распределения интенсивность отказа сначала возрастает, а затем убывает. Для оценивания параметров λ, α в условиях цензурирования наблюдений можно воспользоваться методом максимального правдоподобия, аналогично тому, как делается для *tobit*-модели. Пусть $\tau_1, \ldots, \tau_n$ — наблюдения реализаций «времени жизни». Если обозначить UCO множество нецензурированных наблюдений, а CO — множество цензурированных наблюдений (UnCensored и Censored Observations), то для логарифмической функции правдоподобия имеем следующее представление:

$$l = \ln L = \sum_{i \in \mathrm{UCO}} \ln p(\tau_i) + \sum_{i \in \mathrm{CO}} \ln(1 - F(\tau_i)). \tag{12.50}$$

Иногда удобно записывать логарифмическую функцию правдоподобия, используя функцию интенсивности отказов $\lambda(\cdot)$:

$$l = \sum_{i \in \mathrm{UCO}} \ln \lambda(\tau_i) + \sum_{i \in \mathrm{CO} \cup \mathrm{UCO}} \ln(1 - F(\tau_i)).$$

Во втором слагаемом суммирование берется по всем наблюдениям.

В модели «времени жизни» нетрудно ввести экзогенные переменные. Пусть «время жизни» имеет распределение Вейбулла, и есть набор $\boldsymbol{x} = (x_1, \ldots, x_k)'$ объясняющих переменных. Предполагая некоторую функциональную зависимость параметра λ от $\boldsymbol{x}$, например, $\lambda_i = e^{-\boldsymbol{x}_i'\boldsymbol{\beta}}$ (i — номер наблюдения), и подставляя эти значения в уравнение (12.50), можно построить оценки максимального правдоподобия параметров $\boldsymbol{\beta}$. Подробнее о проблеме интерпретации коэффициентов см. (Greene, 1997).

Упражнения

12.1. Покажите, что если среди регрессоров линейной модели вероятности или *logit*-модели есть константа, то среднее значение прогнозных вероятностей равно доле единиц во всей выборке зависимой переменной.

12.2. Докажите равенства (12.20).
(*Указание.* Воспользуйтесь равенством $\phi'(u) = -u\phi(u)$.)

12.3. Докажите равенство (12.33).

12.4. Проверьте справедливость представления (12.35).

12.5. Рассмотрим модель бинарного выбора $\mathrm{P}(y_t = 1) = F(\alpha + \beta d_t)$, где d — фиктивная переменная (принимающая значения 0 или 1). Ниже представлены результаты 100 наблюдений:

		y	
		0	1
d	0	20	32
	1	36	12

а) Оцените параметры α, β, используя *logit*-модель. Проверьте гипотезу H_0: $\beta = 0$.

б) Повторите а) для *probit*-модели. Изменяются ли ваши выводы?

12.6. Докажите вогнутость (по $\boldsymbol{\beta}$) логарифмической функции правдоподобия для *logit*-модели.

12.7. Рассмотрим простейшую *tobit*-модель: $y_t^* = \alpha + \varepsilon_t$, где $\varepsilon_t \sim N(0, \sigma^2)$, и $y_t = y_t^*$, если $y_t^* > 0$, $y_t = 0$, если $y_t^* \leqslant 0$.

Даны результаты 30 наблюдений переменной y_t:

0.768	2.911	0.461	0.000	0.678	0.000
0.000	1.233	1.868	0.709	0.000	0.000
1.010	0.000	1.422	1.543	4.411	2.385
0.000	2.487	3.469	1.778	2.931	1.283
0.000	0.060	3.198	5.546	1.546	4.680

а) Вычислите МНК-оценку параметра α по цензурированным наблюдениям y_t. Будет ли эта оценка завышена или занижена по сравнению с истинным значением α?

б) Повторите а) для усеченных на уровне 0 наблюдений y_t.

12.8. В таблице приведены 100 наблюдений бинарных переменных (x, y, z).

Количество наблюдений	y	x	z
12	0	0	0
10	0	0	1
20	0	1	0
0	0	1	1
8	1	0	0
28	1	0	1
22	1	1	0
0	1	1	1

а) В рамках *logit*-модели $\mathrm{P}(y = 1) = \Lambda(\alpha + \beta x + \gamma z)$ тестируйте значимость влияния z на y.

б) В рамках линейной модели вероятности $\mathrm{P}(y = 1) = \alpha + \beta x + \gamma z$ тестируйте значимость влияния z на y.

12.9. Запишите функцию правдоподобия для оценки параметров (β, σ^2) модели

$$y_t^* = x_t'\beta + \varepsilon_t, \qquad (y_t^* \text{— ненаблюдаемая переменная}),$$
$$y_t = \max\{y_t^*, c\}, \qquad \text{где } c \neq 0,\ \mathrm{E}\varepsilon = 0,\ \mathrm{V}(\varepsilon) = \sigma^2 I.$$

12.10. Пусть $y_t^* = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + u_t$, где $u_t \sim N(0, \sigma^2)$ и u_t независимы.

1) Бинарная переменная d определяется следующим образом:

$$d_t = \begin{cases} 1, & \text{если } y_t^* > 0, \\ 0, & \text{в противном случае.} \end{cases}$$

а) Выпишите вероятность того, что $d_t = 1$, как функцию переменных x_{ti}.

б) Какие параметры вы можете оценить по наблюдениям (x_{ti}, d_t)?

в) Найдите выражение для асимптотической матрицы ковариаций оценок максимального правдоподобия в этом случае.

г) Вы хотите проверить гипотезу, что переменная x_{t2} незначима. Опишите процедуру проверки этой гипотезы с помощью теста отношения правдоподобия.

2) Пусть

$$y_t = \begin{cases} y_t^*, & \text{если } y_t^* > 0, \\ 0, & \text{в противном случае.} \end{cases}$$

а) Какие параметры можно оценить по наблюдениям (x_t, y_t)?

б) Найдите выражение для предельного эффекта фактора x_{t3} для y_t и y_t^*.

в) Повторите процедуру проверки гипотезы о незначимости x_{t2} из п. 1 г) в данном случае. Какой из этих двух тестов выглядит для вас более предпочтительным?

12.11. Дано $n = n_1 + n_2 + n_3$ наблюдений переменных x и y. Известно, что для n_1 наблюдений $y = 1$ и $x = 1$, для n_2 наблюдений $y = 0$ и $x = 1$, для n_3 наблюдений $y = 0$ и $x = 0$. Покажите, что как для *logit*-, так и для *probit*-модели, уравнение правдоподобия не имеет решения.

12.12. Покажите, что логарифмическая функция правдоподобия для *probit*-модели является вогнутой (по β) функцией.

Указание. Покажите, что при любом x выполнено неравенство $x + \varphi(x)/\Phi(x) > 0$, и воспользуйтесь этим фактом.

12.13. Модель бинарного выбора описывается стандартным образом:

$$y_t = \begin{cases} 1, & \text{если } y_t^* > 0, \\ 0, & \text{в противном случае,} \end{cases}$$

где $y_t^* = x_t'\beta + \varepsilon_t$, а ошибки ε_t имеют распределение Лапласа.

а) Найдите логарифмическую функцию правдоподобия для оценивания вектора β. Является ли эта функция вогнутой по β?

б) Предположим, что вы оценили вектор β, используя *probit*-модель, и эти оценки примерно пропорциональны оценкам, полученным с помощью исходной модели. Чему приблизительно должен быть равен коэффициент пропорциональности?

12.14. Пусть $y_t^* = x_t'\beta + \varepsilon_t$, где ошибки ε_t имеют плотность распределения $f(x)$ и

$$y_t = \begin{cases} \alpha_1, & \text{если } y_t^* \leqslant \alpha_1, \\ y_t^*, & \text{если } \alpha_1 < y_t^* < \alpha_2, \\ \alpha_2, & \text{если } y_t^* \geqslant \alpha_2. \end{cases}$$

а) Найдите распределение y_t.

б) Найдите логарифмическую функцию правдоподобия для оценивания вектора β.

в) Найдите $(\partial \mathrm{E}y)/(\partial x)$.

12.15. Докажите, что в (12.40) случайные величины u_t и η_t независимы.

12.16. Расходы домашних хозяйств в Нидерландах (см. начало в упражнении 6.14).

Расходы на отдых и модели с усеченными переменными. Не все семьи расходуют деньги на отдых. В нашем случае $v3 = 0$ для 22.5% наблюдений. В этом разделе мы рассмотрим модели бинарного выбора для ответа на вопрос, тратит какие-нибудь средства на отдых или нет, игнорируя информацию о размерах этих затрат. Мы рассмотрим также *tobit*-модель, в которой явно учитывается смешанный дискретно-непрерывный тип переменной $v3$.

12.16.1. а) Постройте фиктивную переменную y, такую что $y = 1$, если $v3 > 0$ и $y = 0$, если $v3 = 0$.

б) Оцените линейную модель вероятности для y (выберите подходящий набор объясняющих переменных среди тех, что использовались в упражнениях 6.14.1–6.14.10). Проинтерпретируйте результаты. Как вычислены стандартные ошибки? Почему?

в) Найдите прогноз для y, основываясь на линейной модели вероятности. Находятся ли прогнозные значения в интервале от 0 до 1?

12.16.2. а) Оцените *probit*-модель для y. Проинтерпретируйте результаты.

б) Проверьте совместную значимость переменных в модели п. а).

в) Используя стратегию «от общего к частному», постройте подходящую *probit*-модель для y. Проинтерпретируйте результаты. Используйте эту модель в последующих упражнениях.

г) Найдите прогноз для y_i, т. е. оценки вероятностей событий $y_i = 1$. Вычислите также коэффициент детерминации для этой *probit*-модели.

12.16.3. а) Повторите упражнение 12.16.2, используя *logit*-модель вместо *probit*-модели.

б) Сравните результаты линейной модели вероятности, *probit*-модели и *logit*-модели.

12.16.4. а) Оцените *tobit*-модель для переменной $y = \ln(v3 + 1)$, используя регрессоры, выбранные ранее в упражнении 12.16.2.

б) Сравните результаты линейной модели вероятности, *probit*-модели, *logit*-модели и *tobit*-модели.

12.17. В Великобритании подростки в 16 лет делают важный выбор дальнейшей карьеры. В этом возрасте все они сдают специальный экзамен. Через несколько месяцев после этого они должны решить, продолжать учебу в школе или нет. Те из них, кто решает оставить школу, могут, в свою очередь, работать полный рабочий день или совмещать работу с учебой. В данных упражнениях мы попытаемся выяснить, какие факторы определяют этот выбор.

Здесь используются данные Британского национального опроса (UK National Child and Development Survey). Они содержат информацию о людях, родившихся в Великобритании в марте 1958 года. Детальное описание данных можно найти в статье (Micklewright, 1986). Данные о респондентах собирались в разные моменты их жизни. В упражнениях

используются данные о юношах и девушках (подвыборка тех, кто живет не в Шотландии). Большинство значений переменных относится к шестнадцатилетнему возрасту. Файл `choice.xls`[3] содержит следующие переменные (см. таблицу 12.3):

Таблица 12.3

Переменная	Описание
at16	Решение, принятое в 16-летнем возрасте (1 — продолжать учебу, 2 — совмещение учебы и работы, 3 — работать полный рабочий день)
able7	Результат теста общих способностей, проводящегося в 7-летнем возрасте
loginc	Логарифм дохода семьи (в 16 лет)
ctratio	Число учеников на одного учителя в школе (показатель, отражающий качество школы)
oldsib	Число старших братьев и сестер (в 16 лет)
yngsib	Число младших братьев и сестер (в 16 лет)
etot	Число полученных на выпускных экзаменах высших оценок (экзамен проводится в 16 лет, до принятия решения о продолжении учебы)
female	Пол (1 — для девушек, 0 — для юношей)

12.17.1. Вычислите описательные статистики переменных как для всей выборки, так и отдельно для девушек и юношей. Постройте гистограммы значений переменной *at16* отдельно для юношей и девушек. Интерпретируйте результаты.

12.17.2. Сконструируйте переменную *school*, равную 1, если $at16 = 1$, и равную 0 в противном случае. Оцените *logit*-регрессию переменной *school* на все остальные переменные и интерпретируйте результаты. Согласуются ли они с вашими ожиданиями?

12.17.3. Повторите упражнение 12.17.2, используя *probit* вместо *logit*. Сравните полученные результаты с результатами упражнения 12.17.2.

12.17.4. Выберите из *logit*- и *probit*-моделей более подходящую. Проверьте, одинаково ли влияет на принятие решения о продолжении учебы наличие в семье старших и младших братьев и сестер.

[3] Исходные данные Британского национального опроса можно скачать по адресу `http://www.data-archive.ac.uk/findingData/ncdsNews.asp`

12.17.5. Оцените *logit*- или *probit*-модель раздельно для юношей и для девушек (не забудьте удалить из списка объясняющих переменных переменную *female*). Прокомментируйте различия в результатах оценивания.

12.17.6. Проверьте гипотезу о применимости общей модели для всех подростков против гипотезы о том, что нужно использовать разные модели для юношей и для девушек (используйте тест отношения правдоподобия).

12.17.7. Используя наиболее подходящую, с вашей точки зрения, модель, вычислите прогнозную вероятность продолжить учебу для юноши и девушки со средними характеристиками. Также вычислите влияние на эту вероятность наличия в семье еще одного младшего брата (сестры).

12.17.8. Рассмотрите тех подростков, для которых $school = 0$, постройте переменную *job*, равную 1, если $at16 = 3$ (полный рабочий день), и 0, если $at16 = 2$ (совмещение работы и учебы). Проделайте упражнения 12.17.1–12.17.7 и проанализируйте, какие факторы влияют на выбор между работой и совмещением работы и учебы.

12.17.9. Используя наиболее подходящие, с вашей точки зрения, модели, вычислите прогнозную вероятность выбрать полный рабочий день ($school = 0$ и $job = 1$) для юноши и девушки со средними характеристиками. Также вычислите влияние на эту вероятность одного дополнительного младшего брата (сестры).

Глава 13

Панельные данные

13.1. Введение

Панельные данные (*Panel data*) состоят из наблюдений одних и тех же экономических единиц или объектов (индивидуумы, домашние хозяйства, фирмы, регионы, страны и т. п.), которые осуществляются в последовательные периоды времени. Примерами могут служить ежегодные бюджетные обследования одних и тех же домашних хозяйств, ежеквартальные данные о деятельности фиксированного множества предприятий, ежегодные социально-экономические показатели определенной группы стран. Таким образом, панельные данные сочетают в себе как данные пространственного типа (*cross-sectional data*), так и данные типа временных рядов (*time-series data*): в каждый момент времени имеются данные пространственного типа по экономическим единицам, и для каждого такого объекта соответствующие ему данные образуют один или несколько временных рядов. Благодаря специальной структуре панельные данные позволяют строить более гибкие и содержательные модели и получать ответы на вопросы, которые недоступны только в рамках, например, моделей, основанных на пространственных данных.

В частности, возникает возможность учитывать и анализировать индивидуальные отличия между экономическими единица-

ми, что нельзя сделать в рамках стандартных регрессионных моделей.

Классический пример такой ситуации, вошедший в большинство книг, посвященных панельным данным, привел (Ben-Porah, 1973). Предположим, что ежегодное исследование рынка труда показало, что процент работающих замужних женщин равен 50%. Как можно интерпретировать этот факт? Возможны две крайние точки зрения. Согласно первой, полученный результат означает, что каждая замужняя женщина имеет шанс 50% работать в течение года. Согласно второй, результаты исследования показывают, что 50% всех замужних женщин работают полный рабочий день, а остальные 50% вообще не работают. Ясно, что прогноз состояния рынка труда будет существенно разный в зависимости от того, какая из ситуаций имеет место. Более или менее адекватное представление о реальном положении можно получить, если проследить историю некоторого числа индивидуумов в течение определенного периода времени, т. е. в рамках панельных данных.

Можно привести еще несколько примеров, показывающих, что панельные данные дают возможность учесть эффекты, которые невозможно проследить, оставаясь в рамках обычных моделей. Так, при изучении величины ВВП на душу населения имеется возможность для какой-либо страны в каждый период времени наблюдать уровень инфляции, объем инвестиций, денежную массу и т. п. Но кроме этого существуют факторы, которые либо не наблюдаемы, либо нельзя представить в численной форме, но которые могут оказывать существенное влияние на исследуемый показатель: географическое положение, история, культурные традиции и т. д. При этом действие этих факторов можно считать постоянным (т. е. не зависящим от времени) для каждой национальной экономики. Имея лишь пространственные данные для нескольких стран, можно определить влияние обычных экономических факторов на величину ВВП на душу населения, но нельзя выявить индивидуальные различия между странами. При наличии наблюдений за одними и теми же странами в течение нескольких

периодов времени возникает возможность такие индивидуальные различия оценить.

На микроуровне одна из традиционных задач — объяснение расходов домашних хозяйств на тот или иной товар, например, средства личной гигиены. Можно собрать пространственные данные, включив в них экономические и социально-демографические характеристики семьи и т. п., и получить значимую зависимость расходов на изучаемый товар от семейного дохода. Однако панельные данные могут показать, что доход не всегда оказывает существенное влияние, а разницу в расходах следует объяснять семейными традициями, уровнем культуры и другими факторами, не всегда поддающимися измерению и наблюдению.

Часто индивидуальные факторы коррелированы с другими объясняющими переменными. Так, например, общий уровень культуры семьи и уровень ее дохода естественно считать связанными. В рамках моделей регрессии это означает, что индивидуальный фактор является существенной переменной в модели и ее исключение приводит к смещенным оценкам остальных параметров (см. п. 4.4). Иными словами, модели с панельными данными позволяют получать более точные оценки параметров.

В то же время, поскольку панельные данные содержат наблюдения за одними и теми же объектами в разные периоды времени, предположение о взаимной независимости этих наблюдений становится нереалистичным, поэтому анализ этих моделей может потребовать применения более тонких (по сравнению с обычным методом наименьших квадратов) методов оценивания.

В настоящее время существует много баз панельных данных как на микро-, так и на макроуровне. Например, в National Longitudinal Survey of Labor Market Experience (NLS), `http://www.bls.gov/nls/home.htm`. содержатся данные по рынку труда в США, в Michigan Panel Study of Income Dynamics (PSID), `http://www.isr.umich.edu/src/psid/`, — результаты бюджетных обследований домашних хозяйств в США. В последние годы проводится работа по сбору панельных данных в России (Russia Longitudinal Monitoring Survey, `http://www.cpc.unc.edu/rlms/`), пред-

ставляющих результаты обследований по широкому кругу вопросов (бюджетные, демографические, социальные и т. п.).

Все эти данные содержат наблюдения о большом количестве (несколько тысяч) экономических единиц (семей, индивидуумов) за относительно короткий промежуток времени. Такое соотношение между объемами пространственных и временных наблюдений является типичным для многих панельных данных.

Помимо этого существует большое число финансовых, макроэкономических и т. п. баз панельных данных.

В данной главе основное внимание уделяется статическим моделям с обычными количественными зависимыми переменными. Динамические модели и модели с дискретными зависимыми переменными рассматриваются менее детально. Их подробное изложение выходит за рамки начального курса, поэтому мы даем лишь описание моделей и краткое изложение возможных методов оценивания. В конце главы кратко описывается обобщенный метод моментов, который в настоящее время является одним из основных инструментов оценивания динамических моделей с панельными данными.

13.2. Обозначения и основные модели

Введем обозначения, которые будут использоваться на протяжении всей главы. Пусть y_{it} — зависимая переменная для экономической единицы i в момент времени t, $\boldsymbol{x}_{it}$ — набор объясняющих (независимых) переменных (вектор размерности k) и ε_{it} — соответствующая ошибка, $i = 1, \ldots, n$, $t = 1, \ldots, T$. Обозначим также

$$\boldsymbol{y}_i = \begin{bmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{bmatrix}, \quad \boldsymbol{X}_i = \begin{bmatrix} \boldsymbol{x}'_{i1} \\ \vdots \\ \boldsymbol{x}'_{iT} \end{bmatrix}, \quad \boldsymbol{\varepsilon}_i = \begin{bmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iT} \end{bmatrix}.$$

Введем также «объединенные» наблюдения и ошибки:

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_n \end{bmatrix}, \quad \boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_n \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_n \end{bmatrix}.$$

(Здесь $\boldsymbol{y}$, $\boldsymbol{\varepsilon}$ — $nT \times 1$ векторы, $\boldsymbol{X}$ — $nT \times k$ матрица.)

Простейшая модель — это обычная линейная модель регрессии

$$y_{it} = \boldsymbol{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it} \tag{13.1}$$

или в матричной форме

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{13.2}$$

которая, по существу, не учитывает панельную структуру данных. При этом предполагается, что все ошибки ε_{it} некоррелированы между собой как по i, так и по t, и некоррелированы со всеми объясняющими переменными $\boldsymbol{x}_{it}$. В эконометрической литературе эта модель носит название объединенной модели регрессии (*pooled model*). При выполнении сформулированных выше предположений обычные МНК-оценки $\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ являются состоятельными и эффективными.

Как уже говорилось, панельные данные позволяют учитывать индивидуальные различия между экономическими единицами. Одна из возможных реализаций этой идеи выглядит следующим образом:

$$y_{it} = \alpha_i + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}, \tag{13.3}$$

где величина α_i выражает индивидуальный эффект объекта i, не зависящий от времени t, *при этом регрессоры $\boldsymbol{x}_{it}$ не содержат константу.*

Приведем здесь еще один пример, когда учет индивидуального эффекта позволяет получить более адекватные выводы. Традиционный подход к задаче оценивания производственной функции состоит в оценке уравнения

$$y_{it} = \mu + \boldsymbol{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}, \tag{13.4}$$

где y_{it} — логарифм выпуска, а k-мерный вектор $\boldsymbol{x}'_{it}$ составлен из логарифмов производственных факторов. Однако для получения более правильного представления о производственной функции (особенно для небольших фирм) целесообразно учесть качество

управления, включив его в число факторов производства, т. е. рассмотреть модель

$$y_{it} = \mu + x'_{it}\beta + q_i\beta_{k+1} + \varepsilon_{it}, \tag{13.5}$$

где q_i обозначает качество управления. Если эта переменная является существенной, то МНК-оценки модели (13.4) являются смещенными. Однако величина q_i не наблюдаема, поэтому качество управления можно учесть лишь как индивидуальный эффект в виде (13.3). Этот прием не позволяет оценить параметр β_{k+1}, поскольку в модели (13.5) он не идентифицируем, но позволяет уменьшить смещение оценок параметров β.

В зависимости от предположений относительно характера величины α_i рассматриваются две модели.

Модель с фиксированным эффектом (*fixed effect model*): предполагается, что в уравнении (13.3) величины α_i являются неизвестными параметрами.

Модель со случайным эффектом (*random effect model*): предполагается, что в уравнении (13.3) $\alpha_i = \mu + u_i$, где μ — параметр, общий для всех единиц во все моменты времени, а u_i — ошибки, некоррелированные с ε_{it} и некоррелированные при разных i.

Задача выбора модели в каждом конкретном случае решается индивидуально. Ниже мы более подробно обсудим эту проблему.

13.3. Модель с фиксированным эффектом

Модель с фиксированным эффектом (*fixed effect model*) описывается уравнением (13.3), в котором переменные α_i являются неизвестными параметрами. Предположим, что выполнены следующие условия:

1) ошибки ε_{it} некоррелированы между собой по i и t, $\mathrm{E}(\varepsilon_{it}) = 0$, $\mathrm{V}(\varepsilon_{it}) = \sigma_\varepsilon^2$;

2) ошибки ε_{it} некоррелированы с регрессорами x_{js} при всех i, j, t, s.

Если ввести фиктивные переменные для каждой экономической единицы: $d_{ij} = 1$, если $i = j$, и $d_{ij} = 0$, если $i \neq j$, то

уравнение (13.3) может быть переписано в более привычном виде линейной регрессии

$$y_{it} = \sum_{j=1}^{n} \alpha_j d_{ij} + x'_{it}\beta + \varepsilon_{it}. \tag{13.6}$$

Если объединить все фиктивные переменные в одну большую матрицу

$$D = \begin{bmatrix} \imath_T & 0 & \cdots & 0 \\ 0 & \imath_T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \imath_T \end{bmatrix} = I_n \otimes \imath_T,$$

где вектор $\imath_T = [1, \ldots, 1]'$ имеет размерность T, а I_n — единичная матрица размера n, и обозначить $\alpha = [\alpha_1, \ldots, \alpha_n]'$, уравнение (13.6) можно по аналогии с уравнением (13.2) переписать в следующей матричной форме:

$$y = D\alpha + X\beta + \varepsilon. \tag{13.7}$$

Это соотношение можно рассматривать как стандартную модель регрессии и получать оценки параметров α, β с помощью обычного метода наименьших квадратов. При выполнении сделанных выше предположений 1)–3) относительно модели МНК-оценки будут несмещенными и эффективными. Эти оценки называются МНК-оценками с фиктивными переменными (*Least Squares Dummy Variable estimator, LSDV*). Следует более подробно остановиться на вопросе о состоятельности этих оценок. В панельных данных рост числа наблюдений может происходить как за счет увеличения количества экономических единиц n, так и за счет увеличения длительности наблюдений T. В первом случае происходит рост числа оцениваемых параметров (напомним, что необходимо оценить n параметров α и k параметров β), и гарантировать состоятельность, по крайней мере для оценок параметров α, нельзя. Во втором случае МНК-оценки состоятельны, но большие временные интервалы при небольшом числе экономических единиц в панельных данных встречаются редко.

При реализации этого метода могут возникнуть трудности. Во многих панельных данных число экономических единиц n обычно достаточно велико (несколько сотен или тысяч). Поэтому, применяя непосредственно метод наименьших квадратов к уравнению (13.7), при оценивании параметров можно столкнуться с вычислительными проблемами. К счастью, их можно преодолеть, если интересоваться только оценками параметров $\boldsymbol{\beta}$. Перейдем в уравнении (13.3) к средним по времени величинам:

$$\overline{y}_i = \alpha_i + \overline{\boldsymbol{x}}_i'\boldsymbol{\beta} + \overline{\varepsilon}_i, \tag{13.8}$$

где $\overline{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$, $\overline{\boldsymbol{x}}_i = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}_{it}$, $\overline{\varepsilon}_i = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{it}$. Вычитая почленно (13.8) из (13.3), получаем

$$y_{it} - \overline{y}_i = (\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_i)'\boldsymbol{\beta} + \varepsilon_{it} - \overline{\varepsilon}_i. \tag{13.9}$$

По существу, это — уравнение (13.3), записанное в отклонениях от индивидуальных средних по времени. В матричной форме соотношение (13.9) может быть записано так:

$$\boldsymbol{M_D y} = \boldsymbol{M_D X}\boldsymbol{\beta} + \boldsymbol{M_D}\boldsymbol{\varepsilon}, \tag{13.10}$$

где $\boldsymbol{M_D} = \boldsymbol{I}_{nT} - \boldsymbol{D}(\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}'$ — матрица, осуществляющая вычисление отклонений от индивидуальных средних. Это преобразование называется *внутригрупповым преобразованием* (*within transformation*). Применяя обычный метод наименьших квадратов к уравнению (13.9) (или к уравнению (13.10)), мы получим оценки

$$\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{M_D}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{M_D}\boldsymbol{y}, \tag{13.11}$$

совпадающие с МНК-оценками параметров $\boldsymbol{\beta}$ в исходном уравнении (13.7), т. е. с МНК-оценками с фиктивными переменными (см. упражнение 4.3). Эти оценки также называются *внутригрупповыми оценками* (*within estimator*) или *оценками с фиксированным эффектом* (*fixed effect estimator*): $\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}_{\mathrm{W}} = \widehat{\boldsymbol{\beta}}_{\mathrm{FE}}$. Их можно представить также в следующем виде:

$$\widehat{\boldsymbol{\beta}}_{\mathrm{FE}} = \left(\sum_{i=1}^{n}\sum_{t=1}^{T}(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_i)'\right)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T}(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_i)(y_{it} - \overline{y}_i).$$

Условия 1)–2), наложенные на модель, гарантируют несмещенность и состоятельность оценок с фиксированным эффектом.

В качестве оценок индивидуальных эффектов можно взять $\widehat{\alpha}_i = \overline{y}_i - \overline{\boldsymbol{x}}_i'\widehat{\boldsymbol{\beta}}_{\mathrm{FE}}$, $i = 1, \ldots, n$. Эти оценки, как легко проверить, являются несмещенными и состоятельными для фиксированного n при $t \to \infty$.

Из формулы (13.11) легко вытекает выражение для матрицы ковариаций оценки $\widehat{\boldsymbol{\beta}}_{\mathrm{FE}}$:

$$\begin{aligned}\mathrm{V}(\widehat{\boldsymbol{\beta}}_{\mathrm{FE}}) &= \sigma_\varepsilon^2(\boldsymbol{X}'\boldsymbol{M_D}\boldsymbol{X})^{-1} \\ &= \sigma_\varepsilon^2\left(\sum_{i=1}^{n}\sum_{t=1}^{T}(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_i)(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_i)'\right)^{-1}.\end{aligned} \tag{13.12}$$

Как и в обычной линейной модели, в качестве оценки дисперсии σ_ε^2 можно взять сумму квадратов остатков регрессии (13.9) (или (13.10)), деленную на число степеней свободы:

$$\widehat{\sigma}_\varepsilon^2 = \frac{1}{nT - n - k}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it} - \overline{y}_i - (\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_i)'\widehat{\boldsymbol{\beta}}_{\mathrm{FE}}\right)^2. \tag{13.13}$$

При достаточно слабых условиях регулярности оценки с фиксированным эффектом являются асимптотически нормальными (при $n \to \infty$ или при $T \to \infty$), поэтому можно пользоваться стандартными процедурами (t-тесты, F-тесты) для проверки гипотез относительно параметров $\boldsymbol{\beta}$.

Сделаем одно важное замечание. В панельных данных среди независимых переменных $\boldsymbol{x}_{it}$ могут быть такие, которые не меняются во времени для каждой экономической единицы. Например, при анализе заработной платы в число объясняющих факторов, как правило, включают пол и/или расовую принадлежность индивидуума. Модель с фиксированным эффектом не позволяет идентифицировать соответствующие таким переменным коэффициенты. Формально это объясняется тем, что в уравнении (13.9) один или несколько регрессоров равны нулю (или, что эквивалентно, матрица $[\boldsymbol{D}\ \boldsymbol{X}]$ в уравнении (13.7) имеет неполный ранг), и, следовательно, применять метод наименьших квадратов нельзя.

Если говорить менее формально, то инвариантный во времени фактор является, по существу, частью полного индивидуального эффекта, и выделить влияние только этого фактора нельзя.

Рассмотрим пример оценивания производственной функции российских предприятий топливно-энергетической отрасли. Данные собраны и обработаны сотрудницей Центра экономических и финансовых разработок (ЦЭФИР) Е. А. Бессоновой.

Пример. ***Оценка производственной функции российских предприятий топливно-энергетической отрасли.*** Данные содержат информацию о выпуске, трудозатратах, капитальных вложениях и о некоторых других факторах для 48466 предприятий за период 1993–2000 гг. Из них 1020 относятся к топливно-энергетической отрасли. Попробуем ответить на вопрос, можно ли моделировать работу этих предприятий с помощью производственной функции Кобба–Дугласа $Q = AK^{\alpha}L^{\beta}$, где Q — выпуск, K — капиталовложения, L — трудозатраты, A — константа. Для этого попытаемся оценить эластичности α, β с помощью простой (объединенной) регрессии и на основе модели с фиксированным эффектом.

Приводимые далее результаты получены с помощью пакета STATA. Поэтому форма их представления во многом копирует тот формат, в котором они выдаются этим пакетом.

1. *Простая (объединенная) регрессия.* Для оценки эластичностей α, β можно осуществить простую регрессию $\ln Q_{it}$ на $\ln K_{it}, \ln L_{it}$ и константу:

Dependent Variable: $\ln Q$

Variable	*Coefficient*	*Std.Error*	*t-Statistic*	*Probability*
$\ln K$	0.32957	0.01636	20.148	0.000
$\ln L$	0.92838	0.02465	37.657	0.000
const	−2.48807	0.02465	−31.084	0.000
R-squared	0.5805			

Результаты этой регрессии позволяют считать предположение о том, что производственная функция имеет вид функции Кобба–Дугласа, вполне правдоподобным: оценки получились значимыми, и каждая из эластичностей меньше 1. Однако вполне возможно, что если будут учтены индивидуальные эффекты, то оценки могут существенно измениться.

2. Регрессия с фиксированным эффектом.

Dependent Variable: $\ln Q$

Variable	*Coefficient*	*Std.Error*	*t-Statistic*	*Probability*
$\ln K$	0.11421	0.02391	4.777	0.000
$\ln L$	0.60393	0.03181	18.986	0.000
const	0.75995	0.19401	3.917	0.000
σ_u	2.17317			
σ_ε	0.763405			
ρ	0.890154	*(fraction of variance due to α_i)*		
F-test that all $\alpha_i = 0$: $F(2402, 6013) = 17.62$ $Prob > F = 0.0000$				

Мы видим, что результаты последней регрессии отличаются от первой: оценки эластичностей как по труду, так и по капиталу уменьшились. Забегая немного вперед, следует отметить, что F-тест в данном случае отвергает нулевую гипотезу об отсутствии индивидуальных эффектов (последняя строка таблицы). Иными словами, модель с фиксированным эффектом выглядит более приемлемой по сравнению с моделью объединенной регрессии.

13.4. Модель со случайным эффектом

Модель со случайным эффектом (*random effect model*) описывается уравнением

$$y_{it} = \mu + x'_{it}\beta + u_i + \varepsilon_{it}, \tag{13.14}$$

где μ — константа, а u_i — случайная ошибка, инвариантная по времени для каждой экономической единицы. Иными словами, в модели со случайным эффектом предполагается, что индивидуальные отличия носят случайный характер. Будем считать, что выполнены следующие условия:

1) ошибки ε_{it} некоррелированы между собой, $\mathrm{E}(\varepsilon_{it}) = 0$, $\mathrm{V}(\varepsilon_{it}) = \sigma_\varepsilon^2$;

2) ошибки ε_{it} некоррелированы с регрессорами x_{js} при всех i, j, t, s;

3) ошибки u_i некоррелированы, $\mathrm{E}(u_i) = 0$, $\mathrm{V}(u_i) = \sigma_u^2$;

4) ошибки u_i некоррелированы с регрессорами x_{jt} при всех i, j, t;

5) ошибки u_i и ε_{jt} некоррелированы при всех i, j, t.

Модель со случайным эффектом (13.14) можно рассматривать как линейную модель, в которой ошибка $w_{it} = u_i + \varepsilon_{it}$ имеет некоторую специальную структуру.

Следуя обозначениям раздела 13.2, можно переписать уравнение (13.14) в виде

$$y_i = \mu \imath_T + X_i\beta + w_i \tag{13.15}$$

или, используя объединенные наблюдения, в виде

$$y = \mu \imath_{nT} + X\beta + w. \tag{13.16}$$

Для получения оценок параметров μ, β можно к модели (13.16) применить обычный метод наименьших квадратов. Условия 1)–3) гарантируют несмещенность и состоятельность этих оценок. Однако ошибки в (13.16) не являются гомоскедастичными, поэтому для построения эффективных оценок можно воспользоваться обобщенным методом наименьших квадратов (см. п. 5.2).

Напомним, что $w_i = [u_i + \varepsilon_{i1}, \ldots, u_i + \varepsilon_{iT}]'$, поэтому матрица ковариаций вектора ошибок w_i в (13.15) имеет вид

$$\Sigma = \mathrm{E}(w_i w_i') = \begin{bmatrix} \sigma_u^2 + \sigma_\varepsilon^2 & \sigma_u^2 & \ldots & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2 + \sigma_\varepsilon^2 & \ldots & \sigma_u^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_u^2 & \sigma_u^2 & \ldots & \sigma_u^2 + \sigma_\varepsilon^2 \end{bmatrix} = \sigma_u^2 \imath_T \imath_T' + \sigma_\varepsilon^2 I_T.$$

Для объединенных наблюдений (13.16) тогда имеем:

$$\Omega = \mathrm{E}(ww') = \begin{bmatrix} \Sigma & 0 & \ldots & 0 \\ 0 & \Sigma & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \Sigma \end{bmatrix} = I_n \otimes \Sigma.$$

В соответствии с формулой (5.4) для построения ОМНК-оценок параметров μ, β необходимо обратить матрицу Σ. С помощью непосредственных вычислений можно показать, что

$$\Sigma^{-1} = \frac{1}{\sigma_\varepsilon^2}\left[I_T - \frac{\sigma_u^2}{\sigma_\varepsilon^2 + T\sigma_u^2}\imath_T \imath_T'\right]. \tag{13.17}$$

Эту матрицу можно преобразовать к виду

$$\Sigma^{-1} = \frac{1}{\sigma_\varepsilon^2}\left[\left(I_T - \frac{1}{T}\imath_T \imath_T'\right) + \theta^2 \frac{1}{T}\imath_T \imath_T'\right], \tag{13.18}$$

где

$$\theta^2 = \frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_u^2}. \tag{13.19}$$

Согласно одному из свойств произведений Кронекера (приложение ЛА, п. 18) имеем $\Omega^{-1} = I_n \otimes \Sigma^{-1}$. В силу формулы (5.4) получаем

$$\begin{bmatrix} \widehat{\mu}_{\mathrm{GLS}} \\ \widehat{\beta}_{\mathrm{GLS}} \end{bmatrix} = \left(\begin{bmatrix} \imath_{nT}' \\ X' \end{bmatrix} (I_n \otimes \Sigma^{-1}) \begin{bmatrix} \imath_{nT} & X \end{bmatrix}\right)^{-1} \begin{bmatrix} \imath_{nT}' \\ X' \end{bmatrix} (I_n \otimes \Sigma^{-1})\, y. \tag{13.20}$$

Равенства (13.18), (13.20) в принципе решают проблему оценивания параметров μ, β обобщенным методом наименьших квадратов и соответствующие оценки называются *оценками со случайным эффектом (random effect)*: $\widehat{\beta}_{\mathrm{GLS}} = \widehat{\beta}_{\mathrm{RE}}$. Однако с помощью рутинных вычислений, которые мы оставляем читателю в качестве упражнения, можно получить более наглядное представление об оценках $\widehat{\beta}_{\mathrm{RE}}$.

Ранее (п. 13.3) мы ввели понятие внутригрупповых оценок, перейдя от исходной регрессии (13.3) к регрессии в отклонениях от внутригрупповых средних (13.9). Преобразуем теперь уравнение (13.14), взяв средние значения по времени для каждой экономической единицы:

$$\overline{y}_i = \mu + \overline{x}_i'\beta + u_i + \overline{\varepsilon}_i. \tag{13.21}$$

Оценки, которые получаются применением к уравнению (13.21) обычного метода наименьших квадратов, называются *межгрупповыми оценками* (*between estimator*): $\widehat{\beta} = \widehat{\beta}_{\mathrm{B}}$. Эти оценки, как нетрудно проверить, являются несмещенными и состоятельными

при $n \to \infty$, но неэффективными. Так же, как в разделе 13.3, можно в уравнении (13.21) перейти к отклонениям уже от глобальных средних и представить межгрупповые оценки в следующем виде:

$$\widehat{\beta}_{\mathrm{B}} = \left(\sum_{i=1}^{n}(\overline{x}_i - \overline{x})(\overline{x}_i - \overline{x})'\right)^{-1} \sum_{i=1}^{n}(\overline{x}_i - \overline{x})(\overline{y}_i - \overline{y}),$$

где

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n}\overline{y}_i = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} y_{it}, \quad \overline{x} = \frac{1}{n}\sum_{i=1}^{n}\overline{x}_i = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} x_{it}$$

— глобальные средние зависимой и независимых переменных.

Детальный анализ формулы (13.20) с использованием обращения блочных матриц, структуры матрицы Σ^{-1} и равенств (13.17), (13.18) позволяет получить следующее представление для оценки со случайным эффектом:

$$\begin{aligned}\widehat{\beta}_{\mathrm{RE}} &= W\widehat{\beta}_{\mathrm{B}} + (I_k - W)\widehat{\beta}_{\mathrm{W}} \\ &= W\widehat{\beta}_{\mathrm{B}} + (I_k - W)\widehat{\beta}_{\mathrm{FE}},\end{aligned} \tag{13.22}$$

(см. упражнение 13.2), где W — некоторая матрица, которую можно вычислить в явном виде и которая пропорциональна матрице, обратной матрице ковариаций оценки $\widehat{\beta}_{\mathrm{B}}$ (подробнее см. (Hsiao, 1986)).

Таким образом, оценка со случайным эффектом является средневзвешенной внутри- и межгрупповой оценок. Можно также проверить, что оценка $\widehat{\beta}_{\mathrm{OLS}}$, полученная в обычной линейной модели (уравнение (13.2)), также может быть представлена как средневзвешенное внутри- и межгрупповой оценок (см. упражнение 13.3).

Напомним, что если P — такая матрица, что $P'P = \Sigma^{-1}$, то умножая обе части исходного уравнения (13.16) на матрицу $P \otimes \Sigma^{-1}$ и применяя к преобразованной системе обычный метод наименьших квадратов, можно получить оценки обобщенного метода наименьших квадратов, т. е. оценки β_{RE} (см. п. 5.2). Можно

проверить (см. упражнение 13.4), что в данном случае матрицу P можно выбрать в виде

$$P = \frac{1}{\sigma_\varepsilon}\left(I_T - \frac{1-\theta}{T}\imath_T\imath_T'\right). \tag{13.23}$$

Матрица $\frac{1}{T}\imath\imath'$ вычисляет среднее по времени компонент соответствующего вектора. Поэтому преобразованная система в скалярном виде выглядит следующим образом:

$$\begin{aligned}&y_{it} - (1-\theta)y_i = (1-\theta)\mu + (x_{it} - (1-\theta)\overline{x}_i)'\beta + \eta_{it}, \\ &i = 1, \ldots, n,\ t = 1, \ldots, T,\end{aligned} \tag{13.24}$$

где η_{it} — (гомоскедастичная) ошибка.

Из этого представления следует, что $\widehat{\beta}_{\mathrm{RE}} \to \widehat{\beta}_{\mathrm{W}} (= \widehat{\beta}_{\mathrm{FE}})$ при $\theta \to 0$ и $\widehat{\beta}_{\mathrm{RE}} \to \widehat{\beta}_{\mathrm{OLS}}$ при $\theta \to 1$.

Можно показать (для этого опять требуется обращать блочные матрицы), что

$$\begin{aligned}\mathrm{V}(\widehat{\beta}_{\mathrm{RE}}) = \sigma_\varepsilon^2\Bigg(&\sum_{i=1}^{n}\sum_{t=1}^{T}(x_{it} - \overline{x}_i)(x_{it} - \overline{x}_i)' \\ &+ \theta^2 T\sum_{i=1}^{n}(\overline{x}_i - \overline{x})(\overline{x}_i - \overline{x})'\Bigg)^{-1}\end{aligned} \tag{13.25}$$

(см. упражнение 13.5).

Сопоставляя (13.12) и (13.25), видим, что (как и следовало ожидать) выполняется неравенство $\mathrm{V}(\widehat{\beta}_{\mathrm{RE}}) \leqslant \mathrm{V}(\widehat{\beta}_{\mathrm{W}})$.

На практике дисперсии ошибок ε_{it} и u_i неизвестны. Поэтому чтобы реализовать метод оценивания со случайным эффектом (13.20) (т. е. осуществить доступный обобщенный метод наименьших квадратов, см. гл. 5, п. 5.3), необходимо оценить дисперсии σ_u^2 и σ_ε^2. Для этой цели можно воспользоваться результатами внутри- и межгрупповой регрессий. Оценка $\widehat{\sigma}_\varepsilon^2$ вычисляется по формуле (13.13) по результатам внутригрупповой регрессии. В межгрупповой регрессии (13.21) нетрудно вычислить дисперсию ошибки:

$$\sigma_{\mathrm{B}}^2 = \mathrm{V}(u_i + \overline{\varepsilon}_i) = \sigma_u^2 + \frac{1}{T}\sigma_\varepsilon^2. \tag{13.26}$$

Эта дисперсия оценивается стандартным образом по остаткам межгрупповой регрессии:

$$\widehat{\sigma}_{\mathrm{B}}^2 = \frac{1}{n-k-1}\sum_{i=1}^{n}(\overline{y}_i - \widehat{\mu}_{\mathrm{B}} - \overline{x}_i'\widehat{\beta}_{\mathrm{B}})^2. \qquad (13.27)$$

Наконец, в соответствии с (13.26) для дисперсии σ_u^2 строится оценка

$$\widehat{\sigma}_u^2 = \widehat{\sigma}_{\mathrm{B}}^2 - \frac{1}{T}\widehat{\sigma}_\varepsilon^2. \qquad (13.28)$$

Таким образом, в модели со случайным эффектом удается реализовать доступный обобщенный метод наименьших квадратов.

Формула (13.28) дает несмещенную и состоятельную оценку дисперсии $\widehat{\sigma}_u^2$. Однако на практике, когда объем выборки невелик, может так случиться, что эта оценка окажется отрицательной. В этих ситуациях нужно использовать другие оценки дисперсии σ_u^2, описание которых выходит за рамки данной книги. Подробнее об этом можно прочесть в книге (Greene, 1997, глава 14.4).

Пример. ***Оценка производственной функции российских предприятий топливно-энергетической отрасли.*** Продолжение 1 (см. начало п. 13.3, стр. 366).

Оценка производственной функции Кобба–Дугласа для предприятий топливно-энергетической отрасли в модели со случайным эффектом приводит к следующим результатам:

Dependent Variable: $\ln Q$

Variable	*Coefficient*	*Std.Error*	*z-Statistic*	*Probability*
$\ln K$	0.24671	0.01763	13.990	0.000
$\ln L$	0.77544	0.02578	30.078	0.000
const	−1.13749	0.11982	−9.494	0.000
σ_u	1.94279			
σ_ε	0.76341			

Как и раньше, оценки эластичностей получились значимыми и согласующимися со здравым смыслом. По сравнению с оценками, полученными в модели объединенной регрессии и в модели со случайным эффектом, они занимают промежуточное положение между первыми и вторыми.

Подведем кратко итоги.

При наличии панельных данных возникает несколько оценок вектора коэффициентов $\boldsymbol{\beta}$:

1) МНК-оценка в обычной модели регрессии (13.1).
2) Внутригрупповая оценка — оценка в регрессии, которая оперирует с отклонениями исходных данных от средних по времени для каждой экономической единицы. Эта же оценка называется оценкой с фиксированным эффектом (13.9).
3) Межгрупповая оценка — оценка в регрессии индивидуальных средних по времени (13.21).
4) Оценка со случайным эффектом — оценка, полученная применением обобщенного метода наименьших квадратов в модели (13.14).

При этом как оценка $\widehat{\boldsymbol{\beta}}_{\mathrm{OLS}}$, так и оценка $\widehat{\boldsymbol{\beta}}_{\mathrm{GLS}}$ являются средневзвешенными внутри- и межгрупповой оценок (каждая, естественно, со своей весовой матрицей).

13.5. Качество подгонки

Для стандартных моделей регрессии качество подгонки (при условии, что среди регрессоров есть константа) обычно измеряет коэффициент детерминации R^2 или скорректированный коэффициент детерминации R^2_{adj}. Напомним, что коэффициент детерминации интерпретируется как доля объясненной вариации зависимой переменной. Для моделей с панельными данными это понятие требует уточнения и модификации. Во-первых, внутригрупповая и межгрупповая модели имеют дело с разными вариациями зависимой переменной. Во-вторых, модель со случайным эффектом оценивается с помощью обобщенного метода наименьших квадратов, для которого коэффициент детерминации вообще не является адекватной мерой качества подгонки.

Один из возможных подходов основывается на том факте, что для стандартной линейной модели регрессии коэффициент детерминации R^2 совпадает с квадратом выборочного коэффициента

корреляции между y и прогнозным значением $\widehat{y}$ (см. упражнение 2.7)[1]. Исходя из этого можно для внутригрупповой регрессии (т. е. для модели с фиксированным эффектом) определить коэффициент детерминации равенством

$$R^2_{\text{within}} = r^2(y_{it} - \overline{y}_i,\ \widehat{y}_{it} - \widehat{y}_i), \tag{13.29}$$

где $\widehat{y}_{it} - \widehat{y}_i = (x_{it} - \overline{x}_i)'\widehat{\beta}_{\text{FE}}$ и $r^2(\cdot,\cdot)$ — выборочный коэффициент корреляции (в данном случае объем выборки равен nT).

Аналогично можно определить коэффициент детерминации для межгрупповой регрессии:

$$R^2_{\text{between}} = r^2(\overline{y}_i, \widehat{y}_i), \tag{13.30}$$

где $\widehat{y}_i = \overline{x}_i'\widehat{\beta}_{\text{B}}$ (в данном случае объем выборки равен n).

Наконец, для обычной модели объединенный коэффициент детерминации есть

$$R^2_{\text{overall}} = r^2(y_{it}, \widehat{y}_{it}), \tag{13.31}$$

где $\widehat{y}_{it} = x_{it}'\widehat{\beta}_{\text{OLS}}$.

На самом деле формулами (13.29), (13.30), (13.31) можно определить внутригрупповой, межгрупповой и объединенный коэффициенты детерминации для любой оценки $\widehat{\beta}$ вектора параметров β. При этом, естественно, считают, что $\widehat{y}_{it} = x_{it}'\widehat{\beta}$, $\widehat{y}_i = \frac{1}{T}\sum_{t=1}^{T}\widehat{y}_{it}$, $\widehat{y} = \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}\widehat{y}_{it}$. Именно такой подход реализован, например, в широко используемом эконометрическом пакете STATA: при оценивании моделей с панельными данными вычисляются три коэффициента детерминации в соответствии с формулами (13.29), (13.30), (13.31) независимо от того, какой метод оценивания применяется.

Следует сказать, что коэффициенты детерминации в моделях с панельными данными нецелесообразно использовать для того, чтобы определить, какой метод оценивания лучше. Так, если одну и ту же модель оценить, например, обычным методом наименьших квадратов и с помощью случайного эффекта, то объединенный коэффициент детерминации в первом случае всегда будет

[1] Строго говоря, в этом упражнении речь идет о модели парной регрессии. Однако утверждение легко обобщается на случай множественной регрессии.

больше соответствующего объединенного коэффициента для второго метода, даже если более адекватным является использование случайного эффекта. Тем не менее коэффициенты детерминации можно применять для сравнения моделей, отличающихся набором регрессоров и оцениваемых одним и тем же методом.

13.6. Выбор модели

Фиксированный или случайный эффект?

При работе с реальными панельными данными всегда возникает проблема, какую модель (обычная регрессия, фиксированный или случайный эффект) следует выбрать. На содержательном уровне разницу между моделями можно интерпретировать следующим образом. Обычная модель предполагает, что у экономических единиц нет индивидуальных различий, и в некоторых простых ситуациях такое предположение оправданно. В модели с фиксированным эффектом считается, что каждая экономическая единица «уникальна» и не может рассматриваться как результат случайного выбора из некоторой генеральной совокупности. Такой подход вполне справедлив, когда речь идет о странах, крупных регионах, отраслях промышленности, больших предприятиях. Если же объекты попали в панель «случайно» в результате выборки из большой совокупности, то приемлемой является модель со случайным эффектом. Примером могут служить небольшие фирмы, домашние хозяйства, индивидуумы. Следует, однако, подчеркнуть, что и в подобных ситуациях (особенно для небольшого числа экономических единиц) может возникнуть вопрос о наличии индивидуальных различий, и тогда модель с фиксированным эффектом представляется более предпочтительной.

Сделаем еще одно важное замечание. Модель со случайным эффектом предполагает, что ошибки α_i некоррелированы с регрессорами x_{jt}, т. е. индивидуальный эффект не связан с объясняющими переменными x_{jt}. Это условие выполняется далеко не всегда, даже для выборок из большой совокупности. Так, в

приведенном выше примере оценивания производственной функции (уравнение (13.5)) индивидуальный эффект, связанный с качеством управления, коррелирует с производственными факторами x_{jt}: при прочих равных условиях на предприятии с более высоким качеством управления производственные издержки ниже. В некоторых учебниках по панельным данным именно наличие или отсутствие корреляции между индивидуальным эффектом и регрессорами x_{jt} рассматривается как ключевое различие между моделями с фиксированным и случайным эффектами (см., например, (Johnston and DiNardo, 1997), (Hsiao, 1986)).

Заметим, что независимо от того, коррелированы индивидуальные эффекты с другими объясняющими переменными или нет, оценки с фиксированным эффектом являются несмещенными и состоятельными. Поэтому в любой ситуации модель с фиксированным эффектом дает приемлемые оценки. Однако при отсутствии корреляции эти оценки будут неэффективными, что может быть весьма важным при выборках небольшого объема. В то же время, если в модели присутствует фиксированный эффект, то оценки с помощью случайного эффекта будут несостоятельными. Таким образом, качество оценок существенно зависит от правильной спецификации модели.

Подробнее о выборе моделей с фиксированным или случайным эффектом можно прочесть в книге (Searle, Casella, McCulloch, 1992).

На практике оценивание с помощью разных моделей часто приводит к существенно разным результатам.

Статистические тесты

Помимо содержательных соображений существуют статистические тесты, позволяющие частично решать проблему выбора модели с помощью стандартной техники проверки гипотез. В рассмотренных выше моделях существует определенная иерархия:

1) обычная модель регрессии (13.2) есть частный случай модели с фиксированным эффектом (13.3), когда в последней $\alpha_i = 0$, $i = 1, \ldots, n$;

2) обычная модель регрессии (13.2) есть частный случай модели со случайным эффектом (13.14), когда в последней отсутствуют ошибки u_i или (что эквивалентно) когда $\sigma_u^2 = 0$;

3) модель со случайным эффектом есть частный случай модели с фиксированным эффектом, когда в последней отсутствует корреляция между α_i и x_{jt}.

Поэтому при использовании статистических тестов нулевой гипотезой является, говоря нестрого, более частная модель, а альтернативой — более общая.

1. *Обычная модель против модели с фиксированным эффектом.* Тестирование может быть осуществлено с помощью обычного F-теста, проверяющего гипотезу $\mathrm{H}_0 : \alpha_1 = \cdots = \alpha_n$ в модели с фиктивными переменными (13.7).

В рассматриваемом примере об оценке производственной функции для предприятий топливно-энергетической отрасли результаты этого теста приведены сразу после оценок параметров в модели с фиксированным эффектом: $F(2402, 6013) = 17.62$, $Prob > F = 0.0000$. Иными словами, предположение об отсутствии индивидуальных фиксированных эффектов уверенно отвергается на любом разумном уровне значимости.

2. *Обычная модель против модели со случайным эффектом.* В этом случае требуется в модели со случайным эффектом (13.14) тестировать гипотезу $\mathrm{H}_0 : \sigma_u^2 = 0$. Бреуш и Паган (Breusch and Pagan, 1980) предложили тест множителей Лагранжа, основанный на следующей статистике:

$$\mathrm{LM} = \frac{nT}{2(T-1)} \left(\frac{\sum_{i=1}^{n} \left(\sum_{t=1}^{T} e_{it} \right)^2}{\sum_{i=1}^{n} \sum_{t=1}^{T} e_{it}^2} - 1 \right)^2,$$

где e_{it} — остатки в обычной регрессии (13.2). При гипотезе H_0 величина LM имеет хи-квадрат распределение с одной степенью свободы. Используя матрицу $\boldsymbol{D}$ фиктивных переменных, введенную в разделе 13.3, можно более компактно представить статистику LM:

$$\mathrm{LM} = \frac{nT}{2(T-1)} \left(\frac{\boldsymbol{e}'\boldsymbol{D}\boldsymbol{D}'\boldsymbol{e}}{\boldsymbol{e}'\boldsymbol{e}} - 1 \right)^2.$$

Как обычно, если $\mathrm{LM} > \chi^2_\alpha(1)$, то гипотеза H_0 отвергается при уровне значимости α, где $\chi^2_\alpha(1)$ — α-процентная точка распределения хи-квадрат с одной степенью свободы.

В нашем примере результаты этого теста выглядят так:

Breusch and Pagan LM-test for random effects: H_0: $\mathrm{Var}(u) = 0$

$\chi^2(1)$	6504.28	Probability	0.0000

Вновь видим, что модель простой объединенной регрессии в данном случае уверенно отвергается в пользу модели со случайным эффектом.

3. *Случайный эффект против фиксированного эффекта.* В модели со случайным эффектом предполагается, что индивидуальные эффекты не коррелируют с остальными объясняющими переменными. Таким образом, необходимо проверить гипотезу H_0: $\mathrm{Cov}(\alpha_i, \boldsymbol{x}_{jt}) = 0$. Альтернативная гипотеза состоит в том, что эта ковариация отлична от нуля.

