Оглавление

Введение 11

I Введение в социально-экономическую статистику 15

1. Основные понятия 17

1.1. Краткая историческая справка 17
1.2. Предмет статистики 18
1.3. Экономические величины и статистические показатели 20
1.4. Вероятностная природа экономических величин 22
1.5. Проблемы измерений 24
1.6. Специфика экономических измерений 27
1.7. Адекватность экономических измерений 29
1.8. Типы величин, связи между ними 32
1.9. Статистические совокупности и группировки 36
1.10. Задачи 45

2. Описательная статистика 48

2.1. Распределение частот количественного признака 48
2.2. Средние величины 53
2.3. Медиана, мода, квантили 66

2.4. Моменты и другие характеристики распределения 70
2.5. Упражнения и задачи 83

3. Индексный анализ **89**
3.1. Основные проблемы 89
3.2. Способы построения индексов 93
3.3. Факторные представления приростных величин 100
3.4. Случай, когда относительных факторов более одного 104
3.5. Индексы в непрерывном времени 106
3.6. Прикладные следствия из анализа индексов в непрерывном времени 116
3.7. Факторные представления приростов в непрерывном времени . . . 123
3.8. Задачи 123

4. Введение в анализ связей **129**
4.1. Совместные распределения частот количественных признаков . . . 129
4.2. Регрессионный анализ 141
4.3. Дисперсионный анализ 160
4.4. Анализ временных рядов 167
4.5. Упражнения и задачи 172

II Эконометрия — I: Регрессионный анализ **179**

5. Случайные ошибки **182**
5.1. Первичные измерения 183
5.2. Производные измерения 192
5.3. Упражнения и задачи 194

6. Алгебра линейной регрессии **199**
6.1. Линейная регрессия 199
6.2. Простая регрессия 201
6.3. Ортогональная регрессия 205
6.4. Многообразие оценок регрессии 210

6.5. Упражнения и задачи 216

7. Основная модель линейной регрессии 222

7.1. Различные формы уравнения регрессии 222
7.2. Основные гипотезы, свойства оценок 226
7.3. Независимые факторы: спецификация модели 234
7.4. Прогнозирование 244
7.5. Упражнения и задачи 247

8. Нарушение гипотез основной линейной модели 257

8.1. Обобщенный метод наименьших квадратов (взвешенная регрессия) 257
8.2. Гетероскедастичность ошибок 258
8.3. Автокорреляция ошибок 265
8.4. Ошибки измерения факторов 270
8.5. Метод инструментальных переменных 273
8.6. Упражнения и задачи 278

9. Целочисленные переменные в регрессии 289

9.1. Фиктивные переменные 289
9.2. Модели с биномиальной зависимой переменной 295
9.2.1. Линейная модель вероятности, логит и пробит 296
9.2.2. Оценивание моделей с биномиальной зависимой переменной 298
9.2.3. Интерпретация результатов оценивания моделей с биномиальной зависимой переменной 302
9.3. Упражнения и задачи 304

10. Оценка параметров систем уравнений 314

10.1. Невзаимозависимые системы 314
10.2. Взаимозависимые или одновременные уравнения 318
10.3. Оценка параметров отдельного уравнения 324
10.4. Оценка параметров системы идентифицированных уравнений . . . 331
10.5. Упражнения и задачи 334

III Эконометрия — I: Анализ временных рядов 345

11. Основные понятия в анализе временных рядов 347

11.1. Введение 347
11.2. Стационарность, автоковариации и автокорреляции 351
11.3. Основные описательные статистики для временных рядов 353
11.4. Использование линейной регрессии с детерминированными факторами для моделирования временного ряда 356
11.4.1. Тренды 356
11.4.2. Оценка логистической функции 358
11.4.3. Сезонные колебания 359
11.4.4. Аномальные наблюдения 360
11.5. Прогнозы по регрессии с детерминированными факторами 361
11.6. Критерии, используемые в анализе временных рядов 365
11.6.1. Критерии, основанные на автокорреляционной функции . . 366
11.6.2. Критерий Спирмена 369
11.6.3. Сравнение средних 370
11.6.4. Постоянство дисперсии 372
11.7. Лаговый оператор 373
11.8. Модели регрессии с распределенным лагом 375
11.9. Условные распределения 377
11.10. Оптимальное в среднеквадратическом смысле прогнозирование: общая теория 378
11.10.1. Условное математическое ожидание как оптимальный прогноз 378
11.10.2. Оптимальное линейное прогнозирование 380
11.10.3. Линейное прогнозирование стационарного временного ряда 382
11.10.4. Прогнозирование по полной предыстории. Разложение Вольда 385
11.11. Упражнения и задачи 388

12. Сглаживание временного ряда 391

12.1. Метод скользящих средних 391

12.2. Экспоненциальное сглаживание 398
12.3. Упражнения и задачи 402

13. Спектральный и гармонический анализ **406**
13.1. Ортогональность тригонометрических функций и преобразование Фурье 406
13.2. Теорема Парсеваля 411
13.3. Спектральный анализ 412
13.4. Связь выборочного спектра с автоковариационной функцией . . . 414
13.5. Оценка функции спектральной плотности 417
13.6. Упражнения и задачи 422

14. Линейные стохастические модели ARIMA **426**
14.1. Модель линейного фильтра 426
14.2. Влияние линейной фильтрации на автоковариации и спектральную плотность 429
14.3. Процессы авторегрессии 431
14.4. Процессы скользящего среднего 452
14.5. Смешанные процессы авторегрессии — скользящего среднего . . 457
14.6. Модель ARIMA 463
14.7. Оценивание, распознавание и диагностика модели Бокса—Дженкинса 466
14.8. Прогнозирование по модели Бокса—Дженкинса 475
14.9. Модели, содержащие стохастический тренд 485
14.10. Упражнения и задачи 490

15. Динамические модели регрессии **500**
15.1. Модель распределенного лага: общие характеристики и специальные формы структур лага 500
15.2. Авторегрессионная модель с распределенным лагом 506
15.3. Модели частичного приспособления, адаптивных ожиданий и исправления ошибок 509
15.4. Упражнения и задачи 513

16. Модели с авторегрессионной условной гетероскедастичностью **523**
16.1. Модель ARCH 524

16.2. Модель GARCH 527
16.3. Прогнозы и доверительные интервалы для модели GARCH 531
16.4. Разновидности моделей ARCH 535
16.4.1. Функциональная форма динамики условной дисперсии . . . 535
16.4.2. Отказ от нормальности 536
16.4.3. GARCH-M 537
16.4.4. Стохастическая волатильность 537
16.4.5. ARCH-процессы с долгосрочной памятью 538
16.4.6. Многомерные модели волатильности 539
16.5. Упражнения и задачи 540

17. Интегрированные процессы, ложная регрессия и коинтеграция 546
17.1. Стационарность и интегрированные процессы 546
17.2. Разложение Бевериджа—Нельсона для процесса I(1) 550
17.3. Ложная регрессия 551
17.4. Проверка на наличие единичных корней 553
17.5. Коинтеграция. Регрессии с интегрированными переменными 558
17.6. Оценивание коинтеграционной регрессии: подход Энгла—Грейнджера 560
17.7. Коинтеграция и общие тренды 561
17.8. Упражнения и задачи 563

IV Эконометрия — II 567

18. Классические критерии проверки гипотез 569
18.1. Оценка параметров регрессии при линейных ограничениях 569
18.2. Тест на существенность ограничения 572
18.2.1. Тест Годфрея (на автокорреляцию ошибок) 577
18.2.2. Тест RESET Рамсея (Ramsey RESET test) на функциональную форму уравнения 578
18.2.3. Тест Чоу (Chow-test) на постоянство модели 578
18.3. Метод максимального правдоподобия в эконометрии 582
18.3.1. Оценки максимального правдоподобия 582

18.3.2. Оценки максимального правдоподобия для модели линейной регрессии 584
18.3.3. Три классических теста для метода максимального правдоподобия 587
18.3.4. Сопоставление классических тестов 592
18.4. Упражнения и задачи 593

19. Байесовская регрессия **601**
19.1. Оценка параметров байесовской регрессии 603
19.2. Объединение двух выборок 606
19.3. Упражнения и задачи 607

20. Дисперсионный анализ **611**
20.1. Дисперсионный анализ без повторений 612
20.2. Дисперсионный анализ с повторениями 618
20.3. Упражнения и задачи 621

21. Модели с качественными зависимыми переменными **625**
21.1. Модель дискретного выбора для двух альтернатив 625
21.2. Оценивание модели с биномиальной зависимой переменной методом максимального правдоподобия 627
21.2.1. Регрессия с упорядоченной зависимой переменной 630
21.2.2. Мультиномиальный логит 631
21.2.3. Моделирование зависимости от посторонних альтернатив в мультиномиальных моделях 633
21.3. Упражнения и задачи 635

22. Эффективные оценки параметров модели ARMA **644**
22.1. Оценки параметров модели AR(1) 644
22.2. Оценка параметров модели MA(1) 647
22.3. Оценки параметров модели ARMA(p, q) 651
22.4. Упражнения и задачи 652

23. Векторные авторегрессии **654**
23.1. Векторная авторегрессия: формулировка и идентификация 654
23.2. Стационарность векторной авторегрессии 658

23.3. Анализ реакции на импульсы 660
23.4. Прогнозирование с помощью векторной авторегрессии 662
23.5. Причинность по Грейнджеру 665
23.6. Коинтеграция в векторной авторегрессии 666
23.7. Метод Йохансена 668
23.8. Коинтеграция и общие тренды 674
23.9. Упражнения и задачи 676

A. Вспомогательные сведения из высшей математики 691
A.1. Матричная алгебра 691
A.1.1. Определения 691
A.1.2. Свойства матриц 694
A.2. Матричное дифференцирование 700
A.2.1. Определения 700
A.2.2. Свойства 701
A.3. Сведения из теории вероятностей и математической статистики . . 703
A.3.1. Характеристики случайных величин 703
A.3.2. Распределения, связанные с нормальным 709
A.3.3. Проверка гипотез 712
A.4. Линейные конечно-разностные уравнения 714
A.4.1. Решение однородного конечно-разностного уравнения . . . 714
A.5. Комплексные числа 715

B. Статистические таблицы 717

Введение

Данный учебник написан на основе курсов, читаемых на экономическом факультете Новосибирского государственного университета. С середины 1980-х годов читался спецкурс, в котором излагались основы классической эконометрии, относящиеся к регрессионному анализу. В это же время в рамках «Общей теории статистики» достаточно развернуто начал изучаться материал анализа временных рядов. На базе этих дисциплин в начале 1990-х годов был создан единый курс «Эконометрия», который, постоянно совершенствуясь, читается как обязательный до настоящего времени. Во второй половине 1990-х годов был разработан и введен в практику преподавания обязательный курс «Эконометрия-II» для магистрантов. В конце 1990-х годов на экономическом факультете был восстановлен — на принципиально новом уровне — курс «Общая теория статистики», дающий начальное представление об эмпирических исследованиях.

Эконометрия (другой вариант термина в русском языке — эконометрика) — это инструментальная наука, позволяющая изучать количественные взаимосвязи экономических объектов и процессов с помощью математических и статистических методов и моделей. Дословно этот термин означает «экономическое измерение».

Эконометрия связывает экономическую теорию, прикладные экономические исследования и практику. Благодаря эконометрии осуществляется обмен информацией между этими взаимодополняющими областями, происходит взаимное обогащение и взаимное развитие теории и практики.

Эконометрия дает методы экономических измерений, а также методы оценки параметров моделей микро- и макроэкономики. При этом экономические теории выражаются в виде математических соотношений, а затем проверяются эмпирически статистическими методами. Кроме того, эконометрия активно используется для прогнозирования экономических процессов и позволяет проводить планирование как в масштабах экономики в целом, так и на уровне отдельных предприятий.

В экономике (как и в большинстве других научных дисциплин) не существует и не может существовать абсолютно точных утверждений. Любое эмпирическое утверждение имеет вероятностную природу. В частности, экономические измерения содержат различного рода ошибки. Таким образом, в прикладных экономических исследованиях требуется использовать статистические методы.

Методы эконометрии, позволяющие проводить эмпирическую проверку теоретических утверждений и моделей, выступают мощным инструментом развития самой экономической теории. С их помощью отвергаются одни теоретические концепции и принимаются другие гипотезы. Теоретик, не привлекающий эмпирический материал для проверки своих гипотез и не использующий для этого эконометриче-

ские методы, рискует оказаться в мире своих фантазий. Важно, что эконометрические методы одновременно позволяют оценить ошибки измерений экономических величин и параметров моделей.

Экономист, не владеющий методами эконометрии, не может эффективно работать аналитиком. Менеджер, не понимающий значение этих методов, обречен на принятие ошибочных решений.

Эта книга адресована студентам, магистрантам и аспирантам экономических факультетов классических университетов. Она соответствует требованиям государственного образовательного стандарта по дисциплине «Эконометрика». Кроме того, издание будет полезно преподавателям эконометрии, исследователям, работающим в области прикладной экономики, специалистам по бизнес-планированию и финансовым аналитикам.

Учебник предполагает определенный уровень базовой математической подготовки читателя, владение им основами линейной алгебры, математического анализа, теории вероятностей и математической статистики в объеме курсов для нематематических специальностей вузов. Некоторые наиболее важные сведения из этих разделов высшей математики приведены в приложении к учебнику.

Необходимость в создании учебника по эконометрии вызвана отсутствием отечественного варианта, который бы охватывал все основополагающие позиции современной эконометрической науки. Появившиеся в последние годы учебные издания лишь частично покрывают программу курса, читаемого на экономическом факультете Новосибирского государственного университета. В частности, эти учебники, посвященные в основном регрессионному анализу, не уделяют достаточного внимания теории временных рядов. При создании настоящего учебника авторы стремились систематизировать и объединить в рамках одного источника различные разделы экономической статистики и эконометрии.

Структура учебника примерно соответствует учебному плану экономического факультета НГУ. Соответственно, он состоит из четырех частей: «Введение в социально-экономическую статистику», «Эконометрия-I: регрессионный анализ», «Эконометрия-I: анализ временных рядов», «Эконометрия-II». Каждая часть покрывает семестровый курс. Соответствующие разделы читаются в качестве обязательной дисциплины во втором, четвертом и пятом семестрах бакалавриата и в первом семестре магистратуры. Полный курс эконометрии на ЭФ НГУ (включая «Введение в социально-экономическую статистику») рассчитан на 152 часа аудиторных занятий (45% лекций, 55% семинарских занятий).

В первой части «Введение в социально-экономическую статистику» представлен материал, который более глубоко раскрывается в других частях учебника. В данной части рассмотрены особенности экономических величин, изложены про-

блемы экономических измерений, приводится обсуждение основных описательных статистик, рассмотрен индексный анализ, дан обзор основ анализа связей.

Вторая часть посвящена классическому регрессионному анализу. Здесь рассматривается метод наименьших квадратов в разных вариантах (включая ортогональную регрессию), приведена основная модель линейной регрессии, излагаются методы оценки параметров регрессии в случаях, когда нарушаются требования основной модели (мультиколлинеарность, автокорреляция и гетероскедастичность, наличие ошибок в переменных), рассматриваются способы включения в регрессионное уравнение качественных переменных как для факторов (фиктивные или псевдопеременные), так и для зависимой переменной (модели логит и пробит). Большое внимание уделяется применению основных критериев проверки статистических гипотез в регрессионном анализе (тестированию): критерии Стьюдента, Фишера и Дарбина—Уотсона. Завершается вторая часть изложением некоторых проблем и методов оценки параметров одновременных систем уравнений. Особенность этого раздела учебника состоит в использовании матричного подхода, позволяющего достичь общности и лаконичности изложения материала.

Третья часть посвящена анализу временных рядов. В ней рассматривается как классический инструментарий — выделение трендов, спектральный и гармонический анализ, модели Бокса—Дженкинса, так и более современные методы — динамическая регрессия, ARCH- и GARCH- процессы, единичные корни и коинтеграция, которые недостаточно освещены в отечественной литературе. Классические методы излагаются исходя из стремления дать математическое обоснование множеству утверждений, которые в существующих учебниках просто констатируются, что существенно затрудняет восприятие материала.

Заключительная четвертая часть содержит разделы, в большинстве своем неизвестные русскоязычному читателю, однако без их знания практически невозможно проведение качественного эконометрического исследования. Это классические критерии проверки гипотез, метод максимального правдоподобия, дисперсионный анализ, основы байесовских методов, модели с качественными зависимыми переменными и более сложные разделы анализа временных рядов, в частности, векторная авторегрессия и подход Йохансена к анализу коинтеграционных связей.

Учебник содержит большое количество задач и упражнений. Кроме того, в каждой главе приведен список литературы, которая может быть использована в качестве дополнения к материалу главы.

Подготовка ученика осуществлялась при финансовой и методической поддержке программ TEMPUS (TACIS) JEP 08508–94: «Перестройка и совершенствование подготовки экономистов в НГУ» (1994–1997гг.) и «Совершенствование

преподавания социально-экономических дисциплин в вузах» в рамках «Инновационного проекта развития образования (2002–2004гг.)».

В списке литературы после каждой главы звездочкой отмечены основные источники.

Авторский коллектив благодарит всех, кто помогал в работе над учебником. Особая благодарность Владимиру Шину, который осуществил верстку оригинал-макета в формате LaTeX, а также Марине Шин, проделавшей большую работу по редактированию и согласованию различных частей учебника.

Авторы будут признательны читателям за любые комментарии, сообщения о недочетах и опечатках в этом учебнике.

Часть I

Введение в социально-экономическую статистику

Это пустая страница

Глава 1

Основные понятия

1.1. Краткая историческая справка

Практика статистики зародилась давно, по-видимому, вместе со становлением элементов государственности. Не случайно во многих языках статистика и государство — однокоренные слова. Государству в лице представителей госаппарата всегда надо было хотя бы приблизительно знать численность населения страны, ее экономический потенциал, фактическое состояние дел в разных сферах общественной жизни. Иначе нельзя сколько-нибудь эффективно собирать налоги, проводить крупные строительные работы, вести войны и т.д.

Статистическая теория возникла как результат обобщения уже достаточно развитой статистической практики. Начало ее становления обычно связывают с работами английских политических арифметиков XVII века и, прежде всего, с именем **Вильяма Петти** (1623–1687). В XVIII веке статистическая теория развивалась под флагом государствоведения, зародившегося в Германии. Именно германские ученые в конце XVIII века стали использовать термины «статистика», «статистик», «статистический» в смысле, близком к современному. Хотя слово «статистик» намного старше, его — в ином смысле — можно найти в произведениях Шекспира (начало XVII века). Эти слова, по-видимому, происходят более или менее косвенно от латинского слова «status» в том его смысле, который оно приобрело в средневековой латыни, — **политическое состояние**.

Германские авторы, и вслед за ними известный английский ученый сэр **Джон Синклер**, использовали термин «статистика» в смысле простого изложения заслу-

живающих внимание данных, характеризующих государство. Причем форма изложения являлась преимущественно словесно-текстовой. Для того времени такое понимание было достаточно естественным, т.к. достоверных числовых данных было еще очень мало. Лишь спустя несколько десятилетий с термином «статистика» стали связывать изложение характеристик государства **численным методом**. Но даже после образования в Англии Королевского статистического общества в 1834 году такое понимание статистики еще не стало обычным.

Одним из ярких представителей статистики XIX века является бельгийский ученый **Адольф Кетле** (1796–1874) — создатель первого в мире центрального государственного статистического учреждения в Бельгии, организатор и участник первых международных статистических конгрессов. Он установил, что многие массовые явления (рождаемость, смертность, преступность и т.д.) подчиняются определенным закономерностям, и применил математические методы к их изучению.

В России первым общегосударственным органом статистики явилось Статистическое отделение Министерства полиции (1811), а затем — Министерства внутренних дел (1819). Его начальником был один из первых российских статистиков **Герман К.Ф.** (1767–1838) — автор первого русского оригинального труда по теории статистики — книги «Всеобщая теория статистики» (1809).

Корни современной теории статистики, прежде всего математической статистики, могут быть прослежены в работах **Лапласа** и **Гаусса** по теории ошибок наблюдения, но начало расцвета самой науки относится только к последней четверти XIX века. Значительную роль на этом этапе сыграли работы **Гальтона** и **Карла Пирсона**.

1.2. Предмет статистики

В статистике собираются и систематизируются факты, которые затем анализируются и обобщаются в «содержательных» общественных науках. Поэтому не всегда бывает просто провести границу между собственно статистикой и той общественной наукой, которую она «снабжает» информацией. И многие статистики склонны расширять рамки своей дисциплины за счет «содержательной» тематики. Это — их право, но в строгом смысле **статистика является наукой о методах количественного (численного) отражения фактов общественной жизни**. Именно так понимается статистика в данной книге.

Требуется пояснить, почему в данном определении статистики она связывается именно с науками об обществе.

Любая наука, основываясь на наблюдениях за реальными фактами, стремится их систематизировать, обобщить, выявить закономерности, найти законы, построить теоретические модели, объясняющие наблюдаемую действительность. Други-

ми словами, наука стремится выявить и затем количественно определить структуру причинно-следственных связей. Но события реальной жизни происходят под влиянием многих причин, и простое **пассивное наблюдение** далеко не всегда дает возможность найти эти причины. Более того, такое наблюдение может привести к выводам, прямо противоположным действительности. «Не верь глазам своим» — фраза, резюмирующая многовековой опыт подобных наблюдений.

Однако, в так называемых точных науках научились проводить наблюдения так, чтобы однозначно и, как правило, в количественной форме определять причинно-следственные связи. Такая организация наблюдения называется **экспериментом**. Ученые — физики, химики, биологи могут провести «натурный» эксперимент, на входе которого фиксируются одна-две величины и определяется в результате, на что и как они влияют «при прочих равных условиях». В точных науках анализируются и обобщаются, как правило, наблюдения-результаты экспериментов, т.е. «рафинированные» экспериментальные данные. Прогресс в этих науках самым непосредственным образом связан с целенаправленным развитием возможностей экспериментирования, с развитием «синхрофазотронов».

Возможности проведения управляемых экспериментов в общественной жизни крайне ограничены. Поэтому общественные науки вынуждены опираться на **неэкспериментальные данные**, т.е. на результаты пассивных наблюдений, в потоке которых трудно уловить, а тем более количественно определить причинно-следственные связи. И статистика как раз и занимается методами сбора и подготовки таких данных к анализу, методами их первичного анализа, методами проверки теоретических гипотез на основе таких данных.

Конечно, и во многих необщественных сферах знания остается большое поле для статистики. Метеоролог строит свои прогнозы, основываясь в конечном счете на статистических данных; возможности управляемого эксперимента все еще ограничены в биологии и т.д. Но главным объектом статистики все-таки является общественная жизнь.

Статистикой называют не только науку о методах организации пассивного наблюдения, методах систематизации и первичного анализа таких наблюдений, но и сами массивы этих наблюдений. Статистика рождаемости и смертности, статистика выпуска продукции и т.д. — это совокупности чисел, характеризующих количество рождений и смертей, объемы выпуска продукции и т.д. В этом смысле термин «статистика» эквивалентен термину «информация».

Английским эквивалентом слова «статистика» в указанных смыслах является «statistics», т.е. слово во множественном числе. Это слово используется в английском языке и в единственном числе — «statistic», как определенное число, являющееся результатом некоторого статистического расчета. В этом смысле слово «статистика» используется и в русском языке: статистика Фишера, статистика

Стьюдента, статистика Дарбина-Уотсона — это определенные числа, полученные в результате достаточно сложных расчетов, по величине которых судят о разумности тех или иных статистических гипотез. Например, гипотезы о наличии связи между изучаемыми величинами. Термин «статистики» (во множественном числе), используемый также в русском языке, относится к совокупности таких чисел.

1.3. Экономические величины и статистические показатели

Экономическая величина — есть некоторое количество определенного экономического «качества». Обычно экономические величины обозначают буквами латинского, реже — греческого алфавита. Когда говорят, что x — объем производства или объем затрат, или объем капитала, то подразумевают, что эта буква обозначает некоторое количество произведенной продукции, осуществленных затрат, наличного капитала. Обозначенные таким образом экономические величины используются обычно как переменные и параметры математических моделей экономики, в которых устанавливаются зависимости между экономическими величинами. Примером такой модели может являться межотраслевой баланс:

$$X = AX + Y,$$

где X и Y — вектор-столбцы объемов производства валовой и конечной продукции по отраслям; A — квадратная матрица коэффициентов материальных затрат. Или в покомпонентной записи:

$$x_i = \sum_j a_{ij} x_j + y_i \quad \text{для всех } i,$$

где a_{ij} — коэффициент затрат продукции i-го вида на производство единицы продукции j-го вида.

Эта модель определяет зависимость между валовой, промежуточной и конечной продукцией, а именно: валовая продукция является суммой промежуточной и конечной продукции. Кроме того, в этой модели определяется прямо пропорциональная зависимость текущих материальных затрат от валового выпуска.

Одна из возможных форм зависимости между выпуском продукции и используемыми ресурсами устанавливается производственной функцией Кобба—Дугласа:

$$X = aC^{\alpha} L^{\beta},$$

где X — выпуск продукции; C — затраты основного капитала; L — затраты труда; a, α, β — параметры функции.

В этих записях экономические величины выступают, прежде всего, как некие теоретические понятия, то есть именно как «количества определенного экономического качества». Вопрос об измеримости этих величин непосредственно не ставится, но предполагается, что этот вопрос в принципе разрешим.

Статистическим (экономическим) показателем является **операциональное** определение экономической величины. Такое определение представляет собой исчерпывающий перечень операций, которые необходимо провести, чтобы измерить данную величину. Этот перечень включает обычно и операции по сбору первичной информации — первичных наблюдений. Операциональные определения экономических величин-показателей, особенно обобщающего характера, таких как валовой внутренний или валовой национальный продукт, являются сложными методиками расчетов, далеко не все этапы которых безоговорочно однозначны. Эти операциональные определения являются предметом изучения и построения в социально-экономической статистике.

Одной экономической величине могут соответствовать несколько статистических показателей, которые раскрывают разные стороны соответствующего теоретического понятия. Так, например, понятию «цена» на практике соответствуют: основные цены, цены производителей, оптовые и розничные цены, цены покупателей и т.д. Даже такая, казалось бы, простая величина, как население, имеет несколько «конкретизаций»: население на момент времени или в среднем за период, население постоянное или наличное.

Статистическим показателем называют также конкретное число — результат измерения экономической величины, характеризующей определенный объект в определенный момент времени. Например, чистая прибыль такого-то предприятия в таком-то году составила столько-то миллионов рублей. В этом случае экономическая величина «чистая прибыль» характеризует данное предприятие в данном году. С этой точки зрения понятно, почему в статистике экономические величины в привязке к объекту и времени иногда называют **признаками** этого объекта. В свою очередь статистический показатель-число называют **статистическим наблюдением**. Все множество величин-признаков или показателей-наблюдений можно обозначить следующим образом:

$$X = \{x_{tij}\},$$

где t — индекс времени, i — индекс объекта, j — индекс признака, то есть номер экономической величины в перечне всех экономических величин, которые могут характеризовать изучаемые объекты.

Итак, **экономическая величина-признак** — теоретическое понятие, **статистический показатель-определение** обеспечивает практическую измеримость теоретической величины, **статистический показатель-наблюдение** — результат измерения величины-признака конкретного объекта в конкретный момент времени.

1.4. Вероятностная природа экономических величин

Статистическое исследование строится в предположении, что все экономические величины без исключения являются случайными с вполне определенными, часто неизвестными, законами распределения вероятности. Наблюдаемые значения суть реализации соответствующих случайных величин, выборки из каких-то генеральных совокупностей. Такое отношение к экономическим величинам долгое время отрицалось в отечественной (советской) науке на том основании, что в социалистической экономике, которая сознательно и планомерно организуется, не может быть места случайной компоненте. В настоящее время такую позицию никто практически не занимает, но определенные сомнения в вероятностной природе экономических величин высказываются.

Некоторые экономисты не склонны признавать вероятностный характер немассовых, единичных и уникальных событий. На том основании, что такой немассовый, нерегулярно повторяющийся характер имеет большинство экономических явлений, «отец» кибернетики Норберт Винер вообще отрицал возможность применения количественных методов в экономических и социальных науках. Многие ученые-статистики отрицают необходимость вероятностного подхода к изучению даже массовых явлений, если для них можно провести сплошное наблюдение и получить в свое распоряжение — как они считают — полную генеральную совокупность. Они работают в рамках особого раздела статистики, который называется **анализом данных**.

Нельзя не видеть, что высказываемые сомнения в вероятностной природе экономических явлений имеют основания. Понятие вероятности, вероятностные подходы к анализу зарождались и развивались в естественных науках, а мир физических величин очень сильно отличается от «материи» экономической. В физике, химии генеральные совокупности очень велики, многие из них, по-видимому, можно считать бесконечными. Очень велики и исследуемые выборки, и, что чрезвычайно важно, их, как правило, можно неограниченно увеличивать в управляемом эксперименте, воспроизводя нужные условия в специальных физических установках, в химической аппаратуре. В такой ситуации совершенно естественным кажется определение вероятности как предела относительной частоты появления нужного признака.

Но и в физическом мире многие явления представляются единичными и уникальными, со всеми вытекающими отсюда трудностями для классического, «объективистского», «частотного» понимания вероятности. Например, как может ответить на вопрос о том, какова вероятность жизни на Марсе, «объективист-частотник»? Если он относится к Марсу как к уникальному явлению, единственной в своем роде планете во вселенной, то в лучшем случае его ответ будет 0 или 1.

Если жизнь есть — 1, если ее нет — 0. Но, скорее всего, он просто отметит некорректность этого вопроса, поскольку для него вероятность — это характеристика совокупности, а не единичного явления.

В экономике подобных нарушений классических условий появления вероятности — масса. Можно сказать, что вся экономика состоит из таких нарушений. Мир людей, если к нему относиться «сильно материалистично», без некоторой раскованности в мышлении, уникален и ограничен. Генеральные совокупности конечны и малы, так что многие массивы данных можно интерпретировать как исчерпывающие генеральные совокупности. Ряды наблюдений весьма коротки. И, что сильно отличает экономику от физики, невозможно проведение натурных экспериментов с воспроизводимыми условиями.

В таком положении полезным и продуктивным, по крайней мере внешне, представляется подход субъективной вероятности. Субъективная вероятность — это мера доверия исследователя к утверждению, степень уверенности в его справедливости, наконец, мера готовности действовать в ситуации, связанной с риском. «Субъективист» может давать вероятности любым, даже уникальным событиям, включая их тем самым в строгий анализ. Основываясь на своих знаниях, опыте, интуиции, он может определить вероятность жизни на Марсе, вероятность вхождения России в число развитых стран, вероятность экологической катастрофы на планете или мировой войны к середине столетия. Конечно, его оценки индивидуальны и субъективны, но если их несколько и даже много, то после своего согласования они, несомненно, приобретут элемент объективности. Полезно понимать, что и в таком случае подход к вероятности совершенно отличен от классического «объективистски-частотного».

Направления **субъективной и объективной вероятности** развивались параллельно. Если формальное определение объективной вероятности дано впервые **Пуассоном** во второй четверти прошлого века (1837 год), то **Бернулли** еще в начале XVIII века (1713 год) предположил, что вероятность — это степень доверия, с которой человек относится к случайному событию, и что эта степень доверия зависит от его знаний и у разных людей может быть различной. Во второй половине прошлого века **Байес** доказал известную теорему об условной вероятности и интерпретировал используемые в ней параметры вероятности как степени уверенности. Эти идеи легли в основу современной теории принятия решений в условиях неопределенности и, вообще, подхода субъективной вероятности, который часто называется байесовским.

Бурное развитие этого направления началось в XX веке в связи с усилением интереса к наукам об обществе, к экономической науке. Следует назвать по крайней мере двух ученых, внесших фундаментальный вклад в становление теории субъективной вероятности и связанных с ней теорий полезности — это **Джон М. Кейнс** и **Фрэнк П. Рамсей**. В СССР в 40-х годах прошлого столетия проходила дискуссия

о началах теории вероятностей. Представители субъективной школы потерпели поражение.

Существует подход, объединяющий в определенном смысле идеи субъективной и объективной вероятности. Он основан на понимании многовариантности развития общества вообще и экономики в частности. Имеется множество возможных состояний экономики и путей ее развития, наблюдаемые факты в полном своем объеме являются лишь выборкой из **гипотетической генеральной совокупности**, образованной этим множеством. В рамках такого подхода снимается ряд противоречий частотного понимания вероятностей в экономике. Так, например, вероятность вхождения России в число развитых стран к 2050 году есть относительная частота возможных вариантов развития страны, при которых «вхождение» состоялось к 2050 году, в общем числе возможных вариантов. Вопрос остается только в том, как можно найти эти варианты или хотя бы посчитать их количество, т.е. как можно практически работать с гипотетическими генеральными совокупностями.

Современная экономическая наука располагает соответствующим инструментарием: это математическое моделирование. Всякая математическая модель представляет бесконечное пространство возможных состояний экономики, расчет по модели дает точку или траекторию в этом пространстве. Модель выступает инструментом проведения экономических экспериментов почти в таком же смысле, как и в естественных науках. Конечно, главным при этом является вопрос об адекватности модели. Но все это — темы других курсов.

По-видимому, «субъективист», хотя бы в некоторых ситуациях приписывая вероятности тем или иным событиям, пользуется неявно частотным подходом применительно к некоторым гипотетическим генеральным совокупностям. При этом конструировать эти гипотетические совокупности и работать с ними помогают ему его знания, опыт и интуиция.

1.5. Проблемы измерений

Методы измерения развивались на протяжении всей истории человеческой цивилизации вместе с развитием математики и естественных наук. В прошлом веке математизация социальных и экономических наук дала новый импульс этим процессам. Проводилось серьезное переосмысление феномена измерений, осуществлялись продуктивные попытки разработать общие теории измерения. Шел интенсивный поток литературы, посвященной этой проблематике. Следует назвать таких ученых, как **Н.Р. Кэмпбел**, один из родоначальников современных теорий измерения; **С.С.Стивенс**, одна из его книг, 2-х томная «Экспериментальная психология», в 1969 г. опубликована на русском языке; **И. Пфанцагль**, книга которого в соавтор-

стве с двумя другими учеными «Теория измерений» вышла в нашей стране в 1971 г.; **П. Суппес** и **Дж.Л. Зинес**, их работа «Основы теории измерения» опубликована у нас в 1967 г. Существенный вклад в теорию экономических измерений внесен работой **Дж. фон Неймана** и **О. Моргенштерна** «Теория игр и экономическое поведение», вышедшей у нас в 1970 г. Характерно, что все эти исследователи, кроме Кэмпбела, разрабатывали проблематику нефизических измерений.

Если взять «Большую Советскую Энциклопедию» или «Математическую Энциклопедию» более позднего издания, то можно узнать, что измерение — это процесс сопоставления измеряемого явления с единицей измерения. Такое определение достаточно поверхностно, оно не раскрывает существа возникающих проблем.

В настоящее время практически всеобщим признанием пользуется **репрезентативная теория**, в соответствии с которой измерение есть процесс присваивания числовых выражений объекту измерения для его репрезентации (представления), т.е. для того, чтобы осмысленно выводить заключения о свойствах объекта. Это определение дано Кэмпбелом. Он делает акцент на целях измерения. Измерение осуществляется не ради самого измерения, а с тем, чтобы можно было извлечь пользу из его результатов.

По Стивенсу, измерение — это приписывание чисел вещам в соответствии с определенными правилами. Он акцентирует внимание на измерительных операциях. Теорию измерения, развиваемую им, можно было бы назвать **операциональной**.

Следует привести также определение формальной теории, которое вытекает из теории математических моделей **А. Тарского**. Измерить — значит установить однозначное (гомоморфное) отображение эмпирической реляционной структуры в числовую реляционную структуру. Реляционная структура — это множество объектов вместе со всеми отношениями и операциями на нем. В соответствии с этим определением, если объекты находятся в реальной действительности (в эмпирической реляционной структуре) в некоторых отношениях друг с другом (одинаковы, больше, меньше, лучше, хуже, являются суммой или разностью), то в этих же отношениях должны находиться числа, приписанные им в результате измерения (числовая реляционная структура). Это определение находится в русле репрезентативной теории.

Множества чисел, в которых проводится измерение, образуют измерительные шкалы. В концептуальном отношении Стивенсом выделено 4 основных типа шкал.

1) **Номинальная шкала**, шкала **наименований**, шкала **классификаций**. Объектам приписываются любые числа, которые играют роль простых имен и используются с целью различения объектов и их классов. Примеры: номера футболистов, числовые коды различных классификаторов. Основное правило такого измерения: не приписывать одно число объектам разных классов и разные числа объектам

одного класса. В этой шкале вводится только два отношения: «равно» и «не равно». В ней измеряются объекты, которые пока научились или которые достаточно только различать. Понятно, что в данном случае речь идет об измерении в очень слабом смысле. Результаты измерения X в этой шкале всегда можно изменить, подвергнув их взаимнооднозначному преобразованию f. Говорят, что математическая структура этой шкалы определяется преобразованием f, таким что $f' \neq 0$.

2) **Ординальная** или **порядковая, ранговая шкала**. В этой шкале измеряются объекты, которые одинаковы или предпочтительнее друг друга в каком-то смысле. Принимаются во внимание только три отношения, в которых могут находиться числа этой шкалы: «равно», «больше», «меньше». Математическая структура шкалы определяется монотонно возрастающим преобразованием f: $f' > 0$. Пример такой шкалы дает теория порядковой полезности.

3) **Интервальная шкала**. Шкала используется для измерения объектов, относительно которых можно говорить не только больше или меньше, но и на сколько больше или меньше. Т.е. в ней введено расстояние между объектами и, соответственно, определены единицы измерения, но нет пока нуля, и бессмысленнен вопрос о том, во сколько раз больше или меньше. Математическая структура шкалы: $f = aX + b$, где $a > 0$ (a — коэффициент изменения единицы измерения, b — «сдвиг» нуля). В этой шкале измеряются некоторые физические величины, например, температура. Если ночью по Цельсию было 5 градусов тепла, а днем — 10, то можно сказать, что днем теплее на 5 градусов, но утверждение, что днем в 2 раза теплее, чем ночью, бессмысленно. В шкале Фаренгейта или Кельвина данное отношение совсем другое.

4) **Шкала отношений**. В ней, по сравнению с предыдущей шкалой, введен ноль (естественное начало шкалы) и определено отношение «во сколько раз больше или меньше». Математическая структура шкалы: $f = aX$ (a — коэффициент изменения единицы измерения), $a > 0$. Это обычная шкала, в которой проводится большинство **метрических** измерений.

Первые два типа шкал **неметрические**, они используются в нефизическом измерении (в социологии, психологии, иногда в экономике), которое в этом случае называется обычно **шкалированием**. Метрическими являются шкалы двух последних типов. Экономические величины измерены, как правило, в шкале отношений.

Существуют различные виды измерений. С точки зрения дальнейшего изложения важно выделить два вида: **прямые** или **первичные**, которые в физических измерениях иногда называют **фундаментальными**, и **косвенные** или **производные**. Измерения 1-го вида сводятся к проведению эмпирических операций в непосредственном контакте с измеряемым объектом. Это — опрос, анкетирование, наблюдение, счет, считывание чисел со шкал измерительных приборов. Измерения 2-го

вида связаны с проведением вычислительных операций над первично измеренными величинами.

Таким образом, в измерении используются и эмпирические, и вычислительные операции. Некоторые теоретики измерения склонны минимизировать роль вычисления и отделить собственно измерение, как преимущественно эмпирическую операцию, от вычислений. Однако грань между этими двумя понятиями достаточно расплывчата, особенно при экономических измерениях.

1.6. Специфика экономических измерений

Специфические особенности экономических измерений можно свести в 5 групп:

1) Измеряться могут только операционально определенные величины. В экономике разработка операциональных определений величин — это сложный и неоднозначный исследовательский процесс теоретического характера. Теоретики постоянно дискутируют на темы измерения общих итогов экономического развития, экономической эффективности, производительности общественного труда, экономической динамики, инфляционных процессов, структурных сдвигов и т.д. Не выработано строгой и единой системы операциональных величин, однозначно представляющих эмпирическую экономическую систему. Одно из следствий такого положения, как уже говорилось, заключается в том, что каждому теоретическому понятию, как правило, соответствует несколько операциональных величин, отражающих различные точки зрения и используемых с разными целями.

Очень сильно различались системы статистического учета, сложившиеся в СССР и в мировой практике. В России к концу прошлого столетия в целом завершен переход на западные стандарты, но нельзя не видеть положительных моментов, имевшихся и в отечественной системе статистики. В мировой практике статистики к настоящему времени сложилась более или менее устойчивая, хотя и имеющая национальные особенности, система статистического отображения экономической действительности: Национальные счета на макроуровне, Бухгалтерский учет в фирмах. И эти вопросы не являются предметом активных дискуссий теоретиков. Но нет сомнений, что под воздействием накапливаемых изменений в общественной жизни «взрывы» таких дискуссий ожидают нас впереди.

Таким образом, экономические измерения, в отличие от многих физических, в очень большой степени обусловлены теоретическими моментами.

2) Специфику экономических измерений создают и те особенности экономики, которые обсуждались выше в связи с пониманием особенностей статистики как науки и вероятностной природой экономических явлений. Короткие ряды наблюдений и неэкспериментальный характер данных очень затрудняют процесс измерения и нередко ставят под сомнение научную значимость его результатов.

В процессе управляемого эксперимента можно изменить значение некоторой величины и определить, на что и каким образом она влияет, т.к. остальные величины-факторы остаются неизменными. Неэкспериментальные данные исключают возможность анализа «при прочих равных». В потоке наблюдений за «всеми сразу» величинами, как уже отмечалось, трудно уловить структуру взаимосвязей и измерить их интенсивность. Чисто эмпирически это, пожалуй, невозможно сделать. Это обстоятельство еще в бóльшей степени увеличивает нагрузку на теорию, «силу абстракции» исследователя. И оно не добавляет надежности результатам измерения.

3) В экономике не существует таких объектов и не изобретено таких «линеек», совмещение которых позволило бы путем считывания чисел со шкалы определить объем валового внутреннего продукта или темп инфляции. Экономические измерения почти всегда косвенные, производные. Экономические величины определяются путем расчета, исчисления, формула которого задается операциональным определением величины. Более того, первичные измерения, имеющие в физике фундаментальное значение, в экономике, как правило, экономического характера не имеют. Это — счет, физические измерения веса, объема, длины, первичная регистрация цен, тарифов и т.д. Экономический характер они приобретают лишь после своей **свертки** в экономические величины.

4) В естественных науках единицы измерения: килограмм, метр, джоуль, ватт и т.д. — четко и однозначно определены. Специфические единицы экономических измерений: цены, тарифы, ставки, единицы полезности — постоянно меняются. Важно даже не то, что они меняются во времени, а то, что их изменения зависят от объема и пропорций тех величин, которые они призваны измерять. Если в структуре производства или в потребительском наборе доля какого-то продукта уменьшается, то его цена или полезность, как правило, растет. И наоборот. Учет такого рода зависимостей и изменчивости единиц измерения — очень сложная проблема, совершенно неизвестная в физических измерениях.

5) В процессе измерения инструмент взаимодействует определенным образом с объектом измерения, вследствие чего положение этого объекта может измениться, и результатом измерения окажется не та величина, которая имела место до самого акта измерения. Пример: если попытаться в темной комнате на ощупь определить положение бильярдного шара на столе, то он обязательно сдвинется с места. Эта проблема так или иначе возникает в любых измерениях, но только в экономических и, вообще, социальных измерениях она принимает угрожающие масштабы.

Экономические величины складываются под воздействием определенной деятельности человека и каким-то характеризуют образом эту деятельность. Поэтому люди, как те, кто измеряет, так и те, чья деятельность измеряется, обязательно заинтересованы в результатах измерения. Взаимодействия в процессе измерения,

возникающие по этим причинам, могут приводить к огромным отличиям получаемых значений измеряемых величин от их действительных значений. В физических измерениях влияние этого субъективного фактора практически отсутствует.

1.7. Адекватность экономических измерений

Под адекватностью измерений обычно понимают степень соответствия измеренных значений действительным или истинным. Разность этих значений образует ошибку измерения. Теория ошибок, основанная на теории вероятностей и математической статистике, изучается в следующей части книги. Здесь рассматривается значение учета ошибок экономических измерений, причины этих ошибок и приводятся некоторые примеры.

Любые измерения, а экономические в особенности, содержат ошибки. Точные величины суть не более чем теоретические абстракции. Это происходит хотя бы в силу случайного характера величин. Исследователи располагают выборочными значениями величин и могут лишь приблизительно судить об их истинных значениях в генеральной совокупности. Измерения без указания ошибки достаточно бессмысленны. Фразу: «Национальный доход равен 10 600 млрд. руб.» — если она не содержит сведений о точности или не подразумевает таких знаний у читателя (например, судя по количеству приведенных значащих цифр, ошибка составляет ± 50 млрд. руб.), — всегда можно продолжить: «или любой другой величине». К сожалению, понимание этого элементарного факта в экономике пока еще не достигнуто. Например, можно встретить такие статистические публикации, в которых численность населения бывшего СССР дается с точностью до одного человека. Кстати, «точные» науки знают меру своей неточности, и результаты физических измерений обычно даются с указанием возможной ошибки.

Ошибки обычно подразделяют на **случайные** и **систематические**. Для экономики можно ввести еще один класс ошибок: **тенденциозные**. Случайные ошибки — предмет строгой теории (см. гл. 5), здесь внимание сосредоточено на систематических и тенденциозных ошибках.

В чем причины этих ошибок экономических измерений?

В предыдущем разделе приводились 5 особенностей экономических измерений. Каждая из них вносит в ошибку свою лепту, и немалую, сверх «обычной» ошибки физических измерений.

1) **Ошибки теории**. Операциональные определения экономических величин — продукт теории. И если теория неверна, то, как бы точно в физическом смысле не проводились измерения исходных ингредиентов, какими бы совершенными вычислительными инструментами не пользовались, ошибка — возможно очень большая — обязательно будет присутствовать в результатах измерения.

О величине этих ошибок в практике нашей статистики можно судить лишь косвенно. Если сравнивать показатели совокупного производства, которые использовались в СССР и используемые в мировой практике, то можно отметить две основные компоненты ошибки. В мировой практике используются показатели типа конечной продукции, в советской статистике — типа валовой продукции, которые сильно искажаются повторным счетом и другими «накрутками», содержащимися в промежуточном продукте. И второе: в советской статистике расчет этих показателей проводился только по так называемой материальной сфере. Бóльшая часть продукта, создаваемого в нематериальной сфере, не попадала в итоги.

2) **Ошибки инструмента**, в данном случае — принятых статистических процедур расчета. Наибольшим дефектом в советской статистике страдали процедуры оценки динамики цен. Они скрывали реальные темпы инфляции.

Пример. На практике применяется два метода расчета национального дохода или валового внутреннего продукта (ВВП): потребительский — для определения использованного национального дохода как суммы фактических объемов накопления и непроизводственного потребления, и производственный — для расчета произведенного национального дохода как суммы чистой продукции (добавленной стоимости) по отраслям производства. Эти показатели жестко связаны между собой: их разница равна величине потерь и сальдо экспорта-импорта. Такая зависимость выдерживалась в государственной статистике только в текущих ценах. В сопоставимых ценах произведенный национальный доход устойчиво обгонял использованный ежегодно на несколько миллиардов рублей. Если начать отсчет с начала 70-х годов, то к концу 80-х разрыв между произведенным и используемым национальным доходом достигал $1/5$ последнего. Эти $100 - 150$ млрд. руб. разрыва — одна из оценок ошибки расчета национального дохода в сопоставимых ценах.

В настоящее время в государственной статистике возникла в некотором смысле обратная проблема. ВВП, рассчитанный по производству, оказывается заметно меньшей величиной, чем рассчитанный по использованию. Причем разрыв также достигал в отдельных случаях $1/5$ ВВП. Это происходит потому, что часть продукции производится в так называемой «теневой» экономике и не находит отражения в официальной статистике. Использование же продукции учитывается в более полных объемах.

Страдали и страдают несовершенством и другие статистические процедуры. Еще один пример — из области международных сопоставлений динамики итоговых показателей развития. Если известны темпы роста национального дохода, например, СССР и США, то можно легко установить, как менялось соотношение этих показателей и насколько успешно СССР «догонял» США. Независимо от этого в советской статистике проводились прямые сопоставления национальных доходов, показывающие, какую часть национального дохода США составляет национальный доход СССР. Долгое время оставался незамеченным факт серьезного несоответ-

ствия результатов этих двух расчетов: по данным динамики национального дохода СССР догонял США гораздо быстрее, чем по данным прямых сопоставлений. Можно не сомневаться в том, что искажены были и те и другие данные, но динамика национального дохода была искажена в большей степени.

3) **Тенденциозные ошибки**. Являются следствием субъективного фактора в процессе измерения. Искажение и сокрытие информации — элемент рациональной стратегии экономического поведения. Это общепризнанный факт, но в СССР, в силу значительной идеологической нагрузки на статистику, искажение информации, особенно итоговой, достигало удручающе больших размеров. По оценкам **Г.И. Ханина**, реально национальный доход за период с начала 1-й пятилетки (конца 20-х годов прошлого столетия) до начала 80-х годов прошлого века вырос не в 90 раз, как по официальной статистике, а всего в 7—8 (что тоже, кстати, очень неплохо).

В современной официальной статистике в России такие ошибки также имеют место. Но если во времена СССР совокупные объемы производства преувеличивались, то теперь они занижаются. Это — результат «бартеризации» экономики, выведения хозяйственной деятельности из-под налогообложения. Косвенным подтверждением этих фактов является то, что при резком сокращении общих (официальных) объемов производства в последнем десятилетии прошлого века объемы потребления электроэнергии, топлива, тепла, объемы грузоперевозок уменьшились гораздо в меньшей степени.

4) **Ошибки единиц измерения**. Имеется серьезное отличие понимания точности в физическом и экономическом измерении. Даже если измерения точны в физическом смысле, т.е. правильно взвешены и измерены первичные величины, использована бездефектная теория для свертки этих величин, ошибки в экономическом смысле могут присутствовать и, как правило, присутствуют. Дело в том, что практически всегда искажены по сравнению со своими истинными значениями наблюдаемые экономические единицы измерения: цены, тарифы и т.д. Особенно велик масштаб этих деформаций был в централизованной экономике. Влияние их на результаты измерения и далее на процессы принятия решений в СССР было огромным. Это стало особенно очевидным в конце горбачевской «перестройки», когда разные республики и территории бывшего СССР начали выдвигать взаимные претензии, рассуждая на тему о том, кому, кто и сколько должен. Если взять Западную Сибирь, то по официальным данным на конец 80-х годов XX века ее сальдо вывоза-ввоза было хоть и положительно, но очень невелико. Расчеты же в равновесных ценах давали цифру плюс 15—20 млрд. руб., а в ценах мирового рынка — плюс 25—30.

Доля ошибок такого рода была велика и в реформируемой России, когда ценовые пропорции были неустойчивы и быстро менялись, значительно рос общий

уровень цен. Сложной и не решаемой однозначно оказывается проблема «очистки» итоговых за год показателей от факторов инфляции.

1.8. Типы величин, связи между ними

Экономические величины могут быть двух типов: **экстенсивные**, или **объемные**, и **интенсивные**, или **относительные**. Первые обладают единицами измерения, и их можно складывать, т.е. агрегирование проводится обычным сложением; вторые не имеют единиц измерения, а могут обладать только определенной размерностью, и они не аддитивны, их агрегирование проводится путем расчета средневзвешенных величин.

Экстенсивные величины, в свою очередь, могут иметь тип **запаса** или **потока**. Величины типа запаса регистрируются на конкретный момент времени и имеют элементарные единицы измерения: рубль, штука, тонна, метр и т.д. Примеры: основные фонды, материальные запасы, население, трудовые ресурсы. Величины типа потока определяются только за конкретный период времени и имеют размерность «объем в единицу времени»: рубль в год, штука в час и т.д. К этим величинам относятся выпуск продукции, потребление, затраты, инвестиции, доходы и т.д.

Величины запаса и потока жестко связаны между собой:

$$S_b[v] + P_i[v/t]t = S_e[v] + P_o[v/t]t,$$

где S_b и S_e — запасы на начало и конец периода (v — единица измерения), P_i и P_o — потоки по увеличению и уменьшению запаса (t — период).

Это соотношение лежит в основе большинства балансовых статистических таблиц. Например, в балансе движения основных фондов по полной стоимости S_b и S_e — основные фонды на начало и конец года, P_i и P_o — ввод и выбытие основных фондов; в балансе (межотраслевом) производства и потребления продукции S_b и S_e — материальные запасы на начало и конец года, P_i — производство и импорт продукции, P_o — текущее потребление (производственное и непроизводственное), инвестиции и экспорт.

Интенсивные величины являются отношениями экстенсивных или интенсивных величин. Они могут иметь разное содержание, разную размерность или быть безразмерными.

Примеры интенсивных величин как отношений объемных величин:

- в классе P/S: производительность труда, фондоотдача, коэффициенты рождаемости и смертности населения;
- в классе S/P: трудо- и фондоемкость производства;

– в классе S/S: фондовооруженность труда;

– в классе P/P: материало- и капиталоемкость производства, коэффициенты перевода капитальных вложений во ввод основных фондов.

Размерность этих величин определяется формулой их расчета. Интенсивные величины, получаемые отношением величин одного качества (экстенсивных или интенсивных), размерности не имеют. К ним относятся темпы роста и прироста, коэффициенты пространственного сравнения, показатели отраслевой и территориальной структуры. Такие безразмерные относительные величины могут даваться в процентах или промиллях (если a — относительная величина, то $a \cdot 100\,\%_o$ — ее выражение в процентах, $a \cdot 1000\,\%_{oo}$ — в промиллях).

Если две величины y и x связаны друг с другом, то одним из показателей этой связи является их отношение: y/x — средний коэффициент связи (например, трудо-, материало-, фондоемкость производства).

Иногда пользуются приростным коэффициентом (например, капиталоемкость производства как приростной коэффициент фондоемкости): $\Delta y/\Delta x$, где Δy и Δx — приросты величин y и x за определенный период времени.

Если величины y и x связаны гладкой непрерывной функцией, то непрерывным (моментным) приростным коэффициентом является производная dy/dx.

В этом же ряду находится так называемый **коэффициент эластичности**, показывающий отношение относительных приростов:

$$\left(\frac{\Delta y}{y}\right) : \left(\frac{\Delta x}{x}\right) = \frac{x\Delta y}{y\Delta x} = \frac{\Delta y}{\Delta x} \cdot \frac{x}{y}.$$

Непрерывным (моментным) коэффициентом эластичности является показатель степени при степенной зависимости y от x:

$$y = ax^{\alpha}, \text{ т.к. } \frac{dy}{dx} = a\alpha x^{\alpha-1} = \alpha\frac{y}{x}, \text{ откуда } \alpha = \frac{dy}{dx} \cdot \frac{x}{y}.$$

При наличии такой зависимости y от x моментный коэффициент эластичности рассчитывается как $\dfrac{\ln(y/a)}{\ln x}$.

Это — примеры относительных величин, имеющих размерность. Далее приводятся примеры безразмерных относительных величин.

Пусть $y = \sum\limits_i y_i$. Например, y — совокупный объем производства на определенной территории, y_i — объем производства (в ценностном выражении) в i-й отрасли; или y — общий объем производства какого-то продукта в совокупности

регионов, y_i — объем производства продукта в i-м регионе. Тогда y_i/y — **коэффициент структуры**, отраслевой в первом случае, территориальной (региональной) во втором случае.

Если y_i и y_j — значения некоторого признака (объемного или относительного) двух объектов (i-го и j-го), например, двух отраслей или двух регионов, то y_i/y_j — **коэффициент сравнения**, межотраслевого в первом случае, пространственного (межрегионального) во втором случае.

Пусть y_t — значение величины (объемной или относительной) в момент времени t. Для измерения динамики этой величины используются следующие показатели:

$\Delta y_t = y_{t+1} - y_t$ (или $\Delta y_{t+1} = y_{t+1} - y_t$) — **абсолютный прирост**,

y_{t+1}/y_t — **темп роста**,

$\Delta y_t/y_t = y_{t+1}/y_t - 1$ — **темп прироста**.

В случае, если динамика y задана гладкой непрерывной функцией $y(t)$, то непрерывным темпом прироста в момент времени (моментным темпом прироста) является $\dfrac{d\ln y(t)}{dt}$, поскольку $\dfrac{d\ln y}{dy} = \dfrac{1}{y}$, а непрерывным (моментным) абсолютным приростом выступает $\dfrac{dy(t)}{dt}$. Последнее следует пояснить (почему $\dfrac{dy(t)}{dt}$ выступает моментным абсолютным приростом в единицу времени). Пусть единичный период времени $[t,\ t+1]$ разбит на n равных подпериодов, и в каждом из них одинаков абсолютный прирост. Тогда абсолютный прирост в целом за единичный период равен

$$\frac{y\left(t+\dfrac{1}{n}\right) - y\,(t)}{\dfrac{1}{n}},$$

и предел его при $n \to \infty$, по определению производной, как раз и равен $\dfrac{dy(t)}{dt}$.

Непрерывным (моментным) темпом роста является $e^{d\ln y(t)/dt}$ (e — основание натурального логарифма). Действительно, пусть опять же единичный период времени $[t,\ t+1]$ разбит на n равных подпериодов, и темпы роста во всех них одинаковые. Тогда темп роста за этот период (единицу времени) окажется равным

$$\left(\frac{y\left(t+\frac{1}{n}\right)}{y\,(t)}\right)^n,$$

Таблица 1.1

	За период (единичный)		Моментный
Темп	дискретн.	непрерывный	
Роста	$\frac{y_{t+1}}{y_t}$	$\exp\left(\int\limits_t^{t+1} \frac{d\ln y\,(t')}{dt'}\,dt'\right) = \frac{y\,(t+1)}{y\,(t)}$	$\exp\left(\frac{d\ln y\,(t)}{dt}\right)$
Прироста	$\frac{y_{t+1}}{y_t} - 1$	$\int\limits_t^{t+1} \frac{d\ln y\,(t')}{dt'}\,dt' = \ln\frac{y\,(t+1)}{y\,(t)}$	$\frac{d\ln y\,(t)}{dt} = \frac{1}{y}\frac{dy\,(t)}{dt}$

и переходом к пределу при $n \to \infty$ будет получено искомое выражение для моментного темпа роста. Проще найти предел не этой величины, а ее логарифма. То есть

$$\lim_{n\to\infty} \frac{\ln y\left(t+\frac{1}{n}\right) - \ln y\,(t)}{\frac{1}{n}}.$$

По определению производной, это есть $\dfrac{d\ln y\,(t)}{dt}$, т.е. моментный темп прироста. Следовательно, как и было указано, моментным темпом роста является e в степени $\dfrac{d\ln y\,(t)}{dt}$.