Для проверки подобных гипотез обычно используется тест Хаусмана (Hausman, 1978), о котором уже шла речь в главе 8. Этот тест основан на сравнении оценок параметров $\boldsymbol{\beta}$, полученных в основной и альтернативной моделях. Как уже говорилось выше, при нулевой гипотезе оценка со случайным эффектом $\widehat{\boldsymbol{\beta}}_{\mathrm{RE}}$ состоятельна и эффективна, а при альтернативной гипотезе не состоятельна. Оценка с фиксированным эффектом $\widehat{\boldsymbol{\beta}}_{\mathrm{RE}}$ состоятельна как при нулевой, так и при альтернативной гипотезах. Содержательный смысл теста Хаусмана состоит в том, что при нулевой гипотезе оценки $\widehat{\boldsymbol{\beta}}_{\mathrm{RE}}$ и $\widehat{\boldsymbol{\beta}}_{\mathrm{FE}}$ не должны сильно отличаться, а если справедлива альтернативная гипотеза, то различие должно быть существенным. Чтобы понять, велика ли разница $\widehat{\boldsymbol{\beta}}_{\mathrm{FE}} - \widehat{\boldsymbol{\beta}}_{\mathrm{RE}}$ между оценками, требуется знание ковариационной матрицы $\mathrm{V}(\widehat{\boldsymbol{\beta}}_{\mathrm{FE}} - \widehat{\boldsymbol{\beta}}_{\mathrm{RE}})$ этой разности. Можно показать, что при выполнении нулевой гипотезы из эффективности оценки $\widehat{\boldsymbol{\beta}}_{\mathrm{RE}}$ следует (асимптотическое) равенство

$$\mathrm{V}(\widehat{\boldsymbol{\beta}}_{\mathrm{FE}} - \widehat{\boldsymbol{\beta}}_{\mathrm{RE}}) = \mathrm{V}(\widehat{\boldsymbol{\beta}}_{\mathrm{FE}}) - \mathrm{V}(\widehat{\boldsymbol{\beta}}_{\mathrm{RE}}).$$

Таким образом, статистика

$$\xi_H = (\widehat{\boldsymbol{\beta}}_{\mathrm{FE}} - \widehat{\boldsymbol{\beta}}_{\mathrm{RE}})'(\widehat{\mathrm{V}}(\widehat{\boldsymbol{\beta}}_{\mathrm{FE}}) - \widehat{\mathrm{V}}(\widehat{\boldsymbol{\beta}}_{\mathrm{RE}}))^{-1}(\widehat{\boldsymbol{\beta}}_{\mathrm{FE}} - \widehat{\boldsymbol{\beta}}_{\mathrm{RE}}) \qquad (13.32)$$

при нулевой гипотезе имеет асимптотически хи-квадрат распределение с k степенями свободы, где $\widehat{\mathrm{V}}(\widehat{\boldsymbol{\beta}}_{\mathrm{FE}})$, $\widehat{\mathrm{V}}(\widehat{\boldsymbol{\beta}}_{\mathrm{RE}})$ — оценки соответствующих ковариационных матриц, а k — размерность вектора $\boldsymbol{\beta}$ (ср. с разделом 8.4).

Применяя тест Хаусмана в нашем примере оценки производственной функции, получаем следующие результаты:

Hausman specification test:

	Coefficients		
$\ln Q$	*Fixed Effects*	*Random Effects*	*Difference*
$\ln K$	0.11421	0.24671	−0.13250
$\ln L$	0.60393	0.77544	−0.17151
Test: difference in coefficients not systematic			
$\chi^2(2)$	141.01	*Probability*	0.0000

Таким образом, тест Хаусмана отвергает гипотезу о случайных эффектах в пользу гипотезы о наличии фиксированных индивидуальных эффектов.

При практической реализации теста Хаусмана может возникнуть трудность, связанная с тем, что матрица $\widehat{\mathrm{V}}(\widehat{\boldsymbol{\beta}}_{\mathrm{FE}}) - \widehat{\mathrm{V}}(\widehat{\boldsymbol{\beta}}_{\mathrm{RE}})$ не обязательно является положительно определенной, что может привести к проблемам при ее обращении. В подобной ситуации можно пытаться применять тест Хаусмана лишь к части компонент вектора $\boldsymbol{\beta}$.

В современных эконометрических компьютерных пакетах, как правило, реализованы процедуры работы с панельными данными и оценивание моделей по объединенным данным, с фиксированным или случайным эффектом. Отметим, в частности, уже упоминавшийся выше пакет STATA, где также можно легко осуществлять тест Хаусмана.

13.7. Динамические модели

В предыдущих разделах мы достаточно подробно рассмотрели статические модели с панельными данными. Следующий естественный шаг — это изучение динамических моделей, основанных

на панельных данных, которые, образно говоря, позволяют исследовать динамику «на индивидуальном уровне». Как будет показано ниже, оценвание этих моделей является более сложной задачей по сравнению со статическими моделями, поэтому мы ограничимся в этой книге постановкой задачи и кратким описанием способов ее решения.

Простейшая модель авторегрессии с панельными данными

Наиболее простую динамическую модель, которая является естественным обобщением рассмотренных выше статических моделей и учитывает индивидуальный эффект, можно получить, добавляя в правую часть уравнения (13.3) лагированное значение зависимой переменной:

$$y_{it} = \alpha_i + x'_{it}\beta + \gamma y_{it-1} + \varepsilon_{it}. \tag{13.33}$$

Формально эта модель выглядит почти так же, как и рассмотренные выше статические модели. Поэтому на первый взгляд кажется, что можно использовать те же подходы, что и раньше, и пытаться оценивать эту модель, делая выбор между фиксированным и случайным эффектами. Однако в данном случае очевидно, что переменные y_{it-1} и α_i коррелированы независимо от природы индивидуального эффекта α_i. Чтобы продемонстрировать возникающие здесь новые проблемы, достаточно рассмотреть простейшую авторегрессионную модель без экзогенных переменных x_{it}:

$$y_{it} = \alpha_i + \gamma y_{it-1} + \varepsilon_{it}, \quad |\gamma| < 1. \tag{13.34}$$

Применяя к (13.34) внутригрупповое преобразование, получаем следующее уравнение:

$$y_{it} - \overline{y}_i = \gamma(y_{it-1} - \overline{y}_{i,-1}) + \varepsilon_{it} - \overline{\varepsilon}_i, \tag{13.35}$$

где, как и раньше, $\overline{y}_i, \overline{y}_{i,-1}$ — соответствующие внутригрупповые средние: $\overline{y}_i = (1/T)\sum_{t=1}^T y_{it}$, $\overline{y}_{i,-1} = (1/T)\sum_{t=1}^T y_{it-1}$. Тогда нетрудно заметить, что в уравнении (13.35) регрессоры $y_{it-1} - \overline{y}_{i,-1}$ и ошибки $\varepsilon_{it} - \overline{\varepsilon}_i$ коррелированы, что, как известно, может привести к смещенности и несостоятельности МНК-оценок. Оказывается, в данном случае МНК-оценка параметра γ (оценка с фиксированным эффектом) является несостоятельной при $n \to \infty$ и при фиксированном T: можно показать (Hsiao, 1986; Verbeek, 2000), что

$$\operatorname*{plim}_{n\to\infty} \frac{1}{nT}\widehat{\gamma}_{\mathrm{FE}} = \gamma - \frac{\sigma_\varepsilon^2}{T^2} \cdot \frac{(T-1) - T\gamma + \gamma^T}{(1-\gamma)^2}. \tag{13.36}$$

Таким образом, в данном случае внутригрупповое преобразование в отличие от статических моделей не позволяет получить состоятельную оценку параметра γ.

Как известно (см. гл. 8), при наличии корреляции между ошибками и объясняющими переменными состоятельные оценки параметров в уравнении регрессии можно получить с помощью метода инструментальных переменных. Одна из возможных его реализаций в данном случае выглядит так. Перейдем в уравнении (13.35) к первым разностям:

$$y_{it} - y_{it-1} = \gamma(y_{it-1} - y_{it-2}) + (\varepsilon_{it} - \varepsilon_{it-1}), \tag{13.37}$$

исключая тем самым индивидуальный эффект α_i. В этом уравнении регрессоры и ошибки коррелированы, поскольку коррелированными, очевидно, являются y_{it-1} и ε_{it-1}. В качестве инструментальных переменных можно взять, например, y_{it-2}: эта величина, очевидно, коррелирована с $y_{it-1} - y_{it-2}$ и не коррелирована с $\varepsilon_{it} - \varepsilon_{it-1}$:

$$\mathrm{E}(y_{it-2}(\varepsilon_{it} - \varepsilon_{it-1})) = 0. \tag{13.38}$$

Соответствующая оценка равна

$$\widehat{\gamma}_{\mathrm{IV}} = \frac{\sum_{i=1}^n \sum_{t=2}^T y_{it-2}(y_{it} - y_{it-1})}{\sum_{i=1}^n \sum_{t=2}^T y_{it-2}(y_{it-1} - y_{it-2})}. \tag{13.39}$$

Эта оценка (наряду с другими) была предложена в работе (Andersen and Hsiao, 1981). В этом примере есть и другие инструментальные переменные, например, $y_{it-2} - y_{it-3}$. И вообще можно

предложить довольно много комбинаций лагированных значений зависимой переменной y, которые подходят на роль инструментов.

В связи с этим возникают два естественных вопроса: что же лучше выбрать в качестве инструментов и нельзя ли использовать в каком-то смысле все доступные инструментальные переменные? Ответ заключается в применении *обобщенного метода моментов* (*GMM*), краткое описание которого приводится далее в разделе 13.9.

Основой этого метода являются моментные тождества, или условия ортогональности. Особенность описываемого здесь подхода, который был предложен в работе (Arellano and Bond, 1991), состоит в том, что число моментных тождеств меняется с изменением t. Наше изложение следует, в основном, книге (Verbeek, 2000).

В модели (13.37) при $t = 2$ выполняется моментное тождество

$$\mathrm{E}((\varepsilon_{i2} - \varepsilon_{i1})y_{i0}) = 0.$$

При $t = 3$ выполняются уже два тождества:

$$\mathrm{E}((\varepsilon_{i3} - \varepsilon_{i2})y_{i0}) = 0, \quad \mathrm{E}((\varepsilon_{i3} - \varepsilon_{i2})y_{i1}) = 0.$$

В общем случае для каждого t можно написать $(t-1)$ равенство

$$\begin{gathered}\mathrm{E}((\varepsilon_{it} - \varepsilon_{it-1})y_{i0}) = 0, \quad \mathrm{E}((\varepsilon_{it} - \varepsilon_{it-1})y_{i1}) = 0, \ldots, \\ \mathrm{E}((\varepsilon_{it} - \varepsilon_{it-1})y_{it-2}) = 0.\end{gathered} \tag{13.40}$$

Таким образом, имеется всего $1 + 2 + \cdots + (T-1) = T(T-1)/2$ моментных тождеств. Чтобы более компактно записать соотношения (13.40), введем $(T-1) \times 1$ вектор

$$\Delta\boldsymbol{\varepsilon}_i = \begin{bmatrix} \varepsilon_{i2} - \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iT} - \varepsilon_{iT-1} \end{bmatrix}$$

и $(T-1) \times T(T-1)/2$ матрицу

$$\boldsymbol{Z}_i = \begin{bmatrix} [y_{i0}] & 0 & \ldots & 0 \\ 0 & [y_{i0}, y_{i1}] & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & [y_{i0}, \ldots, y_{iT-2}] \end{bmatrix}$$

Тогда тождества (13.40) равносильны равенству

$$\mathrm{E}(\boldsymbol{Z}_i' \, \Delta\boldsymbol{\varepsilon}_i) = \boldsymbol{0},$$

которое можно переписать в следующем виде:

$$\mathrm{E}\big(\boldsymbol{Z}_i'(\Delta\boldsymbol{y}_i - \gamma\Delta\boldsymbol{y}_i(-1))\big) = \boldsymbol{0} \tag{13.41}$$

(мы воспользовались стандартным обозначением: если $\boldsymbol{a} = (a_0, a_1, \ldots, a_r)'$, то $\boldsymbol{a}(-1) = (b_1, b_2, \ldots, b_r)'$, где $b_t = a_{t-1}$). В соответствии с обобщенным методом моментов оценка параметра γ строится путем решения следующей задачи (см. (13.55)):

$$\begin{aligned}&\left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{Z}_i'(\Delta\boldsymbol{y}_i - \gamma\Delta\boldsymbol{y}_i(-1))\right)' \boldsymbol{S} \\ &\qquad \times \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{Z}_i'(\Delta\boldsymbol{y}_i - \gamma\Delta\boldsymbol{y}_i(-1))\right) \to \min,\end{aligned} \tag{13.42}$$

где $\boldsymbol{S}$ — некоторая весовая матрица. Левая часть (13.42) — квадратичная функция γ, поэтому ее минимум легко находится:

$$\begin{aligned}\widehat{\gamma}_{\text{СММ}} = &\left(\left(\sum_{i=1}^{n} \Delta\boldsymbol{y}_i'(-1)\boldsymbol{Z}_i\right)\boldsymbol{S}\left(\sum_{i=1}^{n} \boldsymbol{Z}_i'\Delta\boldsymbol{y}_i(-1)\right)\right)^{-1} \\ &\times \left(\left(\sum_{i=1}^{n} \Delta\boldsymbol{y}_i'(-1)\boldsymbol{Z}_i\right)\boldsymbol{S}\left(\sum_{i=1}^{n} \boldsymbol{Z}_i'\Delta\boldsymbol{y}_i\right)\right).\end{aligned} \tag{13.43}$$

В соответствии с теорией обобщенного метода моментов для получения асимптотически оптимальной оценки (т. е. с минимальной дисперсией) в качестве весовой матрицы следует взять матрицу (см. (13.56)):

$$S^{opt} = \left(\mathrm{E}(Z_i'\Delta\varepsilon_i'\Delta\varepsilon_i Z_i)\right)^{-1}. \tag{13.44}$$

Опять же в соответствии с теорией обобщенного метода моментов оценкой этой матрицы является матрица

$$\widehat{S}^{opt} = \left(\frac{1}{n}\sum_{i=1}^{n} Z_i'\Delta\widehat{\varepsilon}_i\Delta\widehat{\varepsilon}_i' Z_i\right)^{-1}, \tag{13.45}$$

где $\Delta\widehat{\varepsilon}_i$ — остатки, построенные в модели (13.37) при использовании какой-либо состоятельной оценки параметра γ, например, оценки, полученной решением задачи (13.42) с единичной весовой матрицей $S = I$.

Заметим, что этот способ оценивания явно не накладывает ограничений на ошибки ε_{it}. Однако, чтобы гарантировать адекватность использования обобщенного метода моментов в таком виде, требуется, чтобы ошибки были некоррелированы и имели одинаковую дисперсию σ_ε^2. Нетрудно проверить, что при этих ограничениях

$$\mathrm{E}(\Delta\varepsilon_i\Delta\varepsilon_i') = \sigma_\varepsilon^2 W = \sigma_\varepsilon^2 \begin{bmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & -1 & 2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 2 \end{bmatrix}.$$

Поэтому можно взять оптимальную весовую матрицу в виде

$$\widehat{S}^{opt} = \left(\frac{1}{n}\sum_{i=1}^{n} Z_i' W Z_i\right)^{-1}.$$

Заметим, что эта матрица не зависит от γ, и, следовательно, оценка обобщенного метода моментов может быть получена за один шаг без начального оценивания неизвестного параметра.

Модель с экзогенными переменными

Вернемся теперь к более общей динамической модели (13.33)

$$y_{it} = \alpha_i + x_{it}'\beta + \gamma y_{it-1} + \varepsilon_{it},$$

содержащей экзогенные переменные x_{it}. Переходя к первым разностям

$$y_{it} - y_{it-1} = (x_{it} - x_{it-1})'\beta + \gamma(y_{it-1} - y_{it-2}) + (\varepsilon_{it} - \varepsilon_{it-1}), \tag{13.46}$$

видим, что получается модель, аналогичная (13.37). Экзогенность регрессоров x_{it} означает, что

$$\mathrm{E}(x_{is}\Delta\varepsilon_{it}) = 0$$

при всех s, t. Эти равенства можно рассматривать как моментные тождества, которые аналогичны тождествам (13.40), и использовать в обобщенном методе моментов. Иными словами, в каждый момент времени t переменные $x_{i1}, \ldots, x_{iT}$ можно использовать в качестве инструментов в дополнение к инструментальным переменным, построенным в предыдущем разделе.

В динамических моделях с панельными данными так же, как и в обычных моделях временных рядов, возникают проблемы единичных корней, коинтеграции и т. п., но рассмотрение этих вопросов выходит за рамки нашей книги. Более подробное описание этой темы можно прочесть, например, в книгах (Greene, 1997), (Verbeek, 2000).

Подведем кратко итоги.

1. В динамических моделях с панельными данными регрессоры коррелированы с индивидуальными эффектами независимо от того, являются ли эти эффекты фиксированными или случайными.

2. Внутригрупповая регрессия в отличие от статических моделей не позволяет получить состоятельные (при фиксированном T и при $n \to \infty$) оценки параметров.

3. Состоятельные оценки в динамических моделях можно построить с помощью обобщенного метода моментов.

13.8. Модели бинарного выбора с панельными данными

В этом разделе мы кратко рассмотрим модели с панельными данными, в которых зависимая переменная является бинарной, т. е. принимающей значения 0 или 1. Модель бинарного выбора в случае панельных может быть описана аналогично тому, как это делается для пространственных данных (см. (12.4) и (12.5)):

$$y_{it}^* = x_{it}\beta + \alpha_i + \varepsilon_{it}, \tag{13.47}$$

и

$$\begin{aligned} y_{it} &= 1, \quad \text{если } y_{it}^* \geqslant 0, \\ y_{it} &= 0, \quad \text{если } y_{it}^* < 0, \end{aligned} \tag{13.48}$$

где ошибки ε_{it} независимы по i, t и одинаково распределены, а величины α_i, как и раньше, отражают индивидуальные различия между объектами. Будем считать α_i неизвестными параметрами (модель с фиксированным эффектом). Тогда можно точно так же, как и для обычных моделей, рассмотренных в главе 12, построить оценки максимального правдоподобия параметров α_i, $i = 1, \ldots, n$, β (оставляем в качестве упражнения написать выражение для функции правдоподобия). Эти оценки будут состоятельными лишь при фиксированном n и при $T \to \infty$. Этот факт, конечно же, связан с тем, что с увеличением n растет число оцениваемых параметров α_i, $i = 1, \ldots, n$, поэтому если $n \to \infty$, а T фиксировано, то оценки максимального правдоподобия как параметров α_i, $i = 1, \ldots, n$, так и параметров β будут несостоятельными. Эта проблема возникает и в линейных моделях, но там с помощью внутригруппового преобразования удается исключить α_i и получить состоятельные оценки вектора параметров β. В моделях бинарного выбора формально тоже можно сделать внутригрупповое преобразование (или взять первые разности) для латентной переменной y_{it}^*, что позволит исключить α_i в уравнении (13.47). Однако не существует простого способа, позволяющего связать наблюдаемую бинарную переменную y_{it} с преобразованной латентной переменной $y_{it}^* - \overline{y}_i^*$ (или $y_{it}^* - y_{it-1}^*$). Одно из возможных решений этой проблемы состоит

в максимизации не исходной функции правдоподобия, а некоторой специально подобранной условной функции правдоподобия (Andersen, 1970), (Chamberlain, 1980). Пусть $f(y_{i1}, \ldots, y_{iT} \mid \alpha_i, \beta)$ — совместное распределение величин $y_{i1}, \ldots, y_{iT}$, зависящее от параметров α_i, β. Предположим, что существует такая статистика s_i (т. е. функция, зависящая только от наблюдений), что $f(y_{i1}, \ldots, y_{iT} \mid \alpha_i, \beta) = f(y_{i1}, \ldots, y_{iT} \mid s_i, \beta)$. Тогда, максимизируя условную функцию правдоподобия

$$L^c = \prod_{i=1}^{n} f(y_{i1}, \ldots, y_{iT} \mid s_i, \beta),$$

можно получить состоятельные оценки параметров β. Более того, эти оценки обладают практически теми же свойствами, что и обычные оценки максимального правдоподобия.

Проблема, однако, состоит в том, что далеко не всегда такая статистика s_i существует. Так, например, можно показать, что для *probit*-модели (когда ошибки ε_{it} имеют нормальное распределение) такой статистики построить нельзя. Но если ошибки в (13.47) имеют логистическое распределение (*logit*-модель), такая статистика существует, а именно, $s_i = \sum_{t=1}^{T} y_{it}$. Мы не будем доказывать это в общем случае, а ограничимся иллюстрацией для простейшего случая $T = 2$.

Заметим вначале, что

$$\mathrm{P}(y_{i1} = 0, y_{i2} = 0 \mid s_i = 0) = 1, \mathrm{P}(y_{i1} = 1, y_{i2} = 1 \mid s_i = 2) = 1.$$

Это означает, что объекты, у которых бинарная переменная не менялась в течение всего периода времени, не дают «вклада» в условную функцию правдоподобия. Далее, обозначим для краткости $(y_{i1} = 0, y_{i2} = 1) = (0, 1)$, $(y_{i1} = 1, y_{i2} = 0) = (1, 0)$. Тогда

$$\mathrm{P}((0,1) \mid s_i = 1) = \frac{\mathrm{P}((0,1), s_i = 1)}{\mathrm{P}(s_i = 1)} = \frac{\mathrm{P}((0,1))}{\mathrm{P}((0,1)) + \mathrm{P}((1,0))}.$$

В силу независимости наблюдений имеем

$$\mathrm{P}((0,1)) = \frac{1}{1 + e^{\alpha_i + x_{i1}'\beta}} \cdot \frac{e^{\alpha_i + x_{i2}'\beta}}{1 + e^{\alpha_i + x_{i2}'\beta}},$$

$$\mathrm{P}((1,0)) = \frac{e^{\alpha_i + x_{i1}'\beta}}{1 + e^{\alpha_i + x_i'\beta}} \cdot \frac{1}{1 + e^{\alpha_i + x_{i2}'\beta}}.$$

Следовательно,

$$\mathrm{P}((0,1) \mid s_i = 1) = \frac{e^{x_{i2}'\beta}}{e^{x_{i1}'\beta} + e^{x_{i2}'\beta}},$$
$$\mathrm{P}((1,0) \mid s_i = 1) = \frac{e^{x_{i1}'\beta}}{e^{x_{i1}'\beta} + e^{x_{i2}'\beta}}.$$

Мы видим, что эти условные вероятности не зависят от α_i.

В случае произвольного T доказательство независимости условных распределений от α_i проводится аналогично, но требует более громоздких вычислений (см. (Chamberlain, 1980), (Maddala, 1987)). Таким образом, для *logit*-модели с фиксированным эффектом удается устранить переменные α_i и получить состоятельные оценки параметров β.

Рассмотрим теперь модели со случайным эффектом. Если в уравнении (13.47) обозначить $u_{it} = \alpha_i + \varepsilon_{it}$, то внешне модель (13.47), (13.48) будет выглядеть так же, как модель бинарного выбора для пространственных данных (12.4), (12.5). Однако есть существенное отличие: в данном случае ошибки u_{it}, $t = 1, \ldots, T$, а следовательно, и наблюдения y_{it}, $t = 1, \ldots, T$, уже не являются независимыми по t для каждого объекта i (между объектами эти ошибки, конечно же, независимы). Это означает, что распределение $f(y_{i1}, \ldots, y_{iT} \mid x_{i1}, \ldots, x_{iT}, \beta)$ не распадается в произведение одномерных распределений, а следовательно, и функция правдоподобия для этой модели не представима в виде произведения одномерных распределений, как это было для моделей бинарного выбора с пространственными данными. В общем случае построение функции правдоподобия требует вычисления многомерных интегралов, что делает практически нереализуемым метод максимального правдоподобия.

Однако можно заметить, что поскольку ошибки $\varepsilon_{i1}, \ldots, \varepsilon_{iT}$ независимы как по i, так и по t, то наблюдения y_{it}, $t = 1, \ldots, T$

условно независимы *при фиксированном* α_i. Поэтому

$$\begin{aligned} f(y_{i1}, \ldots, y_{iT} \mid x_{i1}, \ldots, x_{iT}, \beta) &= \int_{-\infty}^{\infty} f(y_{i1}, \ldots, y_{iT} \mid x_{i1}, \ldots, x_{iT}, \beta, \alpha_i) f(\alpha_i) d\alpha_i \\ &= \int_{-\infty}^{\infty} \left(\prod_{t=1}^{T} f(y_{it} \mid x_{it}, \beta, \alpha_i) \right) f(\alpha_i) d\alpha_i. \end{aligned} \tag{13.49}$$

Эта формула позволяет эффективно вычислять функцию правдоподобия и строить оценки максимального правдоподобия параметров β. Заметим, что этот метод может быть реализован для произвольных распределений ошибок α_i, ε_{it}. Но на практике обычно считают, что эти ошибки имеют нормальное распределение, т. е. рассматривают *probit*-модель со случайным эффектом.

Итак, пусть $\alpha_i \sim N(0, \sigma_\alpha^2)$, $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$, случайные величины ε_{it} независимы по t и α_i и ε_{it} также независимы. Будем предполагать, что выполнено условие нормировки $\mathrm{V}(u_{it}) = \mathrm{V}(\alpha_i + \varepsilon_{it}) = 1$, следовательно, $\sigma_\varepsilon^2 = 1 - \sigma_\alpha^2$. Тогда нетрудно проверить, что в формуле (13.49) одномерные условные распределения имеют следующий вид:

$$\begin{aligned} f(y_{it} = 1 \mid x_{it}, \beta, \alpha_i) &= \Phi\left(\frac{x_{it}'\beta + \alpha_i}{\sqrt{1 - \sigma_\alpha^2}} \right), \\ f(y_{it} = 0 \mid x_{it}, \beta, \alpha_i) &= 1 - \Phi\left(\frac{x_{it}'\beta + \alpha_i}{\sqrt{1 - \sigma_\alpha^2}} \right). \end{aligned} \tag{13.50}$$

Наконец, $f(\alpha_i)$ в (13.49) — это обычная плотность нормального распределения:

$$f(\alpha_i) = \frac{1}{\sqrt{2\pi}\sigma_\alpha} \exp\left(-\frac{\alpha_i^2}{2\sigma_\alpha^2} \right).$$

13.9. Обобщенный метод моментов

В этом разделе кратко описывается обобщенный метод моментов, который в настоящее время является одним из наиболее распро-

страненных методов оценивания. Этот метод является достаточно общим и применяется не только в эконометрике, однако наше изложение ориентировано в первую очередь на его применение к моделям регрессии.

Исходным пунктом обобщенного метода моментов являются некоторые теоретические соотношения между переменными и параметрами модели. Идея метода заключается в том, что надо так выбирать параметры, чтобы, говоря нестрого, эти соотношения для заданных наблюдений выполнялись как можно «более точно».

Предположим, что модель включает переменные y_i, $\boldsymbol{x}_i$, $\boldsymbol{z}_i$, $i = 1, \ldots, n$, и пусть выполнены следующие равенства:

$$\mathrm{E}(m_j(y_i, \boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{\theta})) = 0, \quad j = 1, \ldots, l, \tag{13.51}$$

где $m_j(y_i, \boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{\theta})$ — некоторые известные скалярные функции, а $\boldsymbol{\theta}$ — k-мерный вектор параметров. (В применении к моделям регрессии можно считать y_i зависимой переменной, $\boldsymbol{x}_i$ — набором регрессоров, $\boldsymbol{z}_i$ — инструментальными переменными.)

Равенства (13.51) называют моментными тождествами или условиями ортогональности. Если ввести вектор-функцию

$$\boldsymbol{m}(y_i, \boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{\theta}) = (m_1(y_i, \boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{\theta}), \ldots, m_l(y_i, \boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{\theta}))',$$

то соотношения (13.51) можно записать в векторном виде

$$\mathrm{E}(\boldsymbol{m}(y_i, \boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{\theta})) = \boldsymbol{0}. \tag{13.52}$$

Определим вектор-функцию

$$\boldsymbol{g}(\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{m}(y_i, \boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{\theta})$$

и запишем выборочный аналог равенства (13.52):

$$\boldsymbol{g}(\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}) = \boldsymbol{0}. \tag{13.53}$$

Обозначим для краткости $\boldsymbol{g}(\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\theta}) = \boldsymbol{g}(\boldsymbol{\theta})$. Если $k > l$ (число уравнений в (13.53) меньше числа оцениваемых параметров), то модель не идентифицируема. Если размерности $\boldsymbol{g}(\boldsymbol{\theta})$ и $\boldsymbol{\theta}$ совпадают (т. е. число уравнений в (13.53) равно числу параметров), то

(при выполнении условий регулярности) можно разрешить систему уравнений (13.53) относительно $\boldsymbol{\theta}$. Полученные таким образом оценки вектора параметров $\boldsymbol{\theta}$ называются оценками метода моментов (см. раздел 7). Если же количество моментных тождеств больше числа оцениваемых параметров ($k < l$), то система (13.53) оказывается переопределенной. Для получения оценок методом моментов можно было бы отбросить часть моментных тождеств, но можно поступить иначе. А именно, построить оценку параметров $\boldsymbol{\theta}$ таким образом, чтобы, говоря нестрого, вектор $\boldsymbol{g}(\boldsymbol{\theta})$ был как можно ближе к нулю. Например, найти оценку $\widehat{\boldsymbol{\theta}}$ путем решения задачи

$$\boldsymbol{g}'(\boldsymbol{\theta})\boldsymbol{g}(\boldsymbol{\theta}) = \sum_{j=1}^{l} g_j^2(\boldsymbol{\theta}) \to \min. \tag{13.54}$$

Можно показать (см. (Hansen, 1982)), что полученная при этом оценка $\widehat{\boldsymbol{\theta}}$ является состоятельной, но, вообще говоря, неэффективной.

Вместо минимизации суммы квадратов компонент вектора $\boldsymbol{g}(\boldsymbol{\theta})$ можно было бы рассматривать более общую задачу, а именно,

$$\boldsymbol{g}'(\boldsymbol{\theta})\boldsymbol{S}\boldsymbol{g}(\boldsymbol{\theta}) \to \min, \tag{13.55}$$

где $\boldsymbol{S}$ — некоторая симметричная положительно определенная матрица (размера $l \times l$). Оценка, полученная решением задачи (13.55), называется оценкой обобщенного метода моментов или кратко GMM-оценкой (*Generalized Method of Moments, GMM*): $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_{\mathrm{GMM}}$.

Ясно, что разным весовым матрицам $\boldsymbol{S}$ соответствуют разные (состоятельные) оценки $\widehat{\boldsymbol{\theta}}_{\mathrm{GMM}}$. Можно показать, что для получения асимптотически оптимальной оценки (т. е. имеющей минимальную асимптотическую матрицу ковариаций) в качестве $\boldsymbol{S}$ надо взять матрицу, обратную матрице ковариаций вектора моментов, которая (при отсутствии корреляции между наблюдениями) выглядит следующим образом:

$$\boldsymbol{S}^{opt} = \left(\mathrm{E}\left(\boldsymbol{m}(y_i, \boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{\theta})\boldsymbol{m}'(y_i, \boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{\theta})\right)\right)^{-1}. \tag{13.56}$$

В общем случае эта матрица зависит от вектора неизвестных параметров $\boldsymbol{\theta}$, поэтому для построения асимптотически оптимальной оценки обычно используют двухшаговые или многошаговые процедуры. Например, на первом этапе находится оценка $\widehat{\boldsymbol{\theta}}_{(0)}$ путем решения задачи (13.54) (т. е. с единичной весовой матрицей). Затем строится состоятельная оценка матрицы $\boldsymbol{S}^{opt}$:

$$\boldsymbol{S}_n^{opt} = \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{m}(y_i, \boldsymbol{x}_i, \boldsymbol{z}_i, \widehat{\boldsymbol{\theta}}_{(0)})\boldsymbol{m}'(y_i, \boldsymbol{x}_i, \boldsymbol{z}_i, \widehat{\boldsymbol{\theta}}_{(0)})\right)^{-1}.$$

Наконец, решается задача (13.55) с $\boldsymbol{S} = \boldsymbol{S}_n^{opt}$ и в результате получается оценка $\widehat{\boldsymbol{\theta}}_{\mathrm{GMM}}$. Два последних шага можно повторить несколько раз.

Можно показать, что построенная таким образом оценка $\widehat{\boldsymbol{\theta}}_{\mathrm{GMM}}$ является асимптотически нормальной:

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\mathrm{GMM}} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{V}),$$

где $\boldsymbol{V} = \left(\boldsymbol{D}\boldsymbol{S}^{opt}\boldsymbol{D}'\right)^{-1}$, и

$$\boldsymbol{D} = \mathrm{E}\left(\frac{\partial \boldsymbol{m}(y_i, \boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right).$$

Обобщенный метод моментов обладает рядом преимуществ по сравнению с другими методами оценивания: для его использования не требуется знать распределение наблюдений, он работает при наличии гетероскедастичности любого вида и т. п.

Подробнее об обобщенном методе моментов см., например, (Greene, 1997), (Johnston and DiNardo, 1997), (Verbeek, 2000).

Рассмотрим несколько примеров.

1. *Обычный метод наименьших квадратов.* Рассмотрим классическую модель линейной регрессии

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + \varepsilon_t, \quad t = 1, \ldots, n,$$

в которой ошибки ε_t гомоскедастичны и некоррелированы с регрессорами $\boldsymbol{x}_t$:

$$\mathrm{E}(x_{tj}\varepsilon_t) = 0, \quad j = 1, \ldots, k, t = 1, \ldots, n. \tag{13.57}$$

Равенства (13.57) можно переписать в виде следующих (векторных) моментных тождеств:

$$\mathrm{E}(\boldsymbol{x}_t(y_t - \boldsymbol{x}_t'\boldsymbol{\beta})) = \mathbf{0} \tag{13.58}$$

(см. (13.52)). В данном случае размерности вектора моментов и вектора параметров совпадают. Напишем выборочный аналог равенства (13.58) (мы используем обозначения, введенные в гл. 3):

$$\frac{1}{n}\sum_{t=1}^{n}(\boldsymbol{x}_t y_t - \boldsymbol{x}_t\boldsymbol{x}_t'\boldsymbol{\beta}) = \frac{1}{n}(\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta}) = \mathbf{0}. \tag{13.59}$$

Решая уравнение (13.59) относительно $\boldsymbol{\beta}$, получаем:

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{t=1}^{n}\boldsymbol{x}_t\boldsymbol{x}_t'\right)^{-1}\sum_{t=1}^{n}\boldsymbol{x}_t y_t = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} = \widehat{\boldsymbol{\beta}}_{\mathrm{OLS}},$$

т. е. для классической линейной модели регрессии оценка метода моментов совпадает с оценкой метода наименьших квадратов: $\boldsymbol{\beta}_{\mathrm{OLS}} = \boldsymbol{\beta}_{\mathrm{GMM}}$.

2. *Метод инструментальных переменных.* Пусть $\boldsymbol{z}_t$ — $l \times 1$ вектор инструментальных переменных, $l \geq k$, где k — размерность вектора параметров $\boldsymbol{\beta}$. Моментные тождества (в векторном виде) выглядят аналогично соотношению (13.58):

$$\mathrm{E}(\boldsymbol{z}_t\varepsilon_t) = \mathrm{E}(\boldsymbol{z}_t(y_t - \boldsymbol{x}_t'\boldsymbol{\beta})) = \mathbf{0}.$$

Вектор-функция $\boldsymbol{g}$ равна

$$\boldsymbol{g} = \frac{1}{n}\sum_{t=1}^{n}\boldsymbol{z}_t(y_t - \boldsymbol{x}_t'\boldsymbol{\beta}) = \frac{1}{n}(\boldsymbol{Z}'\boldsymbol{y} - \boldsymbol{Z}'\boldsymbol{X}\boldsymbol{\beta}).$$

В данном случае

$$\mathrm{E}(\boldsymbol{m}\boldsymbol{m}') = \mathrm{E}\left(\boldsymbol{z}_t(y_t - \boldsymbol{x}_t'\boldsymbol{\beta})^2\boldsymbol{z}_t'\right) = \sigma^2\boldsymbol{z}_t\boldsymbol{z}_t',$$

поэтому в качестве весовой матрицы можно взять

$$\boldsymbol{S}^{opt} = \left(\sum_{t=1}^{n}\boldsymbol{z}_t\boldsymbol{z}_t'\right)^{-1} = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}.$$

Решая задачу

$$(Z'y - Z'X\beta)'(Z'Z)^{-1}(Z'y - Z'X\beta) \to \min, \qquad (13.60)$$

получаем

$$\beta_{\text{GMM}} = (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'y, \qquad (13.61)$$

что совпадает с оценкой метода инструментальных переменных (см. формулу (8.5) главы 8): $\beta_{\text{GMM}} = \beta_{\text{IV}}$.

Обобщенный метод моментов реализован в большинстве современных эконометрических пакетов.

Упражнения

13.1. Докажите равенства (13.17) и (13.18).

13.2. Докажите равенство (13.22).

13.3. Покажите, что существует $k \times k$ матрица S, такая что $S \geq 0$, $I_k - S \geq 0$, и

$$\widehat{\beta}_{\text{OLS}} = S\widehat{\beta}_{\text{B}} + (I_k - S)\widehat{\beta}_{\text{W}},$$

где $\widehat{\beta}_{\text{OLS}}$ — МНК-оценка вектора параметров β в объединенной регрессии (13.2).

13.4. Докажите равенство (13.23).

13.5. Докажите равенство (13.25).

13.6. Докажите, что решение задачи (13.42) задается равенством (13.43).

13.7. Постройте функцию правдоподобия для модели (13.47), (13.48).

13.8. Рассмотрим модель

$$y_{it} = \gamma y_{it-1} + \alpha_i + \varepsilon_{it},$$

где ошибки $\varepsilon it \sim iid(0, \sigma^2)$ и независимы с y_{js} при $s < t$. В уравнении внутригрупповой регрессии (13.35)

$$y_{it} - \overline{y}_i = \gamma(y_{it-1} - \overline{y}_{i,-1}) + \varepsilon_{it} - \overline{\varepsilon}_i$$

ошибки коррелированы с регрессорами. Покажите, что МНК-оценка $\widehat{\gamma}$ в последнем уравнении несостоятельна (например, покажите, что при $\gamma = 0$ получаем $p\lim\widehat{\gamma} = -1/2$).

13.9. Оценка производственной функции российских предприятий топливно-энергетического комплекса.

В файле `fuel.xls` содержатся ежегодные данные об объемах выпуска, трудозатратах, капитальных вложениях российских предприятий топливно-энергетического комплекса за период 1993–2000 гг. (Е. В. Бессонова, ЦЭФИР). В панель включено около 2400 предприятий, панель не сбалансирована.

Цель примера — оценить производственную функцию предприятий.

Таблица 13.1

Переменная	Описание
OKPO	Номер предприятия по классификации ОКПО
OKONH	Код отрасли ОКОНХ
YEAR	Год
ROUT	Реальный выпуск
EMP	Численность работников
WOR	Промышленно-производственный персонал
RK	Реальные капиталовложения

13.9.1. Вычислите описательные статистики основных переменных.

13.9.2. Оцените производственную функцию Кобба-Дугласа с помощью простой полной регрессии. Выполняется ли условие постоянства отдачи на масштаб?

13.9.3. Повторите упражнение 13.9.2 для регрессий с фиксированным и случайным эффектами. Сравните результаты.

13.9.4. Является ли влияние индивидуальных эффектов существенным? Проверьте гипотезы:

- простая регрессия против регрессии с фиксированным эффектом;
- простая регрессия против регрессии со случайным эффектом;
- регрессия со случайным эффектом против регрессии с фиксированным эффектом.

13.9.5. Повторите предыдущие упражнения для более сложной модели производственной функции путем включения квадратичных и перекрестных членов. Выберите наиболее адекватную, с вашей точки зрения, модель.

13.10. (Tammo Bijmolt, Erwin Charlier)

В этом упражнении модели панельных данных используются для анализа продаж некоторого сорта тунца (обозначенного как A) в нескольких магазинах. Использовались данные о продажах консервированного тунца в 24 магазинах Чикаго в течение 104 недель. Данные для этого примера (описание переменных приведено в таблице 13.2) находятся в файле `brand_a.xls`[2].

Таблица 13.2

Переменная	Описание
$sales_a$	Объем продаж тунца сорта A
$totsales$	Общий объем продаж магазина за весь рассматриваемый период
$regpr_a$	Цена тунца сорта A
$actpr_a$	Цена тунца сорта A с учетом скидки
$feat_a$	Фиктивная переменная (1 — если в магазине была реклама рыбы сорта A, и рыба сорта A не выкладывалась на витрину, 0 — иначе)
$displ_a$	Фиктивная переменная (1 — если в магазине рыба сорта A выкладывалась на витрину, и не было рекламы рыбы сорта A, 0 — иначе)
$ftdpl_a$	Фиктивная переменная (1 — если в магазине рыба сорта A выкладывалась на витрину, и была реклама рыбы сорта A, 0 — иначе)
$regpr_b$	Цена тунца сорта B (B, C, D — конкурирующие сорта)
$actpr_b$	Цена тунца сорта B с учетом скидки
$regpr_c$	Цена тунца сорта C
$actpr_c$	Цена тунца сорта C с учетом скидки
$regpr_d$	Цена тунца сорта D
$actpr_d$	Цена тунца сорта D с учетом скидки

13.10.1. Исследуйте описательные статистики данных. Проверьте (например, с помощью графиков), как связана зависимая переменная с объясняющими.

13.10.2. Оцените простую (pooled) модель зависимости объема продаж от всех остальных переменных. Ввиду того что розничные цены и цены

[2]Оригинальный пример доступен на странице Эрвина Чарлиера по адресу `http://center.uvt.nl/staff/charlier/paneldata.html`

с учетом скидок сильно коррелированы, для каждого сорта i тунца используйте переменные $regpr_i$ и $discount_i = regpr_i - actpr_i$ ($i = a, b, c, d$).

13.10.3. Оцените панельную модель с фиксированными эффектами. Все ли параметры удалось оценить? Если нет, то почему? (В дальнейшем исключите из модели переменную, вызвавшую проблему.)

13.10.4. Приведите оценки стандартных ошибок коэффициентов в модели упражнения 13.10.3, интерпретируйте результаты, сравните с результатами упражнения 13.10.2.

13.10.5. Вычислите межгрупповую (between-group) оценку для модели. Интерпретируйте результаты, сравните их с результатами модели с фиксированными эффектами.

13.10.6. Оцените панельную модель со случайными эффектами. Интерпретируйте результаты и сравните с результатами упражнений 13.10.4 и 13.10.5.

13.10.7. Используя известные вам тесты (тест Хаусмана, LM-тест Бреуша-Пагана), выберите наиболее подходящую модель.

13.10.8. Считая, что издержки продажи $sales_a$ тунца сорта A равны $0.5 \cdot sales_a$, выведите из модели, выбранной в упражнении 13.10.7, оптимальную цену данного сорта рыбы.

Глава 14

Предварительное тестирование: введение*

14.1. Введение

Как правило, в эконометрических исследованиях данные не являются результатом эксперимента, и поэтому мы вынуждены использовать одни и те же данные как для выбора модели, так и для оценки параметров выбранной модели. Конечно, это обстоятельство влияет на свойства полученных оценок. В данной главе нас особенно будет интересовать величина этого эффекта. В конце концов *все*, чем мы занимаемся в эконометрике, неправильно (неверные модели, недостоверные данные, неэффективные и иногда не самые лучшие оценки и т. п.), но это не обязательно является проблемой. Проблема возникает, если эти неточности имеют значительный эффект.

Если параметры модели оцениваются на том же наборе данных, на котором происходил отбор модели, то есть после предварительных тестов, то такие оценки называются *pretest*-оценками[1]

*Данная глава не входит в стандартный набор тем, включаемых обычно в начальный курс эконометрики и может быть опущена при первом чтении.

[1]В русской литературе, насколько нам известно, отсутствует соответствующий термин.

(*оценками, полученными после предварительного тестирования*). Простейшим примером подобной ситуации является стандартная линейная модель

$$y = X\beta + \gamma z + \varepsilon,$$

в которой нас интересуют коэффициенты β, однако мы не уверены, следует ли включать в модель z. Как правило, поступают следующим образом. Вычисляют t-статистику коэффициента γ, и затем, в зависимости от того, является полученное значение $|t|$ «большим» или «малым», выбирают регрессию без ограничения ($\gamma \neq 0$) или с ограничением ($\gamma = 0$). Затем оценивают коэффициенты β по выбранной модели. Полученная оценка $\widehat{\beta}$ и называется *pretest*-оценкой.

Проблема в том, что обычно, описывая свойства полученной оценки, мы действуем так, как будто *не было* предварительного отбора модели. В результате мы (ошибочно) считаем оценку несмещенной и пользуемся неверной оценкой ее дисперсии, поскольку применяемые нами формулы для среднего и дисперсии верны только *условно*, при условии выбранной модели, но не *безусловно*.

Таким образом, нашей задачей является нахождение безусловных моментов *pretest*-оценки, принимая во внимание то, что процедуры выбора модели и оценки параметров интегрированы в одну процедуру. Мы не утверждаем, что следует избегать предварительного тестирования, хотя хорошо известно, что *pretest*-оценки обладают плохими статистическими свойствами, одно из которых — *равномерная неэффективность*[2]. На практике избежать предварительного тестирования почти невозможно. Наша точка зрения состоит в том, что следует вычислять корректно смещение и дисперсию (или среднеквадратичное отклонение) оценки, полностью принимая во внимание то, что оценивание и отбор модели интегрированы в одну процедуру.

[2]Оценка параметра называется равномерно неэффективной, если существует другая оценка, равномерно лучшая данной во всем диапазоне возможных значений параметра (например, в смысле меньшего среднеквадратичного отклонения).

14.2. Постановка задачи

В дальнейшем в данной главе мы будем рассматривать линейную модель множественной регресии (см. п. 3.1)

$$y = X\beta + Z\gamma + \varepsilon, \tag{14.1}$$

где y — $(n \times 1)$ вектор наблюдений зависимой переменной, X и Z — матрицы неслучайных регрессоров размеров $(n \times k)$ и $(n \times m)$ соответственно, ε — $(n \times 1)$ вектор ошибок и β и γ — векторы неизвестных (неслучайных) параметров размеров $(k \times 1)$ и $(m \times 1)$ соответственно. Мы предполагаем, что $k \geqslant 1, m \geqslant 1, n-k-m \geqslant 1$, блочная матрица $[X\ Z]$ имеет полный ранг $k+m$, и ошибки являются независимыми нормальными одинаково распределенными случайными величинами: $\varepsilon \sim N(0, \sigma^2 I_n)$.

Это обычная формулировка стандартной линейной модели с нормальными ошибками, за исключением того, что мы теперь делаем различие между регрессорами X и Z. В (почти) каждой модели есть регрессоры, которые *должны* быть включены в модель. Этому может быть несколько причин. Например, это следует из экономической теории или из предыдущего опыта, возможно, наши коллеги сочтут модель неприемлемой, если в нее не будут включены данные регрессоры, возможно, это как раз те регрессоры, влияние которых на y мы и хотим изучить. Все такие регрессоры называются *основными* регрессорами (представляющими «фокус» интереса) и включаются в набор регрессоров X. Таким образом, регрессоры включаются в модель независимо от полученных значений t-статистик, соответствующих оценкам параметров β.

С другой стороны, матрица Z содержит дополнительные объясняющие переменные («вспомогательные» регрессоры), в необходимости включения которых в модель мы не столь уверены. Они могут быть включены в модель или могут отсутствовать в модели. Нашей целью является оценка параметров β. Переменные Z необходимы только для того, чтобы улучшить оценки параметров β, в то время как γ является вектором *вспомогательных* (излишних, необязательных (*nuisance*)) параметров.

Введем следующие обозначения:

$$\begin{aligned} M &= I_n - X(X'X)^{-1}X'; \\ Q &= (X'X)^{-1}X'Z(Z'MZ)^{-1/2}; \\ \eta &= \frac{1}{\sigma}(Z'MZ)^{1/2}\gamma. \end{aligned} \tag{14.2}$$

Здесь η — нормированный вектор параметров, а матрица Q может быть интерпретирована как (нормированная) матрица корреляций между X и Z. Очевидно, $Q = 0$ тогда и только тогда, когда Z ортогональна X. МНК-оценки параметров β и γ можно представить в виде (см. (4.19))

$$\widehat{\beta}_u = \widehat{\beta}_r - Q\widehat{\theta} \quad \text{и} \quad \widehat{\gamma} = (Z'MZ)^{-1}Z'My,$$

где

$$\widehat{\beta}_r = (X'X)^{-1}Xy \quad \text{и} \quad \widehat{\theta} = (Z'MZ)^{1/2}\widehat{\gamma}.$$

Индексы u и r означают «без ограничения» и «с ограничением» ($\gamma = 0$) соответственно. Пусть $\widehat{\eta} = \widehat{\theta}/\sigma$, тогда $\widehat{\eta} \sim N(\eta, I_m)$. Отметим, что случайный вектор $\widehat{\eta}$ наблюдаем только в том случае, когда дисперсия ошибки σ^2 известна, в то время как случайный вектор $\widehat{\theta}$ наблюдаем независимо от того, известна σ^2 или нет.

14.3. Основной результат

В данной главе мы часто будем использовать следующий результат, касающийся метода наименьших квадратов. Пусть S_i — $m \times r_i$ матрица ранга $r_i \geqslant 0$, такая что $S_i' = [I_{r_i}\ 0]$, или получается из последней перестановкой столбцов. Тогда уравнение $S_i'\gamma = 0$ означает, что несколько компонент вектора γ равны нулю[3]. Справедлива следующая теорема:

[3]Случай $r_i = 0$ соответствует пустой матрице S_i и, соответственно, отсутствию ограничений на вектор γ.

Теорема 14.1. *МНК-оценка параметров β и γ в линейной модели* (14.1) *при ограничении $S_i'\gamma = 0$ имеет вид*

$$\widehat{\beta}_{(i)} = \widehat{\beta}_r - QW_i\widehat{\theta}, \qquad \widehat{\gamma}_{(i)} = (Z'MZ)^{-1/2}W_i\widehat{\theta},$$

где

$$W_i = I_m - P_i,$$
$$P_i = (Z'MZ)^{-1/2}S_i(S_i'(Z'MZ)^{-1}S_i)^{-1}S_i'(Z'MZ)^{-1/2}$$

являются симметричными идемпотентными $m \times m$ матрицами рангов $m - r_i$ и r_i соответственно. ($P_i = 0$ в том случае, если $r_i = 0$) Вектор остатков имеет вид

$$e_{(i)} = y - X\widehat{\beta}_{(i)} - Z\widehat{\gamma}_{(i)} = D_i y,$$

где

$$D_i = M - MZ(Z'MZ)^{-1/2}W_i(Z'MZ)^{-1/2}Z'M$$

является симметричной идемпотентной матрицей ранга $(n - k - m + r_i)$. Оценка $\widehat{\beta}_{(i)}$ имеет нормальное распределение

$$\widehat{\beta}_{(i)} \sim N\left(\beta + \sigma QP_i\eta,\ \sigma^2((X'X)^{-1} + QW_iQ')\right),$$

а величина $s_{(i)}^2 = e_{(i)}'e_{(i)}/(n - k - m + r_i)$ имеет нецентральное χ^2-распределение (см. приложение МС, *п.* 3)

$$\frac{(n - k - m + r_i)\, s_{(i)}^2}{\sigma^2} \sim \chi^2(n - k - m + r_i,\ \eta' P_i \eta).$$

Оставим доказательство этой теоремы в качестве упражнения (см. упражнение 14.2). Отметим, что оценка с частичным ограничением $\widehat{\beta}_{(i)}$ является линейной комбинацией двух независимых (см. упражнение 14.1) векторов $\widehat{\beta}_r$ и $\widehat{\theta}$. Оценка $\widehat{\gamma}_{(i)}$ является линейной функцией только от $\widehat{\theta}$, и поэтому не зависит от $\widehat{\beta}_r$.

14.4. *Pretest*-оценка

Рассмотрим простейший случай, когда у нас есть только один вспомогательный регрессор, т.е. $m = 1$. Мы можем выбирать между двумя моделями: моделью с ограничением ($\gamma = 0$) и моделью без ограничения. В том случае, если мы выбираем модель с ограничением, мы получаем оценку параметра β, равную $\widehat{\beta}_r$; если мы выбираем модель без ограничения, то получаем оценку $\widehat{\beta}_u$. Обычно мы используем t-статистику коэффициента γ для того, чтобы сделать выбор между этими двумя моделями. Таким образом, оценка параметра β имеет следующий вид:

$$\widehat{\beta} = \begin{cases} \widehat{\beta}_r, & \text{если } |t_\gamma| \leqslant c, \\ \widehat{\beta}_u, & \text{если } |t_\gamma| > c, \end{cases}$$

для некоторого порогового значения $c \geqslant 0$. Например, $c = 1.96$ и $c = 2.58$ соответствуют 5%-ному и 1%-ному уровням значимости (для нормального распределения; для распределения Стьюдента значения c несколько выше). Подчеркнем, что оценка $\widehat{\beta}$ не совпадает с оценками $\widehat{\beta}_r$ или $\widehat{\beta}_u$, но равна той или другой в зависимости от критерия, основанного на значении случайной величины t.

Приведем другой способ записи оценки: $\widehat{\beta} = \lambda\widehat{\beta}_u + (1 - \lambda)\widehat{\beta}_r$, где

$$\lambda = \begin{cases} 0, & \text{если } |t_\gamma| \leqslant c, \\ 1, & \text{если } |t_\gamma| > c. \end{cases}$$

Таким образом, оценка $\widehat{\beta}$ является взвешенным средним оценок $\widehat{\beta}_u$ и $\widehat{\beta}_r$ со случайным весом λ.

В случае $m = 2$ у нас есть четыре модели: модель с ограничением ($\gamma_1 = \gamma_2 = 0$), две модели с частичными ограничениями ($\gamma_1 \neq 0$, $\gamma_2 = 0$ или $\gamma_1 = 0$, $\gamma_2 \neq 0$) и модель без ограничений ($\gamma_1 \neq 0$ и $\gamma_2 \neq 0$). В общем случае имеется 2^m различных моделей, по одной для каждого подмножества параметров $\gamma_1, \ldots, \gamma_m$. (Имеется в виду, что параметры из подмножества приравнены к 0.) *Pretest*-оценка вектора параметров β получается в результате 1) выбора одной из этих моделей (на основе t- или F-тестов

или других критериев выбора модели) и, затем, 2) оценивания $\boldsymbol{\beta}$ по выбранной модели.

Мы предположим, что критерий выбора модели зависит от $\boldsymbol{y}$ только через $\boldsymbol{My}$, остатки в модели с ограничениями. Это условие выполнено во всех стандартных процедурах отбора моделей. (Заметим, что остатки в i-й модели всегда выражаются как $\boldsymbol{e}_{(i)} = \boldsymbol{D}_i\boldsymbol{My}$ для некоторой идемпотентной матрицы $\boldsymbol{D}_i$.) Это предположение приводит к существенным упрощениям.

14.5. WALS-оценка

Понятие *pretest*-оценки допускает естественное и, как будет показано ниже, полезное обобщение. Как и ранее, рассмотрим сначала случай $m = 1$. Запишем оценку в виде $\widehat{\boldsymbol{\beta}} = \lambda\widehat{\boldsymbol{\beta}}_u + (1-\lambda)\widehat{\boldsymbol{\beta}}_r$, но теперь пусть λ является гладкой возрастающей функцией t_γ. Это выглядит разумным подходом и позволяет нам произвольно выбирать уровень значимости. (В самом деле, почему следует брать в качестве уровня значимости 5% или 1%, а не, скажем, 50% или 95%?)

В общем случае WALS-оценка (*weighted-average least-squares*) параметра $\boldsymbol{\beta}$ определяется как:

$$\widehat{\boldsymbol{\beta}} = \sum_{i=1}^{2^m} \lambda_i \widehat{\boldsymbol{\beta}}_{(i)},$$

где сумма берется по всем 2^m моделям, получающимся при приравнивании нескольких коэффициентов γ_j нулю.

Мы предполагаем, что весовые коэффициенты λ_i удовлетворяют условиям:

$$\lambda_i = \lambda_i(\boldsymbol{My}), \quad \lambda_i \geqslant 0, \quad \text{и} \quad \sum_i \lambda_i = 1.$$

WALS-оценка тогда может быть записана в следующем виде:

$$\begin{aligned}\widehat{\boldsymbol{\beta}} &= \sum_i \lambda_i \widehat{\boldsymbol{\beta}}_{(i)} = \sum_i \lambda_i(\widehat{\boldsymbol{\beta}}_r - \boldsymbol{QW}_i\widehat{\boldsymbol{\theta}}) \\ &= (\sum_i \lambda_i)\widehat{\boldsymbol{\beta}}_r - \boldsymbol{Q}(\sum_i \lambda_i \boldsymbol{W}_i)\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\beta}}_r - \boldsymbol{QW}\widehat{\boldsymbol{\theta}},\end{aligned}$$

где

$$\boldsymbol{W} = \boldsymbol{I}_m - \boldsymbol{P}, \qquad \boldsymbol{P} = \sum_{i=1}^{2^m} \lambda_i \boldsymbol{P}_i.$$

Заметим, что хотя матрицы $\boldsymbol{P}_i$ и $\boldsymbol{W}_i$ неслучайные, однако матрицы $\boldsymbol{P}$ и $\boldsymbol{W}$ случайные, поскольку $\{\lambda_i\}$ являются случайными величинами.

Очевидно, что *pretest*-оценка является частным случаем WALS-оценки, в случае, когда все λ_i равны 0, за исключением одного, равного 1.

14.6. Теорема эквивалентности

В том случае, когда дисперсия σ^2 известна, любая процедура предварительного тестирования использует t- и F-статистики, которые зависят только от $\hat{\theta}$. В случае, когда σ^2 не известна, можно получить ее оценку s_u^2 (оценка, основанная на МНК-оценках регрессии без ограничения). В этом последнем случае все t- и F-статистики зависят от $(\hat{\theta}, s_u^2)$. В п. 3.1 показано, что s_u^2 не зависит от $(\widehat{\boldsymbol{\beta}}_u, \widehat{\boldsymbol{\gamma}})$. Следовательно, $\widehat{\boldsymbol{\beta}}_r$ не зависит от s_u^2 и поэтому не зависит также и от $(\widehat{\boldsymbol{\theta}}, s_u^2)$. Наконец, в том случае, если дисперсия σ^2 не известна и берется ее оценка $s_{(i)}^2$ по регрессии с (частичным) ограничением, соответствующим матрице ограничений S_i (см. п. 14.3), t- и F-статистики зависят не только от $(\widehat{\boldsymbol{\theta}}, s_u^2)$. Однако они попрежнему зависят только от $\boldsymbol{My}$, поскольку $\widehat{\boldsymbol{\gamma}}_{(i)}$ и $\boldsymbol{e}_{(i)}$ являются линейными функциями $\boldsymbol{My}$. Таким образом, из простого замечания, что $\widehat{\boldsymbol{\beta}}_r$ и $\widehat{\boldsymbol{\theta}}$ независимы, вытекает, что t- и F-статистики не зависят от $\widehat{\boldsymbol{\beta}}_r$. Этот факт является ключевым при доказательстве следующего результата.

Теорема 14.2. (теорема эквивалентности).
Пусть $\widehat{\boldsymbol{\beta}} = \sum_i \lambda_i \widehat{\boldsymbol{\beta}}_{(i)}$, *где* $\lambda_i = \lambda_i(\boldsymbol{My})$, $\lambda_i \geqslant 0$ *и* $\sum_i \lambda_i = 1$. *Тогда*

$$\mathrm{E}\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} - \sigma\boldsymbol{Q}\mathrm{E}(\boldsymbol{W}\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}), \quad \mathrm{V}(\widehat{\boldsymbol{\beta}}) = \sigma^2\left((\boldsymbol{X}'\boldsymbol{X})^{-1} + \boldsymbol{Q}\,\mathrm{V}(\boldsymbol{W}\widehat{\boldsymbol{\eta}})\boldsymbol{Q}'\right),$$

и, следовательно,

$$\mathrm{MSE}(\widehat{\boldsymbol{\beta}}) = \sigma^2\left((\boldsymbol{X}'\boldsymbol{X})^{-1} + \boldsymbol{Q}\,\mathrm{MSE}(\boldsymbol{W}\widehat{\boldsymbol{\eta}})\,\boldsymbol{Q}'\right).$$

Доказательство. Из того, что векторы $\widehat{\beta}_r$ и $\boldsymbol{My}$ независимы, получаем

$$\mathrm{E}(\widehat{\beta}_r \mid \boldsymbol{My}) = \mathrm{E}(\widehat{\beta}_r), \qquad \mathrm{V}(\widehat{\beta}_r \mid \boldsymbol{My}) = \mathrm{V}(\widehat{\beta}_r).$$

Отсюда,

$$\begin{aligned}\mathrm{E}(\widehat{\beta} \mid \boldsymbol{My}) &= \mathrm{E}(\widehat{\beta}_r \mid \boldsymbol{My}) - \boldsymbol{Q}\mathrm{E}(\boldsymbol{W}\hat{\theta} \mid \boldsymbol{My}) \\ &= \mathrm{E}(\widehat{\beta}_r) - \sigma\boldsymbol{QW}\widehat{\eta} = \beta - \sigma\boldsymbol{Q}(\boldsymbol{W}\widehat{\eta} - \eta)\end{aligned}$$

и

$$\mathrm{V}(\widehat{\beta} \mid \boldsymbol{My}) = \mathrm{V}(\widehat{\beta}_r \mid \boldsymbol{My}) = \mathrm{V}(\widehat{\beta}_r) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}.$$

Безусловное математическое ожидание оценки имеет вид

$$\begin{aligned}\mathrm{E}(\widehat{\beta}) &= \mathrm{E}(\mathrm{E}(\widehat{\beta} \mid \boldsymbol{My})) = \mathrm{E}(\beta - \sigma\boldsymbol{Q}(\boldsymbol{W}\widehat{\eta} - \eta)) \\ &= \beta - \sigma\boldsymbol{Q}\mathrm{E}(\boldsymbol{W}\widehat{\eta} - \eta),\end{aligned}$$

а безусловная матрица ковариаций равна

$$\begin{aligned}\mathrm{V}(\widehat{\beta}) &= \mathrm{E}(\mathrm{V}(\widehat{\beta} \mid \boldsymbol{My})) + \mathrm{V}(\mathrm{E}(\widehat{\beta} \mid \boldsymbol{My})) \\ &= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} + \mathrm{V}(\beta - \sigma\boldsymbol{Q}(\boldsymbol{W}\widehat{\eta} - \eta)) \\ &= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} + \sigma^2\boldsymbol{Q}\,\mathrm{V}(\boldsymbol{W}\widehat{\eta})\boldsymbol{Q}'.\end{aligned}$$

Отсюда получаем, что матрица среднеквадратичных отклонений равна

$$\begin{aligned}\mathrm{MSE}(\widehat{\beta}) &= \mathrm{V}(\widehat{\beta}) + \mathrm{E}(\widehat{\beta} - \beta)\mathrm{E}(\widehat{\beta} - \beta)' \\ &= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} + \sigma^2\boldsymbol{Q}\,\mathrm{V}(\boldsymbol{W}\widehat{\eta})\boldsymbol{Q}' \\ &\quad + \sigma^2\boldsymbol{Q}\mathrm{E}(\boldsymbol{W}\widehat{\eta} - \eta)\mathrm{E}(\boldsymbol{W}\widehat{\eta} - \eta)'\boldsymbol{Q}' \\ &= \sigma^2((\boldsymbol{X}'\boldsymbol{X})^{-1} + \boldsymbol{Q}\,\mathrm{MSE}(\boldsymbol{W}\widehat{\eta})\,\boldsymbol{Q}').\end{aligned}$$

Что и требовалось доказать.

Теорема 14.2 доказана в упрощенной форме в (Magnus and Durbin, 1999). Приведенное выше ее обобщение приведено в (Danilov and Magnus, 2002).

Важность этой теоремы состоит в том, что если мы найдем λ_i такие, что $\boldsymbol{W}\widehat{\eta}$ будет оптимальной оценкой η, то *те же самые* λ_i дадут оптимальную WALS-оценку вектора параметров β. Проблема оценивания вектора параметров β в контексте регрессии сводится, таким образом, к задаче оценивания вектора η по единственному вектору наблюдений $\widehat{\eta} \sim N(\eta, \boldsymbol{I}_m)$.

14.7. Предварительное тестирование и эффект «занижения»

Задача нахождения λ_i, таких что $\boldsymbol{W}\widehat{\eta}$ будет оптимальной оценкой η, является интересной и важной, но мы не рассматриваем ее в данной главе. (Анализ этой задачи для случая $m = 1$ можно найти в (Magnus, 2002).) Вместо решения этой задачи мы сосредоточимся в данной главе на обычно используемой *pretest*-оценке.

В рамках стандартной линейной модели $\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Z}\gamma + \varepsilon$, с нормальными ошибками $\varepsilon \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n)$ мы определяем *процедуру* предварительного тестирования как двухшаговую процедуру. На первом шаге происходит выбор модели. В общем случае есть 2^m моделей, которые рассматриваются в процедуре предварительного тестирования. На втором шаге мы оцениваем неизвестные параметры β и σ^2 по выбранной модели. Такая процедура порождает *pretest*-оценку $\widehat{\beta}$ (и s^2). Для определенной таким способом процедуры предварительного тестирования все весовые коэффициенты λ_i равны 0, кроме одного, равного 1. Как и в теореме 14.2, мы накладываем условие, что отбор модели зависит от $\boldsymbol{y}$ только через $\boldsymbol{My}$, кроме того, в дальнейшем мы будем предполагать, что параметр σ^2 известен.

Матрица среднеквадратичных отклонений оценки $\widehat{\beta}$ согласно теореме 14.2 равна

$$\mathrm{MSE}(\widehat{\beta}) = \sigma^2((\boldsymbol{X}'\boldsymbol{X})^{-1} + \boldsymbol{Q}\mathrm{MSE}(\boldsymbol{W}\widehat{\eta})\boldsymbol{Q}').$$

В обычной прикладной эконометрической практике выбирается та же оценка $\widehat{\beta}$, однако эффект предварительного тестирования игнорируется, оценка считается несмещенной и матрица среднеквадратичных отклонений оценки берется равной матрице ковариаций оценки $\widehat{\beta}$, рассчитанной в предположении верности выбранной модели (см. теорему 14.1)[4]

$$\widetilde{\mathrm{MSE}}(\widehat{\beta}) = \sigma^2((\boldsymbol{X}'\boldsymbol{X})^{-1} + \boldsymbol{QWQ}'),$$

[4] $\widetilde{\mathrm{MSE}}$ — «Reported» MSE, т.е. то значение MSE, которое сообщается в результате исследования.

здесь $\boldsymbol{W} = \boldsymbol{W}_i$ если выбрана i-я модель. Заметим, что матрица $\widetilde{\mathrm{MSE}}(\widehat{\boldsymbol{\beta}})$ случайная, поскольку матрица $\boldsymbol{W}$ случайная. Предположим, что целью («фокусом») нашего исследования является (скалярная) величина $\boldsymbol{\omega}'\boldsymbol{\beta}$, где $\boldsymbol{\omega}$ — произвольный ненулевой $k \times 1$ вектор. Для того, чтобы сравнить среднеквадратичное отклонение оценки $\boldsymbol{\omega}'\widehat{\boldsymbol{\beta}}$

$$\mathrm{MSE}(\boldsymbol{\omega}'\widehat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{\omega}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{\omega} + \boldsymbol{\omega}'\boldsymbol{Q}\,\mathrm{MSE}(\boldsymbol{W}\widehat{\boldsymbol{\eta}})\boldsymbol{Q}'\boldsymbol{\omega}) \qquad (14.3)$$

с соответствующим значением, полученным при игнорировании произведенной процедуры предварительного тестирования

$$\widetilde{\mathrm{MSE}}(\boldsymbol{\omega}'\widehat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{\omega}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{\omega} + \boldsymbol{\omega}'\boldsymbol{Q}\boldsymbol{W}\boldsymbol{Q}'\boldsymbol{\omega}), \qquad (14.4)$$

определим *коэффициент занижения* «истинного» MSE по отношению к «сообщаемому» MSE (*underreporting ratio*)[5] UR, как 1 минус отношение (14.4) и (14.3). А именно:

$$\mathrm{UR} = 1 - \frac{\widetilde{\mathrm{MSE}}(\boldsymbol{\omega}'\widehat{\boldsymbol{\beta}})}{\mathrm{MSE}(\boldsymbol{\omega}'\widehat{\boldsymbol{\beta}})} = \frac{\boldsymbol{q}'(\boldsymbol{R} - \boldsymbol{W})\boldsymbol{q}}{\boldsymbol{q}'\boldsymbol{R}\boldsymbol{q} + (1/q_0^2)}, \qquad (14.5)$$

где

$$\boldsymbol{R} = \boldsymbol{R}(\boldsymbol{\eta}) = \mathrm{MSE}(\boldsymbol{W}\widehat{\boldsymbol{\eta}}), \quad \boldsymbol{q} = \frac{\boldsymbol{Q}'\boldsymbol{\omega}}{\sqrt{\boldsymbol{\omega}'\boldsymbol{Q}\boldsymbol{Q}'\boldsymbol{\omega}}}, \quad q_0^2 = \frac{\boldsymbol{\omega}'\boldsymbol{Q}\boldsymbol{Q}'\boldsymbol{\omega}}{\boldsymbol{\omega}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{\omega}}.$$

Заметим, что $\boldsymbol{q}'\boldsymbol{q} = 1$. Величина UR является случайной, так как она зависит от матрицы $\boldsymbol{W}$, которая зависит от $\widehat{\boldsymbol{\eta}}$. Как UR, так и ее математическое ожидание не наблюдаемы, поскольку они зависят от $\boldsymbol{\eta}$ через $\boldsymbol{R}(\boldsymbol{\eta})$.

Можно было бы ожидать, что матрица $\mathrm{MSE}(\widehat{\boldsymbol{\beta}})$ не меньше матрицы $\mathrm{E}(\widetilde{\mathrm{MSE}}(\widehat{\boldsymbol{\beta}}))$ (т. е. их разность есть неотрицательно определенная матрица), поскольку процедура предварительного тестирования вносит дополнительную неопределенность, которая игнорируется в сообщаемом значении MSE. Поскольку

$$\mathrm{MSE}(\boldsymbol{W}\widehat{\boldsymbol{\eta}}) = \sum_{i=1}^{2^m} \mathrm{E}\lambda_i(\boldsymbol{W}_i\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})(\boldsymbol{W}_i\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})'$$

и

$$\mathrm{E}(\boldsymbol{W}) = \sum_{i=1}^{2^m} (\mathrm{E}\lambda_i)\boldsymbol{W}_i,$$

то это условие выполняется, если матрица

$$\sum_{i=1}^{2^m} \mathrm{E}\lambda_i\left((\boldsymbol{W}_i\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})(\boldsymbol{W}_i\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta})' - \boldsymbol{W}_i\right) \qquad (14.6)$$

неотрицательно определена. На самом деле, можно сконструировать процедуру предварительного тестирования (упражнение 14.4), для которой матрица (14.6) не является неотрицательно определенной. Однако такой пример выглядит достаточно нелепо. Назовем процедуру предварительного тестирования *естественной*, если матрица (14.6) неотрицательно определена при всех значениях параметров. Для любой естественной процедуры предварительного тестирования величина E(UR) принимает значения между 0 и 1. В том случае, когда величина q_0^2 (известная исследователю) стремится к 0, то нет «занижения»: $\mathrm{E}(\mathrm{UR}) \to 0$. Однако если значение q_0^2 велико, то ожидаемое значение коэффициента занижения E(UR) может быть близко к 1.

Матрица $\mathrm{E}(\boldsymbol{W})$ (размера $m \times m$) есть взвешенное среднее идемпотентных матриц, и, следовательно, ограничена: все ее элементы по абсолютной величине не превосходят 1, а все диагональные элементы и собственные числа лежат в интервале $[0, 1]$. На самом деле выполняется следующее неравенство:

$$0 \leqslant \pi_u \leq \xi_j(\mathrm{E}\boldsymbol{W}) \leqslant 1 - \pi_r \leqslant 1 \qquad (j = 1, \ldots, m),$$

где $\xi_j(\boldsymbol{A})$ есть j-е собственное число матрицы $\boldsymbol{A}$, π_u — вероятность выбора модели без ограничений ($\boldsymbol{P}_i = \boldsymbol{0}$), а π_r — вероятность выбора модели с ограничением ($\boldsymbol{P}_i = \boldsymbol{I}_m$).

[5] Саму ситуацию, когда сообщаемое значение среднеквадратичного отклонения оценки не учитывает дополнительную неопределенность, связанную с процедурой предварительного отбора модели, и поэтому занижено, будем далее называть «эффект занижения» (*underreporting*).

Математическое ожидание E(UR) является функцией величин $\boldsymbol{q}$ (с нормировкой $\boldsymbol{q}'\boldsymbol{q} = 1$), q_0^2, $\boldsymbol{\eta}$, $\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z}$ (и m). Максимизация по $\boldsymbol{q}$ приводит к неравенству

$$\mathrm{E}(\mathrm{UR}) \leqslant q_0^2 \max_{1\leqslant j\leqslant m} \xi_j \left((\boldsymbol{I}_m + q_0^2\boldsymbol{R})^{-1/2}(\boldsymbol{R} - \mathrm{E}\boldsymbol{W})(\boldsymbol{I}_m + q_0^2\boldsymbol{R})^{-1/2} \right).$$

Вводя следующее обозначение:

$$\mathrm{E}^*(\mathrm{UR}) = \max_{\boldsymbol{q}, q_0^2} \mathrm{E}(\mathrm{UR}),$$

получаем, что при $q_0^2 \to \infty$,

$$\begin{aligned} \mathrm{E}^*(\mathrm{UR}) &= 1 - \min_{1\leqslant j\leqslant m} \xi_j(\boldsymbol{R}^{-1/2}(\mathrm{E}\boldsymbol{W})\boldsymbol{R}^{-1/2}) \\ &\leqslant 1 - \frac{\pi_u}{\max_j \xi_j(\boldsymbol{R})}. \end{aligned} \tag{14.7}$$

Полученное выражение зависит от $\boldsymbol{\eta}$ и $\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z}$ (и m). Из (14.7) видно, что ожидаемое значение UR может быть сколь угодно близко к 1, если матрица среднеквадратичных отклонений $\boldsymbol{R}$ не ограничена по $\boldsymbol{\eta}$. Это не может произойти при $m = 1$ (кроме случая, когда мы *всегда* выбираем модель с ограничением, не обращая внимания на полученные значения t-статистик), но возможно при $m \geqslant 2$.

Поскольку E(UR) зависит от $\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z}$, рассмотрим кратко роль этой матрицы. Без ущерба общности можно нормировать все переменные $\boldsymbol{z}_j$ так, что $\boldsymbol{z}_j'\boldsymbol{M}\boldsymbol{z}_j = 1$ для всех $j = 1, \ldots, m$. Рассмотрим частный случай, когда мы выбираем «ортогональные» переменные $\boldsymbol{z}_j$ (в том смысле, что $\boldsymbol{M}\boldsymbol{z}_i$ и $\boldsymbol{M}\boldsymbol{z}_j$ ортогональны для всех $i \neq j$). Тогда $\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z} = \boldsymbol{I}_m$, что приводит к существенным упрощениям.

Теорема 14.3. *Пусть $\lambda(x) = 1$ если $|x| > c$, иначе $\lambda(x) = 0$, для некоторого $c > 0$. Для частного случая, когда $\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z} = \boldsymbol{I}_m$ и параметр σ^2 известен, имеем:*

(а) *Все модели, включающие регрессор z_j, имеют одинаковое значение t-статистики для γ_j.*

(б) *Предположим, что мы включаем z_j тогда и только тогда, когда t-статистика $\widehat{\eta}_j$ значима, то есть $|\widehat{\eta}_j| > c$, для некоторого $c > 0$. Тогда матрица $\boldsymbol{W}$ — диагональная, с элементами $w_{jj} = \lambda(\widehat{\eta}_j)$, а $\mathrm{MSE}(\boldsymbol{W}\widehat{\boldsymbol{\eta}}) = \boldsymbol{V} + \boldsymbol{d}\boldsymbol{d}'$, где $\boldsymbol{V}$ — диагональная $m \times m$ матрица, а $\boldsymbol{d}$ — $m \times 1$ вектор с элементами соответственно*

$$v_{jj} = \mathrm{V}(\lambda(\widehat{\eta}_j)\widehat{\eta}_j), \qquad d_j = \mathrm{E}(\lambda(\widehat{\eta}_j)\widehat{\eta}_j - \eta_j).$$

Доказательство. Из теоремы 14.1 получаем $\boldsymbol{P}_i = \boldsymbol{S}_i(\boldsymbol{S}_i'\boldsymbol{S}_i)^{-1}\boldsymbol{S}_i'$. Поскольку матрица отбора $\boldsymbol{S}_i'$ имеет вид $[\boldsymbol{I}_{r_i}\ \boldsymbol{0}]$ (или получается из последней перестановкой столбцов), то $\boldsymbol{S}_i'\boldsymbol{S}_i = \boldsymbol{I}_{r_i}$, и, следовательно, матрица $\boldsymbol{P}_i$ диагональная, с r_i единицами и $m - r_i$ нулями на диагонали. Поэтому матрица $\boldsymbol{W}_i$ также диагональная, с $m - r_i$ единицами и r_i нулями на диагонали. Из теоремы 14.1 также следует что $\widehat{\boldsymbol{\gamma}}_{(i)} = \boldsymbol{W}_i\widehat{\boldsymbol{\theta}}$ есть оценка вектора параметров $\boldsymbol{\gamma}$ при ограничении $\boldsymbol{S}_i'\boldsymbol{\gamma} = \boldsymbol{0}$. Поэтому оценкой параметра γ_j при данном ограничении является j-я компонента вектора $\widehat{\boldsymbol{\gamma}}_{(i)}$, которая равна либо 0 (если переменная z_j исключена из модели), либо $\widehat{\theta}_j$ (если переменная z_j включена в модель). Таким образом, все модели, которые включают регрессор z_j, дают *ту же* оценку γ_j, независимо от того, какие еще γ оцениваются. Однако t-статистика для параметра γ_j есть $\widehat{\eta}_j = \widehat{\theta}_j/\sigma$, откуда следует (а). Матрица $\boldsymbol{W}$ — диагональная, поскольку все $\boldsymbol{W}_i$ диагональные. Ее j-й диагональный элемент w_{jj} равен либо 0 (если переменная z_j исключена из модели), либо 1 (если переменная z_j включена в модель). Иначе говоря, $w_{jj} = \lambda(\widehat{\eta}_j)$. Отсюда также следует независимость компонент вектора $\boldsymbol{W}\widehat{\boldsymbol{\eta}}$, откуда вытекает (б). Теорема доказана.

Поскольку процедура отбора модели может оказать существенное влияние на оценки интересующих нас параметров, то желательно (если это возможно) выбрать вспомогательные регрессоры так, что $\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z} = \boldsymbol{I}_m$. В большинстве случаев такой выбор позволяет не только сделать *pretest*-оценку независимой от процедуры отбора модели, но также получить точные аналитические выражения для моментов оценок и гарантировать ограниченность среднеквадратичного отклонения оценок при всех m. (В общем неортогональном случае среднеквадратичные отклонения оценок ограничены при $m = 1$ и не обязательно ограничены при $m \geqslant 2$.)

14.8. Эффект «занижения». Один вспомогательный параметр

В случае одного вспомогательного (необязательного) параметра модель принимает вид: $y = X\beta + \gamma z + \varepsilon$, где скаляр γ — вспомогательный параметр. В этом случае мы сравниваем только две модели: модель без ограничения ($W_1 = 1$, $\widehat{\beta}_{(1)} = \widehat{\beta}_u$, $\lambda_1 = \lambda$) и модель с ограничением ($W_2 = 0$, $\widehat{\beta}_{(2)} = \widehat{\beta}_r$, $\lambda_2 = 1 - \lambda$). В результате мы получаем

$$\widehat{\beta} = \lambda\widehat{\beta}_u + (1-\lambda)\widehat{\beta}_r, \qquad W = \lambda,$$

и

$$\mathrm{MSE}(W\widehat{\eta}) = \mathrm{MSE}(\lambda\widehat{\eta}) = \mathrm{E}(\lambda\widehat{\eta} - \eta)^2, \qquad \mathrm{E}W = \mathrm{E}\lambda.$$

Коэффициент занижения равен

$$\mathrm{UR}(\widehat{\eta}, \eta) = \frac{R(\eta) - \lambda(\widehat{\eta})}{R(\eta) + (1/q_0^2)},$$

где $\lambda(\widehat{\eta}) = 1$, если $|\widehat{\eta}| > c$ для некоторого $c > 0$, и $\lambda(\widehat{\eta}) = 0$, если $|\widehat{\eta}| \leqslant c$, и

$$R(\eta) = \mathrm{E}(\lambda\widehat{\eta} - \eta)^2, \qquad q_0^2 = \frac{(z'X(X'X)^{-1}\omega)^2}{(z'Mz)(\omega'(X'X)^{-1}\omega)}.$$

Снова предположим, что дисперсия σ^2 известна и значение c задано (например, $c = 1.96$). Тогда λ зависит только от $\widehat{\eta}$, R зависит только от η, и поэтому UR зависит от q_0^2 и $\widehat{\eta}$ (известных исследователю) и параметра η (который неизвестен).

Легко видеть, что UR растет с ростом $R(\eta)$. Таким образом, случайная величина $\lambda\widehat{\eta}$, рассматриваемая в качестве оценки η, играет существенную роль в определении коэффициента занижения. Графики ее дисперсии, квадрата смещения и среднеквадратичного отклонения представлены на рис. 14.1.

Смещение оценки $\lambda\widehat{\eta}$ отрицательно при $\eta > 0$ и достигает минимального значения -0.66 при $\eta = 1.46$. Дисперсия достигает минимального значения 0.28 при $\eta = 0$ и максимального значения 2.23 при $\eta = 2.34$. График среднеквадратичного отклонения

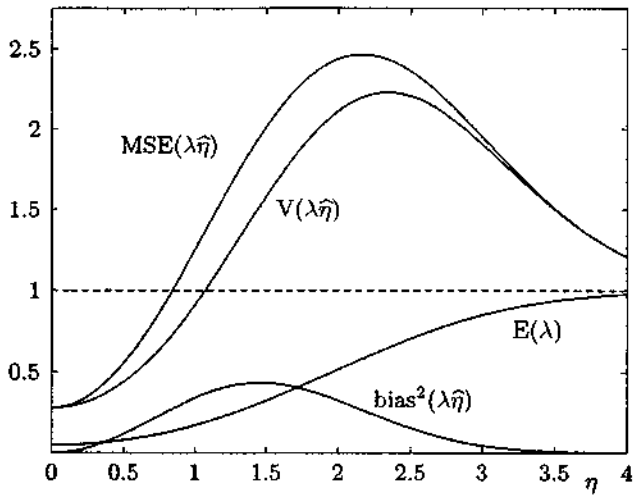


Рис. 14.1. Моменты $\lambda\widehat{\eta}$ и λ ($m = 1$, $c = 1.96$)

имеет форму, похожую на график дисперсии. Среднеквадратичное отклонение принимает минимальное значение при $\eta = 0$ и максимальное значение 2.46 при $\eta = 2.16$. Дисперсия оценки $\lambda\widehat{\eta}$ велика по сравнению с ее смещением, откуда следует, что уменьшение дисперсии важнее, чем уменьшение смещения.

Приведен также график математического ожидания «сообщаемого» значения среднеквадратичного отклонения оценки $\lambda\widehat{\eta}$ как функции параметра η для значения $c = 1.96$ и график среднеквадратичного отклонения оценки без ограничений, $\mathrm{MSE}(\widehat{\eta})$ (пунктирная линия, константа для $\lambda = 1$). Поскольку λ принимает только значения 0 и 1, ее ожидаемое значение $\mathrm{E}(\lambda)$ равно вероятности выбора модели без ограничения ($\lambda = 1$). Видно, что $\mathrm{E}(\lambda) \equiv \mathrm{P}(|\widehat{\eta}| > c)$ монотонно возрастает от 0.05 при $\eta = 0$ до 1 при $\eta = \infty$. Процедура предварительного тестирования является естественной, поскольку $\mathrm{MSE}(\lambda\widehat{\eta}) \geqslant \mathrm{E}(\lambda)$. Величина $\mathrm{E}(\mathrm{UR}) = 0.18$ при $\eta = 0$ и достигает максимального значения 0.57 при $\eta = 1.73$. Величина $\mathrm{E}(\mathrm{UR})$ изменяется в широком диапазоне (от 0 до 0.57) при изменении η, что означает, что среднеквадратичное отклонение *pretest*-оценки может в 2.3 превышать «сообщаемое» значение средне-

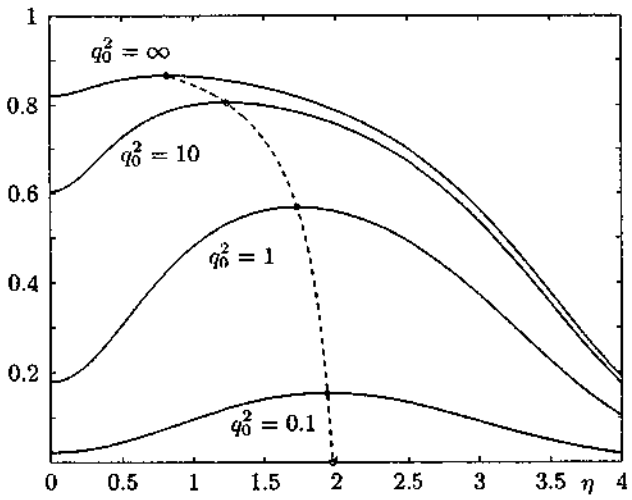


Рис. 14.2. E(UR) и точка максимума max(E(UR)) ($m = 1$, $c = 1.96$)

квадратичного отклонения ($1/(1 - 0.57) = 2.3$). Максимальное значение E(UR) зависит от q_0. График приведен на рис. 14.2.

На рис. 14.2 представлены графики E(UR) для пяти различных значений q_0^2: 0, 0.1, 1, 10 и ∞. При $q_0^2 = 0$ эффект «занижения» отсутствует. При $q_0^2 = \infty$ значение E(UR) достаточно большое; максимум достигается при $\eta = 0.82$ и равен E(UR) = 0.87.

Наконец, поскольку как UR так и E(UR) зависят от η, мы рассмотрим также поведение коэффициента занижения при $\eta = 1$. Значение $\eta = 1$ представляет особый интерес, потому что при таком значении исследователь находится в неопределенности, какую из моделей выбрать — с ограничением или без ограничения (см. упражнение 14.3). Мы приходим к выводу, что эффект «занижения», т. е. то, что не приводятся истинные значения смещения и дисперсии *pretest*-оценки, может привести к ошибочной интерпретации результатов *даже в случае* $m = 1$. Чем больше (известное исследователю) значение q_0^2, тем больше ожидаемое значение коэффициента занижения UR. Для заданного значения q_0^2 мы можем построить график ожидаемого значения UR, как функции η, как это сделано на рис. 14.2, и рассчитать максимум E(UR).

Другой способ — рассчитываем величину E(UR) в точке $\eta = \widehat{\eta}$ и используем ее как оценку степени эффекта «занижения». Максимум величины E(UR), равный 0.87, достигается при $q_0^2 = \infty$ и $\eta = 0.82$. Это означает, что в худшем случае сообщаемая обычно дисперсия *pretest*-оценки составляет лишь 13% от истинного значения ее среднеквадратичного отклонения.

14.9. Выбор модели: от общего к частному и от частного к общему

В случае одного вспомогательного параметра ($m = 1$) процедура предварительного тестирования выглядит просто. Вычисляется t-статистика коэффициента γ в модели без ограничений. Если $|t| > c$, то выбираем модель без ограничения (и получаем оценку $\widehat{\boldsymbol{\beta}}_u$), иначе выбираем модель с ограничением (получаем $\widehat{\boldsymbol{\beta}}_r$). В случае $m > 1$ существует много способов предварительного тестирования. Рассмотрим случай $m = 2$ при следующем условии: выбор модели основан исключительно на t-статистиках. В выбранной модели все t-статистики должны быть «значимые». Предположим также, что известна дисперсия σ^2.

Без потери общности можно нормализовать регрессоры $\boldsymbol{z}_1$ и $\boldsymbol{z}_2$, соответствующие вспомогательным параметрам γ_1 и γ_2, положив $\boldsymbol{z}_1'\boldsymbol{M}\boldsymbol{z}_1 = \boldsymbol{z}_2'\boldsymbol{M}\boldsymbol{z}_2 = 1$. Тогда

$$\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z} = \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix},$$

где $|r| < 1$, и

$$(\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z})^{-1/2} = \frac{1}{\sqrt{1-r^2}} \begin{bmatrix} \alpha & -\rho \\ -\rho & \alpha \end{bmatrix},$$

где

$$\alpha = \frac{\sqrt{1+r} + \sqrt{1-r}}{2}, \qquad \rho = \frac{\sqrt{1+r} - \sqrt{1-r}}{2}.$$

Теперь у нас четыре t-статистики, которые мы принимаем во внимание при отборе модели. Две из регрессии без ограничения (обозначим их t_1 и t_2), одна из модели с ограничением $\gamma_2 = 0$ (обо-

значим ее $t_{(1)}$) и еще одна из модели с ограничением $\gamma_1 = 0$ (обозначим ее $t_{(2)}$). Введем обозначения для компонент вектора $\widehat{\eta} = (\widehat{\eta}_1, \widehat{\eta}_2)'$. В соответствии с теоремой 14.1 каждая из четырех t-статистик является линейной функцией $\widehat{\eta}_1, \widehat{\eta}_2$:

$$t_1 = \alpha\widehat{\eta}_1 - \rho\widehat{\eta}_2, \qquad t_2 = -\rho\widehat{\eta}_1 + \alpha\widehat{\eta}_2,$$

и

$$t_{(1)} = \alpha\widehat{\eta}_1 + \rho\widehat{\eta}_2, \qquad t_{(2)} = \rho\widehat{\eta}_1 + \alpha\widehat{\eta}_2.$$

Поскольку $\alpha^2 + \rho^2 = 1$, все четыре t-статистики имеют нормальное распределение, с единичной дисперсией и при соответствующей нулевой гипотезе нулевым средним. Из тех же соображений, по которым мы ранее получили независимость $\widehat{\beta}_r$ и $\widehat{\eta}$, получаем, что независимы $t_{(1)}$ и t_2, $t_{(2)}$ и t_1. Далее

$$\mathrm{Corr}(t_1, t_{(1)}) = \mathrm{Corr}(t_2, t_{(2)}) = \sqrt{1 - r^2} > 0,$$

и

$$\mathrm{Corr}(t_1, t_2) = -r, \qquad \mathrm{Corr}(t_{(1)}, t_{(2)}) = r.$$

Наконец,

$$|t_1| > |t_2| \iff |t_{(1)}| > |t_{(2)}| \iff |\widehat{\eta}_1| > |\widehat{\eta}_2|.$$

Будем называть t-статистику «значимой», если ее значение по абсолютной величине превосходит некоторое заранее выбранное (положительное) пороговое значение c, например, 1.96.

Рассмотрим две обычно используемые *pretest*-процедуры: «от общего к частному» и «от частного к общему». Введем следующие обозначения: $\mathcal{M}_0$ — модель с ограничением, $\mathcal{M}_1$ — модель только с одной вспомогательной переменной z_1 ($\gamma_2 = 0$), $\mathcal{M}_2$ — другая модель только с одной вспомогательной переменной z_2 ($\gamma_1 = 0$) и $\mathcal{M}_{12}$ — модель без ограничения. Тогда процедура «от общего к частному» (или «сверху вниз») задается следующим порядком действий:

а) оцениваем модель без ограничений, $\mathcal{M}_{12}$. Получаем t-статистики t_1 и t_2;

б) если обе t_1 и t_2 значимы, то выбираем $\mathcal{M}_{12}$;

в) в противном случае:

1) если $|t_1| > |t_2|$, оцениваем $\mathcal{M}_1$ и получаем $t_{(1)}$. Если $t_{(1)}$ значима, то выбираем $\mathcal{M}_1$, иначе — выбираем $\mathcal{M}_0$;

2) если $|t_1| \leqslant |t_2|$, оцениваем $\mathcal{M}_2$ и получаем $t_{(2)}$. Если $t_{(2)}$ значима, то выбираем $\mathcal{M}_2$, иначе — выбираем $\mathcal{M}_0$.

Аналогичным образом задается процедура «от частного к общему» (или «снизу вверх»):

а) оцениваем обе модели с частичным ограничением, $\mathcal{M}_1$ и $\mathcal{M}_2$. Получаем соответственно две t-статистики $t_{(1)}$ и $t_{(2)}$;

б) если обе $t_{(1)}$ и $t_{(2)}$ не значимы, то выбираем $\mathcal{M}_0$;

в) в противном случае оцениваем модель без ограничения, получаем t_1 и t_2 и выбираем $\mathcal{M}_{12}$, если t_1 и t_2 обе значимы;

г) во всех других случаях выбираем $\mathcal{M}_1$ (если $|t_{(1)}| > |t_{(2)}|$) или $\mathcal{M}_2$ (если $|t_{(1)}| \leqslant |t_{(2)}|$).

Соответствующие области принятия решений для значения $r = 0.8$ представлены в координатах $(\widehat{\eta}_1, \widehat{\eta}_2)$ на рис. 14.3 и 14.4.

Поскольку два случая $|t_{(1)}| \leqslant c < |t_1|$, $|t_2| \leqslant c < |t_{(2)}|$ и $|t_{(2)}| \leqslant c < |t_2|$, $|t_1| \leqslant c < |t_{(1)}|$ невозможны, видно, что две процедуры совпадают, за исключением случая, когда t_1 и t_2 значимы, а $t_{(1)}$ и $t_{(2)}$ обе не значимы. В этой ситуации процедура «от общего к частному» приводит к выбору модели без ограничения, а процедура «от частного к общему» приводит к модели с ограничением. В частном случае при $r = 0$, выполняются равенства $t_1 = t_{(1)} = \widehat{\eta}_1$ и $t_2 = t_{(2)} = \widehat{\eta}_2$, и обе процедуры предварительного отбора совпадают. При $|r| \to 1$ различие между двумя процедурами наибольшее. Вопреки кажущемуся незначительному различию между этими двумя процедурами предварительного тестирования размер эффекта «занижения» среднеквадратичного отклонения *pretest*-оценки в них может быть существенно различен.

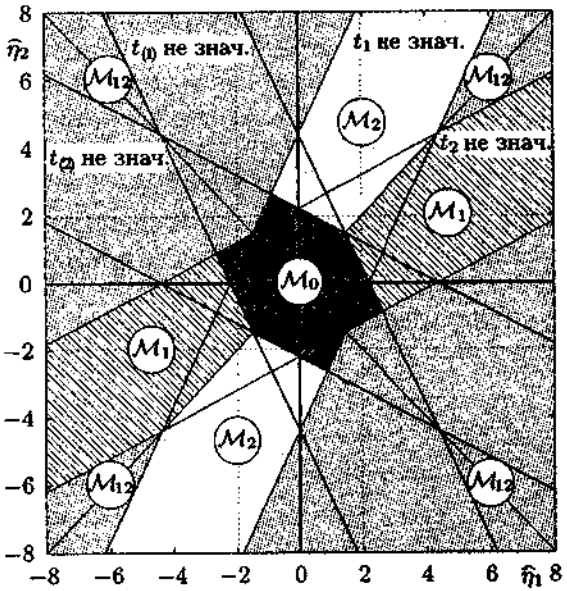


Рис. 14.3. Области выбора модели: «от общего к частному»

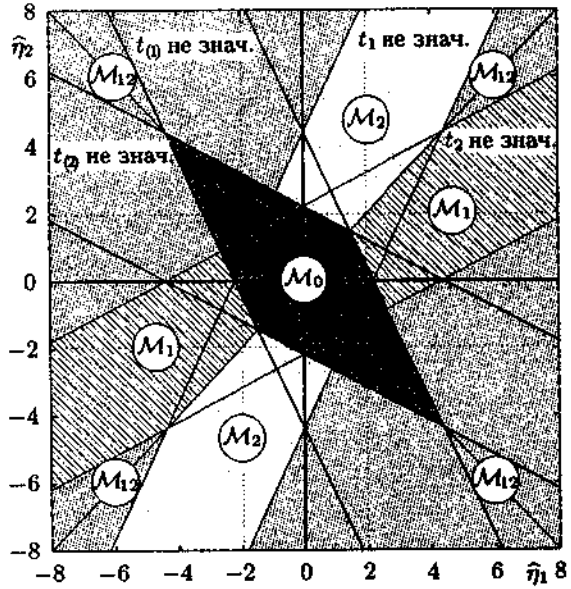


Рис. 14.4. Области выбора модели: «от частного к общему»

14.10. Эффект «занижения». Два вспомогательных параметра

В случае $m = 1$ математическое ожидание коэффициента «занижения» E(UR) зависит (при фиксированном значении c) от двух параметров: q_0^2 (известен исследователю) и η (неизвестен). В случае $m = 2$, коэффициент E(UR) зависит (после нормализации) от 5 параметров. Три из них известны исследователю: q_0^2, $\boldsymbol{q}$ (с нормировкой $\boldsymbol{q}'\boldsymbol{q} = 1$) и r, а два других η_1 и η_2 — неизвестны. Кроме того, значение зависит от процедуры предварительного тестирования.

У нас имеются четыре конкурирующие модели: модель без ограничения, $\mathcal{M}_{12}$; две модели с частичным ограничением, $\mathcal{M}_1$ ($\gamma_2 = 0$) и $\mathcal{M}_2$ ($\gamma_1 = 0$); и также модель $\mathcal{M}_0$ с ограничением ($\gamma_1 = \gamma_2 = 0$).

Поскольку E(UR) зависит от 5 параметров, мы не сможем проанализировать зависимости на графике (понадобился бы 6-мерный график). Поэтому для начала мы рассмотрим матрицу среднеквадратичных отклонений $\boldsymbol{R} = \mathrm{MSE}(\boldsymbol{W}\widehat{\eta})$ и ожидаемое значение «сообщаемого» значения матрицы ковариаций $\mathrm{E}(\boldsymbol{W})$ для двух процедур предварительного тестирования, описанных выше. Обе матрицы зависят от η_1, η_2 и r. Как показано в разделе 14.7, матрица $\mathrm{E}(\boldsymbol{W})$ всегда ограничена. Матрица $\boldsymbol{R}$ также является ограниченной в процедуре «от общего к частному», но может быть не ограничена в процедуре «от частного к общему», а именно при этой процедуре

$$\max_{\eta_1,\eta_2} \boldsymbol{R}(\eta_1, \eta_2, r) \to \infty \qquad \text{при} \quad r \to 1.$$

Такое существенное различие в поведении матрицы $\boldsymbol{R}$ при двух разных процедурах предварительного тестирования отображено на рис. 14.5, на котором построены графики зависимости

$$\mathrm{E}^{**}(\mathrm{UR}) = \max_{\eta_1,\eta_2} \mathrm{E}^{*}(\mathrm{UR}) = 1 - \min_{\eta_1,\eta_2} \min_{1 \leqslant j \leqslant m} \xi_j(\boldsymbol{R}^{-1/2}(\mathrm{E}\boldsymbol{W})\boldsymbol{R}^{-1/2}),$$

как функции r.

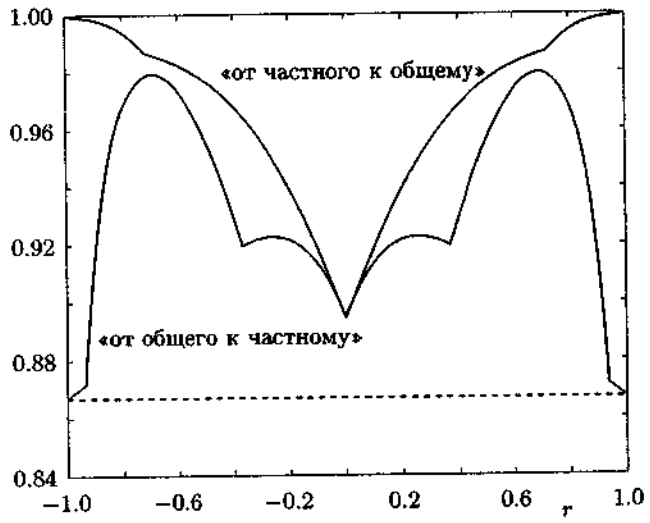


Рис. 14.5. $E^{**}(UR)$ как функция r ($m = 2$)

Для обеих процедур функция $E^{**}(UR)$ симметрична относительно точки $r = 0$. При $r = 0$ значения функций совпадают и примерно равны 0.90. В случае процедуры «от частного к общему» $E^{**}(UR)$ монотонно возрастает до 1 при возрастании r от 0 до 1. В случае процедуры «от общего к частному» функция $E^{**}(UR)$ равномерно меньше соответствующей функции для процедуры «от частного к общему», она не монотонна и при $r \to 1$ сходится к величине 0.87, совпадающей с максимальным значением функции для случая $m = 1$ (отмечено горизонтальной пунктирной линией на графике). Различие между двумя процедурами особенно велико при значениях $|r|$ близких к 1, т.е. в том случае, когда Mz_1 и Mz_2 сильно коррелированы.

Пояснить ситуацию можно следующим образом. Пусть $r = 1$ и пусть, например, $\eta_1 = -\eta_2 = \overline{\eta}$. Тогда при большом $\overline{\eta}$ вероятность выбора одной из моделей с частичным ограничением $\mathcal{M}_1$ или $\mathcal{M}_2$ близка к 0. В случае процедуры «от частного к общему» мы выберем модель с ограничением $\mathcal{M}_0$ с вероятностью, близкой к 0.95, и модель $\mathcal{M}_{12}$ с вероятностью, близкой к 0.05. Следовательно, в случае $r = 1$ и $\overline{\eta} \to \infty$ мы получаем, что E(UR) стремится к 1 при всех q_0^2. (В самом деле, среднеквадратичное отклонение *pretest*-оценки неограничено и пропорционально $\overline{\eta}^2$, при $\overline{\eta} \to \infty$.) Однако в случае процедуры «от общего к частному» среднеквадратичное отклонение *pretest*-оценки всегда ограничено, и, следовательно, $E^*(UR) < 1$ (это следует из уравнения (14.7)).

Хотя рассматриваемые функции непрерывны, у них имеются изломы. Такой вид функций является следствием существования нескольких локальных максимумов. В точках излома происходит переход с одного локального максимума на другой. Очевидно, что эффект «занижения» (среднеквадратичного отклонения) *pretest*-оценки может быть весьма значительным, даже неограниченным, в зависимости от выбора процедуры предварительного тестирования.

Зависимость от m

При $r = 0$ в наихудшем случае мы получаем $E^{**}(UR) = 0.87$ для $m = 1$ и $E^{**}(UR) = 0.90$ для $m = 2$. Теперь нас интересует, как эффект «занижения» зависит от m. У нас 2^m моделей, и можно было бы предположить, что тяжесть проблемы пропорциональна 2^m. С другой стороны, все t-статистики являются функциями только m случайных величин $\widehat{\eta}_1, \ldots, \widehat{\eta}_m$, так что, возможно, тяжесть проблемы пропорциональна только m. Рассмотрим частный случай, когда $\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z} = \boldsymbol{I}_m$. Тогда все векторы $\boldsymbol{M}\boldsymbol{z}_i$ ортогональны, и m-мерная задача распадается на m одномерных задач (см. теорему 14.3). Для этого случая график максимума $E^{**}(UR)$ как функции m приведен на рис. 14.6.

Из графика видно, что $E^{**}(UR)$ растет с ростом m, но медленнее, чем линейная функция. Хотя этот результат получен для частного случая $\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z} = \boldsymbol{I}_m$, однако он позволяет предположить, что тяжесть проблемы возрастает с ростом m не так быстро, как можно было бы опасаться.

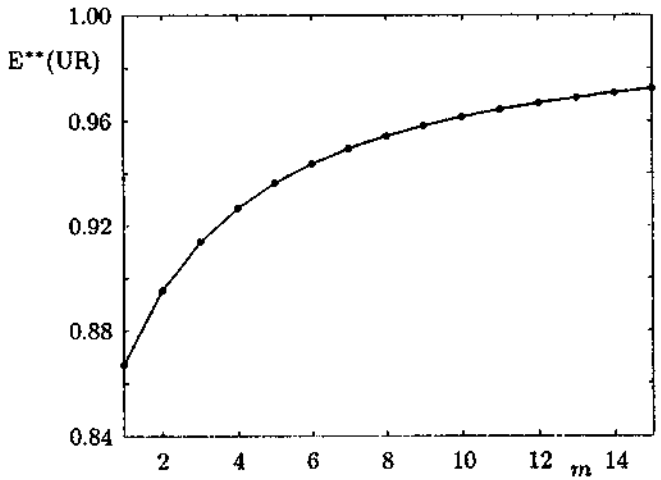


Рис. 14.6. Зависимость E**(UR) от m ($Z'MZ = I_m$)

Частный случай

На практике нам известны q_0^2, q и r, а η_1 и η_2 неизвестны. Рассмотрим одну из подобных ситуаций. Пусть, например, $q_0^2 = 2$, $q = (1/3, (2/3)\sqrt{2})'$ — нормированный вектор ($q'q = 1$) и $r = 0.8$.

На рис. 14.7 и 14.8 построены графики E(UR) как функции η_1 и η_2, для процедур «от общего к частному» и «от частного к общему», соответственно. Значение E(UR) всегда находится между 0 и 1, и каждый график симметричен относительно точки $(\eta_1, \eta_2) = (0, 0)$. Функциональная форма зависимости от (η_1, η_2) довольно сложная и существенно зависит от выбора процедуры. Для процедуры «от общего к частному» (рис. 14.7) E(UR) = 0 при $(\eta_1, \eta_2) = (4, -4)$, но может достигать значения 0.6551 в точке $(0.4, 1.6)$. Для процедуры «от частного к общему» (рис. 14.8) E(UR) изменяется от значения, близкого к 0 в точке $(4, 4)$, до значения 0.8798 вблизи точки $(4, -4)$. В данном частном случае (как и в общем случае) процедура «от частного к общему» более чувствительна к проблеме «занижения», чем процедура «от общего к частному». На рисунках изображены также линии уровня этой функции.

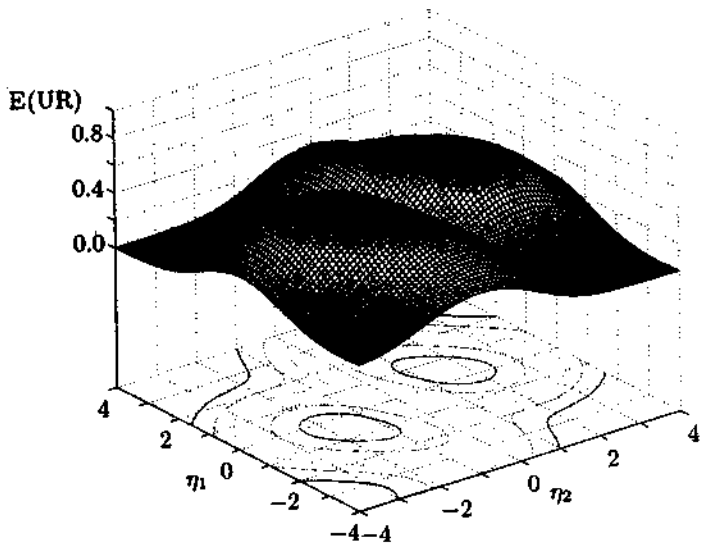


Рис. 14.7. E(UR) как функция η_1 и η_2: «от общего к частному»

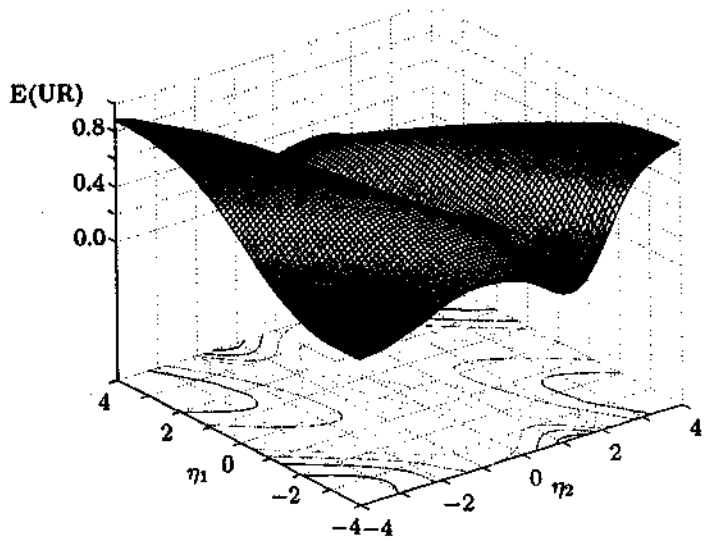


Рис. 14.8. E(UR) как функция η_1 и η_2: «от частного к общему»

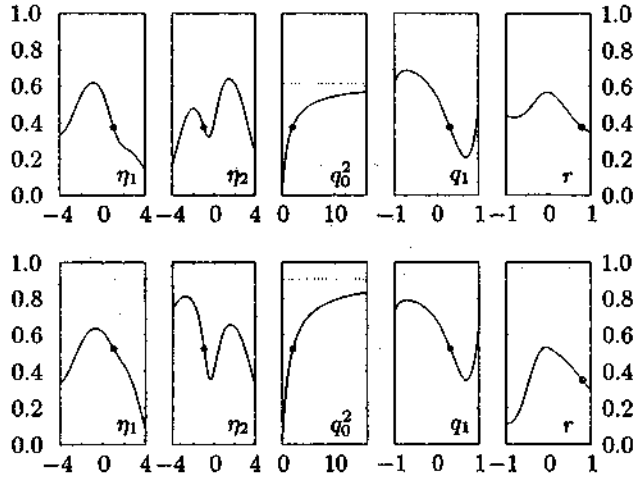


Рис. 14.9. Анализ чувствительности для E(UR). Верхний ряд: «от общего к частному»; нижний ряд: «от частного к общему»

Анализ чувствительности

Рассмотрим теперь точку $(\eta_1, \eta_2) = (1, -1)$. На рис. 14.9 показаны сечения графика в 6-мерном пространстве. В каждом сечении изменяется только один из 5 параметров: η_1, η_2, q_0^2, q_1 (первая компонента вектора $\boldsymbol{q}$), r, в то время как остальные параметры остаются равными своим значениям в рассматриваемой точке.

Для обеих процедур функция E(UR) в выбранной точке возрастает по q_0^2 (и q_2) и убывает по η_1, η_2, q_1, и r. Графики на рис. 14.9 еще раз подтверждают, что E(UR) существенно зависит от η_1 и η_2. Мы уже знаем, что E(UR) является возрастающей функцией q_0^2, однако зависимость значительно менее сильная для процедуры «от общего к частному», чем в случае процедуры «от частного к общему». Величина E(UR) зависит существенно также и от $\boldsymbol{q}$, т. е. от q_1. Следовательно, различные линейные комбинации компонент вектора $\boldsymbol{\beta}$ в различной степени затронуты эффектом процедуры предварительного тестирования. Графики чувствительности, подобные графикам на рис. 14.9, могут быть использованы для оценки степени зависимости E(UR) от неизвестных параметров η_1 и η_2, а также от возможных ошибок в измерении наблюдаемых величин q_0^2, $\boldsymbol{q}$, r.

14.11. Прогнозирование и предварительное тестирование

До сих пор мы рассматривали проблемы, связанные с оценками, получающимися в результате процедуры предварительного тестирования. Конечно, все рассмотренные выше проблемы возникают и при прогнозировании. Рассмотрим, например, стандартную линейную модель множественной регрессии:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \gamma \boldsymbol{z} + \boldsymbol{\varepsilon}.$$

Мы не уверены, следует ли включать $\boldsymbol{z}$ в уравнение. Обычная процедура такова: вычисляем t-статистику коэффициента γ, и в зависимости от того, «большое» или «малое» полученное значение $|t|$, мы решаем использовать модель без ограничения или модель с ограничением. Затем мы вычисляем прогнозное значение величины y_{n+1}, используя выбранную модель. Этот прогноз является, конечно, *pretest*-прогнозом, однако его свойства сообщаются обычно (как и в рассмотренном выше оценивании параметров модели) без учета того, что оцениванию предшествовала процедура предварительного выбора модели. Конечно, это неправильно. Следовало бы сообщать истинные значения смещения и дисперсии прогноза, принимая во внимание то, что прогнозирование и выбор модели интегрированы в одну процедуру.

Предположим, что данные порождены моделью (14.1), возможно, в ней несколько компонент вектора γ равны 0. При ограничении $\boldsymbol{S}_i'\gamma = \boldsymbol{0}$ прогноз, основанный на МНК-оценках, имеет вид

$$\begin{aligned}\widehat{y}_{n+1}^{(i)} &= \boldsymbol{x}_{n+1}'\widehat{\boldsymbol{\beta}}_{(i)} + \boldsymbol{z}_{n+1}'\widehat{\gamma}_{(i)} \\ &= \boldsymbol{x}_{n+1}'(\widehat{\boldsymbol{\beta}}_r - \boldsymbol{Q}\boldsymbol{W}_i\widehat{\boldsymbol{\theta}}) + \boldsymbol{z}_{n+1}'(\boldsymbol{Z}'\boldsymbol{M}\boldsymbol{Z})^{-1/2}\boldsymbol{W}_i\widehat{\boldsymbol{\theta}} \\ &= \boldsymbol{x}_{n+1}'\widehat{\boldsymbol{\beta}}_r - \boldsymbol{\zeta}_{n+1}'\boldsymbol{W}_i\widehat{\boldsymbol{\theta}} = \boldsymbol{x}_{n+1}'\widehat{\boldsymbol{\beta}}_r - \sigma\boldsymbol{\zeta}_{n+1}'\boldsymbol{W}_i\widehat{\boldsymbol{\eta}},\end{aligned}$$

где

$$\zeta_{n+1} = Q'x_{n+1} - (Z'MZ)^{-1/2}z_{n+1},$$

а x_{n+1} и z_{n+1} обозначают набор значений основных и вспомогательных регрессоров, для которого мы ищем прогнозное значение. Поскольку модель, которая выбирается, заранее не фиксирована и зависит от выбора процедуры предварительного тестирования, то прогноз может быть основан на одной из 2^m возможных моделей (или являться линейной комбинацией таких прогнозов). Соответственно, WALS-прогноз имеет вид

$$\widehat{y}_{n+1} = \sum_i \lambda_i \widehat{y}_{n+1}^{(i)} = x'_{n+1}\widehat{\beta}_r - \sigma\zeta'_{n+1}W\widehat{\eta}. \qquad (14.8)$$

Заметим, что вектор $\sigma\widehat{\eta} = \widehat{\theta}$ и, таким образом, является наблюдаемым, но тем не менее $\widehat{y}_{n+1}$ зависит от σ, поскольку W зависит от σ, так как λ_i зависят от σ.

Поскольку $y_{n+1} = x'_{n+1}\beta + z'_{n+1}\gamma + \varepsilon_{n+1}$, ошибка прогноза (FE) равна

$$\begin{aligned}\mathrm{FE} &= \widehat{y}_{n+1} - y_{n+1}\\ &= x'_{n+1}(\widehat{\beta}_r - \beta) - \sigma\zeta'_{n+1}W\widehat{\eta} - \sigma z'_{n+1}(Z'MZ)^{-1/2}\eta - \varepsilon_{n+1}\\ &= x'_{n+1}(\widehat{\beta}_r - \beta - \sigma Q\eta) - \sigma\zeta'_{n+1}(W\widehat{\eta} - \eta) - \varepsilon_{n+1}.\end{aligned}$$

Можно доказать следующие свойства ошибки прогноза.