Непрерывный темп роста за период от t до $t+1$ определяется следующим образом:

$$e^{\int\limits_t^{t+1} \frac{d\ln y(t')}{dt'}dt'}.$$

В этом легко убедиться, если взять интеграл, стоящий в показателе:

$$\int\limits_t^{t+1} \frac{d\ln y\,(t')}{dt'}dt' = \ln y\,(t')\Big|_t^{t+1} = \ln\frac{y\,(t+1)}{y\,(t)},$$

и подставить результат (его можно назвать непрерывным темпом прироста за единичный период времени) в исходное выражение непрерывного темпа роста за период:

$$e^{\ln\frac{y\,(t+1)}{y\,(t)}} = \frac{y\,(t+1)}{y\,(t)}.$$

Построенные относительные показатели динамики сведены в таблице 1.1.

Относительные величины, с точки зрения их измерения, являются производными, т.е. их размер определяется путем расчета. Такой характер относительные величины имеют и в других предметных науках. Но в экономике существуют интенсивные величины особого типа, имеющие первичный или фундаментальный характер. Это экономические единицы (измерения): цены продукции, тарифы за услуги, ставки заработной платы, ставки процента, дивиденды, а также особые управляющие параметры-нормативы, например, ставки налогов и дотаций. Эти величины имеют разную размерность или безразмерны, но регистрируются они как величины запаса — на определенные моменты времени.

1.9. Статистические совокупности и группировки

Статистической совокупностью, или просто совокупностью, называют множество объектов, однородное в определенном смысле. Обычно предполагается, что признаки объектов, входящих в совокупность, измерены (информация имеется) или по крайней мере измеримы (информация может быть получена). Полное множество величин-признаков или показателей-наблюдений было обозначено выше как $\{x_{tij}\}$. Совокупность объектов — это его подмножество по i.

Об однородности совокупности можно говорить в качественном и количественном смысле.

Пусть J_i — множество признаков, которые характеризуют i-й объект.

Совокупность **однородна качественно**, если эти множества для всех входящих в нее объектов идентичны или практически идентичны. Такие совокупности образуют, например, сообщества людей, каждого из которых характеризуют имя, дата и место рождения, пол, возраст, вес, цвет глаз, уровень образования, профессия, место проживания, доход и т.д. В то же время понятно, что, чем большие сообщества людей рассматриваются, тем менее однородны они в этом смысле. Так, например, совокупность, включающая европейцев и австралийских аборигенов, не вполне однородна, поскольку набор признаков для последних включает такие характеристики, которые бессмысленны для первых (например, умение бросать бумеранг), и наоборот.

Совокупность промышленных предприятий качественно достаточно однородна. Но более однородны совокупности предприятий конкретных отраслей, поскольку каждая отрасль имеет свою специфику в наборе всех возможных признаков.

Чем меньше общее пересечение множеств J_i, тем менее однородна в качественном смысле совокупность i-х объектов. Объекты, общее пересечение множеств признаков которых мало, редко образуют совокупности. Так, достаточно бессмыс-

ленна совокупность людей и промышленных предприятий, хотя все они имеют имя, дату и место «рождения», возраст.

Допустимая степень неоднородности совокупности зависит, в конечном счете, от целей исследования. Если, например, изучаются различия средней продолжительности жизни различных представителей животного мира, то в исследуемую совокупность включают и людей, и лошадей, и слонов, и мышей.

Количественная однородность зависит от степени вариации значений признаков по совокупности. Чем выше эта вариация, тем менее однородна совокупность в этом смысле. В разных фрагментах количественно неоднородных совокупностей могут различаться параметры зависимостей между величинами-признаками. Такие совокупности иногда также называют качественно неоднородными. Для них невозможно построить единой количественной модели причинно-следственных связей. Так, например, люди с низким уровнем дохода увеличивают спрос на некоторые товары при снижении своего дохода (малоценные товары) или/и при росте цен на эти товары (товары «Гиффена»). Люди с высоким уровнем дохода реагируют на такие изменения обычным образом — снижают спрос.

Однородные совокупности обычно имеют простое и естественное название: «люди» или «население», «промышленные предприятия». Выделяются эти совокупности с целью изучения, соответственно, человеческого сообщества, промышленности и т.д.

Массив информации по совокупности часто называют **матрицей наблюдений**. Ее строкам соответствуют объекты и/или время, т.е. **наблюдения**, столбцам — величины-признаки или переменные. Обозначают эту матрицу через X, ее элементы — через x_{ij}, где i — индекс наблюдения, j — индекс переменной-признака.

В конкретном исследовании все множество признаков делится на 2 части: **факторные признаки**, или **независимые факторы**, — экзогенные величины и **результирующие (результативные) признаки**, или **изучаемые переменные**, — эндогенные величины. Целью исследования обычно является определение зависимости результирующих признаков от факторных. При использовании развитых методов анализа предполагается, что одни результирующие признаки могут зависеть не только от факторных, но и от других результирующих признаков.

В случае, если факторных признаков несколько, используют методы **регрессионного анализа**, если наблюдениями являются моменты времени, то применяются методы **анализа временных рядов**, если наблюдения даны и по временным моментам, и по территориально распределенным объектам, то целесообразно применить методы **анализа панельных данных**.

Если наблюдений слишком много и/или совокупность недостаточно однородна, а также для изучения внутренней структуры совокупности или при применении

особых методов анализа связи, предварительно проводится группировка совокупности. **Группировка** — деление совокупности на **группы** по некоторым признакам.

Наиболее естественно проводится группировка по **качественным признакам**. Такие признаки измеряются обычно в шкале наименований или в порядковой шкале. Например, признак «пол»: 1 — мужской, 2 — женский (или -1 и 1, 0 и 1, 1 и 0 и т.д.); «академическая группа»: 1 — студент 1-й группы, 2 — студент 2-й группы и т.д. (это — примеры использования шкалы наименований); «образование»: 1 — отсутствует, 2 — начальное, 3 — среднее, 4 — высшее (номинальная шкала с элементами порядковой); оценка, полученная на экзамене: 1 — неудовлетворительно, 2 — удовлетворительно, 3 — хорошо, 4 — отлично (порядковая шкала с элементами интервальной).

Качественный признак принимает определенное количество уровней (например: «пол» — 2 уровня, «образование» — 4 уровня), каждому из которых присваивается некоторое целое число. Перестановка строк матрицы наблюдений по возрастанию или убыванию (если шкала данного признака порядковая, то обычно — по возрастанию) чисел, стоящих в столбце данного фактора, приводит к группировке совокупности по этому фактору. В результате строки матрицы, соответствующие наблюдениям-объектам с одинаковым уровнем данного качественного фактора, оказываются «рядом» и образуют группу.

Группировка по количественному (непрерывному или дискретному) признаку производится аналогичным образом, но после переизмерения этого признака в порядковой (или интервальной) шкале. Для этого проводятся следующие операции.

Пусть x_{ij}, $i = 1, \ldots, N$ — значения j-го количественного признака в матрице N наблюдений, по которому проводится группировка — деление совокупности на k_j групп. Весь интервал значений этого признака $[z_{0j},\ z_{k_jj}]$, где $z_{0j} \leqslant \min\limits_i x_{ij}$, а $z_{k_jj} \geqslant \max\limits_i x_{ij}$, делится на k_j полуинтервалов $[z_{0j},\ z_{1j}]$, $(z_{i_j-1,\,j},\ z_{i_jj}]$, $i_j = 1, \ldots, k_j$. Первый из них закрыт с обеих сторон, остальные закрыты справа и открыты слева. Количество и размеры полуинтервалов определяются целями исследования. Но существуют некоторые рекомендации. Количество полуинтервалов не должно быть слишком малым, иначе группировка окажется малоинформативной. Их не должно быть и слишком много, так, чтобы большинство из них были не «пустыми», т.е. чтобы в них «попадали» хотя бы некоторые значения количественного признака. Часто размеры полуинтервалов принимаются одинаковыми, но это не обязательно.

Теперь j-й столбец матрицы наблюдений замещается столбцом рангов наблюдений по j-му признаку (рангов j-го признака), которые находятся по следующему правилу: i-му наблюдению присваивается ранг i_j, если x_{ij} принадлежит i_j-му полуинтервалу, т.е. если $z_{i_j-1,\,j} < x_{ij} \leqslant z_{i_jj}$ (если $i_j = 1$, условие принадлежности имеет другую форму: $z_{0j} \leqslant x_{ij} \leqslant z_{1j}$). Таким образом, если значение наблюдения

попадает точно на границу двух полуинтервалов (что достаточно вероятно, например, при дискретном характере количественного признака), то в качестве его ранга принимается номер нижнего полуинтервала. В результате данный признак оказывается измеренным (пере-измеренным) в порядковой (ранговой) шкале с элементами интервальной шкалы, или — при одинаковых размерах полуинтервалов — в интервальной шкале. В случае, если исходные значения данного признака потребуются в дальнейшем анализе, столбец рангов не замещает столбец наблюдений за данным признаком, а добавляется в матрицу наблюдений.

Сама группировка осуществляется также перестановкой строк матрицы наблюдений по возрастанию ранга данного признака. В результате i_j-ю группу образуют наблюдения-объекты, имеющие i_j-й ранг, а группы в матрице наблюдений располагаются по возрастанию ранга от 1 до k_j.

Группы, полученные в результате группировки по одному признаку, могут быть разбиты на подгруппы по какому-нибудь другому признаку. Процесс деления совокупности на все более дробные подгруппы по 3-му, 4-му и т.д. признаку может быть продолжен нужное количество раз — в соответствии с целями конкретного исследования. Перестановка строк матрицы наблюдений при группировке по каждому последующему признаку осуществляется в пределах ранее выделенных групп. Некоторые пакеты прикладных программ (электронные таблицы, базы данных) имеют специальную операцию, называемую **сортировкой**. Эта операция переставляет строки матрицы наблюдений по возрастанию (или убыванию) значений ранга (уровня) сначала 1-го, потом 2-го, 3-го и т.д. указанного для этой операции признака. В этом смысле термины группировка и сортировка эквивалентны.

Признаки, по которым группируются объекты совокупности, называются **группирующими**. Если таких признаков больше одного, группировка называется **множественной**, в противном случае — **простой**.

Пусть группирующими являются первые n признаков $j = 1, \ldots, n$, и j-й признак может принимать k_j уровней (может иметь ранги от 1 до k_j). По этим признакам совокупность в конечном итоге будет разбита на K групп, где $K = \prod_{j=1}^{n} k_j$. Это — так называемые **конечные** или **заключающие группы**. Последовательность группирующих признаков определяется целями проводимого исследования, «важностью» признаков. Чем ближе признак к концу общего списка группирующих признаков, тем более **младшим** он считается. Однако с формальной точки зрения последовательность этих признаков не важна, от нее не зависит характер группировки, с ее изменением меняется лишь последовательность конечных групп в матрице наблюдений.

Общее число полученных групп существенно больше количества конечных групп. Каждый j-й признак по отдельности разбивает совокупность на k_j групп, вместе с признаком j' — на $k_j k_{j'}$ групп, вместе с признаком j'' — на $k_j k_{j'} k_{j''}$

групп и т.д. Поэтому, не сложно сообразить, общее число групп, включая саму совокупность, равно $\prod_j (1 + k_j)$.

Действительно:

$$\prod_j (1 + k_j) = 1 + k_1 + k_2 + \cdots + k_1k_2 + k_1k_3 + \cdots + k_1k_2k_3 + \cdots + k_1k_2 \ldots k_n,$$

— слагаемые правой части показывают количества групп, выделяемых всеми возможными сочетаниями группирующих признаков.

Конечные группы можно назвать также группами высшего, в данном случае n-го порядка, имея в виду, что они получены группировкой по всем n признакам. Любое подмножество группирующих признаков, включающее n' элементов, где $0 < n' < n$ делит совокупность на «промежуточные» группы, которые можно назвать группами порядка n'. Каждая такая группа является результатом объединения определенных групп более высокого, в частности, высшего порядка. Конкретное подмножество группирующих признаков, состоящее из n' элементов, образует конкретный класс групп порядка n'. Всего таких классов $C_n^{n'}$ (это — число сочетаний из n по n', равное, как известно, $\frac{n!}{n'!(n-n')!}$). Группой нулевого порядка является исходная совокупность. Общее число всех групп от нулевого до высшего порядка, как отмечено выше, равно $\prod_j (1 + k_j)$.

Дальнейшее изложение материала о группировках будет иллюстрироваться примером, в котором при $n = 2$ первым группирующим признаком является «студенческая группа» с $k_1 = 4$ (т.е. имеется 4 студенческие группы), вторым группирующим признаком — «пол» с $k_2 = 2$, а при $n = 3$ добавляется третий группирующий признак — «оценка», полученная на экзамене, с $k_3 = 4$. В этом примере (при $n = 3$) имеется 32 конечные группы (третьего порядка), образующие класс с именем (все элементы которого имеют имя) «студенты». Существуют 3 класса групп 2-го порядка ($C_3^2 = 3$). Класс А1, образуемый подмножеством группирующих признаков (12), включает 8 групп с именем «юноши или девушки такой-то студенческой группы», А2 — образуемый подмножеством (13), включает 16 групп с именем «студенты такой-то группы, получившие такую-то оценку», и А3 — образуемый подмножеством (23), включает 8 групп с именем «юноши или девушки, получившие такую-то оценку на экзамене». Классов групп первого порядка имеется также 3 ($C_3^1 = 3$). Класс Б1, образуемый подмножеством (1), включающий 4 группы с именем «такая-то студенческая группа», Б2 — подмножеством (2), включающий 2 группы с именем «юноши или девушки», и Б3 — подмножеством (3), включающий 4 группы с именем «студенты, получившие такую-то оценку на экзамене».

Каждой конечной группе соответствует конкретное значение так называемого **мультииндекса** I порядка n (состоящего из n элементов), который имеет следующую структуру: $i_1 i_2 \dots i_n$ ($I = i_1 i_2 \dots i_n$). Для всех наблюдений конечной группы, имеющей такое значение мультииндекса, первый группирующий признак находится на уровне (имеет ранг) i_1, второй группирующий признак — на уровне i_2 и т.д., последний, n-й — на уровне i_n. Линейная последовательность (последовательность в списке) значений мультииндекса совпадает с последовательностью конечных групп в матрице наблюдений. На первом месте стоит значение I_1, все элементы которого равны единице (конечная группа, для всех наблюдений которой все группирующие признаки находятся на первом уровне). Далее работает правило: быстрее меняются элементы мультииндекса, соответствующие более младшим группирующим признакам. Так, в иллюстрационном примере при $n = 2$ последовательность значений мультииндекса такова: **11**, **12**, **21**, **22**, **31**, **32**, **41**, **42**. Последним значением мультииндекса является $I_K = k_1 k_2 \dots k_n$. Поскольку последовательность значений мультииндекса однозначно определена, $\sum\limits_{I'=I_1}^{I}$ означает суммирование по всем значениям мультииндекса от I_1 до I.

В некоторых случаях мультииндексы групп называют **кодами** групп. После завершения группировки столбцы группирующих признаков часто исключаются из матрицы наблюдений, т.к. содержащаяся в них информация сохраняется в мультииндексах-кодах.

Если из «полного» мультииндекса порядка n вычеркнуть некоторые элементы-признаки, то получается мультииндекс более низкого порядка n', который именует определенную группу порядка n'. Операция вычеркивания проводится заменой в исходном мультииндексе вычеркиваемых элементов символом «$*$» (иногда используется символ точки или какой-нибудь другой). Это необходимо для того, чтобы сохранить информацию о том, какие именно признаки вычеркнуты из мультииндекса. В иллюстративном примере группы класса А1 имеют мультииндекс со звездочкой на третьем месте, а класса Б2 — на первом и третьем местах. Для того чтобы подчеркнуть принадлежность мультииндекса I к конечным группам, мультииндексы групп более низкого порядка можно обозначать $I(*)$.

Теперь вводится еще один специальный мультииндекс J, который в «полном формате» (при порядке n) представляет собой последовательность целых чисел от 1 до n и обозначается G. В этом мультииндексе J все элементы, которые заменены звездочкой в мультииндексе $I(*)$, также заменены на звездочку. Пусть J_* — последовательность из n звездочек (все элементы заменены на «$*$»). Для индексации групп можно использовать пару индексов I, J (в этом случае к I излишне приписывать $(*)$). В этом случае из этих мультииндексов можно в действительности вычеркнуть все звездочки, т.к. информация о вычеркнутых признаках сохраняется в J. Так, например, группа «студенты второй группы, получившие «отлично»

на экзамене» именуется мультииндексом $I(*)$, равным $\mathbf{2{*}4}$, или парой мультииндексов $I,\ J$ — $\mathbf{24},\ \mathbf{13}$. Второй способ удобен, когда речь идет о группах низких порядков. В данном изложении будет использоваться первый способ индексации.

Группа $I(*)$ (с мультииндексом $I(*)$) является объединением конечных групп с такими значениями мультииндекса I, что: а) все те их элементы, которые соответствуют элементам, не вычеркнутыми из $I(*)$, совпадают с ними; б) все элементы, соответствующие вычеркнутым из $I(*)$ элементам, пробегают все свои значения. Такую операцию объединения естественно обозначить $\bigcup\limits_{I(*)}$. Так, например, группа $\mathbf{1{*}4}$ является объединением групп $\mathbf{114}$ и $\mathbf{124}$, а группа $\mathbf{42}*$ — объединением групп $\mathbf{421}$, $\mathbf{422}$, $\mathbf{423}$ и $\mathbf{424}$. Если $I(*) = J_*$, объединяются все конечные группы и образуется исходная совокупность, а сам $I(*)$, равный J_*, формально выступает мультииндексом всей совокупности.

Через J обозначается класс групп, образованных подмножеством признаков, не замененных в J звездочками. Так, продолжая пример, А2 является классом $\mathbf{1{*}3}$, а Б2 — классом $*\mathbf{2}*$. Количество групп в J-классе K^J является произведением k_j с такими j, которые не заменены звездочками в J; такую операцию произведения естественно обозначить $\prod\limits_{J}$. При $J = G$ оно равно количеству конечных групп K, а при $J = J_*$ принимается равным 1 (исходная совокупность — одна).

Пусть N_I — число наблюдений-объектов в конечной группе I. Тогда число наблюдений в группе более низкого порядка $I(*)$, которое можно обозначить $N_{I(*)}$, равно $\sum\limits_{I(*)} N_I$, где операция $\sum\limits_{I(*)}$ выполняется аналогично операции $\bigcup\limits_{I(*)}$. Эти числа называются **групповыми численностями**, все они больше либо равны нулю, в случае равенства нулю соответствующая группа пуста. Если $I(*) = K^*$, то $N_{I(*)} = N$.

Каждому наблюдению-объекту можно также поставить в соответствие мультииндекс порядка $n+1$, имеющий структуру Ii_I, где I мультииндекс конечной группы, к которой принадлежит данное наблюдение, а i_I — номер данного наблюдения в этой группе. Так, в иллюстрационном примере $\mathbf{3125}$ — мультииндекс пятой девушки в списке девушек третьей группы, получивших на экзамене «удовлетворительно». Исходный **линейный** индекс i наблюдения с мультииндексом Ii_I равен $\sum\limits_{I'=I_1}^{I_-} N_{I'} + i_I$, где I_- — значение мультииндекса конечной группы, предшествующее I в последовательности всех значений мультииндекса. Так, в примере значение мультииндекса $\mathbf{423}$ предшествует значению $\mathbf{424}$, а значение $\mathbf{314}$ — значению $\mathbf{321}$.

Мультииндекс, в котором $(n+1)$-й элемент замещен звездочкой, обозначает все множество наблюдений группы. Так, $\mathbf{1{*}3}*$ мультииндекс списка всех студентов первой группы, получивших на экзамене «хорошо».

Результаты группировки применяются для решения задач 3-х типов.

1) Используя информацию о групповых численностях, анализируют **распределение частот** или **эмпирических вероятностей признаков,** теоретическим обобщением которых являются **функции распределения вероятностей** и **плотности вероятностей случайных величин**. Потому такие распределения частот иногда называют **эмпирическими** функциями распределения вероятностей и плотностей вероятностей признаков. Если группировка является множественной, то говорят о **совместном распределении признаков** (группирующих), которое может использоваться в анализе зависимостей между этими признаками. В таком случае группирующие признаки делятся на факторные и результирующие. Так, в иллюстрационном примере можно изучать зависимость оценки, полученной на экзамене, от факторов «студенческая группа» и «пол». Приемы построения эмпирических распределений вероятностей и простейшие методы анализа связей с помощью совместных распределений изучаются в этой части книги.

При решении задач этого типа группирующие признаки являются, как правило, количественными.

2) Все группирующие признаки выступают факторными, и исследуется их влияние на некоторые другие — результирующие признаки $x_j,\ j > n$. В этом случае группирующие (факторные) признаки являются обычно качественными, и используются **методы дисперсионного анализа**, элементарные сведения о котором даются в главе 4 этой части (более основательно эти методы рассматриваются в III-й части книги). В иллюстрационном примере при $n = 2$ признак «оценка» не входит в число группирующих, и если взять его в качестве результирующего, то можно также исследовать влияние факторов «студенческая группа» и «пол» на оценку. В пункте 1) говорилось о других методах изучения этого влияния.

3) Анализируются зависимости между признаками внутри выделенных групп и/или между группами, т.е. внутригрупповые и/или межгрупповые связи. Во втором случае в анализе используются средние значения признаков в группах. В обоих случаях факторные и результирующие признаки не входят во множество группирующих признаков. Методы регрессионного анализа, используемые для анализа связей, и методы проверки гипотез о существенности различий параметров связей между различными группами изучаются во II-й и III-й частях книги. В главе 4 настоящей части даются общие сведения о некоторых из этих методов.

Особенность рассмотренных методов группировки заключается в том, что деление на группы всякий раз проводится по значениям строго одного признака. В одну группу попадают наблюдения-объекты с близкими (или — для качественных признаков — совпадающими) значениями признака. Каждый последующий признак лишь «дробит» ранее выделенные группы. Между тем, существуют методы выделения групп сразу по нескольким признакам. При таких группировках используются

различные меры близости векторов. Наблюдения i и i' попадают в одну группу, если по выбранной мере близки вектора x_{ij} и $x_{i'j}$, $j = 1, \ldots, n$. Методы таких группировок используются в **кластерном анализе** (кластер — класс). Существуют и обратные задачи, когда новое наблюдение-объект надо отнести к какому-то известному классу. Такие задачи решаются **методами распознавания образов**, они возникают, например, при машинном сканировании текстов или машинном восприятии человеческой речи.

Признаки также образуют совокупности разной степени однородности, понимаемой в этом случае только в качественном смысле. Как и в анализе совокупности объектов можно обозначить через I_j множество объектов, обладающих j-м признаком. Степень однородности совокупностей признаков тем выше, чем больше общее пересечение этих множеств для признаков, входящих в совокупность. Однородные совокупности признаков часто называют **системами**, акцентируя внимание на наличии связей между признаками совокупности.

Совокупности признаков обычно также группируются. Особенностью их группировок является то, что они имеют строго **иерархический характер**, т.е. последовательность групп признаков разного порядка строго определена. Когда же речь идет о группировках наблюдений-объектов, то их **иерархия** (последовательность групп от низших порядков к высшим) условна, она всегда может измениться при изменении порядка группирующих признаков. Группы признаков обычно называют классами и подклассами или классами разного уровня (иерархии).

На нулевом уровне иерархии признаков размещается имя всей совокупности признаков, например, «показатели развития промышленных предприятий». Далее следуют классы первого уровня с их именами, например, «материальные ресурсы», «затраты», «результаты», «финансовые пассивы», «финансовые активы» и т.д. Эти классы детализируются на втором уровне: например, «материальные ресурсы» делятся на «основной капитал», «запасы готовой продукции», «производственные запасы», «незавершенное производство». На третьем уровне иерархии «запасы готовой продукции», например, делятся по видам продукции. И так далее. Разные направления иерархии могут иметь разное количество уровней детализации (иерархии). Например, «материальные ресурсы» могут иметь 4 уровня, а «финансовые активы» — 3. В исходной матрице наблюдений только признаки низшего уровня иерархии (классов высшего порядка) имеют числовые значения (после группировки признаков и обработки матрицы наблюдений могут быть введены столбцы со значениями итоговых показателей по некоторым или всем классам и подклассам признаков).

Сама группировка формально может быть проведена так же, как и группировка объектов (но с некоторыми отличиями). Разным классам одного уровня, образующим один класс предыдущего уровня, присваиваются различные целые числа-ранги, т.е. классы «измеряются» в номинальной шкале. Как видно, «измерение»

классов одного уровня зависит от результатов «измерения» классов предыдущего уровня, чего не было при группировке совокупностей объектов. Далее, в матрицу наблюдений вводятся строки «классы первого уровня», «классы второго уровня» и т.д. с рангами, присвоенными соответствующим классам, в столбцах признаков. И, наконец, осуществляется перестановка столбцов матрицы наблюдений по возрастанию рангов сначала классов первого уровня, потом второго уровня и т.д. Ранги классов образуют мультииндексы или коды признаков. После завершения группировки введенные строки классов можно убрать.

Обычно эти операции не проводятся, т.к. признаки группируются уже при составлении матрицы наблюдений.

Как исходные массивы и матрицы наблюдений, так и результаты их группировок или других обработок могут изображаться в виде таблиц и графиков. **Таблица** — это визуализированный двухмерный массив с **общим названием-титулом**, названиями строк и названиями столбцов. Первый столбец (столбцы), в котором размещены названия строк, называется **подлежащим** таблицы, первая строка (строки) с названиями столбцов — **сказуемым** таблицы. Подлежащее и сказуемое часто включают мультииндексы-коды соответствующих объектов или признаков. В титул обычно выносится общее имя совокупности элементов (объектов или признаков) сказуемого и/или подлежащего.

Существует несколько вариантов таблиц для массивов типа $\{x_{tij}\}$, имеющих 3 размерности: время t, объекты i и признаки j. Если в подлежащем — время, а в сказуемом — объекты, то в титул должно быть вынесено имя признака; если в подлежащем — объекты, в сказуемом — признаки, то в титуле должно быть указано время и т.д. Всего таких вариантов — 6.

Если в табулируемой матрице не произведено группировок, то таблица является **простой** с простыми именами строк и столбцов. Если строки и/или столбцы сгруппированы, то их имена в таблице являются составными: кроме индивидуальных имен строк и столбцов они включают и имена их групп и классов.

В случае, когда столбцов таблицы не слишком много, информация может быть представлена (визуализирована) **графиком**. Ось абсцисс соответствует обычно подлежащему таблицы, а ось ординат — сказуемому. Сами значения показателей-признаков изображаются в виде различных графических образов, например, в виде «столбиков». Если в подлежащем размещены моменты времени, график выражает траектории изменения показателей.

1.10. Задачи

1. Определить пункты, которые являются выпадающими из общего ряда.

1.1 а) отношений, б) порядковая, в) количественная, г) классификаций;

1.2 а) Пуассон, б) Рамсей, в) Бернулли, г) Байес;

1.3 а) темпы роста, б) относительные, в) производные, г) первичные;

1.4 а) Кейнс, б) Байес, в) Синклер, г) Бернулли;

1.5 а) фондоемкость, б) материалоемкость, в) трудоемкость, г) срок окупаемости инвестиций;

1.6 а) Стивенс, б) Кэмпбел, в) реляционная структура, г) Тарский;

1.7 а) капитал, б) население, в) инвестиции, г) внешний долг;

1.8 а) Пуассон, б) Рамсей, в) Бернулли, г) Байес;

1.9 а) Суппес, б) Стивенс, в) Пуассон, г) Пфанцагль;

1.10 а) величина-признак, б) величина-показатель, в) показатель-определение, г) показатель-наблюдение;

1.11 а) Герман, б) Кетле, в) Моргенштерн, г) Синклер;

1.12 а) Тарский, б) операциональная, в) репрезентативная, г) Кэмпбел;

1.13 а) Зинес, б) Суппес, в) Моргенштерн, г) Петти;

1.14 а) статистика, б) statistics, в) информация, г) statistic;

1.15 а) наименований, б) интервальная, в) ординальная, г) шкалирование;

1.16 а) Суппес, б) интервальная, в) Стивенс, г) порядковая;

1.17 а) Бернулли, б) субъективная, в) Байес, г) объективная;

1.18 а) Пфанцагль, б) Зинес, в) Нейман, г) Кэмпбел;

1.19 а) управляемый эксперимент, б) пассивное наблюдение, в) статистика, г) операциональное определение;

1.20 а) Кетле, б) Кейнс, в) Петти, г) Герман;

1.21 а) производственные мощности, б) выпуск продукции, в) затраты, г) амортизационные отчисления;

1.22 а) Пуассон, б) Рамсей, в) Бернулли, г) Байес;

1.23 а) кластер, б) класс, в) группа, г) совокупность;

1.24 а) абсолютная, б) относительная, в) экстенсивная, г) интенсивная;

1.25 а) дискретный, б) непрерывный, в) моментный, г) интервальный;

1.26 а) подлежащее, б) предлог, в) сказуемое, г) таблица.

2. Какой тип — запаса или потока — имеют следующие величины: а) инвестиции; б) население; в) основные фонды; г) активы?

3. К какому классу относятся и какую размерность имеют следующие интенсивные величины: а) фондоемкость; б) материалоемкость; в) трудоемкость; г) фондоотдача?

Таблица 1.2

Год	Объем производства, млрд. руб.	Абсолютный прирост, млрд.	Темп роста (годовой)	Темп прироста (годовой), %	Абсолютное значение 1% прироста, млрд.
1	2	3	4	5	6
1992	127				
1993			1.102		
1994				7.1	
1995	164.6				
1996					
1997				9.9	1.75

4. Пусть y_t — значение величины в момент времени t. Запишите формулу моментного темпа прироста и непрерывного темпа роста.

5. Имеются данные об объеме производства в отрасли (табл. 1.2).

   Вычислить и вставить в таблицу недостающие показатели.

6. Была проведена группировка студентов НГУ по трем признакам:

   1-й признак: место постоянного жительства (город; село);

   2-й признак: средний балл в аттестате (выше 4.5; от 3.5 до 4.5; ниже 3.5);

   3-й признак: средний балл за вступительные экзамены (выше 4.5; от 3.5 до 4.5; ниже 3.5).

   Определите:

   а) общее число групп и число групп высшего порядка;

   б) количество классов групп 1-го, 2-го и 3-го порядка;

   в) количество групп в классах **2**, **13**, **23**;

   г) число конечных групп в каждой группе класса **2**, **13**, **23**.

   д) Число элементов конечной группы **221** равно 5, в остальных конечных группах по 2 элемента. Каково значение линейного индекса второго элемента конечной группы **232**?

   е) Сколько всего элементов в совокупности?

Глава 2

Описательная статистика

Исходный массив наблюдений может достигать значительных размеров, и непосредственно по его информации трудно делать какие-либо содержательные заключения о свойствах изучаемых совокупностей. Задача описательной статистики — «сжать» исходный массив, представить его небольшим набором числовых характеристик, которые концентрированно выражают свойства изучаемых совокупностей. Граница между описательной статистикой, с одной стороны, и математической статистикой, эконометрией, анализом данных, с другой стороны, достаточно расплывчата. Обычно в описательной статистике даются элементарные сведения, достаточные для проведения начальных этапов экономико-статистического исследования, которые более углубленно и более строго рассматриваются в других научных дисциплинах статистического ряда (в последующих разделах книги).

2.1. Распределение частот количественного признака

Пусть имеются наблюдения x_i, $i = 1, \ldots, N$ за некоторой непрерывной количественной величиной-признаком, т.е. матрица наблюдений имеет размерность $N \times 1$. Такую матрицу наблюдений обычно называют **рядом наблюдений**. В статистике совокупность этих значений иногда называется также **вариационным рядом**. Пусть проведена группировка совокупности по этому признаку с выделением k групп. В соответствии с обозначениями предыдущей главы мультииндексом группы является I, равный i_1, где i_1 — индекс группы. В этом и ряде последующих

пунктов (при $n = 1$) в качестве индекса группы будет использоваться не i_1, чтобы не путать его с линейным индексом i наблюдения, а l. Соответственно, z_l, $l = 0, 1, \ldots, k$ — **границы полуинтервалов**, N_l — групповые численности, которые в этом случае называют **частотами** признака. Следует иметь в виду, что x — случайная величина, но все z — детерминированы.

Размеры полуинтервалов,

$$\Delta_l = z_l - z_{l-1},$$

обычно берут одинаковыми. При выборе размера полуинтервалов можно использовать одно из следующих правил:

$$\Delta = 3.5sN^{-1/3} \quad \text{(правило Скотта)}$$

или

$$\Delta = 2\,IQR\,N^{-1/3} \quad \text{(правило Фридмена—Диакониса)},$$

где s — среднеквадратическое отклонение, $IQR = x_{0.75} - x_{0.25}$ — межквартильное расстояние (определение величин s, $x_{0.25}$ и $x_{0.75}$ дается ниже). В литературе также часто встречается правило Стёрджесса для количества групп:

$$k = 1 + \log_2 N \approx 1 + 1.44 \ln N,$$

однако было показано, что оно некорректно, поэтому использовать его не рекомендуется. В качестве значения признака на l-м полуинтервале можно принять среднее значение признака на этом полуинтервале:

$$\bar{x}_l = \frac{1}{N_l} \sum x_{l*}$$

(использовано введенное в предыдущей главе обозначение x_{l*} всех наблюдений, попавших в l-ю группу). Однако, как правило, в качестве этого значения принимается **середина полуинтервала**:

$$\bar{x}_l = \frac{1}{2}\,(z_l + z_{l-1}) = z_{l-1} + \frac{\Delta_l}{2},$$
$$\alpha_l = \frac{N_l}{N},$$

— **относительные частоты признака** или **оценки вероятностей** (эмпирические вероятности) попадания значений признака в l-й полуинтервал, то есть $\alpha_1 = P\,(z_0 \leqslant x \leqslant z_1)$, $\alpha_l = P\,(z_{l-1} < x \leqslant z_l)$, $l = 2, \ldots, k$.

$$f_l = \frac{\alpha_l}{\Delta_l} \tag{2.1}$$

— **плотности относительной частоты** или **оценки плотности вероятности**.

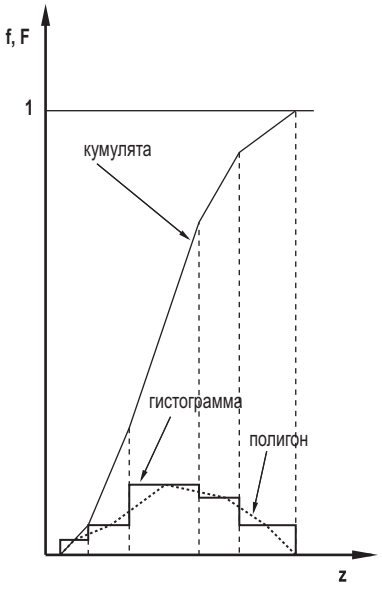

Рис. 2.1. Графическое изображение плотностей частоты

Очевидно, что

$$\sum \alpha_l = 1, \text{ или } \sum f_l \Delta_l = 1. \qquad (2.2)$$

Далее:

$$F_l = \sum_{l'=1}^{l} \alpha_{l'}, \text{ или } F_l = \sum_{l'=1}^{l} f_{l'}\, \Delta_{l'}, \qquad (2.3)$$

— **накопленные относительные частоты** или оценки вероятностей того, что значение признака не превысит z_l, т.е. $F_l = P\,(x \leqslant z_l)$.

Крайние значения этих величин равны 0 и 1: $F_0 = 0,\ F_k = 1$.

Числа $\alpha_l,\ f_l,\ F_l$ $(l = 1, \ldots, k)$ характеризуют разные аспекты распределения частот количественного признака. Понятно, что, если размеры полуинтервалов одинаковы, α_l и f_l различаются с точностью до общей нормировки и являются одинаковыми характеристиками распределения.

Графическое изображение плотностей частоты называется **гистограммой**, а накопленных частот — **кумулятой**. Поскольку плотности частот неизменны на каждом полуинтервале, гистограмма ступенчатая функция (точнее, график ступенчатой функции). Накопленные частоты линейно растут на каждом полуинтервале, поэтому кумулята — кусочно-линейная функция. Вид этих графиков приведен на рисунке 2.1.

Еще один графический образ плотностей частоты называется **полигоном**. Этот график образован отрезками, соединяющими середины ступенек гистограммы. При этом первый отрезок соединяет середину первой ступеньки с точкой z_0 оси абсцисс, последний отрезок — середину последней ступеньки с точкой z_k.

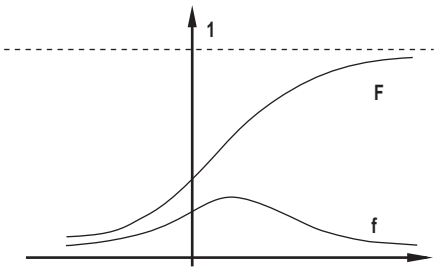

Рис. 2.2

Теоретически можно представить ситуацию, когда N и $k \to \infty$, при этом следует допустить, что $z_0 \to -\infty$, а $z_k \to +\infty$. В результате функции $f(z)$ и $F(z)$, графиками которых были гистограмма и кумулята, станут гладкими (рис. 2.2). В математической статистике их называют, соответственно, **функцией плотности распределения вероятности** и **функцией распределения вероятностей** случайной величины (см. Приложение А.3.1).

Формулы (2.1–2.3) преобразуются, соответственно, в

$$\frac{dF(z)}{dz} = f(z), \quad \int\limits_{-\infty}^{+\infty} f(z)\,dz = 1, \quad F(z) = \int\limits_{-\infty}^{z} f(z')\,dz'.$$

Обычно функции f и F записываются с аргументом, обозначенным символом случайной величины: $f(x)$ и $F(x)$. При этом предполагается, что в такой записи x есть детерминированный «образ» соответствующей случайной величины (в математической статистике для этого часто используют соответствующие прописные символы: $f(X)$ и $F(X)$). Такие функции являются **теоретическими** и выражают различные **законы распределения**, к которым лишь приближаются эмпирические распределения.

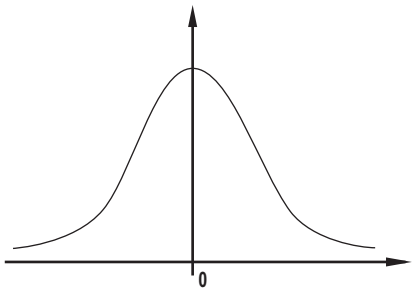

Рис. 2.3

Наиболее распространенным в природе является так называемый закон **нормального распределения**, плотность которого в простейшем случае (при нулевом математическом ожидании и единичной дисперсии) описывается следующей функцией:

$$f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}$$

Ее график, часто называемый **кривой Гаусса**, изображен на рисунке 2.3.

Наиболее вероятное значение величины, имеющей такое распределение, — нуль. Распределение ее **симметрично**, и вероятность быстро падает по мере увеличения ее абсолютной величины. Обычно такое распределение имеют случайные ошибки измерения (при разной дисперсии).

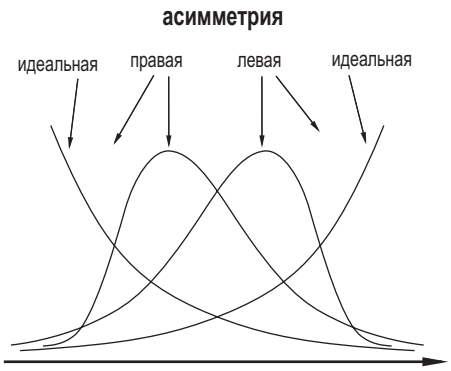

Рис. 2.4

Различают несколько типов распределений признака (случайной величины).

На рисунке 2.4 показаны **асимметричные** или **скошенные** распределения: с **правой** и **левой асимметрией**, **идеальная правая** и **идеальная левая асимметрия**. При правой (левой) асимметрии распределение скошено в сторону больших (меньших)

значений. При идеальной правой (левой) асимметрии вероятность падает (увеличивается) с ростом значения величины на всем интервале ее значений, наиболее вероятно ее минимальное (максимальное) значение. В данном случае идеальными названы распределения с предельной асимметрией.

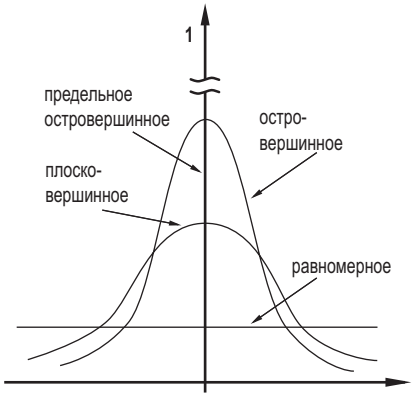

Рис. 2.5

На рисунке 2.5 приведен вид **высоко-** или **островершинных** и **низко-** или **плосковершинных распределений**. В первом случае основная часть значений признака сосредоточена в узкой центральной области распределения, во втором — центральная область распределения «размыта». Плосковершинное распределение в пределе превращается в **равномерное**, плотность которого одинакова на всем интервале значений. Предельным островершинным распределением является вертикальный отрезок единичной длины — распределение детерминированной величины.

Распределения с одним пиком плотности вероятности называют **унимодальными**. На рисунке 2.6 приведен пример **бимодального** распределения и предельного бимодального распределения, называемого **U-образным**. В общем случае распределение с несколькими пиками плотности называют **полимодальным**.

В математической статистике множество всех теоретически возможных значений случайной величины x, характеризуемое функциями f и F, называют **генеральной совокупностью**, а ряд наблюдений $x_1, \ldots, x_N$ — **выборочной совокупностью**, или **выборкой**.

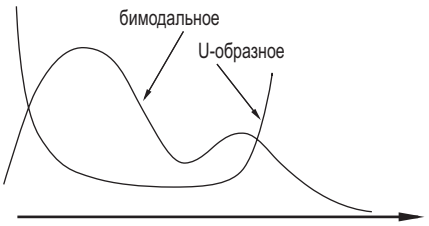

Рис. 2.6

Вообще говоря, гистограмму и кумуляту можно построить непосредственно по данным ряда наблюдений без предварительной группировки. Если предположить для простоты, что все значения в ряде наблюдений различны, то k принимается равным N. В качестве границ полуинтервалов z_i, $i = 1, \ldots, N-1$ принимаются полусуммы двух соседних значений в ряде наблюдений, **упорядоченном по возрастанию** (строго говоря, само упорядочение является операцией группировки в простейшем случае):

$$z_i = \frac{1}{2}(x_i + x_{i+1}).$$

В качестве z_0 и z_N естественно принять, соответственно, $2x_1 - z_1$ и $2x_N - z_{N-1}$, так что первое и последнее значение в ряде наблюдений оказываются в точности на середине своих полуинтервалов. Относительные частоты для всех полуинтервалов одинаковы и равны $1/N$. Однако плотность частоты различается: она тем выше, чем короче полуинтервал, т.е. чем плотнее наблюдения расположены на числовой оси.

2.2. Средние величины

Средние величины, или просто **средние**, являются особым подклассом интенсивных величин, т.к. рассчитываются как отношения других величин. Они выступают наиболее общими характеристиками совокупности объектов. Каждая средняя рассчитывается по конкретному признаку, характеризующему объекты совокупности, и является качественно такой же величиной, имеет те же единицы измерения или ту же размерность (или она безразмерна), что и усредняемый признак. Характер средних по объемным и относительным величинам несколько различается. Ниже рассматриваются сначала **средние объемные** и на их примере — виды средних, затем — **средние относительные** величины.

Пусть x_i — некоторый объемный признак i-го объекта, $1, \ldots, N$, то есть количество объектов в совокупности равно N, как и прежде, $x = \sum\limits_i x_i$, тогда расчет среднего по совокупности значения данного объемного признака, который обычно обозначается тем же символом, но без индекса объекта и с чертой над символом, осуществляется по следующей формуле:

$$\bar{x} = \frac{1}{N}x = \frac{1}{N}\sum_i x_i.$$

Это — **среднее арифметическое (среднеарифметическое) простое** или **средняя арифметическая (среднеарифметическая) простая**. Оно является отношением двух объемных величин: суммарного по совокупности признака и количества объектов в совокупности.

Пусть теперь вся совокупность делится на k групп, N_l — количество объектов в l-й группе, $N = \sum\limits_l N_l$, значение признака внутри каждой группы не варьируется и равняется x_l. Тогда

$$\bar{x} = \frac{1}{N}\sum_l N_l x_l = \sum_l \alpha_l x_l, \text{ где } \alpha_l = \frac{N_l}{N}, \quad \Big(\sum \alpha_l = 1\Big) \text{ — вес } l\text{-й группы.}$$

Это — **среднее арифметическое (среднеарифметическое) взвешенное** (среднеарифметическая взвешенная).

К аналогичной формуле для средней по исходной совокупности можно придти и иначе. Пусть, как и сначала, признак варьирует по всем объектам совокупности, а $\bar{x}_l$ — среднеарифметическое простое по l-й группе. Очевидно, что

$$x = \sum N_l \bar{x}_l, \quad \text{и} \quad \bar{x} = \sum \alpha_l \bar{x}_l.$$

По такой же формуле производится расчет средней по данным эмпирического распределения частот признака (см. предыдущий пункт). В качестве $\bar{x}_l$ в таком случае принимают не среднее по l-й группе, а, как отмечалось выше, середину l-го полуинтервала.

Предполагая, что все объекты совокупности имеют разные веса (вес i-го объекта равен α_i), среднее по совокупности записывается как взвешенное:

$$x = \sum \alpha_i x_i.$$

Это — более общая формула среднеарифметического: при равных весах, то есть в случае, если $\alpha_i = 1/N$ для всех i, она преобразуется в формулу среднеарифметического простого.

Для нахождения средней величины типа запаса за некоторый период времени используется среднее арифметическое взвешенное, называемая **средним хронологическим** (или средней хронологической). Смысл этой величины поясняется рисунком 2.7.

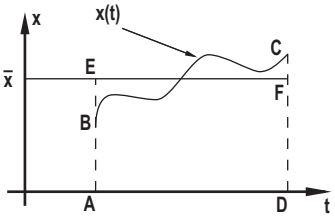

Рис. 2.7

Среднюю хронологическую $\bar{x}$ надо найти так, чтобы площадь $ABCD$ под линией динамики $x(t)$ (BC), т.е. сумма значений показателя за период, равнялась площади прямоугольника $AEFD$ под линией средней EF (см. рис. 2.7).

Другими словами, для расчета средней хронологической используется формула:

$$\bar{x} = \frac{\text{площадь } ABCD}{\text{длина } AD}.$$

На практике в дискретном случае этот расчет можно провести следующим образом.

Пусть $x_0,\ x_1,\ \ldots,\ x_N$ — значения некоторой объемной величины типа запаса в моменты времени $t_0,\ t_1,\ \ldots,\ t_N$, и $\tau_i = t_i - t_{i-1},\ i = 1,\ \ldots,\ N,\ \tau = \sum \tau_i$ (длина AD).

Если предположить, что на каждом временном отрезке τ_i динамика показателя линейна, то его суммарное значение на этом отрезке рассчитывается как $\tau_i \dfrac{x_i + x_{i-1}}{2}$,

и для общей средней хронологической справедливо соотношение:

$$\bar{x} = \frac{1}{2\tau} \sum_{i=1}^{N} \tau_i \, (x_i + x_{i-1}).$$

В выражении этой величины как среднеарифметической взвешенной веса имеют следующие значения:

$$\alpha_0 = \frac{\tau_1}{2\tau}, \; \alpha_i = \frac{\tau_i + \tau_{i+1}}{2\tau}, \; i = 1, \, \ldots, \, N-1, \; \alpha_N = \frac{\tau_N}{2\tau}.$$

Их сумма равна единице.

Если все временные отрезки τ_i одинаковы, то веса первого и последнего x в средней хронологической будут равняться $1/2N$, а веса всех промежуточных «x-ов» — $1/N$.

На практике чаще всего рассчитывают средние величины типа запаса за период времени (обычно за год) по данным на начало и конец этого периода (года). Т.е. решается задача нахождения средней хронологической $\bar{x}$ за некоторый период, для которого известно значение показателя на начало — x_0 и конец периода — x_1. Эта величина, чаще всего, находится как средневзвешенное арифметическое:

$$\bar{x} = (1 - \alpha) \, x_0 + \alpha x_1, \quad \text{или} \quad \bar{x} = x_0 + \alpha\Delta, \quad \text{или} \quad \Delta = x_1 - x_0.$$

Если динамика показателя равномерна (линейна), то $\alpha = 1/2$; если более интенсивные сдвиги в величине показателя происходят в 1-й половине периода, то $\alpha > 1/2$; в противном случае — $\alpha < 1/2$. В советской статистике при расчете, например, среднегодовых основных фондов α принимался в интервале от 0.3 до 0.4, поскольку в плановой экономике вводы и выбытия фондов обычно сдвигаются к концу года — к моменту отчета по плану. Этот параметр иногда называют **среднегодовым коэффициентом**.

При предположении, что на данном отрезке времени неизменным остается относительный прирост (моментный темп прироста), и динамика имеет экспоненциальный характер, справедливы следующие выражения (как и прежде, τ — длина данного временного отрезка, Δ — прирост показателя за период):

$$x_t = x_0 \left(\frac{x_1}{x_0} \right)^{t/\tau}, \quad \text{при} \quad 0 \leqslant t \leqslant \tau,$$

$$\bar{x} = \frac{x_0}{\tau} \int_0^{\tau} \left(\frac{x_1}{x_0} \right)^{t/\tau} dt = \frac{x_1 - x_0}{\ln x_1 - \ln x_0} = \frac{\Delta}{\ln \left(1 + \Delta/x_0\right)}.$$

В знаменателе этого выражения для средней хронологической находится непрерывный темп прироста за период (см. п. 1.8), т.е. средняя хронологическая определяется делением абсолютного прироста на относительный прирост за период. Это — особый вид средней, которую иногда и называют собственно хронологической.

Чтобы лучше понять ее смысл, полезно найти ее предельное значение при $\Delta \to 0$. Для этого логарифм в знаменателе раскладывается в степенной ряд:

$$\ln\left(1+\frac{\Delta}{x_0}\right)=\frac{\Delta}{x_0}-\left(\frac{\Delta}{x_0}\right)^2\frac{1}{2}+\left(\frac{\Delta}{x_0}\right)^3\frac{1}{4}-\left(\frac{\Delta}{x_0}\right)^4\frac{1}{4}+\cdots,$$

затем сокращается Δ в числителе и знаменателе, и он (Δ) приравнивается нулю. Искомый предел равен x_0. Таким образом, на бесконечно малых отрезках времени значение этой величины равно самому показателю, а на конечных отрезках — его среднему значению при предположении, что темп роста на этом отрезке остается неизменным.

Возвращаясь к общему случаю $N+1$ временной точки, среднюю хронологическую при предположении неизменности темпа роста внутри каждого временного периода можно рассчитать следующим образом:

$$\bar{x}=\frac{1}{\tau}\sum_{i=1}^{N}\frac{x_i-x_{i-1}}{\ln x_i-\ln x_{i-1}}\tau_i.$$

Несложно убедиться в том, что в случае, если средние в единицу времени темпы роста $\left(\frac{x_i}{x_{i-1}}\right)^{\frac{1}{\tau_i}}$ на всех временных отрезках одинаковы и равны среднему в единицу темпу роста за весь период $\left(\frac{x_N}{x_0}\right)^{\frac{1}{\tau}}$, среднее хронологическое рассчитывается только по двум крайним значениям:

$$\bar{x}=\frac{x_N-x_0}{\ln x_N-\ln x_0}.$$

Расчет средних хронологических величин типа запаса является необходимой операцией для приведения этих величин к форме, сопоставимой с величинами типа потока, имеющими другое качество. Так, например, производительность труда рассчитывается как отношение выпуска продукции за определенный период времени к средней хронологической занятых в производстве за этот же период. Если величины типа запаса и потока имеют одно качество (потоки выражают изменение запасов за период времени), то используются и показатели отношения потока к запасу на начало или конец периода (или наоборот). Так, например, отношение

выбывших в течение года основных фондов к основным фондам на начало года называется коэффициентом выбытия фондов, а отношение годового ввода фондов к фондам на конец года — коэффициентом обновления фондов.

Среднеарифметическое является частным случаем так называемого **среднестепенного** или **среднего степенного**, которое рассчитывается по следующей формуле:

$$\bar{x} = \left(\sum \alpha_i x_i^k\right)^{\frac{1}{k}}.$$

Следует обратить внимание, что эта величина существует не при всех k, если некоторые из x_i отрицательны. Чтобы избежать непринципиальных уточнений, в дальнейшем предполагается, что все значения признака положительны.

При $k = 1$ среднее степенное превращается в обычное среднеарифметическое, при $k = 2$ это — **среднеквадратическое**, используемое для оценки степени вариации признака по совокупности, при $k = -1$ — среднее **гармоническое**, примеры использования которого приводятся при рассмотрении средних относительных величин, при $k = 0$ — среднее **геометрическое**.

Последнее утверждение доказывается путем нахождения предела среднего степенного при $k \to 0$. Для того чтобы сделать такой предельный переход, обе части формулы среднего степенного возводятся в степень k, затем $\bar{x}^k$ и все x_i^k представляются разложением в степенные ряды:

$$1 + \frac{k \ln \bar{x}}{1!} + \frac{(k \ln \bar{x})^2}{2!} + \cdots = \sum \alpha_i (1 + \frac{k \ln x_i}{1!} + \frac{(k \ln x_i)^2}{2!} + \cdots),$$

далее в обеих частях полученного выражения сокращаются единицы $(1 = \sum \alpha_i)$, и эти обе части делятся на k. Теперь при $k = 0$ получается следующее равенство:

$$\ln \bar{x} = \sum \alpha_i \ln x_i,$$

откуда $\bar{x} = \prod x_i^{\alpha_i}$, что и требовалось доказать.

Средние геометрические используются при построении некоторых специальных индексов. Но это тема следующей главы. Простые примеры использования средней геометрической дает производственная функция.

Пусть в производственной функции Кобба—Дугласа так называемая отдача на масштаб постоянна, т.е. сумма показателей степеней в выражении функции равна единице, и при увеличении использования ресурсов в одинаковое количество раз выпуск продукции растет в такое же количество раз:

$$X = aC^{\alpha} L^{1-\alpha},$$

или в более развернутой форме:

$$X = (Ca_C)^{\alpha}(La_L)^{1-\alpha},$$

где a_C — коэффициент фондоотдачи при нормальном соотношении между основным капиталом и трудом, a_L — коэффициент производительности труда при тех же нормальных условиях.

Нормальное соотношение труда и капитала определяется сложившимся организационно-технологическим уровнем производства. Это — фиксированная величина:

$$s^n = \frac{C}{L}.$$

Откуда $a_C = a\,(s^n)^{\alpha-1}$, $\quad a_L = a\,(s^n)^{\alpha}$.

Таким образом, в общем случае (при любых соотношениях ресурсов) выпуск продукции является средневзвешенной геометрической потенциального выпуска, который мог бы быть обеспечен основным капиталом при нормальном соотношении его с трудом (величины Ca_C), и потенциального выпуска, который обеспечивается трудом при нормальном его соотношении с капиталом (La_L). Коэффициент a в исходной записи производственной функции равен $a_C^{\alpha}a_L^{1-\alpha}$, и он может называться коэффициентом общей производительности ресурсов, поскольку является также среднегеометрической нормальной фондоотдачи и нормальной производительности труда.

Более общая форма связи между выпуском и ресурсами дается производственной функцией с постоянной эластичностью замены ресурсов. В развернутом виде она записывается следующим образом:

$$X = \left[\alpha\,(Ca_C)^{-\rho} + (1-\alpha)\,(La_L)^{-\rho}\right]^{-\frac{1}{\rho}}.$$

Это — пример использования среднего степенного при нецелочисленных значениях параметра степени, поскольку ρ (равный $-k$ в общей формуле среднего степенного) может принимать любые значения на отрезке $[-1, +\infty]$ (при $\rho \to 0$, в силу приведенного выше доказательства, производственная функция с постоянной эластичностью замены преобразуется к форме Кобба—Дугласа). От величины этого параметра зависят возможности взаимного замещения ресурсов, допускаемые в данной модели производства. Чем выше его величина, тем более затруднено это замещение.

Такое свойство производственной функции с постоянной эластичностью замены эквивалентно известному свойству среднего степенного: оно увеличивается с ростом k.

Среднее степенное увеличивается с ростом k, в частности, по возрастанию средние степенные располагаются в следующем порядке: гармоническое, геометрическое, арифметическое, квадратическое. Это свойство иногда называют **мажорантностью** средних.

Пусть $\bar{x}(k)$ — среднее степенное, пусть далее $k_2 > k_1$, и требуется доказать, что $\bar{x}(k_1) > \bar{x}(k_2)$.

Эти средние можно записать в следующем виде:

$$\bar{x}\,(k_1) = \left(\sum \alpha_i \left(x_i^{k_2}\right)^{\frac{k_1}{k_2}}\right)^{\frac{1}{k_1}}, \quad \bar{x}\,(k_2) = \left(\left(\sum \alpha_i x_i^{k_2}\right)^{\frac{k_1}{k_2}}\right)^{\frac{1}{k_1}},$$

и ввести промежуточные обозначения (чтобы не загромождать изложение):

$$\begin{aligned} &y_i = x_i^{k_2}, \\ &q = \frac{k_1}{k_2}, \\ &f\,(y) = y^q, \\ &v = \frac{d^2 f}{dy^2} = q\,(q-1)\,y^{q-2}, \\ &a_1 = \sum \alpha_i f(y_i), \;\; a_2 = f\left(\sum \alpha_i y_i\right). \end{aligned}$$

В этих обозначениях утверждение, которое следует доказать, записывается следующим образом:

$$a_2^{\frac{1}{k_1}} > a_1^{\frac{1}{k_1}}.$$

Далее рассматривается три возможных случая:

1) $k_2 > k_1 > 0$,

2) $k_2 > 0 > k_1$,

3) $0 > k_2 > k_1$.

В первом случае $q < 1$, $v < 0$, т.е. функция f вогнута (выпукла вверх) и $a_2 > a_1$ по определению такой функции. После возведения обеих частей этого неравенства в положительную степень $1/k_1$ знак его сохраняется, что и завершает доказательство в этом случае.

Во втором и третьем случаях $v > 0$, и функция f выпукла (выпукла вниз). Поэтому $a_2 < a_1$, и после возведения обеих частей этого неравенства в отрицательную степень $1/k_1$ оно меняет знак, приобретая тот, который нужно для завершения доказательства.

Свойство мажорантности средних выражается и в том, что предельные значения среднего степенного при $k = \pm\infty$ равны, соответственно, максимальному и минимальному значению признака в выборке.