Теорема 14.4 (теорема эквивалентности для прогноза). *Математическое ожидание, дисперсия и среднеквадратичное отклонение WALS-прогноза равны*

$$\begin{aligned}\mathrm{E(FE)} &= -\sigma\zeta'_{n+1}\mathrm{E}(W\widehat{\eta} - \eta),\\ \mathrm{V(FE)} &= \sigma^2(x'_{n+1}(X'X)^{-1}x_{n+1} + \zeta'_{n+1}\,\mathrm{V}(W\widehat{\eta})\zeta_{n+1} + 1),\\ \mathrm{MSE(FE)} &= \sigma^2(x'_{n+1}(X'X)^{-1}x_{n+1} + \zeta'_{n+1}\mathrm{MSE}(W\widehat{\eta})\zeta_{n+1} + 1).\end{aligned}$$

Доказательство. Существенно то, что $\widehat{\beta}_r$ и My независимы, поскольку они имеют совместное нормальное распределение и некоррелированы, так как $MX = 0$. Отсюда вытекает независимость $\widehat{\beta}_r$ и $W\widehat{\eta}$. Следовательно, величины $\widehat{\beta}_r$, $W\widehat{\eta}$, ε_{n+1} попарно независимы. Доказательство получается из этого утверждения прямым вычислением.

Подчеркнем два важных вывода из теоремы 14.4. Во-первых, она дает точные выражения для двух первых моментов ошибки прогноза, откуда видно, что эти моменты зависят только от η и σ^2 и не зависят от β. Во-вторых, теорема помогает найти *оптимальный* прогноз. Если мы найдем λ_i, такие что $W\widehat{\eta}$ является оптимальной оценкой η (в смысле наименьшей матрицы среднеквадратичных отклонений), то те же λ_i дадут оптимальный прогноз величины y_{n+1}. Это те же самые λ_i, которые дают оптимальную WALS-оценку вектора β.

Теорема 14.4 приводит истинные значения моментов ошибки прогноза, учитывая, что была произведена процедура предварительного тестирования. Однако в прикладных работах предварительное тестирование не принимается во внимание. В работах считается, что прогноз несмещенный, и приводится матрица среднеквадратичных отклонений (ковариационная матрица), рассчитанная по формуле (см. (7.8))

$$\widehat{\mathrm{MSE}}(\mathrm{FE}) = \sigma^2(x'_{n+1}(X'X)^{-1}x_{n+1} + \zeta'_{n+1}W\zeta_{n+1} + 1),$$

соответственно, приводится следующий 95%-ный доверительный интервал для y_{n+1}

$$\widehat{y}_{n+1} \pm 1.96\,\sigma\sqrt{x'_{n+1}(X'X)^{-1}x_{n+1} + \zeta'_{n+1}W\zeta_{n+1} + 1}, \qquad (14.9)$$

где вместо σ используется некоторая состоятельная оценка $\widehat{\sigma}$. Если же правильно учитывать эффект процедуры выбора модели, то мы получаем то же самое значение прогноза $\widehat{y}_{n+1}$, но совсем другие моменты. Определим две функции $\psi_1(\eta)$ и $\psi_2(\eta)$ следующим образом:

$$\begin{aligned}\psi_1(\eta) &= \zeta'_{n+1}\mathrm{E}(W\widehat{\eta} - \eta),\\ \psi_2(\eta) &= x'_{n+1}(X'X)^{-1}x_{n+1} + \zeta'_{n+1}\,\mathrm{V}(W\widehat{\eta})\zeta_{n+1}.\end{aligned}$$

Каждая из функций зависит также и от σ, поскольку W зависит от σ. Тогда по теореме 14.4 получаем

$$\mathrm{E(FE)} = -\sigma\psi_1(\eta), \quad \mathrm{V(FE)} = \sigma^2(\psi_2(\eta)+1).$$

Отсюда 95%-ный доверительный интервал для y_{n+1} приблизительно равен

$$\widehat{y}_{n+1} + \sigma\left(\psi_1(\eta) \pm 1.96\sqrt{\psi_2(\eta)+1}\right). \tag{14.10}$$

Интервал приближенный, поскольку распределение FE не является нормальным (однако оно является асимптотически нормальным, что и позволяет сделать это приближение). Кроме того, в отличие от (14.9) доверительный интервал зависит от неизвестных параметров η и σ. Оценить интервал можно, заменив неизвестные η и σ их оценками $\widehat{\eta}$ и $\widehat{\sigma}$.

Когда число наблюдений n становится большим, $\widehat{\sigma}$ сходится к σ, однако $\widehat{\eta}$ не сходится к η, поскольку $\mathrm{V}(\widehat{\eta}) = I_m$. Таким образом, оценка $\widehat{\eta}$ несмещенная, но не состоятельная. Для того, чтобы застраховаться от «больших» отклонений $\widehat{\eta}$ от η, можно рассмотреть более широкий интервал

$$\widehat{y}_{n+1} + \sigma C_1(\widehat{\eta}) < y_{n+1} < \widehat{y}_{n+1} + \sigma C_2(\widehat{\eta}), \tag{14.11}$$

где

$$C_1(\widehat{\eta}) = \min_{\eta\in\mathcal{H}(\widehat{\eta})}\left(\psi_1(\eta) - 1.96\sqrt{\psi_2(\eta)+1}\right),$$

$$C_2(\widehat{\eta}) = \max_{\eta\in\mathcal{H}(\widehat{\eta})}\left(\psi_1(\eta) + 1.96\sqrt{\psi_2(\eta)+1}\right).$$

Множество $\mathcal{H}$ является m-мерным кубом, заданным условиями $\mathcal{H}(\widehat{\eta}) = \{\eta : |\widehat{\eta}_i - \eta_i| < a_m,\ i = 1,\ldots,m\}$, где a_m определяются так, чтобы для стандартной нормальной величины u выполнялось условие $\mathrm{P}(|u| < a_m)^m = 0.95$.

В работе (Danilov and Magnus, 2003) эта теория применяется на практике к предсказанию доходностей акций. Авторы показывают, что в этом прикладном примере можно достаточно точно учесть эффект предварительного тестирования, *и то, что этот эффект существенен*. Игнорирование эффекта предварительного тестирования приводит к прогнозам, которые менее робастны, чем предполагается. Это и является причиной того, что в прикладной эконометрике прогнозы чаще, чем полагается, выходят за границы обычных доверительных интервалов.

14.12. Обобщения

Рассмотрим кратко три возможных обобщения.

Неизвестная дисперсия σ^2

Хотя теорема 14.1 верна независимо от того, известна дисперсия σ^2 или нет, в большей части данной главы предполагается, что σ^2 известна. Конечно, это предположение не реалистично, и следует коснуться вопроса о том, как изменяются наши результаты в случае, когда дисперсия σ^2 неизвестна. Рассмотрим в качестве примера случай, соответствующий рис. 14.3, где $m = 1$, $q_0^2 = \infty$, $c = 1.96$. В случае, когда σ^2 известна, ожидаемые значения коэффициента «занижения» E(UR) равны 0.82, 0.86, 0.79, 0.19 для η, равного соответственно 0, 1, 2, 4. В случае неизвестной дисперсии σ^2 вычисления сложнее и результат зависит от числа степеней свободы $n-k-m$. В таблице 14.1 приведена сводка результатов.

Таблица 14.1

$n-k-m$	η			
	0	1	2	4
10	0.76	0.83	0.77	0.26
30	0.80	0.85	0.78	0.22
50	0.81	0.86	0.79	0.21
∞	0.82	0.86	0.79	0.19

Видно, что эффект того, что мы оцениваем σ^2, относительно невелик, особенно в области, представляющей интерес, где $|\eta|$ принимает значения около 1 или 2. Этот пример типичен для поведения величины E(UR).

Ошибки спецификации

До сих пор мы предполагали, что модель без ограничения является моделью, порождающей данные. Это также может быть не очень реалистичным предположением. Рассмотрим теперь, что изменится, если модель, порождающая данные, содержит больше параметров. Пусть процесс, порождающий данные, имеет вид

$$y = X\beta + Z_1\gamma_1 + Z_2\gamma_2 + \epsilon,$$

однако у нас нет данных для Z_2. Теорема эквивалентности 14.2 по-прежнему применима и в этой ситуации. Поскольку Z_2 неизвестно, то процедура выбора модели происходит при ограничении $\gamma_2 = 0$. На смещение *pretest*-оценки $\widehat{\beta}$ повлияет тип данной ошибки спецификации, но дисперсия оценки останется прежней. Предположив для простоты, что $Z_1'MZ_2 = 0$, и обозначив $Q_i = (X'X)^{-1}X'Z_i(Z_i'MZ_i)^{-1/2}$ $(i = 1, 2)$, получим

$$\mathrm{E}(\omega'\widehat{\beta}) = \omega'\beta - \sigma\left(\omega'Q_1\mathrm{E}(W_1\widehat{\eta}_1 - \eta_1) - \omega'Q_2\eta_2\right).$$

Следовательно, влияние ошибочной спецификации на смещение и среднеквадратичное отклонение оценки $\omega'\widehat{\beta}$ проявляется через величину $\omega'Q_2\eta_2$, которая, конечно, неизвестна. Заметим, что абсолютная величина смещения оценки и ее среднеквадратичное отклонение в результате ошибки спецификации могут как возрасти, так и уменьшиться.

Асимптотика

Все рассмотренные в данной главе результаты верны для конечных выборок. Рассмотрим теперь асимптотическое поведение оценок при $n \to \infty$. Как обычно (ср. глава 8), предположим, что последовательность матриц $n^{-1}\left[X \quad Z\right]'\left[X \quad Z\right]$ сходится к некоторой положительно определенной матрице. Из теоремы 14.2 следует, что

$$\mathrm{E}\widehat{\beta} = \beta - \sigma(n^{1/2}Q)\mathrm{E}(W(n^{-1/2}\widehat{\eta}) - n^{-1/2}\eta),$$

$$\mathrm{V}(\widehat{\beta}) = \frac{\sigma^2}{n}\left(\left(\frac{X'X}{n}\right)^{-1} + (n^{1/2}Q)\,\mathrm{V}(W\widehat{\eta})(n^{1/2}Q)'\right).$$

Поскольку матрица W является взвешенным средним конечного числа идемпотентных матриц и $\mathrm{V}(\widehat{\eta}) = I_m$, то матрица $\mathrm{V}(W\widehat{\eta})$ ограничена при $n \to \infty$ (см. упражнение 14.5). Следовательно, оценка $\widehat{\beta}$ является состоятельной, если $n^{-1/2}\left(\mathrm{E}\left(W\widehat{\eta} - \eta\right)\right) \to 0$, или, что эквивалентно, если $\mathrm{E}\left(W\widehat{\gamma} - \gamma\right) \to 0$.

Следуя работе (Pötscher, 1991, стр. 164), назовем процедуру предварительного выбора модели (*pretest*-процедуру) *строго состоятельной*, если при увеличении количества наблюдений выбирается правильная «минимальная» модель. В общем случае *pretest*-процедуры не являются строго состоятельными. Назовем *pretest*-процедуру *слабо состоятельной*, если для любого $\gamma_i \neq 0$ $(i = 1, \ldots, m)$ вероятность того, что в результате процедуры будет выбрана модель без γ_i, стремится к 0 при $n \to \infty$. Слабо состоятельная процедура не обязательно исключает из модели γ_i в том случае, когда истинное значение $\gamma_i = 0$. Поэтому в результате мы можем получить модель, которая включает больше переменных, чем необходимо, но никогда не получим модель, в которую не включены необходимые переменные. Все соответствующие здравому смыслу *pretest*-процедуры являются слабо состоятельными. Легко показать, что слабо состоятельные процедуры приводят к состоятельным оценкам (Pötscher 1991, лемма 2). Пусть для простоты $m = 1$. Если $\gamma = 0$, то обе оценки $\widehat{\beta}_r$ и $\widehat{\beta}_u$ являются несмещенными и состоятельными, а следовательно, и их взвешенное среднее $\widehat{\beta} = \lambda\widehat{\beta}_u + (1-\lambda)\widehat{\beta}_r$ также является состоятельной оценкой при любом λ. Если $\gamma \neq 0$, то оценка $\widehat{\beta}_u$ состоятельная, а оценка $\widehat{\beta}_r$ не является состоятельной. Поскольку, однако, процедура является слабо состоятельной, то $p\lim \lambda = 1$ и, следовательно, оценка $\widehat{\beta}$ является состоятельной.

При условии состоятельности матрица ковариаций асимптотического распределения величины $n^{1/2}(\widehat{\beta} - \beta)$ получается из предела $\mathrm{V}(W\widehat{\eta})$ при $n \to \infty$. Рассмотрим случай $m = 1$. Если $\gamma \neq 0$, то $\lambda \xrightarrow{p} 1$ и $\mathrm{V}(\lambda\widehat{\eta}) \to \mathrm{V}(\widehat{\eta}) = 1$. В этом случае распределение случайной величины $\lambda\widehat{\eta} - \eta$ сходится к стандартному нормальному. Если же $\gamma = 0$, то $\widehat{\eta} \sim N(0, 1)$, и, следовательно, при $c = 1.96$ величина λ принимает значение 1 с вероятностью 0.05, а значение 0

с вероятностью 0.95, и $V(\lambda\widehat{\eta}) = 0.28$ (см. рис. 14.1). В этом случае $\lambda\widehat{\eta} - \eta$ сходится к распределению с нулевым средним и дисперсией 0.28, которое, однако, не является нормальным.

Итак, эффект «занижения» может встретиться даже в асимптотическом случае. При $m = 1$ эффект «занижения» исчезает асимптотически при $\gamma \neq 0$. Однако при $\gamma = 0$ эффект «занижения» не исчезает при увеличении количества наблюдений и ожидаемое значение коэффициента «занижения» может достигать значения $(R(0) - 0.05)/R(0) = 0.82$.

14.13. Другие вопросы

Многие вопросы не были рассмотрены в данной главе. Приведем список некоторых из них:

1) Существуют процедуры предварительного выбора моделей, отличные от рассмотренных «от общего к частному» и «от частного к общему». Как ведет себя ожидаемое значение коэффициента «занижения» E(UR) для этих процедур?

2) Все обычно используемые *pretest*-процедуры дискретные; они выбирают одну из 2^m моделей, основываясь на анализе t- и F-статистик. Можно определить также и непрерывную процедуру. Результат теоремы 14.2 верен и для таких процедур, и, возможно, они приведут к оценкам с лучшими свойствами. Первый шаг в этом направлении сделан в работе (Magnus, 2002).

3) Ожидаемое значение коэффициента «занижения» E(UR) зависит от неизвестного параметра $\boldsymbol{\eta}$. Каков наилучший способ оценки значения коэффициента «занижения»? Самый простой способ — заменить $\boldsymbol{\eta}$ на $\widehat{\boldsymbol{\eta}}$. Необходимо проанализировать свойства такой оценки. Другая идея — рассмотреть наихудший случай при наблюденных значениях $\boldsymbol{y}$, $\boldsymbol{X}$ и $\boldsymbol{Z}$, т.е. вычислить $\max_{\boldsymbol{\eta}} E(UR)$. Например, на рис. 14.7 и 14.8 максимумы равны 0.6551 (в случае «от общего к частному») и 0.8798 (в случае «от частного к общему»). Это дает представление о том, насколько серьезной может быть проблема «занижения» в конкретной ситуации.

Несмотря на эти все еще не решенные проблемы, можно тем не менее ответить на несколько вопросов. Во-первых, вопрос, поставленный в начале данной главы: мал или велик эффект игнорирования процедуры предварительного тестирования? Эффект может быть исключительно большим и сильно зависит от выбранной процедуры предварительного тестирования. Во-вторых, можно ли что-нибудь сделать с этой проблемой? Да, методы, развитые в данной главе, могут применяться на практике. Это не означает, что мы перестаем пользоваться процедурами предварительного отбора моделей. Мы просто должны правильным способом принимать во внимание эффект процедуры предварительного тестирования. В-третьих, существует ли альтернатива этой непривлекательной *pretest*-оценке? Да, представленные теоремы применимы к более общей WALS-оценке, в которой весовые коэффициенты λ_i являются гладкими функциями, а не принимают только два значения 0 и 1. Важным вопросом становится задача выбора этих функций. В работах (Magnus, 2002) и (Danilov, 2002) сделан первый шаг в данном направлении, разработаны так называемые «весовые коэффициенты Лапласа».

Обзор литературы

Первой работой о процедуре предварительного тестирования, возможно, является (Bancroft, 1944). В эконометрике процедуры предварительного тестирования изучались в работах Judge и его соавторов, например в (Judge and Bock, 1978). Более поздние обзоры содержатся в работах (Giles and Giles, 1993) и (Magnus, 1999).

Теорема эквивалентности впервые была доказана в работе (Magnus and Durbin, 1999). В работе (Danilov and Magnus, 2002) эта теорема была обобщена и рассмотрены теоретические аспекты игнорирования процедуры предварительного тестирования. В работе (Danilov and Magnus, 2003) было рассмотрено влияние процедуры предварительного тестирования на прогноз и рассмотрены прикладные аспекты. В работах (Magnus, 2002) и (Danilov, 2002) рассмотрен вопрос об альтернативах процедуре предвари-

тельного отбора, введены и изучены «весовые функции Лапласа». Асимптотические свойства *pretest*-оценок рассмотрены в работе (Pötscher, 1991).

Данная глава основана в большей части на работах (Danilov and Magnus, 2002, 2003).

Упражнения

14.1. Покажите, что $\widehat{\beta}_r$ и $\widehat{\theta}$ независимы.

14.2. Докажите теорему 14.1, используя следующие шаги.

а) Пусть $X_* = [X \quad Z]$, $\beta_*' = [\beta' \quad \gamma']$ и $R = [0 \quad S_i']$. Покажите, что МНК-оценка параметра β_* в модели $y = X_*\beta_* + \varepsilon$ при ограничении $R\beta_* = 0$ задается следующей формулой:

$$\widehat{\beta}_* = (X_*'X_*)^{-1}X_*'y - (X_*'X_*)^{-1}R'(R(X_*'X_*)^{-1}R')^{-1}R(X_*'X_*)^{-1}X_*'y.$$

б) Покажите, что

$$(X_*'X_*)^{-1} = \begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z \end{bmatrix}^{-1} = \begin{bmatrix} (X'X)^{-1} + QQ' & -Q(Z'MZ)^{-1/2} \\ -(Z'MZ)^{-1/2}Q' & (Z'MZ)^{-1} \end{bmatrix}.$$

в) Упростите и получите остальные результаты теоремы 14.1.

14.3. Покажите, что в случае $\eta = 1$ исследователю безразлично, какую из двух моделей выбрать (с ограничением или без ограничения).

14.4. Покажите, что не любой выбор функции λ приводит к естественной процедуре.

14.5. Покажите, что матрица $\mathrm{V}(W\widehat{\eta})$ ограничена при $n \to \infty$.

14.6. Выведите плотность распределения $\lambda\widehat{\eta} - \eta$ при условии $\eta = 0$, то есть $\gamma = 0$.

Глава 15

Эконометрика финансовых рынков*

В данной главе мы не будем обсуждать какие-либо теоретические аспекты эконометрики. Главное назначение этой главы — показать пример успешного применения эконометрических методов в реальной жизни.

После либерализации экономики в России возникли финансовые рынки, на которых происходят ежедневные торги по обмену валют, валютным фьючерсам, государственным облигациям, акциям предприятий и т.п. Соответственно, появилась необходимость анализа этого огромного числового материала. Интерес представляют такие вопросы, как: является ли цена того или иного финансового актива предсказуемой? Следует ли западному инвестору вкладывать средства в акции российских предприятий? Следует ли российскому инвестору кроме, скажем, американских покупать также акции и германских фирм?

Мы не ставим цель ответить на все эти вопросы, но иллюстрируем эффективность применения эконометрических методов для анализа финансовых рынков.

*Данная глава не входит в стандартный набор тем, включаемых обычно в начальный курс эконометрики и может быть опущена при первом чтении.

Структура данной главы следующая. Во введении мы вводим необходимые в дальнейшем обозначения. Во втором разделе рассматривается гипотеза эффективности рынка, возможность прогноза цен финансовых активов. В третьем разделе кратко излагается теория Марковица оптимального формирования портфеля, в четвертом и пятом разделах приводится методика тестирования необходимости включения дополнительного актива в эффективный портфель, при отсутствии и наличии безрискового актива соответственно. В последнем, шестом, разделе рассматриваются модели формирования цен финансовых активов (*assets*) и в качестве примера приведен тест эффективности управления взаимными фондами.

15.1. Введение

Мы будем рассматривать модели с дискретным временем. Пусть P_t — цена актива (акции, облигации) в момент t. *Доходностью* актива за один период времени (день, неделю, месяц, год) называется отношение дохода от владения активом $P_{t+1} - P_t$ к его начальной стоимости

$$R_{t+1} = \frac{P_{t+1} - P_t}{P_t}.$$

Часто для удобства вычислений рассматривают «ставку непрерывных процентов», или «логарифмическую доходность»

$$r_{t+1} = \ln \frac{P_{t+1}}{P_t} = p_{t+1} - p_t,$$

где p_t — логарифм цены, $p_t = \ln P_t$. Для коротких периодов времени $|R_t| \ll 1$, и поэтому простая доходность приблизительно совпадает с логарифмической: $r_t = \ln(1 + R_t) \approx R_t$. Логарифмическая доходность удобна для вычисления доходности за n периодов:

$$r_{t+n,n} = \ln \frac{P_{t+n}}{P_t} = \ln\left(\frac{P_{t+n}}{P_{t+n-1}} \cdot \frac{P_{t+n-1}}{P_{t+n-2}} \cdot \ldots \cdot \frac{P_{t+1}}{P_t}\right)$$
$$= r_{t+n} + r_{t+n-1} + \cdots + r_{t+1}.$$

Простые доходности более удобны для вычисления доходности портфеля ценных бумаг. Пусть портфель состоит из n активов,

и w_i — доля средств, вложенных в актив i с доходностью R_{it}. Тогда доходность портфеля равна

$$R_t = \sum_{i=1}^{n} w_i R_{it}.$$

В момент времени t цена актива P_t известна, однако будущая его цена P_{t+1} неизвестна. Поэтому будущую доходность r_t можно считать случайной величиной с некоторым средним $\mu_t = \mathrm{E}(r_t)$ и дисперсией $\sigma_t^2 = \mathrm{V}(r_t)$. Как показывают примеры, распределение случайной величины r_t не является нормальным, а обладает по сравнению с нормальным более «тяжелыми хвостами», т. е. вероятность больших отклонений от среднего больше, чем для нормального распределения с той же дисперсией. На рис. 15.1 представлена гистограмма распределения однодневных доходностей индекса РТС[1] за период 7 апреля 1999 г.–22 июля 2002 г., коэффициент эксцесса 6.5 больше 3, что и означает наличие «тяжелых хвостов».

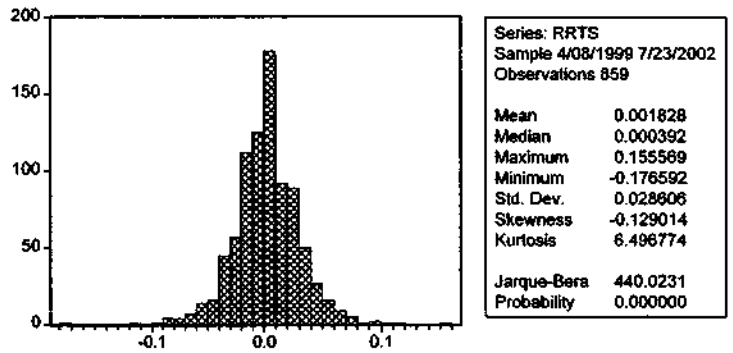


Рис. 15.1. Гистограмма однодневных доходностей индекса РТС

Таким образом, предполагая нормальное распределение доходностей, инвестор недооценивает вероятность аномально больших или аномально малых значений доходности.

[1]Российская торговая система — фондовая биржа, на которой производятся операции купли-продажи акций российских фирм.

15.2. Гипотеза эффективности финансового рынка

Впервые понятие случайного блуждания было введено для описания движения цен активов на рынке в работе (Bachelier[2], 1900), на 5 лет раньше, чем оно появилось в знаменитой работе Эйнштейна. Позднее множество работ было посвящено моделям движения цен и исследованию, является ли наблюдаемое движение цен финансовых активов случайным блужданием.

В многих работах (см., например, (Fama, 1970), (Malkiel, 1992)) обсуждается понятие эффективного рынка. Это понятие не является строго формализованным.

Гипотеза эффективного финансового рынка (*Efficient Markets Hypothesis, EMH*) может быть сформулирована следующим образом. Рынок называется эффективным, если он полностью и верно отражает всю имеющуюся в момент t информацию I_t в ценах активов. Содержательно это означает, что невозможно систематически получать прибыль выше нормальной, строя торговую стратегию на основе информационного множества I_t.

Понятие эффективного рынка близко к понятию равновесия. Здесь имеется в виду, что если бы некоторый актив был недооценен с точки зрения инвесторов, располагающих информацией I_t, то этим немедленно кто-нибудь воспользовался, и цена возросла бы до ее равновесного состояния.

В соответствии с выбором информационного множества I_t существуют три варианта EMH:

1) Слабая форма (*weak form*) EMH: I_t = вся история цен и доходностей данного актива на рынке.

2) Полустрогая форма (*semistrong form*) EMH: I_t = вся публично доступная информация.

3) Строгая форма (*strong form*) EMH: I_t = вся доступная информация, включая частную информацию, доступную лишь ограниченному кругу лиц.

[2]Подробное описание вклада Башелье в математику и теорию финансовых рынков можно найти в работе (Courtault J.-M. et al., 2000).

Можно ли тестировать гипотезу эффективного рынка? Вопрос не так прост. Подробное обсуждение его современного состояния можно найти в (Lo, 2000) и (Campbell, Lo and MacKinlay, 1997).

Дело в том, что сама формулировка EMH носит интуитивный характер и не формализуема. Однако если принять некоторую модель формирования цен (например, CAPM), то в рамках этой модели можно определить, что такое «нормальный» доход от операций на рынке, т. е. такой доход, при котором ни один из участников рынка не получает преимуществ перед другими. Если понятие «нормального» дохода определено, то тест на EMH является на самом деле совместным тестом на EMH и принятую модель формирования цен. Тестирование сводится к выяснению того, может ли инвестор получать избыточный, «сверхнормальный», доход от операций на рынке.

Если, например, определить «нормальную» доходность как некоторую константу μ в модели

$$r_t = \ln P_{t+1} - \ln P_t = \mu + \varepsilon_t, \tag{15.1}$$

то вопрос об эффективности рынка можно сформулировать как вопрос о том, является ли $\ln P_t$ случайным блужданием. Рассмотрим в качестве примера индекс РТС. Для того, чтобы проверить гипотезу случайного блуждания, можно, например, тестировать гипотезу $\beta = 0$ в следующей модели:

$$r_t = \mu + \beta r_{t-1} + \varepsilon_t. \tag{15.2}$$

В том случае, если нулевая гипотеза будет отвергнута, это будет означать, что гипотеза случайного блуждания также отвергается.

В таблице 15.1 приведены результаты оценивания этой модели для однодневных доходностей индекса РТС за период 7.04.1999 г.–22.08.2002 г[3].

[3]Данные по индексу РТС взяты с сайта `http://www.rts.ru/` и содержатся в файле `allrts.xls`.

Таблица 15.1

Dependent Variable: RRTS

Variable	*Coefficient*	*Std.Error*	*t-Statistic*	*Probability*
const	0.001728	0.001016	1.701080	0.0893
$RRTS(-1)$	0.103753	0.034711	2.989094	0.0029

R-squared	0.010766	Mean dependent var	0.001928
S.E. of regression	0.029073	S.D. dependent var	0.029213
Sum squared resid	0.693954	Akaike info criterion	−4.235568
Log likelihood	1744.936	Schwarz criterion	−4.224115

Таким образом, коэффициент β значимо отличается от нуля и гипотеза о случайном блуждании отклоняется. Конечно, надо провести более аккуратно процедуру оценивания уравнения. Тест Уайта (см. п. 6.1) показывает наличие гетероскедастичности в модели (15.2). В таблице 15.2 приведены стандартные ошибки в форме Уайта.

Таблица 15.2

Dependent Variable: RRTS

Variable	*Coefficient*	*Std.Error*	*t-Statistic*	*Probability*
const	0.001728	0.001026	1.683112	0.0927
$RRTS(-1)$	0.103753	0.043025	2.411490	0.0161

LM-тест выявляет наличие в уравнении GARCH-эффекта. В таблице 15.3 приведены результаты оценивания уравнения (15.2) с GARCH(1,1) моделью ошибок (см. п. 11.5) По-прежнему коэффициент β значимо отличается от нуля и гипотеза о случайном блуждании отклоняется.

Итак, мы получили, что однодневная доходность индекса РТС *прогнозируема*. Однако отсюда нельзя сделать вывод, что рынок акций российских предприятий не является эффективным. Дело в том, что одним из оснований современной теории финансовых рынков является необходимость баланса между ожидаемой доходностью и риском. Если инвестор берет на себя дополнительный риск, он «вознаграждается». Поэтому ответить на вопрос о том, указывает ли прогнозируемость доходности актива на неэффективность рынка, можно только после сравнения средней доходности, получаемой при использовании модели со связанным с этим риском (неопределенностью).

Таблица 15.3

Dependent Variable: RRTS

Variable	*Coefficient*	*Std.Error*	*z-Statistic*	*Probability*
const	0.001801	0.000805	2.235704	0.0254
$RRTS(-1)$	0.097411	0.036863	2.642499	0.0082
Variance Equation				
const	$3.07 \cdot 10^{-5}$	$1.39 \cdot 10^{-5}$	2.212160	0.0270
$ARCH(1)$	0.111132	0.029963	3.708923	0.0002
$GARCH(1)$	0.854125	0.034255	24.93406	0.0000

R-squared	0.010721	Mean dependent var	0.001928
S.E. of regression	0.029127	S.D. dependent var	0.029213
Sum squared resid	0.693985	Akaike info criterion	−4.363663
Log likelihood	1800.647	Schwarz criterion	−4.335030

Еще раз подчеркнем, что нет способа тестировать гипотезу эффективности рынка. Можно только тестировать совместную гипотезу эффективности рынка и некоторой модели доходности. В данном случае, в качестве модели мы взяли уравнение (15.1) и отвергли эту совместную гипотезу.

Отметим, что последняя модель (таблица 15.3) объясняет только 1% вариации завтрашнего значения однодневной доходности. Прогноз на один день вперед за границу наблюдений дает прогнозное значение однодневной доходности −0.1% и 95%-ный доверительный интервал прогноза (−4.8%, 4.6%). (В действительности следующее значение доходности, 23 июля 2002 г., было равно −1.6%).

Из литературы известно (см., например, (Lo, 2000)), что поведение доходностей разное для различных временных интервалов, за которые подсчитывается доходность. Для развитых рынков, как правило, корреляция доходностей положительна для коротких временных интервалов (однодневные, недельные доходности) и не отличается достоверно от нуля или отрицательна для более длинных временных интервалов (доходности за месяц, квартал, год).

Выше мы рассматривали однодневные доходности индекса РТС и обнаружили их положительную коррелированность. Рассмотрим теперь недельные доходности. Определим недельную доходность как

$$r_{5,t} = p_t - p_{t-5}. \tag{15.3}$$

Оценим теперь на тех же данных уравнение (15.4), аналог уравнения (15.2)

$$r_{5,t} = \mu + \beta r_{5,t-5} + \eta_t. \tag{15.4}$$

Заметим, однако, что здесь присутствует так называемая «проблема перекрывающихся отрезков» (*overlapping samples problem*). Дело в том, что

$$r_{5,t} = r_t + r_{t-1} + r_{t-2} + r_{t-3} + r_{t-4}, \tag{15.5}$$

поэтому η_t содержит в себе возмущения $\varepsilon_t, \ldots, \varepsilon_{t-4}$. Следовательно, при $k \leqslant 4$ недельные доходности $r_{5,t}$ и $r_{5,t-k}$ содержат общие возмущения, поэтому в уравнении (15.4) можно ожидать автокорреляцию ошибок η_t четвертого порядка. Соответственно, для состоятельного оценивания дисперсий МНК-оценок коэффициентов уравнения (15.4) следует применять стандартные ошибки в форме Ньюи–Веста (см. п. 6.1).

Другая проблема состоит в том, что данные нерегулярны. Есть пропуски в данных (по праздникам биржа не работает), иногда в связи с праздниками переносятся выходные дни и суббота становится рабочим днем и т.д. Выше, когда мы рассматривали однодневные доходности, мы игнорировали эту проблему и просто считали, что торговые дни идут подряд. Например, доходность в понедельник рассчитывалась как прирост индекса по сравнению с пятницей (так часто поступают, когда тонкая временная структура доходностей не является предметом исследования). Однако в построении моделей для принятия реальных финансовых решений эту нерегулярность учитывают тем или другим способом. (Можно, например, предположить, что дисперсия доходности в понедельник больше, чем в другие дни, аргументируя это тем, что на самом деле в выходные экономика не стоит на месте.)

Для начала поступим с недельными доходностями так же, как и с однодневными. Просто вычеркнем пропуски и определим «недельную» доходность как доходность за пять идущих подряд торговых дней (см. (15.3)).

Оценим уравнение (15.4), конечно, при этом используем стандартные ошибки в форме Ньюи–Веста. Результаты приведены в таблице 15.4.

Таблица 15.4

Dependent Variable: *R5RTS*

Newey-West HAC Standard Errors and Covariance (lag truncation=6)

Variable	*Coefficient*	*Std.Error*	*z-Statistic*	*Probability*
const	0.009546	0.004987	1.914190	0.0559
R5RTS(−5)	−0.002585	0.069876	−0.037000	0.9705

R-squared	0.000007	Mean dependent var	0.009522
S.E. of regression	0.071290	S.D. dependent var	0.071247
Sum squared resid	4.172510	Akaike info criterion	−2.441701
Log likelihood	1006.760	Schwarz criterion	−2.430248
Durbin-Watson stat	0.336499		

Из результатов оценивания можно сделать вывод, что недельные доходности некоррелированы с предыдущим значением.

Ошибки в уравнении (15.4) коррелированы (это можно заметить из выборочных значений ACF остатков) и по построению имеют структуру MA(4). Учитывая эту структуру, а также наличие GARCH-эффекта в уравнении, получаем, что наиболее подходящей оказывается модель, приведенная в таблице 15.5.

Уравнение, оценка которого приведена в таблице 15.5, дает более эффективный способ оценивания коэффициента β в уравнении (15.4). На это же указывают и значительно возросшее значение логарифмической функции правдоподобия и меньшие значения критериев Акаике и Шварца (см. п. 11.4). Результат, однако, тот же: нам не удалось выявить прогнозируемость недельных доходностей.

Следует подчеркнуть, что оценивание уравнения (15.4) с ARMA(1,4) моделью для ошибок является более эффективным, если именно эта модель для ошибок и есть верная модель. Как по-

Таблица 15.5

Dependent Variable: $R5RTS$

Variable	*Coefficient*	*Std.Error*	*z-Statistic*	*Probability*
const	0.010331	0.005147	2.007160	0.0447
$R5RTS(-5)$	−0.045259	0.039117	−1.157013	0.2473
$AR(1)$	0.097254	0.040560	2.397804	0.0165
$MA(1)$	0.988829	0.006943	142.4168	0.0000
$MA(2)$	0.984931	0.012031	81.86410	0.0000
$MA(3)$	0.965062	0.011387	84.75081	0.0000
$MA(4)$	0.978367	0.006391	153.0775	0.0000
Variance Equation				
const	$2.77 \cdot 10^{-5}$	$9.95 \cdot 10^{-6}$	2.782892	0.0054
$ARCH(1)$	0.106619	0.020391	5.228640	0.0000
$GARCH(1)$	0.863495	0.025436	33.94805	0.0000

R-squared	0.831180	Mean dependent var	0.009522
S.E. of regression	0.029435	S.D. dependent var	0.071247
Sum squared resid	0.704410	Akaike info criterion	−4.336645
Log likelihood	1794.529	Schwarz criterion	−4.279379
Durbin-Watson stat	1.981054		

казывает результат оценки того же уравнения с ARMA(4,0) моделью для ошибок, приведенный в таблице 15.6, выводы очень чувствительны к тому, какая модель ошибок рассматривается. Поэтому при тестировании прогнозируемости доходностей при наличии перекрывающихся интервалов, как правило, используют стандартные ошибки в форме Ньюи–Веста (таблица 15.4) (см., например, (Fama and French, 1988)).

Вспомним о проблеме пропущенных наблюдений. Возможно, более корректно говорить о недельных доходностях как о доходности понедельник–понедельник, вторник–вторник и т. д. Попробуем учесть это. Для этого заполним пропуски в данных, пользуясь следующими правилами. Если пропущен четверг, заполняем его данными по предыдущему дню — среде, если пропущен понедельник, то заполняем по следующему вторнику, и т. п. Оценив на таким образом скорректированных данных модель (15.4), полу-

Таблица 15.6

Dependent Variable: $R5RTS$

Variable	*Coefficient*	*Std.Error*	*z-Statistic*	*Probability*
const	0.014826	0.007844	1.890206	0.0587
$R5RTS(-5)$	−0.484490	0.035023	−13.83349	0.0000
$AR(1)$	1.009572	0.040131	25.15665	0.0000
$AR(2)$	−0.016341	0.058009	−0.281704	0.7782
$AR(3)$	−0.114422	0.053008	−2.158563	0.0309
$AR(4)$	−0.011019	0.035134	−0.313627	0.7538
Variance Equation				
const	$3.01 \cdot 10^{-5}$	$1.31 \cdot 10^{-5}$	2.299433	0.0215
$ARCH(1)$	0.107456	0.022677	4.738562	0.0000
$GARCH(1)$	0.868575	0.027567	31.50761	0.0000

R-squared	0.769671	Mean dependent var	0.009522
S.E. of regression	0.034361	S.D. dependent var	0.071247
Sum squared resid	0.961055	Akaike info criterion	−4.034120
Log likelihood	1669.040	Schwarz criterion	−3.982581
Durbin-Watson stat	1.978055		

Таблица 15.7

Dependent Variable: $R5RTS$

Newey-West HAC Standard Errors and Covariance (lag truncation=6)

Variable	*Coefficient*	*Std.Error*	*t-Statistic*	*Probability*
const	0.009220	0.004791	1.924524	0.0546
$R5RTS(-5)$	−0.009565	0.071298	−0.134154	0.8933

R-squared	0.000092	Mean dependent var	0.009133
S.E. of regression	0.069733	S.D. dependent var	0.069696
Sum squared resid	4.162506	Akaike info criterion	−2.485948
Log likelihood	1068.472	Schwarz criterion	−2.474865
Durbin-Watson stat	0.339157		

чаем результат, не слишком отличающийся от предыдущего (ср. таблицы 15.7 и 15.4):

Таким образом, мы получили положительную корреляцию однодневных доходностей и статистически незначимую корреляцию недельных доходностей индекса РТС.

В некоторых работах положительную корреляцию однодневных доходностей объясняют как эффект «несинхронной торговли» (*nonsynchronous trading*). Предположим, в индекс входят две акции A и B. В данный день последняя сделка по акциям A была за 2 часа до закрытия биржи, а по акциям B — в последний момент перед закрытием. Однодневные доходности вычисляются по значениям индекса на момент закрытия, где в качестве цены акции на момент закрытия берется цена последней сделки. Соответственно интервалы времени, по которым вычисляются сегодняшняя доходность акции B и завтрашняя доходность акции A, пересекаются, что может приводить к положительной корреляции однодневных доходностей биржевого индекса. Более подробное обсуждение этого вопроса можно найти в (Campbell, Lo and MacKinlay, 1997, глава 3).

15.3. Оптимизация портфеля ценных бумаг

Рассмотрим теперь задачу оптимизации портфеля ценных бумаг. Пусть у инвестора есть портфель из n активов, и w_i — доля средств, вложенных в актив i. Мы предполагаем, что w_i может быть отрицательно[4]. Будем рассматривать только два момента времени: $t = 0$ и $t = 1$. Обозначим через $\boldsymbol{r} = (R_1, \ldots, R_n)'$ вектор доходностей активов за рассматриваемый период времени, а $\boldsymbol{w} = (w_1, \ldots, w_n)'$ — вектор, определяющий структуру портфеля; $\sum_{i=1}^n w_i = \boldsymbol{w}'\boldsymbol{\imath} = 1$.[5] В момент $t = 0$ доходность портфеля за один период является случайной величиной

$$R = \boldsymbol{w}'\boldsymbol{r} = \sum_{i=1}^n w_i R_i \tag{15.6}$$

с математическим ожиданием μ и дисперсией σ^2:

$$\mu = \mathrm{E}(R) = \boldsymbol{w}'\boldsymbol{m}, \qquad \sigma^2 = \mathrm{V}(R) = \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w}, \tag{15.7}$$

[4]Отрицательное значение w_i соответствует операции «короткая продажа», см. стр. 449.

[5]Напомним, что $\boldsymbol{\imath} = (1, 1, \ldots, 1)'$.

где через $\boldsymbol{m}$ и $\boldsymbol{\Sigma}$ обозначены соответственно математическое ожидание и матрица ковариаций вектора $\boldsymbol{r}$

$$\mathrm{E}(\boldsymbol{r}) = \boldsymbol{m}, \qquad \mathrm{V}(\boldsymbol{r}) = \boldsymbol{\Sigma}.$$

Построение оптимального портфеля

Рассмотрим инвестора, который стремится минимизировать свой риск (*variance*) при заданной доходности μ (*mean*) портфеля, так называемого *mean-variance investor*. Тогда задача оптимизации портфеля математически выражается как следующая задача поиска минимума функции при наличии ограничений:

$$\boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} \to \min, \quad \text{при условии} \quad \boldsymbol{w}'\boldsymbol{m} = \mu \quad \text{и} \quad \boldsymbol{w}'\boldsymbol{\imath} = 1. \tag{15.8}$$

Чтобы решить эту задачу, рассмотрим функцию Лагранжа:

$$L(\boldsymbol{w}) = \boldsymbol{w}'\boldsymbol{\Sigma}\boldsymbol{w} - 2\lambda(\boldsymbol{w}'\boldsymbol{m} - \mu) - 2\delta(\boldsymbol{w}'\boldsymbol{\imath} - 1). \tag{15.9}$$

Условия минимума первого порядка выглядят следующим образом:

$$\boldsymbol{\Sigma}\boldsymbol{w} - \lambda\boldsymbol{m} - \delta\boldsymbol{\imath} = 0, \qquad \boldsymbol{w} = \boldsymbol{\Sigma}^{-1}(\lambda\boldsymbol{m} + \delta\boldsymbol{\imath}). \tag{15.10}$$

Подставив это выражение для $\boldsymbol{w}$ в ограничения, получаем систему двух линейных уравнений с двумя неизвестными λ и δ:

$$\boldsymbol{m}'\boldsymbol{\Sigma}^{-1}(\lambda\boldsymbol{m} + \delta\boldsymbol{\imath}) = \mu, \qquad \boldsymbol{\imath}'\boldsymbol{\Sigma}^{-1}(\lambda\boldsymbol{m} + \delta\boldsymbol{\imath}) = 1. \tag{15.11}$$

Решение системы (15.11) имеет вид

$$\lambda = \frac{C\mu - B}{AC - B^2}, \qquad \delta = \frac{A - B\mu}{AC - B^2}, \tag{15.12}$$

где $A = \boldsymbol{m}'\boldsymbol{\Sigma}^{-1}\boldsymbol{m}$, $B = \boldsymbol{m}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\imath}$ и $C = \boldsymbol{\imath}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\imath}$. Подставляя (15.12) в (15.10), получаем оптимальный портфель с ожидаемой доходностью μ:

$$\boldsymbol{w} = \boldsymbol{\Sigma}^{-1}\left(\frac{C\mu - B}{AC - B^2}\boldsymbol{m} + \frac{A - B\mu}{AC - B^2}\boldsymbol{\imath}\right). \tag{15.13}$$

Разным значениям ожидаемой доходности μ соответствуют разные оптимальные портфели $w(\mu)$ и, соответственно, разные значения дисперсий портфелей. Поскольку из (15.13) следует, что $w(\mu)$ является линейной функцией μ, то из (15.7) видно, что дисперсия (риск) портфеля является квадратичным многочленом от μ, принимающим только положительные значения и, значит, имеющим положительный коэффициент при μ^2. Поэтому на плоскости $(\sigma,\ \mu)$ точки, соответствующие различным значениям μ, лежат на гиперболе. Эта гипербола состоит из оптимальных (эффективных) портфелей и называется *«фронт эффективных портфелей»* или *«граница эффективных портфелей»* (*efficient frontier*).

На рис. 15.2 приведены три графика фронтов эффективных портфелей для портфелей, составленных из трех (mu3), четырех (mu4) и пяти (mu5) активов. График составлен на основе данных о месячных доходностях биржевых индексов за период январь 1996 г.–июль 2002 г. Первый график включает в себя индексы DJINDUS (*Dow Jones Industrials*, промышленный индекс Доу-Джонса), NASA100 (NASDAQ 100, индекс высоких технологий) и DJEURS$ (DJ EURO STOXX $, индекс акций европейских компаний). В портфель, состоящий из 4 активов, дополнительно включен RSRTSIN, индекс PTC, а в портфель, состоящий из 5 активов, включен также DJTITAS (Dow Jones Asian Titans 50 ($), индекс акций азиатских фирм)[6].

Следует заметить, что конкретный вид графика зависит от периода наблюдений и способа оценки средней ожидаемой доходности и матрицы ковариаций. Мы здесь использовали выборочные оценки μ и Σ в предположении стационарности временных рядов. Эти оценки могут быть не слишком удачными, если на интервале наблюдений было структурное изменение или какое-то событие, нарушившее стационарность рядов доходностей.

Справа от фронта эффективных портфелей лежит множество достижимых портфелей (*feasible set*). Заметим, что из (15.13) не следует, что все w_i неотрицательны. Отрицательное значение w_i

[6]Данные содержатся в файле `monthly_data.xls`.

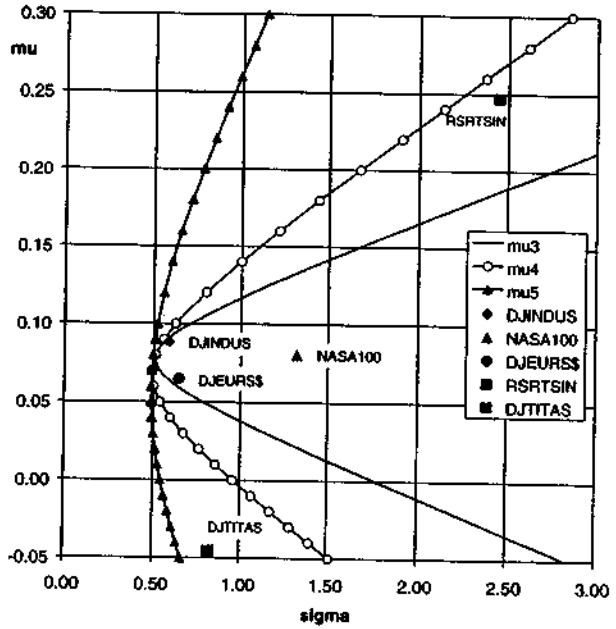


Рис. 15.2. Фронты эффективных портфелей из 3, 4 и 5 активов

соответствует операции «короткая продажа» (*short sale*), которая состоит в следующем. Инвестор занимает актив i (например, у брокерской фирмы), тут же продает его, скажем, за X_0. В следующий период времени инвестор покупает актив i по цене X_1 и возвращает долг. Если цена актива понизилась, то инвестор получает прибыль $X_0 - X_1 > 0$.

В том случае, если «короткая продажа» запрещена, т. е. все $w_i \geqslant 0$, фронт эффективных портфелей состоит из отрезков гипербол, соответствующих эффективным границам, построенных для портфелей, где некоторые $w_i = 0$.

Перечислим некоторые свойства эффективных портфелей.

1. Портфель, составленный из эффективных портфелей, также является эффективным. Любые два эффективных портфеля порождают весь фронт эффективных портфелей. Оба этих утверждения вытекают из линейной зависимости оптимального портфеля w от параметра μ (см. (15.13)).

2. Пусть g — эффективный портфель с наименьшей дисперсией. Тогда среднее, вектор весов и дисперсия его доходности равны

$$\mu_g = \mathrm{E}(R_g) = \frac{B}{C}, \quad \boldsymbol{w}_g = \frac{1}{C}\Sigma^{-1}\boldsymbol{\imath}, \quad \sigma_g^2 = \mathrm{V}(R_g) = \frac{1}{C}. \tag{15.14}$$

Эти результаты вытекают из формул (15.7) и (15.13), в частности, дисперсия доходности эффективного портфеля со средней доходностью μ равна (см. упражнение 15.1)

$$\sigma^2(\mu) = \mathrm{V}(R) = \frac{C\mu^2 - 2B\mu + A}{AC - B^2}. \tag{15.15}$$

3. Пусть есть два эффективных портфеля с ожидаемыми доходностями, μ_1 и μ_2; тогда ковариация их доходностей равна (см. упражнение 15.2)

$$\mathrm{Cov}(R_1, R_2) = \frac{C}{AC - B^2}\left(\mu_1 - \frac{B}{C}\right)\left(\mu_2 - \frac{B}{C}\right) + \frac{1}{C}. \tag{15.16}$$

15.4. Тест на включение новых активов в эффективный портфель

Из рис. 15.2 мы видим, что с включением в портфель новых активов фронт эффективных портфелей сдвигается влево. Это означает, что инвестор может сформировать эффективный портфель с той же средней доходностью μ, но с меньшей дисперсией (риском). Остается, однако, вопрос: является ли этот сдвиг статистически значимым? Имеет ли смысл инвестору включать дополнительные активы в свой портфель? В этом разделе мы приведем тест, помогающий ответить на этот вопрос.

Пусть есть инвестор, который уже инвестировал в множество X из n активов. Обозначим вектор их доходностей через $\boldsymbol{r}_X = (R_{X,1}, \ldots R_{X,n})'$. Множеству X соответствует фронт эффективных портфелей. Зададимся вопросом: сдвигается ли этот фронт, если добавляется возможность дополнительного инвестирования в набор Y из m активов? Ниже мы используем следующие обозначения: все величины, имеющие отношение к портфелю,

составленному из активов множества X (Y), имеют соответствующий индекс; величины без индекса относятся к портфелю, содержащему объединение активов из множеств X и Y. Обозначим

$$\boldsymbol{m} = \begin{bmatrix} \boldsymbol{m}_X \\ \boldsymbol{m}_Y \end{bmatrix} = \mathrm{E}(\boldsymbol{r}) = \mathrm{E}\begin{bmatrix} \boldsymbol{r}_X \\ \boldsymbol{r}_Y \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix}. \tag{15.17}$$

В силу формулы (15.13) для эффективного портфеля, составленного из всех активов, получаем

$$\Sigma\boldsymbol{w} = \frac{C\mu - B}{AC - B^2}\boldsymbol{m} + \frac{A - B\mu}{AC - B^2}\boldsymbol{\imath} = \gamma_\mu \boldsymbol{m} + \theta_\mu \boldsymbol{\imath}. \tag{15.18}$$

В том случае, если активы Y не нужны (т. е. они не участвуют в эффективном портфеле), получаем $\boldsymbol{w}_Y = \boldsymbol{0}$ и

$$\Sigma\boldsymbol{w} = \Sigma\begin{bmatrix} \boldsymbol{m}_X \\ \boldsymbol{0} \end{bmatrix} = \begin{bmatrix} \Sigma_{XX}\boldsymbol{m}_X \\ \Sigma_{XY}\boldsymbol{m}_X \end{bmatrix} = \gamma_\mu \begin{bmatrix} \boldsymbol{m}_X \\ \boldsymbol{m}_Y \end{bmatrix} + \theta_\mu \begin{bmatrix} \boldsymbol{\imath}_X \\ \boldsymbol{\imath}_Y \end{bmatrix}. \tag{15.19}$$

Выразим из первой части векторного уравнения (15.19) $\boldsymbol{m}_X$:

$$\boldsymbol{m}_X = \Sigma_{XX}^{-1}(\gamma_\mu \boldsymbol{m}_X + \theta_\mu \boldsymbol{\imath}_X) \tag{15.20}$$

и подставим полученное выражение во вторую часть уравнения (15.19):

$$\Sigma_{YX}\Sigma_{XX}^{-1}(\gamma_\mu \boldsymbol{m}_X + \theta_\mu \boldsymbol{\imath}_X) = \gamma_\mu \boldsymbol{m}_Y + \theta_\mu \boldsymbol{\imath}_Y. \tag{15.21}$$

Поскольку мы предположили, что фронты эффективных портфелей совпадают для портфелей, составленных из активов X и Y и составленных только из активов X, то равенство (15.21) должно выполняться для всех μ. Отсюда вытекает, что коэффициенты при γ_μ и θ_μ в обеих частях равенства совпадают. Получаем систему уравнений

$$\begin{cases} \Sigma_{YX}\Sigma_{XX}^{-1}\boldsymbol{m}_X = \boldsymbol{m}_Y, \\ \Sigma_{YX}\Sigma_{XX}^{-1}\boldsymbol{\imath}_X = \boldsymbol{\imath}_Y. \end{cases} \tag{15.22}$$

Покажем теперь, как с помощью регрессии можно тестировать нулевую гипотезу H_0, которая состоит в том, что дополнение порт-

феля новыми активами не приводит к изменению фронта эффективных портфелей, или, другими словами, выполняются условия (15.22).

Рассмотрим первое уравнение (15.22). Запишем его в виде $\boldsymbol{m}_Y = \boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{m}_X$, где $\boldsymbol{\alpha} = \boldsymbol{0}$ и $\boldsymbol{B} = \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}$. Заменив математические ожидания доходностей доходностями: $\boldsymbol{m}_Y = \boldsymbol{r}_Y - \boldsymbol{u}_Y$ и $\boldsymbol{m}_X = \boldsymbol{r}_X - \boldsymbol{u}_X$, где, конечно, $\mathrm{E}(\boldsymbol{u}_X) = \mathrm{E}(\boldsymbol{u}_Y) = \boldsymbol{0}$, получим

$$\boldsymbol{r}_Y = \boldsymbol{\alpha} + \boldsymbol{B}\boldsymbol{r}_X + \boldsymbol{\varepsilon}, \tag{15.23}$$

где $\boldsymbol{\varepsilon} = \boldsymbol{u}_Y - \boldsymbol{B}\boldsymbol{u}_X$. Легко проверить, что в силу H_0 векторы $\boldsymbol{\varepsilon}$ и $\boldsymbol{r}_X$ не коррелированы: $\mathrm{Cov}(\boldsymbol{\varepsilon}, \boldsymbol{r}_Y) = \mathrm{Cov}(\boldsymbol{u}_Y - \boldsymbol{B}\boldsymbol{u}_X, \boldsymbol{u}_X) = \boldsymbol{\Sigma}_{YX} - \boldsymbol{B}\boldsymbol{\Sigma}_{XX} = \boldsymbol{0}$.

Рассмотрим теперь уравнение (15.23), как уравнение регрессии $\boldsymbol{r}_Y$ на $\boldsymbol{r}_X$. Взяв ковариацию обеих частей этого уравнения с $\boldsymbol{r}_X$ и математическое ожидание от обеих частей уравнения, получим следующую систему уравнений:

$$\begin{cases} \boldsymbol{B} = \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}, \\ \boldsymbol{\alpha} = \boldsymbol{m}_Y - \boldsymbol{\Sigma}_{YX}\boldsymbol{\Sigma}_{XX}^{-1}\boldsymbol{m}_X. \end{cases} \tag{15.24}$$

Таким образом, чтобы тестировать нулевую гипотезу H_0, нам надо проверить выполнение следующих двух условий на коэффициенты в регрессионном уравнении (15.23):

$$\boldsymbol{\alpha} = \boldsymbol{0}, \qquad \boldsymbol{B}\boldsymbol{\imath}_X = \boldsymbol{\imath}_Y. \tag{15.25}$$

Для простоты рассмотрим случай, когда набор активов Y состоит из одного актива ($m = 1$), и перейдем к выборочному аналогу уравнения (15.23). Тогда, чтобы тестировать гипотезу H_0, нам надо оценить коэффициенты регрессии

$$R_{Y,t} = \alpha + \boldsymbol{\beta}'\boldsymbol{r}_{X,t} + \varepsilon_t \tag{15.26}$$

и тестировать ограничения (15.25), которые в данном случае выглядят следующим образом:

$$\alpha = 0, \qquad \boldsymbol{\beta}'\boldsymbol{\imath}_X = 1. \tag{15.27}$$

Ясен содержательный смысл этой процедуры. Уравнение (15.26) при выполнении ограничений (15.27) означает, что доходность R_Y актива Y может быть получена как доходность некоторого портфеля активов, включенных в исходный портфель, плюс некоторая дополнительная неопределенность (риск) ε. Так как $\mathrm{E}(\varepsilon) = 0$ и $\mathrm{Cov}(\varepsilon, \boldsymbol{r}_X) = \boldsymbol{0}$, то ожидаемая доходность актива R_Y равна ожидаемой доходности портфеля, а дисперсия доходности не меньше, чем дисперсия доходности портфеля, поэтому инвестор, минимизирующий риск, не станет дополнительно включать актив R_Y в свой портфель.

Подробный обзор результатов по проблеме необходимости включения дополнительных активов в портфель инвестора можно найти в (DeRoon F. and Nijman T., 2001).

В качестве иллюстрации рассмотрим представленный ранее пример (см. рис. 15.2). В таблице 15.8 приведен результат регрессии (15.26) и теста (15.27) на включение в качестве четвертого дополнительного актива индекса РТС.

Таблица 15.8

Dependent Variable: RSRTSIN

Variable	*Coefficient*	*Std.Error*	*t-Statistic*	*Probability*
const	0.063431	0.248322	0.255439	0.7991
DJINDUS	1.925889	0.568730	3.386301	0.0011
NASA100	−0.023688	0.273536	−0.086599	0.9312
DJEURS	0.237935	0.570956	0.416731	0.6781

R-squared	0.247658	Mean dependent var	0.247452
S.E. of regression	2.179939	S.D. dependent var	2.464450
Sum squared resid	356.4101	Akaike info criterion	4.445777
Log likelihood	−171.6082	Schwarz criterion	4.565749

Test Statistic	*Value*	*df*	*Probability*
F-statistic	3.557057	(2, 75)	0.0334
Chi-square	7.114113	2	0.0285

Из таблицы следует, что на 5%-ном уровне значимости гипотеза H_0 отвергается, т. е. включение актива РТС значимо сдвигает границу эффективных портфелей.