Для доказательства этого факта в выражении среднего выносится за скобки x_1:

$$\bar{x} = x_1 \left(\alpha_1 + \sum_{i=2}^{N} \alpha_i \left(\frac{x_i}{x_1} \right)^k \right)^{\frac{1}{k}}.$$

Если x_i упорядочены по возрастанию и $x_1 = \min x_i$, то $x_i/x_1 \geqslant 1$, и при $k \to -\infty$ выражение в скобках стремится к $\sum\limits_{i=1}^{k'} \alpha_i$, где k' — число объектов, для которых усредняемый признак минимален (если минимум единственный, то $k' = 1$), т.е. конечно. Это выражение возводится в степень $1/k$, которая стремится к нулю при $k \to -\infty$. Следовательно, среднее степенное при $k \to -\infty$ равно минимальному значению усредняемых признаков.

Предположив теперь, что x_i упорядочены по убыванию, аналогичным образом можно доказать, что среднее степенное при $k \to +\infty$ равно максимальному значению признака по совокупности.

Существует наиболее общая запись средневзвешенного:

$$\bar{x} = f^{-1}\left(\sum \alpha_i f(x_i) \right). \tag{2.4}$$

Если f — степенная функция x^k, то речь идет о средней степенной, если f — логарифмическая функция $\ln x$, то это — средняя логарифмическая, которая является частным случаем средней степенной при $k = 0$, если f — показательная функция a^x, то это — средняя показательная и т.д.

Особенностью **средних относительных** величин является то, что они, как правило, рассчитываются как средние взвешенные.

Пусть i-й объект, $i = 1, \ldots, N$ характеризуется зависимыми друг от друга объемными величинами y_i и x_i. Показателем этой зависимости является относительная величина $a_i = y_i/x_i$. Это может быть производительность, фондовооруженность труда, рентабельность и т.д. Понятно, что средняя по совокупности объектов относительная величина a (знак черты над символом, обозначающим среднее относительное, часто опускается) рассчитывается по следующей формуле:

$$a = \frac{\sum y_i}{\sum x_i},$$

которая легко преобразуется в формулу средней взвешенной:

$$a = \sum \alpha_i^x a_i, \text{ где } \alpha_i^x = \frac{x_i}{\sum x_i}, \quad \text{или}$$
$$a = \frac{1}{\sum \frac{\alpha_i^y}{a_i}}, \text{ где } \alpha_i^y = \frac{y_i}{\sum y_i}.$$

Таким образом, если веса рассчитываются по структуре объемных величин, стоящих в знаменателе, то средняя относительная является средней взвешенной **арифметической**, если эти веса рассчитываются по объемным величинам, стоящим в числителе, то она является средней взвешенной **гармонической**.

Формально можно рассчитать простую среднюю (например, арифметическую)

$$a' = \frac{1}{N} \sum a_i,$$

но содержательного смысла она иметь не будет. Это становится понятным, как только осуществляется попытка привести слагаемые $\sum y_i/x_i$ к общему знаменателю. Тем не менее, такая средняя также может использоваться в анализе. Например, ее иногда полезно сравнить с фактической средней a для выявления некоторых характеристик асимметрии распределения признака по совокупности. Если $a > a'$, то в совокупности преобладают объекты с повышенной величиной a_i, и, по-видимому, имеет место правая асимметрия, в противном случае в совокупности больший удельный вес занимают объекты с пониженной a_i (левая асимметрия). Однако в статистике имеются более четкие критерии асимметрии распределения.

Особое место среди средних относительных занимают средние темпы роста. Темпы роста величин типа потока выражают отношение потока за единицу (период) времени к потоку за некоторую предыдущую единицу (предыдущий период) времени. Темпы роста величин типа запаса показывают отношение запаса в момент времени к запасу в некоторый предыдущий момент времени. Такой же смысл имеют и средние темпы роста. Средние за период темпы роста рассчитываются обычно как средние геометрические.

Пусть $x_0, x_1, \ldots, x_N$ значения некоторой объемной величины в моменты времени $t_0, t_1, \ldots, t_N$, если эта величина типа запаса, или в последнюю единицу времени, соответственно, 0-го, 1-го и т.д., N-го периода времени, если речь идет о величине типа потока (t_0 — последняя единица времени 0-го периода, $[t_{i-1} + 1, t_i]$ — i-й период). Как и прежде, $\tau_i = t_i - t_{i-1}$, $i = 1, \ldots, N$, $\tau = \sum \tau_i$. Предполагается, что i — целые положительные числа.

Тогда

$$\lambda_i = \frac{x_i}{x_{i-1}} \text{ — темп роста за } i\text{-й период времени,}$$

$$\lambda = \frac{x_N}{x_0} = \prod_{i=1}^{N} \lambda_i \text{ — общий темп роста.}$$

Если все периоды одинаковы и равны единице ($\tau_i = 1$), то средний в единицу времени темп роста определяется по формуле:

$$\bar{\lambda} = \left(\frac{x_N}{x_0}\right)^{\frac{1}{N}} = \left(\prod \lambda_i\right)^{\frac{1}{N}},$$

т.е. он равен простому среднему геометрическому темпу по всем периодам.

В общем случае (при разных τ_i) данная формула приобретает вид средневзвешенной геометрической:

$$\bar{\lambda} = \left(\frac{x_N}{x_0}\right)^{\frac{1}{\tau}} = \prod \bar{\lambda}_i^{\alpha_i},$$

где $\bar{\lambda}_i = \lambda_i^{1/\tau_i}$ — средний за единицу времени темп роста в i-м периоде, $\alpha_i = \tau_i/\tau$.

Для величин типа запаса имеется еще одна форма средних темпов роста: отношение средней хронологической за период времени к средней хронологической за некоторый предыдущий период. Такую форму средних можно рассмотреть на следующем примере.

Пусть $x_0,\ x_1,\ x_2$ — значение величины типа запаса в три момента времени: на начало первого периода, конец первого периода, который одновременно является началом второго периода, конец второго периода. Оба периода времени одинаковы. Средние хронологические за первый и второй периоды времени равны, соответственно,

$$\bar{x}_1 = (1-\alpha)\, x_0 + \alpha x_1,\ \ \bar{x}_2 = (1-\alpha)\, x_1 + \alpha x_2.$$

Темп роста средней величины типа запаса $\bar{\lambda} = \bar{x}_2/\bar{x}_1$ можно выразить через средние взвешенные темпов роста за каждый из двух периодов времени $\lambda_1 = x_1/x_0$, $\lambda_2 = x_2/x_1$ следующим образом:

$$\bar{\lambda} = \alpha_1^1\lambda_1 + \alpha_2^1\lambda_2,\ \text{ где }\ \alpha_1^1 = \frac{(1-\alpha)\, x_0}{\bar{x}_1},\ \alpha_2^1 = \frac{\alpha x_1}{\bar{x}_1},\ \alpha_1^2 + \alpha_2^1 = 1,\ \text{ или}$$

$$\bar{\lambda} = \frac{1}{\alpha_1^2/\lambda_1 + \alpha_2^2/\lambda_2},\ \text{ где }\ \alpha_1^2 = \frac{(1-\alpha)\, x_1}{\bar{x}_2},\ \alpha_2^2 = \frac{\alpha x_2}{\bar{x}_2},\ \alpha_1^2 + \alpha_2^2 = 1.$$

Таким образом, темп роста средней хронологической является средней взвешенной **арифметической** темпов роста за отдельные периоды, если веса рассчитываются по информации первого периода, или средней взвешенной **гармонической**, если веса рассчитываются по информации второго периода.

Если коэффициент α, представляющий внутрипериодную динамику, различается по периодам, т.е. динамика величины в разных периодах качественно различна, то темп роста средней хронологической перестает быть средней арифметической или гармонической темпов роста по периодам, т.к. сумма весов при этих темпах роста не будет в общем случае равняться единице.

В разных ситуациях средние темпы роста могут рассчитываться различным образом, что можно проиллюстрировать на простых примерах, взятых из **финансовых расчетов**.

В финансовых расчетах аналогом темпа прироста капитала (величины типа запаса) выступает **доходность** на вложенный (инвестированный) капитал.

Пусть инвестированный капитал x_0 в течение периода τ приносит доход Δ. Тогда капитал к концу периода становится равным $x_1 = x_0 + \Delta$, и доходность капитала за этот период определяется как

$$\delta = \frac{\Delta}{x_0} = \frac{x_1}{x_0} - 1, \quad \text{т.е. совпадает по форме с темпом прироста.}$$

Средняя за период доходность в зависимости от поведения инвестора (субъекта, вложившего капитал) рассчитывается различным образом. Ниже рассматривается три возможные ситуации.

1) Если позиция инвестора **пассивна**, и он не реинвестирует полученные доходы в течение данного периода времени, то средняя доходность в единицу времени определяется простейшим способом:

$$\bar{\delta} = \frac{1}{\tau}\frac{\Delta}{x_0}.$$

Фактически это — средняя арифметическая простая, т.к. Δ/x_0 является общей доходностью за период времени τ. Такой способ расчета средней доходности наиболее распространен.

Эта формула используется и при $\tau < 1$. Так, обычно доходности за разные периоды времени приводятся к среднегодовым, т.е. единицей времени является год. Пусть речь идет, например, о трехмесячном депозите. Тогда $\tau = 0.25$, и среднегодовая доходность получается умножением на 4 доходности Δ/x_0 за 3 месяца.

2) Пусть доходность в единицу времени $\bar{\delta}$ в течение рассматриваемого периода времени не меняется, но доходы полностью **реинвестируются** в начале каждой единицы времени. Тогда за каждую единицу времени капитал возрастает в $1 + \bar{\delta}$ раз, и для нахождения $\bar{\delta}$ используется формула:

$$1 + \frac{\Delta}{x_0} = \left(1 + \bar{\delta}\right)^{\tau}, \text{ т.е. } \bar{\delta} = \left(1 + \frac{\Delta}{x_0}\right)^{\frac{1}{\tau}} - 1.$$

Эта формула справедлива при целых положительных τ. Действительно (предполагается, что начало периода инвестирования имеет на оси времени целую координату), если $\tau < 1$, ситуация аналогична предыдущей, в которой используется формула простой средней арифметической. Если τ не целое, то такая же проблема возникает для последней, неполной единицы времени в данном периоде.

Естественно предположить, что $\bar{\delta} < 1$, тогда $\left(1 + \bar{\delta}\right)^{\tau} > 1 + \tau\bar{\delta}$ (что следует из разложения показательной функции в степенной ряд) и $\dfrac{1}{\tau}\dfrac{\Delta}{x_0} > \bar{\delta}$.

Это соотношение лучше интерпретировать «в обратном порядке»: если по условиям инвестиционного контракта $\bar{\delta}$ фиксирована и допускается реинвестирование

доходов в течение периода, чем пользуется инвестор, то фактическая доходность на инвестированный капитал будет выше объявленной в контракте.

3) Пусть в течение данного периода времени доходы реинвестируются n раз через равные промежутки времени. Тогда для $\bar{\delta}$ справедлива следующая формула:

$$1 + \frac{\Delta}{x_0} = \left(1 + \frac{\tau\bar{\delta}}{n}\right)^n$$

(она совпадает с предыдущей в случае $n = \tau$).

Теоретически можно представить ситуацию **непрерывного реинвестирования**, когда $n \to \infty$. В таком случае

$$\bar{\delta} = \frac{1}{\tau}\ln\left(1 + \frac{\Delta}{x_0}\right), \text{ поскольку } \lim_{n\to\infty}\left(1 + \frac{\tau\bar{\delta}}{n}\right)^n = e^{\tau\bar{\delta}}.$$

В соответствии с введенной ранее терминологией, это — непрерывный темп прироста в единицу времени. Данную формулу можно использовать при любом (естественно, положительном) τ.

Понятно, что средние доходности в единицу времени, полученные в рассмотренных трех случаях, находятся в следующем соотношении друг с другом:

$$\underset{\substack{\text{пассивное}\\\text{поведение}}}{\frac{1}{\tau}\frac{\Delta}{x_0}} > \underset{\substack{\text{дискретное}\\\text{реинвестирование}}}{\left(1 + \frac{\Delta}{x_0}\right)^{\frac{1}{\tau}} - 1} > \underset{\substack{\text{непрерывное}\\\text{реинвестирование}}}{\frac{1}{\tau}\ln\left(1 + \frac{\Delta}{x_0}\right)}.$$

Это соотношение при интерпретации в «обратном порядке» означает, что чем чаще реинвестируется доход, тем выше фактическая доходность на первоначальный капитал. В финансовых расчетах для приведения доходностей к разным единицам времени используется 1-я формула.

Теперь рассматривается общий случай с $N+1$ моментом времени и расчетом средней доходности за N подпериодов.

1) Если позиция инвестора пассивна в течение всего периода времени, то средняя доходность в i-м подпериоде и в целом за период равны:

$$\bar{\delta}_i = \frac{1}{\tau_i}\frac{\Delta_i}{x_0}, \quad \bar{\delta} = \frac{1}{\tau}\frac{\Delta}{x_0},$$

где $\Delta_i = x_i - x_{i-1}$, $\Delta = \sum_{i=1}^{N}\Delta_i = x_N - x_0$ (τ и τ_i определены выше). Средняя доходность в целом за период удовлетворяет формуле средней взвешенной арифметической:

$$\bar{\delta} = \sum \alpha_i\bar{\delta}_i, \text{ где } \alpha_i = \frac{\tau_i}{\tau}.$$

2) Пусть теперь доходы реинвестируются в начале каждого подпериода времени. Тогда в течение i-го подпериода капитал вырастает в $1+\tau_i\bar{\delta}_i$ раз, где $\bar{\delta}_i = \dfrac{1}{\tau_i}\dfrac{\Delta_i}{x_{i-1}}$. Если предположить, что все подпериоды имеют одинаковую длину $\bar{\tau}$, то в среднем за подпериод доход вырастает в $\left(\prod\left(1+\bar{\tau}\bar{\delta}_i\right)\right)^{1/N}$ раз, и это количество раз равно $1+\bar{\tau}\bar{\delta}$. Поэтому

$$\bar{\delta} = \frac{1}{\bar{\tau}}\left(\left(\prod\left(1+\bar{\tau}\bar{\delta}_i\right)\right)^{1/N} - 1\right).$$

Это формула простой средней приведенного выше общего вида $f^{-1}\left(\frac{1}{N}\sum f\left(x_i\right)\right)$, где $f = \ln\left(1+x\right)$.

Аналогичную формулу можно использовать и в случае подпериодов разной длины τ_i:

$$\bar{\delta} = \frac{1}{\bar{\tau}}\left(\left(\prod\left(1+\tau_i\bar{\delta}_i\right)\right)^{\frac{1}{N}} - 1\right), \text{ где } \bar{\tau} = \frac{1}{N}\sum\tau_i.$$

Фактически эти формулы являются вариантами формул простой средней геометрической.

3) Пусть теперь все τ_i являются целыми положительными числами, и реинвестирование доходов происходит в начале каждой единицы времени. Тогда

$$\bar{\delta}_i = \left(1+\frac{\Delta_i}{x_{i-1}}\right)^{1/\tau_i} - 1,\ \bar{\delta} = \left(1+\frac{\Delta}{x_0}\right)^{1/\tau} - 1.$$

Средняя в единицу времени доходность в целом за период равна средней взвешенной геометрической средних доходностей по подпериодам:

$$\bar{\delta} = \prod\left(1+\bar{\delta}_i\right)^{\alpha_i} - 1, \text{ где } \alpha_i = \frac{\tau_i}{\tau}.$$

4) Наконец, в теоретическом случае непрерывного инвестирования

$$\bar{\delta}_i = \frac{1}{\tau_i}\ln\left(1+\frac{\Delta_i}{x_{i-1}}\right), \bar{\delta} = \frac{1}{\tau}\ln\left(1+\frac{\Delta}{x_0}\right),$$

и средняя доходность за весь период, как и в первом случае, равна средней взвешенной арифметической средних доходностей по подпериодам:

$$\bar{\delta} = \sum\alpha_i\bar{\delta}_i, \text{ где } \alpha_i = \frac{\tau_i}{\tau}.$$

В заключение этого раздела следует отметить, что особую роль в статистике играют средние арифметические. Именно они выступают важнейшей характеристикой распределения случайных величин. Так, в обозначениях предыдущего пункта величину $\bar{x} = \sum\alpha_i x_i$ можно записать как $\bar{x} = \sum x_i f_i \Delta_i$ или, при использовании теоретической функции плотности распределения, как $\bar{x} = \int x\, f(x)\, dx$.

Теоретическое арифметическое среднее, определенное последней формулой, называется в математической статистике **математическим ожиданием**. Математическое ожидание величины x обозначают обычно как $\mathbf{E}(x)$, сохраняя обозначение $\bar{x}$ для эмпирических средних (см. Приложение А.3.1).

2.3. Медиана, мода, квантили

Мода и медиана, наряду со средней, являются характеристиками центра распределения признака. **Медиана**, обозначаемая в данном тексте через $x_{0.5}$, — величина (детерминированная), которая «делит» совокупность пополам. Теоретически она такова, что

$$\int_{-\infty}^{x_{0.5}} f(x)\,dx = \int_{x_{0.5}}^{+\infty} f(x)\,dx = 0.5,$$

где $f(x)$ — функция распределения (см. Приложение А.3.1).

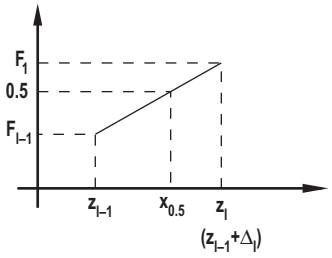

Рис. 2.8

По выборочным данным $x_1, \ldots, x_N$, упорядоченным по возрастанию, за нее принимается $x_{(N+1)/2}$ в случае, если N нечетно, и $(x_{N/2} + x_{N/2+1})/2$, если N четно.

Значение медианы может быть уточнено, если по данным выборки построено эмпирическое распределение частот z_l, $l = 0, \ldots, k$, Δ_l, α_l, f_l, F_l, $l = 1, \ldots, k$. Пусть l-й полуинтервал является медианным, т.е. $F_{l-1} < 0.5 \leqslant F_l$. Тогда, линейно интерполируя значения функции распределения F на этом полуинтервале, медиану определяют по следующей формуле:

$$x_{0.5} = z_{l-1} + \Delta_l \frac{0.5 - F_{l-1}}{\alpha_l}.$$

Ее смысл поясняется на графике (рис. 2.8). Этот график является фрагментом кумуляты.

Мода, обозначаемая в данном тексте через $\overset{o}{x}$, показывает наиболее вероятное значение признака. Это — значение величины в «пике» функции плотности распределения вероятности (см. Приложение А.3.1):

$$f\left(\overset{o}{x}\right) = \max_x f(x).$$

Величины с унимодальным распределением имеют одну моду, полимодальные распределения характеризуются несколькими модами. Непосредственно по выборке, если все ее значения различны, величину моды определить невозможно. Если какое-то значение встречается в выборке несколько раз, то именно его — по определению — принимают за моду. В общем случае моду ряда наблюдений находят по данным эмпирического распределения частот.

Пусть l-й полуинтервал является модальным, т.е. $f_l > f_{l-1}$ и $f_l > f_{l+1}$ (во избежание непринципиальных уточнений случай « $\geqslant$ » не рассматривается). Функция плотности вероятности аппроксимируется параболой, проходящей через середины ступенек гистограммы, и ее максимум определяет положение искомой моды. График (рис. 2.9) поясняет сказанное. В случае если размеры полуинтервалов Δ_{l-1}, Δ_l и Δ_{l+1} одинаковы и равны Δ, такая процедура приводит к определению моды по формуле:

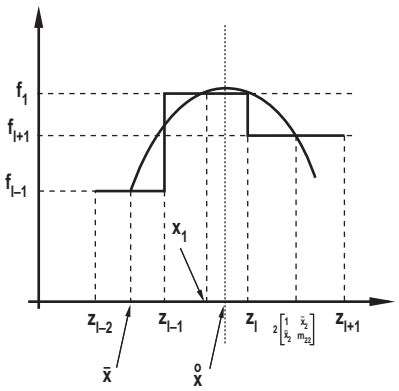

Рис. 2.9

$$\overset{o}{x} = z_{l-1} + \Delta \frac{f_l - f_{l-1}}{(f_l - f_{l-1}) + (f_l - f_{l+1})}.$$

В справедливости этой формулы несложно убедиться. Действительно, коэффициенты a, b и c аппроксимирующей параболы $ax^2 + bx + c$ удовлетворяют следующей системе уравнений:

$$\begin{cases} a\bar{x}_{l-1}^2 + b\bar{x}_{l-1} + c = f_{l-1}, \\ a(\bar{x}_{l-1} + \Delta)^2 + b(\bar{x}_{l-1} + \Delta) + c = f_l, \\ a(\bar{x}_{l-1} + 2\Delta)^2 + b(\bar{x}_{l-1} + 2\Delta) + c = f_{l+1}. \end{cases}$$

Если из второго уравнения вычесть первое, а затем третье, то получится более простая система из двух уравнений:

$$\begin{cases} \Delta(a(2\bar{x}_{l-1} + \Delta) + b) = f_l - f_{l-1}, \\ \Delta(a(-2\bar{x}_{l-1} - 3\Delta) - b) = f_l - f_{l+1}. \end{cases}$$

Первое из этих уравнений дает выражение для b через a:

$$b = \frac{f_l - f_{l-1}}{\Delta} - a\,(2\bar{x}_{l-1} + \Delta),$$

а их сумма — выражение для определения параметра a:

$$-2a\Delta^2 = (f_l - f_{l-1}) + (f_l - f_{l+1}).$$

Очевидно, что a отрицательно, и поэтому парабола имеет максимум в точке $-b/2a$ (в этой точке производная $2ax+b$ равна нулю), т.е. $\overset{o}{x}= -b/2a$, и после подстановки сюда полученных выражений для b и a, учитывая, что $\bar{x}_{l-1} + \dfrac{\Delta}{2} = z_{l-1}$, получается искомая формула.

Все три характеристики центра распределения: мода, медиана, среднее — находятся в определенных соотношениях между собой.

В случае идеальной (теоретически) симметрии

$$f\left(x_{0.5}+\Delta\right) = f\left(x_{0.5}-\Delta\right) \tag{2.5}$$

при любом $\Delta \geqslant 0$, все эти три характеристики совпадают.

Доказательство этого утверждения проводится для теоретической функции плотности распределения $f(x)$, в предположении, что она является гладкой, т.е. непрерывной и непрерывно дифференцируемой.

Дифференцирование выражения (2.5) по Δ в точке 0 дает условие $f'\left(x_{0.5}\right) = = -f'\left(x_{0.5}\right)$, из чего, в силу непрерывной дифференцируемости f, следует равенство нулю производной в точке $x_{0.5}$. И поскольку распределение унимодально, то мода совпадает с медианой.

Теперь доказывается совпадение математического ожидания с медианой. Для случайной величины $x - x_{0.5}$ с той же функцией распределения плотности $f(x)$, в силу того, что $\int\limits_{-\infty}^{+\infty} f\left(x\right) = 1$, имеет место следующее тождество:

$$\mathbf{E}\left(x\right) - x_{0.5} = \int\limits_{-\infty}^{+\infty} \left(x - x_{0.5}\right) f\left(x\right) dx.$$

Его правая часть разбивается на два слагаемых и преобразуется следующим образом:

$$\mathbf{E}(x) - x_{0.5} = \int\limits_{-\infty}^{x_{0.5}} \left(x - x_{0.5}\right) f\left(x\right) dx + \int\limits_{x_{0.5}}^{+\infty} \left(x - x_{0.5}\right) f\left(x\right) dx =$$

(в первом слагаемом производится замена переменных $x - x_{0.5} = -\Delta$ и перестановка пределов интегрирования $\int\limits_{+\infty}^{0} \rightarrow -\int\limits_{0}^{+\infty}$, во 2-м слагаемом — замена переменных $x - x_{0.5} = \Delta$)

$$= -\int\limits_{0}^{+\infty} \Delta f\left(x_{0.5}-\Delta\right) d\Delta + \int\limits_{0}^{+\infty} \Delta f\left(x_{0.5}+\Delta\right) d\Delta =$$

(вводя соответствующие обозначения)

$$= -A^- + A^+. \tag{2.6}$$

Поскольку выполнено условие симметричности распределения (2.5), $A^- = A^+$ и математическое ожидание (среднее) совпадает с медианой. Это завершает рассмотрение случая симметричных распределений.

Для асимметричных распределений указанные три характеристики различаются, но так, что медиана всегда находится между средней и модой. При правой асимметрии

$$\overset{o}{x} < x_{0.5} < \bar{x},$$

при левой асимметрии, наоборот,

$$\bar{x} < x_{0.5} < \overset{o}{x}\,.$$

В этом легко убедиться. Пусть речь идет, например, о правой асимметрии. Распределение скошено в сторону больших значений случайной величины-признака, поэтому $A^- < A^+$ (это соотношение можно рассматривать в качестве определения правой асимметрии), и, в силу выполнения тождества (2.6), среднее должно превышать медиану: $x_{0.5} < \mathbf{E}(x)$, $(x_{0.5} < \bar{x})$.

Условие $A^- < A^+$ может выполняться только в случае, если при достаточно больших Δ имеет место неравенство $f(x_{0.5} + \Delta) > f(x_{0.5} - \Delta)$ (веса больших значений признака больше, чем веса равноудаленных от медианы малых значений). Но тогда для малых Δ, т.е. в окрестности медианы, должно иметь место обратное неравенство (поскольку $\int\limits_0^{+\infty} f(x_{0.5} - \Delta)\, d\Delta = \int\limits_0^{+\infty} f(x_{0.5} + \Delta)\, d\Delta = 0.5$):

$$f(x_{0.5} - \Delta) > f(x_{0.5} + \Delta),$$

а это означает, что мода смещена влево от медианы: $\overset{o}{x} < x_{0.5}$.

Проведенное рассуждение о положении моды относительно медианы не является строгим, оно предполагает как бы «плавный» переход от симметрии к правой асимметрии. При строгом доказательстве существенную роль играет предположение об унимодальности распределения.

Случай левой асимметрии рассматривается аналогично.

Для больших выборок, как правило, подтверждается еще одно утверждение об относительном расположении трех рассматриваемых характеристик: при умеренной асимметрии мода удалена от медианы на расстояние приблизительно в 2 раза большее, чем среднее. То есть

$$\left|\overset{o}{x} - x_{0.5}\right| \approx 2\,|\,\bar{x} - x_{0.5}\,|.$$

Для того чтобы легче запомнить приведенные здесь соотношения, можно использовать следующее мнемоническое правило. Порядок следования среднего, медианы и моды (при левой асимметрии) такой же, как слов **mean**, **median**, **mode** в английском словаре (при правой асимметрии порядок обратный). Причем, как и соответствующие им статистические характеристики, слово **mean** расположено в словаре ближе к **median**, чем **mode**.

Квантилем называют число (детерминированное), делящее совокупность в определенной пропорции. Так, квантиль x_F (используемое в данном тексте обозначение квантиля) делит совокупность в пропорции (верхняя часть к нижней) $1-F$ к F (см. Приложение А.3.1):

$$P(x \leqslant x_F) = F \text{ или } F(x_F) = \int\limits_{-\infty}^{x_F} f(x)\,dx = F.$$

В эмпирическом распределении все границы полуинтервалов являются квантилями: $z_l = x_{F_l}$. По данным этого распределения можно найти любой квантиль x_F с помощью приема, использованного выше при нахождении медианы. Если l-й полуинтервал является квантильным, т.е. $F_{l-1} < F \leqslant F_l$, то

$$x_F = z_{l-1} + \Delta_l \frac{F - F_{l-1}}{\alpha_l}.$$

Иногда квантилями называют только такие числа, которые делят совокупность на равные части. Такими квантилями являются, например, медиана $x_{0.5}$, делящая совокупность пополам, **квартили** $x_{0.25}$, $x_{0.5}$, $x_{0.75}$, которые делят совокупность на четыре равные части, **децили** $x_{0.1}, \ldots, x_{0.9}$, **процентили** $x_{0.01}, \ldots, x_{0.99}$.

Для совокупностей с симметричным распределением и нулевым средним (соответственно, с нулевой модой и медианой) используют понятие двустороннего квантиля $\hat{x}_F$:

$$P(-\hat{x}_F \leqslant x \leqslant \hat{x}_F) = F(\hat{x}_F) - F(-\hat{x}_F) = \int\limits_{-\hat{x}_F}^{\hat{x}_F} f(x)\,dx = F.$$

2.4. Моменты и другие характеристики распределения

Моментом q-го порядка относительно c признака x называют величину (q и c — величины детерминированные)

$$m(q,c) = \frac{1}{N}\sum_{i=1}^{N}(x_i - c)^q,$$

в случае, если она рассчитывается непосредственно по выборке;

$$m(q,c)=\sum_{l=1}^{k}\alpha_l\left(\bar{x}_l-c\right)^q=\sum_{l=1}^{k}f_l\left(\bar{x}_l-c\right)^q\Delta_l,$$

если используются данные эмпирического распределения частот;

$$\mu(q,c)=\int\limits_{-\infty}^{+\infty}f(x)\,(x-c)^q\,dx=\mathbf{E}((x-c)^q)$$

— для теоретического распределения вероятности (см. Приложение А.3.1).

В эконометрии для обозначения теоретических или «истинных» значений величины (в генеральной совокупности) часто используются буквы греческого алфавита, а для обозначения их эмпирических значений (полученных по выборке) или их оценок — соответствующие буквы латинского алфавита. Поэтому здесь в первых двух случаях момент обозначается через m, а в третьем случае — через μ. В качестве общей формулы эмпирического момента (объединяющей первые два случая) будет использоваться следующая:

$$m(q,c)=\sum_{i=1}^{N}\alpha_i\left(x_i-c\right)^q.$$

В принципе, моменты могут рассчитываться относительно любых c, однако в статистике наиболее употребительны моменты, рассчитанные при c, равном нулю или среднему. В первом случае моменты называют **начальными**, во втором — **центральными**. В расчете центральных моментов используются величины $x_i-\bar{x}$, которые часто называют **центрированными** наблюдениями и обозначают через $\hat{x}_i$.

Средняя является начальным моментом 1-го порядка:

$$\bar{x}=m(1,0),$$
$$\mathbf{E}(x)=\mu(1,0).$$

Благодаря этому обстоятельству центральные моменты при целых q всегда можно выразить через начальные моменты. Для этого надо раскрыть скобки (возвести в степень q) в выражении центрального момента.

Центральный момент 2-го порядка или 2-й центральный момент называется **дисперсией** и обозначается через s^2 (эмпирическая дисперсия) или σ^2 (теоретическая дисперсия):

$$s^2=m(2,\,\bar{x}),$$
$$\sigma^2=\mu(2,\,\mathbf{E}(x)).$$

При вычислении дисперсии иногда удобнее пользоваться начальным моментом 2-го порядка. Связь с ним устанавливается следующим образом:

$$s^2 = \sum \alpha_i \left(x_i - \bar{x}\right)^2 = \sum \alpha_i x_i^2 - 2\bar{x} \underbrace{\sum \alpha_i x_i}_{\bar{x}} + \bar{x}^2 =$$
$$= \sum \alpha_i x_i^2 - \bar{x}^2 = m\left(2,0\right) - m^2\left(1,0\right).$$

Корень квадратный из дисперсии — s или σ — является **среднеквадратическим отклонением**, иногда (главным образом, в англоязычной литературе) его называют **стандартным** отклонением.

Величины $\hat{x}_i/s$ называют центрированными и **нормированными** наблюдениями. Они измеряют значения признака в единицах среднеквадратического отклонения от среднего. Такая шкала измерения иногда называется **стандартизованной** или **стандартизированной**.

Дисперсия (и среднеквадратическое отклонение) является мерой **абсолютного** рассеяния или разброса значений признака в совокупности. В принципе такой мерой мог бы служить 2-й момент относительно какого-то другого c, не равного $\bar{x}$, но лежащего в центральной области распределения признака. Однако используют именно дисперсию, поскольку ее величина однозначно определена и, в некотором смысле, не зависит от c. Дисперсия минимальна среди всех 2-х моментов относительно c.

Действительно, производная 2-го момента по c

$$\frac{d\left(\int (x-c)^2 f(x)dx\right)}{dc} = -2\left(\int x f(x)dx - c\int f(x)dx\right) = -2\left(\mathbf{E}(x) - c\right)$$

равна 0 в точке $c = \mathbf{E}(x)$. Это точка минимума, поскольку 2-я производная по c в ней равна 2, т.е. положительна.

В статистике используются и другие показатели разброса. Примерами показателей абсолютного разброса являются:

$\dfrac{\max x_i - \min x_i}{2}$ — **общий размах** вариации,

$\dfrac{x_{1-F} - x_F}{2}$ — **квантильный** размах вариации, где $F < 0.5$ (достаточно часто используется квартильный размах вариации, то есть этот показатель при $F = 0.25$),

$\sum \alpha_i \left|\hat{x}_i\right|$ — **среднее линейное отклонение**.

Среднее линейное отклонение имеет смысл рассчитывать не относительно среднего $\bar{x}$, а относительно медианы $x_{0.5}$, поскольку именно в таком случае оно принимает минимально возможное значение.

Действительно, производная по c среднего линейного отклонения относительно c

$$\frac{d\left(\int |x-c| f(x)\,dx\right)}{dc} = \frac{d\left(\int\limits_{-\infty}^{c} (c-x) f(x)\,dx + \int\limits_{c}^{+\infty} (x-c) f(x)\,dx\right)}{dc} = \int\limits_{-\infty}^{c} f(x)\,dx - \int\limits_{c}^{+\infty} f(x)\,dx$$

равна 0 при $c = x_{0.5}$ (2-я производная в этой точке равна $2f(x_{0.5})$ и положительна по определению функции f).

Для характеристики относительного разброса применяются различные формы коэффициента вариации. Например, он может рассчитываться как отношение среднего квадратичного отношения к среднему, общего или квантильного размаха вариации к медиане. Иногда его рассчитывают как отношения $\max x_i$ к $\min x_i$ или x_{1-F} к x_F (при $F < 0.5$).

Достаточно распространен еще один тип коэффициентов вариации, которые рассчитываются как отношения средней по верхней части совокупности к средней по нижней части совокупности.

Для того чтобы дать определение таким коэффициентам вариации, необходимо ввести понятие среднего по части совокупности.

Математическое ожидание можно представить в следующей форме:

$$\mathbf{E}(x) = F\frac{1}{F}\int\limits_{-\infty}^{x_F} x f(x)\,dx + (1-F)\frac{1}{1-F}\int\limits_{x_F}^{+\infty} x f(x)\,dx = F\,\mathbf{E}_F(x) + (1-F)\,\mathbf{E}_F^+(x).$$

Квантиль x_F делит совокупность на две части, по каждой из которых определяется свое математическое ожидание:

$\mathbf{E}_F(x)$ — по нижней части,

$\mathbf{E}_F^+(x)$ — по верхней части совокупности.

Приведенное тождество определяет связь между двумя этими математическими ожиданиями:

$$\mathbf{E}_F^+(x) = \frac{1}{1-F}\left(\mathbf{E}(x) - F\,\mathbf{E}_F(x)\right).$$

По выборке аналогичные частичные средние рассчитываются следующим образом. Пусть $x_i,\ i = 1, \ldots, N$ ряд наблюдений, упорядоченный по возрастанию. Тогда

$$F_i = \frac{i}{N},\ i = 1, \ldots, N \text{ — накопленные относительные частоты,}$$

$$\bar{x}_i = \frac{1}{i}\sum_{i'=1}^{i} x_{i'} \text{ — } i\text{-я средняя по нижней части, } i = 1, \ldots, N \quad (\bar{x}_0 = 0)\,,$$

$$\bar{x}_i^+ = \frac{1}{N-i}\sum_{i'=i+1}^{N} x_{i'} = \frac{1}{1-F_i}\,(\bar{x} - F_i\,\bar{x}_i) \text{ — } i\text{-я средняя по верхней части,}$$

$$i = 0,\ 1, \ldots, N \quad \left(\bar{x}_N^+ = 0\right).$$

Такой расчет не имеет необходимой иногда степени общности, поскольку позволяет найти частичные средние лишь для некоторых квантилей, которыми в данном случае являются сами наблюдения ($x_i = x_{F_i}$). Для квантилей x_F при любых F частичные средние находятся по данным эмпирического распределения (предполагается, что l-й полуинтервал является квантильным):

$$\bar{x}_F = \frac{1}{F}\Big(\sum_{l'=1}^{l-1}\alpha_{l'}\bar{x}_{l'} + (F - F_{l-1})\,\frac{1}{2}\,(z_{l-1} + x_F)\Big)$$

— средняя по нижней части совокупности (здесь $\frac{1}{2}\,(z_{l-1} + x_F)$ — центр последнего, неполного полуинтервала, $F - F_{l-1}$ — его вес). После подстановки выражения для квантиля x_F, полученного в предыдущем пункте, эта формула приобретает следующий вид:

$$\bar{x}_F = \frac{1}{F}\left(\sum_{l'=1}^{l-1}\alpha_{l'}\bar{x}_{l'} + (F - F_{l-1})\left(z_{l-1} + \frac{F - F_{l-1}}{2\alpha_i}\Delta_l\right)\right).$$

При расчете средней по верхней части совокупности проще воспользоваться полученной выше формулой:

$$\bar{x}_F^+ = \frac{1}{1-F}\,(\bar{x} - F\,\bar{x}_F).$$

Для расчета квантильного коэффициента вариации совокупность делится на 3 части: верхняя часть, объемом не более половины, нижняя часть такого же объема и средняя часть, не используемая в расчете. Данный коэффициент, называемый $F \times 100$-процентным (например, 15-процентным), рассчитывается как отношение средних по верхней и нижней части совокупности:

$$\frac{\bar{x}_{1-F}^+}{\bar{x}_F} = \frac{\bar{x} - (1-F)\,\bar{x}_{1-F}}{F\,\bar{x}_F},$$

$$\left(\frac{\mathbf{E}_{1-F}^+(x)}{\mathbf{E}_F(x)} = \frac{\mathbf{E}(x) - (1-F)\,\mathbf{E}_{1-F}(x)}{F\,\mathbf{E}_F(x)}\right), \text{ где } F \leqslant 0.5$$

При использовании непосредственно данных выборки эта формула имеет другой вид:

$$\frac{\bar{x}^{+}_{N-i}}{\bar{x}_i} = \frac{\bar{x} - (1 - F_i)\,\bar{x}_{N-i}}{F_i\,\bar{x}_i}, \text{ где } i \leqslant \left[\frac{N}{2}\right].$$

Такие коэффициенты вариации называют иногда, как и соответствующие квантили, медианными, если $F = 0.5$, квартильными, если $F = 0.25$, децильными, если $F = 0.1$, процентильными, если $F = 0.01$. Наиболее употребительны децильные коэффициенты вариации.

При расчете коэффициентов вариации в любой из приведенных форм предполагается, что характеризуемый признак может принимать только неотрицательные значения.

Существует еще один — графический — способ представления степени разброса значений признака в совокупности. Он используется для совокупностей объемных признаков, принимающих положительные значения. Это — кривая Лоренца или кривая концентрации. По абсциссе расположены доли накопленной частоты, по ординате — доли накопленного суммарного признака. Она имеет вид, изображенный на графике (рис. 2.10). Чем более выпукла кривая, тем сильнее дифференцирован признак.

Рис. 2.10

По оси абсцисс кривой Лоренца расположены значения величины $F \times 100\%$, по оси ординат — в случае использования теоретического распределения — значения величины:

$$\frac{\int\limits_0^{x_F} x f(x)\,dx}{\int\limits_0^{+\infty} x f(x)\,dx} \times 100\%$$

(предполагается, что $x \geqslant 0$), или, используя введенные выше обозначения для частичных средних,

$$F\frac{\mathbf{E}_F(x)}{\mathbf{E}(x)} \times 100\%.$$

При использовании данных эмпирического распределения по оси ординат расположены значения величины

$$F\frac{\bar{x}_F}{\bar{x}} \times 100\%.$$

При построении кривой непосредственно по данным ряда наблюдений сначала на графике проставляются точки

$$\left(F_i \times 100\%,\ F_i\frac{\bar{x}_i}{\bar{x}} \times 100\%\right),\ \ i = 1, \ldots, N,$$

а затем они соединяются отрезками прямой линии.

В случае, если значение признака в совокупности не варьируется, средние по всем ее частям одинаковы, и кривая Лоренца является отрезком прямой линии (пунктирная линия на рис. 2.10). Чем выше вариация значений признака, тем более выпукла кривая. Степень ее выпуклости или площадь выделенной на рисунке области может являться мерой относительного разброса.

Кривую Лоренца принято использовать для иллюстрации распределения дохода или имущества в совокупностях людей, представляющих собой население отдельных стран или регионов. Отсюда ее второе название — кривая концентрации. Она выражает степень концентрации богатства в руках меньшинства.

В статистике центральные моменты q-го порядка обычно обозначаются через m_q (μ_q — для теоретических величин):

$$m_q = m(q, \bar{x}) \quad (\mu_q = \mu(q, \mathbf{E}(x)).$$

Нормированный центральный момент 3-го порядка

$$d_3 = \frac{m_3}{s^3} \quad \left(\delta_3 = \frac{\mu_3}{\sigma^3}\right)$$

часто используется как мера асимметрии (скошенности) распределения. Если распределение симметрично, то этот показатель равен нулю. В случае его положительности считается, что распределение имеет правую асимметрию, при отрицательности — левую асимметрию (см. Приложение А.3.1).

Следует иметь в виду, что такое определение левой и правой асимметрии может не соответствовать определению, данному в предыдущем пункте. Возможны такие ситуации, когда распределение имеет правую асимметрию, и среднее превышает медиану, но данный показатель отрицателен. И наоборот, среднее меньше медианы (левая асимметрия), но этот показатель положителен.

В этом можно убедиться, рассуждая следующим образом.

Пусть $\varphi(x)$ — функция плотности вероятности симметричного относительно нуля распределения с дисперсией σ^2, т.е.

$$\int\limits_{-\infty}^{+\infty} x\varphi\,(x)\,dx = 0, \quad \int\limits_{-\infty}^{+\infty} x^2\varphi\,(x)\,dx = \sigma^2, \quad \int\limits_{-\infty}^{+\infty} x^3\varphi\,(x)\,dx = 0,$$

$$\int\limits_{-\infty}^{0} \varphi\,(x)\,dx = \int\limits_{0}^{+\infty} \varphi\,(x)\,dx = 0.5, \quad \varphi(x) = \varphi(-x).$$

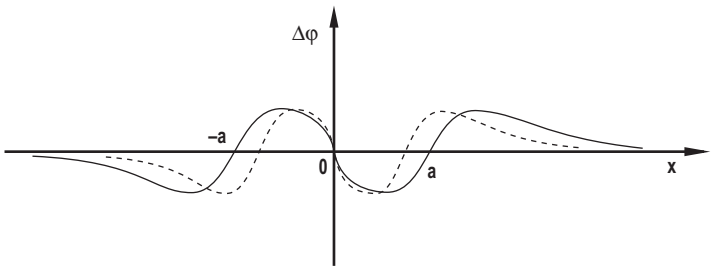

Рис. 2.11

Рассматривается случайная величина x, имеющая функцию плотности вероятности $f(x) = \varphi(x) + \gamma\Delta\varphi(x)$.

Функция $\Delta\varphi$ вносит асимметрию в распределение x. Ее график имеет вид — сплошная линия на рисунке 2.11, а свойства таковы:

$$\Delta\varphi(x) = -\Delta\varphi(-x), \int\limits_{-\infty}^{+\infty} \Delta\varphi\,(x)\,dx = 0, \int\limits_{-\infty}^{0} \Delta\varphi\,(x)\,dx = \int\limits_{0}^{+\infty} \Delta\varphi\,(x)\,dx = 0.$$

Параметр γ не должен быть слишком большим по абсолютной величине, чтобы сохранялась унимодальность распределения (и, конечно же, неотрицательность функции плотности).

Можно обозначить

$$-\int\limits_{0}^{a} \Delta\varphi\,(x)\,dx = \int\limits_{0}^{+\infty} \Delta\varphi\,(a+x)\,dx = S > 0$$

и определить величины a_1 и a_2 :

$$\int\limits_{0}^{a} x\Delta\varphi\,(x)\,dx = -a_1 S, \quad \int\limits_{0}^{+\infty} x\Delta\varphi\,(a+x)\,dx = a_2 S.$$

Понятно, что a_1 — математическое ожидание случайной величины, заданной на отрезке $[0, a]$ и имеющей плотность распределения $-\frac{1}{S}\Delta\varphi\,(x)$, поэтому $0 < a_1 < a$. Аналогично, a_2 — математическое ожидание случайной величины, заданной на отрезке $[0, \infty]$ с плотностью вероятности $\frac{1}{S}\Delta\varphi\,(a+x)$, поэтому $0 < a_2$.

Теперь легко видеть, что (вводя дополнительное обозначение a_3)

$$\int\limits_0^{+\infty} x\Delta\varphi(x)\,dx = \underbrace{\int\limits_0^{a} x\Delta\varphi(x)\,dx}_{-a_1 S} + \int\limits_a^{+\infty} x\Delta\varphi(x)\,dx \overset{x=a+y}{=} -a_1 S + a\underbrace{\int\limits_0^{+\infty}\Delta\varphi(a+y)\,dy}_{S} +$$

$$+ \underbrace{\int\limits_0^{+\infty} y\Delta\varphi\,(a+y)\,dy}_{a_2 S} = S\,(-a_1 + a + a_2) = a_3 > 0.$$

Аналогичным образом можно доказать, что

$$\int\limits_0^{+\infty} x^3\Delta\varphi\,(x)\,dx = a_4 > 0.$$

Прибавление $\gamma\Delta\varphi$ к φ не меняет медиану, т.к.

$$\int\limits_0^{+\infty} f\,(x)\,dx = \underbrace{\int\limits_0^{+\infty}\varphi\,(x)\,dx}_{0.5} + \gamma\underbrace{\int\limits_0^{+\infty}\Delta\varphi\,(x)\,dx}_{0} = 0.5,$$

но сдвигает среднее (из нуля):

$$\mathbf{E}\,(x) = \int\limits_{-\infty}^{+\infty} xf\,(x)\,dx = \underbrace{\int\limits_{-\infty}^{+\infty} x\varphi\,(x)\,dx}_{0} + \gamma\int\limits_{-\infty}^{+\infty} x\Delta\varphi\,(x)\,dx =$$

$$= \gamma\left(\underbrace{\int\limits_{-\infty}^{0} x\Delta\varphi\,(x)\,dx}_{a_3} + \underbrace{\int\limits_0^{+\infty} x\Delta\varphi\,(x)\,dx}_{a_3}\right) = 2\gamma a_3.$$

Таким образом, в соответствии с данным выше определением, если $\gamma > 0$, распределение имеет правую асимметрию (увеличивается плотность вероятности больших значений признака), и среднее, будучи положительным, выше медианы. Если $\gamma < 0$, распределение характеризуется левой асимметрией, и среднее ниже медианы.

Теперь находится 3-й центральный момент:

$$\mu_3 = \int\limits_{-\infty}^{+\infty} (x - \mathbf{E}(x))^3 f(x)\,dx =$$

$$= \int\limits_{-\infty}^{+\infty} x^3 f(x)\,dx - 3\mathbf{E}(x) \int\limits_{-\infty}^{+\infty} x^2 f(x)\,dx + \underbrace{3\mathbf{E}^2(x) \int\limits_{-\infty}^{+\infty} x f(x) dx}_{2\mathbf{E}^3(x)} - \mathbf{E}^3(x) =$$

$$= \underbrace{\int\limits_{-\infty}^{+\infty} x^3 \varphi(x) dx}_{0} + \gamma \underbrace{\int\limits_{-\infty}^{+\infty} x^3 \Delta\varphi(x) dx}_{2a_4} - 3\mathbf{E}(x) \left(\underbrace{\int\limits_{-\infty}^{+\infty} x^2 \varphi(x) dx}_{\sigma^2} + \gamma \underbrace{\int\limits_{-\infty}^{+\infty} x^2 \Delta\varphi(x) dx}_{0} \right) +$$

$$+ 2\mathbf{E}^3(x) \overset{\mathbf{E}(x)=2\gamma a_3}{=} 2\gamma(a_4 - 3a_3\sigma^2 + 8\gamma^2 a_3^3 = 2\gamma\,(D + R),$$

где $D = a_4 - 3a_3\sigma^2,\ R = 8\gamma^2 a_3^3$.

Второе слагаемое в скобках — R — всегда положительно, и, если D (первое слагаемое) неотрицательно, то введенный показатель асимметрии «работает» правильно: если он положителен, то асимметрия — правая, если отрицателен, то — левая. Однако D может быть отрицательным. Это легко показать.

Пусть при заданном $\Delta\varphi$ эта величина положительна (в этом случае $\dfrac{a_4}{3a_3\sigma^2} > 1$). Сжатием графика этой функции к началу координат (пунктирная линия на рис. 2.11) всегда можно добиться смены знака данной величины.

Преобразованная (сжатая) функция асимметрии $\Delta\tilde{\varphi}$ связана с исходной функцией следующим образом:

$$\Delta\tilde{\varphi}(x) = \Delta\varphi(kx), \text{ где } k > 1.$$

Свойства этой новой функции те же, что и исходной, и поэтому все проведенные выше рассуждения для новой случайной величины с функцией плотности $\varphi + \gamma\Delta\tilde{\varphi}$ дадут те же результаты. Новая величина D, обозначаемая теперь $\tilde{D}$, связана с исходными величинами следующим образом:

$$\tilde{D} = \tilde{a}_4 - 3\tilde{a}_3\sigma^2 = \frac{1}{k^2}\left(\frac{1}{k^2}a_4 - 3a_3\sigma^2\right)$$

$$\left(\text{например, } \tilde{a}_3 = \int\limits_0^{+\infty} x\Delta\varphi(kx)\,dx \overset{kx=y,\ x=\frac{1}{k}y,\ dx=\frac{1}{k}dy}{=} \frac{1}{k^2}\int\limits_0^{+\infty} y\Delta\varphi(y)\,dy = \frac{1}{k^2}a_3\right)$$

и при $k > \sqrt{\dfrac{a_4}{3a_3\sigma^2}} > 1$ она отрицательна.

Таблица 2.1

X	-3	-2	-1	0	1	2	3
φ	0.0625	0.125	0.1875	0.25	0.1875	0.125	0.0625
$\Delta\varphi$	0	-1	1	0	-1	1	0
$\Delta\tilde{\varphi}$	-0.2	-1	1	0	-1	1	0.2

В такой ситуации (если γ достаточно мал, и вслед за $\tilde{D}$ отрицательно и $\tilde{D}+\tilde{R}$) 3-й центральный момент оказывается отрицательным при правой асимметрии и положительным при левой асимметрии.

Можно привести числовой пример совокупности с правой асимметрией, 3-й центральный момент которой отрицателен. Исходные данные приведены в таблице 2.1.

При $\gamma = 0.03$ среднее равно 0.06 (превышает медиану, равную 0), а 3-й центральный момент равен -0.187. Но стоит немного растянуть функцию асимметрии от начала координат (последняя строка таблицы), как ситуация приходит в норму. При том же γ среднее становится равным 0.108, а 3-й центральный момент равен $+0.097$.

Проведенный анализ обладает достаточной степенью общности, т.к. любую функцию плотности вероятности f можно представить как сумму функций φ и $\Delta\varphi$ с указанными выше свойствами (при этом $\gamma = 1$). Эти функции определяются следующим образом (предполагается, что медиана для функции f равна 0):

$$\varphi(x) = \frac{1}{2}\left(f(x) + f(-x)\right),\ \Delta\varphi(x) = \frac{1}{2}\left(f(x) - f(-x)\right).$$

Таким образом, если асимметрия «сосредоточена» вблизи от центра распределения (функция асимметрии $\Delta\varphi$ достаточно «поджата» к медиане), то 3-й центральный момент не может играть роль показателя асимметрии.

Надежным показателем асимметрии является величина $\dfrac{(\bar{x} - \overset{o}{x})}{s}$ или, учитывая приведенную в предыдущем пункте эмпирическую закономерность в расположении моды, медианы и среднего, $\dfrac{3(\bar{x} - x_{0.5})}{s}$.

Достаточно употребителен также квартильный коэффициент асимметрии, рассчитываемый как отношение разности квартильных отклонений от медианы к их сумме:

$$\frac{(x_{0.75} - x_{0.5}) - (x_{0.5} - x_{0.25})}{(x_{0.75} - x_{0.5}) + (x_{0.5} - x_{0.25})} = \frac{x_{0.25} + x_{0.75} - 2x_{0.5}}{x_{0.75} - x_{0.25}}.$$

Эти три коэффициента положительны при правой асимметрии и отрицательны при левой. Для симметричных распределений значения этих коэффициентов близки к нулю. Здесь требуется пояснить, что означает «близки к нулю».

Рассчитанные по выборке, значения этих коэффициентов — пусть они обозначаются через K^c (^{c} — calculated) — не могут в точности равняться нулю, даже если истинное распределение в генеральной совокупности симметрично. Как и исходные для их расчета выборочные данные, эти коэффициенты являются случайными величинами K с определенными законами распределения. Эти законы (в частности, функции плотности вероятности) известны в теории статистики, если справедлива **нулевая гипотеза**, в данном случае — если истинное распределение симметрично. А раз известна функция плотности, то можно определить область, в которую с наибольшей вероятностью должно попасть расчетное значение коэффициента K^c в случае справедливости нулевой гипотезы. Эта область, называемая **доверительной**, выделяется квантилем K_F с достаточно большим F. Обычно принимают $F = 0.95$. В данном случае K могут быть как положительными, так и отрицательными, их теоретическое распределение (при нулевой гипотезе) симметрично относительно нуля, и использоваться должен двусторонний квантиль.

Если расчетное значение K^c попадает в доверительную область, т.е. оно по абсолютной величине не превосходит K_F, то нет оснований считать, что истинное распределение не симметрично, и нулевая гипотеза не отвергается. На основании этого не следует делать вывод о симметричности истинного распределения. Установлено только то, что наблюдаемые факты не противоречат симметричности. Другими словами, если распределение симметрично, то расчетное значение попадает в доверительную область. Но обратное может быть не верным.

Если расчетное значение не попадает в доверительную область или, как говорят, попадает в **критическую** область, то маловероятно, что величина K имеет принятое (при нулевой гипотезе) распределение, и нулевая гипотеза отвергается с вероятностью ошибки (1-го рода) $1 - F$ (обычно 0.05). Причем если $K^c > K_F$, то принимается гипотеза о правой асимметрии, если $K^c < -K_F$, то принимается гипотеза о левой асимметрии.

Границы доверительной (критической) области зависят от числа наблюдений. Чем больше наблюдений, тем меньше K_F, при прочих равных условиях, т.е. тем у́же доверительная область — область «нуля». Это означает, что чем больше использовано информации, тем точнее, при прочих равных условиях, сделанные утверждения.

Таким образом, фраза «K^c близко к нулю» означает, что $|K^c| \leqslant K_F$.

Приведенные здесь рассуждения используются в теории статистики при **проверке статистических гипотез**, или **тестировании** (по англоязычной терминологии), а также при построении **доверительных интервалов** (областей).

Подробнее о проверке гипотез см. Приложение А.3.3.

Нормированный центральный момент 4-го порядка

$$d_4 = \frac{m_4}{s^4} \quad \left(\delta_4 = \frac{\mu_4}{\sigma^4}\right)$$

называется **куртозисом** (от греческого слова $\kappa\upsilon\rho\tau o\iota$ — горбатый). По его величине судят о высоковершинности унимодального распределения. Если распределение близко к нормальному, то этот показатель равен приблизительно 3 («приблизительно» понимается в том же смысле, что и «близко к нулю» в предыдущем случае). Если $r_4 > 3$, то распределение высоковершинное, в противном случае — низковершинное. На этом основании вводится показатель, называемый **эксцессом** (см. Приложение А.3.1):

$$d_4 - 3 \quad (\delta_4 - 3).$$

Его используют для оценки высоковершинности распределения, сравнивая с 0.

Граничным для куртозиса является число 3, поскольку для нормального распределения он равен точно 3.

Действительно, плотность $f(x)$ нормально распределенной с математическим ожиданием $\bar{x}$ и дисперсией σ^2 случайной величины x равна

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}.$$

В «Справочнике по математике» И.Н. Бронштейна и К.А. Семендяева (М., 1962) на стр. 407 можно найти следующую формулу:

$$\int_0^\infty x^n e^{-ax^2} dx = \frac{\Gamma\left(\frac{n+1}{2}\right)}{2a^{\frac{n+1}{2}}}, \text{ при } a > 0 \text{ и } n > 1,$$

где Γ — гамма-функция, обладающая следующими свойствами:

$$\Gamma(x+1) = x\Gamma(x),$$
$$\Gamma(n) = (n-1)!, \text{ при } n \text{ целом и положительном},$$
$$\Gamma(x)\Gamma\left(x+\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2^{2x-1}}\Gamma(2x).$$

Отсюда легко установить, что при целом и четном q

$$\mu_q = 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (q-1)\,\sigma^q = (q-1)!! \cdot \sigma^q \text{ и, в частности, } \mu_4 = 3\sigma^4.$$

О свойствах нормального распределения см. Приложение А.3.2.

В практике статистики моменты более высоких порядков используются крайне редко.

2.5. Упражнения и задачи

Упражнение

На основании данных о росте студентов курса построить ряд распределения, дать табличное и графическое его изображение (представив на графике гистограмму, полигон, кумуляту). Какие из графиков соответствуют эмпирической функции плотности распределения вероятности, а какие — эмпирической функции распределения вероятности? Изобразить на графике гистограммы положение моды, медианы и средней арифметической. Подтвердить их соотношения расчетами характеристик центра распределения. Найти дисперсию, коэффициент вариации, а также показатели асимметрии и эксцесса. Оценить степень однородности элементов совокупности.

Задачи

1. Определить пункты, которые являются выпадающими из общего ряда.

   1.1 а) частота, б) плотность, в) гистограмма, г) график;

   1.2 а) арифметическое, б) геометрическое, в) алгебраическое, г) квадратическое;

   1.3 а) мода, б) медиана, в) квантиль, г) квартиль;

   1.4 а) бимодальное, б) нормальное, в) асимметричное, г) U-образное;

   1.5 а) математическое ожидание, б) биномиальное, в) нормальное, г) среднее;

   1.6 а) момент, б) период, в) дисперсия, г) среднее;

   1.7 а) центральный, б) начальный, в) исходный, г) момент.

2. Количественный признак принимает значения 2, 3, 4, 9. Какова плотность относительной частоты 2-го и 3-го элемента?

3. Распределение семей по доходам (в условных единицах в месяц) представлено в группированном виде количеством N_l семей, попавших в l полуинтервал $(z_{l-1}; z_l]$ (табл. 2.2).

   Заполните в таблице недостающие характеристики.

   Изобразите графики гистограммы, полигона и кумуляты.

4. Какова средняя хронологическая величин 1, 2, 5, 9, характеризующих последовательность равных промежутков времени?

Таблица 2.2

$(z_{l-1}; z_l]$	500;700	700;900	900;1100	1100;1300	1300;1500
N_l	4	8	5	2	1
α_l					
F_l					
f_l					

5. На что нужно поделить $y_1 - y_0$, чтобы получить среднюю хронологическую на временном отрезке $[0, 1]$?

6. Чему равны простые средние: геометрическая, арифметическая, гармоническая чисел $1,\ 2,\ 4$?

7. Три объекта характеризуются следующими относительными признаками: $1/6$, $1/3$, $1/4$. Веса этих объектов по числителю равны $0.1,\ 0.2,\ 0.7$, вес первого объекта по знаменателю — 0.15. Чему равен вес второго объекта?

8. Какая из двух величин

$$\frac{(a+b+c)}{3}, \text{ или } \frac{3}{\left(\frac{1}{a}+\frac{1}{b}+\frac{1}{c}\right)}$$

больше и почему?

9. Капитал за первый год не изменился, за второй — вырос на 12%. Среднегодовой коэффициент, одинаковый по годам, равен $3/8$. Каков темп роста среднегодового капитала?

10. За первое полугодие капитал вырос на 12.5%, за второе — в 2 раза. Какова среднегодовая доходность (в процентах), если позиция инвестора была пассивной, или если он реинвестировал доход в середине года?

11. Совокупность предприятий была разделена на группы в зависимости от величины стоимости реализованной продукции. Количество предприятий в каждой группе и среднее значение стоимости реализованной продукции в каждой группе даны в таблице:

Номер группы	1	2	3	4	5
Количество предприятий в группе (ед.)	4	4	5	7	5
Среднее значение стоимости реализованной продукции (ден. ед.)	15	20	25	30	35

Определить среднюю стоимость реализованной продукции по совокупности предприятий в целом.

12. По металлургическому заводу имеются следующие данные об экспорте продукции:

Вид продукции	Доля вида продукции в общей стоимости реализованной продукции, %	Удельный вес продукции на экспорт, %
Чугун	25	35
Прокат листовой	75	25

Определить средний удельный вес продукции на экспорт.

13. Совокупность населенных пунктов области была разделена на группы в зависимости от численности безработных. Количество населенных пунктов в каждой группе и средняя численность безработных в каждой группе даны в таблице:

Номер группы	1	2	3	4	5
Количество населенных пунктов в группе	4	8	2	3	3
Средняя численность безработных	10	12	15	20	30

Определить среднюю численность безработных по совокупности населенных пунктов в целом.

14. В таблице даны величины стоимости основных фондов на конец года за ряд лет:

Год	0	1	2	3	4
Стоимость основных фондов на конец года	100	120	125	135	140

Предположим, что стоимость фондов на конец года t совпадает со стоимостью на начало года $t+1$. Среднегодовой коэффициент равен 0.3. Определить:

а) среднегодовую стоимость основных фондов в 1, 2, 3 и 4 году;

б) среднегодовой темп прироста среднегодовой стоимости основных фондов за период с 1 по 4 годы.

15. В первые два года исленность занятых в экономике возрастала в среднем на 4% в год, за следующие три — на 5% и в последние три года среднегодовые темпы роста составили 103%. Определите среднегодовые темпы роста и базовый темп прироста численности занятых за весь период.

16. В первые три года численность безработных возрастала в среднем на 2% в год, за следующие три — на 4% и в последние два года среднегодовые темпы роста составили 103% . Определите среднегодовые темпы роста и базовый темп прироста численности безработных за весь период.

17. В таблице даны величины дохода (в %), приносимые капиталом за год:

Год	1	2	3	4
Доходность	10	12	8	6

Определить среднегодовую доходность капитала в течение всего периода, если:

а) позиция инвестора пассивна;

б) позиция инвестора активна.

18. В первом квартале капитал возрастает на 20%, во втором — на 15%, в третьем — на 10%, в четвертом — на 20%. Определите среднегодовую доходность капитала, если:

а) позиция инвестора пассивна;

б) позиция инвестора активна, т.е. он ежеквартально реинвестирует доход.

19. Во сколько раз вырастает ваш капитал за год, вложенный в начале года под 20% годовых, если вы

а) не реинвестировали проценты;

б) реинвестировали их один раз в середине года;

в) реинвестировали три раза в начале каждого очередного квартала;

г) реинвестировали в каждый последующий момент времени.

В первом квартале капитал возрастает на 12%, во втором — на 15%, в третьем — на 20%, в четвертом — на 15%. Определите среднегодовую доходность капитала, если:

20. Объем продукции в 1995 г. составил 107% от объема продукции 1990 г., в течение последующих двух лет он снижался на 1% в год, потом за 4 года вырос на 9% и в течение следующих трех лет возрастал в среднем на 2% в год. На сколько процентов возрос объем продукции за вес период? На сколько процентов он возрастал в среднем в год в течение этого периода.

21. Дана функция распределения $F(x) = 1/(1 + e^{-x})$. Найти медиану и моду данного распределения.

22. В эмпирическом распределении $z_0 = 0$, все дельты $= 1$, $F_3 = 0.21$, $F_4 = 0.4$, $F_5 = 0.7$, $F_6 = 0.77$. Чему равны медиана и мода?

23. Известна гистограмма бимодального ряда наблюдений. На каком отрезке лежит медиана?

24. Медиана больше моды, где лежит среднее?

    Какая из трех характеристик центра распределения количественного признака является квантилем и каким?

    Медиана и средняя равны, соответственно, 5 и 6. Каково вероятное значение моды? Почему?

25. На основе информации о возрасте всех присутствующих на занятиях (включая преподавателя) определить характер асимметрии функции распределения?

26. Дать определение 5%-го квантиля и написать интерполяционную формулу расчет 5%-го квантиля для эмпирического распределения. Привести графическое обоснование формулы.

27. В эмпирическом распределении $z_0 = 0$, все дельты $= 1$, $F_4 = 0.4$, $F_5 = 0.7$, $F_6 = 0.8$, среднее равно 4.3. Какова асимметрия: правая $(+)$ или левая $(-)$? Чему равен 75%-ый квантиль?

28. Найти значение 30%-го квантиля, если известно эмпирическое распределение:

Границы интервалов	10–15	15–20	20–25	25–30
Частоты	1	3	4	2

29. Для ряда $1, 2, 3, 6$ найти медианный и квартильный коэффициент вариации.

30. Чему равна ордината кривой Лоренца при абсциссе $1/3$ для ряда $1, 2, 3$?

31. Чему равен медианный коэффициент вариации для ряда $1, 2, 3$?

32. Как посчитать децильный коэффициент вариации?

33. Задан ряд наблюдений за переменной x: 3, 0, 4, 2, 1. Подсчитать основные статистики данного ряда, среднее арифметическое, медиану, дисперсию (смещенную и несмещенную), показатель асимметрии и куртозиса, размах выборки.

34. Для представленных ниже комбинаций значений показателей асимметрии δ_3 и эксцесса δ_4 дать графическое изображение совокупности и указать на графике положение моды, медианы и средней арифметической:

    а) $\delta_3 > 0, \delta_4 > 3$;
    б) $\delta_3 < 0, \delta_4 > 3$;
    в) $\delta_3 < 0, \delta_4 < 3$;
    г) $\delta_3 > 0, \delta_4 = 3$;
    д) $\delta_3 = 0, \delta_4 > 3$;
    е) $\delta_3 < 0, \delta_4 = 3$;
    ж) $\delta_3 = 0, \delta_4 < 3$.

Рекомендуемая литература

1. **Венецкий И.Г., Венецкая В.И.** Основные математико-статистические понятия и формулы в экономическом анализе. — М.: Статистика, 1979. (Разд. 1–4, 6).

2. **Догуерти К.** Введение в эконометрику. — М.: Инфра-М, 1997. (Гл. 1).

3. **Кейн Э.** Экономическая статистика и эконометрия. — М.: Статистика, 1977. Вып. 1. (Гл. 4, 5, 7).

4. **(*) Коррадо Д.** Средние величины. — М.: Статистика, 1970. (Гл. 1).

5. **Judge G.G., Hill R.C., Griffiths W.E., Luthepohl H., Lee T.** Introduction to the Theory and Practice of Econometric. John Wiley & Sons, Inc., 1993. (Ch. 5).

Глава 3

Индексный анализ

До сих пор термин «индекс» использовался исключительно как указатель места элемента в совокупности («мультииндекс» — в сгруппированной совокупности). В данном разделе этот термин применяется в основном для обозначения показателей особого рода, хотя в некоторых случаях он используется в прежнем качестве; его смысл будет понятен из контекста.

3.1. Основные проблемы

В экономической статистике **индексом** называют относительную величину, показывающую, во сколько раз изменяется некоторая другая величина при переходе от одного момента (периода) времени к другому (**индекс динамики**), от одного региона к другому (**территориальный индекс**) или в общем случае — при изменении условий, в которых данная величина измеряется. Так, например, в советской статистике широкое распространение имел индекс выполнения планового задания, который рассчитывается как отношение фактического значения величины к ее плановому значению.

Значение величины, с которым производится сравнение, часто называют **базисным** (измеренным в базисных условиях). Значение величины, которое сравнивается с базисным, называют **текущим** (измеренным в текущих условиях). Эта терминология сложилась в анализе динамики, но применяется и в более общей ситуации. Если y^0 и y^1 — соответственно базисное и текущее значение величины, то индексом ее изменения является $\lambda_y^{01} = \dfrac{y^1}{y^0}$.

В общем случае речь идет о величинах y^t, измеренных в условиях $t = 0, \ldots, T$, и об индексах $\lambda_y^{rs} = \dfrac{y^s}{y^r}$, где r и s принимают значения от 0 до T, и, как правило, $r < s$.

При таком определении система индексов обладает свойством **транзитивности** или, как говорят в экономической статистике, **цепным** свойством (нижний индекс-указатель опущен): $\lambda^{rs} = \lambda^{rt_1}\lambda^{t_1t_2} \cdot \ldots \cdot \lambda^{t_ns}$, где r, s и все t_i, $i = 1, \ldots, n$ также находятся в интервале от 0 до T, и, как следствие, свойством **обратимости**:

$$\lambda^{rs} = \frac{1}{\lambda^{sr}}, \text{ поскольку } \lambda^{tt} = 1.$$

Это — самое общее определение индексов, не выделяющее их особенности среди других относительных величин. Специфика индексов и сложность проблем, возникающих в процессе индексного анализа, определяется следующими тремя обстоятельствами.

1) Задача индексного анализа состоит в количественной оценке не только самого изменения изучаемой величины, но и причин, вызвавших это изменение. Необходимо разложить общий индекс на частные факторные индексы. Пусть (верхний индекс-указатель опущен)

$$y = xa, \tag{3.1}$$

где y и x — объемные величины, a — относительная величина.

Примерами таких «троек» являются:

(a) объем производства продукта в стоимостном выражении, тот же объем производства в натуральном выражении, цена единицы продукта в натуральном выражении;

(b) объем производства, количество занятых, производительность труда;

(c) объем производства, основной капитал, отдача на единицу капитала;

(d) объем затрат на производство, объем производства, коэффициент удельных затрат.

В общем случае формула имеет вид

$$y = x \prod_{j=1}^{n} a_j, \tag{3.2}$$

где все a_j являются относительными величинами.

Примером использования этой формулы при $n = 2$ может явиться сочетание приведенных выше примеров (a) и (b). В этом случае y — объем производства

продукта в стоимостном выражении, x — количество занятых, a_1 — производительность труда, a_2 — цена единицы продукта. Этот пример можно усложнить на случай $n = 3$: a_1 — коэффициент использования труда, a_2 — «технологическая» производительность труда, a_3 — цена.

Дальнейшие рассуждения будут, в основном, проводиться для исходной ситуации ($n = 1$, нижний индекс-указатель у a_1 опускается).

По аналогии с величиной λ_y^{rs}, которую можно назвать **общим** индексом, рассчитываются **частные** или **факторные** индексы для x и a:

$$\lambda_x^{rs} = \frac{x^s}{x^r}, \quad \lambda_a^{rs} = \frac{a^s}{a^r}.$$

Первый из них можно назвать индексом количества, второй — индексом качества.

Оба частных индекса, как и общий индекс, транзитивны и обратимы. Кроме того, вслед за (3.1) выполняется следующее соотношение (верхние индексы-указатели опущены): $\lambda_y = \lambda_x \lambda_a$, и поэтому говорят, что эти три индекса обладают свойством **мультипликативности**. Таким образом, факторные индексы количественно выражают влияние факторов на общее изменение изучаемой величины.

2) Пока неявно предполагалось, что величины y, x, a и, соответственно, все рассчитанные индексы характеризуют отдельный объект, отдельный элемент совокупности. Такие индексы называют **индивидуальными**, и их, а также связанные с ними величины, следует записывать с индексом-указателем i объекта (верхние индексы-указатели t, r, s опущены): y_i, x_i, a_i, λ_{yi}, λ_{xi}, λ_{ai}. До сих пор этот индекс-указатель опускался. Никаких проблем в работе с индивидуальными индексами не возникает, в частности, они по определению обладают свойством транзитивности и мультипликативности.

Предметом индексного анализа являются агрегированные величины. Предполагается, что y_i аддитивны, т.е. выражены в одинаковых единицах измерения, и их можно складывать. Тогда (верхние индексы-указатели опущены)

$$y = \sum_{i=1}^{N} y_i = \sum_{i=1}^{N} x_i a_i.$$

В дальнейшем выражения типа $\sum\limits_{i=1}^{N} x_i a_i$ будут записываться как (x, a), т.е. как скалярные произведения векторов x и a.

Благодаря аддитивности y_i индексы λ_y^{rs} рассчитываются однозначно и являются транзитивными.

Если x_i также аддитивны, их сумму $x = \sum\limits_{i=1}^{N} x_i$ можно вынести за скобки и записать $y = xa$, где a — средняя относительная величина, равная (α_x, a),

Таблица 3.1

	α_x^r	α_x^s	a^r	a^s	λ_a^{rs}	$\lambda_a^{sr} = 1/\lambda_a^{rs}$
1	0.3	0.7	1.25	1.0	0.8	1.25
2	0.7	0.3	0.4	0.5	1.25	0.8
Итого	1.0	1.0	0.66	0.85	1.30	0.7

$\alpha_{xi} = x_i/x$. Такая ситуация имеет место в приведенных выше примерах (b), (c), (d), если объемы производства и затрат измерены в денежном выражении.

В этом случае все формулы, приведенные выше для индивидуальных индексов, остаются справедливыми. Индексы агрегированных величин обладают свойствами транзитивности и мультипликативности.

Индексы агрегированных величин или собственно индексы должны обладать еще одним свойством — свойством **среднего**. Это означает, что их значения не должны выходить за пределы минимального и максимального значений соответствующих индивидуальных индексов. С содержательной точки зрения это свойство весьма желательно. Иногда индексы так и определяются — как средние индивидуальных индексов. Например, индексы динамики — как средние темпы роста.

Легко убедиться в справедливости следующих соотношений (x_i по-прежнему аддитивны):

$$\lambda_y^{rs} = \sum_i \alpha_{yi}^r \lambda_{yi}^{rs}, \text{ где } \alpha_{yi}^r = \frac{y_i^r}{y^r},$$

$$\lambda_x^{rs} = \sum_i \alpha_{xi}^r \lambda_{xi}^{rs}, \text{ где } \alpha_{xi}^r = \frac{x_i^r}{x^r},$$

$$\lambda_a^{rs} = \sum_i \alpha_{ai}^r \lambda_{ai}^{rs}, \text{ где } \alpha_{ai}^r = \frac{\alpha_{xi}^s a_i^r}{\sum \alpha_{xi}^r a_i^r}.$$

Как видно из приведенных соотношений, индексы объемных величин являются средними индивидуальных индексов, т.к. суммы по i весов α_{yi}^r и α_{xi}^r равны единице. Индекс же относительной величины этим свойством не обладает. В частности, он может оказаться больше максимального из индивидуальных индексов, если при переходе от условий r к условиям s резко возрастает вес α_{xi} объекта с высоким показателем λ_a^{rs}. И наоборот, индекс средней относительной величины может оказаться меньше минимального индивидуального индекса, если резко увеличивается вес объекта с низким относительным показателем.

Эту особенность индекса относительной величины можно проиллюстрировать следующим числовым примером при $N = 2$ (см. табл. 3.1).

При переходе от r к s резко увеличивается (с 0.3 до 0.7) доля 1-го объекта с высоким уровнем относительного показателя. В результате значение итогового индекса — 1.43 — оказывается больше значений обоих индивидуальных индексов — 0.8 и 1.25. При переходе от s к r ситуация противоположна (в данном случае индексы обратимы), и итоговый индекс меньше индивидуальных.

Характерно, что этот парадокс возникает в достаточно простой ситуации, когда объемы x_i аддитивны.

3) Собственно проблемы индексного анализа возникают в случае, когда x_i неаддитивны. Такая ситуация имеет место в приведенном выше примере (а). Именно данный пример представляет классическую проблему индексного анализа. В его терминах часто излагается и сама теория индексов. Общий индекс, называемый в этом случае **индексом стоимости**, который рассчитывается по формуле

$$\lambda_y^{rs} = \frac{(x^s, a^s)}{(x^r, a^r)},$$

необходимо разложить на два частных факторных индекса (представить в виде произведения этих частных индексов):

λ_x^{rs} — **индекс объема** (физического объема) и
λ_a^{rs} — **индекс цен**.

В случае аддитивности x_i аналогичные проблемы возникают для индекса не объемной величины y, который раскладывается на факторные индексы естественным образом (как было показано выше), а относительной величины $a = \dfrac{y}{x}$. Общий индекс этой величины, называемый **индексом переменного состава** и удовлетворяющий соотношению

$$\lambda_a^{rs} = \frac{(\alpha_x^s, a^s)}{(\alpha_x^r, a^r)},$$

надо представить как произведение факторных индексов: λ_α^{rs} — **индекс структуры** (структурных сдвигов) и $\lambda_{(a)}^{rs}$ — индекс индивидуальных относительных величин, называемый **индексом постоянного состава**.

3.2. Способы построения индексов

Возникающая проблема разложения общего индекса на факторные индексы может решаться различным образом. Возможны следующие подходы:

$$(1) \quad \lambda_y^{rs} = \frac{(x^s, a^r)}{(x^r, a^r)} \frac{(x^s, a^s)}{(x^s, a^r)} = \lambda_x^{rs} \lambda_a^{rs}.$$

Индекс объема считается как отношение текущей стоимости в базисных ценах к фактической базисной стоимости, а индекс цен — как отношение фактической текущей стоимости к текущей стоимости в базисных ценах.

$$(2)\quad \lambda_y^{rs} = \frac{(x^s, a^s)}{(x^r, a^s)} \frac{(x^r, a^s)}{(x^r, a^r)} = \lambda_x^{rs} \lambda_a^{rs}.$$

В этом случае индекс объема рассчитывается делением фактической текущей стоимости на базисную стоимость в текущих ценах, а индекс цен — делением базисной стоимости в текущих ценах на фактическую базисную стоимость.

Оба эти варианта имеют очевидный содержательный смысл, но результаты их применения количественно различны, иногда — существенно.

(3) Промежуточный вариант, реализуемый, например, если взять среднее геометрическое с равными весами индексных выражений (1) и (2):

$$\lambda_y^{rs} = \sqrt{\frac{(x^s, a^r)}{(x^r, a^r)} \frac{(x^s, a^s)}{(x^r, a^s)}} \sqrt{\frac{(x^s, a^s)}{(x^s, a^r)} \frac{(x^r, a^s)}{(x^r, a^r)}} = \lambda_x^{rs} \lambda_a^{rs}.$$

(4) Индекс объема можно рассчитать как некоторое среднее взвешенное индивидуальных индексов объема:

$$\lambda_x^{rs} = \left(\sum_i \alpha_i \left(\lambda_{xi}^{rs}\right)^k \right)^{\frac{1}{k}}, \quad \sum_i \alpha_i = 1,$$

где k, как правило, принимает значение либо 1 (среднее арифметическое), либо 0 (среднее геометрическое), либо -1 (среднее гармоническое). А индекс цен по формуле $\lambda_a^{rs} = \dfrac{\lambda_y^{rs}}{\lambda_x^{rs}}$, так чтобы выполнялось мультипликативное индексное выражение.

(5) Обратный подход:

$$\lambda_a^{rs} = \left(\sum_i \alpha_i \left(\lambda_{ai}^{rs}\right)^k \right)^{\frac{1}{k}}, \quad \sum_i \alpha_i = 1,$$
$$\lambda_x^{rs} = \frac{\lambda_y^{rs}}{\lambda_a^{rs}}.$$

Индекс объема в подходе (4) и индекс цен в подходе (5) можно находить и другими способами.

(6–7) Например, их можно взять как некоторые средние индексов, определенных в подходах (1) и (2) (т.е. использовать другой вариант подхода (3)):

$$\lambda_x^{rs} = \frac{(x^s, a^r + a^s)}{(x^r, a^r + a^s)}, \quad \left(\lambda_a^{rs} = \frac{\lambda_y^{rs}}{\lambda_x^{rs}}\right),$$
$$\lambda_a^{rs} = \frac{(x^r + x^s, a^s)}{(x^r + x^s, a^r)}, \quad \left(\lambda_x^{rs} = \frac{\lambda_y^{rs}}{\lambda_a^{rs}}\right).$$

(8–9) Или рассчитать по некоторым нормативным ценам a^n и весам x^n:

$$\lambda_x^{rs} = \frac{(x^s, a^n)}{(x^r, a^n)}, \quad \left(\lambda_a^{rs} = \frac{\lambda_y^{rs}}{\lambda_x^{rs}}\right),$$
$$\lambda_a^{rs} = \frac{(x^n, a^s)}{(x^n, a^r)}, \quad \left(\lambda_x^{rs} = \frac{\lambda_y^{rs}}{\lambda_a^{rs}}\right).$$

Подходы (4–5) при определенном выборе типа среднего и весов агрегирования оказываются эквивалентны подходам (1–2).

Так, если в подходе (4) индекс объема взять как среднее арифметическое индивидуальных индексов объема с базисными весами α_y^r, то будет получено индексное выражение подхода (1), поскольку

$$\frac{(x^s, a^r)}{(x^r, a^r)} = \frac{\sum y_i^r \lambda_{xi}^{rs}}{y^r} \quad \text{и, как прежде,} \quad \alpha_{yi}^r = \frac{y_i^r}{y^r}.$$

Аналогично, если в том же подходе (4) индекс объема рассчитать как среднее гармоническое индивидуальных индексов с текущими весами α_y^s, то получится индексное выражение подхода (2).

Подход (5) окажется эквивалентным подходу (1), если в нем индекс цен определить как среднее гармоническое индивидуальных индексов с текущими весами; он будет эквивалентен подходу (2), если индекс цен взять как среднее арифметическое с базисными весами.

Здесь приведено лишь несколько основных подходов к построению мультипликативных индексных выражений. В настоящее время известны десятки (а с некоторыми модификациями — сотни) способов расчета индексов. Обилие подходов свидетельствует о том, что данная проблема однозначно и строго не решается. На этом основании некоторые скептики называли индексы способом измерения неизмеримых в принципе величин и ставили под сомнение саму целесообразность их применения. Такая точка зрения ошибочна.

Во-первых, индексы дают единственную возможность получать количественные макрооценки протекающих экономических процессов (динамика реального

производства, инфляция, уровень жизни и т.д.), во-вторых, они и только они позволяют иметь практические приложения многих абстрактных разделов макроэкономики как научной дисциплины. Так, например, даже самое элементарное макроэкономическое уравнение денежного обмена

$$PQ = MV,$$

где P — уровень цен, Q — товарная масса, M — денежная масса, V — скорость обращения денег, не имеет непосредственно никакого практического смысла, ибо ни в каких единицах, имеющих содержательный смысл, не могут быть измерены P и Q. Можно измерить лишь изменения этих величин и — только с помощью техники индексного анализа. Например, измеримыми могут быть переменные уравнения денежного оборота в следующей форме:

$$Y^0 \lambda_P^{01} \lambda_Q^{01} = M^1 V^1,$$

где Y^0 — валовой оборот (общий объем производства или потребления) базисного периода в фактических ценах.

Проблема выбора конкретного способа построения индексов из всего множества возможных способов решается на практике различным образом.

В советской статистике был принят подход (1). Аргументация сводилась к следующему. Количественный (объемный) признак является первичным по отношению к качественному (относительному) и поэтому при переходе от базисных условий к текущим сначала должен меняться он (количественный признак):

$$\left(x^0, a^0\right) \to \left(x^1, a^0\right) \to \left(x^1, a^1\right).$$

Первый шаг этого «перехода» дает индекс объема, второй — индекс цен. Внятных разъяснений тому, почему количественный признак первичен и почему именно первичный признак должен меняться первым, как правило, не давалось. Тем не менее, применение этого подхода делает весьма наглядным понятие объемов (производства, потребления, ...) в **сопоставимых** или **неизменных** ценах.

Действительно, пусть оценивается динамика в последовательные периоды времени $t = 0, \ldots, N$, и индексы для любого периода $t > 0$ строятся по отношению к одному и тому же базисному периоду $t = 0$. Тогда при использовании подхода (1) указанный выше «переход» для любого периода $t > 0$ принимает форму $(x^0, a^0) \to (x^t, a^0) \to (x^t, a^t)$, и выстраивается следующая цепочка показателей физического объема: (x^0, a^0), (x^1, a^0), $\ldots$, (x^t, a^0), $\ldots$, (x^N, a^0). Очевидна интерпретация этих показателей — это объемы в сопоставимых (базисных) или неизменных ценах. Однако «наглядность» не всегда обеспечивает «правильность». Об этом пойдет речь в пункте 3.6.

В современной индексологии проблема выбора решается в зависимости от того, какому набору требований (аксиом, тестов) должны удовлетворять применяемые индексы. Требования — это свойства, которыми должны обладать индексы. Выше были определены три таких свойства: мультипликативности, транзитивности и среднего. Приведенные выше подходы к построению индексов с этой точки зрения не одинаковы. Все они удовлетворяют требованию мультипликативности — по построению. А транзитивными могут быть, например, только в подходах $(4{-}5)$, при $k = 0$. Свойством среднего могут не обладать индексы цен подходов $(4,\ 6,\ 8)$ и индексы объемов в подходах $(5,\ 7,\ 9)$.

Иногда добавляют еще одно требование — **симметричности**. Это требование означает, что оба факторных индекса должны рассчитываться по одной и той же формуле, в которой лишь меняются местами переменные и нижние индексы с x на a или наоборот. Из всех приведенных выше подходов только (3) приводит к индексам, отвечающим этому требованию. Многие экономисты считают это требование надуманным. Так, например, даже при естественном разложении общего индекса, которое имеет место в случае аддитивности объемного признака, факторные индексы асимметричны.

При выборе способа расчета индексов полезно проводить математический анализ используемых формул. В некоторых случаях эти математические свойства таковы, что результат расчета неизбежно будет содержать систематическую ошибку.

Пусть, например, речь идет о расчете индекса цен как среднего индивидуальных индексов (подход (5)), и веса взвешивания остаются неизменными во времени.

В данном случае (как и в ряде других случаев) имеет смысл проверить, как ведет себя индекс на осциллирующих рядах индивидуальных цен. Цены осциллируют — значит меняются циклически с периодом две единицы времени:

$$\lambda_{ai}^{t,\,t+1} = \frac{1}{\lambda_{ai}^{t+1,\,t+2}}, \quad t = 0,\ 1,\ 2,\ \ldots\ .$$

Поэтому общий индекс цен за период времени, включающий четное количество временных единиц, всегда равен единице:

$$\lambda_{a}^{t,\,t+2T} = 1.$$

Этот результат понятен, поскольку индивидуальные цены лишь колеблются, не изменяя своего общего уровня. Этот же общий индекс можно рассчитать по цепному правилу:

$$\lambda_{a}^{t,\,t+1}\lambda_{a}^{t+1,\,t+2}\ldots\lambda_{a}^{t+2T-1,\,t+2T}.$$

Индекс в такой форме в дальнейшем будет называться **цепным** и обозначаться $\lambda_{a}^{t,\,t+1,\,\ldots,\,t+2T}$ или λ_{a}^{t*t+2T}, где « $*$ » заменяет последовательность временных подпериодов — единиц времени внутри общего периода.

Рассчитанный таким образом индекс равен единице только при использовании среднего геометрического (при $k = 0$) в расчете индексов за каждую единицу времени. Это проверяется непосредственной подстановкой формулы среднего геометрического при неизменных во времени весах индивидуальных индексов. Из свойства мажорантности средних следует, что при использовании средних арифметических общий цепной индекс будет обязательно больше единицы, а при использовании средних гармонических — меньше единицы. Другими словами, результат будет либо преувеличен, либо преуменьшен. Причем ошибка будет тем выше, чем длиннее рассматриваемый период (чем больше T). Из этого следует два вывода:

- при расчете общего индекса как среднего индивидуальных индексов веса не должны оставаться постоянными во времени,
- общий индекс как среднее арифметическое индивидуальных индексов может преувеличить реальный рост изучаемой величины, а как среднее гармоническое — преуменьшить его.

Формальный анализ индексного выражения позволяет выяснить, с какими погрешностями связано его использование при изучении реальных процессов.

Например, полезно исследовать, к каким погрешностям приводит нетранзитивность индексов.

Как уже отмечалось, в общем случае индексы всех приведенных выше подходов не обладают свойством транзитивности. В частности, индекс цен подхода (1) не транзитивен, т.к.

$$\lambda_a^{012} = \frac{(x^1, a^1)}{(x^1, a^0)} \frac{(x^2, a^2)}{(x^2, a^1)} \neq \frac{(x^2, a^2)}{(x^2, a^0)} = \lambda_a^{02}.$$

Вопрос о том, какая из этих величин больше, сводится, как не сложно убедиться, путем элементарных преобразований к вопросу о соотношении следующих двух возможных значений индекса λ_a^{01}:

$$\frac{(x^1, a^1)}{(x^1, a^0)} \text{ и } \frac{(x^2, a^1)}{(x^2, a^0)},$$

которые можно обозначить, соответственно, через $\lambda_a^{01}(1)$ и $\lambda_a^{01}(2)$. Их, в свою очередь, можно представить как средневзвешенные индивидуальных индексов цен λ_{ai}^{01} (индексы-указатели опущены):

$$\lambda(1) = (\alpha(1), \lambda), \ \lambda(2) = (\alpha(2), \lambda),$$
$$\text{где } \alpha_i(1) = \frac{x_i^1 a_i^0}{(x^1, a^0)}, \ \alpha_i(2) = \frac{x_i^2 a_i^0}{(x^2, a^0)}.$$

Для рыночной экономики характерно сокращение объемов покупок товара при росте цен на него. Если предположить, что динамика цен и объемов устойчива в рассматриваемом периоде, и направленность их трендов (вверх или вниз) не меняется на нем

(такое предположение необходимо сделать, т.к. динамика цен на подпериоде 01 связывается в данных индексах с динамикой объемов на подпериоде 12), то в таких условиях

$$\lambda_a^{01}(1) > \lambda_a^{01}(2).$$

Из этого следует, что для рыночной экономики значение цепного индекса λ_a^{012} в подходе (1) больше значения соответствующего обычного индекса λ_a^{02}.

Аналогичный анализ индексов цен подхода (2) показывает, что для них характерно противоположное соотношение: цепной индекс принимает меньшее значение, чем обычный индекс за период времени.

Несколько слов о терминах.

Факторные индексы, используемые в подходах (1–2), называются **агрегатными**. Такие индексы были предложены немецкими экономистами Э. Ласпейресом и Г. Пааше во второй половине XIX века. **Индекс Ласпейреса** строится так, чтобы в числителе и знаменателе неизменными оставались объемы или цены на **базисном** уровне, поэтому знаменателем этого индекса является фактическая базисная стоимость, а числитель образован и базисными, и текущими значениями. Этот индекс является среднеарифметическим индивидуальных индексов с базисными весами. Таковыми являются индекс объема в подходе (1) и индекс цен подхода (2).

В числителе и знаменателе **индекса Пааше** одинаковыми объемы или цены фиксируются на **текущем** уровне. Его числителем является фактическая текущая стоимость, знаменатель имеет смешанный состав. Такой индекс выступает среднегармоническим индивидуальных индексов с текущими весами. Это — индекс цен подхода (1) и индекс объема подхода (2).

В мультипликативном представлении общего индекса стоимости один из факторных индексов — индекс Ласпейреса, другой — Пааше.

В 20-х годах XX века Фишером было предложено рассчитывать индексы как средние геометрические соответствующих индексов Ласпейреса и Пааше с равными весами. Потому индексы подхода (3) называются **индексами Фишера**. Фишер показал, что в его системе тестов (требований, аксиом) они являются наилучшими из всех возможных (им рассмотренных).

Индексы цен, рассчитанные каким-то способом, например, как в подходах (5), (7), (9), или заданные нормативно (при прогнозировании) с целью дальнейшего определения индексов объемов из требования мультипликативности иногда называют **дефляторами** стоимости (например, дефляторами ВВП — валового внутреннего продукта). А такой способ расчета индексов цен и объемов — **дефлятированием**.

На практике при построении индексов цен часто используют нормативный подход (9). Причем структуру весов обычно принимают облегченной — не по всем товарам (их, как правило, бывает много), а по **товарам-представителям**, каждый из которых представляет целую товарную группу. Такой характер имеют, например, индексы цен **по потребительской корзине**, в которую включаются от нескольких десятков до нескольких сотен основных потребительских продуктов.

Итак, рассмотрены основные проблемы и подходы, существующие при проведении индексного анализа, с помощью которого изучается вопрос о том, **во сколько раз** меняется значение величины при переходе от одних условий к другим — в целом и за счет отдельных факторов.

3.3. Факторные представления приростных величин

Во многом схожие проблемы возникают и в анализе вопроса о том, **на сколько** и за счет каких факторов меняется значение изучаемой величины. В таком анализе общее изменение величины во времени или в пространстве требуется разложить по факторам, вызвавшим это изменение:

$$y^s - y^r = \Delta_y^{rs} = \Delta_x^{rs} + \Delta_a^{rs}.$$

В случае, когда y — результат (какая-то результирующая величина, например, объем производства), x — затраты (например, основной капитал или занятые в производстве), a — эффективность использования затрат (отдача на капитал или производительность труда), то говорят о проблеме разложения общего прироста результирующей величины на **экстенсивные** и **интенсивные** факторы.

При изучении изменений относительной величины $a^t = (\alpha^t, a^t)$ во времени или в пространстве — в случае аддитивности объемных признаков x_i — возникает аналогичная проблема разделения прироста этой величины Δ_a^{rs} на факторы изменения структуры Δ_α^{rs} и изменения индивидуальных относительных величин $\Delta_{(a)}^{rs}$. Так, например, общее различие материалоемкости совокупного производства между двумя регионами можно попытаться разбить на факторы различия отраслевых структур производства и отраслевых материалоемкостей производства.

Эти проблемы можно решить так же, как и в подходах (1–3) индексного анализа.

(1′) В подходе (1) индексного анализа общий индекс λ_y^{rs} умножается и делится на величину (x^s, a^r), и после перегруппировки сомножителей получается искомое индексное выражение. Теперь, аналогично, к общему приросту Δ_y^{rs} прибавляется и из него вычитается та же величина (x^s, a^r). После перегруппировки слагаемых

образуется требуемое пофакторное представление:

$$\Delta_y^{rs} = [(x^s, a^r) - (x^r, a^r)] + [(x^s, a^s) - (x^s, a^r)] = \\ = (x^s - x^r, a^r) + (x^s, a^s - a^r) = \Delta_x^{rs} + \Delta_a^{rs}.$$

$(2')$ Теперь работает величина (x^r, a^s):

$$\Delta_y^{rs} = [(x^s, a^s) - (x^r, a^s)] + [(x^r, a^s) - (x^r, a^r)] = \\ = (x^s - x^r, a^s) + (x^r, a^s - a^r) = \Delta_x^{rs} + \Delta_a^{rs}.$$

$(3')$ Берется среднее арифметическое пофакторных представлений $(1')$ и $(2')$ с равными весами:

$$\Delta_y^{rs} = \left(x^s - x^r, \frac{a^r + a^s}{2}\right) + \left(\frac{x^r + x^s}{2}, a^s - a^r\right) = \Delta_x^{rs} + \Delta_a^{rs}.$$

Существует более общий подход, в рамках которого пофакторное представление общего прироста результирующей величины строится на основе определенного мультипликативного индексного выражения $\lambda_y^{rs} = \lambda_x^{rs}\lambda_a^{rs}$.

Для относительного прироста результирующей величины можно записать следующее тождество:

$$\lambda_y^{rs} - 1 = (\lambda_x^{rs} - 1) + (\lambda_a^{rs} - 1) + (\lambda_x^{rs} - 1)(\lambda_a^{rs} - 1).$$

Выражение для общего абсолютного прироста результирующей величины получается умножением обеих частей этого соотношения на y^r, равный (x^r, a^r).

Первое слагаемое правой части этого тождества показывает влияние изменения объемной величины (экстенсивные факторы), второе слагаемое — влияние изменения относительной величины (интенсивные факторы), а третье слагаемое — совместное влияние этих факторов. Эта ситуация иллюстрируется рисунком 3.1.

Общему изменению результирующей величины соответствует площадь фигуры $ABDFGH$, влиянию объемного фактора — площадь $ABCH$, влиянию относительного фактора — площадь $GHEF$, совместному влиянию факторов — площадь $HCDE$. Вопрос получения искомого пофакторного представления общего прироста сводится к тому, как распределить между факторами «вклад» их совместного влияния. Здесь возможны три подхода.

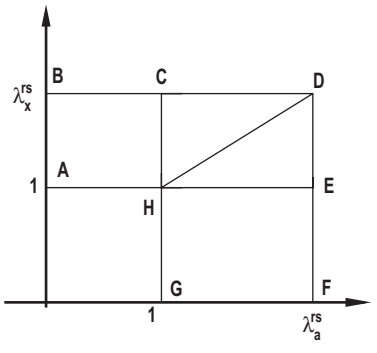

Рис. 3.1

$(1'')$ Все совместное влияние факторов можно отнести на относительный фактор:

$$\lambda_y^{rs} - 1 = (\lambda_x^{rs} - 1) + \lambda_x^{rs}\,(\lambda_a^{rs} - 1) = \frac{1}{y^r}\,(\Delta_x^{rs} + \Delta_a^{rs}).$$

В этом случае влиянию относительного фактора соответствует на рисунке площадь $GCDF$, а влиянию объемного фактора — площадь $ABCH$.

$(2'')$ Совместное влияние факторов относится на количественный фактор:

$$\lambda_y^{rs} - 1 = (\lambda_x^{rs} - 1)\,\lambda_a^{rs} + (\lambda_a^{rs} - 1) = \frac{1}{y^r}\,(\Delta_x^{rs} + \Delta_a^{rs}).$$

Теперь влиянию объемного фактора соответствует на рисунке площадь $ABDE$, а влиянию относительного фактора — площадь $GHEF$.

Несложно убедиться в том, что в подходе $(1'')$ фактически к общему относительному приросту $\lambda_x^{rs}\lambda_a^{rs} - 1$ прибавляется и отнимается индекс объемной величины λ_x^{rs}, а затем нужным образом группируются слагаемые; в подходе $(2'')$ — прибавляется и отнимается индекс относительной величины λ_a^{rs}, а затем также нужным образом группируются слагаемые. Другими словами, имеется определенная аналогия с подходами $(1')$ и $(2')$. Можно сказать, что в подходе $(1'')$ сначала меняет свое значение объемный признак, а затем — относительный:

$$1 \times 1 \to \lambda_x^{rs} \times 1 \to \lambda_x^{rs}\lambda_a^{rs},$$

и первый шаг в этом переходе определяет вклад объемного фактора, второй — относительного фактора. В подходе $(2'')$, наоборот, сначала меняется значение относительного признака, а затем — объемного:

$$1 \times 1 \to 1 \times \lambda_a^{rs} \to \lambda_x^{rs}\lambda_a^{rs},$$

и теперь первый шаг перехода дает вклад относительного фактора, второй — объемного.

$(3'')$ Берется среднее арифметическое с равными весами пофакторных представлений $(1'')$ и $(2'')$:

$$\lambda_y^{rs} - 1 = (\lambda_x^{rs} - 1)\,\frac{1 + \lambda_a^{rs}}{2} + \frac{1 + \lambda_x^{rs}}{2}\,(\lambda_a^{rs} - 1) = \frac{1}{y^r}\,(\Delta_x^{rs} + \Delta_a^{rs}).$$

В этом случае влияние объемного фактора выражает площадь трапеции $ABDH$, а влияние относительного фактора — площадь трапеции $GHDF$.

Итак, на основе каждого мультипликативного индексного выражения можно получить по крайней мере три пофакторных представления прироста изучаемой величины. Причем, если неопределенность (множественность подходов) построения

индексного выражения связана с агрегированным характером изучаемой величины и неаддитивностью объемных факторных величин, то неопределенность пофакторных представлений приростов имеет место и для неагрегированных величин. Это объясняется тем, что она (неопределенность) является следствием наличия компоненты совместного влияния факторов, которую необходимо каким-то образом «разделить» между факторами.

Пусть, например, используется индексное выражение подхода (1). На его основе получается три следующих пофакторных представления общего прироста.

$$(1-1'') \quad \begin{aligned} \Delta_x^{rs} &= (x^s - x^r, a^r)\,, \\ \Delta_a^{rs} &= (x^s, a^s - a^r)\,. \end{aligned}$$

Эти выражения получены в результате подстановки индексов подхода (1) в формулу подхода $(1'')$ и умножения на y^r. Интересно, что результат совпадает с подходом $(1')$.

$$(1-2'') \quad \begin{aligned} \Delta_x^{rs} &= (x^s - x^r, a^r)\,\frac{(x^s, a^s)}{(x^s, a^r)}, \\ \Delta_a^{rs} &= (x^s, a^s - a^r)\,\frac{(x^r, a^r)}{(x^s, a^r)}. \end{aligned}$$

$$(1-3'') \quad \begin{aligned} \Delta_x^{rs} &= (x^s - x^r, a^r)\,\frac{\left(x^s, \frac{a^r+a^s}{2}\right)}{(x^s, a^r)}, \\ \Delta_a^{rs} &= (x^s, a^s - a^r)\,\frac{\left(\frac{x^r+x^s}{2}, a^r\right)}{(x^s, a^r)}. \end{aligned}$$

Теперь используется индексное выражение подхода (2).

$$(2-1'') \quad \begin{aligned} \Delta_x^{rs} &= (x^s - x^r, a^s)\,\frac{(x^r, a^r)}{(x^r, a^s)}, \\ \Delta_a^{rs} &= (x^r, a^s - a^r)\,\frac{(x^s, a^s)}{(x^r, a^s)}. \end{aligned}$$

$$(2-2'') \quad \begin{aligned} \Delta_x^{rs} &= (x^s - x^r, a^s), \\ \Delta_a^{rs} &= (x^r, a^s - a^r). \end{aligned}$$

Этот результат аналогичен подходу $(2')$.

$$(2-3'') \quad \Delta_x^{rs} = (x^s - x^r, a^s)\,\frac{\left(x^r, \frac{a^r+a^s}{2}\right)}{(x^r, a^s)},$$

$$\Delta_a^{rs} = (x^r, a^s - a^r)\,\frac{\left(\frac{x^r+x^s}{2}, a^s\right)}{(x^r, a^s)}.$$

При всем многообразии полученных пофакторных представлений прироста изучаемой величины все они являются «вариациями на одну тему»: вклады объемного и относительного факторов определяются в результате различных скаляризаций, соответственно, векторов $x^s - x^r$ и $a^s - a^r$. Кроме того, несложно установить, что для индивидуальных (неагрегированных) величин подходы $(1') \equiv (1{-}1'')$ и $(2{-}1'')$ эквивалентны, также как подходы $(1{-}2'')$ и $(2') \equiv (2{-}2'')$ и подходы $(3')$, $(1{-}3'')$ и $(2{-}3'')$, т.е. различия между ними, по существу, связаны с разными способами разделения совместного влияния факторов.

3.4. Случай, когда относительных факторов более одного

Теперь можно дать обобщение подходов $(1-3)$, $(1'-3')$ и $(1''-3'')$ на случай, когда относительных факторов в мультипликативном выражении (3.2) два или более. Пусть $n = 2$, т.е.

$$y^t = \sum_i x_i^t a_{1i}^t a_{2i}^t.$$

Для краткости будем далее использовать обозначение $(x^t, a_1^t, a_2^t) = \sum_i x_i^t a_{1i}^t a_{2i}^t$. Речь идет о построении индексного выражения

$$\lambda_y^{rs} = \lambda_x^{rs} \lambda_1^{rs} \lambda_2^{rs}$$

в идеологии подходов $(1-3)$, где λ_1^{rs} и λ_2^{rs} — индексы первого и второго относительного признака.

Построение мультипликативного индексного выражения зависит от того, в какой последовательности факторные величины меняют свои значения от базисных к текущим. Пусть эта последовательность задана такой же, как и в исходном мультипликативном выражении, т.е. сначала меняет свое значение объемный признак, затем первый относительный признак и, в последнюю очередь, второй относительный признак:

$$(x^r, a_1^r, a_2^r) \to (x^s, a_1^r, a_2^r) \to (x^s, a_1^s, a_2^r) \to (x^s, a_1^s, a_2^s).$$

Тогда

$$\lambda_x^{rs} = \frac{(x^s, a_1^r, a_2^r)}{(x^r, a_1^r, a_2^r)}, \quad \lambda_1^{rs} = \frac{(x^s, a_1^s, a_2^r)}{(x^s, a_1^r, a_2^r)}, \quad \lambda_2^{rs} = \frac{(x^s, a_1^s, a_2^s)}{(x^s, a_1^s, a_2^r)}.$$

Такой способ построения индексного выражения полностью аналогичен подходу (1). Пусть теперь последовательность включения факторных величин изменилась. Например, объемный признак по-прежнему меняет свое значение первым, затем — второй и, наконец, первый относительный признак:

$$(x^r, a_1^r, a_2^r) \to (x^s, a_1^r, a_2^r) \to (x^s, a_1^r, a_2^s) \to (x^s, a_1^s, a_2^s).$$

Тогда

$$\lambda_x^{rs} = \frac{(x^s, a_1^r, a_2^r)}{(x^r, a_1^r, a_2^r)}, \quad \lambda_1^{rs} = \frac{(x^s, a_1^s, a_2^s)}{(x^s, a_1^r, a_2^s)}, \quad \lambda_2^{rs} = \frac{(x^s, a_1^r, a_2^s)}{(x^s, a_1^r, a_2^r)}.$$

Общее количество возможных последовательностей включения факторных величин равно числу перестановок из 3 элементов: $3! = 6$, т.е. имеется 6 возможных мультипликативных индексных выражений. Аналогом индексного выражения (3) является среднее геометрическое с равными весами указанных 6-ти вариантов.

Аналогичным образом строятся пофакторные представления типа $(1'-3')$ и типа $(1'' - 3'')$. Во 2-м случае, если принята исходная последовательность «включения» факторных признаков:

$$1 \times 1 \times 1 \to \lambda_x^{rs} \times 1 \times 1 \to \lambda_x^{rs}\lambda_1^{rs} \times 1 \to \lambda_x^{rs}\lambda_1^{rs}\lambda_2^{rs},$$

то

$$\Delta_x^{rs} = y^r\left(\lambda_x^{rs} - 1\right), \quad \Delta_1^{rs} = y^r\lambda_x^{rs}\left(\lambda_1^{rs} - 1\right), \quad \Delta_2^{rs} = y^r\lambda_x^{rs}\lambda_1^{rs}\left(\lambda_2^{rs} - 1\right);$$

если относительные признаки в принятой последовательности меняются местами:

$$1 \times 1 \times 1 \to \lambda_x^{rs} \times 1 \times 1 \to \lambda_x^{rs} \times 1 \times \lambda_2^{rs} \to \lambda_x^{rs}\lambda_1^{rs}\lambda_2^{rs},$$

то

$$\Delta_x^{rs} = y^r\left(\lambda_x^{rs} - 1\right), \quad \Delta_1^{rs} = y^r\lambda_x^{rs}\lambda_2^{rs}\left(\lambda_1^{rs} - 1\right), \quad \Delta_2^{rs} = y^r\lambda_x^{rs}\left(\lambda_2^{rs} - 1\right),$$

и общее количество вариантов таких представлений — 6. Аналогом представления $(3'')$ будет являться среднее арифметическое этих 6-ти вариантов с равными весами.

В общем случае при n относительных величин в мультипликативном представлении результирующей величины имеется $(n+1)!$ вариантов индексных выражений, аналогичных $(1-2)$, и пофакторных представлений, аналогичных $(1'-2')$

и $(1'-2'')$ (в основном случае, рассмотренном в пунктах $3.1-3.2$, $n=1$, и имелось по 2 таких варианта). Усреднение этих вариантов с равными весами дает результаты, аналогичные, соответственно, подходам (3), $(3')$ и $(3'')$.

В пунктах $3.1-3.4$ рассмотрены проблемы, которые возникают в практике построения индексных выражений и пофакторных представлений динамики результирующей величины. Проведенный анализ можно назвать прикладным.

3.5. Индексы в непрерывном времени

Для лучшего понимания проблем, возникающих при индексном анализе, и возможностей решения этих проблем полезно рассмотреть их на примере индексов в непрерывном времени. Анализ индексов в непрерывном времени можно назвать теоретическим. В этом случае динамика объемных и относительных величин задается непрерывными дифференцируемыми функциями $y(t)$, $x(t)$, $a(t)$, и возможны три типа индексов: в момент времени t (моментные), сопоставляющие два момента времени t_1 и t_0 («момент к моменту») и два периода времени $[t_1, t_1+\tau]$ и $[t_0, t_0+\tau]$, $\tau \leqslant |t_1 - t_0|$ («период к периоду»). Ниже рассматриваются эти три типа индексов.

1) Моментные индексы.

Индивидуальными индексами такого типа являются моментные темпы роста, рассмотренные в пункте 1.8 (нижние индексы-указатели объекта опущены):

$$\lambda_{[\,]}(t) = \exp\left(\frac{d\ln[\,](t)}{dt}\right), \quad \ln\lambda_{[\,]}(t) = \frac{d\ln[\,](t)}{dt} = \Delta\lambda_{[\,]}(t),$$

где $\lambda_{[\,]}(t)$ — моментный темп роста, $\Delta\lambda_{[\,]}(t)$ — моментный темп прироста, а на месте $[\,]$ стоит либо y — для объемной результирующей величины (стоимости), либо x — для объемной факторной величины (объема), либо a — для относительной величины (цены).

Легко убедиться в том, что эти индивидуальные индексы вслед за (3.1) обладают свойством мультипликативности или, как говорят, удовлетворяют требованию (тесту) мультипликативности (здесь и далее при описании моментных индексов указатель на момент времени (t) опущен):

$$\ln\lambda_y = \frac{d\ln(xa)}{dt} = \frac{d\ln x}{dt} + \frac{d\ln a}{dt} = \ln\lambda_x + \ln\lambda_a,$$

т.е. $\lambda_y = \lambda_x\lambda_a$.

Вопрос о транзитивности моментных индексов обсуждается ниже в связи с переходом к индексам «момент к моменту». Понять, как перемножаются индексы

в бесконечной последовательности бесконечно малых моментов времени, можно только в интегральном анализе.

Агрегированный моментный индекс или собственно моментный индекс объемной результирующей величины строится следующим образом (возвращаются нижние индексы-указатели объекта):

$$\ln \lambda_y = \frac{1}{y}\frac{d\sum y_i}{dt} = \frac{1}{y}\sum \frac{dy_i}{dt} = \sum \alpha_i \frac{1}{y_i}\frac{dy_i}{dt} = \sum \alpha_i \ln \lambda_{yi},$$

где $\alpha_i = \dfrac{y_i}{y}$, т.е. $\lambda_y = \prod \lambda_{yi}^{\alpha_i}$.

Таким образом, индекс результирующей величины есть средняя геометрическая индивидуальных индексов с весами-долями объектов в этой объемной результирующей величине. Как видно из приведенного доказательства, это — следствие аддитивности результирующей величины.

Не сложно провести разложение общего индекса на факторные:

$$\ln \lambda_y = \frac{1}{y}\sum \frac{dx_i a_i}{dt} = \frac{1}{y}\sum \left(a_i \frac{dx_i}{dt} + x_i \frac{da_i}{dt}\right) =$$
$$= \frac{1}{y}\sum \left(y_i \frac{1}{x_i}\frac{dx_i}{dt} + y_i \frac{1}{a_i}\frac{da_i}{dt}\right) = \sum \alpha_i \ln \lambda_{xi} + \sum \alpha_i \ln \lambda_{ai},$$

т.е.

$$\lambda_y = \left(\prod \lambda_{xi}^{\alpha_i}\right)\left(\prod \lambda_{ai}^{\alpha_i}\right),$$

и, если факторные индексы определить как

$$\lambda_x = \prod \lambda_{xi}^{\alpha_i}, \quad \lambda_a = \prod \lambda_{ai}^{\alpha_i},$$

то получается искомое мультипликативное выражение $\lambda_y = \lambda_x \lambda_a$.

Чрезвычайно интересно, что и факторный индекс объемной величины, которая может быть неаддитивной, и факторный индекс относительной величины, которая принципиально неаддитивна, рассчитываются так же, как общий индекс аддитивной результирующей величины — как средние геометрические индивидуальных индексов. Причем во всех этих трех индексах используются одинаковые веса — доли объектов в результирующей величине.

Итак, моментные индексы мультипликативны, транзитивны, что будет показано ниже, обладают свойством среднего и симметричны по своей форме. Следовательно, обсуждаемые выше проблемы являются следствием не принципиальных особенностей индексов, а разных способов привязки их ко времени.

2) Индексы «момент к моменту» (индексы за период времени).

Индивидуальные индексы такого типа рассмотрены в пункте 1.8 как непрерывные темпы роста за период (нижние индексы-указатели объектов опущены):

$$\lambda_{[\,]}(t_0,t_1)=e^{\int_{t_0}^{t_1}\ln\lambda_{[\,]}(t)dt}=\frac{[\,](t_1)}{[\,](t_0)},$$

где $\lambda_{[\,]}(t_0,t_1)$ — индекс за период $[t_0,t_1]$, а на месте $[\ ]$, как и прежде, стоит либо y — для объемной результирующей величины (стоимости), либо x — для объемной факторной величины (объема), либо a — для относительной величины (цены).

Это выражение, прежде всего, означает транзитивность моментных индексов. Чтобы убедиться в этом, можно провести следующие рассуждения (указатель $[\,]$ в этих рассуждениях опущен).

Пусть моментный индекс в периоде $[t_0,t_1]$ неизменен и равен $\lambda(t_1)$, тогда, вычислив $\int\limits_{t_0}^{t_1}\ln\lambda(t)dt$, можно увидеть, что

$$\lambda(t_0,t_1)=\lambda(t_1)^{t_1-t_0},$$

т.е. для того, чтобы привести моментные индексы к форме, сопоставимой с индексами за период, надо их возводить в степень, равную длине периода.

Теперь, разбив общий период $[t_0,\ t_1]$ на n равных подпериодов длиной $\tau=\dfrac{t_1-t_0}{n}$ и обозначив $t_{(j)}=t_0+j\tau$, можно записать исходное выражение связи индекса за период с моментными индексами в следующем виде:

$$\ln\lambda(t_0,t_1)=\sum_{j=1}^{n}\int\limits_{t_{(j-1)}}^{t_{(j)}}\ln\lambda(t)\,dt.$$

Пусть в каждом j-м подпериоде $[t_{(j-1)},t_{(j)}]$ моментный индекс неизменен и равен $\lambda(t_{(j)})$. Тогда из этого выражения следует, что

$$\lambda(t_0,t_1)=\left(\prod_{j=1}^{n}\lambda\left(t_{(j)}\right)\right)^{\tau}.$$

В результате перехода к пределу при $n\to\infty$ получается соотношение, которое можно интерпретировать как свойство транзитивности моментных индексов. Возведение цепного моментного индекса в степень τ необходимо, как было только что показано, для приведения его к форме, сопоставимой с индексом за период.

Индивидуальные индексы «момент к моменту» транзитивны по своему определению:

$$\ln \lambda_{[\,]}(t_0, t_2) = \int_{t_0}^{t_2} \ln \lambda_{[\,]}(t)\, dt = \int_{t_0}^{t_1} \ln \lambda_{[\,]}(t)\, dt + \int_{t_1}^{t_2} \ln \lambda_{[\,]}(t)\, dt = \\ = \ln \lambda_{[\,]}(t_0, t_1) + \ln \lambda_{[\,]}(t_1, t_2),$$

т.е. $\lambda_{[\,]}(t_0, t_2) = \lambda_{[\,]}(t_0, t_1) \cdot \lambda_{[\,]}(t_1, t_2)$.

Их мультипликативность следует непосредственно из мультипликативности моментных индексов. Действительно:

$$\ln \lambda_y(t_0, t_1) = \int_{t_0}^{t_1} \underbrace{(\ln \lambda_x(t) + \ln \lambda_a(t))}_{\ln \lambda_x(t)\lambda_a(t)} dt = \underbrace{\ln \lambda_x(t_0, t_1) + \ln \lambda_a(t_0, t_1)}_{\lambda_y(t)},$$

т.е.

$$\lambda_y(t_0, t_1) = \lambda_x(t_0, t_1) \cdot \lambda_a(t_0, t_1).$$

Теперь рассматриваются агрегированные индексы «момент к моменту» (возвращаются нижние индексы-указатели объекта). Индексы такого типа были предложены в конце 20-х годов XX века французским статистиком Ф. Дивизиа, и поэтому их называют **индексами Дивизиа**.

Как было показано выше, моментный индекс результирующей величины является средним геометрическим индивидуальных индексов. Для индекса Дивизиа результирующей величины такое свойство в общем случае не выполняется. Действительно:

$$\ln \lambda_y(t_0, t_1) = \sum_i \int_{t_0}^{t_1} \alpha_i(t) \ln \lambda_{yi}(t)\, dt,$$

и, если бы веса $\alpha_i(t)$ не менялись во времени, их можно было бы вынести за знак интеграла и получить выражение индекса как среднего геометрического индивидуальных индексов. Однако веса меняются во времени, и такую операцию провести нельзя. Можно было бы ввести средние за период веса по следующему правилу:

$$\alpha_i(t_0, t_1) = \frac{\int \alpha_i(t) \ln \lambda_{yi}(t)\, dt}{\int \ln \lambda_{yi}(t)\, dt},$$

и получить выражение

$$\ln \lambda_y(t_0, t_1) = \sum \alpha_i(t_0, t_1) \ln \lambda_{yi}(t_0, t_1), \tag{3.3}$$

которое являлось бы средним геометрическим, если бы сумма средних за период весов равнялась единице. Но равенство единице их суммы в общем случае не гарантировано.