В таблице 15.9 приведен результат регрессии (15.26) и теста (15.27) на включение в качестве пятого дополнительного актива индекса DJTITAS.

Таблица 15.9

Dependent Variable: DJTITAS

Variable	*Coefficient*	*Std.Error*	*t-Statistic*	*Probability*
const	−0.122446	0.069360	−1.765383	0.0816
DJINDUS	0.476694	0.170492	2.795995	0.0066
*NASA*100	0.129076	0.076373	1.690075	0.0952
DJEURS	0.223752	0.159591	1.402036	0.1651
RSRTSIN	0.037811	0.032238	1.172867	0.2446

R-squared	0.485761	Mean dependent var	−0.045929
S.E. of regression	0.608621	S.D. dependent var	0.826671
Sum squared resid	27.41104	Akaike info criterion	1.905957
Log likelihood	−70.28531	Schwarz criterion	2.055922

Test Statistic	*Value*	*df*	*Probability*
F-statistic	2.284456	(2, 74)	0.1090
Chi-square	4.568912	2	0.1018

Из таблицы следует, что на 5%-ном уровне значимости гипотеза H_0 не отвергается, т. е. включение актива DJTITAS не сдвигает значимо границу эффективных портфелей.

Тест при фиксированной ожидаемой доходности

Рассмотрим вопрос о необходимости включения нового актива в портфель с другой точки зрения. Теперь нас будет интересовать статистическая значимость сдвига фронта эффективных портфелей при включении нового актива только в точке, соответствующей оптимальному портфелю с заданной ожидаемой доходностью μ. Значимо ли уменьшится риск при включении нового актива в портфель?

В этом случае повторяются все рассуждения, использованные для вывода уравнения (15.21), однако теперь оно должно выполняться только для одного заданного значения μ. Обозначив $\eta = -\theta_\mu/\gamma_\mu$, получим

$$\boldsymbol{m}_Y - \eta \boldsymbol{\imath}_Y = \Sigma_{YX}\Sigma_{XX}^{-1}(\boldsymbol{m}_X - \eta \boldsymbol{\imath}_X). \tag{15.28}$$

Рассмотрим следующее регрессионное уравнение (ср. (15.23)):

$$\boldsymbol{r}_Y - \eta \boldsymbol{\imath}_Y = \boldsymbol{\alpha} + \boldsymbol{B}(\boldsymbol{r}_X - \eta \boldsymbol{\imath}_X) + \boldsymbol{\varepsilon}. \tag{15.29}$$

Пусть, как и ранее, нулевая гипотеза состоит в том, что риск не изменяется при включении нового актива, т. е. его доля в новом портфеле равна 0. Легко проверить, что при выполнении H_0 векторы $\boldsymbol{\varepsilon}$ и $\boldsymbol{r}_X$ некоррелированы. Взяв математическое ожидание от обеих частей уравнения (15.29), получаем

$$\boldsymbol{\alpha} = \boldsymbol{m}_Y - \eta \boldsymbol{\imath}_Y - \boldsymbol{B}(\boldsymbol{m}_X - \eta \boldsymbol{\imath}_X), \tag{15.30}$$

т. е. при H_0 получаем $\boldsymbol{\alpha} = 0$. Как и ранее (ср. (15.26)), для простоты рассмотрим случай, когда набор активов Y состоит из одного актива ($m = 1$), и перейдем к выборочному аналогу уравнения (15.29). Для тестирования H_0 нам надо оценить регрессию

$$R_{Y,t} - \eta = \alpha + \boldsymbol{\beta}'(\boldsymbol{r}_{X,t} - \eta \boldsymbol{\imath}_X) + \varepsilon_t \tag{15.31}$$

и тестировать ограничение

$$\alpha = 0. \tag{15.32}$$

В качестве примера рассмотрим эффективный портфель с ожидаемой доходностью $\mu = 0.10$, состоящий из трех активов DJINDUS, NASA100 и DJEURS$, и тестируем необходимость включения в него дополнительного актива RSRTSIN. Получаем $\eta = 0.0534$, и результаты регрессии (15.31) приведены в таблице 15.10.

Таблица 15.10

Dependent Variable: RSRTSIN − ETA

Variable	*Coefficient*	*Std.Error*	*t-Statistic*	*Probability*
const	0.124346	0.245787	0.505908	0.6144
DJINDUS − ETA	1.925889	0.568730	3.386301	0.0011
*NASA*100 *− ETA*	−0.023688	0.273536	−0.086599	0.9312
DJEURS − ETA	0.237935	0.570956	0.416731	0.6781

R-squared	0.247658	Mean dependent var	0.194025
S.E. of regression	2.179939	S.D. dependent var	2.464450
Sum squared resid	356.4101	Akaike info criterion	4.445777
Log likelihood	−171.6082	Schwarz criterion	4.565749

Как видно из таблицы, гипотеза $\alpha = 0$ не отвергается (p-значение равно 0.61), таким образом, у нас нет необходимости

включать в эффективный портфель с ожидаемой доходностью $\mu = 0.10$ четвертый актив RSRTSIN. Однако, как мы видели ранее, фронт эффективных портфелей при этом сдвигается статистически значимо.

15.5. Оптимальный портфель при наличии безрискового актива

Рассмотрим формирование оптимального портфеля при возможности включения в портфель «безрискового» актива. Поскольку мы рассматриваем инвестиции на один период по времени, то безрисковым активом (*risk-free*) называется такой, ставка на следующий период которого, R^f, заранее фиксированна. Соответственно, дисперсия доходности σ^2 равна 0. В качестве безрискового актива обычно рассматриваются государственные ценные бумаги, например, краткосрочные государственные облигации США.

Один из способов найти фронт эффективных портфелей в случае присутствия безрискового актива аналогичен предыдущему (15.8)–(15.12); мы предоставляем читателю проделать его самостоятельно (см. упражнение 15.3).

Здесь мы приведем геометрически более наглядный вывод. На рис. 15.3 представлен фронт эффективных портфелей для тех же четырех активов, что и на рис. 15.2. Справа от гиперболы находится множество *достижимых* портфелей, т. е. таких точек плоскости (σ, μ), которые соответствуют хотя бы одному портфелю. Пусть w_p — портфель активов с ожидаемой доходностью μ_p и дисперсией доходности σ_p^2.

Рассмотрим теперь новый портфель, в котором доля средств α вложена в безрисковый актив, а доля $1-\alpha$ вложена в портфель w_p. (Очевидно, что любой портфель, включающий в себя безрисковый актив, можно представить таким образом.)

Для этого нового портфеля имеем:

$$\begin{cases} \mu = \alpha R^f + (1-\alpha)\mu_p, \\ \sigma^2 = (1-\alpha)^2\sigma_p^2; \quad \sigma = (1-\alpha)\sigma_p. \end{cases} \tag{15.33}$$

Поскольку μ и σ линейно зависят от α, то на плоскости (σ, μ) всем подобным портфелям соответствует прямая, проходящая через точки $(0, R^f)$ и (σ_p, μ_p) (см. рис. 15.3).

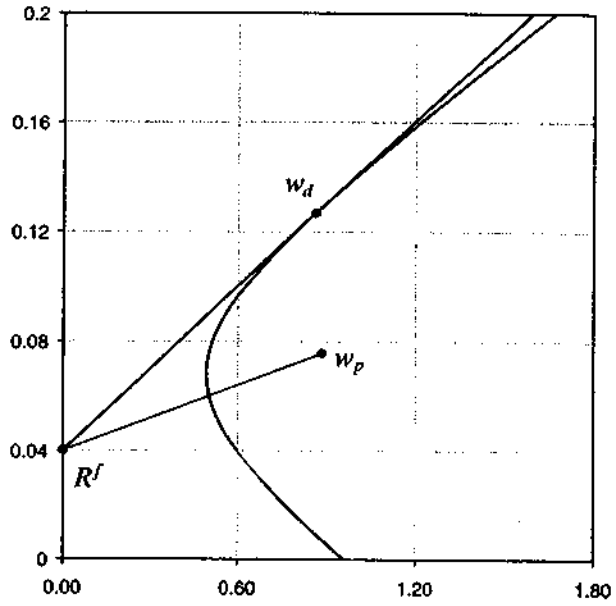


Рис. 15.3. Фронт эффективных портфелей при наличии безрискового актива

В том случае, когда w_p пробегает всю область допустимых портфелей, соответствующие прямые «заметают» на плоскости угол с вершиной в точке $(0, R^f)$, содержащий фронт эффективных портфелей. Поскольку $\mu_g > R^f$, то верхний луч, ограничивающий угол, касается гиперболы (фронта эффективных портфелей) в некоторой точке, соответствующей так называемому «тангенциальному» портфелю w_d.

Для того чтобы найти эту касательную, которая и является фронтом эффективных портфелей при наличии безрискового актива, нам теперь осталось только найти параметры тангенциального портфеля. Поступим следующим образом.

Обозначим через θ угол между касательной и осью σ. Получаем:

$$\operatorname{tg}\theta = \frac{\mu_d - R^f}{\sigma_d}. \tag{15.34}$$

Обозначим для краткости через $\boldsymbol{w} = \boldsymbol{w}_d$ — распределение ресурсов (по рисковым активам) в тангенциальном портфеле. Тогда из (15.34) получаем

$$\operatorname{tg}\theta = \frac{\boldsymbol{w}'\boldsymbol{m} - R^f}{(\boldsymbol{w}'\Sigma\boldsymbol{w})^{1/2}} = \frac{\boldsymbol{w}'(\boldsymbol{m} - R^f\boldsymbol{\imath})}{(\boldsymbol{w}'\Sigma\boldsymbol{w})^{1/2}}. \tag{15.35}$$

Заметим, что в выражении (15.35) нет необходимости в нормировке $\boldsymbol{w}'\boldsymbol{\imath} = 1$. При умножении всех компонент w_i на константу эта константа сокращается.

Поскольку угол θ соответствует касательной, то для каждого i должно выполняться условие

$$\frac{\partial \operatorname{tg}\theta}{\partial w_i} = 0.$$

Отсюда, дифференцируя (15.35), получаем

$$\begin{aligned}\frac{\partial \operatorname{tg}\theta}{\partial \boldsymbol{w}} &= \frac{\boldsymbol{m} - R^f\boldsymbol{\imath}}{(\boldsymbol{w}'\Sigma\boldsymbol{w})^{1/2}} + \left(-\frac{1}{2}\right)\frac{\boldsymbol{w}'(\boldsymbol{m} - R^f\boldsymbol{\imath})}{(\boldsymbol{w}'\Sigma\boldsymbol{w})^{3/2}}\,2\Sigma\boldsymbol{w}\\ &= \frac{(\boldsymbol{w}'\Sigma\boldsymbol{w})(\boldsymbol{m} - R^f\boldsymbol{\imath}) - \boldsymbol{w}'(\boldsymbol{m} - R^f\boldsymbol{\imath})\,\Sigma\boldsymbol{w}}{(\boldsymbol{w}'\Sigma\boldsymbol{w})^{3/2}} = \boldsymbol{0}.\end{aligned} \tag{15.36}$$

Из (15.36) вытекает, что $\Sigma\boldsymbol{w} = \lambda(\boldsymbol{m} - R^f\boldsymbol{\imath})$, где λ — некоторая константа. После нормировки получаем распределение ресурсов в тангенциальном портфеле:

$$\boldsymbol{w}_d = \frac{\Sigma^{-1}(\boldsymbol{m} - R^f\boldsymbol{\imath})}{\boldsymbol{\imath}'\Sigma^{-1}(\boldsymbol{m} - R^f\boldsymbol{\imath})}. \tag{15.37}$$

Ожидаемая доходность и дисперсия доходности тангенциального портфеля равны соответственно

$$\begin{aligned}\mu_d &= \boldsymbol{m}'\boldsymbol{w} = \frac{\boldsymbol{m}'\Sigma^{-1}(\boldsymbol{m} - R^f\boldsymbol{\imath})}{\boldsymbol{\imath}'\Sigma^{-1}(\boldsymbol{m} - R^f\boldsymbol{\imath})},\\ \sigma_d^2 &= \boldsymbol{w}_d'\Sigma\boldsymbol{w}_d = \frac{(\boldsymbol{m} - R^f\boldsymbol{\imath})'\Sigma^{-1}(\boldsymbol{m} - R^f\boldsymbol{\imath})}{\left(\boldsymbol{\imath}'\Sigma^{-1}(\boldsymbol{m} - R^f\boldsymbol{\imath})\right)^2}.\end{aligned} \tag{15.38}$$

На рис. 15.3 представлен луч, соответствующий фронту эффективных портфелей, включающих безрисковый актив с доходностью $R^f = 0.04$.

Рассмотрим теперь вопрос о том, как при наличии безрискового актива можно тестировать необходимость включения дополнительного (рискового) актива в оптимальный портфель.

Поскольку в данном случае положение эффективного фронта (касательной) определяется единственным параметром — углом θ, то вопросы тестирования значимости сдвига фронта для данного μ и сдвига фронта как целого совпадают.

Очевидно, что положение касательной не изменится тогда и только тогда, когда при включении дополнительного актива не изменится тангенциальный портфель, что означает, что старый и новый фронты эффективных портфелей без безрискового актива (гиперболы) касаются друг друга в точке, соответствующей тангенциальному портфелю $\boldsymbol{w}_d$.

Поэтому, как и выше, для тестирования нулевой гипотезы об отсутствии сдвига нам надо оценить регрессию (15.31) и тестировать ограничение H_0: $\alpha = 0$. Что взять в качестве η в уравнении (15.31)?

Заметим, что

$$\eta = -\frac{\theta_\mu}{\gamma_\mu} = \frac{B\mu - A}{C\mu - B}. \tag{15.39}$$

Покажем, что это значение равно ординате точки пересечения касательной с осью μ.

Проведем касательную к фронту эффективных портфелей (гиперболе) в некоторой точке (σ_p, μ_p), принадлежащей фронту и, следовательно, удовлетворяющей уравнению (15.15). Уравнение касательной имеет вид

$$\mu - \mu_p = \left.\frac{d\mu}{d\sigma}\right|_p (\sigma - \sigma_p). \tag{15.40}$$

Дифференцируя (15.15) по μ, получаем

$$2\sigma\frac{d\sigma}{d\mu} = \frac{1}{D}(2C\mu - 2B), \qquad D = AC - B^2.$$

Вычислив это выражение в точке (σ_p, μ_p), получаем

$$\left.\frac{d\mu}{d\sigma}\right|_p = \frac{1}{D\sigma_p}(C\mu_p - B) \tag{15.41}$$

и после подстановки в (15.40) получаем уравнение касательной:

$$\mu = \mu_p + \frac{D\sigma_p}{C\mu_p - B}(\sigma - \sigma_p). \tag{15.42}$$

Подставляя в (15.42) $\sigma = 0$ и используя равенство (15.15) в точке (σ_p, μ_p), получаем ординату пересечения касательной с осью μ:

$$\mu\Big|_{\sigma=0} = \mu_p - \frac{D\sigma_p^2}{C\mu_p - B} = \frac{B\mu_p - A}{C\mu_p - B}. \tag{15.43}$$

Полученное выражение совпадает с выражением (15.39) для η, что и требовалось показать.

Поскольку ордината пересечения касательной к точке, соответствующей тангенциальному портфелю, с осью μ равна R^f, то уравнение (15.31) в данном случае выглядит следующим образом:

$$R_{Y,t} - R^f = \alpha + \beta'(r_{X,t} - R^f \iota_X) + \varepsilon_t. \tag{15.44}$$

В качестве примера приведем в таблице 15.11 результаты теста на необходимость включения четвертого актива RSRTSIN, если в портфель уже включены три рисковых актива (DJINDUS, NASA100 и DJEURS$) и безрисковый актив с доходностью $R^f = 0.04$.

Таблица 15.11

Dependent Variable: $RSRTSIN - 0.04$

Variable	*Coefficient*	*Std.Error*	*t-Statistic*	*Probability*
const	0.109037	0.246174	0.442924	0.6591
$DJINDUS - 0.04$	1.925889	0.568730	3.386301	0.0011
$NASA100 - 0.04$	−0.023688	0.273536	−0.086599	0.9312
$DJEURS - 0.04$	0.237935	0.570956	0.416731	0.6781

Как видно из таблицы, и в этом случае гипотеза $\alpha = 0$ не отвергается (p-значение равно 0.66), таким образом, при наличии безрискового актива с доходностью $R^f = 0.04$ у нас нет необходимости включать в эффективный портфель четвертый актив RSRTSIN.

15.6. Модели оценки финансовых активов

Начнем с нового, более современного способа изложения материала, рассмотренного в предыдущих разделах[7].

Введем следующие обозначения. Пусть есть рынок ценных бумаг, каждая из них приносит какой-то доход в конце периода, при $t = 1$. Для простоты предполагается, что все бумаги бесконечно делимы. Через X будем обозначать капитал, кратное (возможно, дробное или отрицательное) некоторого портфеля. Этой же буквой будем обозначать и валовой доход, выплату (*payoff*), который этот капитал принесет в конце периода. Этот доход еще не известен в начале периода (в момент $t = 0$) и поэтому является случайной величиной. Обозначим через $\mathcal{X} = \{X\}$ множество всех мыслимых выплат (пространство выплат). Через $p(X)$ обозначим сегодняшнюю ($t = 0$) цену капитала X, т. е. p — функция на множестве $\mathcal{X}$. Сформулируем два естественных предположения:

A1) $\mathcal{X}$ является линейным пространством, т. е. если $X_1, X_2 \in \mathcal{X}$, то для любых вещественных a, b выплата $X = aX_1 + bX_2 \in \mathcal{X}$.

A2) $p(aX_1 + bX_2) = ap(X_1) + bp(X_2)$ (закон одной цены, *law of one price*). Другими словами, p — линейная функция на пространстве $\mathcal{X}$.

Эквивалентная формулировка этого утверждения такова. Если есть две выплаты, которые одинаковы при всех возможных состояниях экономики в конце периода, то их сегодняшние цены совпадают (см. упражнение 15.4). Фактически A2) означает, что мы имеем дело с рынком, уже достигшим равновесия.

Из утверждений A1) и A2) также следует, что $p(0) = 0$, т. е. существует выплата, равная 0 в конце периода, и ее цена должна быть равна 0.

Поскольку линейное пространство $\mathcal{X}$ состоит из случайных величин X, то на нем определена функция $\mathrm{E}(X)$ — математическое ожидание выплаты X. Можно показать, что функция $\mathrm{E}(XY)$ удовлетворяет всем свойствам скалярного произведения в $\mathcal{X}$.

[7] В этом разделе мы будем в основном следовать обозначениям из (Cohrane, 2001).

В самом деле, рассмотрим для простоты случай, когда количество возможных состояний экономики в момент $t = 1$ конечно, $s = 1, 2, \ldots, S$. Каждому состоянию соответствует его вероятность $p_s > 0$. Каждая выплата полностью описывается набором ее значений в каждом из возможных состояний экономики: $X = (x_1, x_2, \ldots, x_S)$, который можно интерпретировать как вектор в евклидовом пространстве R^S. Тогда

$$\mathrm{E}(XY) = \sum_{s=1}^{S} p_s x_s y_s,$$

конечно, является скалярным произведением векторов $X = (x_1, x_2, \ldots, x_S)$ и $Y = (y_1, y_2, \ldots, y_S)$ в R^S. В этом смысле можно говорить, что случайные величины X и Y *ортогональны*, если $\mathrm{E}(XY) = 0$.

Определение. Рынок называется *полным*, если $\mathcal{X}$ совпадает с R^S (это определение можно распространить и на бесконечномерный случай).

Определение. Случайная величина m называется *стохастическим дисконтирующим множителем* (*stochastic discount factor*), если цена каждого капитала X представляется в виде $p(X) = \mathrm{E}(mX)$.

Из условий A1) и A2) следует существование стохастического дисконтирующего множителя.

Теорема. *Если выполняются условия* A1) и A2), *то существует единственная выплата $X^* \in \mathcal{X}$, такая что для любой выплаты $X \in \mathcal{X}$ ее цена равна*

$$p(X) = \mathrm{E}(X^* X). \tag{15.45}$$

Доказательство. Для конечномерного случая данная теорема является хорошо известной из линейной алгебры теоремой о том, что любая линейная функция $f(\boldsymbol{x})$, определенная на евклидовом пространстве, представима в виде скалярного произведения $f(\boldsymbol{x}) = (\boldsymbol{x}, \boldsymbol{a})$ (см. упражнение 15.5). Обобщение на бесконечномерный случай называется теоремой Рисса о представлении линейного функционала.

Замечание. Пусть m — некоторый другой стохастический дисконтирующий множитель. Тогда имеет место разложение

$$m = X^* + \varepsilon, \quad \text{где для всякого } X \in \mathcal{X}, \quad \mathrm{E}(X\varepsilon) = 0. \tag{15.46}$$

Таким образом, ортогональная проекция $m_{\mathcal{X}}$ любого стохастического дисконтирующего множителя m на пространство $\mathcal{X}$ равна X^*.

Определение. (Отсутствие *арбитража*[8]). Говорят, что в пространстве выплат $\mathcal{X}$ с ценой $p(X)$ отсутствуют арбитражные возможности, если цена всякой выплаты $X \in \mathcal{X}$, такой, что $X \geqslant 0$ и $\mathrm{P}\{X > 0\} > 0$, положительна, $p(X) > 0$.

Другими словами, отсутствие арбитража означает запрет на существование такого портфеля, цена которого равна 0, отсутствует возможность отрицательных выплат, а вероятность положительных выплат положительна. Приведем без доказательства следующую теорему.

Теорема. *Арбитраж отсутствует тогда и только тогда, когда существует строго положительный стохастический дисконтирующий множитель $m > 0$.*

Прежде чем перейти к моделям оценки финансовых активов, покажем, как с помощью приведенного выше формализма получается вывод теории фронта эффективных портфелей.

Фронт эффективных портфелей

В этой секции, для краткости, будем обозначать через R «валовой доход», $R = X/p(X)$. По определению R является элементом пространства выплат $\mathcal{X}$ с ценой $p(R) = 1$. Такие элементы пространства выплат $\mathcal{X}$ будем называть «доходностями» (*returns*).

[8] Современное математическое изложение теории арбитража можно найти в (Kabanov, 2001), (Ширяев, 1998)

Введем следующие обозначения.

$\mathcal{R}^e$ — пространство «избыточных выплат»,

$$\mathcal{R}^e = \{X \in \mathcal{X} : p(X) = 0\}.$$

Выделим два специальных элемента пространства $\mathcal{X}$:

$$R^* = \frac{X^*}{p(X^*)} = \frac{X^*}{\mathrm{E}(X^{*2})}, \qquad R^{e*} = 1_{\mathcal{R}^e}.$$

Последнее равенство означает проекцию такого элемента $\mathcal{X}$, который дает выплату 1 в любом состоянии в момент $t = 1$, на подпространство $\mathcal{R}^e \subset \mathcal{X}$.

Для любого $R^e \in \mathcal{R}^e$ имеем

$$\mathrm{E}(R^e) = \mathrm{E}(1 \cdot R^e) = \mathrm{E}(1_{\mathcal{R}^e} \cdot R^e) = \mathrm{E}(R^{e*}R^e). \tag{15.47}$$

Теперь сформулируем следующую, очевидную с геометрической точки зрения, теорему.

Теорема. *Для любой «доходности» R существует следующее ортогональное разложение:*

$$R = R^* + \gamma R^{e*} + n, \tag{15.48}$$

где γ — число, n — «избыточная выплата», $\mathrm{E}(n) = 0$, и все компоненты разложения взаимно ортогональны

$$\mathrm{E}(R^*R^{e*}) = \mathrm{E}(R^*n) = \mathrm{E}(R^{e*}n) = 0.$$

Доказательство. Рассмотрим $R - R^*$. Поскольку $p(R - R^*) = p(R) - p(R^*) = 1 - 1 = 0$, то $R - R^* \in \mathcal{R}^e$. Пусть γR^{e*} есть ортогональная проекция вектора $R - R^*$ на вектор R^{e*}, тогда вектор $n = (R - R^*) - \gamma R^{e*}$ по построению ортогонален R^{e*}. Вектор R^* ортогонален подпространству $\mathcal{R}^e$, поскольку для всякого $R^e \in \mathcal{R}^e$ имеем $0 = p(R^e) = \mathrm{E}(X^*R^e) = p(X^*)\mathrm{E}(R^*R^e)$, поэтому R^* ортогонален n и R^{e*}. Из (15.47) получаем $\mathrm{E}(n) = \mathrm{E}(R^{e*}n) = 0$. Теорема доказана.

Используем разложение (15.47) для подсчета ожидаемого значения и дисперсии любой «доходности» R:

$$\begin{aligned} \mathrm{E}(R) &= \mathrm{E}(R^*) + \gamma\mathrm{E}(R^{e*}), \\ \sigma^2(R) &= \mathrm{V}(R) = \mathrm{V}(R^* + \gamma R^{e*}) + \mathrm{V}(n). \end{aligned} \tag{15.49}$$

Мы видим, что γ определяет однозначно ожидаемую доходность портфеля, а дисперсия доходности минимальна при $n = 0$. Таким образом, все эффективные портфели описываются равенством $R = R^* + \gamma R^{e*}$, и фронт эффективных портфелей (гипербола на плоскости (σ, μ)) задается парой уравнений

$$\begin{aligned} \mu &= \mathrm{E}(R^*) + \gamma\mathrm{E}(R^{e*}), \\ \sigma^2 &= \mathrm{V}(R^* + \gamma R^{e*}). \end{aligned} \tag{15.50}$$

Перейдем теперь непосредственно к теме данного раздела.

Факторные модели оценки финансовых активов

Идея факторных моделей оценки финансовых активов состоит в том, что мы используем формулу $p(X) = \mathrm{E}(mX)$ для цены актива и постулируем вид стохастического дисконтирующего фактора $m = a + \boldsymbol{b}'\boldsymbol{f}$, где a и $\boldsymbol{b}$ — постоянные число и вектор, а $\boldsymbol{f}$ — вектор *факторов*. Выбор факторов определяет модель. Например, в качестве вектора факторов можно взять

$$\boldsymbol{f} = \begin{bmatrix} R_1 - R^f \\ \vdots \\ R_K - R^f \end{bmatrix}, \tag{15.51}$$

где R_i — K активов, а R^f — ставка безрискового актива.

Модель оценки финансовых активов CAPM

Рассмотрим инвестора, обладающего портфелем с выплатой $X \in \mathcal{X}$. Возьмем ортогональную проекцию X на 1 (безрисковый актив) и X^*:

$$X = \beta_0 + \beta_1 X^* + \varepsilon, \qquad \mathrm{E}(\varepsilon) = 0, \quad \mathrm{E}(\varepsilon X^*) = 0. \tag{15.52}$$

Имеем:

$$\begin{aligned} &\mathrm{E}(X) = \beta_0 + \beta_1 \mathrm{E}(X^*), \\ &p(X) = \mathrm{E}(XX^*) = \mathrm{E}(\beta_0 X^* + \beta_1 X^{*2} + \varepsilon X^*) = p(\beta_0 + \beta_1 X^*), \\ &\mathrm{V}(X) = \mathrm{V}(\beta_0 + \beta_1 X^*) + \mathrm{V}(\varepsilon). \end{aligned} \tag{15.53}$$

Из (15.53) видно, что инвестор, минимизирующий риск, предпочтет выплату $\beta_0 + \beta_1 \mathrm{E}(X^*)$ выплате X.

Рассмотрим *всех* инвесторов $j = 1, \ldots, J$. Пусть $X^j = \beta_0^j + \beta_1^j X^*$ — выбор инвестора j. Сложив эти уравнения для всех инвесторов, получим уравнение для рыночного портфеля (*market portfolio*)

$$M = \sum_{j=1}^{J} X^j = \sum_{j=1}^{J} \beta_0^j 1 + \sum_{j=1}^{J} \beta_0^j X^* = \beta_0 + \beta_1 X^*. \tag{15.54}$$

Выразим X^* из (15.54):

$$\begin{aligned} X^* &= -\frac{\beta_0}{\beta_1} + \frac{1}{\beta_1} M = -\frac{\beta_0}{\beta_1} + \frac{p(M)}{\beta_1}\left(\frac{M}{p(M)} - R^f\right) + \frac{p(M)}{\beta_1} R^f \\ &= a + b(R^M - R^f). \end{aligned} \tag{15.55}$$

Таким образом, мы получили однофакторную модель вида (15.51), которая называется CAPM (*Capital Asset Pricing Model*).

Выведем ее так называемое β-представление. Пусть m — стохастический дисконтирующий множитель. Тогда для «доходности» R имеем

$$1 = p(R) = \mathrm{E}(mR) = \mathrm{Cov}(m, R) + \mathrm{E}(m)\mathrm{E}(R). \tag{15.56}$$

Поскольку по определению

$$R^f = \frac{1}{p(1)} = \frac{1}{\mathrm{E}(m \cdot 1)} = \frac{1}{\mathrm{E}(m)},$$

то из (15.56) получаем

$$\mathrm{E}(R) - R^f = -\frac{1}{\mathrm{E}(m)} \mathrm{Cov}(m, R) = -\frac{\mathrm{E}(m^2)}{\mathrm{E}(m)} \mathrm{Cov}(R^m, R). \tag{15.57}$$

Подставив в (15.57) R^m в качестве R, получим

$$\mathrm{E}(R^m - R^f) = -\frac{\mathrm{E}(m^2)}{\mathrm{E}(m)} \mathrm{V}(R^m). \tag{15.58}$$

И, наконец, сравнивая (15.57) и (15.58), получаем

$$\mathrm{E}(R - R^f) = \frac{\mathrm{Cov}(R^m, R)}{\mathrm{V}(R^m)} \mathrm{E}(R^m - R^f) = \beta\, \mathrm{E}(R^m - R^f). \tag{15.59}$$

Поскольку, в силу (15.54), M и стохастический дисконтирующий множитель X^* связаны линейным соотношением, мы получаем β-представление CAPM-модели

$$\mathrm{E}(R - R^f) = \beta\, \mathrm{E}(R^M - R^f), \tag{15.60}$$

где

$$\beta = \frac{\mathrm{Cov}(R^M, R)}{\mathrm{V}(R^M)} = \frac{\mathrm{Cov}(R^M - R^f, R - R^f)}{\mathrm{V}(R^M - R^f)}. \tag{15.61}$$

Возвращаясь к вопросу о «нормальной» доходности при обсуждении гипотезы эффективности рынка, мы видим, что в рамках CAPM «нормальной» доходностью является $\mathrm{E}(R) = R^f + \beta \mathrm{E}(R^M - R^f)$. Все, что выше этой величины, считается «сверхнормальной» доходностью.

Коэффициент β в (15.61) совпадает с коэффициентом β в регрессии

$$R - R^f = \alpha + \beta(R^m - R^f) + \varepsilon, \tag{15.62}$$

поэтому можно тестировать CAPM, проверяя ограничение $\alpha = 0$ в регрессии (15.62).

В качестве аппроксимации рыночного портфеля (который не наблюдаем) при тестировании обычно используют какой-нибудь индекс, включающий в себя большое количество акций, например, S&P500.

Многофакторная модель оценки финансовых активов

Многофакторная модель отличается от CAPM тем, что мы постулируем зависимость стохастического дисконтирующего множителя не от одного, а от K факторов (15.51).

Для многофакторной модели справедливы многие выводы, полученные для однофакторной. Например, аналог равенства (15.59):

$$\mathrm{E}(R - R^f) = \sum_{k=1}^{K} \beta_k \mathrm{E}(R_k - R^f). \tag{15.63}$$

Интерпретация уравнения (15.63) такая же, как и (15.59). Запишем (15.63) в виде

$$R - R^f = \sum_{k=1}^{K} \beta_k (R_k - R^f) + \varepsilon, \tag{15.64}$$

где слагаемое ε ортогонально факторам и $\mathrm{E}(\varepsilon) = 0$. Как и ранее,

$$\mathrm{V}(R - R^f) = \mathrm{V}\left(\sum_{k=1}^{K} \beta_k (R_k - R^f)\right) + \mathrm{V}(\varepsilon), \tag{15.65}$$

а

$$p(\varepsilon) = \mathrm{E}(m\varepsilon) = \mathrm{E}((a + \boldsymbol{b}'\boldsymbol{f})\varepsilon) = 0. \tag{15.66}$$

Таким образом, $\sum_{k=1}^{K} \beta_k (R_k - R^f)$ является систематической частью доходности, единственным риском, который оценивается рынком. Несистематическая часть ε увеличивает дисперсию доходности, оставляя цену портфеля неизменной. Поэтому эффективный инвестор предпочтет $\varepsilon = 0$.

Коэффициенты β_k можно оценивать из регрессионного уравнения

$$R - R^f = \alpha + \sum_{k=1}^{K} \beta_k (R_k - R^f) + \varepsilon. \tag{15.67}$$

Если модель верна, то должно выполняться равенство $\alpha = 0$.

Если применить (15.67) к активу R, цена которого еще, возможно, не пришла к равновесной, то получим

$$\alpha = \mathrm{E}(R) - \left(R^f + \sum_{k=1}^{K} \beta_k \mathrm{E}(R_k - R^f)\right). \tag{15.68}$$

Таким образом, α равно превышению действительно наблюдаемой средней доходностью актива над «нормальной» доходностью, определенной в рамках данной модели.

Можно получить оценку $\widehat{\alpha}$, оценивая коэффициент регрессии (15.67) по предыстории. Эта оценка называется «α-коэффициент Йенсена» (*Jensen's alpha*) и служит для оценки успеха данного актива R на рынке. Если $\widehat{\alpha} > 0$, то актив показывает лучшую доходность по сравнению с «нормальной» доходностью (*overperformance*) (и его стоит включить в свой портфель). Значение $\widehat{\alpha} \approx 0$ соответствует «нормальной» доходности, а $\widehat{\alpha} < 0$ соответствует доходности, меньшей, чем «нормальная» (*underperformance*).

Другой показатель, часто используемый для оценки финансового актива, — коэффициент Шарпа (*Sharpe ratio*)

$$Sh(R) = \frac{\mathrm{E}(R) - R^f}{\sigma(R)}, \tag{15.69}$$

(ср. с выражением (15.34)). Содержательный смысл его — доля ожидаемой доходности актива, приходящейся на единицу риска. Чем выше эта величина, тем лучше. В частности, исходя из этого критерия, и выбирался выше тангенциальный портфель.

Приведем теперь эмпирический пример. Часто менеджеры взаимных фондов (*mutual funds*) утверждают, что благодаря их опыту и искусству управления портфелем фонд (и его вкладчики) получает доход выше, чем рыночный. Это утверждение можно тестировать в рамках однофакторной (15.62) или многофакторной (15.67) моделей, оценивая α-коэффициент Йенсена. Ниже приведены две оценки такого рода для фонда *Bull & Bear U.S. & Overseas*, в качестве безрискового актива взяты 5-летние

государственные облигации США, в качестве рыночных портфелей MSCI индексы (США, Европа, Япония, мировой)[9]. В таблице 15.12 приведен результат оценивания модели (15.62), а в таблице 15.13 — результат оценивания модели (15.67). Оценивание производилось на интервале январь 1996 г.–декабрь 1998 г. для месячных доходностей.

Таблица 15.12

Dependent Variable: $BBUO - RFUSM$

Variable	*Coefficient*	*Std.Error*	*t-Statistic*	*Probability*
const	−1.998535	0.610759	−3.272213	0.0025
$USA - RFUSM$	1.135910	0.123267	9.215072	0.0000
R-squared	0.714088	*Durbin–Watson stat*		1.811360

Таблица 15.13

Dependent Variable: $BBUO - RFUSM$

Variable	*Coefficient*	*Std.Error*	*t-Statistic*	*Probability*
const	−2.010744	0.746870	−2.692230	0.0113
$USA - RFUSM$	0.034784	0.984426	0.035335	0.9720
$EUROPE - RFUSM$	−0.172594	0.710937	−0.242769	0.8098
$JAPAN - RFUSM$	−0.360471	0.337891	−1.066825	0.2943
$WORLD - RFUSM$	1.749027	1.939281	0.901895	0.3741
R-squared	0.761283	*Durbin–Watson stat*		1.805166

Как видно из таблиц, α-коэффициент Йенсена отрицательный и значимо отличается от 0. Аналогичный результат получается и для других взаимных фондов из данного набора. Коэффициент α получается либо отрицательный, либо незначимо отличающийся от 0. Таким образом, на этих данных, моделях и периоде времени утверждение менеджеров не подтверждается данными. Аналогичный результат дает и вычисление коэффициентов Шарпа — для фондов они оказываются ниже, чем для фондовых индексов.

[9] Данные содержатся в файлах `data_description.doc`, `mutual_fund.xls`, `index.xls`, `interest_rates.xls`.

Упражнения

15.1. Выведите формулы (15.14) и (15.15).

15.2. Выведите формулу (15.16).

15.3. Выведите формулу для оптимального портфеля при наличии безрискового актива, аналогичную формуле (15.13), повторив шаги (15.8)–(15.12) вывода формулы (15.13), т. е. записав целевую функцию, затем ограничения и решив задачу оптимизации при наличии ограничений методом Лагранжа.

15.4. Докажите эквивалентность двух определений закона одной цены (*law of one price*) (стр. 461).

15.5. Докажите, что из закона одной цены следует теорема о существовании дисконтирующего множителя (стр. 462).

15.6. В файле `index.xls` содержатся данные по четырем фондовым индексам.

а) Рассчитайте средние месячные доходности, дисперсии и корреляции доходностей индексов. В дальнейшем предположите, что доходности имеют нормальное распределение и выборочные моменты достаточно точно описывают распределение будущих доходностей.

б) Постройте эффективные портфели, состоящие из фондовых индексов США, Германии, Великобритании, с годовой ожидаемой доходностью в 6%, 9%, 12%, 15% и 18%.

в) Покажите, как информация о двух эффективных портфелях с годовой доходностью, например, 12% и 15% может быть использована для построения других эффективных портфелей.

г) Найдите портфель с наименьшей дисперсией.

д) Тот же вопрос, что в б), но теперь в случае, когда «короткие продажи» запрещены.

е) Повторите б), в), г), д) для случая портфеля, в который добавлен четвертый актив (фондовый индекс Гонконга).

ж) Повторите предыдущие пункты в предположении, что доступен также безрисковый актив с годовой доходностью 6%.

Глава 16

Перспективы эконометрики

16.1. Введение

Невозможно стать хорошим моряком, сидя дома и изучая книги по мореходному делу. Так и студент, изучивший материал этой книги, не станет сразу же хорошим эконометристом. Прикладные эконометрические исследования включают в себя гораздо больше проблем и задач, чем мы рассмотрели в этой книге. Возможно, лучшим способом обучения является стажировка у более опытного коллеги. Не следует думать, что практика эконометрики состоит просто в механическом применении известных процедур. Прикладная эконометрика — это весьма тонкое балансирование между экономической теорией, доступностью данных, предварительными идеями и, конечно, эконометрической теорией.

В этой, заключительной, главе мы обсудим, чем собственно занимается эконометрист, рассмотрим связь между эконометрикой и физикой, эконометрикой и математической экономикой, эконометрикой и математической статистикой, разрыв между теорией и практикой, методологиями «сверху вниз» и «снизу вверх», слабые звенья, агрегирование и как использовать опыт других исследований. Это попытка суммировать все то, что могло бы быть образно названо патологией эконометрики.

16.2. Чем собственно занимается эконометрист?

Представим себе океанский лайнер водоизмещением 58 000 тонн. Корабль построен 30 лет назад, на нем есть залы отдыха, рестораны, бары, кинотеатр, конференц-зал, библиотека, ночной клуб, художественная галерея и торговый центр. Сколько лет капитану?

На первый взгляд, мы не в состоянии ответить на этот вопрос, поскольку доступная нам информация не содержит никакого упоминания о возрасте капитана. Однако это не совсем так. Лайнер водоизмещением 58 000 тонн — это очень большой корабль. На его капитане лежит огромная ответственность. И только человек с большим опытом в состоянии справиться с этой работой. В то же время капитан должен обладать хорошей физической формой и быть способным справляться со стрессовыми ситуациями. Здравый смысл подсказывает, что возраст капитана — между 40 и 60 годами. Эконометрист, который хочет выглядеть наукообразно, сказал бы, что его оценка возраста капитана равна 50 со стандартным отклонением 5.

Это хороший пример эконометриста за работой. Он показывает, что доступная нам информация обычно недостаточна для решения задачи, и требуются также здравый смысл и знание жизни, а «решение» (возраст = 50, стандартное отклонение = 5) сформулировано в вероятностных терминах без достаточных на то статистических оснований.

Реальные примеры из экономики всегда труднее. Большинство эконометристов полагают, что главная цель прикладной эконометрики — сопоставление экономических теорий с наблюдаемыми явлениями. Это включает в себя проверку гипотез, например, теории монетаризма или рационального поведения потребителя. Задачей эконометриста (в идеале) было бы проверить, верна ли данная экономическая теория или нет, основываясь на экономических данных и статистическом аппарате. Никто не скажет, что это легко. Но возможно ли это? Недавно обсуждая этот вопрос,

Кезенкамп и Магнус в конце своей статьи обратились к читателям с просьбой назвать опубликованную работу, которая, по их мнению, изменила представление экономистов о какой-либо экономической проблеме (Keuzenkamp, Magnus, 1995). Если бы такая работа нашлась, она была бы примером успешной проверки теории. Никто не ответил!

16.3. Эконометрика и физика

Индивидуумы, семьи, фирмы ведут себя так иррационально и их групповое поведение настолько мало предсказуемо, что трудно предположить существование какого-либо закона, претендующего на универсальность. Это сильное утверждение, но оно верно. Причем более огорчительно даже не столько отсутствие этого закона, сколько отсутствие тех выводов, которые могли бы быть получены из него. Давайте коротко обсудим последствия одного универсального закона. Это знаменитый пример, взятый из физики и показывающий, как один верный закон может быть использован для открытия другого.

Наблюдая спутники Юпитера, можно заметить, что иногда они на восемь минут впереди, а иногда на восемь минут отстают от времени, рассчитанного по законам Ньютона. Можно также заметить, что спутники опережают расчетное время, когда Юпитер ближе к Земле, и отстают, когда Юпитер удален. Если мы доверяем теории Ньютона, то должны заключить отсюда, что свет имеет конечную скорость, и то положение спутников, которое мы наблюдаем, является вовсе не их текущим положением, а тем, в котором они находились некоторое время назад, необходимое, чтобы свет дошел от Юпитера до Земли. Олаф Ремер таким образом показал в 1675 г., что свет имеет конечную скорость, и годом позже он привел оценку скорости света 214 000 км/с, что было замечательным достижением — лишь на 30% меньше истинного значения.

Ничего подобного этой строгой красоте невозможно в эконометрике. Люди, фирмы, организации, их взаимодействие на различных уровнях агрегирования настолько богаче и интереснее

планет, что неизбежно являются гораздо более сложными объектами для моделирования и прогноза.

16.4. Эконометрика и математическая статистика

Данные в эконометрике никогда не являются экспериментальными. Не правда ли, было бы интересно удвоить цены на сахар, оставив все остальные цены неизменными, и посмотреть на реакцию потребителя? Вероятно, это и было возможно в Советском Союзе, но определенно невозможно в современной России, и, следовательно, эконометрист не может ставить подобные эксперименты.

Все параметры изменяются одновременно. Данные, с которыми мы должны работать, не являются результатом контролируемого эксперимента. Они — не экспериментальные.

В физике, химии, биологии, медицине можно проводить контролируемые эксперименты, но только не в экономике. (Астрономические данные также не являются экспериментальными: мы не можем изменить орбиту Марса, чтобы посмотреть, как это повлияет на орбиту Земли.) Отсюда следуют серьезные последствия для эконометрической теории. Традиционные методы математической статистики — теория оценивания и проверки гипотез — были развиты для экспериментальных наук, но не для экономики. Эти методы, таким образом, не могут быть без какой-либо модификации применены в эконометрике.

Так как эконометрист имеет дело с данными не экспериментальной природы, то, как правило, он не может получить больше данных, чем у него имеется, по крайней мере, при анализе временных рядов. Физики-экспериментаторы и представители других экспериментальных наук могут поступать так, как предписывают книги по статистике. У них есть теория, они собирают данные и формируют новую теорию, основанную на этих данных, затем они выбрасывают старые данные и собирают новые данные, затем проверяют новую теорию и так далее. Эконометрист

не может поступать таким образом. Если он выбросит старые данные, у него ничего не останется.

В традиционной математической статистике проверка гипотез и оценивание являются двумя разными темами, излагаемыми в разных главах разных томов. Прикладной статистик либо проверяет гипотезу, либо оценивает некоторые параметры, но никогда не делает и то и другое одновременно. Эконометрист, напротив, вынужден оценивать параметры и проверять гипотезу одновременно. Еще в 1949 г. Купманс высказал идею, что необходима совершенно новая теория статистического вывода, которая позволяла бы делать это, но такая теория до сих пор не создана (Koopmans T., 1949).

16.5. Теория и практика

Нам кажется, что разрыв между теорией и практикой в эконометрике больше, чем в физике, медицине и других науках. Когда Эд Лимер около тридцати лет назад был студентом Мичиганского университета, там работала очень активная группа, занятая построением эконометрической модели США. Группа работала на первом этаже, а теоретические курсы по эконометрике читались на последнем, четвертом, этаже. «Я был ошеломлен тем, – писал Лимер, – что один и тот же язык использовался в обоих местах. Еще более забавно было наблюдать метаморфозы отдельных индивидуумов, которые буйно грешили на первом этаже и превращались в святейших жрецов, поднимаясь на четвертый этаж» (Leamer, 1978).

За последние тридцать лет ситуация не изменилась. Конечно, есть некоторое количество экономистов и эконометристов, которые нашли верный баланс между тем, что может быть сделано теоретически, и тем, что может быть сделано на практике. Они знают данные, знают экономическую теорию, владеют необходимыми эконометрическими методами и они способны смешать эти ингредиенты в питательную и вкусную похлебку. Но их меньшинство.

Подавляющее большинство принадлежит к тому или другому лагерю. Два лагеря оторваны друг от друга и почти не общаются.

Этот разрыв между теорией и практикой имеет довольно интересные последствия. Одно из них то, что прикладные эконометристы чувствуют необходимость проверки гипотез, потому что они проходили курс «Теория эконометрики» и хотят использовать свои знания. Однако они редко могут объяснить, почему они тестируют конкретную гипотезу, скажем, однородность или выпуклость. Если гипотеза отклоняется, как и происходит в большинстве случаев, они видят в этом свидетельство ошибки спецификации. Зачем же тогда проводить тестирование, если его логические следствия игнорируются? Размышление о последствиях тестирования перед его выполнением было бы разумным, но редко встречается в эконометрической практике.

Одна история, известная среди эконометристов, иллюстрирует этот разрыв между теорией и практикой.

Ночь. Некто А идет по улице и видит другого человека, Б, что-то ищущего под уличным фонарем. «Что Вы ищете?» – спрашивает А. «Я потерял ключи», – отвечает Б. «Где Вы их потеряли?» – снова спрашивает А. «Там», – отвечает Б, указывая на тридцать метров в сторону. «Тогда почему же Вы их ищете здесь?» – спрашивает А, являясь, очевидно, человеком с практическим складом ума. «Там слишком темно», – отвечает Б и продолжает свои поиски под уличным фонарем.

16.6. Эконометрический метод

Существует ли эконометрический метод? Конечно, не такой метод, который берет за руку неискушенного исследователя и ведет его от А к В, затем к С и так далее. К счастью, такого метода не существует, иначе все эконометристы остались бы без работы. Имеется в виду нечто менее амбициозное. Существует ли единый подход к прикладным эконометрическим исследованиям? Если просмотреть статьи, опубликованные в ведущих журналах, может показаться, что такой подход существует. В самом деле,

большинство работ имеют одну и ту же структуру: введение, обзор литературы, модель, описание данных, некоторые эконометрические трудности и как автор их преодолел, прикладные результаты, заключение. Таким образом, существует некий единый подход к тому, как должны быть изложены результаты прикладной работы, но это имеет мало отношения к тому, как проводить собственно исследования. Ведь статьи никогда не содержат раздела «журнал экспериментов», в котором автор рассказал бы о том, как и в каком порядке он проводил исследование, какие ошибки делал и т. д.

Традиционная эконометрика предписывает исследователю построить модель, собрать данные, выбрать подходящий метод оценивания и затем оценить модель. После того как получена подходящая модель, можно делать шаги в разных направлениях: оценивать функции параметров (например, эластичности), проверять гипотезы, представляющие интерес, делать прогнозы или давать рекомендации по экономической политике. Это хороший метод, но он не работает. Он слишком амбициозный. В отличие от физики в экономике нет моделей, которые были бы справедливы во всех случаях. Лучшее, на что можно надеяться, — то, что модель будет справедлива локально. Это означает, что модель должна зависеть от того, на какой главный вопрос собирается ответить исследователь. Назовем его центром (фокусом) исследования. По нашему мнению, выбор центра исследования — это единственный здравый путь его начала. Все остальное: модели, необходимые данные, метод оценивания — зависит от этого. Вышесказанное может многим показаться очевидным, но это не очевидно для большинства эконометристов.

Эконометристы, конечно, имеют разные представления о том, как проводить прикладные исследования. Недавно было несколько дискуссий на эту тему. Всемирный конгресс Эконометрического общества в Кембридже (Массачусетс) в 1985 г. и Австралоазиатский конгресс в Канберре в 1988 г. содержали в своих программах пленарные дискуссии по методологии эконометрики. В ходе этих дискуссий выяснилось, что есть только один пункт,

с которым согласны все эконометристы, — это подход «сверху вниз». В этом «сверху вниз» методе нам советуют начать с большой модели, включающей большое количество переменных, и затем тестировать их значимость. Если в результате проверки статистических гипотез переменная оказывается незначимой, ее удаляют из модели. Несколько таких шагов приводят к меньшей модели, в которой остались только значимые переменные. В принципе это звучит прекрасно и имеет определенные теоретические преимущества. Но есть некоторые трудности. Первая — то, что это не работает. Если вы пытаетесь оценить модель, содержащую все мыслимые переменные, вы получаете бессмысленный результат. Это знакомо каждому, связанному с прикладной эконометрикой. Вторая трудность состоит в том, что вы не можете получить что-то новое этим методом, так как самое интересное и есть построение «верхней» модели, а о том, как это делать, ничего не сказано. Поэтому никто из прикладных эконометристов не идет таким путем. Вместо этого они используют подход «снизу вверх». При этом подходе сначала выбирается простая модель, которая затем усложняется. В сущности, так же поступают исследователи и в других областях знаний.

Таким образом, существует некоторая теория для подхода «сверху вниз», но нет теории для подхода «снизу вверх», который используется на практике. Нет теории, подсказывающей прикладному эконометристу, как переходить от малого к большому, например, как выбрать новую переменную, если он чувствует, что небольшое число переменных в простейшей модели недостаточно. Это непростая задача для эконометриста-теоретика, но это, по крайней мере, то, что прикладные исследователи могли бы использовать.

Печальным следствием этого столкновения между теорией «сверху вниз» и практикой «снизу вверх» является представление результатов. Здесь прикладной эконометрист встречает некоторое затруднение, так как свои результаты он получил с использованием подхода «снизу вверх», в то время как журнальные традиции предписывают придерживаться в статьях подхода «сверху вниз».

На практике это происходит таким образом. Сначала прикладной эконометрист «играет» с данными, начиная с небольшой модели, затем добавляет в нее компоненты до тех пор, пока не будет удовлетворен моделью. (Снизу вверх!) Затем, чтобы удовлетворить требованиям редактора журнала, он добавляет в модель переменные, которые, по его мнению, в ней должны отсутствовать. Эту расширенную модель он и представляет в начале статьи. Далее, следуя подходу «сверху вниз», он проверяет, должны ли присутствовать в модели переменные, которые он только что добавил, удаляет их и приходит к ранее полученной модели. Не слишком-то почтенный способ заниматься наукой!

16.7. Слабое звено

Попробуем представить себе, какую теорию мог бы вызвать к жизни подход «снизу вверх», учитывая, что такая теория пока отсутствует. Такая теория должна была бы, например, помогать ответить на вопрос, какие звенья модели являются наиболее слабыми. Прикладная часть работы имеет несколько аспектов: данные, экономическая теория, эконометрический метод оценивания и др. Исследователь начинает с простейшей модели и с данными, которые есть у него под рукой. Он получает некоторые результаты, которыми не вполне удовлетворен. Что дальше? Следует ли расширить модель, применить более сложную процедуру оценивания или необходимо собрать больше данных (или данные лучшего качества)? Эконометристы редко прибегают к этому простому, зачастую самому разумному, средству — улучшить данные. Вместо этого внимание сосредоточивается на модели и методах оценивания. Зачастую встречаются утонченные улучшения (например, метод максимума правдоподобия с полной информацией вместо инструментальных переменных, непараметрические методы вместо *probit*-модели) даже в тех случаях, когда используемые данные содержат много «шума», и было бы гораздо полезнее потратить усилия на добывание более надежных данных. Очевидно, должен быть баланс между различными ингредиентами в

практической работе, и самые слабые звенья должны быть найдены и укреплены. Существует ли тест, указывающий на слабое звено? Нет, такого теста не существует. Конечно, его нет, как нет и теории «снизу вверх». Поэтому и тест «слабых звеньев», как один из аспектов теории «снизу вверх», также не существует. Такой тест был бы более общим, чем поиск спецификации модели, потому что он должен был бы работать в ситуации, когда модель разумна, но в данных слишком много «шума» или же неверно выбран метод оценивания.

16.8. Агрегирование

С проблемой «слабых звеньев» связана и проблема агрегирования. Многих исследователей интересуют оценки функций параметров в макроэкономических уравнениях (например, эластичности цен). Очевидно, их можно получить из макромодели. Эти оценки также можно получить усреднением оценок, полученных из микроэкономических соотношений. На первый взгляд, второй подход предпочтительнее, хотя и требует больше данных. Оказывается, микроподход не всегда является более предпочтительным, возможно, потому что микросоотношения могут значительно отличаться от макросоотношений. Даже в том случае, когда микроподход кажется предпочтительнее, он требует больше усилий и средств. Проблема здесь в следующем: основываясь на макроданных и неполных микроданных, решить, необходимо ли собрать более полные микроданные. Решающее правило (тест) для этой ситуации в принципе можно построить. Его было бы полезно иметь, но пока оно не существует.

16.9. Как использовать другие работы

Теперь перейдем к весьма фундаментальному вопросу, связанному с агрегированием и иерархическим моделированием в том смысле, что он касается соотношения между различными частями данных. Предположим, вы собираетесь изучать задачу взаимного

замещения между капиталом, трудом и энергией в российской промышленности. Есть три тесно связанные с предметом вашего исследования работы: одна с данными по Канаде, другая — по США и третья — по Нидерландам. Чем эти работы могут помочь в вашем исследовании? Как содержащуюся в них информацию можно применить для изучения российских данных? Решением мог бы быть байесовский подход, но он имеет свои трудности. К тому же большинство из нас не являются ярыми приверженцами этого подхода. Итак, что делать? Обычная процедура такова: поскольку мы хотим, чтобы в нашей работе было как можно больше нового, мы не копируем в точности модель или методы оценивания, а пытаемся сделать что-то немного другое. Мы упоминаем имеющиеся работы во введении, но далее мы их игнорируем, пока не доходим до заключения. А тут есть два варианта: или наши результаты близки к предыдущим, или нет. В первом случае мы говорим читателю, что наша работа хорошая и основательная, подтверждающая на примере Российской Федерации то, что уже было найдено для других стран. Во втором случае, если наши результаты не совпадают с ранее полученными, мы рады сообщить читателю, что это неудивительно, так как мы используем другой метод, а то, что было сделано другими раньше, безусловно, ошибочно!

16.10. Заключение

В этой главе мы затронули некоторые аспекты, которые кажутся нам странными или ошибочными в теории и практике эконометрики. Можно было бы рассмотреть и другие вопросы, например, роль парадоксов, влияние катастроф и структурных изменений (нам необходима теория, которая могла бы быть использована для объяснения того, как наши оценки и прогнозы изменятся после структурного изменения); процедура формирования данных (почему мы не предполагаем ответственности за данные, которые используем? Если нам указывают на изъян в данных, мы отвечаем, что это вина не наша, а того, кто предоставил эти данные);

проблема линейности (в чем роль линейных моделей, кроме той очевидной, что они являются приближениями первого порядка к гладким нелинейным моделям?).

Эконометрика имеет великолепные достижения, и эконометрическая теория быстро развивается вширь и вглубь. Тем не менее, к счастью, еще многое осталось сделать.

Приложение ЛА

Линейная алгебра

Данное приложение не претендует на полноту изложения линейной алгебры. В нем лишь приведены определения основных понятий линейной алгебры, используемые в курсе эконометрики, и (без доказательств) некоторые их свойства.

1. Векторное пространство

Определение. *Вещественным векторным пространством* называется множество L, элементы которого называются векторами, удовлетворяющее следующим условиям (аксиомам).

1. Определена операция сложения векторов, результатом которой является вектор: $\boldsymbol{a}, \boldsymbol{b} \in L \Rightarrow \boldsymbol{a} + \boldsymbol{b} \in L$.
2. $\boldsymbol{a} + \boldsymbol{b} = \boldsymbol{b} + \boldsymbol{a}$ для всех $\boldsymbol{a}, \boldsymbol{b} \in L$ (коммутативность).
3. $\boldsymbol{a}+(\boldsymbol{b}+\boldsymbol{c}) = (\boldsymbol{a}+\boldsymbol{b})+\boldsymbol{c}$ для всех $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in L$ (ассоциативность).
4. Существует нулевой вектор $\mathbf{0}$, такой, что $\mathbf{0} + \boldsymbol{a} = \boldsymbol{a} + \mathbf{0} = \boldsymbol{a}$ для любого $\boldsymbol{a} \in L$.
5. Для всякого вектора $\boldsymbol{a} \in L$ и вещественного числа $\alpha \in R$ определено их произведение $\alpha\boldsymbol{a} \in L$.
6. $(\alpha + \beta)\boldsymbol{a} = \alpha\boldsymbol{a} + \beta\boldsymbol{a}$ для всех $\alpha, \beta \in R$ и $\boldsymbol{a} \in L$.