Имеется один частный случай, когда общий индекс является средним геометрическим индивидуальных. Если индивидуальные моментные индексы не меняются во времени и, как было показано выше, равны $(\lambda_{yi}(t_0,t_1))^{1/(t_1-t_0)}$, то их можно вынести за знак интеграла и получить выражение, аналогичное по форме (3.3):

$$\ln \lambda_y(t_0,t_1) = \sum \alpha_i(t_0,t_1) \ln \lambda_{yi}(t_0,t_1),$$

где теперь $\alpha_i(t_0,t_1) = \dfrac{1}{t_1-t_0}\int\limits_{t_0}^{t_1} \alpha_i(t)\,dt$ — средние хронологические весов. Их сумма равна единице, т.к. $\sum \alpha_i(t) = 1$:

$$\sum \alpha_i(t_0,t_1) = \frac{1}{t_1-t_0}\int\limits_{t_0}^{t_1} \sum \alpha_i(t)\,dt = \frac{1}{t_1-t_0}\int\limits_{t_0}^{t_1} dt = 1.$$

Тем не менее, индекс Дивизиа результирующей величины обладает свойством среднего в общем случае. В силу аддитивности y_i, этот индекс является обычной средней относительной и, как отмечалось в пункте 2.2, может быть представлен как среднее арифметическое индивидуальных индексов с базисными весами (по знаменателю) или среднее гармоническое индивидуальных индексов с текущими весами (по числителю).

Вслед за мультипликативностью моментных индексов, индексы Дивизиа также мультипликативны. В этом не сложно убедиться, если определить факторные индексы Дивизиа естественным образом:

$$\ln \lambda_x(t_0,t_1) = \int\limits_{t_0}^{t_1} \ln \lambda_x(t)\,dt = \sum \int\limits_{t_0}^{t_1} \alpha_i(t) \ln \lambda_{xi}(t)\,dt,$$

$$\ln \lambda_a(t_0,t_1) = \int\limits_{t_0}^{t_1} \ln \lambda_a(t)\,dt = \sum \int\limits_{t_0}^{t_1} \alpha_i(t) \ln \lambda_{ai}(t)\,dt.$$

Действительно:

$$\ln \lambda_y(t_0,t_1) = \int\limits_{t_0}^{t_1} \ln \underbrace{\lambda_x(t)\,\lambda_a(t)}_{\lambda_y(t)}\,dt = \int\limits_{t_0}^{t_1} \ln \lambda_x(t)\,dt + \int\limits_{t_0}^{t_1} \ln \lambda_a(t)\,dt =$$

$$= \ln \lambda_x(t_0,t_1) + \ln \lambda_a(t_0,t_1),$$

т.е. $\lambda_y(t_0,t_1)=\lambda_x(t_0,t_1)\cdot\lambda_a(t_0,t_1)$.

Факторные индексы не могут быть представлены как средние геометрические индивидуальных индексов — кроме частного случая, когда индивидуальные моментные индексы неизменны во времени. Это было показано на примере индекса результирующей величины. В случае аддитивности x_i факторный индекс объема все-таки обладает свойством среднего (как и индекс результирующей величины). В общем случае факторные индексы требованию среднего не удовлетворяют.

Непосредственно из определения индексов Дивизиа следует их транзитивность. Но факторные индексы этим свойством обладают в специфической, не встречавшейся ранее форме. До сих пор при наличии транзитивности общий за период индекс можно было рассчитать двумя способами: непосредственно по соотношению величин на конец и начало периода или по цепному правилу — произведением аналогичных индексов по подпериодам:

$$\lambda(t_0,t_N)=\lambda(t_0,t_1)\cdot\lambda(t_1,t_2)\cdot\ldots\cdot\lambda(t_{N-1},t_N),\quad t_0<t_1<\ldots<t_N.$$

Именно выполнение этого равенства трактовалось как наличие свойства транзитивности. Теперь (для факторных индексов Дивизиа) это равенство — определение общего индекса (t_0,t_n), т.к. другого способа его расчета — непосредственно по соотношению факторных величин на конец и начало общего периода — не существует. В частности, общий за период факторный индекс зависит не только от значений факторных величин на начало и конец периода, но и от всей внутрипериодной динамики этих величин.

Эту особенность факторных индексов Дивизиа можно проиллюстрировать в случае, когда моментные темпы роста всех индивидуальных величин неизменны во времени. В этом случае, как было показано выше (указатели периода времени (t_0,t_1) опущены),

$$\lambda_y=\prod\lambda_{yi}^{\alpha_i},\quad \lambda_x=\prod\lambda_{xi}^{\alpha_i},\quad \lambda_a=\prod\lambda_{ai}^{\alpha_i}, \tag{3.4}$$

где α_i — средние хронологические веса по результирующей величине y.

Пусть $N=2$, тогда выражение для этих средних хронологических весов можно найти в аналитической форме. Для периода $(0,1)$ ($t_0=0$, $t_1=1$, из таблицы неопределенных интегралов: $\int\frac{dx}{b+ce^{ax}}=\frac{x}{b}-\frac{1}{ab}\ln(b+ce^{ax})$):

$$\alpha_1=\int_0^1\frac{y_1(0)\left(\lambda_{y1}\right)^t}{y_1(0)\left(\lambda_{y1}\right)^t+y_2(0)\left(\lambda_{y2}\right)^t}dt=\frac{\ln\frac{\lambda_{y2}}{\lambda_y}}{\ln\frac{\lambda_{y2}}{\lambda_{y1}}}, \tag{3.5}$$

Таблица 3.2. Объемы производства и цены в три последовательных момента времени

	Моменты времени								
	0			1			2		
i	y	x	a	y	x	a	y	x	a
1	20	10	2	30/45	12/18	2.5	60	20	3
2	10	10	1	30	15	2	90	30	3
Итого	30			60/75			150		

$$\alpha_2 = \int_0^1 \frac{y_2(0)\left(\lambda_{y2}\right)^t}{y_1(0)\left(\lambda_{y1}\right)^t + y_2(0)\left(\lambda_{y2}\right)^t} dt = \frac{\ln\dfrac{\lambda_{y1}}{\lambda_y}}{\ln\dfrac{\lambda_{y1}}{\lambda_{y2}}},$$

где λ_y, λ_{y1}, λ_{y2} — общий (агрегированный) и индивидуальные индексы «момент к моменту» для результирующей величины y. Указанная особенность факторных индексов Дивизиа иллюстрируется на примере, исходные данные для которого приведены в двух таблицах 3.2[1] и 3.3.

Динамика физических объемов дана в 2 вариантах (через знак «/»). Физический объем 1-го продукта в момент времени « 1 » в варианте (а) составляет 12 единиц, в варианте (б) — 18. Это — единственное отличие вариантов.

Результаты расчетов сведены в двух таблицах 3.4 и 3.5.

Расчет средних хронологических весов за периоды $(0,1)$ и $(1,2)$ в 1-й результирующей таблице проводился по формулам (3.5), индексы 2-й таблицы за периоды $(0,1)$

[1] Физические объемы производства продуктов имеют разные единицы измерения (например, тонны и штуки) и не могут складываться, т.е. x неаддитивен.

Таблица 3.3. Индексы наблюдаемых величин

	Периоды времени								
	(0,1)			(1,2)			(0,2)		
i	λ_y	λ_x	λ_a	λ_y	λ_x	λ_a	λ_y	λ_x	λ_a
1	1.5/2.25	1.2/1.8	1.25	2/1.333	1.667/1.111	1.2	3	2	1.5
2	3	1.5	2	3	2	1.5	9	3	3
Итого	2/2.5			2.5/2			5		

Таблица 3.4. Веса индивидуальных индексов

	Варианты									
	(а)					(б)				
	Моменты и периоды времени					Моменты и периоды времени				
i	0	$(0,1)$	1	$(1,2)$	2	0	$(0,1)$	1	$(1,2)$	2
1	0.667	0.585	0.5	0.450	0.4	0.667	0.634	0.6	0.5	0.4
2	0.333	0.415	0.5	0.550	0.6	0.333	0.366	0.4	0.5	0.6
Итого	1	1	1	1	1	1	1	1	1	1

и $(1,2)$ рассчитывались по формулам (3.4), а за период $(0,2)$ — в соответствии с определением по цепному правилу.

Данный пример показывает, что даже относительно небольшое изменение «внутренней» динамики — увеличение физического объема 1-го товара в «средний» момент времени « 1 » с 12 до 18 единиц — привело к увеличению индекса физического объема за весь период $(0,2)$ с 2.426 до 2.510 и к соответствующему снижению индекса цен с 2.061 до 1.992. «Концевые» (на начало и конец периода) значения факторных величин при этом оставались неизменными. В обоих вариантах факторные индексы транзитивны, поскольку индексы за период $(0,2)$ равны произведению индексов за периоды $(0,1)$ и $(1,2)$.

Можно сказать, что факторные индексы Дивизиа обладают свойством транзитивности в усиленной дефинитивной форме, т.к. это свойство определяет сам способ расчета индексов за периоды, включающие подпериоды. Такая особенность факторных индексов в конечном счете является следствием того, что физический объем $x(t)$, как таковой, и относительная величина $a(t)$ в общем случае не на-

Таблица 3.5. Индексы Дивизиа

	Варианты					
	(а)			(б)		
Периоды	λ_y	λ_x	λ_a	λ_y	λ_x	λ_a
$(0,1)$	2.0	1.316	1.519	2.5	1.684	1.485
$(1,2)$	2.5	1.843	1.357	2.0	1.491	1.342
$(0,2)$	5.0	2.426	2.061	5.0	2.510	1.992

блюдаемы, и для их измерения, собственно говоря, и создана теория индексов, в частности индексов Дивизиа. Полезно напомнить, что индекс Дивизиа результирующей величины и все индивидуальные индексы Дивизиа удовлетворяют требованию транзитивности в обычной форме.

Итак, индексы «момент к моменту» продолжают удовлетворять требованиям мультипликативности, транзитивности (в дефинитивной форме), симметричности, но теряют свойство среднего.

Факторные индексы Дивизиа обычно записывают в следующей форме:

$$\lambda_x(t_0,t_1)=\exp\left(\int_{t_0}^{t_1}\frac{\sum a_i(t)\,dx_i(t)}{\sum x_i(t)\,a_i(t)}\right),\quad \lambda_a(t_0,t_1)=\exp\left(\int_{t_0}^{t_1}\frac{\sum x_i(t)\,da_i(t)}{\sum x_i(t)\,a_i(t)}\right).$$

В том, что это форма эквивалентна используемой выше, легко убедиться. Для этого достаточно вспомнить, что, например для индекса объемной величины:

$$\alpha_i(t)=\frac{x_i(t)\,a_i(t)}{\sum x_i(t)\,a_i(t)},\quad \ln\lambda_{xi}(t)=\frac{d\ln x_i(t)}{dt}=\frac{1}{x_i(t)}\frac{dx_i(t)}{dt}.$$

Индексы Дивизиа могут служить аналогом прикладных индексов, рассмотренных в пунктах $1-3$ данного раздела, в случае, если речь идет о величинах x и y типа запаса, поскольку такие величины измеряются на моменты времени.

3) Индексы «период к периоду».

Чаще всего предметом индексного анализа является динамика величин типа потока, поэтому именно непрерывные индексы «период к периоду» являются наиболее полным аналогом прикладных индексов, рассмотренных в пунктах $1-3$ этого раздела.

Сначала необходимо определить следующие индивидуальные величины (здесь и далее нижний индекс-указатель объекта i опущен):

$y(t,\tau)=\int\limits_{t}^{t+\tau}y(t')\,dt'$ — результирующая величина в периоде $[t,t+\tau]$,

$x(t,\tau)=\int\limits_{t}^{t+\tau}x(t')\,dt'$ — объемная величина в периоде $[t,t+\tau]$,

$a(t,\tau)=\dfrac{y(t,\tau)}{x(t,\tau)}=\int\limits_{t}^{t+\tau}\alpha^x(t')\,a(t')\,dt'$ — относительная величина в периоде $[t,t+\tau]$,

где $\alpha^x(t')=\dfrac{x(t')}{x(t,\tau)}$ — временные веса относительной величины.

Таким образом, при переходе к суммарным за период величинам проявилось принципиальное различие объемных и относительных величин. Первые аддитивны во времени и складываются по последовательным моментам времени, вторые — неаддитивны и рассчитываются за период как средние хронологические с весами, определенными динамикой объемной факторной величины. Именно с этим обстоятельством связана возможная несимметричность факторных индексов, которая имеет место для большинства прикладных индексов, рассмотренных в пункте 3.2.

Индивидуальные индексы «период к периоду» строятся естественным способом:

$$\lambda_{[\,]}(t_0, t_1, \tau) = \frac{[\,](t_1, \tau)}{[\,](t_0, \tau)},$$

где $\lambda_{[\,]}(t_0, t_1, \tau)$ — индекс, сопоставляющий периоды $[t_1, t_1 + \tau]$ (текущий) и $[t_0, t_0 + \tau]$ (базисный), а на месте $[\,]$, как и прежде, стоит либо y — для объемной результирующей величины (стоимости), либо x — для объемной факторной величины (объема), либо a — для относительной величины (цены).

Если динамика (траектория изменения) показателя $[\,](t)$ в базисном и текущем периодах одинакова, то для любого $t \in [t_0, t_0 + \tau]$ индекс «момент к моменту» $\lambda_{[\,]}(t, t+t_1-t_0)$ неизменен и равен $\lambda_{[\,]}(t_0, t_1)$. Тогда для любого $t \in [t_1, t_1 + \tau]$ имеет место равенство $[\,](t) = [\,](t - t_1 + t_0)\,\lambda_{[\,]}(t_0, t_1)$, и индекс «период к периоду» объемной величины ($[\,]$ — есть либо y, либо x) можно представить следующим образом:

$$\lambda_{[\,]}(t_0, t_1, \tau) = \frac{\overbrace{\int_{t_1}^{t_1+\tau} [\,](t - t_1 + t_0)\,\overbrace{\lambda_{[\,]}(t_0, t_1)}^{=const}\,dt}^{=[\,](t_1,\tau)}}{\underbrace{\int_{t_0}^{t_0+\tau} [\,](t)\,dt}_{=[\,](t_0,\tau)}} =$$

$$= \lambda_{[\,]}(t_0, t_1)\,\underbrace{\frac{\int_{t_1}^{t_1+\tau} [\,](t - t_1 + t_0)\,dt}{\int_{t_0}^{t_0+\tau} [\,](t)\,dt}}_{=1} = \lambda_{[\,]}(t_0, t_1)\,,$$

т.е. он совпадает с индексом «момент к моменту».

Для того чтобы такое же равенство имело место для индексов относительной величины, необходима идентичность динамики в базисном и текущем периодах времени не только самой относительной величины, но и объемной факторной величины. Иначе веса $\alpha^x(t)$ в базисном и текущем периодах времени будут различны

и интегралы в числителе и знаменателе выражения индекса «период к периоду» относительной величины (после выноса $\lambda_a(t_0, t_1)$ за знак интеграла в числителе) не будут равны друг другу.

Тогда, если в базисном и текущем периодах времени одинакова динамика всех индивидуальных величин, то индексы «период к периоду» совпадают с индексами Дивизиа. Чаще всего считается, что различия в динамике индивидуальных величин в базисном и текущем периодах времени не существенны, и в качестве непрерывных аналогов прикладных индексов поэтому принимают индексы Дивизиа. Именно на таком допущении построено изложение материала в следующем пункте.

В случае, если указанные различия в динамике величин принимаются значимыми, приходится вводить поправочные коэффициенты к индексам Дивизиа, чтобы приблизить их к индексам «период к периоду». Теоретический анализ таких индексных систем в непрерывном времени затруднен и не дает полезных для практики результатов.

3.6. Прикладные следствия из анализа индексов в непрерывном времени

Теоретически «правильными» в этом пункте принимаются индексы Дивизиа. Это предположение можно оспаривать только с той позиции, что внутренняя динамика сопоставляемых периодов времени существенно различается. Здесь предполагается, что эти различия не значимы. Из проведенного выше анализа индексов Дивизиа следует по крайней мере три обстоятельства, важные для построения прикладных индексов.

1) Факторные индексы за период, включающий несколько «единичных» подпериодов, правильно считать по цепному правилу, а не непосредственно из сопоставления величин на конец и на начало всего периода. Для иллюстрации разумности такого подхода проведены расчеты в условиях примера, приведенного в конце предыдущего пункта. Результаты этих расчетов сведены в таблицу 3.6.

Из приведенных данных видно, что

– во-первых, агрегатные индексы, рассчитанные в целом за период (по «концам»), не реагируют, по понятным причинам, на изменение внутренней динамики и одинаковы для вариантов (а) и (б); индекс Ласпейреса — особенно в варианте (а) — заметно преуменьшает реальный (по Дивизиа) рост физического объема, индекс Пааше, наоборот, преувеличивает этот рост. В другой (числовой) ситуации индекс Ласпейреса мог бы преувеличивать, а индекс Пааше преуменьшать реальную динамику. Важно то, что оба эти индекса дают оценки динамики существенно отличные от реальной.

Таблица 3.6

	Варианты					
	(а)			(б)		
Индексы:	λ_y	λ_x	λ_a	λ_y	λ_x	λ_a
Дивизиа	5.0	2.426	2.061	5.0	2.510	1.992
В целом за период — (02)						
(1) Ласпейрес—Пааше	5.0	2.333	2.143	5.0	2.333	2.143
(2) Пааше—Ласпейрес	5.0	2.500	2.000	5.0	2.500	2.000
(3) Фишер	5.0	2.415	2.070	5.0	2.415	2.070
По цепному правилу — (012)						
(1) Ласпейрес—Пааше	5.0	2.383	2.098	5.0	2.493	2.005
(2) Пааше—Ласпейрес	5.0	2.469	2.025	5.0	2.525	1.980
(3) Фишер	5.0	2.426	2.061	5.0	2.509	1.993

— во-вторых, рассчитанные по цепному правилу индексы имеют более реалистичные значения. Так, например, реальный рост физического объема в варианте (а), равный 2.426, заметно преуменьшенный индексом Ласпейреса в целом за период — 2.333, получает более точную оценку тем же индексом Ласпейреса, рассчитанным по цепному правилу, — 2.383. Цепные индексы дают более правильные оценки динамики. Но, вообще говоря, это свойство цепных индексов не гарантировано. Так, в варианте (б) физический рост 2.510 преуменьшен индексом Пааше в целом за период — 2.500 (хотя и в меньшей степени, чем индексом Ласпейреса — 2.333), и преувеличен этим же индексом по цепному правилу — 2.525.

Принимая предпочтительность цепного правила, следует с сомнением отнестись к принятым правилам расчета объемов в неизменных (сопоставимых) ценах: (x^0, a^0), (x^1, a^0), ..., (x^t, a^0), ..., (x^N, a^0) (см. п. 3.2). Правильнее считать физический объем в году t в ценах, сопоставимых с базисным периодом, как $y^0 \lambda_x^{01} \cdot \ldots \cdot \lambda_x^{t-1,t}$ или $y^t / \left(\lambda_a^{01} \cdot \ldots \cdot \lambda_a^{t-1,t}\right)$. В этом случае теряется наглядность, но приобретается соответствие теории. Следует отметить, что в действующей сейчас Системе национальных счетов, рекомендованных ООН в 1993 г. для использования национальными правительствами, при расчете индексов применяется цепное правило, но при расчете физических объемов в неизменных ценах — обычный под-

ход, основанный на индексах Пааше в целом за период. Это противоречие остается, по-видимому, для сохранения принципа наглядности.

2) Индексы Дивизиа рассчитываются как средние индивидуальных индексов с некоторыми весами, занимающими промежуточное положение между базисным и текущим моментами (периодами) времени. Из рассмотренных прикладных индексов такому подходу в большей степени удовлетворяют индексы Фишера.

Действительно, в рассматриваемом примере индекс физического объема Фишера в целом за период — 0.415 — более точно отражает реальную динамику, чем индекс Пааше или Ласпейреса — в варианте (а). В варианте (б) более точным оказывается индекс физического объема Пааше. Зато индексы Фишера, рассчитанные по цепному правилу, дают практически точное приближение к реальной динамике.

3) Если предположить (как это делалось в предыдущем пункте), что индивидуальные моментные индексы всех величин не меняются во времени в отдельных периодах, то расчет индексов Дивизиа как средних геометрических индивидуальных индексов становиться вполне операциональным. Сложность заключается лишь в определении средних хронологических весов по результирующей величине. В случае двух продуктов соответствующие интегралы, как это показано в предыдущем пункте, берутся аналитически. В общем случае их всегда можно найти численным приближением. Однако такой подход вряд ли применим в практике, поскольку он достаточно сложен с точки зрения вычислений и не обладает наглядностью хоть в какой-нибудь степени. Возможен компромисс, при котором веса для средних геометрических индивидуальных индексов находятся как средние базисных и текущих долей объектов в результирующей величине по формуле, более простой и наглядной, чем интеграл теоретической средней хронологической.

Для индекса результирующей величины, которая аддитивна по объектам, справедливы следующие соотношения:

$$\lambda_y^{rs} = \sum \alpha_i^r \lambda_{yi}^{rs} = \frac{1}{\sum \left(\alpha_i^s \big/ \lambda_{yi}^{rs} \right)},$$

где α_i^r, α_i^s — доли объектов в результирующей величине, соответственно, в базисном и текущем периодах времени.

Теперь рассчитываются два индекса результирующей величины $\lambda_y^{rs}(r)$, $\lambda_y^{rs}(s)$ как средние геометрические индивидуальных индексов по весам, соответственно, базисного и текущего периодов:

$$\lambda_y^{rs}(r) = \prod \left(\lambda_{yi}^{rs} \right)^{\alpha_i^r}, \quad \lambda_y^{rs}(s) = \prod \left(\lambda_{yi}^{rs} \right)^{\alpha_i^s}.$$

По свойству мажорантности средних степенных:

$$\lambda_y^{rs}(r) < \lambda_y^{rs} < \lambda_y^{rs}(s),$$

и уравнение относительно γ^{rs}:

$$\lambda_y^{rs} = \left(\lambda_y^{rs}(r)\right)^{\gamma^{rs}} \left(\lambda_y^{rs}(s)\right)^{1-\gamma^{rs}},$$

будет иметь решение $0 < \gamma^{rs} < 1$.

Тогда $\alpha_i^{rs} = \gamma^{rs}\alpha_i^r + (1-\gamma^{rs})\,\alpha_i^s$ могут сыграть роль средних хронологических весов в формулах индексов Дивизиа (соотношения, аналогичные (3.4)):

$$\lambda_y^{rs} = \prod \left(\lambda_{yi}^{rs}\right)^{\alpha_i^{rs}}, \quad \lambda_x^{rs} = \prod (\lambda_{xi}^{rs})^{\alpha_i^{rs}}, \quad \lambda_a^{rs} = \prod (\lambda_{ai}^{rs})^{\alpha_i^{rs}}.$$

Теперь эти соотношения являются формулами расчета прикладных индексов, обладающих всеми свойствами теоретических индексов Дивизиа: они мультипликативны, транзитивны (в дефинитивной форме), симметричны и являются средними индивидуальных индексов.

В прикладном анализе иногда используют похожие индексы, называемые по имени автора **индексами Торнквиста**. В их расчете в качестве γ^{rs} всегда принимают 0.5, и потому индекс результирующей величины Торнквиста не равен в общем случае его реальному значению. Предложенные здесь индексы можно назвать **модифицированными индексами Торнквиста**.

Для того чтобы оценить качество прикладных индексов, проводился численный эксперимент, в котором значения факторных признаков (объем и цена) задавались случайными числами (случайными величинами, равномерно распределенными на отрезке $[0,1]$), и определялись отклонения прикладных индексов от значения теоретического индекса Дивизиа (по абсолютной величине логарифма отношения прикладного индекса к теоретическому). Рассматривались три системы: 2 продукта — 2 периода (как в приводимом выше примере), 2 продукта — 3 периода, 3 продукта — 2 периода. В случае двух продуктов значения модифицированного индекса Торнквиста и индекса Дивизиа совпадают, т.к. уравнение

$$\lambda_y^{rs} = \left(\lambda_{y1}^{rs}\right)^{\alpha_1^{rs}} \left(\lambda_{y2}^{rs}\right)^{1-\alpha_1^{rs}}$$

имеет относительно α_1^{rs} единственное решение. Поэтому в этих случаях индекс Дивизиа сравнивался с индексами Ласпейреса, Пааше и Фишера, рассчитанными в целом за период и по цепному правилу. В случае 3-х продуктов индекс Дивизиа, рассчитанный с использованием численной оценки интеграла среднехронологических весов (для этого единичный период времени делился на 100 подпериодов), сравнивался также и с модифицированным индексом Торнквиста. В каждом из этих трех случаев проводилось около $1\,000\,000$ численных расчетов, поэтому полученные оценки вероятностей достаточно точны.

Оценки вероятности для случая «2 продукта — 2 периода» приведены в таблице 3.7. В этой же таблице стрелочками вверх и вниз отмечено, как меняются

Таблица 3.7. Вероятности того, что индекс в подлежащем дает большую ошибку, чем индекс в сказуемом таблицы (для индексов объемной факторной величины)

	В целом за период			По цепному правилу	
	Ласпейрес	Пааше	Фишер	Ласпейрес	Пааше
В целом за период					
Пааше	0.500	0	—	—	—
Фишер	0.415 ↓↓	0.415 ↓↓	0	—	—
По цепному правилу					
Ласпейрес	0.482 ↑↓	0.479 ↑↓	0.524 ↑↑	0	—
Пааше	0.479 ↑↓	0.482 ↑↓	0.524 ↑↑	0.500	0
Фишер	0.052 ↑↑	0.052 ↑↑	0.060 ↑↑	0.053 ↑↑	0.053 ↑↑

соответствующие показатели при переходе к ситуации «2 продукта — 3 периода» и далее «3 продукта — 2 периода».

По данным этой таблицы преимущество цепного правила проявляется не столь очевидно. Цепные индексы Ласпейреса и Пааше лишь в 48% случаев (чуть меньше половины) дают более высокую ошибку, чем те же индексы, рассчитанные в целом за период. Это преимущество растет (падает соответствующий показатель вероятности) с увеличением числа объектов (продуктов) в агрегате и исчезает с увеличением числа периодов (при 3-х периодах соответствующие вероятности становятся больше 0.5). Зато преимущество индекса Фишера становится явным. Рассчитанные в целом за период, эти индексы хуже соответствующих индексов Ласпейреса и Пааше только в 41.5% случаев, причем их качество повышается с ростом как числа объектов, так и количества периодов. Особенно «хороши» цепные индексы Фишера: они лишь в 5–6% случаев дают ошибку большую, чем любые другие индексы. К сожалению, с ростом числа объектов и количества периодов их качество снижается.

В ситуации «3 продукта — 2 периода» рассчитывались модифицированные индексы Торнквиста. Они оказались самыми лучшими. Вероятность того, что они дают более высокую ошибку, чем индексы Ласпейреса и Пааше, а также Фишера, рассчитанного в целом за период, на 2–3% ниже, чем для цепного индекса Фишера.

Итак, можно сказать, что модифицированные индексы Торнквиста, рассчитываемые как средние геометрические индивидуальных индексов с особыми весами,

в наилучшей степени соответствуют теории. Тем не менее, в существующей практике статистики индексы как средние геометрические величины фактически не применяются. В действующей (рекомендованной ООН в 1993 г.) Системе национальных счетов применение индексов Торнквиста (обычных, не модифицированных) рекомендуется лишь в весьма специфических ситуациях.

Индексы как средние геометрические индивидуальных применялись в практике статистики, в том числе в России и СССР, в первой трети XX века. Затем практически всеобщее распространение получили агрегатные индексы. Это произошло по крайней мере по двум причинам. Первая: агрегатные индексы наглядны и поэтому понятны. Вторая: средние геометрические величины, если веса взвешивания принять за константы, весьма чувствительны к крайним значениям индивидуальных индексов. Так, например, очень большое значение какого-то одного индивидуального индекса приведет к существенному преувеличению общего индекса (в крайней ситуации, когда базисное значение индивидуальной величины равно нулю, т.е., например, какой-то продукт в базисном периоде еще не производился, общий индекс окажется бесконечным). Наоборот, очень малое значение единственного индивидуального индекса существенно преуменьшит общий индекс (обратит его в ноль, если текущее значение соответствующей индивидуальной величины равно нулю — данный продукт уже не производится в текущем периоде).

Указанные доводы против среднегеометрических индексов вряд ли серьезны. По поводу первого из них следует еще раз напомнить, что наглядность и понятность нельзя считать критерием истины. Второй довод не выдерживает критики, поскольку резким изменениям могут подвергаться малые индивидуальные величины, которые входят в среднюю с малыми весами и поэтому не могут заметно повлиять на ее уровень. В крайних ситуациях, когда индивидуальный индекс по какому-то объекту принимает нулевое или бесконечное значение, такой объект вообще не должен участвовать в расчете общего индекса (его вес в среднем геометрическом равен нулю).

Действительно,

$$\lambda_y(0,1) = \prod_{i=1}^{N} \left(\frac{y_i(1)}{y_i(0)} \right)^{\alpha_i(0,1)},$$

где по определению $\dfrac{y_i(1)}{y_i(0)} = \lambda_{yi}(0,1)$, а

$$\alpha_i(0,1) = \int_0^1 \frac{y_i(0) \left(\frac{y_i(1)}{y_i(0)} \right)^t}{\sum\limits_{i=1}^{N} y_i(0) \left(\frac{y_i(1)}{y_i(0)} \right)^t} dt = \int_0^1 \frac{y_i(0)^{1-t} y_i(1)^t}{\sum\limits_{i=1}^{N} y_i(0)^{1-t} y_i(1)^t} dt.$$

Далее рассматривается только компонента i-го объекта $\left(\frac{y_i(1)}{y_i(0)}\right)^{\alpha_i(0,1)}$ (обозначаемая ниже $\tilde{\lambda}_{yi}$), для которого либо $y_i(0)$, либо $y_i(1)$ равны нулю (продукт либо еще не производился в базисном моменте, либо уже не производится в текущем моменте времени).

Пусть период времени $[0,1]$ делится на n равных подпериодов, и t_j — середина j-го подпериода. Тогда рассматриваемую компоненту $\tilde{\lambda}_{yi}$ можно приближенно представить выражением (в силу аддитивности интеграла)

$$\prod_{j=1}^{n}\tilde{\lambda}_{yij},\ \text{где}\ \tilde{\lambda}_{yij}=\left(\frac{y_i(1)}{y_i(0)}\right)^{\frac{y_i(0)^{1-t_j}\,y_i(1)^{t_j}}{\sum_{i=1}^{N} y_i(0)^{1-t_j}\,y_i(1)^{t_j}}\frac{1}{n}},$$

которое в пределе при $n\to\infty$ совпадет с исходным значением этой компоненты. При конкретном $n<\infty$ и любом t_j, которое в таком случае обязательно больше нуля и меньше единицы,

$$\tilde{\lambda}_{yij}\to 1,$$

при $y_i(0)\to 0$ или $y_i(1)\to 0$. Это можно доказать аналитически, но проще показать численно. В первом случае ($y_i(0)\to 0$) указанная величина $\tilde{\lambda}_{yij}$ стремится к единице сверху, во втором — снизу, т.е. в крайней ситуации, когда либо $y_i(0)$, либо $y_i(1)$ равны нулю, $\prod_{j=1}^{n}\tilde{\lambda}_{yij}$ равно единице. И в результате перехода в этом выражении к пределу при $n\to\infty$ оказывается, что компонента i-го объекта $\tilde{\lambda}_{yi}$ также равна единице. Это означает, что данный i-й объект не участвует в расчете среднегеометрического индекса.

Индексы Дивизиа при гипотезе неизменности во времени всех индивидуальных моментных индексов, а вслед за ними — модифицированные индексы Торнквиста — должны рассчитываться по **сопоставимому набору** объектов (продуктов). В такой набор входят только такие объекты, которые существовали как в базисном, так и в текущем периодах времени (только те продукты, которые производились и в базисном, и в текущем периодах). Это правило выступает дополнительным аргументом в пользу цепных индексов, поскольку за длительные периоды времени наборы объектов (продуктов) могут меняться заметно, тогда как их изменения за короткие единичные периоды не так существенны.

В заключение следует заметить, что мультипликативные индексные выражения, построенные на основе индексов Дивизиа и модифицированных индексов Торнквиста, естественным образом обобщаются на случай более одного относительного фактора в мультипликативном представлении результирующей величины.

3.7. Факторные представления приростов в непрерывном времени

Моментные приросты делятся на факторы естественным и однозначным образом:

$$\Delta\lambda_y(t) = \frac{d\ln y(t)}{dt} = \frac{d\ln x(t)}{dt} + \frac{d\ln a(t)}{dt} = \Delta\lambda_x(t) + \Delta\lambda_a(t).$$

Принимая во внимание, что непрерывным за период темпом прироста $\Delta\lambda_y(t_0,t_1)$ является $\ln\lambda_y(t_0,t_1)$, аналогично делятся на факторы и приросты за период (т.к. индексы «момент к моменту» мультипликативны):

$$\Delta\lambda_y(t_0,t_1) = \ln\lambda_y(t_0,t_1) = \ln\lambda_x(t_0,t_1) + \ln\lambda_a(t_0,t_1) = \\ = \Delta\lambda_x(t_0,t_1) + \Delta\lambda_a(t_0,t_1).$$

В прикладном анализе такое правило деления приростов на факторы также вполне операционально, и его имеет смысл использовать.

Каждому мультипликативному индексному выражению $\lambda_y^{rs} = \lambda_x^{rs}\lambda_a^{rs}$ следует сопоставить не три варианта факторных разложений $(1''-3'')$, как в пункте 3.3, а одно:

$$\ln\lambda_y^{rs} = \ln\lambda_x^{rs} + \ln\lambda_a^{rs}.$$

Однако, поскольку $\ln\lambda_y^{rs} \neq \dfrac{\Delta_y^{rs}}{y^r}$, правильнее из этого факторного разложения определять лишь доли экстенсивных и интенсивных факторов:

$$\gamma_x^{rs} = \frac{\ln\lambda_x^{rs}}{\ln\lambda_y^{rs}}, \quad \gamma_a^{rs} = \frac{\ln\lambda_a^{rs}}{\ln\lambda_y^{rs}},$$

которые, в свою очередь, использовать в расчете вкладов факторов:

$$\Delta_x^{rs} = \gamma_x^{rs}\Delta_y^{rs}, \quad \Delta_a^{rs} = \gamma_a^{rs}\Delta_y^{rs}.$$

Такой подход успешно работает при любом количестве относительных факторов в мультипликативном представлении результирующей величины.

3.8. Задачи

1. Определить пункты, которые являются выпадающими из общего ряда.

1.1. а) Ласпейрес, б) Пирсон, в) Фишер, г) Пааше;

1.2. а) Ласпейрес, б) Пааше, в) Фишер, г) Торнквист;

1.3. а) индекс, б) дефлятор, в) корзина, г) коробка;

1.4. а) Ласпейрес, б) транзитивность, в) мультипликативность, г) Пааше;

1.5. а) коммутативность, б) транзитивность, в) мультипликативность, г) симметричность;

1.6. а) Торнквист, б) цепное правило, в) транзитивность, г) Фишер;

1.7. а) прирост, б) экстенсивные, в) интегральные, г) интенсивные;

1.8. а) дефлятор, б) темп роста, в) индекс, г) темп прироста;

1.9. а) постоянного состава, б) относительная величина, в) структуры, г) стоимости;

1.10. а) цепное, б) обратимости, в) симметрии, г) среднего;

1.11. а) среднего, б) транзитивности, в) обратимости, г) цепное;

1.12. а) дефлятор, б) темп роста, в) симметрии, г) среднего;

1.13. а) частный, б) факторный, в) цен, г) стоимости;

1.14. а) непрерывность, б) Дивизиа, в) геометрическое, г) дискретность;

1.15. а) сопоставимый набор, б) цепное правило, в) Торнквист, г) Фишер.

2. Индексы стоимости и объема для совокупности из 2 товаров равны соответственно 1.6 и 1.0. Стоимость в текущий период распределена между товарами поровну. Индивидуальный индекс цен для одного из товаров равен 1.3, чему он равен для другого товара?

3. Объемы производства 2 товаров в базисном и текущем периодах равны 10, 20 и 30, 40 единиц, соответствующие цены — 2, 1 и 4, 3. Чему равны индексы объема Ласпейреса и Пааше? Чему равны те же индексы цен?

4. Объемы производства 2 товаров в базисном и текущем периодах равны 10, 20 и 30, 40 единиц, соответствующие цены — 2, 1 и 4, 3. Чему равен вес 1-го товара в индексе Торнквиста? Чему равны факторные индексы Дивизиа?

5. Объемы производства 2 товаров в базисном и текущем периодах равны 10, 20 и 20, 10 тыс. руб., материалоемкости их производства — 0.6, 0.5 и 0.7, 0.6. Чему равны индексы структурных сдвигов Ласпейреса и Пааше? Чему равны те же индексы постоянного состава?

6. Объемы производства 2 товаров в базисном и текущем периодах равны 10, 20 и 20, 10 тыс. руб., материалоемкости их производства — 0.6, 0.5 и 0.7, 0.6. Чему равны факторные индексы («количества» и «качества») в «тройке»: материальные затраты равны объемам производства, умноженным на материалоемкость?

7. В 1999 году ВВП в текущих ценах составил 200 млрд. руб. В 2000 году ВВП в текущих ценах вырос на 25%, а в сопоставимых снизился на 3%. Определите дефлятор ВВП.

8. Сумма удорожания продукции за счет повышения цен составила 200 млн. руб., прирост физического объема продукции составил 300 млн. руб. На сколько процентов повысились цены и возрос физический объем продукции, если стоимость продукции в базисном периоде составила 3 млрд. руб.?

9. Физический объем продукции возрос на 300 млн. руб., или на 20%. Цены снизились на 10%. Найти прирост стоимости продукции с учетом роста физического объема продукции и снижении цен.

10. Стоимость продукции в текущем периоде в текущих ценах составила 1600 млн. руб. Индекс цен равен 0.8, индекс физического объема — 1.2. Определить прирост стоимости продукции, в том числе обусловленный ростом физического объема продукции и снижением цен.

11. Расходы на потребительские товары составили 20 тыс. руб., что в текущих ценах больше соответствующих расходов прошлого года в 1.2 раза, а в сопоставимых ценах на 5% меньше. Определите индекс цен на потребительские товары и изменение их физического объема (абсолютно и относительно).

12. По данным, приведенным в таблице, рассчитайте:

Показатель	Продукт	Базовый период	Текущий период
Объем производства, тыс. т	сталь	2400	3800
	чугун	3700	4800
Цена, тыс. руб./т	сталь	1.5	3.0
	чугун	1.0	0.8

   а) индивидуальные и общие индексы изменения стоимости;

   б) индексы Ласпейреса, Пааше, Фишера цен и физического объема.

13. По данным, приведенным в таблице:

Показатель	Отрасль	Базовый период	Текущий период
Валовый выпуск	растениеводство	720	1800
	животноводство	600	900
Численность занятых	растениеводство	200	250
	животноводство	300	330

а) рассчитайте производительность труда по отраслям и сельскому хозяйству в целом;

б) рассчитайте одним из методов влияние изменения отраслевых показателей численности занятых и производительности труда на динамику валового выпуска сельского хозяйства.

14. По данным, приведенным в таблице, рассчитайте:

Годы	ВВП (текущие цены, трлн. руб.)	Индексы-дефляторы ВВП (в разах к предыдущему году)
1990	0.644	1.2
1991	1.398	2.3
1992	19.006	15.9
1993	171.510	9.9
1994	610.592	4.1
1995	1630.956	2.8

а) ВВП России в 1991–1995 гг. в сопоставимых ценах 1990 г.;

б) базовые индексы-дефляторы.

15. Используя один из подходов, вычислите индексы товарооборота, физического объема и цен в целом по мясопродуктам на основании данных из таблицы.

Мясопродукты	Розничный товарооборот, млрд. руб.		Индекс цен, %
	март	апрель	
Мясо	1128		1517
Колбасные изделия	2043		3120
Мясные консервы	815		1111

16. Вычислите общие индексы стоимости, физического объема и цен по закупкам мяса на основании данных из таблицы:

	Год	Мясо		
		Говядина	Свинина	Баранина
Количество проданного мяса, тыс. т	базисный отчетный	238 245	183 205	40 48
Закупочная цена за 1 т, тыс. руб.	базисный	35	30	28

Закупочные цены в отчетном году по сравнению с базисным возросли на говядину — на 160%, свинину — на 80%, на баранину — на 50%.

17. По данным, приведенным в таблице, рассчитайте:

Показатель	Регион	Базовый период	Текущий период
Валовой выпуск	Западная Сибирь	3600	4000
	Восточная Сибирь	2700	2500
Производственные затраты	Западная Сибирь	2400	2500
	Восточная Сибирь	2000	2200

а) материалоемкость производства по регионам и Сибири в целом;

б) индексы переменного и постоянного состава и структурных сдвигов материалоемкости.

18. По данным, приведенным в таблице, рассчитайте:

Показатель	Подразделение	Базовый период	Текущий период
Валовой выпуск	1-й цех	80	160
	2-й цех	120	90
Основной капитал	1-й цех	50	75
	2-й цех	240	240

а) фондоотдачу по цехам предприятия и заводу в целом;

б) индексы переменного и постоянного состава и структурных сдвигов фондоотдачи.

19. Используя один из подходов, вычислите общие индексы стоимости, физического объема и цен по закупкам зерновых на основании следующих данных:

	Год	Зерновые		
		пшеница	рожь	гречиха
Количество проданного зерна, тыс. т	базисный отчетный	548 680	385 360	60 75
Закупочная цена за 1 т, тыс. руб.	отчетный	7.2	7.0	12

Закупочные цены в отчетном году по сравнению с базисным возросли на пшеницу — на 60%, рожь — на 40%, гречиху — 50%.

Рекомендуемая литература

1. **Аллен Р.** Экономические индексы. — М.: «Статистика», 1980. (Гл. 1, 5).
2. **(*) Зоркальцев В.И.** Индексы цен и инфляционные процессы. — Новосибирск: «Наука», 1996. (Гл. 1, 4–6, 15).
3. **Кёвеш П.** Теория индексов и практика экономического анализа. — М.: «Финансы и статистика», 1990.

Глава 4

Введение в анализ связей

Одна из задач статистики состоит в том, чтобы по данным наблюдений за признаками определить, связаны они между собой (зависят ли друг от друга) или нет. И если зависимость есть, то каков ее вид (линейный, квадратичный, логистический и т.д.) и каковы ее параметры. Построенные зависимости образуют эмпирические (эконометрические) модели, используемые в анализе и прогнозировании соответствующих явлений. Часто задача ставится иначе: используя данные наблюдений, подтвердить или опровергнуть наличие зависимостей, следующих из теоретических моделей явления. Математические методы решения этих задач во многом идентичны, различна лишь содержательная интерпретация их применения.

В этой главе даются самые элементарные сведения об этих методах. Более развернуто они представлены в следующих частях книги.

4.1. Совместные распределения частот количественных признаков

Пусть имеется группировка совокупности по n признакам (см. п. 1.9), где $n > 1$, и N_I — количество объектов в I-й конечной группе (групповая численность), т.е. частота одновременного проявления 1-го признака в i_1-м полуинтервале, 2-го признака в i_2-м полуинтервале и т.д., n-го признака в i_n-м полуинтервале (уместно напомнить, что $I = i_1 i_2 \dots i_n$, см. п. 1.9). Как и прежде, $\alpha_I = \dfrac{N_I}{N}$ — относительные частоты или оценки вероятности того,

что $z_{i_1-1,\,1} < x_1 \leqslant z_{i_1 1},\ \ldots,\ z_{i_n-1,\,n} < x_n \leqslant z_{i_n n}$ (если $i_j = 1$, то левые строгие неравенства записываются как $\leqslant$).

Пусть $\Delta_{i_j(j)}$ — длина i_j-го полуинтервала в группировке по j-му фактору, а $\Delta_I = \prod\limits_{j=1}^{n} \Delta_{i_j(j)}$. Тогда $f_I = \dfrac{\alpha_I}{\Delta_I}$ — плотности относительной частоты совместного распределения или оценки плотности вероятности.

Очевидно[1], что $\sum\limits_{I=I_1}^{I_K} \alpha_I = 1$, или

$$\sum_I f_I \Delta_I = 1. \tag{4.1}$$

Далее: $F_I = \sum\limits_{I' \leqslant I} \alpha_{I'}$ ($F_I = \sum\limits_{i'_1=1}^{i_1} \ldots \sum\limits_{i'_n=1}^{i_n} \alpha_{I'}$ — новая по сравнению с п. 1.9 операция суммирования) или

$$F_I = \sum_{I' \leqslant I} f_{I'} \Delta_{I'} \tag{4.2}$$

— накопленные относительные частоты совместного распределения, или оценки вероятностей того, что $x_j \leqslant z_{i_j j},\ j = 1,\ \ldots,\ n$. F_0 — оценка вероятности того, что $x_j < z_{0j},\ j = 1,\ \ldots,\ n$, т. е. $F_0 = 0$. $F_{I_K} = 1$.

Введенные таким образом совместные распределения частот признаков являются прямым обобщением распределения частоты одного признака, данного в пункте 2.1.

Аналогичным образом можно ввести совместные распределения любого подмножества признаков, которое обозначено в пункте 1.9 через J, т.е. по группам более низкого порядка, чем конечные, образующим класс J. Для индексации этих групп в этом разделе будет использован 2-й способ (см. п. 1.9) — составной мультииндекс $I(J)$, в котором и из I, и из J исключены все $*$. Так, индекс $\mathbf{51}(\mathbf{13})$ именует группу, в которой 1-й признак находится на 5-м уровне, 3-й — на 1-м, а остальные признаки «пробегают» все свои уровни. При 1-м способе (используемом в п. 1.9) и при $n = 3$ эта группа именуется двумя мультииндексами $\mathbf{5{*}1}$ и $\mathbf{1{*}3}$. Введенное выше обозначение длин полуинтервалов $\Delta_{i_j(j)}$ построено по этому 2-му способу.

Распределение частот признаков множества J, т.е. по группам класса J определяется следующим образом.

[1] Операция такого суммирования объясняется в пункте 1.9, тогда же через I_K был обозначен мультииндекс, в котором все факторы находятся на последнем уровне; в данном случае эту операцию можно записать так: $\sum\limits_{i_1=1}^{k_1} \ldots \sum\limits_{i_n=1}^{k_n} \alpha_I = 1$.

$N_{I(J)}$ — частота, количество объектов, попавших в группу $I(J)$. Если вернуться к обозначениям пункта 1.9 для мультииндекса этой группы — $I(*)$ (в полном мультииндексе I все те позиции, которые соответствуют не вошедшим в J признакам, заменены на $*$, например: $\mathbf{51(13)} \to \mathbf{5{*}1}$, и воспользоваться введенной в том же пункте операцией $\sum\limits_{I(*)}$, то

$$N_{I(J)} = \sum_{I(*)} N_I.$$

Но в данном случае обозначение этой операции следует уточнить. Пусть $\bar{J}$ — множество тех признаков, которые не вошли в J, а операция ‘ $+$ ’ в соответствующем контексте такова, что $J + \bar{J} = G$ (через G в п. 1.9 было обозначено полное множество факторов $\{12\ldots n\}$) и $I(J) + I(\bar{J}) = I$ (например, $\mathbf{13} + \mathbf{2} = \mathbf{123}$ и $\mathbf{51(13)} + \mathbf{3(2)} = \mathbf{531}$), тогда

$$N_{I(J)} = \sum_{\bar{J}} N_{I(J)+I(\bar{J})},$$

где суммирование ведется по всем уровням признаков указанного под знаком суммирования множества (далее операция $\sum\limits_{\text{мн-во призн.}}$ будет пониматься именно в этом смысле).

$\alpha_{I(J)} = \dfrac{N_{I(J)}}{N}$ — относительные частоты, которые, очевидно, удовлетворяют условию: $\sum\limits_{J} \alpha_{I(J)} = 1$,

$f_{I(J)} = \dfrac{\alpha_{I(J)}}{\Delta_{I(J)}}$ — плотности относительной частоты, где $\Delta_{I(J)} = \prod\limits_{J} \Delta_{i_j(j)}$ (операция такого перемножения объясняется в п. 1.9),

$F_{I(J)} = \sum\limits_{I'(J) \leqslant I(J)} \alpha_{I'(J)}$ накопленные относительные частоты, где $I'(J)$ — текущие («пробегающие») значения уровней признаков J.

Такие распределения по отношению к исходному распределению в полном множестве признаков называются **маргинальными** (**предельными**), поскольку накопленные относительные частоты (эмпирический аналог функции распределения вероятностей) таких распределений получаются из накопленных относительных частот исходного распределения заменой в них на предельные уровни k_j факторов, не вошедших в множество J:

$$F_{I(J)} = F_{I(J)+I_K(\bar{J})}. \tag{4.3}$$

Действительно, поскольку вслед за $N_{I(J)}$

$$\alpha_{I(J)} = \sum_{\bar{J}} \alpha_{I(J)+I(\bar{J})}, \tag{4.4}$$

то

$$F_{I(J)} = \sum_{I'(J)\leqslant I(J)} \alpha_{I'(J)} = \sum_{I'(J)\leqslant I(J)} \sum_{\bar{J}} \alpha_{I'(J)+I(\bar{J})} = \\ = \sum_{\substack{I'(J)\ \leqslant I(J) \\ I'(\bar{J})\ \leqslant I_K(\bar{J})}} \alpha_{I'(J)+I'(\bar{J})} = F_{I(J)+I_K(\bar{J})}.$$

Кроме того,

$$f_{I(J)} = \sum_{\bar{J}} f_{I(J)+I(\bar{J})} \Delta_{I(\bar{J})}, \tag{4.5}$$

т.к. $\Delta_I = \Delta_{I(J)} \Delta_{I(\bar{J})}$.

Действительно:

$$\sum_{\bar{J}} f_{I(J)+I(\bar{J})} \Delta_{I(\bar{J})} = \sum_{\bar{J}} \frac{\alpha_{I(J)+I(\bar{J})}}{\Delta_{I(J)} \Delta_{I(\bar{J})}} \Delta_{I(\bar{J})} = \frac{1}{\Delta_{I(J)}} \sum_{\bar{J}} \alpha_{I(J)+I(\bar{J})} = f_{I(J)}.$$

Крайним случаем предельных распределений являются распределения частот отдельных признаков (см. п. 2.1), которые получаются, если множества J включают лишь один элемент (признак) из $j = 1, \ldots, n$. Для таких распределений $I(J) \to i_j(j)$.

В частном, но достаточно важном случае при $n = 2$ частоты распределения обычно представляют в **таблице сопряженности**, или **корреляционной таблице**:

	1	$\cdots$	i_2	$\cdots$	k_2	Y
1	N_{11}	$\cdots$	N_{1i_2}	$\cdots$	N_{1k_2}	$N_{1(1)}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
i_1	$N_{i_1 1}$	$\cdots$	$N_{i_1 i_2}$	$\cdots$	$N_{i_1 k_2}$	$N_{i_1(1)}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
k_1	$N_{k_1 1}$	$\cdots$	$N_{k_1 i_2}$	$\cdots$	$N_{k_1 k_2}$	$N_{k_1(1)}$
Y	$N_{1(2)}$	$\cdots$	$N_{i_2(2)}$	$\cdots$	$N_{k_2(2)}$	N

В этом случае существует только два маргинальных распределения частот — отдельно для 1-го признака (итоговый столбец таблицы сопряженности) и для 2-го признака (итоговая строка). Для частот и других параметров этих распределений удобнее и нагляднее 1-й способ обозначения: вместо $N_{i_1(1)}$ и $N_{1_2(2)}$ используется, соответственно, N_{i_1*} и N_{*i_2}. Этот способ обозначений удобен, если n мало, но описать общий случай, как это сделано выше, с его помощью весьма затруднительно. Формулы (4.3) в случае двух признаков принимают вид (после запятой эти же формулы даются в обозначениях 1-го способа):

$$F_{i_1(1)} = F_{i_1k_2}, \quad F_{i_1*} = F_{i_1k_2};$$
$$F_{i_2(2)} = F_{k_1i_2}, \quad F_{*i_2} = F_{k_1i_2}.$$

Аналогично, для формул (4.5):

$$f_{i_1(1)} = \sum_{i_2=1}^{k_2} f_{i_1i_2}\Delta_{i_2(2)}, \quad f_{i_1*} = \sum_{i_2=1}^{k_2} f_{i_1i_2}\Delta_{*i_2};$$
$$f_{i_2(2)} = \sum_{i_1=1}^{k_1} f_{i_1i_2}\Delta_{i_1(1)}, \quad f_{*i_2} = \sum_{i_1=1}^{k_1} f_{i_1i_2}\Delta_{i_1*}.$$

Если в таблице сопряженности разместить не частоты, а плотности относительных частот, и на каждой клетке таблицы построить параллелепипед высотой, равной соответствующему значению плотности, то получится трехмерный аналог гистограммы, который иногда называют **стереограммой**. Ее верхнюю поверхность называют **поверхностью двухмерного распределения**.

Если предположить, что $N,\ k_1,\ k_2 \to \infty$, допуская при этом, что $z_{01},\ z_{02} \to -\infty$, а $z_{k_11},\ z_{k_22} \to \infty$, то f и F станут гладкими функциями $f(x_1,\ x_2)$ и $F(x_1,\ x_2)$, соответственно, распределения плотности вероятности и распределения вероятности. Это — теоретические функции распределения. Формулы (4.1−4.3, 4.5) записываются для них следующим образом:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f\,(x_1, x_2)\,dx_1dx_2 = 1,$$
$$F\,(x_1, x_2) = \int_{-\infty}^{x_1}\int_{-\infty}^{x_2} f\,(x_1', x_2')\,dx_1'dx_2',$$
$$F\,(x_1) = F\,(x_1, \infty), \quad F\,(x_2) = F\,(\infty, x_2),$$
$$f\,(x_1) = \int_{-\infty}^{\infty} f\,(x_1, x_2)\,dx_2, f\,(x_2) = \int_{-\infty}^{\infty} f\,(x_1, x_2)\,dx_1.$$

Легко представить возможные обобщения таблицы сопряженности на случай $n > 2$. Ее аналогом является n-мерный прямоугольный параллелепипед, в итоговых гранях которого (в таблице сопряженности таких граней две — итоговые столбец и строка) даны все возможные маргинальные распределения частот. Итоговые грани — крайние, предельные, маргинальные части параллелепипеда. Это дает еще одно объяснение используемому термину — «маргинальные распределения».

Исходное распределение и любое маргинальное распределение частот строятся по всей совокупности. Однако важное значение имеют и распределения, построенные по отдельным частям выборки. Так, наряду с рассмотренным распределением частот признаков J по группам класса J, можно говорить о распределении частот признаков $\bar{J}$ (всех оставшихся признаков) по конечным группам в каждой отдельной группе класса J. Это — **условные распределения** частот. Они показывают распределения частот признаков $\bar{J}$ при условии, что все остальные признаки J зафиксированы на определенных уровнях $I(J)$. В таблице сопряженности таковыми являются распределения 1-го признака в каждом отдельном столбце, если $J = 2$, и распределения 2-го признака в каждой отдельной строке, если $J = 1$.

$\alpha_{I(\bar{J})\,|\,I(J)} = \dfrac{N_{I(J)+I(\bar{J})}}{N_{I(J)}}$ — относительные частоты условного распределения признаков $\bar{J}$ по $I(J)$. Если числитель и знаменатель правой части этой формулы поделить на N, то получится

$$\alpha_{I(\bar{J})\,|\,I(J)} = \frac{\alpha_{I(J)+I(\bar{J})}}{\alpha_{I(J)}} \quad \text{или}$$

$$\alpha_{I(\bar{J})\,|\,I(J)}\alpha_{I(J)} = \alpha_{I(J)+I(\bar{J})}. \tag{4.6}$$

$f_{I(\bar{J})\,|\,I(J)} = \dfrac{\alpha_{I(\bar{J})\,|\,I(J)}}{\Delta_{I(\bar{J})}}$ — плотности относительных частот условного распределения. Если левую часть равенства (4.6) разделить на $\Delta_{I(\bar{J})}\Delta_{I(J)}$, а правую — на Δ_I (оба этих делителя, как отмечено выше, равны), то получится аналогичное (4.6) равенство для плотностей:

$$f_{I(\bar{J})\,|\,I(J)}f_{I(J)} = f_{I(J)+I(\bar{J})}. \tag{4.7}$$

В случае двух признаков и при использовании 1-го способа индексации:

$$f_{i_1*\,|\,*i_2} = \frac{N_{i_1i_2}}{N_{*i_2}}\frac{1}{\Delta_{i_1*}}, \quad f_{*i_2\,|\,i_1*} = \frac{N_{i_1i_2}}{N_{i_1*}}\frac{1}{\Delta_{*i_2}},$$

Δ_{i_1*} и Δ_{*i_2} — результат использования первого способа индексации для $\Delta_{i_1(1)}$ и $\Delta_{i_2(2)}$;

$$f_{i_1*\,|\,*i_2}f_{*i_2} = f_{i_1i_2}, \quad f_{*i_2\,|\,i_1*}f_{i_1*} = f_{i_1i_2}.$$

В результате объединения двух последних равенств и перехода к непрерывному случаю получаются известные формулы математической статистики об условных распределениях:

$$f(x_1 \mid x_2)\, f(x_2) = f(x_1, x_2) = f(x_2 \mid x_1)\, f(x_1),$$

из которых, в частности, следует тождество **теоремы Байеса**:

$$f(x_1 \mid x_2)\, f(x_2) = f(x_2 \mid x_1)\, f(x_1).$$

Далее, по определению,

$$F_{I(\bar{J}) \mid I(J)} = \sum_{I'(\bar{J}) \leqslant I(\bar{J})} \alpha_{I'(\bar{J}) \mid I(J)}$$

— накопленные относительные частоты условного распределения. Правую часть этого равенства можно преобразовать:

$$F_{I(\bar{J}) \mid I(J)} = \sum_{I'(\bar{J}) \leqslant I(\bar{J})} \frac{N_{I(J)+I'(\bar{J})}}{N_{I(J)}} = \frac{N}{N_{I(J)}} \sum_{I'(\bar{J}) \leqslant I(\bar{J})} \frac{N_{I(J)+I'(\bar{J})}}{N} = \frac{F_{I(J)+I(\bar{J})}}{F_{I(J)}},$$

т.е. для накопленных относительных частот получается соотношение такое же, как и для плотностей относительных частот f:

$$F_{I(\bar{J}) \mid I(J)} F_{I(J)} = F_{I(J)+I(\bar{J})}. \tag{4.8}$$

В непрерывном случае для двух признаков:

$$F(x_1 \mid x_2)F(x_2) = F(x_1, x_2) = F(x_2 \mid x_1)F(x_1),$$
$$F(x_1 \mid x_2)F(x_2) = F(x_2 \mid x_1)F(x_1).$$

Количество параметров относительной частоты (также как и плотности относительной частоты и накопленной относительной частоты) $\alpha_{I(\bar{J}) \mid I(J)}$ условного распределения признаков $\bar{J}$ по $I(J)$ равно $K^{\bar{J}} = (\prod_{\bar{J}} k_j)$ — числу всех возможных сочетаний уровней признаков $\bar{J}$. Таких условных распределений признаков $\bar{J}$ имеется K^J — для каждого возможного сочетания уровней факторов J. Так, при $n = 2$ в таблице сопряженности структура каждого столбца (результат деления элементов столбца на итоговый — сумму элементов) показывает относительные частоты условного распределения 1-го признака по уровням 2-го признака (если $J = 2$). Количество параметров относительной частоты каждого такого условного

распределения — k_1, а число столбцов — условных распределений — k_2. Аналогично — для строк таблицы сопряженности (если $J = 1$).

Маргинальное распределение признаков $\bar{J}$ может быть получено из этой совокупности условных распределений (для плотностей относительных частот):

$$f_{I(\bar{J})} = \sum_{J} f_{I(\bar{J}) \mid I(J)} \alpha_{I(J)} \tag{4.9}$$

или

$$f_{I(\bar{J})} = \sum_{J} f_{I(\bar{J}) \mid I(J)} f_{I(J)} \Delta_{I(J)}.$$

Действительно, в соответствии с (4.5)

$$f_{I(\bar{J})} = \sum_{J} f_{I(J)+I(\bar{J})} \Delta_{I(J)},$$

а, учитывая (4.7),

$$\sum_{J} f_{I(J)+I(\bar{J})} \Delta_{I(J)} = \sum_{J} f_{I(\bar{J}) \mid I(J)} \alpha_{I(J)}.$$

Соотношение, аналогичное (4.9), выполняется и для самих относительных частот:

$$\alpha_{I(\bar{J})} = \sum_{J} \alpha_{I(\bar{J}) \mid I(J)} \alpha_{I(J)} \tag{4.10}$$

(оно получается умножением обеих частей соотношения (4.9) на $\Delta_{I(\bar{J})}$), а вслед за ним и для накопленных относительных частот:

$$F_{I(\bar{J})} = \sum_{J} F_{I(\bar{J}) \mid I(J)} \alpha_{I(J)}. \tag{4.11}$$

Такая связь условных и маргинального распределений наглядно иллюстрируется таблицей сопряженности (для относительных частот). Очевидно, что средневзвешенный, по весам итоговой строки, вектор структур столбцов этой матрицы алгебраически есть вектор структуры итогового столбца. Аналогично — для строк этой матрицы (для условных и маргинального распределений 2-го признака).

В непрерывном случае при $n = 2$ соотношение (4.9) имеет вид:

$$f(x_1) = \int_{-\infty}^{\infty} f(x_1 \mid x_2) f(x_2)\, dx_2, \quad f(x_2) = \int_{-\infty}^{\infty} f(x_2 \mid x_1) f(x_1)\, dx_1.$$

Если итоговые грани n-мерного прямоугольного параллелепипеда параметров распределения (обобщения таблицы сопряженности), как отмечалось выше, дают все возможные маргинальные распределения, то ортогональные «срезы» этого параллелепипеда (как строки и столбцы таблицы сопряженности) представляют все возможные условные распределения.

Условные распределения, сопоставляющие в определенном смысле вариации признаков двух разных групп $\bar{J}$ и J, используются в анализе связей между этими двумя группами признаков. При этом чрезвычайно важно понимать следующее. Речь в данном случае не идет об анализе причинно-следственных связей, хотя формально изучается поведение признаков $\bar{J}$ при условии, что признаки J принимают разные значения, т.е. признаки J выступают как бы «причиной», а признаки $\bar{J}$ — «следствием». Направление влияния в таком анализе не может быть определено. Это — предмет более тонких и сложных методов анализа. Более того, содержательно признаки этих групп могут быть не связаны, но, если они одновременно зависят от каких-то других общих факторов, то в таком анализе связь между ними может проявиться. Такие связи в статистике называют **ложными корреляциями** (или ложными регрессиями). Поэтому всегда желательно, чтобы формальному анализу зависимостей предшествовал содержательный, в котором были бы сформулированы теоретические гипотезы и построены теоретические модели. А результаты формального анализа использовались бы для проверки этих гипотез. То есть из двух задач статистического анализа связей, сформулированных в преамбуле к этому разделу, предпочтительней постановка второй задачи.

Если признаки двух множеств $\bar{J}$ и J не зависят друг от друга, то очевидно, что условные распределения признаков $\bar{J}$ не должны меняться при изменении уровней признаков J. Верно и обратное: если условные распределения признаков $\bar{J}$ одинаковы для всех уровней $I(J)$, то признаки двух множеств $\bar{J}$ и J не зависят друг от друга. Таким образом, необходимым и достаточным **условием независимости** признаков двух множеств $\bar{J}$ и J является неизменность совместных распределений признаков $\bar{J}$ при вариации уровней признаков J. Это условие можно сформулировать и в симметричной форме: неизменность совместных распределений признаков J при вариации уровней признаков $\bar{J}$.

Для таблицы сопряженности это условие означает, что структуры всех ее столбцов одинаковы. Одинаковы и структуры всех ее строк.

Итак, в случае независимости данных множеств признаков относительные частоты $\alpha_{I(\bar{J})\,|\,I(J)}$ не зависят от $I(J)$ и их можно обозначить через $\tilde{\alpha}_{I(\bar{J})}$. Тогда из соотношения (4.10) следует, что относительные частоты этого распределения совпадают с относительными частотами соответствующего маргинального распределения: $\tilde{\alpha}_{I(\bar{J})} = \alpha_{I(\bar{J})}$, т.к. $\sum\limits_{J} \alpha_{I(J)} = 1$, и соотношения (4.6) приобретают вид:

$$\alpha_{I(\bar{J})}\alpha_{I(J)} = \alpha_{I(J)+I(\bar{J})}. \tag{4.12}$$

В случае двух признаков при использовании первого способа индексации: $\alpha_{i_1 *}\alpha_{* i_2} = \alpha_{i_1 i_2}$.

Не сложно убедиться в том, что аналогичные соотношения в случае независимости признаков выполняются и для f и F:

$$f_{I(\bar{J})} f_{I(J)} = f_{I(J)+I(\bar{J})}, \tag{4.13}$$

$f_{i_1 *} f_{* i_2} = f_{i_1 i_2}$, а в непрерывном случае: $f(x_1)f(x_2) = f(x_1, x_2)$,

$$F_{I(\bar{J})} F_{I(J)} = F_{I(J)+I(\bar{J})}. \tag{4.14}$$

$F_{i_1 *} F_{* i_2} = F_{i_1 i_2}$, $F(x_1)F(x_2) = F(x_1, x_2)$.

Любое из соотношений (4.12), (4.13), (4.14) является необходимым и достаточным условием независимости признаков $\bar{J}$ и J. Необходимость следует из самого вывода этих соотношений. Достаточность легко показать, например, для (4.12). Так, если выполняется (4.12), то в соответствии с (4.4):

$$\alpha_{I(\bar{J}) \mid I(J)} = \frac{\alpha_{I(\bar{J})+I(J)}}{\alpha_{I(J)}} = \frac{\alpha_{I(\bar{J})}\alpha_{I(J)}}{\alpha_{I(J)}} = \alpha_{I(\bar{J})},$$

т.е. условные распределения признаков $\bar{J}$ не зависят от уровней, которые занимают признаки J, а это означает, что признаки $\bar{J}$ и J не зависят друг от друга.

Можно доказать, что из независимости признаков $\bar{J}$ и J следует взаимная независимость признаков любого подмножества $\bar{J}$ с признаками любого подмножества J.

Пусть $J = J_1 + J_2$, тогда соотношение (4.12) можно переписать в форме:

$$\alpha_{I(\bar{J})}\alpha_{I(J_1)+I(J_2)} = \alpha_{I(J_1)+I(J_2)+I(\bar{J})},$$

и, просуммировав обе части этого выражения по J_2 (т.е., в соответствии с введенной операцией $\sum_{J_2}$, — по всем уровням признаков J_2), получить следующее:

$$\alpha_{I(\bar{J})}\alpha_{I(J_1)} \overset{(4.4)}{=} \sum_{J_2} \alpha_{I(\bar{J})}\alpha_{I(J_1)+I(J_2)} \overset{(4.12)}{=} \sum_{J_2} \alpha_{I(J_1)+I(J_2)+I(\bar{J})} \overset{(4.4)}{=} \alpha_{I(J_1)+I(\bar{J})},$$

$$\text{т.е.} \quad \alpha_{I(\bar{J})}\alpha_{I(J_1)} = \alpha_{I(J_1)+I(\bar{J})}, \tag{4.15}$$

что означает независимость признаков $\bar{J}$ и J_1 в рамках маргинального распределения признаков $\bar{J} + J_1$.

Пусть теперь $\bar{J} = \bar{J}_1 + \bar{J}_2$. После проведения аналогичных операций с (4.15) (в частности операции суммирования по $\bar{J}_2$) получается соотношение

$\alpha_{I(\bar{J}_1)}\alpha_{I(J_1)} = \alpha_{I(J_1)+I(\bar{J}_1)}$, что означает независимость признаков $\bar{J}_1$ и J_1 в рамках маргинального распределения $\bar{J}_1 + J_1$. Что и требовалось доказать, т.к. $\bar{J}_1$ и J_1 — любые подмножества $\bar{J}$ и J.

Пока речь шла о независимости двух множеств признаков. Точно так же можно говорить и о независимости трех множеств.

Пусть $G = \bar{J} + J_1 + J_2$, где $J = J_1 + J_2$. Необходимым и достаточным условием взаимной независимости этих трех множеств признаков является следующее равенство:

$$\alpha_{I(\bar{J})}\alpha_{I(J_1)}\alpha_{I(J_2)} = \alpha_{I(J_1)+I(J_2)+I(\bar{J})}. \tag{4.16}$$

Это соотношение получается, если в левой части (4.12) вместо $\alpha_{I(J)}$ записать $\alpha_{I(J_1)}\alpha_{I(J_2)}$, т.к. $\alpha_{I(J_1)}\alpha_{I(J_2)} = \alpha_{I(J_1)+I(J_2)} \equiv \alpha_{I(J)}$ — известное условие независимости двух множеств признаков в рамках маргинального распределения признаков J.

Необходимым и достаточным условием взаимной независимости всех признаков, входящих в множество J служит следующее соотношение:

$$\alpha_I = \prod_J \alpha_{i_j(j)}. \tag{4.17}$$

Это соотношение — результат завершения процесса дробления множеств признаков, который начат переходом от (4.12) к (4.16).

Соотношения (4.12–4.14, 4.16–4.17) являются теоретическими. Оцененные по выборочной совокупности параметры совместных распределений, даже если соответствующие множества признаков независимы друг от друга, не могут обеспечить точное выполнение этих соотношений, поскольку они (параметры эмпирических распределений) являются случайными величинами. Критерий независимости строится как определенный показатель (статистика), характеризующий степень нарушения равенств в указанных соотношениях. Использование этого критерия осуществляется как проверка статистической гипотезы (нулевая гипотеза: признаки данных групп не зависимы), логика которой описана в конце пункта 2.4. Данный критерий входит в группу **критериев согласия** и называется **критерием Пирсона**, или χ^2 (**критерием хи-квадрат**).

Показатели (статистики) этого критерия — $\chi_l^{2^c}$ («^{c}» — calculated, «$_l$» — количество множеств признаков),— называемые иногда **выборочными среднеквадратическими сопряженностями признаков**, рассчитываются на основе (4.12), (4.16), (4.17) следующим образом:

$$\chi_2^{2^c} = N \sum_{J,\bar{J}} \frac{\left(\alpha_{I(J)+I(\bar{J})} - \alpha_{I(\bar{J})}\alpha_{I(J)}\right)^2}{\alpha_{I(\bar{J})}\alpha_{I(J)}},$$

$$\chi_3^{2^c} = N \sum_{J_1,J_2,\bar{J}} \frac{\left(\alpha_{I(J_1)+I(J_2)+I(\bar{J})} - \alpha_{I(\bar{J})}\alpha_{I(J_1)}\alpha_{I(J_2)}\right)^2}{\alpha_{I(\bar{J})}\alpha_{I(J_1)}\alpha_{I(J_2)}},$$

$$\chi_n^{2^c} = N \sum_{G} \frac{\left(\alpha_I - \prod\limits_J \alpha_{i_j(j)}\right)^2}{\prod\limits_J \alpha_{i_j(j)}}.$$

Если признаки не зависимы, то соответствующая статистика критерия имеет известное распределение, называемое **χ^2-распределением** (см. Приложение А.3.2). Данное распределение имеет один параметр — **число степеней свободы df** (*degrees free*), показывающее количество независимых случайных величин, квадраты которых входят в сумму. Так, в статистику $\chi_2^{2^c}$ входят квадраты $K(K^{\bar{J}}K^J)$ величин $\alpha_{I(J)+I(\bar{J})} - \alpha_{I(\bar{J})}\alpha_{I(J)}$, но не все они независимы, т.к. удовлетворяют целому ряду линейных соотношений.

Действительно, например:

$$\sum_{\bar{J}} (\alpha_{I(J)+I(\bar{J})} - \alpha_{I(\bar{J})}\alpha_{I(J)}) = 0_{K^J},$$

где 0_{K^J} — матричный нуль, имеющий размерность K^J. То есть K^J величин $\alpha_{I(J)+I_K(\bar{J})} - \alpha_{I_K(\bar{J})}\alpha_{I(J)}$ линейно выражаются через другие величины. Пусть множество этих величин обозначается $\chi_{I(J)}$.

Аналогично, исходные величины $\alpha_{I(J)+I(\bar{J})} - \alpha_{I(\bar{J})}\alpha_{I(J)}$ можно суммировать по J и установить, что $K^{\bar{J}}$ величин $\alpha_{I_K(J)+I(\bar{J})} - \alpha_{I(\bar{J})}\alpha_{I_K(J)}$ линейно выражаются через остальные; их множество можно обозначить $\chi_{I(\bar{J})}$.

Эти два множества $\chi_{I(J)}$ и $\chi_{I(\bar{J})}$ имеют один общий элемент: $\alpha_{I_K(J)+I_K(\bar{J})} - -\alpha_{I_K(\bar{J})}\alpha_{I_K(J)}$. Таким образом, количество степеней свободы $\mathbf{df}_2$ (при $l=2$) равно $K - K^{\bar{J}} - K^J + 1 = (K^{\bar{J}} - 1)(K^J - 1)$. Аналогично рассуждая, можно установить, что $\mathbf{df}_3 = (K^{\bar{J}} - 1)(K^{J_1} - 1)(K^{J_2} - 1)$, $\mathbf{df}_L = \prod\limits_J (k_j - 1)$.

Итак, чтобы ответить на вопрос, являются ли независимыми изучаемые множества признаков, необходимо расчетное значение статистики $\chi_l^{2^c}$ сравнить со значением 95-процентного квантиля $\chi^2_{\mathbf{df}_l}$-распределения (в п. 2.4 отмечалось, что в статистике вполне приемлемым считается 95-процентный уровень доверия), который обозначается $\chi^2_{\mathbf{df}_l,\,0.95}$ (это — односторонний квантиль, так как плотность χ^2-распределения расположена в положительной области значений случайной величины и не симметрична). Значения этих квантилей находят в соответствующих статистических таблицах и называют теоретическими, или табличными. Если расчетное значение не превышает табличное (т.е. является достаточно малым), то нулевая гипотеза не отвергается и данные множества признаков считаются незави-

симыми. Если расчетное значение больше табличного, то множества признаков определяются как зависимые между собой с уровнем ошибки 5%.

Современные пакеты прикладных статистических программ избавляют от необходимости пользоваться статистическими таблицами, т.к. расчет статистики критерия сопровождается оценкой уровня его значимости sl (*significance level*). Для некоторых критериев этот показатель называется значением вероятности pv (*probability value*). Уровень значимости sl — это такое число, что

$$\chi_l^{2^c} = \chi^2_{\mathbf{df}_l, 1-sl}.$$

То есть нулевая гипотеза отвергается с вероятностью ошибки 0.05, если $sl < 0.05$.

В случае 2-х признаков среднеквадратичная сопряженность имеет следующий вид (здесь и ниже используется 1-й способ обозначений):

$$\chi_2^{2^c} = N \sum_{i_1, i_2} \frac{(\alpha_{i_1 i_2} - \alpha_{i_1 *}\alpha_{* i_2})^2}{\alpha_{i_1 *}\alpha_{* i_2}},$$

а соответствующее ей χ^2-распределение имеет $(k_1 - 1)(k_2 - 1)$ степеней свободы; множество $\chi_{i_1 *}$ образовано величинами $\alpha_{i_1 k_2} - \alpha_{i_1 *}\alpha_{* k_2}$, $i_1 = 1, \ldots, k_1$, множество $\chi_{* i_2}$ — величинами $\alpha_{k_1 i_2} - \alpha_{k_1 *}\alpha_{* i_2}$, $i_2 = 1, \ldots, k_2$, общим для них является элемент $\alpha_{k_1 k_2} - \alpha_{k_1 *}\alpha_{* k_2}$.

Далее в этой главе рассматривается в основном случай двух признаков.

4.2. Регрессионный анализ

В качестве значений признаков $x_{i_1 *}$ и $x_{* i_2}$ на полуинтервалах, как и прежде, принимаются середины этих полуинтервалов. Средние и дисперсии признаков рассчитываются по известным формулам:

$$\bar{x}_1 = \sum x_{i_1 *}\alpha_{i_1 *}, \quad \bar{x}_2 = \sum x_{* i_2}\alpha_{* i_2};$$
$$s_1^2 = \sum (x_{i_1 *} - \bar{x}_1)^2 \alpha_{i_1 *}, \quad s_2^2 = \sum (x_{* i_2} - \bar{x}_2)^2 \alpha_{* i_2} \quad \text{или, более компактно,}$$
$$s_1^2 = \sum \hat{x}^2_{i_1 *}\alpha_{i_1 *}, \quad s_2^2 = \sum \hat{x}^2_{* i_2}\alpha_{* i_2}.$$

Важной характеристикой совместного распределения двух признаков является **ковариация** — совместный центральный момент 2-го порядка:

$$m_{12} = \sum \hat{x}_{i_1 *}\hat{x}_{* i_2}\alpha_{i_1 i_2}.$$

Дисперсия — частный случай ковариации (ковариация признака с самим собой), поэтому для обозначения дисперсии j-го признака часто используется m_{jj}.

В случае независимости признаков, когда $\alpha_{i_1 i_2} = \alpha_{i_1 *}\alpha_{* i_2}$, как несложно убедиться, ковариация равна нулю. Равенство ковариации нулю[2] является необходимым, но не достаточным условием независимости признаков, т.к. ковариация — характеристика только **линейной** связи. Если ковариация равна нулю, признаки линейно независимы, но какая-то другая форма зависимости между ними может существовать.

Мерой линейной зависимости является относительная ковариация, называемая **коэффициентом корреляции**:

$$r_{12} = \frac{m_{12}}{\sqrt{m_{11} m_{22}}}.$$

Этот коэффициент по абсолютной величине не превышает единицу (этот факт доказывается ниже). Если его значение близко к нулю, то признаки линейно независимы, если близко к плюс единице — между признаками существует прямая линейная зависимость, если близко к минус единице — существует обратная линейная зависимость. В частности, легко убедиться в том, что если $\hat{x}_{i_1 *} = \pm a_{12} \hat{x}_{* i_2}$ (т.е. между признаками имеет место линейная зависимость), то $r_{12} = \pm 1$.

Значения ковариаций и коэффициентов корреляции симметричны: $m_{12} = m_{21}$, $r_{12} = r_{21}$.

В дальнейшем рассуждения проводятся так, как будто 1-й признак зависит от 2-го (хотя с тем же успехом можно было бы говорить о зависимости 2-го признака от 1-го). В таком случае переменная x_1 (значения 1-го признака) называется **объясняемой**, **моделируемой**, **эндогенной**, а переменная x_2 (значения 2-го признака) — **объясняющей**, **факторной**, **экзогенной**.

Наряду с общей средней 1-го признака $\bar{x}_1$ полезно рассчитать **условные средние** $\bar{x}_{1 \mid *i_2}$[3] — средние 1-го признака при условии, что 2-й признак зафиксирован на определенном уровне i_2. При расчете таких средних усреднение значений признака на полуинтервалах проводится по относительным частотам не маргинального ($\alpha_{i_1 *}$), а соответствующих условных распределений ($\alpha_{i_1 * \mid * i_2}$):

$$\bar{x}_{1 \mid * i_2} = \sum x_{i_1 *} \alpha_{i_1 * \mid * i_2}.$$

Усреднение этих величин по весам маргинального распределения 2-го признака дает общее среднее:

$$\bar{x}_1 = \sum_{i_1} x_{i_1 *} \alpha_{i_1 *} = \sum_{i_2} \sum_{i_1} x_{i_1 *} \alpha_{i_1 i_2} = \sum_{i_2} \sum_{i_1} x_{i_1 *} \alpha_{i_1 * \mid * i_2} \alpha_{* i_2} = \sum_{i_2} \bar{x}_{1 \mid * i_2} \alpha_{* i_2}.$$

[2] Равенство или неравенство нулю понимается в статистическом смысле: не отвергается или отвергается соответствующая нулевая гипотеза.

[3] В общем случае вектор условных средних признаков $\bar{J}$ обозначается $\bar{x}_{\bar{J}/I(J)}$.

В непрерывном случае эти формулы принимают вид:

$$\mathbf{E}\left(x_1|x_2\right) = \int\limits_{-\infty}^{\infty} x_1 f\left(x_1|x_2\right) dx_1, \quad \mathbf{E}\left(x_1\right) = \int\limits_{-\infty}^{\infty} \mathbf{E}\left(x_1|x_2\right) f\left(x_2\right) dx_2.$$

(Об условных и маргинальных распределениях см. Приложение А.3.1.)

Условные дисперсии признака рассчитываются следующим образом:

$$s^2_{1\,|\,*i_2} = \sum \left(x_{i_1*} - \bar{x}_{1\,|*\,i_2}\right)^2 \alpha_{i_1*\,|\,*i_2}.$$

Отклонения фактических значений признака от условных средних

$$e_{i_1*\,|\,*i_2} = x_{i_1*} - \bar{x}_{1\,|\,*i_2}$$

обладают, по определению, следующими свойствами:

а) их средние равны нулю:

$$\sum e_{i_1*\,|\,*i_2}\alpha_{i_1*\,|\,*i_2} = 0,$$

б) их дисперсии, совпадающие с условными дисперсиями признака, минимальны (суммы их квадратов минимальны среди сумм квадратов отклонений от каких-либо фиксированных значений признака — наличие этого свойства у дисперсий доказывалось в п. 2.4):

$$s^2_{e1\,|\,*i_2} = \sum e^2_{i_1*\,|\,*i_2}\alpha_{i_1*\,|\,*i_2} = s^2_{1\,|\,*i_2} = \min_c \sum \left(x_{i_1*} - c\right)^2 \alpha_{i_1*\,|\,*i_2}.$$

Общая дисперсия связана с условными дисперсиями более сложно:

$$\begin{aligned}
s^2_1 &= \sum \hat{x}^2_{i_1*}\alpha_{i_1*} = \sum_{i_1}\sum_{i_2} \hat{x}^2_{i_1*}\alpha_{i_1i_2} = \\
&= \sum_{i_1}\sum_{i_2} \left(\left(x_{i_1*} - \bar{x}_{1\,|\,*i_2}\right) + \left(\bar{x}_{1\,|\,*i_2} - \bar{x}_1\right)\right)^2 \alpha_{i_1i_2} = \\
&= \sum_{i_1}\sum_{i_2} \left(x_{i_1*} - \bar{x}_{1\,|\,*i_2}\right)^2 \alpha_{i_1i_2} + 2\sum_{i_1}\sum_{i_2} \left(x_{i_1*} - \bar{x}_{1\,|\,*i_2}\right)\left(\bar{x}_{1\,|\,*i_2} - \bar{x}_1\right)\alpha_{i_1i_2} + \\
&\qquad + \sum_{i_1}\sum_{i_2} \left(\bar{x}_{1\,|\,*i_2} - \bar{x}_1\right)^2 \alpha_{i_1i_2} =
\end{aligned}$$

$$= \sum_{i_2} \alpha_{*i_2} \underbrace{\sum_{i_1} \left(x_{i_1*} - \bar{x}_{1\,|\,*i_2}\right)^2 \overbrace{\frac{\alpha_{i_1 i_2}}{\alpha_{*i_2}}}^{\alpha_{i_1*|*i_2}}}_{s^2_{e1\,|*\,i_2}} +$$

$$+ \underbrace{2 \sum_{i_2} \alpha_{*i_2} \left(\bar{x}_{1\,|\,*i_2} - \bar{x}_1\right) \overbrace{\sum_{i_1} \left(x_{i_1*} - \bar{x}_{1\,|\,*i_2}\right) \alpha_{i_1*\,|\,*i_2}}^{=0}}_{=0} +$$

$$+ \sum_{i_2} \left(\bar{x}_{1\,|\,*i_2} - \bar{x}_1\right)^2 \underbrace{\sum_{i_1} \alpha_{i_1 i_2}}_{\alpha_{*i_2}} = s^2_{e1} + s^2_{q1}.$$

Равенство нулю среднего слагаемого в этой сумме означает, что отклонения фактических значений 1-го признака от условных средних не коррелированы (линейно не связаны) с самими условными средними.

В терминах регрессионного анализа

s^2_{q1} — **объясненная дисперсия**, т.е. та дисперсия 1-го признака, которая объясняется вариацией 2-го признака (в частности, когда признаки независимы и условные распределения 1-го признака одинаковы при всех уровнях 2-го признака, то условные средние не варьируют и объясненная дисперсия равна нулю);

s^2_{e1} — **остаточная дисперсия**.

Чем выше объясненная дисперсия по сравнению с остаточной, тем вероятнее, что 2-й признак влияет на 1-й. Количественную меру того, насколько объясненная дисперсия должна быть больше остаточной, чтобы это влияние можно было признать существенным (значимым), дает **критерий Фишера**, или **F-критерий**. Статистика этого критерия F^c рассчитывается следующим образом:

$$F^c = \frac{s^2_{q1} k_2 (k_1 - 1)}{s^2_{e1} (k_2 - 1)}.$$

В случае если влияние 2-го признака на 1-й не существенно, эта величина имеет **F-распределение** (см. Приложение А.3.2). Такое распределение имеет случайная величина, полученная отношением двух случайных величин, имеющих χ^2-распределение, деленных на количество своих степеней свободы:

$$F_{\mathbf{df}_1,\,\mathbf{df}_2} = \frac{\chi^2_{\mathbf{df}_1} \mathbf{df}_2}{\chi^2_{\mathbf{df}_2} \mathbf{df}_1}.$$

Количество степеней свободы в числителе ($\mathbf{df}_1$) и знаменателе ($\mathbf{df}_2$) относится к параметрам F-распределения.

Рассуждая аналогично тому, как это сделано в конце предыдущего пункта, можно установить, что объясненная дисперсия (в числителе F-статистики) имеет k_2-1 степеней свободы, а остаточная дисперсия (в знаменателе) — $k_2(k_1-1)$ степеней свободы. Это объясняет указанный способ расчета данной статистики.

Чтобы проверить гипотезу о наличии влияния 2-го признака на 1-й, необходимо сравнить расчетное значение статистики F^c с теоретическим — взятым из соответствующей статистической таблицы 95-процентным квантилем (односторонним) F-распределения с k_2-1 и $k_2(k_1-1)$ степенями свободы $F_{k_2-1,k_2(k_1-1),\,0.95}$. Если расчетное значение не превышает теоретическое, то нулевая гипотеза не отвергается, и влияние считается не существенным. В противном случае (объясненная дисперсия достаточно велика по сравнению с остаточной) нулевая гипотеза отвергается и данное влияние принимается значимым. Современные статистические пакеты прикладных программ дают уровень значимости расчетной статистики, называемый в данном случае значением вероятности pv:

$$F^c = F_{k_2-1,\,k_2(k_1-1),\,1-pv}.$$

Если $pv < 0.05$, то нулевая гипотеза отвергается с вероятностью ошибки 5%.

Линия, соединяющая точки $\left(x_{*i_2}, \bar{x}_{1\,|\,*i_2}\right)$ в пространстве значений признаков (абсцисса — 2-й признак, ордината — 1-й) называется **линией регрессии**, она показывает зависимость 1-го признака от 2-го. Условные средние, образующие эту линию, являются расчетными (модельными) или объясненными этой зависимостью значениями 1-го признака. Объясненная дисперсия показывает вариацию значений 1-го признака, которые расположены на этой линии, остаточная дисперсия — вариацию фактических значений признака вокруг этой линии.

Линию регрессии можно провести непосредственно в таблице сопряженности. Это линия, которая соединяет клетки с максимальными в столбцах плотностями относительных частот. Понятно, что о такой линии имеет смысл говорить, если имеются явные концентрации плотностей относительных частот в отдельных клетках таблицы сопряженности. Критерием наличия таких концентраций как раз и является F-критерий.

В непрерывном случае уравнение

$$x_1 = \mathbf{E}\,(x_1|x_2)$$

называют **уравнением регрессии** x_1 по x_2, т.е. уравнением статистической зависимости 1-го признака от 2-го (о свойствах условного математического ожидания см. Приложение А.3.1). Это уравнение выражает статистическую зависимость, поскольку показывает наиболее вероятное значение, которое принимает 1-й признак при том или ином уровне 2-го признака. В случае если 2-й признак является единственным существенно влияющим на 1-й признак, т.е. это уравнение выражает

теоретическую, истинную зависимость, эти наиболее вероятные значения называют **теоретическими**, а отклонения от них фактических значений — **случайными ошибками** измерения. Для фактических значений x_1 это уравнение записывают со стохастическим членом, т.е. со случайной ошибкой, остатками, отклонением фактических значений от теоретических:

$$x_1 = \mathbf{E}\left(x_1 | x_2\right) + \varepsilon_1.$$

Случайные ошибки по построению уравнения регрессии имеют нулевое математическое ожидание и минимальную дисперсию при любом значении x_2, они взаимно независимы со значениями x_2. Эти факты обсуждались выше для эмпирического распределения.

В рассмотренной схеме регрессионного анализа уравнение регрессии можно построить лишь теоретически. На практике получают линию регрессии, по виду которой можно лишь делать предположения о форме и, тем более, о параметрах зависимости.

В эконометрии обычно используется другая схема регрессионного анализа. В этой схеме используют исходные значения признаков $x_{i1},\ x_{i2},\ i = 1, \ldots, N$ без предварительной группировки и построения таблицы сопряженности, выдвигают гипотезу о форме зависимости f: $x_1 = f\left(x_2, A\right)$, где A — параметры зависимости, и находят эти параметры так, чтобы была минимальной остаточная дисперсия $s_{e1}^2 = \dfrac{1}{N}\sum\limits_i \left(x_{i1} - f\left(x_{i2}, A\right)\right)^2$.

Такой метод называется **методом наименьших квадратов** (МНК).

Ковариация и коэффициент корреляции непосредственно по данным выборки рассчитываются следующим образом:

$$m_{jj'} = \frac{1}{N}\sum \left(x_{ij} - \bar{x}_j\right)\left(x_{ij'} - \bar{x}_{j'}\right), \quad r_{jj'} = \frac{m_{jj'}}{\sqrt{m_{jj} m_{j'j'}}}, \quad j, j' = 1, 2.$$

Далее в этом пункте рассматривается случай **линейной регрессии**, т.е. случай, когда

$$x_1 = \alpha_{12} x_2 + \beta_1 + \varepsilon_1, \tag{4.18}$$

где $\alpha_{12}, \beta_1, \varepsilon_1$ — истинные значения параметров регрессии и остатков.

Следует иметь в виду, что регрессия линейна, если форма зависимости признаков линейна относительно оцениваемых параметров, а не самих признаков,

и уравнения

$$x_1 = \alpha_{12}\sqrt{x_2} + \beta_1 + \varepsilon_1,$$
$$x_1 = \alpha_{12}\frac{1}{x_2} + \beta_1 + \varepsilon_1,$$
$$\ln x_1 = \alpha_{12}\ln x_2 + \ln\beta_1 + \ln\varepsilon_1 \quad \left(x_1 = x_2^{\alpha_{12}}\beta_1\varepsilon_1\right),$$

и т.д. также относятся к линейной регрессии. Во всех этих случаях метод наименьших квадратов применяется одинаковым образом. Поэтому можно считать, что в записи (4.18) x_1 и x_2 являются результатом какого-либо функционального преобразования исходных значений.

Оценки параметров регрессии и остатков обозначаются соответствующими буквами латинского алфавита, и уравнение регрессии, записанное по наблюдениям i, имеет следующий вид:

$$x_{i1} = a_{12}x_{i2} + b_1 + e_{i1}, \quad i = 1, \ldots, N, \tag{4.19}$$

а в матричной форме:

$$X_1 = X_2 a_{12} + 1_N b_1 + e_1, \tag{4.20}$$

где X_1, X_2 — вектор-столбцы наблюдений размерности N, соответственно, за 1-м и 2-м признаками, e_1 — вектор-столбец остатков; 1_N — вектор-столбец размерности N, состоящий из единиц.

Прежде чем переходить к оценке параметров регрессии (применению метода наименьших квадратов), имеет смысл объяснить происхождение термина «регрессия». Этот термин введен английским статистиком Ф. Гальтоном в последней четверти XIX века при изучении зависимости роста сыновей от роста отцов. Оказалось, что если по оси абсцисс расположить рост отцов (x_2), а по оси ординат — рост сыновей (x_1), то точки, соответствующие проведенным наблюдениям (облако точек наблюдений), расположатся вокруг некоторой прямой (рис. 4.1).

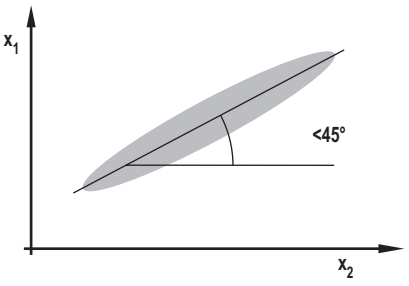

Рис. 4.1

Это означает, что зависимость между ростом сыновей и отцов существует, и эта зависимость близка к линейной. Но угол наклона соответствующей прямой меньше 45°. Другими словами, имеет место «возврат» — регрессия — роста сыновей к некоторому среднему росту. Для этой зависимости и был предложен термин «регрессия». Со временем он закрепился за любыми зависимостями статистического характера, т.е. такими, которые выполняются «по математическому ожиданию», с погрешностью.

Остаточная дисперсия из (4.19) получает следующее выражение:

$$s_{e1}^2 = \frac{1}{N}\sum_i (x_{i1} - a_{12}x_{i2} - b_1)^2,$$

или в матричной форме:

$$s_{e1}^2 = \frac{1}{N} e_1' e_1,$$

где

$$e_1 = X_1 - X_2 a_{12} - 1_N b_1, \text{ — остатки регрессии,}$$

штрих — знак транспонирования. Величина $e_1' e_1$ называется **суммой квадратов остатков**.

Для минимизации этой дисперсии ее производные по искомым параметрам (сначала по b_1, потом по a_{12}) приравниваются к нулю.

$$\frac{\partial s_{e1}^2}{\partial b_1} = -\frac{2}{N}\sum (x_{i1} - a_{12}x_{i2} - b_1) = 0, \quad \text{откуда:}$$

$$\sum e_{i1} = 0,$$

$$b_1 = \bar{x}_1 - a_{12}\bar{x}_2. \tag{4.21}$$

Это означает, что $\bar{e}_1 = 0$, т.е. сумма остатков равна нулю, а также, что линия регрессии проходит через точку средних.

После подстановки полученной оценки свободного члена форма уравнения регрессии и остаточной дисперсии упрощается:

$$\hat{x}_{i1} = a_{12}\hat{x}_{i2} + e_{i1}, \quad i = 1, \ldots, N, \tag{4.22}$$

$$\hat{X}_1 = \hat{X}_2 a_{12} + e_1, \text{ — сокращенная запись уравнения регрессии,} \tag{4.23}$$

$$s_{e1}^2 = \frac{1}{N}\sum (\hat{x}_{i1} - a_{12}\hat{x}_{i2})^2. \tag{4.24}$$

Далее:

$$\frac{\partial s_{e1}^2}{\partial a_{12}} = -\frac{2}{N}\sum \hat{x}_{i2} \overbrace{(\hat{x}_{i1} - a_{12}\hat{x}_{i2})}^{e_{i1}} = 0. \tag{4.25}$$

Отсюда следует, во-первых, то, что вектора e_1 и X_2 ортогональны, т.к. ковариация между ними равна нулю $(\sum \hat{x}_{i2} e_{i1} = 0)$; во-вторых — выражение для оценки углового коэффициента:

$$a_{12} = \frac{m_{12}}{m_{22}}. \tag{4.26}$$

Матрица вторых производных остаточной дисперсии в найденной точке равна

$$2\begin{bmatrix} 1 & \bar{x}_2 \\ \bar{x}_2 & m_{22}^0 \end{bmatrix},$$

где m_{22}^0 — 2-й начальный (а не центральный, как m_{22}) момент для x_2. Тот же результат можно получить, если не переходить к сокращенной записи уравнения регрессии перед дифференцированием остаточной дисперсии по a_{12}.

Эта матрица положительно определена (ее определитель равен $2m_{22}$, то есть всегда неотрицателен), поэтому найденная точка является действительно точкой минимума остаточной дисперсии.

Таким образом, построен оператор МНК-оценивания (4.21, 4.26) и выявлены свойства МНК-остатков: они ортогональны факторной переменной x_2, стоящей в правой части уравнения регрессии, и их среднее по наблюдениям равно нулю.

«Теоретические» значения моделируемой переменной x_1, лежащие на линии оцененной регрессии:

$$\begin{aligned} x_{i1}^c &= a_{12}x_{i2} + b_1, \\ \hat{x}_{i1}^c &= a_{12}\hat{x}_{i2}, \end{aligned} \tag{4.27}$$

где « ^{c} » — calculated, часто называют **расчетными**, или **объясненными**. Это — математические ожидания моделируемой переменной.

Вторую часть оператора МНК-оценивания (4.26) можно получить, используя другую логику рассуждений, часто применяемую в регрессионном анализе.

Обе части уравнения регрессии, записанного в сокращенной матричной форме (4.23) умножаются слева на транспонированный вектор $\widehat{X}_2'$ и делятся на N:

$$\frac{1}{N}\widehat{X}_2'\widehat{X}_1 = \frac{1}{N}\widehat{X}_2'\widehat{X}_2 a_{12} + \frac{1}{N}\widehat{X}_2' e_1.$$

Второе слагаемое правой части полученного уравнения отбрасывается, так как в силу отмеченных свойств МНК-остатков оно равно нулю, и получается искомое выражение: $m_{12} = m_{22}a_{12}$.

Пользуясь этой логикой, оператор МНК-оценивания можно получить и в полном формате. Для этого используют запись регрессионного уравнения в форме без свободного члена (со скрытым свободным членом):

$$X_1 = \widetilde{X}_2\tilde{a}_{12} + e_1, \tag{4.28}$$

где $\widetilde{X}_2$ — матрица $[X_2, 1_N]$ размерности $N \times 2$, $\tilde{a}_{12}$ — вектор $\begin{bmatrix} a_{12} \\ b_1 \end{bmatrix}$.

Как и прежде, обе части этого уравнения умножаются слева на транспонированную матрицу $\widetilde{X}_2'$ и делятся на N, второе слагаемое правой части отбрасывается по тем же причинам. Получается выражение для оператора МНК-оценивания:

$$\widetilde{m}_{12} = \widetilde{M}_{22}\widetilde{a}_{12}, \quad \text{т.е.} \quad \widetilde{a}_{12} = \widetilde{M}_{22}^{-1}\widetilde{m}_{12}, \tag{4.29}$$

где $\widetilde{m}_{12} = \dfrac{1}{N}\widetilde{X}_2'X_1$, $\widetilde{M}_{22} = \dfrac{1}{N}\widetilde{X}_2'\widetilde{X}_2$.

Это выражение эквивалентно полученному выше. Действительно, учитывая, что $X_j = \hat{X}_j + 1_N\bar{x}_j$, $1_N'\hat{X}_j = 0$, $\quad j = 1, 2$,

$$\tilde{m}_{12} = \frac{1}{N}\begin{bmatrix} X_2'X_1 \\ 1_N'X_1 \end{bmatrix} = \begin{bmatrix} m_{12} + \bar{x}_1\bar{x}_2 \\ \bar{x}_1 \end{bmatrix},$$

$$\tilde{M}_{22} = \frac{1}{N}\begin{bmatrix} X_2'X_2 & X_2'1_N \\ 1_N'X_2 & 1_N'1_N \end{bmatrix} = \begin{bmatrix} \overbrace{m_{22} + \bar{x}_2^2}^{m_{22}^0} & \bar{x}_2 \\ \bar{x}_2 & 1 \end{bmatrix}.$$

Тогда матричное уравнение (4.29) переписывается следующим образом:

$$m_{12} + \bar{x}_1\bar{x}_2 = m_{22}a_{12} + \bar{x}_2^2a_{12} + \bar{x}_2b_1,$$
$$\bar{x}_1 = \bar{x}_2a_{12} + b_1.$$

Из 2-го уравнения сразу следует (4.21), а после подстановки b_{12} в 1-е уравнение оно преобразуется к (4.26). Что и требовалось доказать.

Таким образом, выражение (4.29) представляет собой компактную запись оператора МНК-оценивания.

Из проведенных рассуждений полезно, в частности, запомнить, что уравнение регрессии может быть представлено в трех формах: в исходной — (4.19, 4.20), сокращенной — (4.22, 4.23) и со скрытым свободным членом — (4.28). Третья форма имеет только матричное выражение.

Оцененное уравнение линейной регрессии «наследует» в определенном смысле свойства линии регрессии, введенной в начале этого пункта по данным совместного распределения двух признаков: минимальность остаточной дисперсии, равенство нулю средних остатков и ортогональность остатков к объясняющей переменной — в данном случае к значениям второго признака. (Последнее для регрессии, построенной по данным совместного распределения, звучало как линейная независимость отклонений от условных средних и самих условных средних.) Отличие в том, что теперь линия регрессии является прямой, условными средними являются расчетные значения моделируемой переменной, а условными дисперсиями — остаточная

дисперсия, которая принимается при таком методе оценивания одинаковой для всех наблюдений.

Теперь рассматривается остаточная дисперсия (4.24) в точке минимума:

$$s_{e1}^2 = \frac{1}{N}\sum\left(\hat{x}_{i1}^2 - 2\hat{x}_{i1}\hat{x}_{i2}a_{12} + \hat{x}_{i2}^2 a_{12}^2\right) \overset{(4.26)}{=} m_{11} - \frac{m_{12}^2}{m_{22}}. \tag{4.30}$$

Поскольку остаточная дисперсия неотрицательна,

$$m_{11} \geqslant \frac{m_{12}^2}{m_{22}}, \quad \text{т.е.} \quad r_{12}^2 \leqslant 1.$$

Это доказывает ранее сделанное утверждение о том, что коэффициент корреляции по абсолютной величине не превышает единицу.

Второе слагаемое (взятое с плюсом) правой части соотношения (4.30) является дисперсией расчетных значений моделируемой переменной (**var** — обозначение дисперсии):

$$\begin{aligned}\mathbf{var}\,(x_1^c) = \frac{1}{N}\sum\left(x_{i1}^c - \bar{x}_1^c\right)^2 &\overset{\bar{e}=0}{=} \frac{1}{N}\sum\left(x_{i1}^c - \bar{x}_1\right)^2 \overset{(4.27)}{=} \\ &= \frac{1}{N}\sum\left(a_{12}\hat{x}_{i2}\right)^2 = a_{12}^2 m_{22} \overset{(4.26)}{=} \frac{m_{12}^2}{m_{22}}.\end{aligned} \tag{4.31}$$

Эту дисперсию, как и в регрессии, построенной по данным совместного распределения признаков, естественно назвать объясненной и обозначить s_{q1}^2. Тогда из (4.30) следует, что общая дисперсия моделируемого признака, как и прежде, распадается на две части — объясненную и остаточную дисперсии:

$$s_1^2 = m_{11} = s_{q1}^2 + s_{e1}^2.$$

Доля объясненной дисперсии в общей называется **коэффициентом детерминации**, который обозначается R^2. Такое обозначение не случайно, поскольку этот коэффициент равен квадрату коэффициента корреляции:

$$R^2 = \frac{s_{q1}^2}{s_1^2} = \frac{m_{12}^2}{m_{11}m_{22}}.$$

Коэффициент детерминации является показателем точности аппроксимации фактических значений признаков линией регрессии: чем ближе он к единице, тем точнее аппроксимация. При прочих равных его значение будет расти с уменьшением числа наблюдений. Так, если наблюдений всего два, этот коэффициент всегда будет равен единице, т.к. через две точки можно провести единственную прямую. Поэтому

данный коэффициент выражает скорее «алгебраическое» качество построенного уравнения регрессии.

Показатель статистической значимости оцененного уравнения дает статистика Фишера — как и для регрессии, построенной по данным совместного распределения признаков. В данном случае остаточная дисперсия имеет $N-2$ степени свободы, а объясненная — одну степень свободы (доказательство этого факта дается во II части книги):

$$F^c = \frac{s_{q1}^2 (N-2)}{s_{e1}^2} = \frac{R^2 (N-2)}{(1-R^2)}.$$

Если переменные не зависят друг от друга, т.е. $\alpha_{12} = 0$ (нулевая гипотеза), то эта статистика имеет распределение Фишера с одной степенью свободы в числителе и $N-2$ степенями свободы в знаменателе. Логика использования этой статистики описана выше. Статистическая значимость (качество) полученного уравнения тем выше, чем ниже значение показателя pv для расчетного значения данной статистики F^c.

Оценки параметров α_{12}, β_1 и остатков ε_{i1} можно получить иначе, из регрессии x_2 по x_1:

$$\hat{x}_{i2} = a_{21}\hat{x}_{i1} + e_{i2}, \quad i = 1, \ldots, N.$$

В соответствии с (4.26) оценка углового коэффициента получается делением ковариации переменных, стоящих в левой и правой частях уравнения, на дисперсию факторной переменной, стоящей в правой части уравнения:

$$a_{21} = \frac{m_{21}}{m_{11}}.$$

Поскольку $\hat{x}_{i1} = \dfrac{1}{a_{21}}\hat{x}_{i2} - \dfrac{1}{a_{21}}e_{i2}$,

$$\begin{aligned} a_{12}(2) &= \frac{1}{a_{21}} = \frac{m_{11}}{m_{21}}, \\ b_1(2) &= \bar{x}_1 - a_{12}(2)\,\bar{x}_2, \\ e_{i1}(2) &= a_{12}(2)\,e_{i2}, \quad i = 1, \ldots, N. \end{aligned} \tag{4.32}$$

Это — новые оценки параметров. Легко убедиться в том, что $a_{12}(2)$ совпадает с a_{12} (а вслед за ним $b_1(2)$ совпадает с b_1 и $e_{i1}(2)$ — с e_{i1}) тогда и только тогда, когда коэффициент корреляции r_{12} равен единице, т.е. зависимость имеет функциональный характер и все остатки равны нулю.

При оценке параметров α_{12}, β_1 и остатков e_{i1} регрессия x_1 по x_2 иногда называется **прямой**, регрессия x_1 по x_2 — **обратной**.

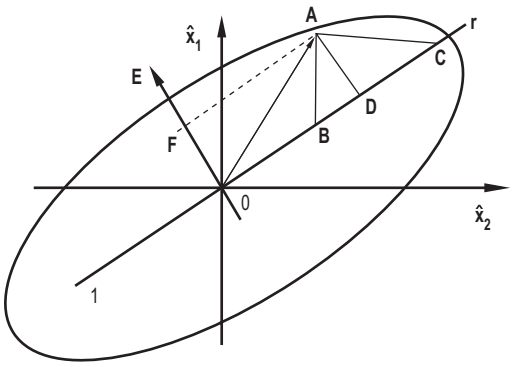

Рис. 4.2

На рисунке 4.2 в плоскости (в пространстве) переменных x_1, x_2 применение прямой регрессии означает минимизацию суммы квадратов расстояний от точек облака наблюдений до линии регрессии, измеренных параллельно оси x_1. При применении обратной регрессии эти расстояния измеряются параллельно оси x_2.

lr — линия регрессии,

OA — вектор-строка i-го наблюдения $\hat{x}_i = (\hat{x}_{i1}, \hat{x}_{i2})$,

AB — расстояние до линии регрессии, измеренное параллельно оси $\hat{x}_1$, равное величине e_{i1},

AC — расстояние, измеренное параллельно оси $\hat{x}_2$, равное величине e_{i2},

AD — расстояние, измеренное перпендикулярно линии регрессии, равное e_i,

OE — вектор-строка a' параметров ортогональной регрессии.

Очевидно, что оценить параметры регрессии можно, измеряя расстояния до линии регрессии перпендикулярно самой этой линии (на рисунке — отрезок AD). Такая регрессия называется **ортогональной**. В уравнении такой регрессии обе переменные остаются в левой части с коэффициентами, сумма квадратов которых должна равняться единице (длина вектора параметров регрессии должна равняться единице):

$$\begin{aligned} &a_1\hat{x}_{i1} + a_2\hat{x}_{i2} = e_i, \quad i = 1, \ldots, N \\ &a_1^2 + a_2^2 = 1. \end{aligned} \tag{4.33}$$

В матричной форме:

$$\begin{aligned} &\hat{X}a = e, \\ &a'a = 1, \end{aligned} \tag{4.34}$$

где $\hat{X}$ — матрица наблюдений за переменными, размерности $N * 2$, a — вектор-столбец параметров регрессии.

Само уравнение регрессии можно записать еще и так:

$$\hat{x}_i a = e_i, \quad i = 1, \ldots, N. \tag{4.35}$$

Чтобы убедиться в том, что такая регрессия является ортогональной, достаточно вспомнить из линейной алгебры, что скалярное произведение вектора на вектор

единичной длины равно длине проекции этого вектора на единичный вектор. В левой части (4.35) как раз и фигурирует такое скалярное произведение. На рисунке вектором параметров a является OE, проекцией вектора наблюдений $OA(\hat{x}_i)$ на этот вектор — отрезок OF, длина которого $(\hat{x}_i a)$ в точности равна расстоянию от точки облака наблюдений до линии регрессии, измеренному перпендикулярно этой линии (e_i).

Следует иметь в виду, что и в «обычной» регрессии, в левой части которой остается одна переменная, коэффициент при этой переменной принимается равным единице, т.е. фактически используется аналогичное ортогональной регрессии требование: вектор параметров при переменных в левой части уравнения должен иметь единичную длину.

В противоположность ортогональной «обычные» регрессии называют **простыми**. В отечественной литературе **простой** часто называют «обычную» регрессию с одной факторной переменной. А регрессию с несколькими факторными переменными называют **множественной**.

Теперь остаточную дисперсию в матричной форме можно записать следующим образом:

$$s_e^2 = \frac{1}{N} e'e = \frac{1}{N} a'\hat{X}'\hat{X}a = a'Ma,$$

где $M = \frac{1}{N}\hat{X}'\hat{X}$ — матрица ковариации переменных, равная $\begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}$.

Для минимизации остаточной дисперсии при ограничении на длину вектора параметров регрессии строится функция Лагранжа:

$$L(a, \lambda) = a'Ma - \lambda a'a,$$

где λ — множитель Лагранжа (оценка ограничения).

Далее находятся производные этой функции по параметрам регрессии, и эти производные приравниваются к нулю. Результат таких операций в матричной форме представляется следующим образом (поскольку M — симметричная матрица: $M' = M$):

$$(M - \lambda I)\, a = 0. \tag{4.36}$$

Таким образом, множитель Лагранжа есть собственное число матрицы ковариации M, а вектор оценок параметров регрессии — соответствующий правый собственный вектор этой матрицы (см. Приложение А.1.2).

Матрица M является вещественной, симметричной и положительно полуопределенной (см. Приложение А.1.2).

Последнее справедливо, т.к. квадратичная форма $\mu' M \mu$ при любом векторе μ неотрицательна. Действительно, эту квадратичную форму всегда можно представить как сумму квадратов компонент вектора $\eta = \frac{1}{\sqrt{N}} \hat{X} \mu$:

$$\mu' M \mu = \frac{1}{N} \mu' \hat{X}' \hat{X} \mu = \eta' \eta \geqslant 0.$$

Из линейной алгебры известно, что все собственные числа такой матрицы вещественны и неотрицательны, следовательно λ неотрицательно.

После умножения обеих частей уравнения (4.36) слева на a' из него следует, что

$$s_e^2 = a' M a = \lambda a' a \overset{a'a=1}{=} \lambda,$$

т.е. минимизации остаточной дисперсии соответствует поиск минимального собственного числа матрицы ковариации переменных M. Соответствующий этому собственному числу правый собственный вектор этой матрицы есть вектор оценок параметров ортогональной регрессии a (см. Приложение А.1.2). Кроме того, в соответствии со свойствами матрицы M, сумма ее собственных чисел равна сумме ее диагональных элементов (следу матрицы), и, т.к. λ — меньшее из двух собственных чисел, то $\lambda < \frac{1}{2}(m_{11} + m_{12})$ (случай двух одинаковых собственных чисел не рассматривается, т.к. он имеет место, когда связь между переменными отсутствует, и $m_{12} = 0$).

Оценка свободного члена b, как и прежде, получается из условия прохождения линии регрессии через точку средних: $b = \bar{x} a$, где $\bar{x}$ — вектор-строка средних значений переменных.

Расчетное значение $\hat{x}_i$ дает вектор OD (см. рис. 4.2), который равен разности векторов OA и OF, т.е. (в матричной форме):

$$\hat{X}^c = \hat{X} - e a'.$$

Теперь можно дать еще одну оценку параметров уравнения (4.18):

$$\begin{aligned} a_{12}(\perp) &= -\frac{a_2}{a_1}, \\ b_1(\perp) &= \bar{x}_1 - a_{12}(\perp)\, \bar{x}_2, \\ e_{i1}(\perp) &= \frac{1}{a_1} e_i. \end{aligned}$$

Полученная оценка углового коэффициента $a_{12}(\perp)$ лежит между его оценками по прямой и обратной регрессиям. Действительно, из (4.36) следует, что

$$a_{12}(\perp) = -\frac{a_2}{a_1} = \frac{m_{12}}{m_{22} - \lambda} = \frac{m_{11} - \lambda}{m_{12}}.$$

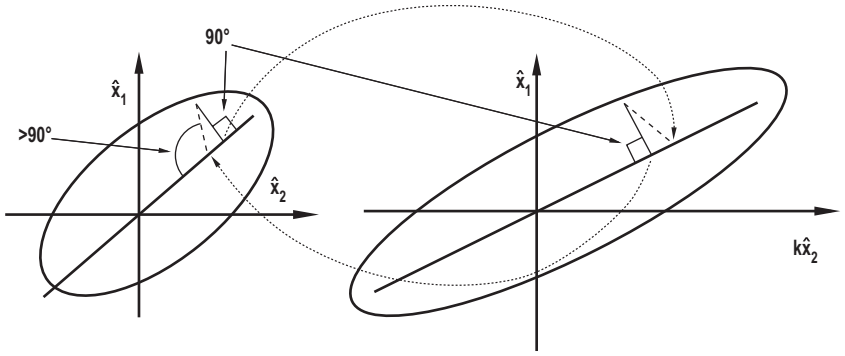

Рис. 4.3

Отсюда, в частности, следует, что величины $m_{11}-\lambda$ и $m_{22}-\lambda$ имеют один знак, и, т.к. $\lambda<\frac{1}{2}(m_{11}+m_{12})$, то обе эти величины положительны.

Поэтому, если $m_{12}\geqslant 0$, то

$$\frac{m_{11}}{m_{12}}\overset{(4.32)}{=}a_{12}(2)>a_{12}(\perp)>a_{12}\overset{(4.26)}{=}\frac{m_{12}}{m_{22}},$$

а если $m_{12}\leqslant 0$, то $a_{12}(2)<a_{12}(\perp)<a_{12}$.

Понятно, что эти 3 оценки совпадают тогда и только тогда, когда $\lambda\,(=s_e^2)=0$, т.е. зависимость функциональна.

В действительности любое число, лежащее на отрезке с концами $a_{12}, a_{12}(2)$ (т.е. либо $[a_{12}, a_{12}(2)]$, если $m_{12}\geqslant 0$, либо $[a_{12}(2), a_{12}]$, если $m_{12}\leqslant 0$), может являться МНК-оценкой параметра α_{12}, т.е. оценкой этого параметра является $\gamma_1 a_{12}+\gamma_2 a_{12}(2)$ при любых γ_1 и γ_2, таких что $\gamma_1\geqslant 0$, $\gamma_2\geqslant 0$, $\gamma_1+\gamma_2=1$. Каждая из этих оценок может быть получена, если расстояния от точек облака наблюдения до линии регрессии измерять под определенным углом, что достигается с помощью предварительного преобразования в пространстве переменных.

Убедиться в этом можно, рассуждая следующим образом.

Пусть получена оценка углового коэффициента по ортогональной регрессии (рис. 4.3, слева). Теперь проводится преобразование в пространстве переменных: $\hat{x}_2$ умножается на некоторое число $k>1$, и снова дается оценка этого коэффициента по ортогональной регрессии (рис. 4.3, справа). После возвращения в исходное пространство получается новая оценка углового коэффициента, сопоставимая со старой (возвращение в исходное пространство осуществляется умножением оценки коэффициента, полученной в преобразованном пространстве, на число k).

Этот рисунок не вполне корректен, т.к. переход в новое пространство переменных и возвращение в исходное пространство ведет к смещению линии регрессии. Однако

смысл происходящего он поясняет достаточно наглядно: новая оценка получена так, как будто расстояния от точек облака наблюдений до линии регрессии измеряются под углом, не равным $90°$. Должно быть понятно, что в пределе, при $k \to \infty$, расстояния до линии регрессии будут измеряться параллельно оси $\hat{x}_1$ и полученная оценка углового коэффициента совпадет с a_{12}. Наоборот, в пределе при $k \to 0$ эта оценка совпадет с a_{12} (2).