7. $\alpha(\boldsymbol{a} + \boldsymbol{b}) = \alpha\boldsymbol{a} + \alpha\boldsymbol{b}$ для всех $\alpha \in R$ и $\boldsymbol{a}, \boldsymbol{b} \in L$.
8. $0\boldsymbol{a} = \mathbf{0}$ для всех $\boldsymbol{a} \in L$.
9. $1\boldsymbol{a} = \boldsymbol{a}$ для всех $\boldsymbol{a} \in L$.

2. Векторное пространство R^n

Элементами (точками, векторами) вещественного векторного пространства R^n являются векторы-столбцы, состоящие из n вещественных чисел

$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix};$$

операции сложения и умножения на число определены следующим образом:

$$\boldsymbol{a} + \boldsymbol{b} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} + \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} = \begin{bmatrix} a_1 + b_1 \\ \vdots \\ a_n + b_n \end{bmatrix}, \quad \alpha\boldsymbol{a} = \begin{bmatrix} \alpha a_1 \\ \vdots \\ \alpha a_n \end{bmatrix}.$$

Нулевой вектор имеет все координаты, равные 0.

Упражнение. Проверьте, что таким образом определенное множество в самом деле является векторным пространством, т. е. удовлетворяет аксиомам (1)–(9) векторного пространства.

3. Линейная зависимость

Определение. Векторы $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k$ называются *линейно независимыми*, если из того, что

$$\alpha_1\boldsymbol{a}_1 + \alpha_2\boldsymbol{a}_2 + \cdots + \alpha_k\boldsymbol{a}_k = \mathbf{0}, \qquad \alpha_i \in R, \tag{ЛА.1}$$

следует, что все $\alpha_i = 0$.

Векторы $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k$ называются *линейно зависимыми*, если существует набор α_i, $i = 1, \ldots, k$, где хотя бы одно α_i отлично от 0, удовлетворяющий условию (ЛА.1).

4. Линейное подпространство

Определение. Линейным подпространством линейного пространства L называется подмножество K векторов пространства L, замкнутое относительно операций сложения и умножения на число, т. е. из того, что векторы $\boldsymbol{a}, \boldsymbol{b} \in K$, следует, что $\boldsymbol{a} + \boldsymbol{b}$ и $\alpha\boldsymbol{a}$ принадлежат K.

Пример. Рассмотрим множество векторов из R^n, состоящее из таких векторов, у которых последние $n - k$ координат равны 0. Нетрудно проверить, что это множество является линейным подпространством пространства R^n, совпадающим с R^k.

Определение. Множество всех *линейных комбинаций* векторов $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k \in L$ $\alpha_1\boldsymbol{a}_1 + \alpha_2\boldsymbol{a}_2 + \cdots + \alpha_k\boldsymbol{a}_k$, $\alpha_i \in R$ называется *пространством, порожденным векторами* $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k$. (Проверьте, что оно является линейным подпространством векторного пространства L.)

Если линейное подпространство K векторного пространства L не совпадает с ним, то его часто называют *гиперплоскостью*.

5. Базис. Размерность

Определение. Набор векторов $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n \in L$ называется *базисом* пространства L, если выполняются два условия:

1. векторы $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$ линейно независимы,
2. пространство, порожденное векторами $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$, совпадает с L.

Пример. Набор векторов $\boldsymbol{e}_i$, $i = 1, \ldots, n$, у которых все координаты, кроме i-й, равны 0, а i-я координата равна 1, является базисом в пространстве R^n.

Для записи таких векторов удобно использовать *символ Кронекера*

$$\delta_{ij} = \begin{cases} 1, & \text{если } i = j, \\ 0, & \text{если } i \neq j. \end{cases} \qquad \text{(ЛА.2)}$$

Тогда векторы из предыдущего примера можно записать как

$$\boldsymbol{e}_i = \begin{bmatrix} \delta_{i1} \\ \vdots \\ \delta_{in} \end{bmatrix}.$$

Предложение. Все базисы векторного пространства L содержат одно и то же число векторов, которое называется размерностью $\dim(L)$ векторного пространства L.

Пример. Размерность R^n равна $\dim(R^n) = n$.

Предложение. Любой вектор $\boldsymbol{a}$ линейного пространства можно *единственным* способом *разложить по базису*, т. е. представить в виде линейной комбинации базисных векторов: $\boldsymbol{a} = \alpha_1\boldsymbol{e}_1 + \alpha_2\boldsymbol{e}_2 + \cdots + \alpha_n\boldsymbol{e}_n$, $\alpha_i \in R$.

6. Линейные операторы

Определение. Линейным оператором называется отображение векторного пространства L в векторное пространство M, $\mathcal{A}\colon L \to M$, которое инвариантно относительно операций сложения векторов и умножения вектора на число:

1. $\mathcal{A}(\boldsymbol{a} + \boldsymbol{b}) = \mathcal{A}(\boldsymbol{a}) + \mathcal{A}(\boldsymbol{b})$ для всех $\boldsymbol{a}, \boldsymbol{b} \in L$,
2. $\mathcal{A}(\alpha\boldsymbol{a}) = \alpha\mathcal{A}(\boldsymbol{a})$ для всех $\alpha \in R$, $\boldsymbol{a} \in L$.

Образом оператора $\mathcal{A}\colon L \to M$ называется множество $\mathrm{Im}(\mathcal{A})$, состоящее из всех векторов $\mathcal{A}(x)$, $x \in L$. $\mathrm{Im}(\mathcal{A})$ является линейным подпространством пространства M.

Ядром оператора $\mathcal{A}$: $L \to M$ называется множество $\mathrm{Ker}(\mathcal{A})$, состоящее из всех векторов $\boldsymbol{x} \in L$, таких, что $\mathcal{A}(\boldsymbol{x}) = \mathbf{0}$. $\mathrm{Ker}(\mathcal{A})$ является линейным подпространством пространства L.

7. Матрицы

Линейный оператор полностью определяется своими значениями на базисных векторах. В самом деле, любой вектор $\boldsymbol{x}$, принадлежащий n-мерному векторному пространству L, можно разложить по базисным векторам $\boldsymbol{e}_i$: $\boldsymbol{x} = \alpha_1 \boldsymbol{e}_1 + \alpha_2 \boldsymbol{e}_2 + \cdots \alpha_n \boldsymbol{e}_n$, $\alpha_i \in R$. Тогда, используя определение линейного оператора, получаем: $\mathcal{A}(\boldsymbol{x}) = \alpha_1 \mathcal{A}(\boldsymbol{e}_1) + \alpha_2 \mathcal{A}(\boldsymbol{e}_2) + \cdots + \alpha_n \mathcal{A}(\boldsymbol{e}_n)$.

Каждый из n векторов $\mathcal{A}(\boldsymbol{e}_j)$, $j = 1, \ldots, n$, также можно разложить по базису $\boldsymbol{l}_i$, $i = 1, \ldots, m$ (в m-мерном пространстве M): $\mathcal{A}(\boldsymbol{e}_j) = \sum_{i=1}^{m} a_{ij} \boldsymbol{l}_i$.

Таким образом, линейному оператору $\mathcal{A}$ и выбранным базисам $\{\boldsymbol{e}_j\}$, $\{\boldsymbol{l}_i\}$ в пространствах L, M соответствует таблица $m \times n$ чисел

$$\boldsymbol{A} = (a_{ij}) = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix} \qquad \text{(ЛА.3)}$$

Определение. $m \times n$ *матрицей* называется прямоугольная таблица чисел (ЛА.3), где первый индекс означает номер строки, а второй — номер столбца.

Очевидно, верно и обратное: каждой $m \times n$ матрице соответствует линейный оператор, отображающий n-мерное векторное пространство в m-мерное векторное пространство (предполагается, что базисы в обоих пространствах фиксированы).

Матрица $n \times 1$ называется *вектором-столбцом*; матрица $1 \times n$ — *вектором-строкой*; 1×1 матрица называется *скалярной* матрицей. Далее, если не оговорено противное, мы будем везде рассматривать вектор как вектор-столбец.

Квадратная матрица, в которой все элементы, не лежащие на *главной диагонали*, равны 0, называется *диагональной*:

$$\boldsymbol{\Lambda} = (\lambda_i \delta_{ij}) = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & \lambda_n \end{bmatrix}.$$

Диагональная матрица $\boldsymbol{I}_n = (\delta_{ij})$, у которой все диагональные элементы равны 1, называется *единичной* (индекс n здесь обозначает размерность матрицы и может быть опущен).

Нулевой матрицей называется матрица, состоящая из одних нулей.

8. Операции с матрицами

Определение операций с матрицами (сложение, умножение и т. п.) следует из определения операций с линейными операторами.

Определение. Две матрицы $\boldsymbol{A}$ и $\boldsymbol{B}$ *равны*, если совпадают их размерности и равны их соответствующие элементы.

Определение. *Суммой двух матриц* $\boldsymbol{A} = (a_{ij})$ и $\boldsymbol{B} = (b_{ij})$ размерностей $m \times n$ называется матрица $\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{C} = (c_{ij})$ размерности $m \times n$ с элементами $c_{ij} = a_{ij} + b_{ij}$, т. е. при сложении матриц складываются соответствующие элементы.

Определение. *Произведением* $m \times n$ *матрицы* $\boldsymbol{A} = (a_{ij})$ *на число* $\alpha \in R$ называется матрица $\alpha \boldsymbol{A} = \boldsymbol{C} = (c_{ij})$ размерности $m \times n$ с элементами $c_{ij} = \alpha a_{ij}$, т. е. при умножении матрицы на число все элементы матрицы умножаются на это число.

Предложение. Операция сложения матриц удовлетворяет следующим свойствам:

$$\begin{gathered} \boldsymbol{A} + \boldsymbol{B} = \boldsymbol{B} + \boldsymbol{A}, \\ (\boldsymbol{A} + \boldsymbol{B}) + \boldsymbol{C} = \boldsymbol{A} + (\boldsymbol{B} + \boldsymbol{C}), \\ \alpha(\boldsymbol{A} + \boldsymbol{B}) = \alpha \boldsymbol{A} + \alpha \boldsymbol{B}, \\ \boldsymbol{A} + \mathbf{0} = \boldsymbol{A}. \end{gathered} \qquad \text{(ЛА.4)}$$

Определение. *Транспонированной* матрицей называется матрица, у которой строки и столбцы поменялись местами, а именно для $m \times n$ матрицы $\boldsymbol{A} = (a_{ij})$ транспонированной является $n \times m$ матрица $\boldsymbol{A}' = (a'_{ij})$, где $a'_{ij} = a_{ji}$. Например,

$$\begin{bmatrix} 1 & 4 & -3 \\ 2 & 5 & 0 \end{bmatrix}' = \begin{bmatrix} 1 & 2 \\ 4 & 5 \\ -3 & 0 \end{bmatrix}, \quad \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}' = \begin{bmatrix} a_1 & a_2 & \dots & a_n \end{bmatrix}.$$

Предложение. Свойства операции транспонирования матриц:

$$(\boldsymbol{A} + \boldsymbol{B})' = \boldsymbol{A}' + \boldsymbol{B}', \quad (\boldsymbol{A}')' = \boldsymbol{A}.$$

Определение. Пусть мы имеем матрицы $\boldsymbol{A}$ размерности $m \times n$ и $\boldsymbol{B}$ размерности $n \times k$, т. е. число столбцов у матрицы $\boldsymbol{A}$ равно числу строк у матрицы $\boldsymbol{B}$. *Произведением двух матриц* $\boldsymbol{A}$, $\boldsymbol{B}$ называется $m \times k$ матрица $\boldsymbol{C} = \boldsymbol{AB}$, элементы которой определяются следующим образом:

$$c_{ij} = \sum_{s=1}^{n} a_{is} b_{sj}, \qquad i = 1, \dots, m, \quad j = 1, \dots, k.$$

Примеры.

$$\begin{aligned} \boldsymbol{AB} &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} 1 & -1 & 2 \\ 0 & 3 & 4 \end{bmatrix} \\ &= \begin{bmatrix} 1 \cdot 1 + 0 \cdot 0 & 1 \cdot (-1) + 0 \cdot 3 & 1 \cdot 2 + 0 \cdot 4 \\ 0 \cdot 1 + 1 \cdot 0 & 0 \cdot (-1) + 1 \cdot 3 & 0 \cdot 2 + 1 \cdot 4 \\ 2 \cdot 1 + 3 \cdot 0 & 2 \cdot (-1) + 3 \cdot 3 & 2 \cdot 2 + 3 \cdot 4 \end{bmatrix} \\ &= \begin{bmatrix} 1 & -1 & 2 \\ 0 & 3 & 4 \\ 2 & 7 & 16 \end{bmatrix}. \end{aligned}$$

Перемножим эти же матрицы в обратном порядке.

$$\begin{aligned} \boldsymbol{BA} &= \begin{bmatrix} 1 & -1 & 2 \\ 0 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 3 \end{bmatrix} \\ &= \begin{bmatrix} 1 \cdot 1 + (-1) \cdot 0 + 2 \cdot 2 & 1 \cdot 0 + (-1) \cdot 1 + 2 \cdot 3 \\ 0 \cdot 1 + 3 \cdot 0 + 4 \cdot 2 & 0 \cdot 0 + 3 \cdot 1 + 4 \cdot 3 \end{bmatrix} \\ &= \begin{bmatrix} 5 & 5 \\ 8 & 15 \end{bmatrix}. \end{aligned}$$

Этот пример является иллюстрацией того, что операция произведения матриц, вообще говоря, *некоммутативна*: $\boldsymbol{AB} \neq \boldsymbol{BA}$. Более того, $\boldsymbol{AB}$ может быть определено, а $\boldsymbol{BA}$ — не определено вовсе.

Определение. *Скалярным произведением* двух векторов $\boldsymbol{a}$, $\boldsymbol{b}$ размерности n называется число, равное $\boldsymbol{a}'\boldsymbol{b} = \boldsymbol{b}'\boldsymbol{a} = a_1 b_1 + a_2 b_2 + \dots + a_n b_n$.

Пример.

$$\boldsymbol{a}'\boldsymbol{b} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}' \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix} = 1 \cdot 2 + 2 \cdot 0 + 3 \cdot (-1) = -1.$$

Замечание 1. Элемент с координатами i, j в произведении матриц $\boldsymbol{AB}$ равен скалярному произведению i-го вектора-строки матрицы $\boldsymbol{A}$ на j-й вектор-столбец матрицы $\boldsymbol{B}$.

Замечание 2. Важным частным случаем произведения матриц является произведение квадратной $n \times n$ матрицы $\boldsymbol{A}$ на вектор $\boldsymbol{b}$. Например,

$$\begin{aligned} \boldsymbol{Ab} &= \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 2 & 4 & 6 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} 1 \cdot b_1 + 2 \cdot b_2 + 3 \cdot b_3 \\ 1 \cdot b_1 + 0 \cdot b_2 + 1 \cdot b_3 \\ 2 \cdot b_1 + 4 \cdot b_2 + 6 \cdot b_3 \end{bmatrix} \\ &= b_1 \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} + b_2 \begin{bmatrix} 2 \\ 0 \\ 3 \end{bmatrix} + b_3 \begin{bmatrix} 3 \\ 1 \\ 6 \end{bmatrix}. \end{aligned}$$

Как видно из примера, вектор $\boldsymbol{Ab}$ является *линейной комбинацией* столбцов матрицы $\boldsymbol{A}$ с коэффициентами b_i.

Аналогично при умножении матрицы $\boldsymbol{A}$ на вектор-строку (слева) $\boldsymbol{b}'\boldsymbol{A}$ мы получаем вектор-строку, являющийся линейной комбинацией *строк* матрицы $\boldsymbol{A}$ с коэффициентами b_i.

Предложение. Свойства операции умножения матриц:

$\boldsymbol{AI} = \boldsymbol{A}$, $\boldsymbol{IA} = \boldsymbol{A}$ ($\boldsymbol{I}$ — единичная матрица подходящей размерности),

$\boldsymbol{A}(\boldsymbol{B}+\boldsymbol{C}) = \boldsymbol{AB}+\boldsymbol{AC}$,

$(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{C} = \boldsymbol{AC}+\boldsymbol{BC}$,

$\boldsymbol{A}(\boldsymbol{BC}) = (\boldsymbol{AB})\boldsymbol{C}$,

$(\boldsymbol{AB})' = \boldsymbol{B}'\boldsymbol{A}'$, $(\boldsymbol{ABC})' = \boldsymbol{C}'\boldsymbol{B}'\boldsymbol{A}'$,

$\boldsymbol{A0} = \boldsymbol{0}$.

9. Инварианты матриц: след, определитель

В дальнейшем часто используются две числовые функции, определенные только для *квадратных* матриц: *след* матрицы и *определитель* (*детерминант*) матрицы.

Определение. *След* (*trace*) матрицы равен сумме ее диагональных элементов

$$\operatorname{tr}(\boldsymbol{A}) = a_{11} + a_{22} + \cdots + a_{nn} = \sum a_{ii}. \qquad \text{(ЛА.5)}$$

Предложение. Свойства следа матриц:

$$\begin{aligned}
&\operatorname{tr}(\boldsymbol{AB}) = \operatorname{tr}(\boldsymbol{BA}),\\
&\operatorname{tr}(\boldsymbol{I}_n) = n,\\
&\operatorname{tr}(\alpha\boldsymbol{A}) = \alpha\operatorname{tr}(\boldsymbol{A}),\\
&\operatorname{tr}(\boldsymbol{A}') = \operatorname{tr}(\boldsymbol{A}),\\
&\operatorname{tr}(\boldsymbol{A}+\boldsymbol{B}) = \operatorname{tr}(\boldsymbol{A}) + \operatorname{tr}(\boldsymbol{B}).
\end{aligned} \qquad \text{(ЛА.6)}$$

Если $\boldsymbol{a}$ — вектор-столбец, то $\operatorname{tr}(\boldsymbol{aa}') = \operatorname{tr}(\boldsymbol{a}'\boldsymbol{a}) = \boldsymbol{a}'\boldsymbol{a}$ ($\boldsymbol{a}'\boldsymbol{a}$ — скалярный квадрат вектора $\boldsymbol{a}$).

Определение. *Определителем* (*детерминантом*) $\det(\boldsymbol{A}) = |\boldsymbol{A}|$ квадратной $n \times n$ матрицы $\boldsymbol{A}$ называется числовая функция матриц, удовлетворяющая следующим условиям:

1. $n = 1$, $\det(\boldsymbol{A}) = a_{11}$;
2. *«разложение определителя по строке»* при $n > 1$: $|\boldsymbol{A}| = \sum_{j=1}^{n} a_{ij}(-1)^{i+j}|\boldsymbol{A}_{ij}|$, где $\boldsymbol{A}_{ij}$ — $(n-1)\times(n-1)$ матрица, получающаяся из исходной вычеркиванием i-й строки и j-го столбца. Определитель $|\boldsymbol{A}_{ij}|$ называется *минором* порядка $n-1$ матрицы $\boldsymbol{A}$.

Условия 1, 2 дают рекуррентное определение детерминанта матрицы. Для малых размерностей удобно пользоваться формулами:

$$n = 2:\quad \det\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{12}a_{21},$$

$$n = 3:\quad \det\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32}$$
$$- a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32}.$$

Предложение. Свойства определителя матриц:

$\det(\boldsymbol{AB}) = \det(\boldsymbol{A})\det(\boldsymbol{B})$,

$\det(\boldsymbol{\Lambda}) = \det(\lambda_i\delta_{ij}) = \prod_{i=1}^{n}\lambda_i$,

$\det(\boldsymbol{I}_n) = 1$,

$\det(\alpha\boldsymbol{A}) = \alpha^n\det(\boldsymbol{A})$,

$\det(\boldsymbol{A}') = \det(\boldsymbol{A})$,

при перестановке двух строк (столбцов) определитель меняет знак,

определитель равен 0, если в нем есть две одинаковые строки (столбца),

определитель не меняется, если к одной строке (столбцу) матрицы добавить линейную комбинацию других строк (столбцов),

определитель равен 0 тогда и только тогда, когда строки (столбцы) линейно зависимы.

10. Ранг матрицы

Пусть $\boldsymbol{A}$ — $m \times n$ матрица (не обязательно квадратная).

Определение. *Рангом по строкам* матрицы $\boldsymbol{A}$ называется размерность линейного подпространства в R^n, порожденного m векторами-строками матрицы $\boldsymbol{A}$.

Определение. *Рангом по столбцам* матрицы $\boldsymbol{A}$ называется размерность линейного подпространства в R^m, порожденного n векторами-столбцами матрицы $\boldsymbol{A}$.

Определение. *Рангом по минорам* матрицы $\boldsymbol{A}$ называется наибольший порядок ненулевого минора матрицы $\boldsymbol{A}$. (Минор порядка k матрицы — определитель квадратной $k \times k$ матрицы, получающейся из исходной матрицы вычеркиванием некоторого количества строк и столбцов.)

Предложение. Все три приведенных выше определения дают одно и то же число, называемое *рангом матрицы*: $\operatorname{rank}(\boldsymbol{A})$.

Предложение. Свойства ранга матрицы:

$\operatorname{rank}(\boldsymbol{A}) \leqslant \min(m, n)$,

$\operatorname{rank}(\boldsymbol{AB}) \leqslant \min(\operatorname{rank}(\boldsymbol{A}), \operatorname{rank}(\boldsymbol{B}))$,

если $\boldsymbol{B}$ — $n \times n$ квадратная матрица ранга n, то $\operatorname{rank}(\boldsymbol{AB}) = \operatorname{rank}(\boldsymbol{A})$,

если $\boldsymbol{B}$ — $m \times m$ квадратная матрица ранга m, то $\operatorname{rank}(\boldsymbol{BA}) = \operatorname{rank}(\boldsymbol{A})$,

$\operatorname{rank}(\boldsymbol{A}) = \operatorname{rank}(\boldsymbol{AA}') = \operatorname{rank}(\boldsymbol{A}'\boldsymbol{A})$, причем $\boldsymbol{AA}'$ — $m \times m$ матрица, а $\boldsymbol{A}'\boldsymbol{A}$ — $n \times n$ матрица.

Замечание. Пусть $\mathcal{A}$ — линейный оператор $\mathcal{A}$: $R^n \to R^m$, соответствующий $m \times n$ матрице $\boldsymbol{A}$. Образом $\operatorname{Im}(\mathcal{A})$ оператора $\mathcal{A}$ называется множество всех векторов из R^m, которые являются образами векторов из R^n при отображении $\mathcal{A}$. Тогда размерность образа оператора равна рангу матрицы: $\dim(\operatorname{Im}(\mathcal{A})) = \operatorname{rank}(\boldsymbol{A})$.

11. Обратная матрица

Пусть $\boldsymbol{A}$ — квадратная $n \times n$ матрица.

Определение. Матрица $\boldsymbol{A}$ называется *невырожденной*, если она имеет максимальный возможный ранг: $\operatorname{rank}(\boldsymbol{A}) = n$.

Определение. Матрицей, *обратной* к матрице $\boldsymbol{A}$, называется матрица, обозначаемая $\boldsymbol{A}^{-1}$, такая, что $\boldsymbol{AA}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}$.

Предложение. Для всякой *невырожденной* квадратной $n \times n$ матрицы $\boldsymbol{A}$ существует (единственная) обратная матрица A^{-1}.

Предложение. Обозначим через a^{ij} элементы обратной матрицы A^{-1}. Тогда

$$a^{ij} = \frac{(-1)^{i+j}|\boldsymbol{M}_{ji}|}{|\boldsymbol{A}|},$$

где $\boldsymbol{M}_{ij}$ — матрица, получающаяся из $\boldsymbol{A}$ вычеркиванием i-й строки и j-го столбца.

Примеры.

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix},$$

$$\begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_n \end{bmatrix}^{-1} = \begin{bmatrix} \lambda_1^{-1} & 0 & \dots & 0 \\ 0 & \lambda_2^{-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_n^{-1} \end{bmatrix}.$$

Предложение. Свойства обратной матрицы:

$|\boldsymbol{A}^{-1}| = |\boldsymbol{A}|^{-1}$,

$(\boldsymbol{A}^{-1})^{-1} = \boldsymbol{A}$,

$(\boldsymbol{A}^{-1})' = (\boldsymbol{A}')^{-1}$,

если существуют $\boldsymbol{A}^{-1}$ и $\boldsymbol{B}^{-1}$, то $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$.

(Отметим, что в последней формуле все матрицы *квадратные* и невырожденные.)

12. Системы линейных уравнений

Систему n линейных уравнений с n неизвестными

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1, \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n = b_n, \end{cases} \qquad (\text{ЛА}.7)$$

удобно записать в матричном виде:

$$\boldsymbol{Ax} = \boldsymbol{b},$$

где $\boldsymbol{A} = (a_{ij})$ — квадратная $n \times n$ матрица, $\boldsymbol{x} = (x_1, \dots, x_n)'$ и $\boldsymbol{b} = (b_1, \dots, b_n)'$ — векторы-столбцы.

Предложение. Если матрица $\boldsymbol{A}$ невырожденная, то система (ЛА.7) имеет единственное решение: $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$.

Предложение. Однородная система $\boldsymbol{Ax} = \boldsymbol{0}$ имеет ненулевое решение тогда и только тогда, когда матрица $\boldsymbol{A}$ вырожденная: $|\boldsymbol{A}| = 0$.

13. Собственные числа и векторы

Пусть $\mathcal{A}$ — линейный оператор $\mathcal{A}$: $R^n \to R^n$.

Определение. Вектор $\boldsymbol{a}$, не равный $\boldsymbol{0}$, называется *собственным вектором* (характеристическим вектором) линейного оператора $\mathcal{A}$, а λ — *собственным числом* (собственным корнем, характеристическим числом), соответствующим собственному вектору $\boldsymbol{a}$, если выполняется равенство

$$\mathcal{A}\boldsymbol{a} = \lambda\boldsymbol{a}. \qquad (\text{ЛА}.8)$$

Собственный вектор определен с точностью до коэффициента пропорциональности.

Перепишем равенство (ЛА.8) в виде:

$$\mathcal{A}\boldsymbol{a} - \lambda\boldsymbol{a} = \boldsymbol{0}, \quad \text{или} \quad (\mathcal{A} - \lambda\mathcal{I})\boldsymbol{a} = \boldsymbol{0}.$$

Выберем базис в R^n и перейдем от операторов к матрицам. Однородная система уравнений имеет нетривиальное решение, если определитель системы равен 0: $|\boldsymbol{A} - \lambda\boldsymbol{I}| = 0$.

Заметим, что определитель $|\boldsymbol{A} - \lambda\boldsymbol{I}|$ является многочленом степени n от λ.

Определение. Уравнение $|\boldsymbol{A} - \lambda\boldsymbol{I}| = 0$ называется *характеристическим уравнением* матрицы. Корнями этого многочлена являются *характеристические числа* матрицы (или соответствующего оператора — ниже будет показано, что при другом выборе базиса характеристический многочлен тот же).

Пример.

$$\boldsymbol{A} = \begin{bmatrix} 1 & 4 \\ 1 & 1 \end{bmatrix},$$

$$\det(\boldsymbol{A} - \lambda\boldsymbol{I}) = \begin{vmatrix} 1-\lambda & 4 \\ 1 & 1-\lambda \end{vmatrix} = (1-\lambda)^2 - 4 = \lambda^2 - 2\lambda - 3 = 0,$$

$$\lambda_1 = -1, \quad \lambda_2 = 3.$$

Предложение. Пусть есть два базиса в R^n: $\{\boldsymbol{e}_i\}$ и $\{\boldsymbol{l}_j\}$. Обозначим через $\boldsymbol{C} = \{c_{ij}\}$ матрицу перехода от базиса $\{\boldsymbol{l}_j\}$ к базису $\{\boldsymbol{e}_i\}$: $\boldsymbol{l}_j = \sum_s c_{sj}\boldsymbol{e}_s$. Пусть линейному оператору $\mathcal{A}$ соответствуют матрицы $\boldsymbol{A}$ и $\boldsymbol{B}$ в базисах $\{\boldsymbol{e}_i\}$ и $\{\boldsymbol{l}_j\}$ соответственно. Тогда

$$\boldsymbol{B} = \boldsymbol{C}^{-1}\boldsymbol{AC}. \qquad (\text{ЛА}.9)$$

Определение. Матрицы $\boldsymbol{A}$ и $\boldsymbol{B}$, для которых существует матрица $\boldsymbol{C}$, такая что выполняется (ЛА.9), называются *подобными*.

Вычислим характеристический многочлен оператора $\mathcal{A}$ в базисе $\{\boldsymbol{l}_j\}$.

$$\begin{aligned} \det(\boldsymbol{B} - \lambda\boldsymbol{I}) &= \det(\boldsymbol{C}^{-1}\boldsymbol{AC} - \lambda\boldsymbol{I}) = \det(\boldsymbol{C}^{-1}\boldsymbol{AC} - \lambda\boldsymbol{C}^{-1}\boldsymbol{IC}) \\ &= \det(\boldsymbol{C}^{-1}(\boldsymbol{A} - \lambda\boldsymbol{I})\boldsymbol{C}) = \det(\boldsymbol{C}^{-1})\det(\boldsymbol{A} - \lambda\boldsymbol{I})\det(\boldsymbol{C}) \\ &= \det(\boldsymbol{A} - \lambda\boldsymbol{I})\det(\boldsymbol{C})\det(\boldsymbol{C})^{-1} = \det(\boldsymbol{A} - \lambda\boldsymbol{I}). \end{aligned}$$

Таким образом, характеристический многочлен зависит только от линейного оператора и не зависит от выбора базиса. Сформулируем этот результат иначе: характеристические многочлены подобных матриц совпадают.

Характеристический многочлен — многочлен с вещественными коэффициентами — имеет не более n вещественных корней (часть корней могут быть комплексными).

Предложение. Разным собственным числам соответствуют *линейно независимые* собственные векторы.

Предложение. Пусть характеристический многочлен матрицы $\boldsymbol{A}$ имеет n *различных вещественных* корней. Тогда матрица $\boldsymbol{A}$ может быть представлена в виде

$$\boldsymbol{A} = \boldsymbol{C}^{-1}\boldsymbol{\Lambda}\boldsymbol{C}, \qquad \text{(ЛА.10)}$$

где матрица $\boldsymbol{\Lambda}$ — *диагональная*, а матрица $\boldsymbol{C}$ — невырожденная.

В самом деле, так как все n собственных чисел разные, то им соответствуют n линейно независимых собственных векторов, образующих базис. Теперь воспользуемся (ЛА.9), где $\boldsymbol{C}$ — матрица перехода к базису, состоящему из собственных векторов, в котором матрица оператора, очевидно, имеет диагональный вид.

14. Симметричные матрицы

Определение. Матрица $\boldsymbol{A}$ называется *симметричной*, если $\boldsymbol{A}' = \boldsymbol{A}$.

Пример.

$$\begin{bmatrix} 1 & -2 & 4 \\ -2 & 2 & 5 \\ 4 & 5 & 3 \end{bmatrix}.$$

Предложение. Для любой матрицы $\boldsymbol{A}$ матрица $\boldsymbol{A}'\boldsymbol{A}$ — симметричная.

В самом деле, $(\boldsymbol{A}'\boldsymbol{A})' = \boldsymbol{A}'(\boldsymbol{A}')' = \boldsymbol{A}'\boldsymbol{A}$.

Предложение. Симметричная $n \times n$ матрица $\boldsymbol{A}$ имеет n собственных чисел (некоторые из них могут совпадать), которым соответствуют n собственных векторов $\boldsymbol{c}_1, \ldots, \boldsymbol{c}_n$, которые могут быть выбраны *попарно ортогональными*. (Собственные векторы, соответствующие *разным* собственным значениям симметричной матрицы, всегда ортогональны.)

Более того, поскольку собственный вектор определяется с точностью до коэффициента пропорциональности, то можно нормировать собственные векторы $\{\boldsymbol{c}_i\}$ так, что они будут *ортонормированной системой*, т. е. попарно ортогональны и единичной длины:

$$\boldsymbol{A}\boldsymbol{c}_i = \lambda_i \boldsymbol{c}_i, \quad \boldsymbol{c}_i'\boldsymbol{c}_j = 0, \quad \boldsymbol{c}_i'\boldsymbol{c}_i = 1, \qquad i \neq j.$$

Тогда (ЛА.10) матрица $\boldsymbol{A}$ приводится к диагональному виду при помощи матрицы $\boldsymbol{O}$, столбцы которой являются векторами $\boldsymbol{c}_i$.

$$\boldsymbol{\Lambda} = \boldsymbol{O}^{-1}\boldsymbol{A}\boldsymbol{O}, \qquad \text{(ЛА.11)}$$

где на диагонали матрицы $\boldsymbol{\Lambda}$ стоят собственные числа матрицы $\boldsymbol{A}$.

Определение. Матрица, столбцы которой составляют ортонормированную систему векторов, называется *ортогональной*.

Предложение. Ортогональная матрица удовлетворяет соотношению:

$$\boldsymbol{O}'\boldsymbol{O} = \boldsymbol{I}.$$

Предложение. Если $\boldsymbol{O}$ — ортогональная матрица, то $\boldsymbol{O}' = \boldsymbol{O}^{-1}$.

Предложение. Ортогональная матрица имеет определитель, равный $+1$ или -1.

Предложение. В ортогональной матрице строки также образуют ортонормированную систему векторов.

Предложение. Симметричная матрица $\boldsymbol{A}$ может быть приведена к диагональному виду при помощи ортогонального преобразования $\boldsymbol{O}$

$$\boldsymbol{O}'\boldsymbol{A}\boldsymbol{O} = \boldsymbol{\Lambda}. \qquad \text{(ЛА.12)}$$

Предложение. Соотношение (ЛА.12) можно записать в виде разложения симметричной матрицы $\boldsymbol{A}$ на ортогональную и диагональную:

$$A = O\Lambda O' = \sum_{s=1}^{n} \lambda_s c_s c_s'. \qquad (ЛА.13)$$

В разложении (ЛА.13) диагональная матрица Λ состоит из собственных значений матрицы A.

15. Положительно определенные матрицы

Определение. Симметричная $n \times n$ матрица A *называется положительно определенной*, если для каждого ненулевого вектора x выполняется неравенство

$$x'Ax > 0. \qquad (ЛА.14)$$

Определение. Симметричная $n \times n$ матрица A *называется неотрицательно определенной*, если для каждого вектора x выполняется неравенство

$$x'Ax \geqslant 0. \qquad (ЛА.15)$$

Пример. Матрица $A'A$ неотрицательно определена для любой матрицы A. В самом деле, для любого вектора x

$$x'(A'A)x = (Ax)'(Ax) = y'y \geqslant 0.$$

Здесь вектор $y = Ax$, а скалярный квадрат любого вектора, конечно, неотрицателен.

Для симметричных матриц можно ввести отношение порядка.

Определение. Будем говорить, что $A \geqslant 0$, если A — неотрицательно определена, и $A > 0$, если A положительно определена. Будем говорить, что $A \geqslant B$ ($A > B$), если матрица $A - B$ неотрицательно определена (положительно определена).

Предложение. Если $A \geqslant B$, то $a_{ii} \geqslant b_{ii}$ для всех i.

(Для доказательства достаточно рассмотреть $x_i'(A-B)x_i$ при $x_i = (\delta_{i1}, \delta_{i2}, \ldots, \delta_{in})'$.)

Предложение. Если $A > B$ и $C \geqslant 0$, то $A + C > B$.

Предложение. Если $A > B$ и обе матрицы обратимы, то $B^{-1} > A^{-1}$.

Предложение. У положительно определенной (неотрицательно определенной) матрицы A все собственные числа положительны (неотрицательны).

В самом деле, пусть x — собственный вектор, соответствующий собственному числу λ, т. е. $Ax = \lambda x$. Так как матрица положительно определена, то $x'Ax > 0$. Но $x'Ax = x'\lambda x = \lambda x'x > 0$, следовательно, $\lambda > 0$ ($x'x > 0$, как скалярный квадрат ненулевого вектора).

Для положительно определенных матриц можно определить дробные степени и другие функции от матриц следующим образом. Представим положительно определенную симметричную матрицу A в виде разложения на ортогональную и диагональную (ЛА.13):

$$A = O\Lambda O'.$$

Диагональные элементы Λ являются собственными числами матрицы A, следовательно, неотрицательны (см. выше). Тогда можно определить

$$A^{1/2} = O\Lambda^{1/2}O',$$
$$(A^{1/2})^2 = O\Lambda^{1/2}O'O\Lambda^{1/2}O' = O\Lambda^{1/2}\Lambda^{1/2}O' = O\Lambda O' = A,$$

и аналогично для любой другой дробной степени (в том числе отрицательной).

Здесь, конечно,

$$\Lambda^{1/2} = \begin{bmatrix} \lambda_1^{1/2} & 0 & \ldots & 0 \\ 0 & \lambda_2^{1/2} & \ddots & \vdots \\ \ldots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \lambda_n^{1/2} \end{bmatrix}.$$

16. Идемпотентные матрицы

Определение. Матрица M называется *идемпотентной*, если она совпадает со своим квадратом: $M = M^2$. Мы далее будем считать матрицу M также и *симметричной*, так как именно такие матрицы встречаются в эконометрике. Однако многие приведенные ниже результаты верны и без предположения симметричности матрицы M. Часто требование симметричности включают в определение идемпотентной матрицы.

Предложение. Собственные числа идемпотентной матрицы могут принимать значения только 0 или 1.

В самом деле, если x — собственный вектор идемпотентной матрицы M, а λ — соответствующее собственное значение, то $\lambda x = Mx = M^2x = M\lambda x = \lambda Mx = \lambda^2 x$, или $(\lambda - \lambda^2)x = 0$, откуда $\lambda(1-\lambda) = 0$.

Предложение. Ранг идемпотентной матрицы равен ее следу.

Пусть M — идемпотентная (симметричная) матрица. В силу (ЛА.13) ее можно представить в виде $M = O\Lambda O'$, где на диагонали Λ стоят нули и единицы (собственные числа матрицы M). Из свойств ранга матрицы (см. ЛА, п. 10) следует, что $\operatorname{rank}(M) = \operatorname{rank}(\Lambda)$, т. к. ортогональная матрица O невырождена ($\det(O) = \pm 1$). Ранг матрицы Λ равен, очевидно, числу ненулевых элементов на диагонали, т. е. числу собственных значений матрицы M, равных 1. След матрицы M равен $\operatorname{tr}(M) = \operatorname{tr}(O\Lambda O') = \operatorname{tr}(O'O\Lambda) = \operatorname{tr}(\Lambda)$, также равен числу собственных значений матрицы M, равных 1, что и требовалось показать.

Пример. Обозначим через $\imath$ $n \times 1$ — вектор-столбец, состоящий из одних единиц. Рассмотрим матрицу $M = I - \frac{1}{n}\imath\imath'$. Проверим, что она идемпотентная.

$$M^2 = \left(I - \frac{1}{n}\imath\imath'\right)\left(I - \frac{1}{n}\imath\imath'\right) = I - \frac{2}{n}\imath\imath' + \frac{1}{n^2}\imath\imath'\imath\imath' = I - \frac{1}{n}\imath\imath' = M.$$

Эта матрица обладает следующим свойством (вычисление отклонений от среднего значения):

$$Mx = x - \overline{x}\imath = \begin{bmatrix} x_1 - \overline{x} \\ \dots \\ x_n - \overline{x} \end{bmatrix}, \quad \text{где} \quad \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

С геометрической точки зрения идемпотентная матрица соответствует оператору *проектирования* на векторное подпространство. Так, например, матрица $M = I - \frac{1}{n}\imath\imath'$ является проектором на подпространство, ортогональное вектору $\imath = (1, \dots, 1)'$.

17. Блочные матрицы

Часто, в соответствии со смыслом задачи, удобно разбить матрицу на подматрицы (блоки). Например, $m \times n$ матрицу A можно разбить на блоки:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \tag{ЛА.16}$$

где A_{11} — $m_1 \times n_1$ матрица, A_{12} — $m_1 \times n_2$ матрица, A_{21} — $m_2 \times n_1$ матрица, A_{22} — $m_2 \times n_2$ матрица, $m = m_1 + m_2$, $n = n_1 + n_2$.

Две матрицы, разбитые на блоки одинаковым образом (т. е. размерности матриц A_{ij} и B_{ij} совпадают), можно складывать

$$A + B = \begin{bmatrix} A_{11} + B_{11} & A_{12} + B_{12} \\ A_{21} + B_{21} & A_{22} + B_{22} \end{bmatrix}.$$

Можно также умножать матрицы, разбитые на блоки подходящим способом, т. е. так, чтобы все операции в приведенной ниже формуле были корректны (для этого необходимо, чтобы количество столбцов в матрице A_{ij} и строк в матрице B_{jk} совпадали для всех i, j, k)

$$\begin{aligned} AB &= \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \\ &= \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix}. \end{aligned}$$

Определитель блочной матрицы

Пусть $n \times n$ матрица $\boldsymbol{A}$ разбита на блоки (ЛА.16), такие что $\boldsymbol{A}_{11}$ и $\boldsymbol{A}_{22}$ являются квадратными матрицами. Тогда верна следующая формула для определителя матрицы $\boldsymbol{A}$:

$$\begin{vmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{vmatrix} = |\boldsymbol{A}_{11}||\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}| = |\boldsymbol{A}_{22}||\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21}|.$$

В частном случае, когда $\boldsymbol{A}_{12}$ и $\boldsymbol{A}_{21}$ — нулевые матрицы (такая матрица $\boldsymbol{A}$ называется *блочно-диагональной*)

$$\begin{vmatrix} \boldsymbol{A}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_{22} \end{vmatrix} = |\boldsymbol{A}_{11}||\boldsymbol{A}_{22}|.$$

Матрица, обратная к блочной матрице

Для блочно-диагональной матрицы $\boldsymbol{A}$ обратная матрица равна

$$\begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{A}_{11}^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A}_{22}^{-1} \end{bmatrix}.$$

Для блочной матрицы общего вида можно получить (см. упражнение ЛА.9)

$$\begin{bmatrix} \boldsymbol{A}_{11} & \boldsymbol{A}_{12} \\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \boldsymbol{A}^{11} & -\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12}\boldsymbol{A}^{22} \\ -\boldsymbol{A}^{22}\boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1} & \boldsymbol{A}^{22} \end{bmatrix}, \qquad \text{(ЛА.17)}$$

где

$$\boldsymbol{A}^{22} = (\boldsymbol{A}_{22} - \boldsymbol{A}_{21}\boldsymbol{A}_{11}^{-1}\boldsymbol{A}_{12})^{-1}, \quad \boldsymbol{A}^{11} = (\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})^{-1}. \quad \text{(ЛА.18)}$$

18. Произведение Кронекера

Некоторые вычисления, связанные с системами регрессионных уравнений, значительно сокращаются, если использовать понятие произведения Кронекера.

Определение. *Произведением Кронекера* двух матриц $\boldsymbol{A}$ (размерности $m \times n$) и $\boldsymbol{B}$ (размерности $k \times l$) называется блочная матрица

$\boldsymbol{A} \otimes \boldsymbol{B}$ размерности $km \times ln$:

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{11}\boldsymbol{B} & a_{12}\boldsymbol{B} & \dots & a_{1n}\boldsymbol{B} \\ a_{21}\boldsymbol{B} & a_{22}\boldsymbol{B} & \dots & a_{2n}\boldsymbol{B} \\ \dots & \dots & \dots & \dots \\ a_{m1}\boldsymbol{B} & a_{m2}\boldsymbol{B} & \dots & a_{mn}\boldsymbol{B} \end{bmatrix}. \qquad \text{(ЛА.19)}$$

Пример.

$$\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix} \otimes \begin{bmatrix} 2 & 0 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 1\begin{bmatrix} 2 & 0 \\ 1 & 4 \end{bmatrix} & 2\begin{bmatrix} 2 & 0 \\ 1 & 4 \end{bmatrix} \\ 0\begin{bmatrix} 2 & 0 \\ 1 & 4 \end{bmatrix} & 3\begin{bmatrix} 2 & 0 \\ 1 & 4 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} 2 & 0 & 4 & 0 \\ 1 & 4 & 2 & 8 \\ 0 & 0 & 6 & 0 \\ 0 & 0 & 3 & 12 \end{bmatrix}.$$

Произведение Кронекера обладает следующими свойствами.

1. Если матрицы $\boldsymbol{A}$, $\boldsymbol{B}$ обратимы, то $[\boldsymbol{A} \otimes \boldsymbol{B}]^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}$.
2. Если $\boldsymbol{A}$ — $m \times m$ матрица и $\boldsymbol{B}$ — $n \times n$ матрица, то $\det(\boldsymbol{A} \otimes \boldsymbol{B}) = (\det(\boldsymbol{A}))^n(\det(\boldsymbol{B}))^m$.
3. $(\boldsymbol{A} \otimes \boldsymbol{B})' = (\boldsymbol{A}' \otimes \boldsymbol{B}')$.
4. $\operatorname{tr}(\boldsymbol{A} \otimes \boldsymbol{B}) = \operatorname{tr}(\boldsymbol{A})\operatorname{tr}(\boldsymbol{B})$.

19. Дифференцирование по векторному аргументу

Мы следуем подходу, изложенному в книге (Magnus, Neudecker, 1988).

Определение. Производной скалярной функции $\varphi(\boldsymbol{x})$ от векторного $n \times 1$ аргумента $\boldsymbol{x} = (x_1, \dots, x_n)'$ называется $1 \times n$ вектор (вектор-строка)

$$\frac{\partial \varphi(\boldsymbol{x})}{\partial \boldsymbol{x}'} = \left(\frac{\partial \varphi(\boldsymbol{x})}{\partial x_1}, \dots, \frac{\partial \varphi(\boldsymbol{x})}{\partial x_n} \right). \qquad \text{(ЛА.20)}$$

Определение. Производной векторной $m \times 1$ функции $\boldsymbol{f}(\boldsymbol{x})$ от векторного $n \times 1$ аргумента $\boldsymbol{x} = (x_1, \dots, x_n)'$ называется $m \times n$

матрица, составленная из производных

$$\frac{\partial f(x)}{\partial x'} = \begin{bmatrix} \frac{\partial f_1(x)}{\partial x_1} & \dots & \frac{\partial f_1(x)}{\partial x_n} \\ \dots & \dots & \dots \\ \frac{\partial f_m(x)}{\partial x_1} & \dots & \frac{\partial f_m(x)}{\partial x_n} \end{bmatrix} \quad \text{(матрица Якоби).} \qquad \text{(ЛА.21)}$$

Заметим, что при $m = 1$ второе определение совпадает с первым.

Важные примеры

1. $\varphi(x) = a'x$, $a = (a_1, \dots, a_n)'$, $x = (x_1, \dots, x_n)'$ — $n \times 1$ векторы.

$$\frac{\partial \varphi(x)}{\partial x'} = \frac{\partial (a'x)}{\partial x'} = a'. \qquad \text{(ЛА.22)}$$

2. $\varphi(x) = x'Ax$, A — $n \times n$ матрица.

$$\frac{\partial \varphi(x)}{\partial x'} = \frac{\partial (x'Ax)}{\partial x'} = x'(A + A'). \qquad \text{(ЛА.23)}$$

Докажем последнее равенство.

$$\begin{aligned} \frac{\partial \varphi(x)}{\partial x_i} &= \frac{\partial (\sum_s \sum_t x_s A_{st} x_t)}{\partial x_i} = \sum_s \sum_t \frac{\partial (x_s A_{st} x_t)}{\partial x_i} \\ &= \sum_s \sum_t (\delta_{is} A_{st} x_t + x_s A_{st} \delta_{ti}) = \sum_t A_{it} x_t + \sum_s x_s A_{si} \\ &= \big(x'(A' + A)\big)_i \end{aligned}$$

(здесь δ_{ij} — символ Кронекера (ЛА.2)).

В том случае, когда матрица A симметричная, (ЛА.23) принимает вид

$$\frac{\partial \varphi(x)}{\partial x'} = \frac{\partial (x'Ax)}{\partial x'} = 2x'A. \qquad \text{(ЛА.24)}$$

3. $f(x) = Ax$, A — $m \times n$ матрица.

$$\frac{\partial f(x)}{\partial x'} = \frac{\partial Ax}{\partial x'} = A. \qquad \text{(ЛА.25)}$$

Подробное и последовательное изложение аппарата векторного и матричного дифференциального исчисления можно найти в книге (Magnus, Neudecker, 1988).

Упражнения

ЛА.1. Докажите, что $\operatorname{tr}(AB) = \operatorname{tr}(BA)$. (*Указание.* Напишите это условие в координатах.)

ЛА.2. Докажите, что $\operatorname{tr}(A \otimes B) = \operatorname{tr}(A)\operatorname{tr}(B)$.

ЛА.3. Покажите, что матрица $X(X'X)^{-1}X'$ — идемпотентная.

ЛА.4. Матрица A — идемпотентная. Докажите, что матрица $B = I - A$ также идемпотентная и $BA = 0$.

ЛА.5. Пусть A — $n \times n$ матрица $A = (1-\alpha)I + \alpha\iota\iota'$, где $\iota = [1 \dots 1]'$ — $n \times 1$ вектор. Найдите собственные числа и собственные векторы матрицы A.

ЛА.6. Пусть $n \times n$ матрица A зависит от скалярной переменной t, $A = A(t)$, x — вектор $n \times 1$.

а) Найдите $\partial(x'Ax)/\partial t$.

б) Найдите $\partial(x'Ax)/\partial t$, если x зависит от t: $x = x(t)$.

ЛА.7. Пусть $\mathcal{A}$: $L \to M$ линейный оператор, $\dim(L) = n$, $\dim(M) = m$. Докажите, что

а) $\operatorname{Im}(\mathcal{A})$, $\operatorname{Ker}(\mathcal{A})$ — векторные пространства;

б) $\dim(\operatorname{Im}(\mathcal{A})) \leqslant \min(n, m)$;

в) $\dim(\operatorname{Ker}(\mathcal{A})) + \dim(\operatorname{Im}(\mathcal{A})) = n$.

ЛА.8. Покажите, что многочлены $p(x)$ с вещественными коэффициентами и естественными операциями сложения и умножения на число образуют векторное пространство L.

а) Найдите размерность L.

б) Выберите базис в L. Какие матрицы соответствуют в выбранном вами базисе операторам $\mathcal{A}$: $p(x) \to dp(x)/dx$; $\mathcal{B}$: $p(x) \to p(x-1)$. Найдите собственные числа и собственные векторы этих операторов.

ЛА.9. Докажите формулы (ЛА.17), (ЛА.18) для обращения блочной матрицы. (*Указание.* Используйте формулу для произведения блочных матриц из п. 17 и определение обратной матрицы).

ЛА.10. Покажите, что:

а) $\dfrac{\partial a' f(x)}{\partial x'} = a' \dfrac{\partial f(x)}{\partial x'}$;

б) $\dfrac{\partial (f(x))' g(x)}{\partial x'} = (g(x))' \dfrac{\partial f(x)}{\partial x'} + (f(x))' \dfrac{\partial g(x)}{\partial x'}$;

в) $\dfrac{\partial (f(x))' A g(x)}{\partial x'} = (g(x))' A \dfrac{\partial f(x)}{\partial x'} + (f(x))' A \dfrac{\partial g(x)}{\partial x'}$.

Приложение МС

Теория вероятностей и математическая статистика

В этом приложении приводится краткий обзор основных понятий и результатов теории вероятностей и математической статистики, которые используются в основном тексте книги. Приложение носит справочный характер и не может служить заменой стандартному курсу теории вероятностей и математической статистики.

1. Случайные величины, случайные векторы

На содержательном уровне случайной величиной мы называем какую-либо числовую характеристику, связанную с изучаемым объектом, значение которой принципиально не может быть предсказано точно и зависит от случая.

Формально *случайная величина* X — это числовая функция, заданная на некотором вероятностном пространстве (Ω, P): $X = X(\omega)$, $\omega \in \Omega$. *Функцией распределения* случайной величины X называется числовая функция числового аргумента, определяемая

равенством

$$F(x) = \mathrm{P}(X \leqslant x), \qquad x \in R \tag{МС.1}$$

(R — множество действительных чисел). Часто, чтобы подчеркнуть связь функции распределения со случайной величиной, используют обозначение $F(x) = F_X(x)$. Каждая функция распределения обладает следующими свойствами:

F1) $0 \leqslant F(X) \leqslant 1$ при любом $x \in R$;

F2) $F(x)$ является неубывающей, непрерывной справа функцией;

F3) $\lim_{x\to-\infty} F(x) = 0$, $\lim_{x\to\infty} F(x) = 1$.

Верно и обратное, т. е. любая функция, удовлетворяющая условиям F1)–F3), является функцией распределения некоторой случайной величины. Нестрого говоря, функция распределения содержит всю вероятностную информацию о случайной величине X. В частности, $\mathrm{P}(X \in (a, b]) = F(b) - F(a)$ для любых чисел $a \leqslant b$.

Можно выделить два класса случайных величин — дискретные и непрерывные. Случайная величина X называется *дискретной*, если множество ее значений конечно или счетно. Несмотря на то что функция распределения универсальным образом описывает вероятностный характер случайной величины, дискретную случайную величину удобно представлять в виде таблицы

$$X = \begin{pmatrix} x_1 & x_2 & \dots & x_k & \dots \\ p_1 & p_2 & \dots & p_k & \dots \end{pmatrix}, \qquad p_k = \mathrm{P}(X = x_k). \tag{МС.2}$$

Случайная величина X называется *непрерывной*, если ее функция распределения дифференцируема, т. е. существует производная $p(x) = F'(x)$, называемая *плотностью распределения* случайной величины X. В этом случае для любого (измеримого) множества $A \subseteq R$ справедливо равенство

$$\mathrm{P}(X \in A) = \int_A p(x)\,dx. \tag{МС.3}$$

В частности, $F(x) = \int_{-\infty}^{x} p(y)\,dy$. Плотность распределения обладает следующими свойствами:

p1) $p(x) \geqslant 0$ при любом $x \in R$;

p2) $\int_{-\infty}^{+\infty} p(y)\,dy = 1$.

Обратно, любая функция, удовлетворяющая условиям p1), p2), является плотностью распределения некоторой случайной величины.

Непрерывная случайная величина с плотностью $p(x)$ имеет *симметричное распределение*, если $p(x) = p(-x)$ для всех x.

Если случайная величина не принадлежит ни одному из этих двух классов, то говорят, что она имеет *смешанный тип*.

Математическое ожидание (среднее значение) дискретной случайной величины X, имеющей распределение (МС.2), есть по определению ряд

$$\mathrm{E}(X) = \sum_k x_k p_k \tag{МС.4}$$

при условии его абсолютной сходимости. Для непрерывной случайной величины X с плотностью распределения $p(x)$ математическое ожидание — это интеграл

$$\mathrm{E}(X) = \int_{-\infty}^{+\infty} x p(x)\,dx \tag{МС.5}$$

также при условии, что он абсолютно сходится. В общем случае математическое ожидание определяется как интеграл Лебега–Стилтьеса $\int_{-\infty}^{+\infty} x\,dF(x)$, но для наших целей достаточно определений (МС.4) и (МС.5). Математическое ожидание имеет следующие свойства (X, Y — произвольные случайные величины, a, b — константы):

E1) $\mathrm{E}(aX + bY) = a\mathrm{E}(X) + b\mathrm{E}(Y)$;

E2) если $X \geqslant Y$ при всех реализациях, то $\mathrm{E}(X) \geqslant \mathrm{E}(Y)$;

E3) если X — непрерывная случайная величина с плотностью распределения $p(x)$, а $g(x)$, $x \in R$ — числовая функция, то для случайной величины $Y = g(X)$ справедливо равенство

$$\mathrm{E}(Y) = \int_{-\infty}^{+\infty} g(x)p(x)\,dx. \qquad \text{(МС.6)}$$

Другой важнейшей числовой характеристикой случайной величины X является *дисперсия*, отражающая степень «разброса» случайной величины относительно среднего значения. Она определяется равенством

$$\mathrm{V}(X) = \mathrm{E}(X - \mathrm{E}X)^2.$$

Величину $\sigma = \sqrt{\mathrm{V}(X)}$ называют *стандартным отклонением* случайной величины X.

Из свойства E1) непосредственно вытекает, что

$$\mathrm{V}(X) = \mathrm{E}(X^2) - (\mathrm{E}X)^2.$$

Рассмотрим задачу минимизации по c среднеквадратичного отклонения $f(c) = \mathrm{E}(X - c)^2$. С помощью элементарных вычислений получаем следующий результат.

E4) Функция $f(c)$ достигает минимума при $c = \mathrm{E}(X)$, причем $\min f(c) = \mathrm{V}(X)$.

В математической статистике широко используются квантили и процентные точки распределения случайной величины. Пусть случайная величина X имеет непрерывную функцию распределения $F(x)$ и пусть задано число $0 < q < 1$. *Квантилью уровня* q (или *q-квантилью*) распределения $F(x)$ называется такое число u_q, что $F(u_q) = \mathrm{P}(X \leqslant u_q) = q$. Если величина X дискретна, то может случиться, что такого числа u_q либо не существует, либо их бесконечно много. Но всегда можно найти два числа $x'_q < x''_q$ такие, что $F(x'_q) \leqslant q$, $F(x''_q) > q$. Тогда q-квантиль определяется как любое число, лежащее между x'_q и x''_q.