Выбор оценок параметров регрессии на имеющемся множестве зависит от характера распределения ошибок измерения переменных. Это — предмет изучения во II части книги. Пока можно предложить некоторые эмпирические критерии. Например, следующий.

Общая совокупность (множество наблюдений) делится на две части: обучающую и контрольную. Оценка параметров производится по обучающей совокупности. На контрольной совокупности определяется сумма квадратов отклонений фактических значений переменных от расчетных. Выбирается та оценка, которая дает минимум этой суммы. В заключение выбранную оценку можно дать по всей совокупности.

Рассмотренный случай двух переменных легко обобщить на n переменных (без доказательств: они даются во II части книги). Основное уравнение регрессии записывается следующим образом: $x_1 = x_{-1}\alpha_{-1} + \beta_1 + \varepsilon_1$, где $x_{-1} = [x_2, \ldots, x_n]$ — вектор-строка всех переменных кроме первой, вектор факторных переменных,

$$\alpha_{-1} = \begin{bmatrix} \alpha_{12} \\ \vdots \\ \alpha_{1n} \end{bmatrix}$$

— вектор-столбец параметров регрессии при факторных переменных, а в матричной форме: $\hat{X}_1 = \hat{X}_{-1}a_{-1} + e_1$, где $\hat{X}_{-1}$ — матрица размерности $N \times (n-1)$ наблюдений за факторными переменными.

По аналогии с (4.21, 4.26):

$$\begin{aligned} a_{-1} &= M_{-1}^{-1}m_{-1}, \\ b_1 &= \bar{x}_1 - \bar{x}_{-1}a_{-1}, \end{aligned} \tag{4.37}$$

где $M_{-1} = \frac{1}{N}\hat{X}'_{-1}\hat{X}_{-1}$ — матрица ковариации факторных переменных между собой,

$m_{-1} = \frac{1}{N}\hat{X}'_{-1}\hat{X}_1$ — вектор-столбец ковариации факторных переменных с моделируемой переменной,

$\bar{x}_{-1} = \frac{1}{N} 1'_N \hat{X}_{-1}$ — вектор-строка средних значений факторных переменных.

Расчетные значения моделируемой переменной, т.е. ее математические ожидания, есть

$$\hat{X}_1^c = \hat{X}_{-1} a_{-1}.$$

Как и в случае двух переменных объясненной дисперсией является дисперсия расчетных значений моделируемой переменной:

$$s_{q1}^2 = \frac{1}{N} a'_{-1} \hat{X}'_{-1} \hat{X}_{-1} a_{-1} = a'_{-1} M_{-1} a_{-1} \overset{(4.37)}{=} a'_{-1} m_{-1} \overset{(4.37)}{=} m'_{-1} M_{-1}^{-1} m_{-1}. \tag{4.38}$$

Коэффициент множественной корреляции $r_{1,-1}$ есть коэффициент корреляции между моделируемой переменной и ее расчетным значением (**cov** — обозначение ковариации):

$$\mathbf{cov}\,(x_1^c, x_1) = \frac{1}{N} a'_{-1} \hat{X}'_{-1} \hat{X}_1 = a'_{-1} m_{-1} \overset{(4.38)}{=} s_{q1}^2,$$

$$r_{1,-1} = \frac{\mathbf{cov}\,(x_1^c, x_1)}{\sqrt{\mathbf{var}\,(x_1^c)\,\mathbf{var}\,(x_1)}} = \frac{s_{q1}^2}{s_{q1} s_1} = \frac{s_{q1}}{s_1},$$

Коэффициент детерминации, равный квадрату коэффициента множественной корреляции:

$$R^2 = \frac{s_{q1}^2}{s_1^2},$$

показывает долю объясненной дисперсии в общей.

Если связь отсутствует и $\alpha_{-1} = 0$ (нулевая гипотеза), то расчетная статистика Фишера

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

имеет F-распределение с $n - 1$ степенями свободы в числителе и $N - n$ степенями свободы в знаменателе — $F_{n-1,N-n}$. Логика использования этой статистики сохраняется прежней.

При использовании в общем случае записи уравнения регрессии в форме со скрытым свободным членом

$$X_1 = \tilde{X}_{-1} \tilde{a}_{-1} + e,$$

где $\tilde{X}_{-1}$ — матрица $[X_{-1}, 1_N]$ размерности $N \times (n+1)$, $\tilde{a}_{-1}$ — вектор $\begin{bmatrix} a_{-1} \\ b_1 \end{bmatrix}$, оператор МНК-оценивания записывается как

$$\tilde{a}_{-1} = \tilde{M}_{-1}^{-1} \tilde{m}_{-1}, \tag{4.39}$$

где $\tilde{m}_{-1} = \frac{1}{N} \tilde{X}'_{-1} X_1$, $\tilde{M}_{-1} = \frac{1}{N} \tilde{X}'_{-1} \tilde{X}_{-1}$.

Достаточно простые алгебраические преобразования показывают, что этот оператор эквивалентен (4.37).

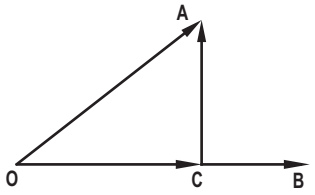

Рис. 4.4

Полезной является еще одна геометрическая иллюстрация регрессии — в пространстве наблюдений (см. рис. 4.4 и 4.5).

При $n = 2$ (n — количество переменных), OA — вектор $\hat{x}_1$, OB — вектор $\hat{x}_2$, OC — вектор проекции $\hat{x}_1$ на $\hat{x}_2$, равный расчетному значению $\hat{x}_1^c$, CA — вектор остатков e_1, так что: $\hat{x}_1 = a_{12}\hat{x}_2 + e_1$. Косинус угла между OA и OB равен коэффициенту корреляции.

При $n = 3$, OA — вектор $\hat{x}_1$, OB — вектор $\hat{x}_2$, OC — вектор $\hat{x}_3$, OD — вектор проекции $\hat{x}_1$ на плоскость, определяемую $\hat{x}_2$ и $\hat{x}_3$, равный расчетному значению $\hat{x}_1^c$, DA — вектор остатков e_1, OE — вектор проекции $\hat{x}_1^c$ на $\hat{x}_2$, равный $a_{12}\hat{x}_2$, OF — вектор проекции $\hat{x}_1^c$ на $\hat{x}_3$, равный $a_{13}\hat{x}_3$, так что $\hat{x}_1 = a_{12}\hat{x}_2 + a_{13}\hat{x}_3 + e_1$. Косинус угла между OA и плоскостью, определенной $\hat{x}_2$ и $\hat{x}_3$, (т.е. между OA и OD) равен коэффициенту множественной корреляции.

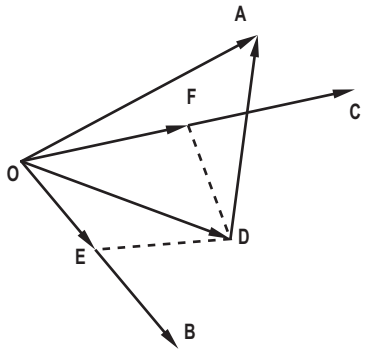

Рис. 4.5

Кроме оценки a_{-1} можно получить оценки $a_{-1}(j)$, $j = 2, \ldots, n$, последовательно переводя в левую часть уравнения переменные $\hat{x}_j$, применяя МНК и алгебраически возвращаясь к оценкам исходной формы уравнения.

Для представления ортогональной регрессии в общем случае подходят формулы (4.34, 4.36) и другие матричные выражения, приведенные выше при описании ортогональной регрессии. Необходимо только при определении векторов и матриц, входящих в эти выражения, заменить «2» на «n».

С помощью преобразований в пространстве переменных перед использованием ортогональной регрессии и последующего возвращения в исходное пространство

в качестве оценок a_{-1} можно получить любой вектор из множества (симплекса)

$$\gamma_1 a_{-1} + \sum_{j=2}^{n} \gamma_j a_{-1}(j), \quad \gamma_j \geqslant 0, \quad j = 1, \ldots, n, \quad \sum_{j=1}^{n} \lambda_j = 1.$$

Это — подмножество всех возможных МНК-оценок истинных параметров α_{-1}.

4.3. Дисперсионный анализ

Дисперсионный анализ заключается в представлении (разложении) дисперсии изучаемых признаков по факторам и использовании F-критерия для сопоставления факторных «частей» общей дисперсии с целью определения степени влияния факторов на изучаемые признаки. Примеры использования дисперсионного анализа даны в предыдущем пункте при рассмотрении общей дисперсии моделируемой переменной как суммы объясненной и остаточной дисперсии.

Дисперсионный анализ может быть **одномерным** или **многомерным**. В первом случае имеется только один изучаемый (моделируемый) признак, во втором случае их несколько. В данном курсе рассматривается только первый случай. Применение методов этого анализа основывается на определенной группировке исходной совокупности (см. п. 1.9). В качестве факторных выступают группирующие признаки. То есть изучается влияние группирующих признаков на моделируемый. Если группирующий (факторный) признак один, то речь идет об **однофакторном** дисперсионном анализе, если этих признаков несколько — о **многофакторном** анализе. Если в группировке для каждого сочетания уровней факторов имеется строго одно наблюдение (численность всех конечных групп в точности равна единице), говорят о дисперсионном анализе **без повторений**; если конечные группы могут иметь любые численности — **с повторениями**. Многофакторный дисперсионный анализ может быть **полным** или **частичным**. В первом случае исследуется влияние всех возможных сочетаний факторов (смысл этой фразы станет понятным ниже). Во втором случае принимаются во внимание лишь некоторые сочетания факторов.

В этом пункте рассматриваются две модели: однофакторный дисперсионный анализ с повторениями и полный многофакторный анализ без повторений.

Пусть исходная совокупность x_i, $i = 1, \ldots, N$ сгруппирована по одному фактору, т.е. она разделена на k групп:

$x_{i_l l}$ — значение изучаемого признака в i_l-м наблюдении ($i_l = 1, \ldots, N_l$) в l-й группе ($l = 1, \ldots, k$); $\sum N_l = N$.

Рассчитываются общая средняя и средние по группам:

$$\bar{x} = \frac{1}{N}\sum_{l=1}^{k}\sum_{i_l=1}^{N_l} x_{i_l l} = \frac{1}{N}\sum_{l=1}^{k} N_l \bar{x}_l,$$

$$\bar{x}_l = \frac{1}{N_l}\sum_{i_l=1}^{N_l} x_{i_l l},$$

общая дисперсия, дисперсии по группам и межгрупповая дисперсия (s_q^2):

$$s^2 = \frac{1}{N}\sum_{l=1}^{k}\sum_{i_l=1}^{N_l} (x_{i_l l} - \bar{x})^2,$$

$$s_l^2 = \frac{1}{N_l}\sum_{i_l=1}^{N_l} (x_{i_l l} - \bar{x}_l)^2,$$

$$s_q^2 = \frac{1}{N}\sum_{l=1}^{k} N_l\,(\bar{x}_l - \bar{x})^2.$$

Общую дисперсию можно разложить на групповые и межгрупповую дисперсии:

$$s^2 = \frac{1}{N}\sum_{l=1}^{k}\sum_{i_l=1}^{N_l} \left((x_{i_l l} - \bar{x}_l) + (\bar{x}_l - \bar{x})\right)^2 =$$

$$= \frac{1}{N}\sum_{l=1}^{k}\sum_{i_l=1}^{N_l} (x_{i_l l} - \bar{x}_l)^2 + \frac{2}{N}\sum_{l=1}^{k}\sum_{i_l=1}^{N_l} (x_{i_l l} - \bar{x}_l)\,(\bar{x}_l - \bar{x}) + \frac{1}{N}\sum_{l=1}^{k}\sum_{i_l=1}^{N_l} (\bar{x}_l - \bar{x})^2 =$$

$$= \frac{1}{N}\sum_{l=1}^{k} N_l \frac{1}{N_l}\sum_{i_l=1}^{N_l} (x_{i_l l} - \bar{x}_l)^2 + \frac{2}{N}\sum_{l=1}^{k}(\bar{x}_l - \bar{x})\underbrace{\sum_{i_l=1}^{N_l}(x_{i_l l} - \bar{x}_l)}_{=0} + \frac{1}{N}\sum_{l=1}^{k} N_l\,(\bar{x}_l - \bar{x})^2 =$$

$$= \frac{1}{N}\sum_{l=1}^{k} N_l s_l^2 + s_q^2 = s_e^2 + s_q^2.$$

Данное представление общей дисперсии изучаемого признака аналогично полученному в начале предыдущего пункта при рассмотрении регрессии, построенной по данным совместного эмпирического распределения признаков. В том случае «группами» выступали значения первого признака при тех или иных значениях второго признака. В данном случае (в терминах дисперсионного анализа)

s_e^2 — внутригрупповая дисперсия;

s_q^2 — межгрупповая дисперсия.

Тот факт, что среднее слагаемое в вышеприведенном выражении равно нулю, означает линейную независимость внутригрупповой и межгрупповой дисперсий.

Чем выше межгрупповая дисперсия по сравнению с внутригрупповой, тем вероятнее, что группирующий (факторный) признак влияет на изучаемый признак. Степень возможного влияния оценивается с помощью F-статистики:

$$F^c = \frac{s_q^2 (N-k)}{s_e^2 (k-1)}.$$

В случае если влияние отсутствует (нулевая гипотеза), эта статистика имеет распределение $F_{k-1,N-k}$ (межгрупповая дисперсия имеет $k-1$ степеней свободы, внутригрупповая — $N-k$), что объясняет указанный способ расчета F-статистики. Логика проверки нулевой гипотезы та же, что и в предыдущих случаях.

Рассмотрение модели однофакторного дисперсионного анализа с повторениями завершено.

Пусть теперь имеется группировка исходной совокупности x_i, $i = 1, \ldots, N$ по n факторам; j-й фактор может принимать k_j уровней, $j = 1, \ldots, n$. Все численности конечных групп равны единице: $N_I = 1$, для любого I. Такая совокупность может быть получена по результатам проведения управляемого эксперимента. В экономических исследованиях она может быть образована в расчетах по математической модели изучаемой переменной: для каждого сочетания уровней факторов проводится один расчет по модели.

В этом случае

$$N = \prod_{j=1}^{n} k_j = \prod_{G} k_j,$$

где через G, как и в пункте 1.9, обозначено полное множество факторов $J = \{12 \ldots n\}$, x_I — значение изучаемого признака при сочетании уровней факторов $I = \{i_1 i_2 \ldots i_n\}$.

Общая средняя изучаемого признака:

$$b_0 = \bar{x} = \frac{1}{N} \sum_{I} x_I.$$

Каждый j-й фактор делит исходную совокупность на k_j групп по N/k_j элементов. Для каждого из уровней i_j j-го фактора (для каждой из таких групп) рассчитывается среднее значение изучаемого признака:

$$x_{i_j(j)} = \frac{k_j}{N} \sum_{I - i_j(j)} x_I,$$

где $\sum\limits_{I-i_j(j)}$ означает суммирование по всем наблюдениям, в которых j-й фактор находится на уровне i_j.

Если бы тот факт, что j-й фактор находится на уровне i_j, не влиял на изучаемый признак, означало бы, что

$$x_{i_j(j)} = b_0.$$

Потому $b_{i_j(j)} = x_{i_j(j)} - b_0$ — коэффициент влияния на изучаемый признак того, что j-й фактор находится на уровне i_j. Это — **главные эффекты**, или **эффекты 1-го порядка**.

Очевидно, что

$$\sum_{i_j=1}^{k_j} b_{i_j(j)} = 0$$

и дисперсия, определенная влиянием j-го фактора, равна

$$s_j^2 = \frac{1}{k_j} \sum_{i_j=1}^{k_j} \left(b_{i_j(j)}\right)^2.$$

Каждые два фактора j и j' делят совокупность на $K^{jj'} = k_j k_{j'}$ групп по $\dfrac{N}{K^{jj'}}$ элементов. Для каждой из таких групп рассчитывается среднее изучаемого признака:

$$x_{i_j i_{j'}(jj')} = \frac{K^{jj'}}{N} \sum_{I-i_j i_{j'}(jj')} x_I,$$

где $\sum\limits_{I-i_j i_{j'}(jj')}$ означает суммирование по всем наблюдениям, в которых j-й фактор находится на уровне i_j, а j'-й фактор — на уровне $i_{j'}$.

Если бы тот факт, что одновременно j-й фактор находится на уровне i_j, а j'-й фактор — на уровне $i_{j'}$, не влиял на изучаемый признак, то это означало бы, что

$$x^{jj'}_{i_j i_{j'}(jj')} = b_0 + b_{i_j(j)} + b_{i_{j'}(j')}.$$

Поэтому

$$b_{i_j i_{j'}(jj')} = x_{i_j i_{j'}(jj')} - \left(b_0 + b_{i_j(j)} + b_{i_{j'}(j')}\right)$$

— коэффициент влияния на изучаемый признак того, что одновременно j-й фактор находится на уровне i_j, а j'-й фактор — на уровне $i_{j'}$. Это **эффекты взаимодействия** (или **сочетания**) факторов j и j', **парные эффекты**, или **эффекты 2-го порядка**.

Легко убедиться в том, что

$$\sum_{i_j=1}^{k_j} b_{i_j i_{j'}(jj')} = \sum_{i_{j'}=1}^{k_{j'}} b_{i_j i_{j'}(jj')} = 0,$$

и тогда

$$s^2_{jj'} = \frac{1}{Kjj'} \sum_{i_j, i_{j'}} \left(b_{i_j i_{j'}(jj')} \right)^2$$

— дисперсия, определенная совместным влиянием факторов j и j'.

Рассмотрим общий случай.

Факторы $J = \{j_1 j_2 \dots j_{n'}\}$, $n' \leqslant n$ делят совокупность на $K^J = \prod\limits_J k_j$ групп по $\dfrac{N}{K^J}$ элементов (выделяют группы класса J порядка n'). Мультииндексом таких групп является $I(J) = \{i_1 i_2 \dots i_{n'}\}$ $(\{j_1 j_2 \dots j_{n'}\}) = \{i_{j_1} i_{j_2} \dots i_{j_{n'}}\}$; конкретно данный мультииндекс именует группу, в которой фактор j_1 находится на уровне i_{j_1} и т.д. По каждой такой группе рассчитывается среднее изучаемого признака:

$$x_{I(J)} = \frac{K^J}{N} \sum_{I - I(J)} x_I,$$

где $\sum\limits_{I-I(J)}$ — означает суммирование по всем наблюдениям, в которых фактор j_1 находится на уровне i_{j_1} и т.д.

Как и в двух предыдущих случаях:

$$b_{I(J)} = x_{I(J)} - \left(b_0 + \sum_{\bar{J} \in J^-} b_{I(\bar{J})} \right) \tag{4.40}$$

— **эффекты взаимодействия (или сочетания) факторов J**, **эффекты порядка** n'. Здесь $\sum\limits_{\bar{J} \in J^-}$ — суммирование по всем подмножествам множества J без самого множества J.

Суммирование этих коэффициентов по всем значениям любого индекса, входящего в мультииндекс $I(J)$ дает нуль.

$$s^2_J = \frac{1}{K^J} \sum_{I(J)} b^2_{I(J)}$$

— дисперсия, определенная совместным влиянием факторов J.

При определении эффектов наивысшего порядка

$$J = G,\ x_{I(G)} = x_I,\ K^G = N.$$

Из способа получения коэффициентов эффектов должно быть понятно, что

$$x_I = b_0 + \sum_{J=1}^{G} b_{I(J)}.$$

Все факторные дисперсии взаимно независимы и общая дисперсия изучаемого признака в точности раскладывается по всем возможным сочетаниям факторов:

$$s^2 = \sum_{J=1}^{G} s_J^2. \tag{4.41}$$

Данное выражение называют **дисперсионным представлением**, или **тождеством**.

Этот факт доказывается в IV части книги.

Пока можно его только проверить, например, при $n = 2$.

Используя 1-й способ обозначений (см. п. 4.1):

$$b_0 = \frac{1}{k_1 k_2} \sum_{i_1, i_2} x_{i_1 i_2},$$

$$x_{i_1 *} = \frac{1}{k_2} \sum_{i_2} x_{i_1 i_2}, \quad b_{i_1 *} = x_{i_1 *} - b_0, \quad s_1^2 = \frac{1}{k_1} \sum_{i_1} b_{i_1 *}^2,$$

$$x_{* i_2} = \frac{1}{k_1} \sum_{i_1} x_{i_1 i_2}, \quad b_{* i_2} = x_{* i_2} - b_0, \quad s_2^2 = \frac{1}{k_2} \sum_{i_2} b_{* i_2}^2,$$

$$b_{i_1 i_2} = x_{i_1 i_2} - b_0 - b_{i_1 *} - b_{* i_2}, \quad s_{12}^2 = \frac{1}{k_1 k_2} \sum_{i_1, i_2} b_{i_1 i_2}^2.$$

Теперь, раскрывая скобки в выражении для s_{12}^2 и учитывая, что $\hat{x}_{i_1 i_2} = x_{i_1 i_2} - b_0$, получаем:

$$s_{12}^2 = \frac{1}{k_1 k_2} \sum_{i_1, i_2} \hat{x}_{i_1 i_2}^2 + \frac{1}{k_1} \sum_{i_1} b_{i_1 *}^2 + \frac{1}{k_2} \sum_{i_2} b_{* i_2}^2 - \frac{2}{k_1 k_2} \sum_{i_1} b_{i_1 *} \underbrace{\sum_{i_2} \hat{x}_{i_1 i_2}}_{= k_2 b_{i_1 *}} -$$

$$- \frac{2}{k_1 k_2} \sum_{i_2} b_{* i_2} \underbrace{\sum_{i_1} \hat{x}_{i_1 i_2}}_{= k_1 b_{* i_2}} + \underbrace{\frac{2}{k_1 k_2} \sum_{i_1} b_{i_1 *} \sum_{i_2} b_{* i_2}}_{= 0 \quad = 0} = s^2 - s_1^2 - s_2^2.$$

$= 0$

Т.е. $s^2 = s_1^2 + s_2^2 + s_{12}^2$, что и требовалось показать.

В силу взаимной независимости эффектов оценки коэффициентов и дисперсий эффектов остаются одинаковыми в любой модели частичного анализа (в котором рассматривается лишь часть всех возможных сочетаний факторов) и совпадают с оценками полного анализа.

Дисперсия s_J^2 имеет $K_-{}^J$ степеней свободы:

$$K_-{}^J = \prod_J (k_j - 1)\,.$$

Сумма этих величин по всем J от 1 до G равна $N-1$. В этом легко убедиться, если раскрыть скобки в следующем тождестве:

$$N = \prod_G ((k_j - 1) + 1).$$

Процедура определения степени влияния факторов на изучаемый признак может быть следующей.

На 1-м шаге выбирается сочетание факторов J_1, оказывающих наибольшее влияние на изучаемый признак. Этими факторами будут такие, для которых минимума достигает показатель pv статистики Фишера

$$F_1^c = \frac{s_{J_1}^2 \left(N - K_-^{J_1} - 1\right)}{\left(s^2 - s_{J_1}^2\right) K_-^{J_1}}.$$

На 2-м шаге выбирается сочетание факторов J_2, для которого минимума достигает показатель pv статистики Фишера

$$F_2^c = \frac{\left(s_{J_1}^2 + s_{J_2}^2\right)\left(N - K_-^{J_1} - K_-^{J_2} - 1\right)}{\left(s^2 - s_{J_1}^2 - s_{J_2}^2\right)\left(K_-^{J_1} + K_-^{J_2}\right)}.$$

И так далее. Процесс прекращается, как только показатель pv достигнет заданного уровня ошибки, например, 0.05. Пусть этим шагом будет t-й. Оставшиеся сочетания факторов формируют остаточную дисперсию. Как правило, в таком процессе сначала выбираются главные эффекты, затем парные и т.д., так что остаточную дисперсию образуют эффекты высоких порядков.

Расчетные значения изучаемого признака определяются по следующей формуле:

$$x_I^c = b_0 + \sum_{l=1}^{t} b_{I(J_l)}.$$

Этим завершается рассмотрение модели полного многофакторного дисперсионного анализа без повторений.

Несколько слов можно сказать о многофакторном дисперсионном анализе с повторениями.

Если все $N_I \geqslant 1$, можно попытаться свести этот случай к предыдущему.

Для каждой конечной группы рассчитываются среднее $\bar{x}_I$ и дисперсия s_I^2. Используя приведенные выше формулы можно рассчитать коэффициенты и дисперсии всех эффектов, заменяя x_I на $\bar{x}_I$. К сожалению, в общем случае эффекты перестают быть взаимно независимыми, и в представлении общей дисперсии (4.41) кроме дисперсий эффектов различных сочетаний факторов появляются слагаемые с нижним индексом $J\bar{J}$. Возникает неопределенность результатов и зависимость их от того набора сочетаний факторов, которые включены в анализ. Поэтому разные модели частичного анализа дают разные результаты, отличные от полного анализа.

Имеется несколько частных случаев, в которых «хорошие» свойства оценок сохраняются. Один из них — случай, когда все численности конечных групп одинаковы. Тогда дисперсионное тождество записывается следующим образом:

$$s^2 = \sum_{J=1}^{G} s_J^2 + \underbrace{\sum_{I=I_1}^{I_K} s_I^2}_{s_e^2},$$

причем последнее слагаемое — остаточная, или внутригрупповая дисперсия — имеет $N - K_-{}^G - 1$ степеней свободы.

4.4. Анализ временных рядов

Временным или динамическим рядом называется совокупность наблюдений x_i в последовательные моменты времени $i = 1, \ldots, N$ (обычно для индексации временных рядов используется t, в этом пункте для целостности изложения материала сохранено i). Задача анализа временного ряда заключается в выделении и моделировании 3-х его основных компонент:

$$x_i = \delta_i + \gamma_i + \varepsilon_i, \quad i = 1, \ldots, N,$$

или в оценках:

$$x_i = d_i + c_i + e_i, \quad i = 1, \ldots, N,$$

где

$\delta_i,\ d_i$ — **тренд**, **долговременная тенденция**,

$\gamma_i,\ c_i$ — **цикл**, **циклическая составляющая**,

$\varepsilon_i,\ e_i$ — **случайная компонента**,

с целью последующего использования построенных моделей в прикладном экономическом анализе и прогнозировании.

Для выявления долгосрочной тенденции используют различные методы.

Наиболее распространено использование **полиномиального** тренда. Такой тренд строится как регрессия x_i на полином определенной степени относительно времени:

$$x_i = a_1 i + a_2 i^2 + \ldots + b + e_i, \quad i = 1, \ldots, N.$$

Для выбора степени полинома можно использовать F-критерий: оценивают тренд как полином, последовательно увеличивая его степень до тех пор, пока удается отвергнуть нулевую гипотезу.

Тренд может быть **экспоненциальным**. Он строится как регрессия $\ln x_i$ на полином от времени, так что после оценки параметров регрессии его можно записать в следующем виде:

$$x_i = e^{a_1 i + a_2 i^2 + \ldots + b + e_i}, \quad i = 1, \ldots, N.$$

Иногда тренд строится как **сплайн**, т.е. как некоторая «гладкая» композиция разных функций от времени на разных подпериодах.

Пусть, например, на двух подпериодах $[1, \ldots, N_1]$ и $[N_1 + 1, \ldots, N]$ тренд выражается разными квадратическими функциями от времени (в момент времени N_1 происходит смена тенденции):

$$\begin{aligned} x_i &= a_1 i + a_2 i^2 + b_1 + e_{i1}, \quad i = 1, \ldots, N_1, \\ x_i &= a_3 i + a_4 i^2 + b_2 + e_{i2}, \quad i = N_1 + 1, \ldots, N. \end{aligned}$$

Для того чтобы общий тренд был «гладким» требуют совпадения самих значений и значений первых производных двух полиномов в точке «перелома» тенденции:

$$\begin{aligned} a_1 N_1 + a_2 N_1^2 + b_1 &= a_3 N_1 + a_4 N_1^2 + b_2, \\ a_1 + 2 a_2 N_1 &= a_3 + 2 a_4 N_1. \end{aligned}$$

Отсюда выражают, например, a_3 и b_2 через остальные параметры и подставляют полученные выражения в исходное уравнение регрессии. После несложных преобразований уравнение приобретает следующий вид:

$$\begin{aligned} x_i &= a_1 i + a_2 i^2 + b_1 + e_{i1}, \quad i = 1, \ldots, N_1, \\ x_i &= a_1 i + a_2 \left(i^2 - (i - N_1)^2 \right) + b_1 + a_4 (i - N_1)^2 + e_{i2}, \quad i = N_1 + 1, \ldots, N. \end{aligned}$$

Параметры полученного уравнения оцениваются, и, тем самым, завершается построение тренда как полиномиального сплайна.

Для выявления долговременной тенденции применяют также различные приемы **сглаживания** динамических рядов с помощью **скользящего среднего**.

Один из подходов к расчету скользящей средней заключается в следующем: в качестве сглаженного значения x_i, которое по аналогии с расчетным значением можно обозначить через x_i^c, принимается среднее значений $x_{i-p}, \ldots, x_i, \ldots, x_{i+p}$, где p — полупериод сглаживания. Сам процесс сглаживания заключается в последовательном расчете (скольжении средней) $x_{p+1}^c, \ldots, x_{N-p}^c$. При этом часто теряются первые и последние p значений исходного временного ряда.

Для сглаживания могут использоваться различные средние. Так, например, при **полиномиальном сглаживании** средние рассчитываются следующим образом.

Пусть сглаживающим является полином q-й степени. Оценивается регрессия вида:

$$x_{i+l} = a_1 l + a_2 l^2 + \ldots + a_q l^q + b + e_{i+l}, \quad l = -p, \ldots, p,$$

и в качестве сглаженного значения x_i^c принимается b (расчетное значение при $l = 0$).

Так, при $q = 2$ и $p = 2$ уравнение регрессии принимает следующий вид (исключая i как текущий индекс):

$$\begin{bmatrix} x_{-2} \\ x_{-1} \\ x_0 \\ x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -2 & 4 & 1 \\ -1 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \\ 2 & 4 & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ b \end{bmatrix} + \begin{bmatrix} e_{-2} \\ e_{-1} \\ e_0 \\ e_1 \\ e_2 \end{bmatrix}.$$

По аналогии с (4.29), можно записать:

$$\begin{bmatrix} a_1 \\ a_2 \\ b \end{bmatrix} = \left[\begin{bmatrix} -2 & -1 & 0 & 1 & 2 \\ 4 & 1 & 0 & 1 & 4 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} -2 & 4 & 1 \\ -1 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \\ 2 & 4 & 1 \end{bmatrix} \right]^{-1} \begin{bmatrix} -2 & -1 & 0 & 1 & 2 \\ 4 & 1 & 0 & 1 & 4 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_{-2} \\ x_{-1} \\ x_0 \\ x_1 \\ x_2 \end{bmatrix} =$$

$$= \frac{1}{70}\begin{bmatrix} -14 & -7 & 0 & 7 & 14 \\ 10 & -5 & -10 & -5 & 10 \\ -6 & 24 & 34 & 24 & -6 \end{bmatrix}\begin{bmatrix} x_{-2} \\ x_{-1} \\ x_0 \\ x_1 \\ x_2 \end{bmatrix}.$$

Таким образом, в данном случае веса скользящей средней принимаются равными

$$\frac{1}{35}[-3,\ 12,\ 17,\ 12,\ -3].$$

При полиномиальном сглаживании потеря первых и последних p наблюдений в сглаженном динамическом ряду не является неизбежной; их можно взять как расчетные значения соответствующих наблюдений по первому и последнему полиному (в последовательности скольжения средней).

Так, в рассмотренном примере при $p = q = 2$:

$$\begin{bmatrix} x_1^c \\ x_2^c \end{bmatrix} = \begin{bmatrix} -2a_1 + 4a_2 + b \\ -a_1 + a_2 + b \end{bmatrix} = \frac{1}{35}\begin{bmatrix} 31 & 9 & -3 & -5 & 3 \\ 9 & 13 & 12 & 6 & -5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix},$$

$$\begin{bmatrix} x_{N-1}^c \\ x_N^c \end{bmatrix} = \begin{bmatrix} a_1 + a_2 + b \\ 2a_1 + 4a_2 + b \end{bmatrix} = \frac{1}{35}\begin{bmatrix} -5 & 6 & 12 & 13 & 9 \\ 3 & -5 & -3 & 9 & 31 \end{bmatrix}\begin{bmatrix} x_{N-4} \\ x_{N-3} \\ x_{N-2} \\ x_{N-1} \\ x_N \end{bmatrix}.$$

Как видно, все эти расчетные значения являются средними взвешенными величинами с несимметричными весами.

Для выбора параметров сглаживания p и q можно воспользоваться F-критерием (применение этого критерия в данном случае носит эвристический

характер). Для каждой проверяемой пары p и q рассчитывается сначала остаточная дисперсия:

$$s_e^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - x_i^c)^2,$$

а затем F-статистика:

$$F^c = \frac{(s_x^2 - s_e^2)\,(2p-q)}{s_e^2 q},$$

где s_x^2 — полная дисперсия ряда.

Выбираются такие параметры сглаживания, при которых эта статистика (q степеней свободы в числителе и $2p-q$ степеней свободы в знаменателе) имеет наименьший показатель pv.

Другой способ сглаживания называется **экспоненциальным**. При таком способе в качестве сглаженного (расчетного) значения принимается среднее всех предыдущих наблюдений с экспоненциально возрастающими весами:

$$x_{i+1}^c = (1-a)\sum_{l=0}^{\infty} a^l x_{i-l},$$

где $0 < a < 1$ — параметр экспоненциального сглаживания (x_i^c является на самом деле средней, т.к. $\sum\limits_{l=0}^{\infty} a^l = \dfrac{1}{1-a}$).

В такой форме процедура сглаживания неоперациональна, поскольку требует знания всей предыстории — до минус бесконечности. Но если из x_{i+1}^c вычесть ax_i^c, то весь «хвост» предыстории взаимно сократится:

$$x_{i+1}^c - ax_i^c = (1-a)x_i + \underbrace{(1-a)\sum_{l=1}^{\infty} a^l x_{i-l}}_{\longleftarrow\!\!\!

случае является достаточно сложной задачей, поскольку относительно a она (остаточная дисперсия) является полиномом степени $2(N-1)$ (по x_1^c — квадратичной функцией).

Пусть долговременная тенденция выявлена. На ее основе можно попытаться сразу дать прогноз моделируемой переменной (прогноз, по-видимому, будет точнее, если в нем учесть все компоненты временного ряда).

В случае тренда как аналитической функции от времени i, прогнозом является расчетное значение переменной в моменты времени $N+1,\ ,N+2,\ \ldots$.

Процедура экспоненциального сглаживания дает прогноз на один момент времени вперед:

$$x_{N+1}^c = (1-a)\, x_N + a x_N^c.$$

Последующие значения «прогноза» не будут меняться, т.к. отсутствуют основания для определения ошибки e_{N+1} и т.д. и, соответственно, для наблюдения различий между x_{N+1}^c и x_{N+1} и т.д.

При полиномиальном сглаживании расчет x_{N+1}^c проводится по последнему полиному (в последовательности скольжения средней) и оказывается равным некоторой средней последних $2p+1$ наблюдений во временном ряду.

В приведенном выше примере ($p=q=2$):

$$x_{N+1}^c = (b + 3a_1 + 9a_2) = \frac{1}{35}\begin{bmatrix} 21 & -21 & -28 & 0 & 63 \end{bmatrix}\begin{bmatrix} x_{N-4} \\ x_{N-3} \\ x_{N-2} \\ x_{N-1} \\ x_N \end{bmatrix}.$$

Определение циклической и случайной составляющей временного ряда дается во II части книги.

4.5. Упражнения и задачи

Упражнение 1

На основании информации о весе и росте студентов вашего курса:

1.1. Сгруппируйте студентов по росту и весу (юношей и девушек отдельно).

1.2. Дайте табличное и графическое изображение полученных совместных распределений частот, сделайте выводы о наличии связи между признаками.

1.3. С помощью критерия Пирсона проверьте нулевую гипотезу о независимости роста и веса студентов.

1.4. С помощью дисперсионного анализа установите, существенно ли влияние роста на их вес.

1.5. На основе построенной таблицы сопряженности рассчитайте средние и дисперсии роста и веса, а также абсолютную и относительную ковариацию между ними.

1.6. На основе исходных данных, без предварительной группировки (для юношей и девушек отдельно):

- Оцените с помощью МНК параметры линейного регрессионного уравнения, предположив, что переменная «рост» объясняется переменной «вес». Дайте интерпретацию полученным коэффициентам уравнения регрессии.
- Повторите задание, предположив, что переменная «вес» объясняется переменной «рост».
- Оцените с помощью МНК параметры ортогональной регрессии.
- Изобразите диаграмму рассеяния признаков роста и веса и все три линии регрессии. Объясните почему, если поменять экзогенные и эндогенные переменные местами, получаются различные уравнения.
- Для регрессионной зависимости роста от фактора веса вычислите объясненную и остаточную дисперсию, рассчитайте коэффициент детерминации и с помощью статистики Фишера проверьте статистическую значимость полученного уравнения.

Упражнение 2

Дана таблица (табл. 4.1, индекс Доу—Джонса средних курсов на акции ряда промышленных компаний).

2.1. Изобразить данные, представленные в таблице, графически.

2.2. Найти оценки параметров линейного тренда. Вычислить и изобразить графически остатки от оценки линейного тренда.

2.3. На основе данных таблицы

Таблица 4.1

Год	Индекс	Год	Индекс	Год	Индекс
1897	45.5	1903	55.5	1909	92.8
1898	52.8	1904	55.1	1910	84.3
1899	71.6	1905	80.3	1911	82.4
1900	61.4	1906	93.9	1912	88.7
1901	69.9	1907	74.9	1913	79.2
1902	65.4	1908	75.6		

- произвести сглаживание ряда с помощью процедуры, основывающейся на $q = 1$ и $p = 3$ (q — степень полинома, p — полупериод сглаживания);
- произвести сглаживание ряда с помощью процедуры, основывающейся на $q = 2$ и $p = 2$.

2.4. Сравнить сглаженный ряд с трендом, подобранным в упражнении 2.2.

Задачи

1. Используя интенсивность цвета для обозначения степени концентрации элементов в группах, дайте графическое изображение совокупности, характеризующейся:

   а) однородностью и прямой зависимостью признаков (x_1, x_2);

   б) однородностью и обратной зависимостью признаков (x_1, x_2);

   в) неоднородностью и прямой зависимостью признаков (x_1, x_2);

   г) неоднородностью и обратной зависимостью признаков (x_1, x_2);

   д) неоднородностью и отсутствием связи между признаками (x_1, x_2).

2. Пусть заданы значения (x_1, x_2). Объясните, какие приемы следует применять для оценки параметров следующих уравнений, используя обычный метод наименьших квадратов:

   а) $x_1 = \beta x_2^{\alpha}$;

   б) $x_2 = \beta e^{x_1 \alpha}$;

   в) $x_1 = \beta + \alpha \ln(x_2)$;

   г) $x_1 = x_2/(\beta + \alpha x_2)$;

   д) $x_1 = \beta + \alpha/(\pi - x_2)$.

3. Может ли матрица

   а) $\begin{bmatrix} 2 & 3 \\ 3 & 4 \end{bmatrix}$ б) $\begin{bmatrix} 4 & 3 \\ 2 & 3 \end{bmatrix}$

   являться ковариационной матрицей переменных, для которых строятся уравнения регрессии? Ответ обосновать.

4. Наблюдения трех пар (x_1, x_2) дали следующие результаты:

$$\sum_i x_{i1}^2 = 41,\ \sum_i x_{i2}^2 = 14,\ \sum_i x_{i1}x_{i2} = 23,\ \sum_i x_{i1} = 9,\ \sum_i x_{i2} = 6.$$

   Оценить уравнения прямой, обратной и ортогональной регрессии.

5. Построить уравнения прямой, обратной и ортогональной регрессии, если

   а) $X_1 = (1,\ 2,\ 3)'$, $X_2 = (1,\ 0,\ 5)'$;
   б) $X_1 = (0,\ 2,\ 0,\ 2)'$, $X_2 = (0,\ 0,\ 2,\ 2)'$;
   в) $X_1 = (0,\ 1,\ 1,\ 2)'$, $X_2 = (1,\ 0,\ 2,\ 1)'$.

   Нарисовать на графике в пространстве переменных облако наблюдений и линии прямой, обратной и ортогональной регрессии. Вычислить объясненную, остаточную дисперсию и коэффициент детерминации для каждого из построенных уравнений регрессии.

6. Какая из двух оценок коэффициента зависимости баллов, полученных на экзамене, от количества пропущенных занятий больше другой: по прямой или по обратной регрессии.

7. В регрессии $x_1 = a_{12}x_2 + 1_N b_1 + e_1$, фактор x_1 равен $(1,\ 3,\ 7,\ 1)'$. Параметры регрессии найдены по МНК. Могут ли остатки быть равными:

   а) $(1,\ -2,\ 2,\ 1)'$;
   б) $(1,\ -2,\ 1,\ -1)'$.

8. Для рядов наблюдений x_1 и x_2 известны средние значения, которые равны соответственно 10 и 5. Коэффициент детерминации в уравнениях регрессии x_1 на x_2 равен нулю. Найти значения параметров простой регрессии x_1 по x_2.

9. В регрессии $x_1 = a_{12}x_2 + 1_4 b_1 + e_1$, где $x_2 = (5,\ 3,\ 7,\ 1)'$, получены оценки $a_{12} = 2,\ b_1 = 1$, а коэффициент детерминации оказался равным 100%. Найти вектор фактических значений x_1.

10. Изобразите на графике в пространстве двух переменных облако наблюдений и линию прямой регрессии, если коэффициент корреляции между переменными:

    а) положительный;

    б) равен единице;

    в) отрицательный;

    г) равен минус единице;

    д) равен нулю.

11. Существенна ли связь между зарплатой и производительностью труда по выборке из 12 наблюдений, если матрица ковариаций для этих показателей имеет вид $\begin{bmatrix} 9 & 6 \\ 6 & 16 \end{bmatrix}$.

12. Оцените параметры ортогональной регрессии и рассчитайте остаточную дисперсию и коэффициент детерминации для переменных, у которых матрица ковариаций равна $\begin{bmatrix} 9 & 6 \\ 6 & 16 \end{bmatrix}$, а средние значения равны 3 и 4.

13. Имеются данные об объемах производства по четырем предприятиям двух отраслей, расположенным в двух регионах (млн. руб):

Регион \ Отрасль	1	2
1	48	60
2	20	40

    Рассчитать эффекты взаимодействия, факторную и общую дисперсии.

14. Имеются данные об инвестициях на предприятиях двух отраслей:

	Предприятие	Инвестиции (млн. руб.)
Отрасль 1	1	50
	2	60
	3	40
Отрасль 2	4	110
	5	160
	6	150

Рассчитать групповую, межгрупповую и общую дисперсии.

15. Имеются данные об урожайности культуры (в ц/га) в зависимости от способа обработки земли и сорта семян:

Сорт семян (A)	**Способы обработки земли (B)**			
	1	**2**	**3**	**4**
1	16	18	20	21
2	20	21	23	25
3	23	24	26	27

С помощью двухфакторного дисперсионного анализа оценить, зависит ли урожайность культуры от сорта семян (A) или от способа обработки земли.

16. Запишите систему нормальных уравнений оценивания параметров полиномиального тренда первой, второй и третей степеней.

17. Перенесите систему отсчета времени в середину ряда, т.е. $i = \ldots -3;\ -2;\ -1;\ 0;\ 1;\ 2;\ 3\ldots$, и перепишите систему нормальных уравнений для полиномиального тренда первой, второй и третей степеней. Как изменится вид системы? Найдите оценку параметров многочленов в явном виде из полученной системы уравнений.

18. По данным о выручке за 3 месяца: 11, 14, 15 — оцените параметры полиномиального тренда первой степени и сделайте прогноз выручки на четвертый месяц.

19. Имеются данные об ежедневных объемах производства (млн. руб.):

День	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Объем	9	12	27	15	33	14	10	26	18	24	38	28	45	32	41

Проведите сглаживание временного ряда, используя различные приемы скользящего среднего:

а) используя полиномиальное сглаживание;

б) используя экспоненциальное сглаживание.

20. Оценена регрессия $x_i = \beta + \alpha_s \sin(\omega i) + \alpha_c \cos(\omega i) + \varepsilon_i$ для частоты $\pi/2$. При этом $\alpha_s = 4$ и $\alpha_c = 3$. Найти значения амплитуды, фазы и периода.

21. Что называется гармоническими частотами? Записать формулу с расшифровкой обозначений.

22. Что такое частота Найквиста? Записать одним числом или символом.

23. Строится регрессия с циклическими компонентами:

$$x_i = \beta + \sum_{j=1}^{k} (\alpha_{sj} \sin(\omega_j i) + \alpha_{cj} \cos(\omega_j i)) + \varepsilon_i, \quad i = 1, \ldots, 5, \quad k = 2.$$

Запишите матрицу ковариаций факторов для данной регрессии.

Рекомендуемая литература

1. **Доугерти К.** Введение в эконометрику. — М.: «Инфра-М», 1997. (Гл. 2).
2. **Кендэл М.** Временные ряды. — М.: «Финансы и статистика», 1981. (Гл. 3–5, 8).
3. **Магнус Я.Р., Катышев П.К., Пересецкий А.А.** Эконометрика — начальный курс. — М.: «Дело», 2000. (Гл. 2).

Часть II

Эконометрия — I: Регрессионный анализ

Это пустая страница

В этой части развиваются положения 4-й главы «Введение в анализ связей» I-й части книги. Предполагается, что читатель знаком с основными разделами теории вероятностей и математической статистики (функции распределения случайных величин, оценивание и свойства оценок, проверка статистических гипотез), линейной алгебры (свойства матриц и квадратичных форм, собственные числа и вектора). Некоторые положения этих теорий в порядке напоминания приводятся в тексте.

В частности, в силу особой значимости здесь дается краткий обзор функций распределения, используемых в классической эконометрии (см. также Приложение А.3.2).

Пусть ε — случайная величина, имеющая нормальное распределение с нулевым математическим ожиданием и единичной дисперсией ($\varepsilon \sim N(0,1)$). Функция плотности этого распределения прямо пропорциональна $e^{-\frac{\varepsilon^2}{2}}$; 95-процентный двусторонний квантиль $\hat{\varepsilon}_{0.95}$ равен 1.96, 99-процентный квантиль — 2.57.

Пусть теперь имеется k таких взаимно независимых величин $\varepsilon_l \sim N(0,1)$, $l = 1, \ldots, k$. Сумма их квадратов $\sum_{l=1}^{k} \varepsilon_l^2$ является случайной величиной, имеющей распределение χ^2 с k степенями свободы (обозначается χ_k^2). Математическое ожидание этой величины равно k, а отношение χ_k^2/k при $k \to \infty$ стремится к 1, т.е. в пределе χ^2 становится детерминированной величиной. 95-процентный (односторонний) квантиль $\hat{\chi}^2_{k,0.95}$ при $k = 1$ равен 3.84 (квадрат 1.96), при $k = 5$ — 11.1, при $k = 20$ — 31.4, при $k = 100$ — 124.3 (видно, что отношение $\chi^2_{k,0.95}/k$ приближается к 1).

Если две случайные величины ε и χ_k^2 независимы друг от друга, то случайная величина $t_k = \dfrac{\varepsilon}{\sqrt{\chi_k^2/k}}$ имеет распределение t-Стьюдента с k степенями свободы. Ее функция распределения пропорциональна $\left(1 + \dfrac{t_k^2}{k}\right)^{-\frac{k+1}{2}}$; в пределе при $k \to \infty$ она становится нормально распределенной. 95-процентный двусторонний квантиль $\hat{t}_{k,\,0.95}$ при $k = 1$ равен 12.7, при $k = 5$ — 2.57, при $k = 20$ — 2.09, при $k = 100$ — 1.98, т.е. стремится к $\hat{\varepsilon}_{0.95}$.

Если две случайные величины $\chi^2_{k_1}$ и $\chi^2_{k_2}$ не зависят друг от друга, то случайная величина $F_{k_1,k_2} = \frac{\chi^2_{k_1}/k_1}{\chi^2_{k_2}/k_2}$ имеет распределение F-Фишера с k_1 и k_2 степенями свободы (соответственно, в числителе и знаменателе). При $k_2 \to \infty$ эта случайная величина стремится к $\chi^2_{k_1}/k_1$, т.е. $k_1 F_{k_1,\infty} = \chi^2_{k_1}$. Очевидно также, что $F_{1,k_2} = t^2_{k_2}$. 95-процентный (односторонний) квантиль $\hat{F}_{1,k_2,0.95}$ при $k_2 = 1$ равен 161, при $k_2 = 5$ — 6.61, при $k_2 = 20$ — 4.35, при $k_2 = 100$ — 3.94 (квадраты соответствующих $t_{k,0.95}$); квантиль $\hat{F}_{2,k_2,0.95}$ при $k_2 = 1$ равен 200, при $k_2 = 5$ — 5.79, при $k_2 = 20$ — 3.49, при $k_2 = 100$ — 3.09; квантиль $\hat{F}_{k_1,20,0.95}$ при $k_1 = 3$ равен 3.10, при $k_1 = 4$ — 2.87, при $k_1 = 5$ — 2.71, при $k_1 = 6$ — 2.60.

Глава 5

Случайные ошибки

Задачей регрессионного анализа является построение зависимости изучаемой случайной величины x от факторов z:

$$x = f(z, A) + \varepsilon,$$

где A — параметры зависимости.

Если z — истинный набор факторов, полностью определяющий значение x, а f — истинная форма зависимости, то ε — **случайные ошибки измерения** x. Однако в экономике весьма ограничены возможности построения таких истинных моделей, прежде всего потому, что факторов, влияющих на изучаемую величину, слишком много. В конкретных моделях в лучшем случае наборы z включают лишь несколько наиболее значимых факторов, и влияние остальных, неучтенных, факторов определяет ε. Поэтому ε называют просто **случайными ошибками** или **остатками**.

В любом случае считают, что ε — случайные величины с нулевым математическим ожиданием и, как правило, нормальным распределением. Последнее следует из центральной предельной теоремы теории вероятностей, поскольку ε по своему смыслу является результатом (суммой) действия многих мелких малозначимых по отдельности факторов случайного характера.

> Действительно, в соответствии с этой теоремой, случайная величина, являющаяся суммой большого количества других случайных величин, которые могут иметь различные распределения, но взаимно независимы и не слишком различаются между собой, имеет асимптотически нормальное распределение, т.е. чем больше случайных величин, тем ближе распределение их суммы к нормальному.

5.1. Первичные измерения

Пусть имеется N измерений x_i, $i = 1, \ldots, N$, случайной величины x, т.е. N наблюдений за случайной величиной. Предполагается, что измерения проведены в неизменных условиях (факторы, влияющие на x, не меняют своих значений), и систематические ошибки измерений исключены. Тогда различия в результатах отдельных наблюдений (измерений) связаны только с наличием случайных ошибок измерения:

$$x_i = \beta + \varepsilon_i, \quad i = 1, \ldots, N, \tag{5.1}$$

где β — истинное значение x, ε_i — случайная ошибка в i-м наблюдении. Такой набор наблюдений называется **выборкой**.

Понятно, что это — идеальная модель, которая может иметь место в естественнонаучных дисциплинах (в управляемом эксперименте). В экономике возможности измерения одной и той же величины в неизменных условиях практически отсутствуют. Определенные аналогии с этой моделью возникают в случае, когда некоторая экономическая величина измеряется разными методами (например, ВВП — по производству или по использованию), и наблюдениями выступают результаты измерения, осуществленные этими разными методами. Однако эта аналогия достаточно отдаленная, хотя бы потому, что в модели N предполагается достаточно большим, а разных методов расчета экономической величины может быть в лучшем случае два-три. Тем не менее, эта модель полезна для понимания случайных ошибок.

Если X и ε — вектор-столбцы с компонентами, соответственно, x_i и ε_i, а 1_N — N-мерный вектор-столбец, состоящий из единиц, то данную модель можно записать в матричной форме:

$$X = 1_N\beta + \varepsilon. \tag{5.2}$$

Предполагается, что ошибки по наблюдениям имеют нулевое математическое ожидание в каждом наблюдении: $\mathbf{E}(\varepsilon_i) = 0$, $i = 1, \ldots, N$; линейно не зависят друг от друга: $\mathbf{cov}(\varepsilon_i, \varepsilon_j) = 0$, $i \neq j$; а их дисперсии по наблюдениям одинаковы: $\mathbf{var}(\varepsilon_i) = \sigma^2$, $i = 1, \ldots, N$ или, в матричной форме: $\mathbf{E}(\varepsilon\varepsilon') = I_N\sigma^2$, где σ^2 — дисперсия случайных ошибок или остаточная дисперсия, I_N — единичная матрица размерности N. Это — обычные **гипотезы относительно случайных ошибок**.

Требуется найти b и e_i — оценки, соответственно, β и ε_i. Для этого используется метод наименьших квадратов (МНК), т.е. искомые оценки определяются так, чтобы $\sum\limits_{i=1}^{N}(x_i - b)^2 = \sum\limits_{i=1}^{N} e_i^2 = e'e \to \min!$, где e вектор-столбец оценок e_i.

В результате,

$$b = \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{1}{N} 1'_N X, \quad e = X - 1_N b,$$

т.к. $\dfrac{de'e}{db} = -2\sum (x_i - b) = 0$. Кроме того, $\dfrac{d^2 e'e}{db^2} = 2N > 0$, следовательно, в данной точке достигается минимум, т.е. МНК-оценкой истинного значения измеряемой величины является, как и следовало ожидать, среднее арифметическое по наблюдениям, а среднее МНК-оценок остатков равно нулю:

$$\bar{e} = \frac{1}{N} 1'_N (X - 1_N b) = \bar{x} - b = 0.$$

Оценка b относится к классу линейных, поскольку линейно зависит от наблюдений за случайной величиной.

Полученная оценка истинного значения является **несмещенной** (т.е. ее математическое ожидание равно истинному значению оцениваемого параметра), что можно легко показать.

Действительно:

$$b = \frac{1}{N}\sum x_i \overset{(5.1)}{=} \frac{1}{N}\sum (\beta + \varepsilon_i) = \beta + \frac{1}{N}\sum \varepsilon_i, \tag{5.3}$$

$$\mathbf{E}(b) \overset{\beta \text{ — детерминировано}}{=} \beta + \frac{1}{N}\sum \mathbf{E}(\varepsilon_i) \overset{\mathbf{E}(\varepsilon_i)=0}{=} \beta.$$

Что и требовалось доказать.

Однако несмещенной оценкой β является и любое наблюдение x_i, т.к. из (5.1) следует, что $\mathbf{E}(x_i) = \beta$.

Легко установить, что оценка b лучше, чем x_i, т.к. имеет меньшую дисперсию (меньшую ошибку), то есть является эффективной. Более того, b — наилучшая в этом смысле оценка во множестве всех возможных линейных несмещенных оценок. Ее дисперсия минимальна в классе линейных несмещенных оценок и определяется следующим образом:

$$\sigma_b^2 = \frac{1}{N}\sigma^2, \tag{5.4}$$

т.е. она в N раз меньше, чем дисперсия x_i, которая, как это следует из (5.1), равна σ^2.

Действительно, множество всех линейных оценок по определению представляется следующим образом:

$$b^* = \sum_{i=1}^{N} d_i x_i,$$

где d_i — любые детерминированные числа.

Из требования несмещенности,

$$\mathbf{E}\,(b^*) = \beta,$$

следует, что $\sum d_i = 1$, т.к.

$$\mathbf{E}\,(b^*) = \mathbf{E}\left(\sum d_i x_i\right) \overset{\substack{d_i \text{ — детер-}\\ \text{минировано}}}{=} \sum d_i \underbrace{\mathbf{E}\,(x_i)}_{\beta} = \beta \sum d_i.$$

Таким образом, множество всех линейных несмещенных оценок описывается так:

$$b^* = \sum_{i=1}^{N} d_i x_i, \quad \sum_{i=1}^{N} d_i = 1.$$

В этом множестве надо найти такую оценку (такие d_i), которая имеет наименьшую дисперсию,

$$b^* = \sum d_i x_i \overset{(5.1)}{=} \beta \underbrace{\sum d_i}_{=1} + \sum d_i \varepsilon_i,$$

откуда $b^* - \beta = \sum d_i \varepsilon_i$, и можно рассчитать дисперсию b^*:

$$\mathbf{var}(b^*) = \sigma^2_{b^*} = \mathbf{E}\left((b^* - \mathbf{E}(b^*)^2\right) =$$
$$= \mathbf{E}\left(\left(\sum d_i \varepsilon_i\right)^2\right) \overset{\mathbf{E}(\varepsilon_i \varepsilon_{i'})=0}{=} \sum d_i^2 \mathbf{E}(\varepsilon_i^2) \overset{\mathbf{E}\left(\varepsilon_i^2\right)=\sigma^2}{=} \sigma^2 \sum d_i^2.$$

Минимум $\sum d_i^2$ при ограничении $\sum d_i = 1$ достигается, если все d_i одинаковы и равны $\frac{1}{N}$, т.е. если $b^* = b$. Отсюда, в частности, следует, что $\sigma_b^2 = \frac{1}{N}\sigma^2$.

Что и требовалось доказать.

Такие оценки относятся к классу **BLUE** — **B**est **L**inear **U**nbiased **E**stimators.

Кроме того, оценка b **состоятельна** (стремится при $N \to \infty$ к истинному значению параметра), т.к. она несмещена и ее дисперсия, как это следует из (5.4), при $N \to \infty$ стремится к 0.

Чтобы завершить рассмотрение данного случая, осталось дать оценку остаточной дисперсии. Естественный «кандидат» на эту «роль» — дисперсия x:

$$s^2 = \frac{1}{N}\sum (x_i - b)^2 = \frac{1}{N}\sum e_i = \frac{1}{N}e'e,$$

— дает смещенную оценку. Для получения несмещенной оценки остаточной дисперсии сумму квадратов остатков надо делить не на N, а на $N-1$:

$$\hat{s}^2 = \frac{1}{N-1}e'e, \tag{5.5}$$

поскольку в векторе остатков e и, соответственно, в сумме квадратов остатков $e'e$ линейно независимых элементов только $N-1$ (т.к. $1'_N e = 0$). Этот факт можно доказать строго.