Для симметричных распределений $F(x)$ вводится также понятие *двусторонней q-квантили*, а именно, число t_q $(0 < q < 1)$ называется двусторонней q-квантилью распределения $F(x)$, если $\mathrm{P}(|X| \leqslant t_q) = q$.

Пусть теперь задано число $0 < \alpha < 1$. *100α-процентной точкой* непрерывного распределения $F(x)$ называется такое число w_α, что $1 - F(w_\alpha) = \mathrm{P}(X > w_\alpha) = \alpha$. Для дискретной величины

это определение модифицируется аналогично тому, как это сделано для квантили. Нетрудно проверить, что $u_q = w_{1-q}$.

Упорядоченный набор $\boldsymbol{x} = (X_1, \ldots, X_n)'$ случайных величин $X_1, \ldots, X_n$ называется *n-мерным случайным вектором*, а случайная величина X_i — его i-й компонентой. По аналогии с (МС.1) определяется функция распределения случайного вектора $\boldsymbol{x}$:

$$F(x_1, \ldots, x_n) = \mathrm{P}(X_1 \leqslant x_1, \ldots, X_n \leqslant x_n).$$

Как и в одномерном случае, случайный вектор $\boldsymbol{x}$ называется непрерывным, если его функция распределения имеет смешанную частную производную n-го порядка по всем переменным, а сама эта производная называется *плотностью распределения* случайного вектора $\boldsymbol{x}$:

$$p(x) = p(x_1, \ldots, x_n) = \frac{\partial^n F}{\partial x_1 \cdots \partial x_n}(x).$$

Для непрерывного случайного вектора справедливо равенство, аналогичное (МС.3):

$$\mathrm{P}(\boldsymbol{x} \in A) = \int_A p(x)\,dx = \int_A p(x_1, \ldots, x_n)\,dx_1 \cdots dx_n$$

для любого (измеримого) множества $A \subseteq R^n$. Свойства F1)–F3) и p1), p2) легко обобщаются на многомерный случай. Это мы оставляем читателю. Нетрудно проверить, что если n-мерный случайный вектор $\boldsymbol{x}$ имеет плотность $p_{\boldsymbol{x}}(x) = p_{\boldsymbol{x}}(x_1, \ldots, x_n)$, то любой его подвектор, например (для простоты), $\boldsymbol{y} = (X_1, \ldots, X_k)'$, $k < n$, также имеет плотность

$$p_{\boldsymbol{y}}(x_1, \ldots, x_k) = \int_{R^{n-k}} p_{\boldsymbol{x}}(x_1, \ldots, x_k, x_{k+1}, \ldots, x_n)\,dx_{k+1} \cdots dx_n.$$

Одним из основных понятий теории вероятностей является понятие независимости случайных событий и случайных величин. Случайные величины $X_1, \ldots, X_n$ называются *независимыми*, если

$$\mathrm{P}(X_1 \in A_1, \ldots, X_n \in A_n) = \mathrm{P}(X_1 \in A_1) \cdots \mathrm{P}(X_n \in A_n)$$

для любых (измеримых) множеств $A_i \subseteq R$, $i = 1, \ldots, n$. Можно легко установить, что если компоненты случайного вектора $\boldsymbol{x} = (X_1, \ldots, X_n)'$ независимы, то его функция распределения

распадается в произведение функций распределения компонент, а если, кроме того, каждая случайная величина X_i непрерывна, то вектор $\boldsymbol{x}$ также непрерывен и его плотность есть произведение плотностей компонент:

$$F_{\boldsymbol{x}}(x_1,\dots,x_n) = F_{X_1}(x_1)\cdots F_{X_n}(x_n),$$
$$p_{\boldsymbol{x}}(x_1,\dots,x_n) = p_{X_1}(x_1)\cdots p_{X_n}(x_n).$$

Математическим ожиданием случайного вектора $\boldsymbol{x} = (X_1,\dots,X_n)'$ называется вектор в пространстве R^n, составленный из математических ожиданий компонент X_i: $\mathrm{E}\boldsymbol{x} = (\mathrm{E}X_1,\dots,\mathrm{E}X_n)'$. В многомерном случае свойство Е3) выглядит следующим образом. Пусть вектор $\boldsymbol{x}$ имеет плотность $p(x)$ и пусть $\boldsymbol{g}$: $R^n \to R^m$ — некоторая вектор-функция. Тогда

$$\mathrm{E}\boldsymbol{g}(\boldsymbol{x}) = \int_{R^n} \boldsymbol{g}(x)p(x)\,dx.$$

Аналогом дисперсии для случайного вектора является *матрица ковариаций* или *ковариационная матрица*, определяемая равенством

$$\mathrm{V}(\boldsymbol{x}) = \mathrm{E}\big((\boldsymbol{x}-\mathrm{E}\boldsymbol{x})(\boldsymbol{x}-\mathrm{E}\boldsymbol{x})'\big).$$

Матрица ковариаций n-мерного вектора имеет размерность $n\times n$, ее (i,j)-й элемент есть *ковариация*

$$\mathrm{Cov}(X_i,X_j) = \mathrm{E}((X_i-\mathrm{E}X_i)(X_j-\mathrm{E}X_j))$$

случайных величин X_i и X_j. В частности, на главной диагонали расположены дисперсии компонент вектора $\boldsymbol{x}$. Нетрудно проверить, что матрица ковариаций симметрична и неотрицательно определена.

В дальнейшем часто будет применяться следующий легко доказываемый результат. Пусть $\boldsymbol{x}$ — n-мерный случайный вектор со средним значением $\mathrm{E}\boldsymbol{x}$ и матрицей ковариаций $\mathrm{V}(\boldsymbol{x})$ и пусть $\boldsymbol{b}$ — m-мерный вектор и $\boldsymbol{A}$ — $m\times n$ матрица. Тогда для случайного вектора $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}$ справедливы равенства

$$\mathrm{E}(\boldsymbol{y}) = \boldsymbol{A}\mathrm{E}(\boldsymbol{x})+\boldsymbol{b}, \quad \mathrm{V}(\boldsymbol{y}) = \boldsymbol{A}\mathrm{V}(\boldsymbol{x})\boldsymbol{A}'. \tag{МС.7}$$

Для двух случайных величин X и Y мерой их линейной зависимости служит *коэффициент корреляции*

$$r(X,Y) = \mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{V}(X)\mathrm{V}(Y)}}.$$

Коэффициент корреляции обладает следующими свойствами:

г1) $-1 \leqslant r(X,Y) \leqslant 1$ для любых случайных величин X и Y;

г2) если $r(X,Y)=1$, то $Y = aX+b$, $a>0$;
если $r(X,Y)=-1$, то $Y = aX+b$, $a<0$;

г3) если случайные величины X и Y независимы, то $r(X,Y)=0$.

Случайные величины X и Y, для которых $r(X,Y)=0$, называются *некоррелированными*. Таким образом, в силу г3) независимые случайные величины некоррелированы, обратное, вообще говоря, неверно. Из свойства Е1) легко следует, что для любых случайных величин X и Y

$$\mathrm{V}(X+Y) = \mathrm{V}(X)+\mathrm{V}(Y)+2\,\mathrm{Cov}(X,Y);$$

в частности, для некоррелированных (а тем более для независимых) X и Y выполнено равенство

$$\mathrm{V}(X+Y) = \mathrm{V}(X)+\mathrm{V}(Y). \tag{МС.8}$$

В многомерном случае обобщением ковариации является матрица взаимных ковариаций. Пусть $\boldsymbol{x}$ — n-мерный, а $\boldsymbol{y}$ — k-мерный случайные векторы. *Матрицей взаимных ковариаций* называется $n\times k$ матрица

$$\mathrm{Cov}(\boldsymbol{x},\boldsymbol{y}) = \mathrm{E}\big((\boldsymbol{x}-\mathrm{E}\boldsymbol{x})(\boldsymbol{y}-\mathrm{E}\boldsymbol{y})'\big).$$

Если $\mathrm{Cov}(\boldsymbol{x},\boldsymbol{y}) = 0$, то говорят, что векторы $\boldsymbol{x}$ и $\boldsymbol{y}$ *некоррелированы*. Пусть $\boldsymbol{z} = \boldsymbol{A}\boldsymbol{x}+\boldsymbol{a}$, $\boldsymbol{u} = \boldsymbol{B}\boldsymbol{y}+\boldsymbol{b}$, где $\boldsymbol{A}$: $R^n \to R^p$, $\boldsymbol{B}$: $R^k \to R^q$ — некоторые линейные преобразования и $\boldsymbol{a}\in R^p$, $\boldsymbol{b}\in R^q$ — произвольные (неслучайные) векторы. Проводя непосредственные вычисления, получаем, что

$$
\begin{aligned}
\operatorname{Cov}(\boldsymbol{z},\boldsymbol{u}) &= \operatorname{Cov}(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{a},\boldsymbol{B}\boldsymbol{y}+\boldsymbol{b}) \\
&= \mathrm{E}\Big(\big(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{a}-\mathrm{E}(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{a})\big)\big(\boldsymbol{B}\boldsymbol{y}+\boldsymbol{b}-\mathrm{E}(\boldsymbol{B}\boldsymbol{y}+\boldsymbol{b})\big)'\Big) \\
&= \mathrm{E}\Big(\boldsymbol{A}\big(\boldsymbol{x}-\mathrm{E}(\boldsymbol{x})\big)\big(\boldsymbol{B}(\boldsymbol{y}-\mathrm{E}(\boldsymbol{y}))\big)'\Big) \\
&= \mathrm{E}\Big(\boldsymbol{A}\big(\boldsymbol{x}-\mathrm{E}(\boldsymbol{x})\big)\big(\boldsymbol{y}-\mathrm{E}(\boldsymbol{y})\big)'\boldsymbol{B}'\Big) \\
&= \boldsymbol{A}\mathrm{E}\Big(\big(\boldsymbol{x}-\mathrm{E}(\boldsymbol{x})\big)\big(\boldsymbol{y}-\mathrm{E}(\boldsymbol{y})\big)'\Big)\boldsymbol{B}' \\
&= \boldsymbol{A}\operatorname{Cov}(\boldsymbol{x},\boldsymbol{y})\boldsymbol{B}'. \qquad \text{(МС.9)}
\end{aligned}
$$

2. Условные распределения

Пусть $\boldsymbol{x} \in R^m$ и $\boldsymbol{y} \in R^n$ — случайные векторы, имеющие плотность совместного распределения $p_{\boldsymbol{xy}}(x,y)$, $x \in R^m$, $y \in R^n$. Функция

$$p_{\boldsymbol{x}|\boldsymbol{y}}(x \mid y) = \frac{p_{\boldsymbol{xy}}(x,y)}{\int_{R^m} p_{\boldsymbol{xy}}(x,y)\,dx} = \frac{p_{\boldsymbol{xy}}(x,y)}{p_{\boldsymbol{y}}(y)}$$

называется *плотностью условного распределения* $\boldsymbol{x}$ *при условии, что* $\boldsymbol{y} = y$. Нетрудно проверить, что $p_{\boldsymbol{x}|\boldsymbol{y}}(x \mid y)$ при каждом y, таком что $p_{\boldsymbol{y}}(y) \neq 0$, удовлетворяет (по x) условиям p1), p2). Пусть dV_x, dV_y — малые окрестности точек x и y в пространствах R^m и R^n соответственно. Тогда, используя элементарные понятия условных вероятностей и свойства плотности, можно показать, что если совместная плотность непрерывна в точке (x,y), то имеет место приближенное равенство $\mathrm{P}(\boldsymbol{x} \in dV_x \mid \boldsymbol{y} \in dV_y) \approx p_{\boldsymbol{x}|\boldsymbol{y}}(x \mid y)|dV_x|$, где $|dV_x|$ — объем области dV_x, и равенство становится тем точнее, чем меньше области dV_x, dV_y. Это обстоятельство оправдывает применение термина «плотность распределения».

Условным математическим ожиданием (*conditional expectation*) случайного вектора $\boldsymbol{x}$ при условии, что $\boldsymbol{y} = y$, называется функция

$$\boldsymbol{h}(y) = \mathrm{E}(\boldsymbol{x} \mid \boldsymbol{y} = y) = \int_{R^m} x p_{\boldsymbol{x}|\boldsymbol{y}}(x \mid y)\,dx.$$

Условное математическое ожидание при каждом y удовлетворяет тем же условиям E1)–E3), что и обычное математическое ожидание. В прикладных областях теории вероятностей $\boldsymbol{h}(y)$ называют *функцией регрессии* $\boldsymbol{x}$ *на* $\boldsymbol{y}$. Если у $\boldsymbol{h}(y)$ в качестве аргумента взять случайный вектор $\boldsymbol{y}$, то получим случайный вектор $\boldsymbol{h}(\boldsymbol{y})$, называемый условным математическим ожиданием $\boldsymbol{x}$ при условии $\boldsymbol{y}$ и обозначаемый $\mathrm{E}(\boldsymbol{x} \mid \boldsymbol{y})$. Перечислим его основные свойства.

CE1) Правило повторного условного математического ожидания: если $\boldsymbol{g}\colon R^n \to R^k$ — некоторая (неслучайная) функция и $\boldsymbol{z} = \boldsymbol{g}(\boldsymbol{y})$, то

$$\mathrm{E}(\mathrm{E}(\boldsymbol{x} \mid \boldsymbol{y}) \mid \boldsymbol{z}) = \mathrm{E}(\boldsymbol{x} \mid \boldsymbol{z}).$$

В частности, $\mathrm{E}(\mathrm{E}(\boldsymbol{x} \mid \boldsymbol{y})) = \mathrm{E}(\boldsymbol{x})$.

CE2) Если $g(\boldsymbol{y})$— скалярная случайная величина, то

$$\mathrm{E}(g(\boldsymbol{y})\boldsymbol{x} \mid \boldsymbol{y}) = g(\boldsymbol{y})\mathrm{E}(\boldsymbol{x} \mid \boldsymbol{y}).$$

CE3) Если векторы $\boldsymbol{x}$ и $\boldsymbol{y}$ независимы, то

$$\mathrm{E}(\boldsymbol{x} \mid \boldsymbol{y}) = \mathrm{E}(\boldsymbol{x}).$$

Условное математическое ожидание обладает оптимизационным свойством, аналогичным свойству обычного математического ожидания E4). А именно, пусть X — скалярная случайная величина, $\boldsymbol{y}$ — n-мерный случайный вектор и $f\colon R^n \to R^1$ — произвольная функция. Рассмотрим среднеквадратичное отклонение $\mathrm{E}(X - f(\boldsymbol{y}))^2$ и поставим задачу нахождения функции f, минимизирующей это отклонение. С помощью простых выкладок, используя свойства CE1), CE2), получаем следующий результат.

CE4) Минимальное значение величины $\mathrm{E}(X - f(\boldsymbol{y}))^2$ достигается при $f(\boldsymbol{y}) = \mathrm{E}(X \mid \boldsymbol{y})$.

3. Некоторые специальные распределения

В этом разделе рассматриваются некоторые конкретные случайные величины и случайные векторы, часто используемые в теории вероятностей, математической статистике и их приложениях.

1. *Биномиальное распределение.* Дискретная случайная величина $\nu_n(p)$, принимающая значения $k = 0, \dots, n$ с вероятностями

$$p_k = \mathrm{P}\big(\nu_n(p) = k\big) = C_n^k p^k (1-p)^{n-k}, \quad 0<p<1, \quad C_n^k = \frac{n!}{k!(n-k)!},$$

называется *биномиальной* случайной величиной с параметрами n и p. Случайная величина с таким распределением возникает в *схеме Бернулли*. Предположим, что проводится серия n независимых испытаний, причем каждое испытание имеет два исхода — «успех» или «неудача», и пусть p — вероятность успеха в отдельном испытании. Тогда общее число успехов в этих n испытаниях является биномиальной случайной величиной. Свяжем с каждым i-м испытанием случайную величину $\varepsilon_i = 1$, если в i-м испытании успех, и $\varepsilon_i = 0$, если в i-м испытании неудача. Тогда случайные величины ε_i, $i = 1, \dots, n$, независимы и

$$\nu_n(p) = \sum_{i=1}^{n} \varepsilon_i.$$

Из E1) и (МС.8) легко следует, что

$$\mathrm{E}\nu_n(p) = np, \quad \mathrm{V}\big(\nu_n(p)\big) = np(1-p).$$

2. *Пуассоновское распределение.* Дискретная случайная величина $\Pi(\lambda)$, принимающая значения $k = 0, 1, \dots$ с вероятностями

$$p_k = \frac{\lambda^k}{k!} e^{-\lambda}, \qquad k = 0, 1, \dots, \quad \lambda > 0,$$

называется *пуассоновской* случайной величиной с параметром λ. Пуассоновское распределение широко используется в теории массового обслуживания. Число λ носит название *интенсивность*. Непосредственные вычисления показывают, что

$$\mathrm{E}\big(\Pi(\lambda)\big) = \mathrm{V}\big(\Pi(\lambda)\big) = \lambda.$$

3. *Равномерное распределение.* Непрерывная случайная величина X, плотность распределения которой задается формулой

$$p(x) = \begin{cases} \dfrac{1}{b-a}, & x \in [a, b], \\ 0, & x \notin [a, b], \end{cases}$$

называется *равномерной* на отрезке $[a, b]$. Нетрудно проверить, что

$$\mathrm{E}(X) = \frac{b+a}{2}, \quad \mathrm{V}(X) = \frac{(b-a)^2}{12}.$$

4. *Показательное (экспоненциальное) распределение.* Непрерывная случайная величина X, плотность распределения которой задается формулой

$$p(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geqslant 0, \\ 0, & x < 0, \end{cases}$$

называется *показательной* или *экспоненциальной* с параметром λ. В широком числе случаев показательное распределение описывает время безотказной работы прибора, при этом число λ интерпретируется как интенсивность отказа. Это распределение находит также широкое применение в демографии. Как и раньше, с помощью несложных вычислений получаем

$$\mathrm{E}(X) = \frac{1}{\lambda}, \quad \mathrm{V}(X) = \frac{1}{\lambda^2}.$$

5. *Нормальное (гауссовское) распределение.* Непрерывная случайная величина X, плотность распределения которой задается формулой

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right), \qquad x \in R, \quad m \in R, \quad \sigma > 0.$$

называется *нормальной* или *гауссовской* с параметрами m и σ^2. Часто используется обозначение $X \sim N(m, \sigma^2)$. Нормальная случайная величина с $m = 0$ и $\sigma^2 = 1$ называется *стандартной нормальной величиной*. Если $X \sim N(m, \sigma^2)$, то случайная величина $Z = (X - m)/\sigma$ является стандартной нормальной. Гауссовское

распределение играет исключительно важную роль в теории вероятностей и математической статистике. Если исследуемая случайная величина формируется под воздействием большого числа независимых аддитивных случайных факторов, то, согласно приводимой ниже центральной предельной теореме, можно считать, что ее распределение является приближенно нормальным. Можно показать, что

$$\mathrm{E}(X) = m, \quad \mathrm{V}(X) = \sigma^2.$$

Ниже мы более подробно опишем свойства нормального распределения.

6. *Логарифмически нормальное распределение.* Пусть $X \sim N(m, \sigma^2)$. Случайная величина $Y = e^X$ называется *логарифмически нормальной.* Можно показать, что плотность распределения этой величины определяется формулой

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma x} \exp\left(-\frac{(\ln x - m)^2}{2\sigma^2}\right), \quad x > 0; \quad p(x) = 0, \quad x \leqslant 0.$$

Логарифмически нормальное распределение возникает в ситуации, когда исследуемая случайная величина формируется под воздействием большого числа мультипликативных случайных факторов. Можно показать, что

$$\mathrm{E}(Y) = \exp(m + \sigma^2/2), \quad \mathrm{V}(Y) = \exp(2m + \sigma^2)(\exp(\sigma^2) - 1).$$

7. *χ^2-распределение.* Пусть $\varepsilon_1, \ldots, \varepsilon_n$ — независимые стандартные нормальные случайные величины (т.е. $\varepsilon_i \sim N(0,1)$, $i = 1, \ldots, n$). Говорят, что случайная величина $\chi^2(n) = \sum_{i=1}^n \varepsilon_i^2$ имеет *χ^2-распределение с n степенями свободы.* Плотность распределения величины $\chi^2(n)$ задается формулой

$$p(x) = \frac{1}{2^{n/2}\Gamma(n/2)} \left(\frac{x}{2}\right)^{n/2-1} \exp\left(-\frac{x}{2}\right), \quad x > 0; \quad p(x) = 0, \quad x \leqslant 0,$$

где Γ — гамма-функция, определяемая равенством

$$\Gamma(t) = \int_0^{+\infty} x^{t-1} e^{-x}\, dx, \qquad t > 0,$$

в частности, $\Gamma(k) = (k-1)!$ для целых k.

Нетрудно проверить, что

$$\mathrm{E}\chi^2 = n, \quad \mathrm{V}(\chi^2) = 2n.$$

8. *Распределение Стьюдента (t-распределение).* Пусть $\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_n$ — независимые стандартные нормальные случайные величины. Распределение случайной величины

$$t(n) = \frac{\varepsilon_0}{\sqrt{\frac{1}{n}\sum_{i=1}^n \varepsilon_i^2}} = \frac{\varepsilon_0}{\sqrt{\frac{1}{n}\chi^2(n)}}$$

называется *распределением Стьюдента* или *t-распределением с n степенями свободы.* Плотность распределения величины $t(n)$ задается формулой

$$p(x) = \frac{\Gamma((n+1)/2)}{(n\pi)^{1/2}\Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2}.$$

При $n = 1$ соответствующее распределение называют *распределением Коши.*

Можно показать, что при $n > 2$

$$\mathrm{E}t(n) = 0, \quad \mathrm{V}(t(n)) = \frac{n}{n-2}.$$

9. *Распределение Фишера (F-распределение).* Пусть $\varepsilon_1, \ldots, \varepsilon_m$, $\eta_1, \ldots, \eta_n$ — совокупность независимых стандартных нормальных случайных величин. Распределение случайной величины

$$F(m,n) = \frac{\frac{1}{m}\sum_{i=1}^m \varepsilon_i^2}{\frac{1}{n}\sum_{i=1}^n \eta_i^2} = \frac{\frac{1}{m}\chi^2(m)}{\frac{1}{n}\chi^2(n)}$$

называется *распределением Фишера* или *F-распределением с (m,n) степенями свободы.* Плотность распределения величины $F(m,n)$ задается формулой

$$p(x) = \frac{\Gamma[(m+n)/2]m^{m/2}n^{n/2}}{\Gamma(m/2)\Gamma(n/2)} \frac{x^{m/2-1}}{(mx+n)^{(m+n)/2}}, \qquad x > 0;$$
$$p(x) = 0, \qquad x \leqslant 0.$$

Можно показать, что при $n > 4$

$$\mathrm{E}F(m,n) = \frac{n}{n-2}, \quad \mathrm{V}(F(m,n)) = \frac{2n^2(m+n-2)}{m(n-4)(n-2)^2}.$$

Распределения χ^2, Стьюдента и Фишера применяются в статистике при построении доверительных интервалов для оцениваемых параметров и при проверке гипотез.

В некоторых областях математической и прикладной статистики используются *нецентральные* аналоги распределений χ^2, Стьюдента и Фишера.

10. *Нецентральное χ^2-распределение.* Пусть $X_1, \ldots, X_n$ — независимые нормальные случайные величины: $X_i \sim N(m_i, 1)$. Тогда говорят, что случайная величина $Y = X_1^2 + \ldots + X_n^2$ имеет нецентральное χ^2-распределение. Это распределение зависит только от двух параметров: n — число степеней свободы и $\lambda = \sum_{i=1}^n m_i^2$ — параметр сдвига (параметр нецентральности). Это распределение обозначается $\chi^2(n, \lambda)$.

Покажем, что распределение $\chi^2(n, \lambda)$ действительно зависит только от двух параметров. Обозначим через $\boldsymbol{m} = (m_1, \ldots, m_n)'$ — вектор средних значений и через $\|\boldsymbol{m}\|$ — его длину. Пусть $\boldsymbol{Q}$ — ортогональная матрица, у которой первая строчка является вектором $\boldsymbol{m}'/\|\boldsymbol{m}\|$, а остальные дополняют вектор $\boldsymbol{m}'/\|\boldsymbol{m}\|$ до ортонормированного базиса. Обозначим вектор, состоящий из случайных величин X_i, через $\boldsymbol{x} = (X_1, \ldots, X_n)'$, и пусть $\boldsymbol{z} = \boldsymbol{Q}\boldsymbol{x}$ его линейное преобразование; $\boldsymbol{z} = (Z_1, \ldots, Z_n)'$. В силу ортогональности матрицы $\boldsymbol{Q}$ имеем $Y = X_1^2 + \ldots + X_n^2 = \|\boldsymbol{x}\|^2 = \|\boldsymbol{z}\|^2 = Z_1^2 + \ldots + Z_n^2$. В силу (приложение МС, п. 4, N5) получаем, что $Z_1, \ldots, Z_n$ — независимые нормальные случайные величины, такие что $Z_1 \sim N(\|\boldsymbol{m}\|, 1)$ и $Z_i \sim N(0, 1)$, $i = 2, \ldots, n$. Отсюда следует, что распределение зависит только от $\|\boldsymbol{m}\|$.

Нетрудно видеть, что при $\lambda = 0$ нецентральное распределение $\chi^2(n, \lambda)$ совпадает с распределением $\chi^2(n)$.

Распределение $\chi^2(n, \lambda)$ обладает следующими свойствами:

1) Пусть Y_1, Y_2 — две независимых случайных величины с распределениями $Y_i \sim \chi^2(n_i, \lambda_i)$, $i = 1, 2$, тогда случайная величина $Y = Y_1 + Y_2$ имеет тоже нецентральное χ^2-распределение: $Y \sim \chi^2(n, \lambda)$, где $\lambda = \lambda_1 + \lambda_2$ и $n = n_1 + n_2$.

2) Пусть $Y \sim \chi^2(n, \lambda)$, тогда $\mathrm{E}(Y) = n + \lambda$ и $\mathrm{V}(Y) = 2n + 4\lambda$.

3) Плотность распределения $\chi^2(n, \lambda)$ задается формулой

$$p(x) = \exp\left(-\frac{\lambda}{2}\right) \exp\left(-\frac{x}{2}\right) \frac{x^{n/2-1}}{2^{n/2}} \sum_{i=0}^{\infty} \frac{x^i}{i!} \frac{\lambda^i}{2^{2i}} \frac{1}{\Gamma(i + n/2)}$$

при $x > 0$ и $p(x) = 0$ при $x \leqslant 0$.

11. *Нецентральное распределение Стьюдента.* Нецентральным распределением Стьюдента с n степенями свободы и параметром сдвига m называется распределение случайной величины

$$Z = \frac{X}{\sqrt{(Y/n)}},$$

где случайные величины X и Y независимы и имеют распределения соответственно $N(m, 1)$ и $\chi^2(n)$.

Плотность нецентрального распределения Стьюдента задается следующей формулой:

$$p(x) = \frac{n^{n/2}}{\Gamma(n/2)\Gamma(1/2)} \frac{\exp(-m^2/2)}{(n + x^2)^{(n+1)/2}} \times \sum_{i=0}^{\infty} \Gamma\left(\frac{n+i+1}{2}\right) \frac{m^i}{i!} \left(\frac{2x^2}{n + x^2}\right)^{i/2} (\operatorname{sign} x)^i.$$

12. *Нецентральное распределение Фишера.* Нецентральным распределением Фишера $F(n_1, n_2, \lambda_1, \lambda_2)$ со степенями свободы n_1 и n_2 и параметрами сдвига λ_1 и λ_2 называется распределение случайной величины

$$Z = \frac{Y_1/n_1}{Y_2/n_2},$$

где Y_1 и Y_2 — независимые случайные величины с распределениями соответственно $\chi^2(n_1, \lambda_1)$ и $\chi^2(n_2, \lambda_2)$.

Обычно в приложениях используется распределение с $\lambda_2 = 0$. При $n_1 > 4$ математическое ожидание и дисперсия распределения $F(n_1, n_2, \lambda_1, 0)$ равны:

$$\mathrm{E}(Z) = \frac{n_2(n_1 + \lambda_1)}{n_1(n_2 - 2)},$$

$$\mathrm{V}(Z) = 2\left(\frac{n_2}{n_1}\right)^2 \frac{(n_1 + \lambda_1)^2 + (n_1 + 2\lambda_1)(n_2 - 2)}{(n_2 - 2)^2(n_2 - 4)}.$$

4. Многомерное нормальное распределение

Случайный вектор $x = (X_1, \ldots, X_n)'$ называется *невырожденным нормальным (гауссовским) случайным вектором*, если плотность его распределения задается равенством

$$p(x) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-m)'\Sigma^{-1}(x-m)\right), \qquad x \in R^n, \tag{МС.10}$$

где $m \in R^n$ — (произвольный) вектор, Σ — симметричная, положительно определенная матрица (т. е. $\Sigma = \Sigma'$ и $\Sigma > 0$). В случае $n = 1$ получаем нормальную случайную величину, введенную в п. 3. Вектор m и матрица Σ являются параметрами распределения, и обычно используется обозначение $X \sim N(m, \Sigma)$. Нормальный вектор ε, у которого $m = 0$, а $\Sigma = I$ (единичная матрица), называется *стандартным нормальным вектором*. Если матрица Σ вырождена (но неотрицательно определена), то можно определить вырожденное нормальное распределение. Пусть $\operatorname{rank}(\Sigma) = k < n$ и $L = m + \Sigma R^n$ — k-мерная гиперплоскость в R^n, где Σ — линейный оператор в R^n, задаваемый матрицей Σ. Известно, что в подпространстве ΣR^n оператор Σ обратим, обозначим через $|\Sigma|$ ($\neq 0$) определитель матрицы этого оператора в ΣR^n, а через Σ^{-1} — матрицу соответствующего обратного оператора. Тогда распределение, сосредоточенное в L и имеющее плотность (МС.10) (относительно k-мерного объема в L), называется *вырожденным нормальным распределением*. Перечислим без доказательства основные свойства многомерного нормального распределения.

N1) Если $x \sim N(m, \Sigma)$, то $\mathrm{E}(x) = m$, $\mathrm{V}(x) = \Sigma$.

N2) Любой подвектор нормального вектора также является нормальным вектором.

N3) Пусть x и y — два независимых нормальных вектора. Тогда объединенный вектор $z = \begin{bmatrix} x \\ y \end{bmatrix}$ также является нормальным.

N4) Если $z = \begin{bmatrix} x \\ y \end{bmatrix}$ — нормальный вектор и его компоненты x и y некоррелированы, то они независимы.

N5) Пусть B: $R^n \to R^k$ — линейное преобразование пространства R^n в R^k, B — его матрица и l — произвольный вектор в R^k. Тогда если $x \sim N(m, \Sigma)$, то случайный вектор $y = Bx + l$ является нормальным с параметрами $Bm + l$ и $B\Sigma B'$. (Преобразование пространства R^n в R^k вида $y = Bx + l$, являющееся композицией линейного преобразования B и параллельного переноса на вектор l, называется *аффинным преобразованием*.)

В частности,

а) линейная комбинация компонент гауссовского вектора есть гауссовская случайная величина;

б) ортогональное линейное преобразование стандартного нормального вектора есть стандартный нормальный вектор.

Пусть $x \sim N(m, \Sigma)$. Поскольку матрица Σ симметрична и неотрицательно определена, то, как известно (приложение ЛА, п. 15), все ее собственные значения λ_i, $i = 1, \ldots, n$, неотрицательны и существует ортогональная матрица P, такая что $\Lambda = P'\Sigma P$, где Λ — диагональная матрица, на главной диагонали которой стоят числа λ_i, $i = 1, \ldots, n$. Тогда вектор $s = P'x - P'm$ в силу N5) является гауссовским, а из (МС.7) следует, что $\mathrm{E}s = 0$ и $\mathrm{V}(s) = \Lambda$. Это означает, что компоненты вектора s некоррелированы, а в силу N4) и независимы. Таким образом,

$$x = Ps + m,$$

где матрица P ортогональна, а вектор s имеет нулевое среднее и независимые компоненты. Обозначим через $\Lambda^{1/2}$ диагональную матрицу, полученную из Λ извлечением квадратных корней из ее элементов, и пусть ε — стандартный нормальный вектор. Тогда, вновь используя (МС.7), получаем, что вектор $y = P\Lambda^{1/2}\varepsilon + m$ имеет среднее m и матрицу ковариаций Σ, т. е. совпадает по распределению с вектором x. Итак,

N6) Любой гауссовский вектор может быть получен аффинным преобразованием из стандартного гауссовского вектора и ортогональным аффинным преобразованием из вектора с независимыми компонентами.

Из (МС.9) и N4) вытекает следующее свойство:

N7) Пусть $\boldsymbol{\varepsilon}$ — стандартный n-мерный нормальный вектор и $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{\varepsilon} + \boldsymbol{a}$, $\boldsymbol{y} = \boldsymbol{B}\boldsymbol{\varepsilon} + \boldsymbol{b}$, где $\boldsymbol{A}$: $R^n \to R^p$, $\boldsymbol{B}$: $R^n \to R^q$ — некоторые линейные преобразования и $\boldsymbol{a} \in R^p$, $\boldsymbol{b} \in R^q$ — произвольные (неслучайные) векторы. Тогда $\mathrm{Cov}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{A}\boldsymbol{B}'$, в частности, векторы $\boldsymbol{x}$ и $\boldsymbol{y}$ независимы тогда и только тогда, когда $\boldsymbol{A}\boldsymbol{B}' = \boldsymbol{0}$.

Пусть $\boldsymbol{M}$ — идемпотентная $n \times n$ матрица, $\mathrm{rank}(\boldsymbol{M}) = r$ (см. приложение ЛА, п. 16), а $\boldsymbol{\varepsilon}$ — стандартный n-мерный гауссовский вектор. Как известно (см. приложение ЛА, п. 13, п. 16), матрицу $\boldsymbol{M}$ можно представить в виде $\boldsymbol{M} = \boldsymbol{O}'\boldsymbol{\Lambda}\boldsymbol{O}$, где $\boldsymbol{O}$ — ортогональная матрица, а $\boldsymbol{\Lambda}$ — диагональная матрица, на главной диагонали которой расположены единицы и нули, причем число единиц равно рангу $\boldsymbol{M}$. Рассмотрим случайную величину $\chi^2 = \boldsymbol{\varepsilon}'\boldsymbol{M}\boldsymbol{\varepsilon}$. Имеем

$$\chi^2 = \boldsymbol{\varepsilon}'\boldsymbol{M}\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}'\boldsymbol{O}'\boldsymbol{\Lambda}\boldsymbol{O}\boldsymbol{\varepsilon} = (\boldsymbol{O}\boldsymbol{\varepsilon})'\boldsymbol{\Lambda}\boldsymbol{O}\boldsymbol{\varepsilon} = \boldsymbol{s}'\boldsymbol{\Lambda}\boldsymbol{s},$$

где в силу N5) вектор $\boldsymbol{s}$ является стандартным гауссовским вектором. Отсюда следует, что χ^2 представляет сумму квадратов независимых стандартных нормальных случайных величин в количестве, равном рангу матрицы $\boldsymbol{M}$. Таким образом,

N8) Случайная величина $\chi^2 = \boldsymbol{\varepsilon}'\boldsymbol{M}\boldsymbol{\varepsilon}$ имеет распределение $\chi^2(r)$, где $r = \mathrm{rank}(\boldsymbol{M})$.

Аналогичным образом устанавливается следующий результат.

N9) Пусть $\boldsymbol{x} \sim N(\boldsymbol{m}, \boldsymbol{\Sigma})$ и $n \times n$ матрица $\boldsymbol{\Sigma}$ невырождена. Тогда случайная величина $(\boldsymbol{x} - \boldsymbol{m})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{m})$ имеет распределение $\chi^2(n)$.

Выделим два важных частных случая применения свойств N7) и N8). Рассмотрим $n \times n$ матрицу

$$\boldsymbol{N} = \frac{1}{n}\begin{bmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{bmatrix}.$$

Нетрудно проверить, что $\boldsymbol{N}$ идемпотентная матрица: $\boldsymbol{N}^2 = \boldsymbol{N}$, и $\mathrm{rank}(\boldsymbol{N}) = 1$. Легко также видеть, что для любого $n \times 1$ вектора $\boldsymbol{x} = (X_1, \dots, X_n)'$ вектор $\boldsymbol{N}\boldsymbol{x}$ имеет одинаковые компоненты, равные $\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i$. Обозначим $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{N}$, где $\boldsymbol{I}$ — единичная матрица n-го порядка. Тогда матрица $\boldsymbol{M}$ также идемпотентная, $\mathrm{rank}(\boldsymbol{M}) = n - 1$ и $\boldsymbol{M}\boldsymbol{N}' = \boldsymbol{M}\boldsymbol{N} = \boldsymbol{0}$. Заметим, наконец, что $\boldsymbol{x}'\boldsymbol{M}\boldsymbol{x} = \boldsymbol{x}'\boldsymbol{M}^2\boldsymbol{x} = ((\boldsymbol{I} - \boldsymbol{N})\boldsymbol{x})'((\boldsymbol{I} - \boldsymbol{N})\boldsymbol{x}) = \sum_{i=1}^n (X_i - \overline{X})^2$. Предположим теперь, что $X_1, \dots, X_n$ — независимые одинаково распределенные нормальные случайные величины с параметрами m и σ^2. Тогда вектор $\boldsymbol{\varepsilon}$ с компонентами $\varepsilon_i = \dfrac{X_i - m}{\sigma}$, $i = 1, \dots, n$ является стандартным нормальным вектором и $\overline{\varepsilon} = \dfrac{\overline{X} - m}{\sigma}$, $\displaystyle\sum_{i=1}^n (\varepsilon - \overline{\varepsilon})^2 = \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \overline{X})^2$. Таким образом, в силу N7) и N8) имеем

N10) Случайные величины $\overline{X}$ и $\sum_{i=1}^n (X_i - \overline{X})^2$ независимы.

N11) Случайная величина $\displaystyle\frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \overline{X})^2$ имеет распределение $\chi^2(n-1)$.

Замечая, что $\mathrm{V}(\overline{\varepsilon}) = \dfrac{1}{\sqrt{n}}$ и

$$\frac{\overline{\varepsilon}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n (\varepsilon_i - \overline{\varepsilon})^2}} = \frac{(\overline{X} - m)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2}},$$

получаем следующее свойство:

N12) Случайная величина

$$\frac{(\overline{X}-m)\sqrt{n}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2}}$$

имеет распределение Стьюдента с $n-1$ степенями свободы.

Свойства N10)–N12) широко используются в статистике при построении интервальных оценок неизвестных параметров и проверке статистических гипотез.

Пусть $\boldsymbol{z}=\begin{bmatrix}\boldsymbol{x}\\ \boldsymbol{y}\end{bmatrix}$ — нормальный вектор, $\mathrm{E}\boldsymbol{x}=\boldsymbol{m}_{\boldsymbol{x}}$, $\mathrm{E}\boldsymbol{y}=\boldsymbol{m}_{\boldsymbol{y}}$, и пусть матрица $\mathrm{V}(\boldsymbol{y})$ невырождена.

N13) Условное распределение $\boldsymbol{x}$ при условии $\boldsymbol{y}=y$ является нормальным. При этом

$$\mathrm{E}(\boldsymbol{x}\mid \boldsymbol{y}=y)=\mathrm{Cov}(\boldsymbol{x},\boldsymbol{y})[\mathrm{V}(\boldsymbol{y})]^{-1}(y-\boldsymbol{m}_{\boldsymbol{y}})+\boldsymbol{m}_{\boldsymbol{x}}.$$

Иными словами, функция регрессии $\boldsymbol{x}$ на $\boldsymbol{y}$ является линейной.

5. Закон больших чисел. Центральная предельная теорема

Результаты, касающиеся асимптотического поведения последовательностей случайных величин, в теории вероятностей принято называть *предельными теоремами*. Простейшими из них являются закон больших чисел (ЗБЧ) и центральная предельная теорема (ЦПТ). Введем необходимые понятия.

Определение 1. Последовательность случайных величин X_1, $X_2,\ldots$ сходится *почти наверное* к случайной величине X ($\lim_{n\to\infty}X_n=X$ п. н.), если при каждой случайной реализации ω числовая последовательность $X_1(\omega),X_2(\omega),\ldots$ сходится к числу $X(\omega)$.[1]

[1] В общепринятом определении требуется, чтобы сходимость имела место не для каждой случайной реализации ω, а для всех ω из множества полной вероятности. Однако определение этого понятия потребовало бы изложения более общей математической теории меры.

Определение 2. Последовательность случайных величин X_1, $X_2,\ldots$ сходится к случайной величине X *по вероятности* ($p\lim_{n\to\infty}X_n=X$, $p=\mathrm{probability}$), если

$$\lim_{n\to\infty}\mathrm{P}(|X_n-X|>\varepsilon)=0$$

для любого $\varepsilon>0$.

Определение 3. Последовательность случайных величин X_1, $X_2,\ldots$, имеющих функции распределения $F_1(x),F_2(x),\ldots$, сходится *по распределению* к случайной величине X с функцией распределения $F(x)$, ($d\lim_{n\to\infty}X_n=X$) или $X_n\xrightarrow{d}X$ ($d=\mathrm{distribution}$), если $\lim_{n\to\infty}F_n(x)=F(x)$ в каждой точке x, где функция F непрерывна.

Можно показать, что из сходимости почти наверное следует сходимость по вероятности, а из сходимости по вероятности следует сходимость по распределению. Кроме того, доказывается, что из последовательности, сходящейся по вероятности, можно извлечь подпоследовательность, сходящуюся почти наверное.

Приведем достаточно простую, но важную теорему.

Неравенство Чебышева. *Пусть X — неотрицательная случайная величина. Тогда для любого числа $c>0$ справедливо неравенство*

$$\mathrm{P}(X\geqslant c)\leqslant\frac{\mathrm{E}(X)}{c}. \qquad (\text{МС.11})$$

Действительно, предполагая для простоты, что у случайной величины X есть плотность распределения $p(x)$, имеем

$$\mathrm{E}(X)=\int_0^{+\infty}xp(x)\,dx\geqslant\int_c^{+\infty}xp(x)\,dx\geqslant c\int_c^{+\infty}p(x)\,dx\geqslant c\mathrm{P}(X\geqslant c),$$

что эквивалентно (МС.11).

Следствие. *Для любой случайной величины Y и для любого числа $\varepsilon>0$ справедливо неравенство*

$$\mathrm{P}\big(|Y-\mathrm{E}(Y)|>\varepsilon\big)\leqslant\frac{\mathrm{V}(Y)}{\varepsilon^2}. \qquad (\text{МС.12})$$

Действительно, полагая в (МС.11) $X = (Y - \mathrm{E}(Y))^2$, $c = \varepsilon^2$ имеем:

$$\mathrm{P}(|Y - \mathrm{E}Y| > \varepsilon) = \mathrm{P}((Y - \mathrm{E}Y)^2 > \varepsilon^2) \leqslant \frac{\mathrm{V}(Y)}{\varepsilon^2}.$$

Пусть $X_1, X_2, \ldots$ — последовательность независимых одинаково распределенных случайных величин, $\mathrm{E}(X_i) = m$, $\mathrm{V}(X_i) = \sigma^2$. Обозначим

$$S_n = \sum_{i=1}^{n} X_i.$$

Закон больших чисел в форме Чебышева. *Имеет место равенство*

$$p\lim_{n\to\infty} \frac{1}{n} S_n = m.$$

В самом деле, из свойств E1) и (МС.8) следует, что $\mathrm{E}\left(\frac{1}{n}S_n\right) = m$, $\mathrm{V}\left(\frac{1}{n}S_n\right) = \frac{\sigma^2}{n}$. Поэтому, полагая в (МС.12) $Y = \frac{1}{n}S_n$, получаем:

$$\mathrm{P}\left(\left|\frac{1}{n}S_n - m\right| > \varepsilon\right) \leqslant \frac{\sigma^2}{n\varepsilon^2} \to 0 \quad \text{при} \quad n \to \infty$$

при любом $\varepsilon > 0$.

С помощью более сложных рассуждений можно установить более общий результат.

Усиленный закон больших чисел. *Имеет место равенство*

$$\lim_{n\to\infty} \frac{1}{n} S_n = m \quad \text{п.н.}$$

Введем случайную величину

$$T_n = \frac{S_n - mn}{\sigma\sqrt{n}}.$$

Нетрудно проверить, что $\mathrm{E}(T_n) = 0$, $\mathrm{V}(T_n) = 1$. Справедлив следующий фундаментальный результат.

Центральная предельная теорема. *Последовательность $T_1, T_2, \ldots$ сходится по распределению к стандартной нормальной случайной величине. В частности,*

$$\mathrm{P}(a < T_n \leqslant b) \to \frac{1}{\sqrt{2\pi}} \int_a^b \exp\left(-\frac{x^2}{2}\right) dx \quad \text{при} \quad n \to \infty$$

для любых чисел $a < b$.

Эта теорема объясняет ту исключительную роль, которую играет в теории вероятностей и прикладных вопросах нормальное распределение. Согласно ЦПТ, сумма большого числа случайных величин имеет приближенно нормальное распределение независимо от индивидуального распределения слагаемых.

Сформулируем еще одну общую теорему, касающуюся сходимости по вероятности. Предположим, что задано k последовательностей случайных величин $\{X_n^j,\ n = 1, 2, \ldots\}$, $j = 1, \ldots, k$, и числовая функция $g(x_1, \ldots, x_k)$.

Теорема Слуцкого. *Пусть каждая из последовательностей $\{X_n^j\}$ сходится по вероятности к константе: $p\lim_{n\to\infty} X_n^j = c_j$, $j = 1, \ldots, k$, и пусть функция g непрерывна в точке $(c_1, \ldots, c_k)$. Тогда $p\lim_{n\to\infty} g(X_n^1, \ldots, X_n^k) = g(c_1, \ldots, c_k)$.*

6. Основные понятия и задачи математической статистики

Случайная выборка. Последовательность наблюдений $X_1, \ldots, X_n$ называется *случайной выборкой объема n*, если $X_1, \ldots, X_n$ получены как независимые реализации некоторой случайной величины X с распределением $F(x)$. При этом также говорят, что $X_1, \ldots, X_n$ есть *выборка из генеральной совокупности X* (или $F(x)$). С теоретико-вероятностной точки зрения случайная выборка $X_1, \ldots, X_n$ может рассматриваться как последовательность независимых случайных величин, имеющих одно и то же распределение $F(x)$.

Выборочные статистики. Пусть $X_1, \ldots, X_n$ — случайная выборка. С ней связаны следующие величины, называемые *выборочными статистиками*:

выборочное среднее: $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$;

выборочная дисперсия: $\mathrm{Var}(X) = s_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$;

размах: $d = \max_{1\leqslant i\leqslant n}\{X_i\} - \min_{1\leqslant i\leqslant n}\{X_i\}$.

Если есть еще одна случайная выборка $Y_1, \ldots, Y_n$, то определяются также:

выборочная ковариация:

$$\mathrm{Cov}(X,Y) = c_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y});$$

выборочный коэффициент корреляции: $r_{XY} = \frac{c_{XY}}{s_X s_Y}$.

То обстоятельство, что в формулах для выборочной дисперсии и ковариации присутствует множитель $1/(n-1)$, а не $1/n$, будет пояснено ниже.

Эти понятия легко обобщаются на случай многомерных наблюдений.

Оценивание параметров. Предположим, что распределение случайной величины X (генеральной совокупности) зависит от некоторого (возможно, многомерного) неизвестного параметра $\boldsymbol{\theta}$: $F(x) = F(x;\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta \subseteq R^m$. Общая задача оценивания заключается в получении каких-либо выводов о параметре $\boldsymbol{\theta}$ на основании наблюдений $X_1, \ldots, X_n$. Различают точечное и интервальное оценивание. Любая функция φ_n: $R^n \to \Theta$ называется *точечной оценкой* (или просто оценкой) параметра $\boldsymbol{\theta}$. Часто используется обозначение $\widehat{\boldsymbol{\theta}} = \varphi_n(X_1, \ldots, X_n)$. В русскоязычной литературе по статистике, как правило, одним и тем же термином «оценка» называют как функцию φ_n, так и ее значение $\widehat{\boldsymbol{\theta}}$ для конкретных наблюдений $X_1, \ldots, X_n$. В английском языке эти объекты различают, называя φ_n *estimator*, а величину $\widehat{\boldsymbol{\theta}}$ — *estimate*. Поэтому правильнее было бы называть функцию φ_n методом оценивания, сохранив название «оценка» за величиной $\widehat{\boldsymbol{\theta}}$, однако такая терминология не является общепринятой. Заметим, что оценка, являясь функцией случайных наблюдений, также есть случайная величина.

Множество $D_n(X_1, \ldots, X_n) \subseteq \Theta$ называется *доверительным множеством с уровнем доверия* $1-\alpha$ (или $100(1-\alpha)$%-ным доверительным множеством), если $\mathrm{P}(\boldsymbol{\theta} \in D_n(X_1, \ldots, X_n)) = 1-\alpha$, где $0 < \alpha < 1$. Часто это множество называют *интервальной оценкой* параметра $\boldsymbol{\theta}$ с уровнем доверия $1-\alpha$. Термин «интервальная» связан с тем, что в случае одномерного параметра в качестве доверительных множеств рассматриваются, как правило, интервалы.

Проверка гипотез. Одной из традиционных задач статистики является проверка гипотез. Простейшая схема выглядит следующим образом. Выдвигается основная, или *нулевая*, гипотеза о том, что неизвестный параметр принадлежит некоторому заданному подмножеству $Z_0 \subset \Theta$ и *альтернативная* гипотеза о том, что параметр принадлежит другому подмножеству $Z_1 \subset \Theta$. Обычно используются обозначения H_0: $\boldsymbol{\theta} \in Z_0$ и H_1: $\boldsymbol{\theta} \in Z_1$. Требуется на основании наблюдений $X_1, \ldots, X_n$ принять (проверить) нулевую гипотезу H_0 или отвергнуть ее в пользу альтернативной гипотезы H_1.

Ниже мы более подробно рассмотрим задачи оценивания и проверки гипотез.

7. Оценивание параметров

Пусть $\widehat{\boldsymbol{\theta}} = \varphi_n(X_1, \ldots, X_n)$ — некоторая оценка параметра $\boldsymbol{\theta}$. Она называется *несмещенной*, если

$$\mathrm{E}(\varphi_n(X_1, \ldots, X_n)) = \boldsymbol{\theta} \qquad \text{(МС.13)}$$

при любом $\boldsymbol{\theta} \in \Theta$. В (МС.13) математическое ожидание берется по распределению, порожденному функцией $F(x;\boldsymbol{\theta})$. Чтобы подчеркнуть это, условие несмещенности часто записывают в виде $\mathrm{E}_{\boldsymbol{\theta}}(\widehat{\boldsymbol{\theta}}) = \boldsymbol{\theta}$. Выборочное среднее $\overline{X}$ является примером несмещенной оценки математического ожидания случайной величины X, определяющей генеральную совокупность. Действительно, по-

скольку $\mathrm{E}(X_i) = m$ для каждого i, то из свойства E1) следует, что $\mathrm{E}(\overline{X}) = m$. Сложнее проверяется, что $\mathrm{E}(s^2) = \mathrm{V}(X)$, т. е. выборочная дисперсия дает несмещенную оценку теоретической дисперсии. Этим обстоятельством объясняется наличие сомножителя $1/(n-1)$. Аналогичное утверждение справедливо и для выборочной ковариации. Рекомендуем читателю самостоятельно провести соответствующие выкладки.

Оценка $\widehat{\theta} = \varphi_n(X_1, \ldots, X_n)$ называется *состоятельной*, если $p\lim_{n\to\infty} \varphi_n(X_1, \ldots, X_n) = \theta$. Закон больших чисел в форме Чебышева утверждает, что выборочное среднее есть состоятельная оценка математического ожидания генеральной совокупности.

Оценка $\widehat{\theta} = \varphi_n(X_1, \ldots, X_n)$ называется *эффективной*, если

$$\begin{aligned}\mathrm{E}\big((\varphi_n(X_1, \ldots, X_n) - \theta)(\varphi_n(X_1, \ldots, X_n) - \theta)'\big)\\ = \min_g \mathrm{E}\big((g(X_1, \ldots, X_n) - \theta)(g(X_1, \ldots, X_n) - \theta)'\big),\end{aligned}$$

где минимум берется по всем возможным оценкам g параметра θ. Если θ — одномерный параметр, то эффективность в классе несмещенных оценок означает минимальность дисперсии.

Содержательно несмещенность оценки означает, что при ее использовании мы не получаем систематической ошибки; состоятельность оценки гарантирует приближение оценки к истинному значению параметра при увеличении объема выборки, а эффективная оценка является наилучшей в смысле минимума среднеквадратичного отклонения. Отметим, что несмещенность и эффективность — это свойства, не зависящие от объема выборки n, в то время как состоятельность является асимптотическим свойством при стремлении n к бесконечности.

Смещенность или несмещенность конкретной оценки проверяется, как правило, непосредственными вычислениями. Для установления состоятельности можно пользоваться предельными теоремами типа закона больших чисел. Проверить эффективность оценки обычно существенно труднее. В данном приложении мы сформулируем один общий результат, имеющий непосредственное отношение к проблеме эффективности.

Неравенство Рао–Крамера. *Пусть $p(x;\theta)$ — плотность распределения случайного вектора $x = (X_1, \ldots, X_n)$, $x \in R^n$, зависящая от одномерного параметра θ. Предположим, что выполнены следующие условия:*

1) *множество $G = \{x \in R^n\colon\ p(x;\theta) > 0\}$ не зависит от θ (условие регулярности),*

2) *плотность $p(x;\theta)$ дифференцируема по θ при каждом $x \in G$ и*

$$I_n = \int_G \left(\frac{\partial \ln p(x;\theta)}{\partial\theta}\right)^2 p(x;\theta)\,dx = \mathrm{E}\left(\frac{\partial \ln p(X;\theta)}{\partial\theta}\right)^2 < \infty. \tag{МС.14}$$

Тогда для любой несмещенной оценки $\widehat{\theta} = \varphi_n(X_1, \ldots, X_n)$ параметра θ выполнено неравенство (Рао–Крамера)

$$\mathrm{E}(\widehat{\theta} - \theta)^2 = \mathrm{V}(\widehat{\theta}) \geqslant \frac{1}{I_n}. \tag{МС.15}$$

Число I_n в (МС.14) называется *информационным количеством* θ в x. Если компоненты $X_1, \ldots, X_n$ вектора x независимы и одинаково распределены с плотностью $p(x;\theta)$, $x \in R^1$, то можно показать, что $I_n = nI_1$, где I_1 — информационное количество θ в одной компоненте X_k: $I_1 = \mathrm{E}[\partial \ln p(X_k;\theta)/\partial\theta]^2$. Неравенство Рао–Крамера устанавливает нижнюю границу для дисперсии оценки, поэтому если для какой-то несмещенной оценки $\widehat{\theta}$ в (МС.15) достигается равенство, можно утверждать, что оценка $\widehat{\theta}$ эффективна (в классе несмещенных оценок). Именно таким образом можно доказать, что выборочное среднее $\overline{X}$ есть эффективная оценка среднего значения для нормальной генеральной совокупности. Неравенство Рао–Крамера обобщается на случай смещенных оценок, а также на случай многомерного параметра $\boldsymbol{\theta}$ (число I_n при этом заменяется на соответствующую матрицу). Отметим, что условие регулярности является существенным — можно привести примеры, когда его отсутствие приводит к нарушению неравенства (МС.15).

В этом приложении мы кратко опишем лишь два общих метода оценивания неизвестных параметров.

Метод максимального правдоподобия. Рассмотрим случайный вектор $\boldsymbol{x} = (X_1, \ldots, X_n)$, имеющий плотность распределения $p(\boldsymbol{x}; \boldsymbol{\theta})$, которая зависит от неизвестного параметра $\boldsymbol{\theta}$. *Функцией правдоподобия* (*likelihood function*) называется случайная величина $L = L(\boldsymbol{x}; \boldsymbol{\theta}) = p(\boldsymbol{x}; \boldsymbol{\theta})$. Наряду с L рассматривают также *логарифмическую функцию правдоподобия* $l = \ln L$. *Оценкой максимального правдоподобия* называется величина $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_{\mathrm{ML}}$, максимизирующая (при каждом фиксированном значении $\boldsymbol{x}$) функцию правдоподобия L (или, что эквивалентно, логарифмическую функцию правдоподобия l), т.е. такая функция $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}(\boldsymbol{x})$, что $L(\boldsymbol{x}; \widehat{\boldsymbol{\theta}}(\boldsymbol{x})) = \max_{\boldsymbol{\theta}} L(\boldsymbol{x}; \boldsymbol{\theta})$. Индекс ML соответствует английскому термину *Maximum Likelihood estimation*. В регулярном случае необходимым условием максимума является уравнение $\partial L(\boldsymbol{x}; \boldsymbol{\theta})/\partial \boldsymbol{\theta} = 0$, которое называется *уравнением правдоподобия*. Легко видеть, что если компоненты $X_1, \ldots, X_n$ вектора $\boldsymbol{x}$ независимы и одинаково распределены с плотностью $p(x; \boldsymbol{\theta})$, $x \in R^1$, то функция правдоподобия есть произведение функций правдоподобия каждой компоненты (соответственно логарифмическая функция правдоподобия распадается в сумму индивидуальных логарифмических функций).

Для широкого класса задач оценки максимального правдоподобия являются состоятельными и асимптотически эффективными. В то же время они могут быть смещенными. Например, с помощью непосредственных вычислений можно показать, что для нормальной генеральной совокупности оценки максимального правдоподобия среднего значения и дисперсии есть соответственно $\widehat{m}_{\mathrm{ML}} = \overline{X}$ и $\widehat{\sigma}^2_{\mathrm{ML}} = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X})^2 = \frac{n-1}{n}s^2$, и $\mathrm{E}(\widehat{\sigma}^2_{\mathrm{ML}}) = \frac{n-1}{n}\sigma^2$. Недостатком метода является необходимость знать распределение вектора $\boldsymbol{x}$.

Метод моментов. Пусть $X_1, \ldots, X_n$ — случайная выборка из генеральной совокупности X, распределение которой зависит от некоторого неизвестного параметра $\boldsymbol{\theta}$:

$$F(x) = F(x; \boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \Theta \subseteq R^m.$$

Предположим, что заданы функции $g_1(x), \ldots, g_m(x)$ такие, что

а) в явном виде вычисляются математические ожидания

$$f_k(\boldsymbol{\theta}) = f_k(\theta_1, \ldots, \theta_m) = \mathrm{E}g_k(X), \quad k = 1, \ldots, m; \qquad \text{(МС.16)}$$

б) система (МС.16), рассматриваемая как система уравнений относительно $\boldsymbol{\theta}$, имеет единственное решение.

Для получения оценки параметра $\boldsymbol{\theta}$ в правой части системы (МС.16) математические ожидания заменяют их выборочными аналогами $\overline{g_k(X)} = \frac{1}{n}\sum_{i=1}^n g_k(X_i)$, $k = 1, \ldots, m$, т.е. строят систему уравнений (относительно $\boldsymbol{\theta}$)

$$f_k(\boldsymbol{\theta}) = f_k(\theta_1, \ldots, \theta_m) = \overline{g_k(X)}, \quad k = 1, \ldots, m, \qquad \text{(МС.17)}$$

решение которой и дает оценку параметра $\boldsymbol{\theta}$ методом моментов. Функции $g_1(x), \ldots, g_m(x)$, участвующие в оценивании, носят название *моментных функций*. Термин «момент» связан с тем обстоятельством, что часто используют функции $g_k(x) = x^k$, иными словами, правыми частями системы (МС.16) являются моменты случайной величины X. Оценки, полученные методом моментов, в широком числе случаев состоятельны, но, как правило, менее эффективны, чем оценки максимального правдоподобия.

Доверительные интервалы. Помимо получения точечной оценки неизвестного параметра часто ставится задача построения такой области, в которую параметр попадает с заданной вероятностью или с заданным *уровнем доверия*. Иными словами, требуется для случайной выборки $\boldsymbol{x} = (X_1, \ldots, X_n)$ и числа $0 < \alpha < 1$ найти такое множество $D_n = D_n(X_1, \ldots, X_n) \subseteq \Theta$, что $\mathrm{P}(\boldsymbol{\theta} \in D_n(X_1, \ldots, X_n)) = 1 - \alpha$. Это множество D_n называется *доверительным множеством с уровнем доверия* $1 - \alpha$ (или $100(1 - \alpha)$%-ным доверительным множеством). Для одномерного параметра θ доверительное множество обычно называют *доверительным интервалом*. Часто этот термин распространяют и на многомерный случай. Для построения доверительных множеств используется следующий общий подход. Предположим, что существует такое преобразование $h_n(x_1, \ldots, x_n; \boldsymbol{\theta})$, что распределение случайной величины $h_n(X_1, \ldots, X_n; \boldsymbol{\theta})$ уже не зависит от $\boldsymbol{\theta}$ и может быть эффективно найдено (например, табулировано).

Тогда для заданного $0 < \alpha < 1$ можно найти интервал I_n такой, что $\mathrm{P}(h_n(X_1,\ldots,X_n;\boldsymbol{\theta}) \in I_n) = 1-\alpha$. Если при этом включение $h_n(X_1,\ldots,X_n;\boldsymbol{\theta}) \in I_n$ можно «разрешить» относительно $\boldsymbol{\theta}$, т.е. найти такое множество $D_n = D_n(X_1,\ldots,X_n) \subseteq \Theta$, что $h_n(X_1,\ldots,X_n;\boldsymbol{\theta}) \in I_n \Leftrightarrow \boldsymbol{\theta} \in D_n(X_1,\ldots,X_n)$, то из самой конструкции следует, что D_n будет доверительным множеством с уровнем доверия $1-\alpha$. Подчеркнем, что доверительное множество не единственно.

Рассмотрим два примера построения доверительных интервалов.

Двусторонний доверительный интервал для среднего значения нормальной генеральной совокупности. Пусть $X_1,\ldots,X_n$ — случайная выборка из нормальной генеральной совокупности с параметрами (m,σ^2). В силу свойства N12) величина

$$\frac{(\overline{X}-m)\sqrt{n}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n (X_i-\overline{X})^2}} = \frac{(\overline{X}-m)\sqrt{n}}{s_X}$$

имеет распределение Стьюдента с $n-1$ степенями свободы (преобразование h_n). Для заданного $0<\alpha<1$ найдем (например, по таблицам) $100(\alpha/2)\%$-ную точку $t_{\alpha/2}$. Тогда, в силу симметричности распределения Стьюдента,

$$\mathrm{P}\left(\left|\frac{\sqrt{n}(\overline{X}-m)}{s_X}\right| \leqslant t_{\alpha/2}\right) = 1-\alpha.$$

Разрешая относительно m, получаем для этого параметра доверительный интервал $\left(\overline{X} - \frac{s_X}{\sqrt{n}}t_{\alpha/2}, \overline{X} + \frac{s_X}{\sqrt{n}}t_{\alpha/2}\right)$.

Двусторонний доверительный интервал для дисперсии нормальной генеральной совокупности. В силу свойства N11) случайная величина

$$\frac{1}{\sigma^2}\sum_{i=1}^n (X_i-\overline{X})^2 = \frac{n-1}{\sigma^2}s_X^2$$

имеет распределение $\chi^2(n-1)$ (преобразование h_n). Поэтому для заданного $0<\alpha<1$, взяв процентные точки $\chi^2_{1-\alpha/2}$, $\chi^2_{\alpha/2}$, имеем

$\mathrm{P}(\chi^2_{1-\alpha/2} \leqslant \frac{n-1}{\sigma^2}s_X^2 \leqslant \chi^2_{\alpha/2}) = 1-\alpha$. Разрешая относительно σ^2, получаем для этого параметра доверительный интервал $\left(\frac{(n-1)s_X^2}{\chi^2_{\alpha/2}}, \frac{(n-1)s_X^2}{\chi^2_{1-\alpha/2}}\right)$.

8. Проверка гипотез

Проверка гипотез и построение на их основе статистических выводов является одной из центральных задач математической и прикладной статистики. В рамках параметрического подхода общая схема проверки гипотезы может быть описана так. Пусть $X_1,\ldots,X_n$ — случайная выборка из некоторой генеральной совокупности с функцией распределения $F(x) = F(x;\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta \subseteq R^m$. Относительно параметра $\boldsymbol{\theta}$ выдвигаются две гипотезы, а именно, H_0: $\boldsymbol{\theta} \in Z_0$ и H_1: $\boldsymbol{\theta} \in Z_1$, где $Z_0 \subset \Theta$, $Z_1 \subset \Theta$ — некоторые заданные множества. Гипотезу H_0 называют основной или *нулевой*, а гипотезу H_1 — *альтернативной*. Если множество Z состоит из одной точки ($Z = \{\boldsymbol{\theta}_0\}$), то соответствующая гипотеза называется простой, в противном случае она называется сложной. Если альтернативная гипотеза явно не указана, то это означает, что $Z_1 = \Theta \backslash Z_0$.

Статистическим тестом или просто *тестом* называется любая процедура, основанная на наблюдениях $X_1,\ldots,X_n$, результатом которой является одно из двух возможных решений:

1) не отвергать (принять) нулевую гипотезу H_0;
2) отвергнуть нулевую гипотезу H_0 в пользу альтернативной гипотезы H_1.

Поскольку тест использует случайную выборку $X_1,\ldots,X_n$, то, естественно, могут возникать ошибочные решения. В связи с этим возникают две ошибки теста:

ошибка первого рода: нулевая гипотеза отвергается, когда она верна;

ошибка второго рода: нулевая гипотеза принимается, когда верна альтернативная гипотеза.

Вероятности ошибок первого и второго рода можно обозначить $\alpha = \mathrm{P}(\mathrm{H}_1 \mid \mathrm{H}_0)$ и $\beta = \mathrm{P}(\mathrm{H}_0 \mid \mathrm{H}_1)$ соответственно. Величину α называют *значимостью* теста, а величину $1-\beta$ — его *мощностью*.

Естественно при построении теста стремиться уменьшить эти ошибки, однако нетрудно понять, что невозможно минимизировать их одновременно. Поэтому обычно поступают следующим образом: фиксируют значимость теста и стараются найти такой тест, у которого мощность максимальна (именно здесь в явном виде проявляется несимметричность гипотез, деление их на основную и альтернативную).

На практике для построения тестов часто используют следующий подход. Предположим, что можно найти такую статистику $t_n = t_n(X_1, \ldots, X_n)$, что если гипотеза H_0 верна, то распределение случайной величины t_n известно (например, табулировано). Тогда для заданного значения α ошибки первого рода можно найти такую область K_α, что $\mathrm{P}(t_n \in K_\alpha) = 1-\alpha$ (подчеркнем, что вероятность вычисляется в предположении, что верна нулевая гипотеза). Тогда тест определяется следующим образом:

1) на основании наблюдений $X_1, \ldots, X_n$ вычисляется значение статистики t_n;
2) для заданного уровня значимости α находится область K_α;
3) если $t_n \in K_\alpha$, то нулевая гипотеза не отвергается (принимается); если $t_n \notin K_\alpha$, то нулевая гипотеза отвергается в пользу альтернативной.

Статистику t_n называют *критической статистикой*, а область K_α — *критической областью*. На практике часто критические статистики имеют распределения стандартное нормальное, χ^2, Стьюдента и Фишера. В этих случаях при использовании подобного рода тестов для каждого значения критической статистики, полученной в эксперименте, находится еще так называемое *P-значение*. Если статистика t_n, распределение которой при нулевой гипотезе принадлежит к одному из указанных четырех типов, приняла значение c, то соответствующим *P-значением* называется число $\mathrm{P}(|t_n| > |c|)$ — для нормального распределения и

распределения Стьюдента и число $\mathrm{P}(t_n > c)$ — для распределений χ^2 и Фишера. Таким образом, P-значение сразу позволяет судить о значимости нулевой гипотезы. Все компьютерные пакеты, как правило, приводят P-значения вычисленных статистик.

Легко установить связь между описанной выше процедурой и построением доверительных интервалов. Действительно, предположим, что для неизвестного параметра $\boldsymbol{\theta}$ построен доверительный интервал D_n с уровнем доверия $1-\alpha$, и предположим, что нулевая гипотеза является простой, т. е. H_0: $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Тогда нулевая гипотеза не отвергается, если $\boldsymbol{\theta}_0 \in D_n$, и отвергается в противном случае.

Приложение ЭП

Обзор эконометрических пакетов

Последнее десятилетие характеризовалось чрезвычайно быстрым прогрессом и конкуренцией в области эконометрического программного обеспечения. Новые методы, описанные в эконометрической литературе, через несколько лет уже включаются в пакеты и становятся общедоступными методами анализа. Поэтому обзоры эконометрических пакетов, на которые мы ссылались в первом издании нашей книги (MacKie-Mason (1992), Айвазян и др. (1989), Кузнецов, Халилеев (1991), Greene (1997)), в значительной степени устарели. Прогресс в мощности персональных компьютеров привел к тому, что ограничения на объем оперативной памяти, жесткого диска, представляются сейчас несущественными. Большинство пакетов в настоящее время имеют Windows — версию (а многие имеют версии также для Unix и Macintosh).

В данном обзоре использован также личный опыт работы авторов с экономическими и финансовыми данными с пакетами *GAUSS*, *MicroTSP*, *Microfit386*, *Econometric Views*, *STATA*, *Statgraphics*, *TSP*, а также непосредственное знакомство с пакетами *SAS*, *SPSS*, *RATS*.

1. Происхождение пакетов. Windows-версии. Графика

Большинство пакетов — такие, как *RATS, SAS, STATA, TSP, SPSS, SORITEC, SYSTAT*, были написаны до эры персональных компьютеров и ориентировались на технологию работы с большими ЭВМ. Фактически они представляли собой набор процедур на языке FORTRAN. Многие из этих пакетов и до сих пор включают части, написанные на FORTRANe. При перенесении на персональные компьютеры к набору процедур добавлялась надстройка в виде командного языка (или достаточно примитивной системы меню, за которой стоит командный язык) для управления вызовом процедур. В силу этого они уступают пакетам, специально написанным для PC (например, *GAUSS, Microfit, MicroTSP, EViews,*), по таким параметрам, как быстродействие, дизайн, интерфейс с пользователем, эффективность использования PC.

Пакеты, предназначенные для PC, написаны, как правило, на языке C, который позволяет запрограммировать более гибкий и эффективный интерфейс с пользователем. Однако пакеты, изначально написанные на FORTRANe, такие, как *TSP, STATA, SPSS, SYSTAT, SAS*, создавались и развивались в течение многих лет большими коллективами разработчиков и включают в себя огромное количество самых разнообразных методов.

Графические возможности представляются не очень существенным фактором при выборе пакета. Достаточно иметь графические средства, необходимые для анализа и понимания данных, моделей (например, графики остатков регрессии, автокорреляционная функция остатков, гистограмма остатков и т. п.), а их предоставляют практически все статистические пакеты. Больше внимания, на наш взгляд, следует уделить легкости получения необходимых графиков (например, сразу из меню пострегрессионного анализа) и интерактивным возможностям графического интерфейса (графический курсор, графический редактор и т. д.). Если же для отчета необходима презентационная графика, то лучше обратиться к специализированным графическим пакетам или к мощным табличным процессорам, например к *Excel*.

2. О некоторых пакетах

SAS (Windows)

Самый объемный из существующих пакетов. Охватывает практически все области прикладной статистики и анализа данных. Имеет лучшие средства для работы с данными. В настоящее время является мощным средством для сбора, доставки, хранения и отображения информации. Однако за это приходится платить. PC-версия работает крайне медленно. Это один из наиболее распространенных пакетов, однако большинство пользователей не являются эконометристами, а представляют другие области социальных и естественных наук. Документация не ориентирована на экономистов. По набору современных эконометрических методов уступает специализированным эконометрическим пакетам. *SAS Institute* издает журнал для пользователей *SAS*. Впрочем, аналогичные периодические журналы для своих пользователей издают также *SPSS, GAUSS, STATA*. Подробную информацию о *SAS* можно найти по адресу `http://www.sas.com/`.

GAUSS

Фактически является языком программирования с очень развитыми возможностями работы с матрицами. Непосредственно в языке отсутствуют эконометрические методы, но они доступны в виде обширных библиотек программ, написанных на *GAUSS* (см., например, `http://www.aptech.com/links.html`). Многие эконометристы предпочитают составлять программы для нестандартных методов на языке *GAUSS*. Отрицательной чертой является неразвитая диагностика ошибок. Подробную информацию о *GAUSS* можно найти по адресу `http://www.aptech.com/`.

STATA

Пакет в значительной степени ориентирован на экономистов. Обладает широкими возможностями работы с данными. Удобен для работы с панельными данными. Легко осваивается. Подробную информацию о *STATA* можно найти по адресу `http://www.stata.com/`.

TSP

Пакет предоставляет широкие возможности для анализа временных рядов. Хороший раздел систем одновременных уравнений. Содержит полный раздел нелинейных моделей, обобщенный метод моментов GMM. Производит аналитическое дифференцирование. Хорошо документирован. Подробную информацию о *TSP* можно найти по адресу `http://www.tsp.com/`.

SPSS

Отличается удобной структурой Windows-интерфейса. Содержит большой набор разнообразных статистических методов. Имеет дилеров в Москве (`http://www.spss.ru`). Проводятся курсы для пользователей в Москве. Подробную информацию о *SPSS* можно найти по адресу `http://www.spss.com/`.

Microfit386

Авторы — Hashem Pesaran, Bahram Pesaran (специалист по программированию и профессор статистики). Программа содержит ряд современных эконометрических методов. Некоторые из них являются особенностью программы. Легко осваивается.

Econometric Views

Является Windows-версией пакета *MicroTSP*, значительно превосходя DOS-версию по набору методов. Благодаря стройной и логичной идеологии построения Windows-интерфейса очень прост в освоении. Содержит развитую подсказку (help), являющуюся, по существу, справочником по эконометрическим методам. Подробную информацию о *EViews* можно найти по адресу `http://www.eviews.com/`.

3. Опыт практической работы

Как показывает опыт, одного пакета часто бывает явно недостаточно для работы. Выбор здесь во многом зависит от конкретных условий, набора задач, квалификации, опыта и вкусов пользователя.

В своей работе авторы обычно используют пакеты *EViews, TSP, STATA, Microfit.* Каждый из этих пакетов содержит огромное количество эконометрических методов. *EViews* очень легок в освоении, поэтому именно его авторы используют в программе «*outreach — эконометрика*», проводимой Российской экономической школой и Министерством образования РФ, целью которой является помощь преподавателям вузов России и СНГ в подготовке современного курса эконометрики. (Более подробно про эту программу семинаров можно узнать из `http://www.nes.ru/` или по адресу `outreach@nes.ru`).

Для начинающих изучение эконометрики студентов бакалавриата вполне подходят простые пакеты, работающие в среде ДОС, такие, как *MicroTSP, Microfit, Minitab.*

Приложение СТ

Краткий англо-русский словарь терминов

adaptive expectation — адаптивное ожидание

adjusted R^2 — скорректированный R^2

augmented test — расширенный, пополненный тест

autocorrelation function (ACF) — автокорреляционная функция

autoregressive conditional heteroscedasticity (ARCH) model — авторегрессионная условно гетероскедастичная модель

autoregressive (AR) model — авторегрессионная модель

autoregressive integrated moving average (ARIMA) model — интегрированная модель авторегрессии и скользящего среднего

autoregressive moving average (ARMA) model — модель авторегрессии и скользящего среднего

best linear unbiased estimator (BLUE) — наилучшая (с минимальной дисперсией) оценка в классе несмещенных линейных оценок

binary variable — бинарная переменная, принимающая значения 0 или 1

Box–Jenkins model = ARIMA — модель Бокса–Дженкинса = интегрированная модель авторегрессии и скользящего среднего

censored model — модель с цензурированными наблюдениями, т. е. модель, в которой значения некоторых переменных ограничиваются, как правило, некоторыми пороговыми значениями

central limit theorem (CLT) — центральная предельная теорема

classical normal regression (CNR) — модель классической регрессии, в которой ошибки имеют совместное нормальное распределение

classical regression (CR) — модель регрессии, в которой ошибки независимы, одинаково распределены, имеют нулевое среднее значение и постоянную дисперсию

coefficient of determination (R-squared) — коэффициент детерминации R^2

cointegrated processes — коинтегрированные процессы, нестационарные процессы, линейная комбинация которых стационарна

conditional distribution — условное распределение

conditional expectation — условное среднее, условное математическое ожидание

confidence interval — доверительный интервал

consistent estimator — состоятельная оценка

convergence in distribution (law) — сходимость по распределению

convergence in probability — сходимость по вероятности

correlation — корреляция

correllogram — график (выборочной) автокорреляционной функции

correlation coefficient — коэффициент корреляции

covariance — ковариация

cross-section data — данные, не имеющие временной природы, порядок их расположения несуществен

curve fitting — подгонка кривой

density function — плотность распределения

dependent (endogenous) variable — зависимая (эндогенная) переменная

distributed lags model — модель распределенных лагов

distribution — распределение

distribution function — функция распределения

dummy variable — фиктивная независимая переменная, принимающая, как правило, два значения — 0 или 1

dummy trap — ситуация, когда сумма нескольких фиктивных переменных, включенных в регрессию, равна константе, также включенной в регрессию

duration model — модель «времени жизни», модель продолжительности какого-либо процесса

efficient estimator — эффективная оценка, оптимальный (в некотором классе) метод оценивания

efficient frontier — граница эффективных портфелей

endogenous (dependent) variable — эндогенная (зависимая) переменная

error correction model — модель коррекции ошибок

estimate — величина оценки при заданных выборочных значениях

estimator — метод оценивания, функция выборочных значений

exogenous (independent) variable — экзогенная (независимая) переменная, регрессор

expectation (mean) — среднее значение, математическое ожидание

explanatory variables — объясняющие переменные, регрессоры, независимые переменные

explained (unexplained) variance — объясняемая (необъясняемая) дисперсия

exponential smoothing — экспоненциальное сглаживание

fitted value — прогнозное значение

first-order condition (FOC) — необходимые условия экстремума

generalized autoregressive conditional heteroscedasticity (GARCH) model — обобщенная авторегрессионная условно гетероскедастичная модель

generalized least squares (GLS) estimation — обобщенный метод наименьших квадратов

goodness-of-fit — качество приближения данных моделью

hazard rate — интенсивность отказов, коэффициент смертности

heteroscedasticity — гетероскедастичность

homoscedasticity — гомоскедастичность

idempotent matrix — идемпотентная матрица

independence of irrelevant alternatives — независимость от посторонних альтернатив

independent (exogenous) variable — независимая (экзогенная) переменная

indirect least squares — косвенный метод наименьших квадратов

information matrix — информационная матрица

instrumental variable (IV) — инструментальная переменная

instrumental variables estimator (IV-estimator) — метод оценивания с помощью инструментальных переменных

intercept — свободный член в уравнении регрессии

joint distribution — совместное распределение

lag operator — оператор сдвига

lagged variable — лагированная переменная, переменная с запаздыванием

latent variable — скрытая, ненаблюдаемая переменная

law of large numbers (LLN) — закон больших чисел

likelihood function — функция правдоподобия

linear probability model — линейная модель вероятности, линейная регрессионная модель для бинарной зависимой переменной

linear regression model — линейная регрессионная модель

logit model — *logit*-модель, нелинейная модель для бинарной зависимой переменной с использованием функции логистического распределения

loglikelihood function — логарифмическая функция правдоподобия

loss function — функция потерь, измеряющая степень отклонения данных от «желаемых» значений

marginal distribution — маргинальное распределение, т. е. распределение одной или нескольких компонент случайного вектора

maximum likelihood (ML) — метод максимального правдоподобия

maximum likelihood estimate — оценка максимального правдоподобия

maximum likelihood estimator — оценивание с помощью метода максимального правдоподобия

mean absolute deviation — среднее абсолютное отклонение

mean absolute percentage error — среднее относительное отклонение

mean squared error — среднеквадратичная ошибка

model specification — спецификация модели

moving average — скользящее среднее

moving average (MA) model — модель скользящего среднего

multicollinearity — мультиколлинеарность

multinomial logit model — *logit*-модель множественного выбора

multiple regression model — модель множественной регрессии

normal (Gaussian) distribution — нормальное (гауссовское) распределение

nuisance parameter — вспомогательный (излишний) параметр

OLS-estimator, OLS-estimate — оценивание с помощью метода наименьших квадратов и значения этих оценок

omitted variables — пропущенные переменные (независимые переменные, не включенные в модель)
ordinary least squares (OLS) method — метод наименьших квадратов, МНК
outliers — выбросы (данные, имеющие большие отклонения от большинства других)

panel data — панельные данные, данные имеющие как пространственную, так и временную структуру (например, данные по показателям нескольких фирм за несколько лет)
partial adjustment model — модель частичного приспособления
partial autocorrelation function (PACF) — частная автокорреляционная функция
partial correlation coefficient — коэффициент частной корреляции
pretest estimator — оценка коэффициента регрессии, полученная после процедуры предварительного отбора модели
probit model — *probit*-модель, нелинейная модель для бинарной зависимой переменной с использованием функции стандартного нормального распределения

qualitative variable — качественная, номинальная переменная

random walk — случайное блуждание
ranking variable — ранговая, порядковая, ординальная переменная
reduced form of the model — приведенная форма модели
residuals — остатки
restricted regression — регрессия с ограничениями, модель регрессии с ограничениями на параметры
return — доходность актива

sample — выборка
sample mean (variance, covariance, moment etc.) — выборочное среднее (дисперсия, ковариация, момент и т. д.)
seemingly unrelated regression (SUR) — система внешне не связанных между собой уравнений

sensitivity analysis — анализ чувствительности
serial correlation — (для временных рядов) наличие корреляции между ошибками, относящимися к разным моментам времени
significance level — уровень значимости
simultaneous equations — одновременные уравнения
slope — коэффициент при независимой переменной в уравнении регрессии
spurious regression — мнимая, ложная регрессия
standard deviation — стандартное отклонение (корень из дисперсии)
stationary time series — стационарный временной ряд (статистические свойства которого не зависят от времени)
stochastic discount factor — стохастический дисконтирующий множитель
strictly stationary process — строго стационарный процесс, стационарный в узком смысле процесс

time-series data — временные ряды, параметры системы, наблюдаемые в последовательные моменты времени
time series, integrated of order p — интегрированный временной ряд порядка p, нестационарный временной ряд, p-я последовательная разность которого является стационарным процессом
truncated model — модель с усеченными наблюдениями, т. е. модель, из которой исключены некоторые наблюдения
two-stage least squares (2SLS) — двухшаговый метод наименьших квадратов

unbiased estimator — несмещенное оценивание (оценка)
under-, overestimation — оценка параметра, имеющая отрицательное (соответственно положительное) смещение
unit root hypothesis — гипотеза о наличии единичного корня
unrestricted regression — регрессия без ограничений, модель регрессии без ограничений на параметры

variance — дисперсия

variance (covariance) matrix — матрица ковариаций случайного вектора

weakly stationary process — слабо стационарный процесс, стационарный в широком смысле процесс

weighted least squares — метод взвешенных наименьших квадратов

white noise — белый шум, процесс с независимыми одинаково распределенными значениями с нулевыми средними

Приложение ТА

Таблицы

Таблица 1. Функция стандартного нормального распределения $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}}\, dt$

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767

Таблица 1. Функция стандартного нормального распределения $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}}\, dt$ (окончание)

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9980	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

Пример. Пусть z —распределенная по стандартному нормальному закону $N(0,1)$. $\mathrm{P}(z < 0.31) = \Phi(0.31) = 0.6217$ (см. четвертую строку, второй столбец в первой части таблицы).

Таблица 2. Процентные точки распределения Стьюдента $t_\alpha(n)$

n \ α	0.40	0.30	0.25	0.20	0.10	0.05	0.025	0.01	0.005
1	0.325	0.727	1.000	1.376	3.078	6.314	12.706	31.821	63.657
2	0.289	0.617	0.816	1.061	1.886	2.920	4.303	6.965	9.925
3	0.277	0.584	0.765	0.978	1.638	2.353	3.182	4.541	5.841
4	0.271	0.569	0.741	0.941	1.533	2.132	2.776	3.747	4.604
5	0.267	0.559	0.727	0.920	1.476	2.015	2.571	3.365	4.032
6	0.265	0.553	0.718	0.906	1.440	1.943	2.447	3.143	3.707
7	0.263	0.549	0.711	0.896	1.415	1.895	2.365	2.998	3.499
8	0.262	0.546	0.706	0.889	1.397	1.860	2.306	2.896	3.355
9	0.261	0.543	0.703	0.883	1.383	1.833	2.262	2.821	3.250
10	0.260	0.542	0.700	0.879	1.372	1.812	2.228	2.764	3.169

n — число степеней свободы, $100\alpha\%$ — уровень процентной точки.

Таблица 2. Процентные точки распределения Стьюдента $t_\alpha(n)$ (окончание)

n \ α	0.40	0.30	0.25	0.20	0.10	0.05	0.025	0.01	0.005
11	0.260	0.540	0.697	0.876	1.363	1.796	2.201	2.718	3.106
12	0.259	0.539	0.695	0.873	1.356	1.782	2.179	2.681	3.055
13	0.259	0.538	0.694	0.870	1.350	1.771	2.160	2.650	3.012
14	0.258	0.537	0.692	0.868	1.345	1.761	2.145	2.624	2.977
15	0.258	0.536	0.691	0.866	1.341	1.753	2.131	2.602	2.947
16	0.258	0.535	0.690	0.865	1.337	1.746	2.120	2.583	2.921
17	0.257	0.534	0.689	0.863	1.333	1.740	2.110	2.567	2.898
18	0.257	0.534	0.688	0.862	1.330	1.734	2.101	2.552	2.878
19	0.257	0.533	0.688	0.861	1.328	1.729	2.093	2.539	2.861
20	0.257	0.533	0.687	0.860	1.325	1.725	2.086	2.528	2.845
21	0.257	0.532	0.686	0.859	1.323	1.721	2.080	2.518	2.831
22	0.256	0.532	0.686	0.858	1.321	1.717	2.074	2.508	2.819
23	0.256	0.532	0.685	0.858	1.319	1.714	2.069	2.500	2.807
24	0.256	0.531	0.685	0.857	1.318	1.711	2.064	2.492	2.797
25	0.256	0.531	0.684	0.856	1.316	1.708	2.060	2.485	2.787
26	0.256	0.531	0.684	0.856	1.315	1.706	2.056	2.479	2.779
27	0.256	0.531	0.684	0.855	1.314	1.703	2.052	2.473	2.771
28	0.256	0.530	0.683	0.855	1.313	1.701	2.048	2.467	2.763
29	0.256	0.530	0.683	0.854	1.311	1.699	2.045	2.462	2.756
30	0.256	0.530	0.683	0.854	1.310	1.697	2.042	2.457	2.750
40	0.255	0.529	0.681	0.851	1.303	1.684	2.021	2.423	2.704
60	0.254	0.527	0.679	0.848	1.296	1.671	2.000	2.390	2.660
100	0.254	0.526	0.677	0.845	1.290	1.660	1.984	2.364	2.626
200	0.254	0.525	0.676	0.843	1.286	1.652	1.972	2.345	2.601
∞	0.253	0.524	0.675	0.842	1.282	1.645	1.960	2.326	2.576

n — число степеней свободы, $100\alpha\%$ — уровень процентной точки.

Пример. Пусть t — случайная величина, распределенная по закону Стьюдента с 5 степенями свободы. $t_{0.025}(5) = 2.571$, т.е. $\mathrm{P}(t > 2.571) = 0.025$ (см. пятую строку, третий справа столбец в первой части таблицы).

Таблица 3. Процентные точки распределения $\chi^2(n)$

n \ α	0.995	0.990	0.975	0.950	0.900	0.100	0.050	0.025	0.010	0.005
1	0.000039	0.00016	0.00098	0.0039	0.0158	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.1026	0.2107	4.61	5.99	7.38	9.21	10.60
3	0.0717	0.115	0.216	0.352	0.584	6.25	7.81	9.35	11.34	12.84
4	0.207	0.297	0.484	0.711	1.064	7.78	9.49	11.14	13.28	14.86
5	0.412	0.554	0.831	1.15	1.61	9.24	11.07	12.83	15.09	16.75
6	0.676	0.872	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	0.989	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
24	9.89	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69	66.77
60	35.53	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38	91.95
80	51.17	53.54	57.15	60.39	64.28	96.58	101.88	106.63	112.33	116.32
100	67.33	70.06	74.22	77.93	82.36	118.50	124.34	129.56	135.81	140.17
120	83.85	86.92	91.58	95.70	100.62	140.23	146.57	152.21	158.95	163.64

n — число степеней свободы, $100\alpha\%$ — уровень процентной точки.

Пример. Пусть χ^2 — случайная величина, распределенная по закону χ^2 с 5 степенями свободы. $\chi^2_{0.05}(5) = 11.07$, т. е. $P(\chi^2 > 11.07) = 0.05$ (см. пятую строку, седьмой столбец).

Таблица 4. 5%-ные точки распределения Фишера $F(k_1, k_2)$

k_2 \ k_1	1	2	3	4	5	6	7	8	9	10
1	161	200	216	225	230	234	237	239	241	242
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83

k_1 — число степеней свободы числителя, k_2 — число степеней свободы знаменателя.

Таблица 4. 5%-ные точки распределения Фишера $F(k_1, k_2)$ (окончание)

k_2 \ k_1	12	15	20	24	30	40	60	120	∞
1	244	246	248	249	250	251	252	253	254
2	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5
3	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
6	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	2.69	2.62	2.54	2.51	2.47	2.43	2.38	2.34	2.30
13	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.92
19	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
30	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

k_1 — число степеней свободы числителя, k_2 — число степеней свободы знаменателя.

Пример. Пусть F — случайная величина, распределенная по закону Фишера $F(3,5)$. $F_{0.05}(3,5) = 5.41$, т.е. $P(F > 5.41) = 0.05$ (см. пятую строку, третий столбец).

Литература

Айвазян С. А., Енюков И. С., Мешалкин Л. Д. (1983). *Прикладная статистика. Основы моделирования и первичная обработка данных.* М., Финансы и статистика.

Айвазян С. А., Енюков И. С., Мешалкин Л. Д. (1985). *Прикладная статистика. Исследование зависимостей.* М., Финансы и статистика.

Айвазян С. А., Бухштабер В. М., Енюков С. А., Мешалкин Л. Д. (1989). *Прикладная статистика. Классификация и снижение размерности.* М., Финансы и статистика.

Айвазян С. А., Мхитарян В. С. (1998). *Прикладная статистика и основы эконометрики.* М., ЮНИТИ.

Айвазян С. А. (2002). *Прикладная статистика. Основы эконометрики,* том 2. М., ЮНИТИ.

Бородич С. А. (2001). *Эконометрика.* Минск, Новое знание.

Вентцель Е. С. (1964). *Теория вероятностей,* 3-е изд. М., Наука.

Гельфанд И. М. (1971). *Лекции по линейной алгебре,* 4-е изд. М., Наука.

Доугерти К. (1997). *Введение в эконометрику* (перевод с издания 1992 г.). М., ИНФРА-М.

Джонстон Дж. (1980). *Эконометрические методы.* М., Статистика.

Елисеева И. И. (2001a). *Эконометрика.* М., Финансы и статистика.

Елисеева И. И. (2001b). *Практикум по эконометрике.* М., Финансы и статистика.

Замков О. О. (2001). *Эконометрические методы в макроэкономическом анализе: Курс лекций.* М., ГУ-ВШЭ.

Ильин В. А., Позняк Э. Г. (1984). *Линейная алгебра*, 3-е изд. М., Наука.

Каргин В., Онацкий А. (1996). *Рынок квартир в Москве.* Рукопись.

Карминский А. М., Пересецкий А. А., ван Суст А. Г. О. (2003). Моделирование рейтингов надежности российских банков. *Экономика и математические методы*, т. 39, № 4.

Крамер Г. (1975). *Математические методы статистики*, 2-е изд. (перевод с издания 1946 г.). М., Мир.

Кремер Н. Ш., Путко Б. А. (2002). *Эконометрика.* М., ЮНИТИ.

Кузнецов С. Е., Халилеев А. А. (1991). *Обзор специализированных статистических пакетов по анализу временных рядов.* М., Статдиалог.

Магнус Я.Р., Нейдеккер Х. (2002). *Матричное дифференциальное исчисление с приложениями к статистике и эконометрике.* М., Физматлит.

Малюгин В. И. (2003). *Рынок ценных бумаг: количественные методы анализа.* М., Дело.

Рао М. М. (1968). *Линейные статистические методы и их применения.* М., Наука.

Ширяев А. Н. (1998). *Основы стохастической финансовой математики.* М., Фазис.

Яковлев А. А., Бессонов В. А. (1995а). Валютные фьючерсы: пределы развития. *КоммерсантЪ*, т. 7, стр. 70–73.

Яковлев А. А., Бессонов В. А. (1995б). *Очерк развития российского рынка валютных фьючерсов.* М., Аргус.

Akaike H. (1973). Information theory and an extension of the maximum likelihood principle, in B. Petrov and F. Csake, eds. *2nd International Symposium on Informational Theory.* Budapest, Akademiai Kiado.

Andersen E. (1970). Asymptotic properties of conditional maximum likelihood estimation. *Journal of the Royal Statistical Society, Series B*, v. 32, pp. 283–301.

Anderson T.W., Hsiao C. (1981). Estimation of dynamic models with error components. *Journal of the American Statistical Association*, v. 76, pp. 598–606.

Arellano M., Bond S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, v. 58, pp. 277–294.

Bachelier L. (1900). Théorie de la spéculation. *Annales scientifiques de l'École Normale Supérieur*, v. 17, pp. 21–86.

Bancroft, T. A. (1944). On biases in estimation due to the use of preliminary tests of significance. *Annals of Mathematical Statistics* v. 15, pp. 190–204.

Ben-Porah Y. (1973). Labor force participation rates and the supply of labor. *Journal of Political Economy*, v. 81, pp. 697–704.

Bollerslev T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, v. 31, pp. 307–327.

Box G.E.P. and Pierce D.A. (1970). Distribution of residual autocorrelations in autoregressive–integrated moving average time series models. *Journal of the American Statistical Association*, v. 65, pp. 1509–1526.

Box G. and Jenkins G. (1976). *Time Series Analysis: Forecasting and Control.* San Francisco, Holden-Day.

Breusch T. S. (1979). Conflict among criteria for testing hypotheses: extensions and comments. *Econometrica*, v. 47, pp. 203–207.

Breusch T. S. and Pagan A. R. (1979). A Simple test for heteroscedasticity and random coefficient variation. *Econometrica*, v. 47, pp. 1287–1294.

Breusch T. S. and Pagan A. R. (1980). The LM test and its applications to model specification in econometrics. *Review of Economic Studies*, v. 47, pp. 239–254.

Campbell J. Y., Lo A. W. and MacKinlay A. G. (1997). *The Econometrics of Financial Markets.* Princeton University Press.

Chamberlain G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies*, v. 47, pp. 225–238.

Cohrane J. H. (2001). *Asset Pricing.* Princeton University Press, Princeton.

Courtault J.-M., Kabanov Yu. M., Bru B., Crepel P., Lebon I. and Le Mashand A. (2000). To the centenary of Théorie de la Spéculation. *Mathematical Finance.* v. 10, n. 3, pp. 341-353.

Danilov D. (2002). Estimation of the mean of a univariate normal distribution when the variance is not known, submitted for publication.

Danilov D. and Magnus J. R. (2002). On the harm that pretesting can cause, submitted for publication.

Danilov D. and Magnus J. R. (2003). Forecasting accuracy after pretesting with an application to the stock market, submitted for publication.

Davidson R. and MacKinnon J. G. (1981). Several tests for model specification in the presence of alternative hypotheses. *Econometrica*, v. 49, pp. 781–793.

Davidson R. and MacKinnon J. G. (1993). *Estimation and Inference in Econometrics.* Oxford University Press, New York.

DeRoon F. A. and Nijman T. E. (2001). Testing for mean-variance spanning: a survey. *Journal of Empirical Finance*, v. 8, pp. 111-155.

Durbin J. and Watson G. S. (1951). Testing for serial correlation in least-squares regression. *Biometrika*, v. 38, pp. 159–177.

Engle R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of variance of United Kingdom inflation. *Econometrica*, v. 50, pp. 987–1008.

Engle R. F. and Granger C. W. J. (1987). Cointegration and error correction: representation, estimating and testing. *Econometrica*, v. 55, pp. 251–276.

Evans G. B. A. and Savin N. E. (1982). Conflict among the criteria revisited; the W, LR and LM tests. *Econometrica*, v. 50, pp. 737–748.

Fama E. F. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, v. 25, pp. 383-417.

Fama E. F. and French K. R. (1988). Permanent and temporary components of stock prices. *Journal of Political Economy*, v. 96, pp. 246-273.

Fuller W. A. (1976). *Introduction to Statistical Time Series.* Wiley, New York, pp. 366–382.

Giles J. A. and Giles D. E. A. (1993). Pre-test estimation and testing in econometrics: recent developments. *Journal of Economic Surveys* v. 7, pp. 145–197.

Godfrey L. G. (1988). *Misspecification Tests in Econometrics.* Cambridge University Press, Cambridge, UK.

Goldberger A. (1990). *A Course in Econometrics.* Cambridge, MA: Harvard University Press.

Gourieroux C. (1997). *ARCH Models and Financial Applications.* Springer Series in Statistics.

Granger C. W. J. (1969). Investigating casual relations by econometric methods and cross-spectral methods. *Econometrica*, v. 37, pp. 424–438.

Granger C. W. J. and Newbold P. (1974). Spurious regression in econometrics. *Journal of Econometrics*, v. 2, pp. 1045–1066.

Greene W. H. (1997). *Econometric Analysis*, 3rd edition. Prentice-Hall, Upper Saddle River, New Jersey.

Gronau R. (1974). Wage Comparisons: A Selectivity Bias. *Journal of Political Economy*, v. 82, pp. 1119–1143.

Griffiths W. E., Hill R. C. and Judge G. G. (1993). *Learning and Practicing Econometrics.* Wiley.

Hamilton J. D. (1994). *Time Series Analysis.* Princeton University Press, Princeton.

Hansen L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, v. 50, pp. 1029–1054.

Hausman J. (1978). Specification tests in econometrics. *Econometrica*, v. 46, pp. 1251–1271.

Heckman J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, v. 47, pp. 153–161.

Heckman J. and Singer B. (1984). Econometric duration analysis. *Journal of Econometrics*, v. 24, pp. 63–132.

Hsiao C. (1986). *Analysis of Panel Data*. Cambridge University Press, New York.

Johnston J. and DiNardo J. (1997) *Econometric Methods*, 4th edition. McGraw-Hill, New York.

Judge G. G. and Bock M. E. (1978). *The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics*. North-Holland, Amsterdam.

Kabanov Yu. M. (2001) Arbitrage theory. In: *Handbooks in Mathematical Finance. Option Pricing: Theory and Practice.*. Cambridge University Press, pp. 3-42.

Kennedy P. (1992). *A Guide to Econometrics*. 3rd edition, M.I.T. Press, Cambridge, MA.

Kennedy P. (1998). *A Guide to Econometrics*. 4th edition, Blackwell Publishers.

Keuzenkamp H. A. and Magnus J. R. (1995). On tests and significance in econometrics. *Journal of Econometrics*, v. 67, pp. 5–24.

Kiefer N. (1988). Economic duration data and hazard function. *Journal of Economic Literature*, v. 26, pp. 646–679.

Koopmans T. (1949). Identification problems in economic model construction. *Econometrica*, v. 17, pp. 125–144.

Lancaster T. (1974). Econometric Models for the duration of unemployment. *Econometrica*, v. 47, pp. 939–956.

Lancaster T. (1985). Generalized residuals and heterogeneous duration models: With application to the Weibull model. *Journal of Econometrics*, v. 28, pp. 155–169.

Leamer E. E. (1978). *Specification Searches*. Wiley, New York.

LeRoy S. F. and Werner J. (2001). *Principles of Financial Economics*. Cambridge University Press.

Ljung G.M. and Box G.E.P. (1978). On a measure of lack of fit in time series models. *Biometrika*, v. 65, pp. 297–303.

Lo A. W. (2000). Finance: A selective survey. *Journal of American Statistical Association*, v. 95, pp. 629–635.

Luenberger D. G. (1998). *Investment Science*. Oxford University Press.

MacKie-Mason J. K. (1992). Econometric software: A user's view. *Journal of Economic Perspectives*, v. 6, n. 4, pp. 165-187.

MacKinnon J. G. (1991). Critical values for cointegration tests, Chapter 13. *Long-Run Economic Relationship*, eds. R. F. Engle and C. W. J. Granger, Oxford University Press.

MacKinnon J. G., White, H. and Davidson R. (1983). Test for model specification in the presence of alternative hypotheses: Some further results. *Journal of Econometrics*, v. 21, pp. 53–70.

Maddala G. (1987). Limited dependent variable models using panel data. *The Journal of Human Resources*, v. 22, pp. 307–338.

Magnus J. R. and Neudecker H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Revised Edition. Wiley, New York.

Magnus J. R. and Durbin J. (1999). Estimation of regression coefficients of interest when other regression coefficients are of no interest. *Econometrica*, v. 67, pp. 639–643.

Magnus J. R. (1999). The traditional pretest estimator. *Теория вероятностей и ее применения*, v. 44, pp. 401–418.

Magnus J. R. (2002). Estimation of the mean of a univariate normal distribution with known variance. *The Econometrics Journal*, v.5, pp. 225–236.

Malkiel B. (1992). Efficient market hypothesis, in Newman P., Milgate M., and Eatwell J. (eds.). *New Palgrave Dictionary of Money and Finance*. Macmillan, London.

Mankiw N. G., Romer D. and Weil D. N. (1992). A Contribution to the empirics of economic growth. *Quarterly Journal of Economics*, v. 107(2), pp. 407–437.

Mills T. M. (1999). *The Econometric Modelling of Financial Time Series*. 2nd edition. Cambridge University Press, New York.

Newey W. and West K. (1987). A Simple positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica*, v. 55, pp. 703–708.

Peresetsky A. and de Roon F. (1997). Risk premia in the ruble/dollar futures market. *The Journal of Futures Markets*, v. 17, n. 2, pp. 191–214.

Peresetsky A. A. and Ivanter A. I. (2000). Interaction of the Russian financial markets. *Economics of Planning*, v. 33, pp. 103–140.

Peresetsky A., Turmuhambetova G. and Urga G. (2001) The development of the GKO futures market in Russia. *Emerging Markets Review*, v. 2, pp. 1–16.

Phillips P. C. B. (1986) Understanding spurious regressions in Econometrics. *Journal of Econometrics*, v. 33, pp. 311–340.

Phillips P. C. B. and Ouliaris S. (1990). Asymptotic properties of residual based tests for cointegration. *Econometrica*, v. 58, pp. 165–193.

Pindyck R. S. and Rubinfeld D. L. (1991). *Econometric Models and Economic Forecasts*, 3rd edition. McGraw-Hill, New York.

Pötscher, B. M. (1991). Effects of model selection on inference. *Econometric Theory* v. 7, pp. 163–185.

Ramsey J. B. (1969). Tests for specification errors in classical least squares regression analysis. *Journal of Royal Statistical Society*, B, v. 32, pp. 350–371.

Romer D. (2001). *Advanced Macroeconomics*. 2nd edition. McGraw-Hill, New York.

Schwarz G. (1978). Estimating the dimension of a model. *Annals of Statistics*, v. 6, pp. 461–464.

Searle S.R., Casella G., McCulloch C.E. (1992). *Variance Components*. Wiley, New York.

Stewart J. (1991). *Econometrics*. Philip Allan.

Tobin J. (1958). Estimation of relationships for limited dependent variables. *Econometrica*, v. 26, pp. 24–36.

Verbeek M. (2000). *A Guide to Modern Econometrics*. Wiley, Chichester.

White H. (1980). A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. *Econometrica*, v. 48, n. 4, pp. 817–838.

Предметный указатель

Автокорреляционная функция (ACF), 277
– – выборочная, 290
Автокорреляция ошибок, 40
Авторегрессионный процесс (AR), 294
– – первого порядка, 184, 209
Анализ чувствительности, 424
Арбитраж, 463
Асимптотическая информационная матрица, 247
Асимптотическая матрица ковариаций, 247

Базис, 486
Безрисковый актив, 456
Безусловное прогнозирование, 205
Белый шум, 277
Блочная матрица, 503

Вектор остатков, 51, 69, 70, 72
Вектор-столбец, 488
Вектор-строка, 488
Векторное пространство, 484
Внешне не связанные уравнения, 221
Внутригрупповое преобразование, 364
Временной ряд, 30
Выборочные статистики, 531
Выбросы, 34

Гауссовский случайный вектор, 524
Гетероскедастичность, 40, 149, 154, 168
Гиперплоскость, 486
Гипотеза эффективного финансового рынка, 438
Гомоскедастичность, 40, 154
Граница эффективных портфелей, 448, 463

Двухшаговый метод наименьших квадратов (2SLS), 216, 237, 238
Детерминант, 493
Диагональная матрица, 489
Динамическая модель, 268
Дисперсия, 512
Длинная регрессия, 83, 125
Доверительное множество, 533, 537
Доверительный интервал, 50, 79, 537
Доступный обобщенный метод наименьших квадратов, 149, 158, 160, 161, 221

Единичная матрица, 489

Единичный корень, 279

Закон больших чисел, 530
Значимость теста, 540

Идемпотентная матрица, 72, 502
Идентифицируемость, 109, 229, 233, 234
Инструментальные переменные, 212, 214–217
Интенсивность отказов, 348
Информационная матрица, 247
Информационное количество, 535

Квантиль, 512
– двусторонняя, 512
Ковариационная матрица, 514
Ковариация, 514
Коинтеграция, 283
Коинтегрирующий вектор, 283
Короткая регрессия, 84, 125
Коррелограмма, 290
Косвенный метод наименьших квадратов, 227, 228, 237
Коэффициент авторегрессии, 185
– детерминации R^2, 52, 74, 75
– – внутригрупповой, 374
– – межгрупповой, 374
– – объединенный, 374
– – скорректированный, 74, 76
– Йенсена, 470
– корреляции, 515
– смертности, 348
– частной корреляции, 119, 122
– Шарпа, 470
Коэффициенты приведенной формы, 226, 233, 234
Критерий Акаике, 307
– Шварца, 307

Линейная зависимость, 485
Линейная модель вероятности, 322
Линейная независимость, 485
Линейное подпространство, 486
Линейный оператор, 487
Лямбда Хекмана, 344

Математическое ожидание, 511, 514
Матрица, 488
– взаимных ковариаций, 515
– Якоби, 506
Метод Алмон, 266
– взвешенных наименьших квадратов, 168, 169
– инструментальных переменных, 393
– максимального правдоподобия, 244
– – – условный, 304
– моментов, 536
– – обобщенный, 382, 389
– наименьших квадратов (МНК), 34, 55
– – – двухшаговый, 216, 237, 238
– – – косвенный, 227, 228, 237
– – – обобщенный, 148, 154, 156, 157, 222
Мнимая регрессия, 282
Модель авторегрессии и скользящего среднего (ARMA), 293

– адаптивных ожиданий, 273
– Бокса–Дженкинса (ARIMA), 285, 298
– без ограничения, 129
– времени жизни, 348
– геометрических лагов, 267
– динамическая, 268
– Клейна, 240
– Койка, 266
– коррекции ошибок, 274
– множественного выбора, 329
– множественной регрессии, 67
– оценки финансовых активов многофакторная, 468
– – – – факторная, 465
– – – – CAPM, 466
– полиномиальных лагов, 267
– распределенных лагов, 265, 266
– – – авторегрессионная, 265
– с ограничением, 129
– с фиксированным эффектом, 362
– скользящего среднего (MA), 295
– со случайным эффектом, 362, 367
– Хекмана, 342
– частичного приспособления, 272
– *logit*, 321, 323, 324
– – множественного выбора, 332
– – условная, 332
– *probit*, 321, 323, 324
– *tobit*, 339, 340
Мощность теста, 540
Мультиколлинеарность, 109–111

Не вложенные модели, 132
Невырожденная матрица, 495
Независимые случайные величины, 513
Неидентифицируемость, 228
Нелинейные ограничения, 258
Необходимые условия экстремума (минимума), 35, 37, 69
Неотрицательно определенная матрица, 500
Неравенство Рао–Крамера, 535
Несущественные переменные, 124, 127, 131
Нормальный случайный вектор, 524
Нулевая матрица, 489

Обобщенный метод моментов, 382, 389
Обобщенный метод наименьших квадратов, 148, 154, 156, 157, 222
Образ оператора, 487
Обратимость, 295
Обратная матрица, 495
Обратное отношение Миллса, 348
Объединенная модель регрессии, 361
Оператор сдвига, 265
Определитель, 493
Ортогональная матрица, 499
Ортогональная система векторов, 499
Ортонормированная система векторов, 499

Остатки регрессии, 43, 51, 72
Оценка максимального правдоподобия, 55, 57, 162, 246, 536
– межгрупповая, 369
– метода инструментальных переменных, 213, 227, 270
– метода наименьших квадратов, 69
– параметра, 532
– – несмещенная, 533
– – состоятельная, 534
– – эффективная, 534
Ошибка второго рода, 539
– первого рода, 539
Ошибки в измерениях зависимой переменной, 214
– – – независимой переменной, 214

Панельные данные, 357
– – динамические модели, 380
– – модели бинарного выбора, 386
Переменная бинарная, 319
– зависимая, 29
– независимая, 29
– номинальная, 319
– объясняемая, 29
– объясняющая, 29
– ординальная, 319
– переопределенная, 232, 234
– порядковая, 319
– фиктивная, 108, 113, 115–117
– экзогенная, 224, 232, 234
– эндогенная, 224, 232, 235
Плотность распределения, 510, 513
– условного распределения, 516
Подобные матрицы, 497
Полная коллинеарность, 109, 238
Полный рынок, 462
Положительно определенная матрица, 500
Портфель ценных бумаг, 446
Порядковое условие, 236
Пошаговый отбор регрессоров, 122
Предварительное тестирование, 398
Приведенная форма модели, 226, 233
Проблема идентификации, 228, 231
Прогноз, прогнозные значения, 43, 72
Прогнозирование безусловное, 205
– условное, 208
Произведение Кронекера, 504
Произведение матриц, 490
Пропущенные переменные, 38
Пространственные данные, 30
Процедура Дарбина, 188
– Кохрейна–Оркатта, 187
– Хилдрета–Лу, 188
Процентная точка, 512
Процесс, порождающий данные, 124

Ранг матрицы, 494
Ранговое условие, 236
Распределение биномиальное, 518

– гауссовское, 519
– логарифмически нормальное, 520
– нормальное, 519
– показательное, 519
– пуассоновское, 518
– равномерное, 519
– Стьюдента (t-распределение), 521
– – нецентральное, 523
– Фишера (F-распределение), 521
– – нецентральное, 523
– экспоненциальное, 519
– χ^2, 520
– – нецентральное, 522
Регрессионная модель линейная, 39
– – множественная, 67
– – нормальная линейная, 39, 46, 47, 54, 68
Регрессионное уравнение, 38
Регрессор, 38, 67, 68

Сверхидентифицируемость, 236
Сезонность, 286
Симметричная матрица, 498
Системы одновременных уравнений, 221, 224
Скалярная матрица, 488
Скалярное произведение векторов, 491
След матрицы, 492
Случайная величина, 509
– – дискретная, 510
– – непрерывная, 510
Случайная выборка, 531
Случайное блуждание, 278
Случайные величины некоррелированные, 515
Случайный вектор, 513
Собственное число, 496
Собственный вектор, 496
Спецификация модели, 39, 67, 68, 122, 124
Среднее значение, 511, 514
Стандартная форма нормальных уравнений, 35
Стандартное отклонение, 512
Стандартные ошибки в форме Ньюи–Веста, 174
– – в форме Уайта, 173
Статистика Бокса–Пирса (Q-статистика), 306
– Дарбина–Уотсона, 189
– Дики–Фуллера, 280
Статистический тест, 539
Стационарность, 276
– слабая, 277
– строгая, 276
Стохастические регрессоры, 149
Стохастический дисконтирующий множитель, 462, 466
Структурная форма модели, 226, 231
Структурные коэффициенты, 226, 233, 234
Сумма матриц, 489
Существенные переменные, 124, 125, 131
Сходимость по вероятности, 529
– по распределению, 529
– почти наверное, 528

Теорема Айткена, 156
– Гаусса–Маркова, 41, 43, 69, 70, 151
– Слуцкого, 531

Тест Бреуша–Пагана, 179
– Вальда, 255
– Голдфелда–Куандта, 178
– Гранжера, 275
– Дарбина–Уотсона, 189, 192
– Дики–Фуллера, 281
– Льюнга–Бокса, 306
– множителей Лагранжа, 255, 272
– отношения правдоподобия, 255
– Уайта, 177
– Хаусмана, 378
– Чоу, 85
– J-тест, 132
– $RESET$-тест, 133
Точная идентификация, 236
Транспонирование матрицы, 490
Тренд, 285

Уравнение правдоподобия, 536
Уравнения в отклонениях, 36
– Юла–Уолкера, 294
Урезанная выборка, 320, 337
Уровень доверия, 537
Условная *logit*-модель, 332
Условное математическое ожидание, 516
Условное прогнозирование, 208
Условный метод максимального правдоподобия, 304

Фиктивная переменная, 108, 113, 115–117
Фронт эффективных портфелей, 448, 463
Функция правдоподобия, 56, 245, 246, 325, 536
– – логарифмическая, 246, 253
– распределения, 509, 512, 513
– регрессии, 517
– Хубера, 33, 34

Характеристическое уравнение матрицы, 497
– число матрицы, 497

Цензурированная выборка, 320, 339
Центральная предельная теорема, 530

Частная автокорреляционная функция (PACF), 290

Эффект несинхронной торговли, 446
Эффективный портфель, 447

Ядро оператора, 488

ARCH, 311, 312

CAPM, 466

Dummy trap, 114, 181

GARCH, 311, 313

Heckman lambda, 344

Multinomial *logit* model, 332

P-значение, 540
Payoff, 461
Pooled model, 361
Pretest-оценка, 399, 403, 405, 407
Pretesting, 398

Q-статистика Бокса–Пирса, 306

WALS-оценка, 404

Учебник

Ян Р. МАГНУС,
Павел Константинович КАТЫШЕВ,
Анатолий Абрамович ПЕРЕСЕЦКИЙ

ЭКОНОМЕТРИКА

Начальный курс

Гл. редактор *Ю.В. Луизо*
Зав. редакцией *Г.Г. Кобякова*
Художник *Н.В. Пьяных*
Компьютерная подготовка оригинал-макета *С.В. Головань*
Технический редактор *Л.А. Зотова*
Корректор *Н.В. Андрианова*

Санитарно-эпидемиологическое заключение № 77.99.02.953.Д.008700.12.02
от 24.12.2002 г.

Подписано в печать 12.11.2003. Формат 60 × 90 1/16. Бумага офсетная. Гарнитура Таймс.
Печать офсетная. Усл. печ. л. 36,0. Тираж 4000 экз. Заказ № 915. Изд. № 437.

Издательство "Дело"
119571, Москва, пр-т Вернадского, 82
Коммерческий отдел — тел. 433-2510, 433-2502
E-mail: delo@ane.ru
Internet: http://www.delo.ane.ru

Интернет-магазин: www.delokniga.ru

ОАО «Московская типография № 6»
115088, Москва, Южнопортовая ул., 24

ISBN 5-7749-0055-X

9 785774 900558 >