Если просуммировать по i соотношения (5.1) и поделить обе части полученного выражения на N, то окажется, что $b = \beta + \frac{1}{N}\sum \varepsilon_i$. Кроме того, известно, что $x_i = \beta + \varepsilon_i = b + e_i$. Объединяя эти два факта можно получить следующее выражение:

$$e_i = \varepsilon_i - \frac{1}{N}\sum \varepsilon_i, \tag{5.6}$$

(т.е. оценки остатков равны центрированным значениям истинных случайных ошибок), и далее получить

$$e'e = \sum \left(\varepsilon_i - \frac{1}{N}\sum \varepsilon_i\right)^2 = \sum \varepsilon_i^2 - \frac{2}{N}\left(\sum \varepsilon_i\right)^2 + \frac{1}{N}\left(\sum \varepsilon_i\right)^2 = \\ = \sum \varepsilon_i^2 - \frac{1}{N}\left(\sum \varepsilon_i\right)^2.$$

Наконец:

$$\mathbf{E}\,(e'e) \overset{\mathbf{E}\left(\varepsilon_i^2\right)=\sigma^2,\ \mathbf{E}(\varepsilon_i\varepsilon_{i'})=0}{=} (N-1)\,\sigma^2,$$

т.е.

$$\mathbf{E}\left(\frac{1}{N-1}e'e\right) = \sigma^2.$$

Что и требовалось доказать.

Теперь относительно случайных ошибок вводится дополнительное предположение: они взаимно независимы (а не только линейно независимы) и распределены нормально: $\varepsilon_i \sim NID\left(0, \sigma^2\right)$. *NID* расшифровывается как *normally and*

independently distributed (нормально и независимо распределенные случайные величины). Тогда становится известной функция плотности вероятности ε_i:

$$f(\varepsilon_i) = (2\pi)^{-\frac{1}{2}} \sigma^{-1} e^{-\frac{1}{2\sigma^2}\varepsilon_i^2} = (2\pi)^{-\frac{1}{2}} \sigma^{-1} e^{-\frac{1}{2\sigma^2}(x_i-\beta)^2},$$

и функция совместной плотности вероятности (произведение отдельных функций плотности, так как случайные ошибки по наблюдениям взаимно независимы) (см. Приложение А.3.2):

$$f(\varepsilon_1, \ldots, \varepsilon_N) = (2\pi)^{-\frac{N}{2}} \sigma^{-N} e^{-\frac{1}{2\sigma^2}\sum(x_i-\beta)^2}.$$

Эта функция рассматривается как **функция правдоподобия** $L(\sigma, \beta)$, значения которой показывают «вероятность» (правдоподобность) появления наблюдаемых x_i, $i = 1, \ldots, N$, при тех или иных значениях σ и β. Имея такую функцию, можно воспользоваться для оценки параметров σ и β **методом максимального правдоподобия** (ММП): в качестве оценок принять такие значения σ и β, которые доставляют максимум функции правдоподобия (фактически предполагая, что, раз конкретные x_i, $i = 1, \ldots, N$ реально наблюдаются, то вероятность их появления должна быть максимальной).

Обычно ищется максимум не непосредственно функции правдоподобия, а ее логарифма (значения этой функции при конкретных x_i и конечных σ положительны, и их можно логарифмировать; эта операция, естественно, не меняет точки экстремума), что проще аналитически.

$$\ln L(\sigma, \beta) = -\frac{N}{2}\ln 2\pi - N\ln\sigma - \frac{1}{2\sigma^2}\sum(x_i-\beta)^2.$$

Ищутся производные этой функции по σ и β, приравниваются нулю и определяются искомые оценки:

$$\frac{\partial \ln L}{\partial \beta} = \frac{1}{\sigma^2}\sum(x_i-\beta) = 0 \qquad \Rightarrow \quad \beta = \bar{x} = b,$$
$$\frac{\partial \ln L}{\partial \sigma} = -\frac{N}{\sigma} + \frac{1}{\sigma^3}\sum e_i^2 = 0 \qquad \Rightarrow \quad \sigma^2 = \frac{1}{N}\sum e_i^2 = s^2.$$

Это точка минимума, поскольку матрица 2-х производных

$$-\frac{N}{s^2}\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

в ней отрицательно определена.

Таким образом, ММП-оценки β и ε_i совпадают с МНК-оценками, но ММП-оценка σ^2 равна не $\hat{s}^2$, а s^2, т.е. является смещенной. Тем не менее, эта оценка состоятельна, т.к. при $N \to \infty$ различия между $\hat{s}^2$ и s^2 исчезают.

Известно, что метод максимального правдоподобия гарантирует оценкам состоятельность и **эффективность**, т.е. они обладают минимально возможными дисперсиями (вообще, а не только в классе линейных несмещенных, как оценки класса **BLUE**).

В рамках гипотезы о нормальности ошибок ε можно построить **доверительный интервал** для истинного значения параметра, т.е. интервал, в который это значение попадает с определенной вероятностью $1-\theta$, где θ — уровень ошибки (аналогичен величинам sl и pv, введенным во 2-й и 4-й главах I части книги; в прикладных исследованиях уровень ошибки принимается обычно равным 0.05). Он называется $(1-\theta)100$-процентным (например, при $\theta = 0.05$ — 95-процентным) доверительным интервалом.

Следствием нормальности ε является нормальность b: $b \sim N\left(\beta, \dfrac{\sigma^2}{N}\right)$. Поэтому

$$\frac{(b-\beta)\sqrt{N}}{\sigma} \sim N(0,1), \tag{5.7}$$

и, по определению двустороннего квантиля (см. п. 2.3),

$$\left|\frac{(b-\beta)\sqrt{N}}{\sigma}\right| \leqslant \hat{\varepsilon}_{1-\theta},$$

где $\hat{\varepsilon}_{1-\theta}$ — $(1-\theta)100$-процентный двусторонний квантиль нормального распределения.

Откуда

$$\beta \in \left[b \pm \frac{\sigma}{\sqrt{N}}\hat{\varepsilon}_{1-\theta}\right] \tag{5.8}$$

— искомый $(1-\theta)100$-процентный доверительный интервал.

К сожалению, на практике этой формулой доверительного интервала воспользоваться невозможно, т.к. она предполагает знание остаточной дисперсии σ^2. Известна же только ее оценка $\hat{s}^2$.

Простая замена в (5.8) σ на $\hat{s}$ будет приводить к систематическим ошибкам — к преуменьшению доверительного интервала, т.е. к преувеличению точности расчета.

Чтобы получить правильную формулу расчета, необходимо провести дополнительные рассуждения.

Прежде всего, доказывается, что

$$\frac{e'e}{\sigma^2} \sim \chi^2_{N-1}. \tag{5.9}$$

Справедливость этого утверждения достаточно очевидна, поскольку, как было показано выше, сумма квадратов $e'e$ имеет $N-1$ степень свободы, но может быть доказана строго.

В матричной форме выражение (5.6) записывается следующим образом:

$$e = B\varepsilon, \tag{5.10}$$

где $B = I_N - \frac{1}{N}1_N 1'_N$.

Матрица B размерности $N \times N$:

а) вещественна и симметрична ($B' = B$), поэтому она имеет N вещественных корней, которые можно «собрать» в диагональной матрице Λ, и N взаимно ортогональных вещественных собственных векторов, образующих по столбцам матрицу Y. Пусть проведена надлежащая нормировка и длины этих собственных векторов равны 1. Тогда:

$$Y'Y = I_N, \quad Y' = Y^{-1}, \quad BY = Y\Lambda, \quad B = Y\Lambda Y'; \tag{5.11}$$

б) вырождена и имеет ранг $N-1$. Действительно, имеется один и только один (с точностью до нормировки) вектор $\xi \neq 0$, который дает равенство $B\xi = 0$. Все компоненты этого единственного вектора одинаковы, т.к., как было показано выше, $B\xi$ — центрированный ξ. В частности,

$$B1_N = 0. \tag{5.12}$$

Это и означает, что ранг B равен $N-1$;

в) идемпотентна, т.е. $B^2 = B$ (см. Приложение А.3.2):

$$B^2 = \left(I_N - \frac{1}{N}1_N 1'_N\right)\left(I_N - \frac{1}{N}1_N 1'_N\right) =$$
$$= I_N - \frac{1}{N}1_N 1'_N - \underbrace{\frac{1}{N}1_N 1'_N + \frac{1}{N^2}1_N \overbrace{1'_N 1_N}^{=N} 1'_N}_{=0} = B.$$

Далее, пусть

$$u = \frac{1}{\sigma}Y'\varepsilon, \quad u_j = \frac{1}{\sigma}Y'_j\varepsilon, \tag{5.13}$$

где Y_j — j-й собственный вектор матрицы B.

Очевидно, что $\mathbf{E}(u_j) = 0$, дисперсии u_j одинаковы и равны 1:

$$\mathbf{E}\left(u_j^2\right) = \frac{1}{\sigma^2}\mathbf{E}\left(Y_j'\varepsilon\varepsilon' Y_j\right) \overset{\mathbf{E}(\varepsilon\varepsilon')=\sigma^2 I_N}{=} Y_j'Y_j \overset{(5.11)}{=} 1,$$

и u_j взаимно независимы (при $j \neq j'$):

$$\mathbf{E}\left(u_j u_{j'}\right) \overset{\text{аналогично}}{=} Y_j'Y_{j'} \overset{(5.11)}{=} 0.$$

Тогда

$$\frac{e'e}{\sigma^2} \overset{(5.10)}{=} \frac{1}{\sigma^2}\varepsilon' B'B\varepsilon \overset{B'=B,\ B^2=B}{=} \frac{1}{\sigma^2}\varepsilon' B\varepsilon \overset{(5.11)}{=} \frac{1}{\sigma^2}\varepsilon' Y\Lambda Y'\varepsilon \overset{(5.13)}{=} u'\Lambda u. \qquad (5.14)$$

Собственные числа матрицы B, как и любой другой идемпотентной матрицы, равны либо 1, либо 0 (λ — любое собственное число, ξ — соответствующий собственный вектор):

$$\begin{array}{l} B\xi = \lambda\xi, \\ B\xi = B^2\xi = B\xi\lambda = \lambda^2\xi \end{array} \quad \Rightarrow \quad \lambda^2 = \lambda \quad \Rightarrow \quad \lambda = \left\{ \begin{array}{l} 0, \\ 1. \end{array} \right.$$

и, поскольку ранг матрицы B равен $N-1$, среди ее собственных чисел имеется $N-1$, равных 1, и одно, равное 0. Поэтому (5.14), в соответствии с определением случайной величины, имеющей распределение χ^2, дает требуемый результат (см. также Приложение А.3.2).

Случайные величины, определенные соотношениями (5.7, 5.9), некоррелированы, а, следовательно, и взаимно независимы по свойствам многомерного нормального распределения (см. Приложение А.3.2).

Действительно:

$$b - \beta \overset{(5.3)}{=} \frac{1}{N}1_N'\varepsilon,$$

$$\mathbf{cov}(e,b) = \mathbf{E}(e\,(b-\beta)') \overset{(5.10)}{=} \mathbf{E}(B\varepsilon\varepsilon' 1_N \frac{1}{N}) \overset{\mathbf{E}(\varepsilon\varepsilon')=\sigma^2 I_N}{=} \frac{\sigma^2}{N}B1_N \overset{(5.12)}{=} 0.$$

Что и требовалось доказать.

Поэтому, в соответствии с определением случайной величины, имеющей t-распределение (см. также Приложение А.3.2):

$$\frac{(b-\beta)\sqrt{N}}{\sigma} \Bigg/ \sqrt{\frac{e'e}{\sigma^2}/(N-1)} \sim t_{N-1},$$

и после элементарных преобразований (сокращения σ и замены (5.5)) получается следующий результат:

$$\frac{(b-\beta)\sqrt{N}}{\hat{s}} \sim t_{N-1}.$$

Откуда:

$$\beta \in \left[b \pm \frac{\hat{s}}{\sqrt{N}} \hat{t}_{N-1,1-\theta} \right], \tag{5.15}$$

где $\hat{t}_{N-1,1-\theta}$ — $(1-\theta)100$-процентный двусторонний квантиль t_{N-1}-распределения.

Это — операциональная (допускающая расчет) форма доверительного интервала для β. Как видно, для ее получения в (5.8) надо заменить не только σ на $\hat{s}$, но и $\hat{\varepsilon}_{1-\theta}$ на $\hat{t}_{N-1,1-\theta}$. Т. к. $\hat{t}_{N-1,1-\theta} > \hat{\varepsilon}_{1-\theta}$, использование (5.8) с простой заменой σ на $\hat{s}$ действительно преуменьшает доверительный интервал (преувеличивает точность расчета). Но по мере роста N (объема информации), в соответствии со свойствами t-распределения, доверительный интервал сужается (растет точность расчета), и в пределе при $N \to \infty$ он совпадает с доверительным интервалом (5.8) (с простой заменой σ на $\hat{s}$).

Важным является **вопрос содержательной интерпретации доверительных интервалов**.

Понятно, что в рамках подхода объективной вероятности непосредственно утверждения (5.8, 5.15) не могут считаться корректными. Величина β — детерминирована и не может с какой-либо вероятностью $0 < 1-\theta < 1$ принадлежать конкретному интервалу. Она может либо принадлежать, либо не принадлежать этому интервалу, т.е. вероятность равна либо 1, либо 0. Потому в рамках этого подхода интерпретация может быть следующей: если процедуру построения доверительного интервала повторять многократно, то $(1-\theta)\cdot 100$ процентов полученных интервалов будут содержать истинное значение измеряемой величины.

Непосредственно утверждения (5.8, 5.15) справедливы в рамках подхода субъективной вероятности.

Рассмотренная модель (5.1) чрезвычайно идеализирует ситуацию: в экономике условия, в которых измеряются величины, постоянно меняются. Эти условия представляются некоторым набором факторов z_j, $j = 1, \ldots, n$, и модель «измерения» записывается следующим образом:

$$x_i = \sum_{j=1}^{n} z_{ij}\alpha_j + \beta + \varepsilon_i, \quad i = 1, \ldots, N,$$

где z_{ij} — наблюдения за значениями факторов, α_j, $j = 1, \ldots, n$, β — оцениваемые параметры.

Такая модель — это предмет регрессионного анализа. Рассмотренная же модель (5.1) является ее частным случаем: формально — при $n = 0$, по существу — при неизменных по наблюдениям значениях факторов $z_{ij} = z_j^c$ (c — **const**), так что оцениваемый в (5.1) параметр β в действительности равен $\sum z_j^c \alpha_j + \beta$.

Прежде чем переходить к изучению этой более общей модели, будут рассмотрены проблемы «распространения» ошибок первичных измерений (в этой главе) и решены алгебраические вопросы оценки параметров регрессии (следующая глава).

5.2. Производные измерения

Измеренные первично величины используются в различных расчетах (в производных измерениях), и результаты этих расчетов содержат ошибки, являющиеся следствием ошибок первичных измерений. В этом пункте изучается связь между ошибками первичных и производных измерений, или проблема «распространения» ошибок первичных измерений. Возможна и более общая трактовка проблемы: влияние ошибок в исходной информации на результаты расчетов.

Пусть x_j, $j = 1, \ldots, n$, — выборочные (фактические) значения (наблюдения, измерения) n различных случайных величин, β_j — их истинные значения, ε_j — ошибки измерений. Если x, β, ε — соответствующие n-компонентные вектор-строки, то $x = \beta + \varepsilon$. Предполагается, что $\mathbf{E}(\varepsilon) = 0$ и ковариационная матрица ошибок $\mathbf{E}(\varepsilon'\varepsilon)$ равна Ω.

Пусть величина y рассчитывается как $f(x)$. Требуется найти дисперсию σ_y^2 ошибки $\varepsilon_y = y - f(\beta)$ измерения (расчета) этой величины.

Разложение функции f в ряд Тэйлора в фактической точке x по направлению $\beta - x$ ($= -\varepsilon$), если в нем оставить только члены 1-го порядка, имеет вид: $f(\beta) = y - \varepsilon g$ (заменяя «$\approx$» на «$=$») или $\varepsilon_y = \varepsilon g$, где g — градиент f в точке x (вектор-столбец с компонентами $g_j = \frac{\partial f}{\partial x_j}(x)$).

Откуда $\mathbf{E}(\varepsilon_y) = 0$ и

$$\sigma_y^2 = \mathbf{E}\left(\varepsilon_y^2\right) = \mathbf{E}\left(g'\varepsilon'\varepsilon g\right) \overset{\mathbf{E}(\varepsilon'\varepsilon)=\Omega}{=} g'\Omega g. \tag{5.16}$$

Это — общая формула, частным случаем которой являются известные формулы для дисперсии среднего, суммы, разности, произведения, частного от деления и др.

Пусть $n = 2$, $\Omega = \begin{bmatrix} \sigma_1^2 & \omega \\ \omega & \sigma_2^2 \end{bmatrix}$.

а) если $y = x_1 \pm x_2$, то: $g = \begin{bmatrix} 1 \\ \\ \pm 1 \end{bmatrix}$, $\sigma_y^2 = \sigma_1^2 + \sigma_2^2 \pm 2\omega$.

б) если $y = x_1 x_2$, то: $g = \begin{bmatrix} x_2 \\ \\ x_1 \end{bmatrix}$, $\sigma_y^2 = x_2^2\sigma_1^2 + x_1^2\sigma_2^2 + 2x_1x_2\omega$ или $\frac{\sigma_y^2}{y^2} = \frac{\sigma_1^2}{x_1^2} +$

$+ \frac{\sigma_2^2}{x_2^2} + 2\frac{\omega}{x_1x_2}$.

в) если $y = \frac{x_1}{x_2}$, то: $g = \begin{bmatrix} \frac{1}{x_2} \\ \\ -\frac{x_1}{x_2^2} \end{bmatrix}$, $\sigma_y^2 = \frac{1}{x_2^2}\sigma_1^2 + \frac{x_1^2}{x_2^4}\sigma_2^2 - 2\frac{x_1^2}{x_2^3}\omega$ или $\frac{\sigma_y^2}{y^2} = \frac{\sigma_1^2}{x_1^2} + \frac{\sigma_2^2}{x_2^2} -$

$- 2\frac{\omega}{x_1x_2}$.

Случаи (б) и (в) можно объединить: если $y = x_1x_2^{\pm 1}$, то $\frac{\sigma_y^2}{y^2} = \frac{\sigma_1^2}{x_1^2} + \frac{\sigma_2^2}{x_2^2} \pm 2\frac{\omega}{x_1x_2}$

Можно назвать $\sigma_y, \sigma_1, \sigma_2$ абсолютными, а $\frac{\sigma_y}{y}, \frac{\sigma_1}{x_1}, \frac{\sigma_2}{x_2}$ — относительными ошибками, и, как только что показано, сделать следующие утверждения.

Если ошибки аргументов не коррелированы ($\omega = 0$), то квадрат абсолютной ошибки суммы или разности равен сумме квадратов абсолютных ошибок аргументов, а квадрат относительной ошибки произведения или частного от деления равен сумме квадратов относительных ошибок аргументов.

Если ошибки аргументов коррелированы положительно ($\omega > 0$), то ошибка суммы или произведения возрастает (предполагается, что $x_1x_2 > 0$), а разности или частного от деления — сокращается. Влияние отрицательной корреляции ошибок аргументов противоположное.

Выражение (5.4), которое фактически дает формулу ошибки среднего, также является частным случаем (5.16).

Действительно, в данном случае $y = \frac{1}{N}\sum\limits_{i=1}^{N} x_i$, $\Omega = \sigma^2 I_N$, и поскольку

$$g = \frac{1}{N}\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \text{то} \quad \sigma_y^2 = \frac{1}{N}\sigma^2.$$

В случае, если ошибки величин x_j не коррелированы друг с другом и имеют одинаковую дисперсию σ^2 ($\Omega = \sigma^2 I_n$), то

$$\sigma_y^2 = \sigma^2 g'g, \tag{5.17}$$

т.е. чем резче меняется значение функции в точке расчета, тем в большей степени ошибки исходной информации влияют на результат расчета. Возможны ситуации, когда результат расчета практически полностью определяется ошибками «на входе».

В случае, если известны дисперсии ошибок ε_j, а информация о их ковариациях отсутствует, можно воспользоваться формулой, дающей верхнюю оценку ошибки результата вычислений:

$$\sigma_y \leqslant \sum_{j=1}^{n} |\sigma_j g_j| = \Delta_y,$$

где σ_j — среднеквадратическое отклонение ε_j.

Пусть в данном случае σ — диагональная матрица $\{\sigma_j\}$, тогда $\Omega = \sigma R \sigma$, где R — корреляционная матрица ($r_{jj'} = \frac{\omega_{jj'}}{\sigma_j \sigma_{j'}}$).

Тогда (5.16) преобразуется к виду:

$$\sigma_y^2 = g' \sigma R \sigma g.$$

Пусть далее $|\sigma g|$ — вектор-столбец $\{|\sigma_j g_j|\}$, а W — диагональная матрица $\{\pm 1\}$ такая, что $\sigma g = W |\sigma g|$.

Тогда

$$\sigma_y^2 = |g' \sigma| \, WRW \, |\sigma g| \,. \tag{5.18}$$

По сравнению с R в матрице WRW лишь поменяли знаки некоторые недиагональные элементы, и поэтому все ее элементы, как и в матрице R, не превышают единицы:

$$WRW \leqslant 1_n 1_n'.$$

Умножение обеих частей этого матричного неравенства справа на вектор-столбец $|\sigma g|$ и слева на вектор строку $|g' \sigma|$ сохранит знак « $\leqslant$ », т.к. эти векторы, по определению, неотрицательны. Следовательно:

$$|g' \sigma| \, WRW \, |\sigma g| \overset{(5.18)}{=} \sigma_y^2 \leqslant |g' \sigma| \, 1_n 1_n' \, |\sigma g| = \left(\sum |\sigma_j g_j| \right)^2 .$$

Что и требовалось доказать.

5.3. Упражнения и задачи

Упражнение 1

Дана модель $x_i = \beta + \varepsilon_i = 12 + \varepsilon_i$, $i = 1, \ldots, N$. Используя нормальное распределение, в котором каждое значение ошибки ε_i независимо, имеет среднее 0 и дисперсию 2, получите 100 выборок вектора ε размерности $(N \times 1)$, $k = 1, \ldots, 100$, где $N = 10$ (в каждой выборке по 10 наблюдений). Прибавив к каждому элементу этой выборки число 12 получите 100 выборок вектора x.

1.1. Используйте 20 из 100 выборок, чтобы получить выборочную оценку b_k для β ($b_k = \frac{1}{10}\sum\limits_{i=1}^{10} x_{ik},\ k = 1, \ldots, 20$).

1.2. Вычислите среднее и дисперсию для 20 выборок оценок параметра β $\left(b = \frac{1}{20}\sum\limits_{k=1}^{20} b_k,\ s^2 = \frac{1}{20-1}\sum\limits_{k=1}^{20}(b_k - b)^2\right)$. Сравните эти средние значения с истинными параметрами.

1.3. Для каждой из 20 выборок оцените дисперсию, используя формулу

$$\hat{s}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - b)^2.$$

Пусть $\hat{s}_k^2$ — это оценка σ^2 в выборке k. Рассчитайте $\dfrac{1}{20}\sum\limits_{k=1}^{20}\hat{s}_k^2$ и сравните с истинным значением.

1.4. Объедините 20 выборок по 10 наблюдений каждая в 10 выборок по 20 наблюдений и повторите упражнение 1.1–1.3. Сделайте выводы о результатах увеличения объема выборки.

1.5. Повторите упражнение 1.1–1.3 для всех 100 и для 50 выборок и проанализируйте разницу в результатах.

1.6. Постройте распределения частот для оценок, полученных в упражнении 1.5, сравните и прокомментируйте результаты.

1.7. Постройте 95 % доверительный интервал для параметра β в каждой выборке, сначала предполагая, что σ^2 известно, а потом при условии, что истинное значение σ^2 неизвестно. Сравните результаты.

Задачи

1. При каких условиях средний за ряд лет темп инфляции будет несмещенной оценкой истинного значения темпа инфляции?

2. В каком случае средняя за ряд лет склонность населения к сбережению будет несмещенной оценкой истинного значения склонности к сбережению?

3. Пусть $x_1,\ x_2, \ldots, x_N$ — независимые случайные величины, распределенные нормально с математическим ожиданием β и дисперсией σ^2.

Пусть $b^* = \dfrac{\sum\limits_{i=1}^{N} i x_i}{\sum\limits_{i=1}^{N} i}$ — это оценка β,

- покажите, что b^* — относится к классу несмещенных линейных оценок;
- рассчитайте дисперсию b^*;
- проверьте b^* на состоятельность;
- сравните b^* с простой средней $b = \frac{1}{N}\sum\limits_{i=1}^{N} x_i$;

4. Случайная величина измерена три раза в неизменных условиях. Получены значения: 99, 100, 101. Дать оценку истинного значения этой величины и стандартную ошибку данной оценки.

5. Измерения веса студента Иванова на четырех весах дали следующие результаты: 80.5 кг, 80 кг, 78.5 кг, 81 кг. Дайте оценку веса с указанием ошибки измерения.

6. Пусть β — величина ВВП в России в 1998 г. Несколько различных экспертов рассчитали оценки ВВП x_i. Какие условия для ошибок этих оценок $x_i - \beta$ должны выполнятся, чтобы среднее x_i было несмещенной и эффективной оценкой β?

7. Проведено пять измерений некоторой величины. Результаты этих измерений следующие: 5.83, 5.87, 5.86, 5.82, 5.87 . Как бы вы оценили истинное значение этой величины при доверительной вероятности 0.95 ? А при вероятности 0.99 ?

8. Предположим, что исследователь, упоминавшийся в задаче 7, полагает, что истинное стандартное отклонение измеряемой величины равно 0.02. Сколько независимых измерений он должен сделать, чтобы получить оценку значения величины, отличающуюся от истинного значения не более чем на 0.01:

   а) при 95%-ном доверительном уровне?

   б) при 99%-ном доверительном уровне?

9. Случайная величина измерена три раза в неизменных условиях. Получена оценка истинного значения этой величины 5.0 и стандартная ошибка этой оценки $\dfrac{1}{\sqrt{3}}$. Каким мог быть исходный ряд?

10. Пусть имеется 25 наблюдений за величиной x, и по этим данным построен 95%-ный доверительный интервал для x: [1.968; 4.032]. Найдите по этим данным среднее значение и дисперсию ряда.

11. Пусть x_i — продолжительность жизни i-го человека ($i = 1, \ldots, N$), $\overline{x}$ — средняя продолжительность жизни, элементы выборки случайны и независимы. Ошибка измерения исходного показателя для всех i составляет 5%,

какова ошибка $\overline{x}$? Вывести формулу $\sigma^2_{\bar{x}}$, рассчитать коэффициент вариации для $\overline{x}$, если $x_1 = 50, x_2 = 60, x_3 = 70$.

12. Пусть объем экспорта равен 8 условных единиц, а импорта — 7 условных единиц. Показатели некоррелированы, их дисперсии одинаковы и равны 1 условной единице. На каком уровне доверия можно утверждать, что сальдо экспорта-импорта положительно?

13. Средние рентабельности двух разных фирм равны соответственно 0.4 и 0.2, стандартные отклонения одинаковы и составляют 0.2. Действительно ли первая фирма рентабельнее и почему?

14. Наблюдаемое значение некоторой величины в предыдущий и данный момент времени одинаково и равно 10. Ошибки наблюдений не коррелированы и имеют одинаковую дисперсию. Какова относительная ошибка темпа роста?

15. Пусть величина ВНП в I и II квартале составляла соответственно 550 и 560 млрд. долларов. Ошибки при расчетах ВНП в I и II квартале не коррелированы и составляют 1%. Какова относительная ошибка темпа прироста ВНП во II квартале? К каким последствиям в расчетах темпов роста и темпов прироста приведут ошибки измерения ВНП, равные 5%?

16. Стандартная ошибка измерения показателя труда и показателя капитала составляет 1%, ошибки измерений не коррелированы. Найти относительную ошибку объема продукции, рассчитанного по производственной функции Кобба—Дугласа: $Y = CK^{\alpha}L^{\beta}$.

17. Доля бюджетного дефицита в ВВП вычисляется по формуле $(R - E)/Y$, где $R = 600$ условных единиц — доходы бюджета, $E = 500$ условных единиц — расходы, $Y = 1000$ условных единиц — ВВП. Известно, что дисперсии R и E равна 100, дисперсия Y равна 25. Оценить сверху дисперсию доли дефицита.

Рекомендуемая литература

1. **Венецкий И.Г., Венецкая В.И.** Основные математико-статистические понятия и формулы в экономическом анализе. — М.: «Статистика», 1979. (Разд. 7).

2. **Езекиэл М., Фокс К.** Методы анализа корреляций и регрессий. — М.: «Статистика», 1966. (Гл. 2).

3. **Кейн Э.** Экономическая статистика и эконометрия. — М.: «Статистика», 1977. Вып. 1. (Гл. 8, 9).

4. **Моргенштерн О.** О точности экономико-статистических наблюдений. — М.: «Статистика», 1968. (Гл. 2, 6).

5. **Тинтер Г.** Введение в эконометрию. — М.: «Статистика», 1965. (Гл. 1).

6. **Frees Edward W.** Data Analysis Using Regression Models: The Business Perspective, Prentice Hall, 1996. (Ch. 2).

7. **(*) Judge G.G., Hill R.C., Griffiths W.E., Luthepohl H., Lee T.** Introduction to the Theory and Practice of Econometric. John Wiley & Sons, Inc., 1993. (Ch. 3, 5).

8. **William E.Griffiths, R. Carter Hill., George G. Judge** Learning and Practicing econometrics, N 9 John Wiley & Sons, Inc., 1993. (Ch. 14).

Глава 6

Алгебра линейной регрессии

6.1. Линейная регрессия

В этой главе предполагается, что между переменными x_j, $j = 1, \ldots, n$ существует линейная зависимость:

$$\sum_{j=1}^{n} x_j \alpha_j = \beta + \varepsilon, \tag{6.1}$$

где α_j, $j = 1, \ldots, n$, β (угловые коэффициенты и свободный член) — **параметры (коэффициенты) регрессии** (их истинные значения), ε — случайная ошибка; или в векторной форме:

$$x\alpha = \beta + \varepsilon, \tag{6.2}$$

где x и α — соответственно вектор-строка переменных и вектор-столбец параметров регрессии.

Как уже отмечалось в пункте 4.2, регрессия называется линейной, если ее уравнение линейно относительно параметров регрессии, а не переменных. Поэтому предполагается, что x_j, $j = 1, \ldots, n$, могут являться результатом каких-либо функциональных преобразований исходных значений переменных.

Для получения оценок a_j, $j = 1, \ldots, n$, b, e, соответственно, параметров регрессии α_j, $j = 1, \ldots, n$, β и случайных ошибок ε используется N наблюдений за переменными x, $i = 1, \ldots, N$, которые образуют матрицу наблюдений X

размерности $N \times n$ (столбцы — переменные, строки — наблюдения). Уравнение регрессии по наблюдениям записывается следующим образом:

$$X\alpha = 1_N\beta + \varepsilon, \tag{6.3}$$

где, как и прежде, 1_N — вектор-столбец размерности N, состоящий из единиц, ε — вектор-столбец размерности N случайных ошибок по наблюдениям; или в оценках:

$$Xa = 1_N b + e. \tag{6.4}$$

Собственно уравнение регрессии (без случайных ошибок) $x\alpha = \beta$ или $xa = b$ определяет, соответственно, истинную или расчетную **гиперплоскость** (**линию**, **плоскость**, ...) **регрессии**.

Далее применяется метод наименьших квадратов: оценки параметров регрессии находятся так, чтобы минимального значения достигла **остаточная дисперсия**:

$$s_e^2 = \frac{1}{N}e'e = \frac{1}{N}\left(a'X' - b1'_N\right)\left(Xa - 1_N b\right).$$

Из равенства нулю производной остаточной дисперсии по свободному члену b следует, что

$$\bar{x}a = b \tag{6.5}$$

и

$$1'_N e = 0. \tag{6.6}$$

Действительно,

$$\frac{\partial s_e^2}{\partial b} = -\frac{2}{N}1'_N\left(Xa - 1_N b\right) = \begin{cases} -2\left(\bar{x}a - b\right), \\ -\dfrac{2}{N}1'_N e. \end{cases}$$

Вторая производная по b равна 2, т.е. в найденной точке достигается минимум.

Здесь и ниже используются следующие правила матричной записи результатов дифференцирования линейных и квадратичных форм.

Пусть x, a — вектор-столбцы, α — скаляр, а M — симметричная матрица. Тогда:

$$\frac{dx\alpha}{d\alpha} = x, \quad \frac{\partial x'a}{\partial x} = a, \quad \frac{\partial x'M}{\partial x} = M, \quad \frac{\partial x'Mx}{\partial x} = 2Mx.$$

(См. Приложение А.2.2.)

Этот результат означает, что точка средних значений переменных лежит на расчетной гиперплоскости регрессии.

В результате подстановки выражения b из (6.5) через a в (6.4) получается другая форма записи уравнения регрессии:

$$\hat{X}a = e, \tag{6.7}$$

где $\hat{X} = X - 1_N\bar{x}$ — матрица центрированных значений наблюдений.

(6.3, 6.4) — **исходная**, (6.7) — **сокращенная запись уравнения регрессии**.

Минимизация остаточной дисперсии по a без дополнительных условий приведет к тривиальному результату: $a = 0$. Чтобы получать нетривиальные решения, на вектор параметров α и их оценок a необходимо наложить некоторые ограничения. В зависимости от формы этих ограничений возникает регрессия разного вида — **простая** или **ортогональная**.

6.2. Простая регрессия

В случае, когда ограничения на вектор a (α) имеют вид $a_j = 1$ ($\alpha_j = 1$), возникают **простые регрессии**. В таких регрессиях в левой части уравнения остается одна переменная (в данном случае j-я), а остальные переменные переносятся в правую часть, и уравнение в исходной форме приобретает вид (регрессия j-й переменной по остальным, j-я регрессия):

$$X_j = X_{-j}a_{-j} + 1_N b_j + e_j, \tag{6.8}$$

где X_j — вектор-столбец наблюдений за j-й переменной — **объясняемой**, X_{-j} — матрица наблюдений размерности $N \times (n-1)$ за остальными переменными — **объясняющими** (композиция X_j и X_{-j} образует матрицу X), a_{-j} — вектор a без j-го элемента (равного 1), взятый с обратным знаком (композиция 1 и $-a_{-j}$ образует вектор a), b_j и e_j — соответственно свободный член и вектор-столбец остатков в j-й регрессии. В сокращенной форме:

$$\hat{X}_j = \hat{X}_{-j}a_{-j} + e_j. \tag{6.9}$$

В таких регрессиях ошибки e_{ij} — расстояния от гиперплоскости регрессии до точек облака наблюдения — измеряются параллельно оси x_j.

Остаточная дисперсия приобретает следующую форму:

$$s^2_{ej} = \frac{1}{N}e'_j e_j = \frac{1}{N}\left(\hat{X}'_j - a'_{-j}\hat{X}'_{-j}\right)\left(\hat{X}_j - \hat{X}_{-j}a_{-j}\right). \tag{6.10}$$

Из равенства нулю ее производных по параметрам a_{-j} определяется, что

$$a_{-j} = M_{-j}^{-1} m_{-j}, \tag{6.11}$$

где $M_{-j} = \frac{1}{N}\hat{X}'_{-j}\hat{X}_{-j}$ — матрица ковариации объясняющих переменных x_{-j} между собой, $m_{-j} = \frac{1}{N}\hat{X}'_{-j}\hat{X}_j$ — вектор-столбец ковариации объясняющих переменных с объясняемой переменной x_j; и

$$\mathbf{cov}\,(X_{-j}, e_j) = \frac{1}{N}\hat{X}'_{-j}e_j = 0. \tag{6.12}$$

Действительно,

$$\frac{\partial s_{ej}^2}{\partial a_{-j}} = -\frac{2}{N}\hat{X}'_{-j}\left(\hat{X}_j - \hat{X}_{-j}a_{-j}\right) = \begin{cases} -2(m_{-j} - M_{-j}a_{-j}), \\ -\dfrac{2}{N}\hat{X}'_{-j}e_j. \end{cases}$$

Кроме того, очевидно, что матрица вторых производных равна $2M_{-j}$, и она, как всякая ковариационная матрица, положительно полуопределена. Следовательно, в найденной точке достигается минимум остаточной дисперсии.

Справедливость утверждения о том, что любая матрица ковариации (теоретическая или ее оценка) положительно полуопределена, а если переменные линейно независимы, то — положительно определена, можно доказать в общем случае.

Пусть x — случайный вектор-столбец с нулевым математическим ожиданием. Его теоретическая матрица ковариации по определению равна $\mathbf{E}\,(xx')$. Пусть $\xi \neq 0$ — детерминированный вектор-столбец. Квадратичная форма

$$\xi'\mathbf{E}(xx')\xi = \mathbf{E}(\xi' xx'\xi) = \mathbf{E}\left((\xi'x)^2\right) \geqslant 0,$$

т.е. матрица положительно полуопределена. Если не существует такого $\xi \neq 0$, что $\xi' x = 0$, т.е. переменные вектора x линейно не зависят друг от друга, то неравенство выполняется строго, и соответствующая матрица положительно определена.

Пусть X — матрица N наблюдений за переменными x. Оценкой матрицы ковариации этих переменных является $\frac{1}{N}\hat{X}'\hat{X}$. Квадратичная форма $\frac{1}{N}\xi'\hat{X}'\hat{X}\xi = \frac{1}{N}u'u \geqslant 0$, где $u = \hat{X}\xi$, т.е. матрица положительно полуопределена. Если не существует такого $\xi \neq 0$, что $\hat{X}\xi = 0$, т.е. переменные x линейно не зависят друг от друга, то неравенство выполняется строго, и соответствующая матрица положительно определена.

Оператор МНК-оценивания образуется соотношениями (6.11) и (6.5), которые в данном случае записываются следующим образом:

$$b_j = \bar{x}_j - \bar{x}_{-j}a_{-j} \tag{6.13}$$

(соотношения МНК-оценивания (4.37), данные в пункте 4.2 без доказательства, являются частным случаем этого оператора).

Уравнения

$$m_{-j} = M_{-j}a_{-j}, \tag{6.14}$$

решение которых дает первую часть оператора МНК-оценивания (6.11), называется **системой нормальных уравнений**.

МНК-оценки остатков имеют нулевую среднюю (6.6) и не коррелированы (ортогональны) с объясняющими переменными уравнения (6.12).

Систему нормальных уравнений можно вывести, используя иную логику. Если обе части уравнения регрессии (6.9) умножить слева на $\hat{X}'_{-j}$ и разделить на N, то получится условие $m_{-j} = M_{-j}a_{-j} + \frac{1}{N}\hat{X}'_{-j}e_j$, из которого получается искомая система при требованиях $\bar{e}_j = 0$ и $\mathbf{cov}(X_{-j}, e_j) = 0$, следующих из полученных свойств МНК-оценок остатков.

Такая же логика используется в **методе инструментальных переменных**. Пусть имеется матрица Z размерности $N \times (n-1)$ наблюдений за некоторыми величинами z, называемыми инструментальными переменными, относительно которых известно, что они линейно не зависят от ε_j и коррелированы с переменными X_{-j}. Умножение обеих частей уравнения регрессии слева на $\hat{Z}'$ и деление их на N дает условие $\frac{1}{N}\hat{Z}'\hat{X}_j = \frac{1}{N}\hat{Z}'\hat{X}_{-j}a_{-j} + \frac{1}{N}\hat{Z}'e_j$, из которого — после отбрасывания второго члена правой части в силу сделанных предположений — следует система нормальных уравнений метода инструментальных переменных:

$$m^z_{-j} = M^z_{-j}a^z_{-j}, \tag{6.15}$$

где $m^z_{-j} = \mathbf{cov}\,(z, x_j)$, $M^z_{-j} = \mathbf{cov}\,(z, x_{-j})$.

Значения j-й (объясняемой) переменной, лежащие на гиперплоскости регрессии, называются **расчетными** (по модели регрессии):

$$X^c_j = X_{-j}a_{-j} + 1_N b_j, \tag{6.16}$$

$$\hat{X}^c_j = \hat{X}_{-j}a_{-j}. \tag{6.17}$$

Их дисперсия называется **объясненной** (дисперсия, объясненная регрессией) и может быть представлена в различных вариантах:

$$s^2_{qj} = \frac{1}{N}\hat{X}'^c_j\hat{X}^c_j \overset{(6.17)}{=} a'_{-j}M_{-j}a_{-j} \overset{(6.11)}{=} a'_{-j}m_{-j} = m'_{-j}a_{-j} = m'_{-j}M^{-1}_{-j}m_{-j}. \tag{6.18}$$

Если раскрыть скобки в выражении остаточной дисперсии (6.10) и провести преобразования в соответствии с (6.11, 6.18), то получается $s_{ej}^2 = s_j^2 - s_{qj}^2$, где s_j^2 — дисперсия j-й (объясняемой) переменной, или

$$s_j^2 = s_{qj}^2 + s_{ej}^2. \tag{6.19}$$

Это — **дисперсионное тождество**, показывающее разложение общей дисперсии объясняемой переменной на две части — объясненную (регрессией) и остаточную.

Доля объясненной дисперсии в общей называется **коэффициентом детерминации**:

$$R_j^2 = \frac{s_{qj}^2}{s_j^2} = 1 - \frac{s_{ej}^2}{s_j^2}, \tag{6.20}$$

который является показателем точности аппроксимации исходных значений объясняемой переменной гиперплоскостью регрессии (объясняющими переменными). Он является квадратом **коэффициента множественной корреляции** между объясняемой и объясняющими переменными $r_{j,-j}$, который, по определению, равен коэффициенту парной корреляции между исходными и расчетными значениями объясняемой переменной:

$$r_{j,-j} = \frac{\mathbf{cov}\left(x_j, x_j^c\right)}{s_j s_{qj}} = \frac{1}{N}\frac{\hat{X}_j' \hat{X}_j^c}{s_j s_{qj}} \overset{(6.17)}{=} \frac{1}{N}\frac{\hat{X}_j' \hat{X}_{-j} a_{-j}}{s_j s_{qj}} = \\ = \frac{m_{-j}' a_{-j}}{s_j s_{qj}} \overset{(6.18)}{=} \frac{s_{qj}^2}{s_j s_{qj}} \overset{(6.20)}{=} \sqrt{R_j^2}.$$

Из (6.19) следует, что коэффициент корреляции по абсолютной величине не превышает единицы.

Эти утверждения, начиная с (6.16), обобщают положения, представленные в конце пункта 4.2.

Композиция 1 и $-a_j$ обозначается $a(j)$ и является одной из оценок вектора α. Всего таких оценок имеется n — по числу простых регрессий, в левой части уравнения которых по очереди остаются переменные x_j, $j = 1, \ldots, n$. Эти векторстолбцы образуют матрицу A. По построению ее диагональные элементы равны единице ($a_{jj} = 1$ вслед за $a_j(j) = 1$).

Все эти оценки в общем случае различны, т.е. одну из другой нельзя получить алгебраическим преобразованием соответствующих уравнений регрессии:

$$a(j) \neq \frac{1}{a_j(j')} a(j'), \quad j \neq j'. \tag{6.21}$$

Это утверждение доказывалось в пункте 4.2 при $n = 2$. В данном случае справедливо утверждение, что соотношение (6.21) может (при некоторых j, j') выполняться как равенство в том и только том случае, если среди переменных x_j, $j = 1, \ldots, n$ существуют линейно зависимые.

Достаточность этого утверждения очевидна. Действительно, пусть переменные некоторого подмножества J линейно зависимы, т.е. существует такой вектор ξ, в котором $\xi_j \neq 0$ при $j \in J$ и $\xi_j = 0$ при $j \notin J$, и $\hat{X}\xi = 0$. Тогда для любого $j \in J$ справедливо: $a(j) = \frac{1}{\xi_j}\xi$, причем $a_{j'}(j) = 0$ при $j' \notin J$, и $e_j = 0$, т.е. некоторые соотношения (6.21) выполняются как равенства.

Для доказательства необходимости утверждения предполагается, что существует такой $\xi \neq 0$, что

$$A\xi = 0 \tag{6.22}$$

(т.е., в частности, некоторые соотношения из (6.21) выполняются как равенства).

Сначала следует обратить внимание на то, что вслед за (6.14) все компоненты вектора $Ma(j)$ (M — матрица ковариации всех переменных x: $M = \frac{1}{N}\hat{X}'\hat{X}$), кроме j-й, равны нулю, а j-я компонента этого вектора в силу (6.18, 6.19) равна s^2_{ej}, т.е.

$$MA = S^2_e, \tag{6.23}$$

где S^2_e — диагональная матрица $\{s^2_{ej}\}$.

Теперь, после умножения обеих частей полученного матричного соотношения справа на вектор ξ, определенный в (6.22), получается соотношение: $0 = S^2_e\xi$, которое означает, что для всех j, таких, что $\xi_j \neq 0$, $s^2_{ej} = 0$, т.е. переменные x_j линейно зависят друг от друга.

Что и требовалось доказать.

Все возможные геометрические иллюстрации простых регрессий в пространстве наблюдений и переменных даны в пункте 4.2.

6.3. Ортогональная регрессия

В случае, когда ограничения на вектор a (или α) состоят в требовании равенства единице длины этого вектора

$$a'a = 1 \quad (\alpha'\alpha = 1), \tag{6.24}$$

и все переменные остаются в левой части уравнения, получается **ортогональная регрессия**, в которой расстояния от точек облака наблюдений до гиперплоскости регрессии измеряются перпендикулярно этой гиперплоскости. Разъяснения этому факту давались в пункте 4.2.

Оценка параметров регрессии производится из условия минимизации остаточной дисперсии:

$$s_e^2 \overset{(6.7)}{=} \frac{1}{N} a' \hat{X}' \hat{X} a = a' M a \to \min!,$$

где $M = \frac{1}{N} \hat{X}' \hat{X}$ — ковариационная матрица переменных регрессии, при условии (6.24).

Из требования равенства нулю производной по a соответствующей функции Лагранжа следует, что

$$(M - \lambda I_n)\, a = 0, \tag{6.25}$$

где λ — множитель Лагранжа ограничения (6.24), причем

$$\lambda = s_e^2. \tag{6.26}$$

Действительно, функция Лагранжа имеет вид:

$$L(a, \lambda) = a' M a - \lambda a' a,$$

а вектор ее производных по a:

$$\frac{\partial L}{\partial a} = 2\,(Ma - \lambda a)\,.$$

Откуда получается соотношение (6.25). А если обе части этого соотношения умножить слева на a' и учесть (6.24), то получается (6.26).

Таким образом, применение МНК сводится к поиску минимального собственного числа λ ковариационной матрицы M и соответствующего ему собственного (правого) вектора a (см. также Приложение А.1.2). Благодаря свойствам данной матрицы (вещественность, симметричность и положительная полуопределенность), искомые величины существуют, они вещественны, а собственное число неотрицательно (предполагается, что оно единственно). Пусть эти оценки получены.

В ортогональной регрессии все переменные x выступают объясняемыми, или моделируемыми, их расчетные значения определяются по формуле:

$$\hat{X}^c = \hat{X} - e a'. \tag{6.27}$$

Действительно: $\hat{X}^c a = \underbrace{\hat{X}a}_{e} - e\,\underbrace{a'a}_{1} = 0$, т.е. вектор-строки $\hat{x}_i^c$, соответствующие наблюдениям, лежат на гиперплоскости регрессии и являются проекциями на нее вектор-строк фактических наблюдений $\hat{x}_i$ (вектор a по построению ортогонален гиперплоскости регрессии, а $e_i a'$ — вектор нормали $\hat{x}_i^c$ на $\hat{x}_i$), а аналогом коэффициента детерминации выступает величина $1 - \dfrac{\lambda}{s_\Sigma^2}$, где $s_\Sigma^2 = \sum\limits_{j=1}^{n} s_j^2$ — суммарная дисперсия переменных x, равная следу матрицы M.

Таким образом, к n оценкам вектора a простой регрессии добавляется оценка этого вектора ортогональной регрессии, и общее количество этих оценок становится равным $n+1$.

Задачу простой и ортогональной регрессии можно записать в единой, обобщенной форме:

$$(M - \lambda W)\, a = 0, \quad a'Wa = 1, \quad \lambda \to \min!, \tag{6.28}$$

где W — диагональная $n \times n$-матрица, на диагонали которой могут стоять 0 или 1.

В случае, если в матрице W имеется единственный ненулевой элемент $w_{jj} = 1$, то это — задача простой регрессии x_j по x_{-j} (действительно, это следует из соотношения (6.23)); если W является единичной матрицей, то это — задача ортогональной регрессии. Очевидно, что возможны и все промежуточные случаи, когда некоторое количество n_1, $1 < n_1 < n$, переменных остается в левой части уравнения, а остальные n_2 переменных переносятся в правую часть уравнения регрессии:

$$\hat{X}^1 a^1 = \hat{X}^2 a^2 + e^1, \quad \left(a^1\right)' a^1 = 1.$$

Если J — множество переменных, оставленных в левой части уравнения, то в записи (6.28) такой регрессии $w_{jj} = 1$ для $j \in J$ и $w_{jj} = 0$ для остальных j. Оценка параметров регрессии производится следующим образом:

$$a^2 = M_{22}^{-1} M_{21} a^1, \quad \left(M_{11} - M_{12} M_{22}^{-1} M_{21} - \lambda I_{n_1}\right) a^1 = 0$$

(a^1 находится как правый собственный вектор, соответствующий минимальному собственному числу матрицы $M_{11} - M_{12} M_{22}^{-1} M_{21}$), где

$$\begin{aligned} M_{11} &= \frac{1}{N}\left(\hat{X}^1\right)' \hat{X}^1, \\ M_{12} = M_{21}' &= \frac{1}{N}\left(\hat{X}^1\right)' \hat{X}^2, \\ M_{22} &= \frac{1}{N}\left(\hat{X}^2\right)' \hat{X}^2 \end{aligned}$$

— соответствующие ковариационные матрицы.

Таким образом, общее количество оценок регрессии — $(2^n - 1)$. В рамках любой из этих оценок λ в (6.28) является остаточной дисперсией.

Задача ортогональной регрессии легко обобщается на случай нескольких уравнений и альтернативного представления расчетных значений изучаемых переменных.

Матрица M, как уже отмечалось, имеет n вещественных неотрицательных собственных чисел, сумма которых равна s^2_Σ, и n соответствующих им вещественных взаимноортогональных собственных векторов, дающих ортонормированный базис в пространстве наблюдений (см. также Приложение А.1.2). Пусть собственные числа, упорядоченные по возрастанию, образуют диагональную матрицу Λ, а соответствующие им собственные вектора (столбцы) — матрицу A. Тогда

$$A'A = I_n, \quad MA = A\Lambda. \tag{6.29}$$

Собственные вектора, если их рассматривать по убыванию соответствующих им собственных чисел, есть **главные компоненты** облака наблюдений, которые показывают направления наибольшей «вытянутости» (наибольшей дисперсии) этого облака. Количественную оценку степени этой «вытянутости» (дисперсии) дают соответствующие им собственные числа.

Пусть первые k собственных чисел «малы».

s^2_E — сумма этих собственных чисел;

A^E — часть матрицы A, соответствующая им (ее первые k стоблцов); это — коэффициенты по k уравнениям регрессии или k младших главных компонент;

A^Q — остальная часть матрицы A, это — $n - k$ старших главных компонент или собственно главных компонент;

$A = [A^E, A^Q]$;

$xA^E = 0$ — гиперплоскость ортогональной регрессии размерности $n - k$;

$[E, Q] = \hat{X}\left[A^E, A^Q\right]$ — координаты облака наблюдений в базисе главных компонент;

E — матрица размерности $N \times k$ остатков по уравнениям регрессии;

Q — матрица размерности $N \times (n - k)$, столбцы которой есть значения так называемых **главных факторов**.

Поскольку $A' = A^{-1}$, можно записать $\hat{X} = E\left(A^E\right)' + Q\left(A^Q\right)'$. Откуда получается два возможных представления расчетных значений переменных:

$$\hat{X}^c \underset{(6.27)}{\overset{(1)}{=}} \hat{X} - E\left(A^E\right)' \overset{(2)}{=} Q\left(A^Q\right)'. \tag{6.30}$$

Первое из них — по уравнениям ортогональной регрессии, второе (альтернативное) — по главным факторам (факторная модель).

$1 - {s_E^2}/{s_\Sigma^2}$ — аналог коэффициента детерминации, дающий оценку качества обеих этих моделей.

Факторная модель представляет n переменных через $n - k$ факторов и, тем самым, «сжимает» информацию, содержащуюся в исходных переменных. В конкретном исследовании, если k мало, то предпочтительнее использовать ортогональные регрессии, если k велико (соответственно $n - k$ мало), целесообразно применить факторную модель. При этом надо иметь в виду следующее: главные факторы — расчетные величины, и содержательная интерпретация их является, как правило, достаточно сложной задачей.

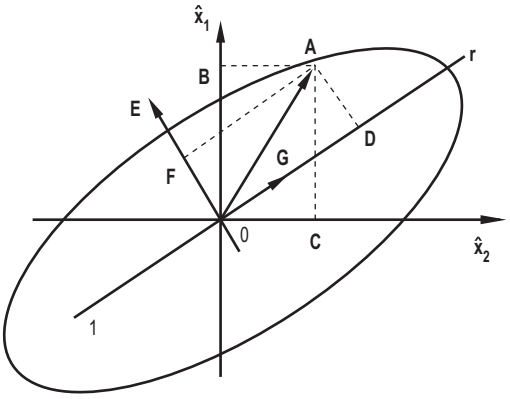

Рис. 6.1

Сделанные утверждения можно проиллюстрировать на примере $n = 2$, предполагая, что $\lambda_1 \ll \lambda_2$, и упрощая обозначения (введенные выше матрицы являются в данном случае векторами):

$a_1 = A^E$ — вектор параметров ортогональной регрессии,

$a_2 = A^Q$ — вектор первой (в данном случае — единственной) главной компоненты,

$e = E$ — остатки в уравнении ортогональной регрессии,

$q = Q$ — значения первого (в данном случае — единственного) главного фактора.

На рисунке: OA — вектор-строка i-го наблюдения $\hat{x}_i = (\hat{x}_{i1}, \hat{x}_{i2})$, OD — вектор-строка расчетных значений $\hat{x}_i^c$, длина OC — $\hat{x}_{i1}$, длина OB — $\hat{x}_{i2}$, OE — вектор-строка a_1', OG — вектор-строка a_2', длина OF — e_i, длина OD — q_i.

Как видно из рисунка 6.1, квадрат длины вектора $\hat{x}_i$ равен (из прямоугольных треугольников OAC и OAD) $\hat{x}_{i1}^2 + \hat{x}_{i2}^2 = e_i^2 + q_i^2$, и если сложить все эти уравнения по i и разделить на N, то получится $s_1^2 + s_2^2 = s_e^2 + s_q^2$. Понятно, что $s_e^2 = \lambda_1$, $s_q^2 = \lambda_2$, и это равенство означает, что след матрицы ковариации равен сумме ее собственных чисел. Кроме того, как видно из рисунка, s_1^2 показывает дисперсию облака наблюдений (суммарную дисперсию переменных регрессии) в направлении a_1 наименьшей «вытянутости» облака, s_2^2 — дисперсию облака наблюдений в направлении a_2 его наибольшей «вытянутости».

Вектор OF есть $e_i a_1'$, а вектор OD — $q_i a_2'$, и рисунок наглядно иллюстрирует выполнение соотношения (6.30):

$$\hat{x}_i^c = \hat{x}_i - e_i a_1' = q_i a_2'.$$

Пусть теперь $n = 3$, и λ_1, λ_2, λ_3, a_1, a_2, a_3 — собственные числа и вектора ковариационной матрицы переменных.

1) Если $\lambda_1 \approx \lambda_2 \approx \lambda_3$, то облако наблюдений не «растянуто» ни в одном из направлений. Зависимости между переменными отсутствуют.

2) Если $\lambda_1 \ll \lambda_2 \approx \lambda_3$ и $k = 1$, то облако наблюдений имеет форму «блина». Плоскость, в которой лежит этот «блин», является плоскостью ортогональной регрессии, которую описывает уравнение $\hat{x}a_1 = 0$, а собственно уравнением регрессии является $\hat{X}a_1 = e$.

Эту же плоскость представляют вектора a_2 и a_3, являясь ее осями координат. В этих осях координат можно выразить любую точку данной плоскости, в том числе все точки расчетных значений переменных (6.30):

$$\hat{X}^c = \begin{bmatrix} q_1 & q_2 \end{bmatrix} \begin{bmatrix} a_2' \\ a_3' \end{bmatrix} = q_1 a_2' + q_2 a_3',$$

где $q_1 = \hat{X}a_2$, $q_2 = \hat{X}a_3$ — вектора значений главных факторов или вектора координат расчетных значений переменных в осях a_2, a_3.

3) Если $\lambda_1 \approx \lambda_2 \ll \lambda_3$ и $k = 2$, то облако наблюдений имеет форму «веретена». Ось этого «веретена» является линией регрессии, образованной пересечением двух плоскостей $\hat{x}a_1 = 0$ и $\hat{x}a_2 = 0$. И уравнений ортогональной регрессии в данном случае два: $\hat{X}a_1 = e_1$ и $\hat{X}a_2 = e_2$.

Данную линию регрессии представляет вектор a_3, и через него можно выразить все расчетные значения переменных:

$$\hat{X}^c = q a_3',$$

где $q = \hat{X}a_3$ — вектор значений главного фактора.

6.4. Многообразие оценок регрессии

Множество оценок регрессии не исчерпывается $2^n - 1$ отмеченными выше элементами. Перед тем как получать любую из этих оценок, можно провести преобразование в пространстве наблюдений или переменных.

Преобразование в пространстве наблюдений проводится с помощью матрицы D размерности $N' \times N$, $N' \leqslant N$. Обе части исходного уравнения (6.3) умножаются слева на эту матрицу:

$$DX\alpha = D1_N\beta + D\varepsilon, \tag{6.31}$$

после чего проводится оценка параметров любым из указанных $2^n - 1$ способов. Понятно, что полученные оценки будут новыми, если только $D'D \neq cI_N$, где c — любая константа.

В результате такого преобразования β может перестать являться свободным членом, если только $D1_N \neq c1_{N'}$ (c — любая константа). Но, главное, меняется распределение ошибок по наблюдениям. Именно с целью изменить это распределение в нужную сторону (с помощью подбора матрицы D) и проводятся такие преобразования (см. гл. 8).

Преобразование в пространстве переменных осуществляется с помощью квадратной невырожденной матрицы C размерности $n \times n$: $Y = XC$ — преобразованные значения переменных регрессии. И затем оцениваются параметры регрессии в новом пространстве: $Yf = 1_N g + u$.

Это преобразование можно проводить в пространстве центрированных переменных, т.к. $\hat{Y} = \hat{X}C$.

Действительно: $\hat{X}C = \left(I_N - \frac{1}{N}1_N 1'_N\right) XC = \left(I_N - \frac{1}{N}1_N 1'_N\right) Y = \hat{Y}$.

То есть исходное уравнение регрессии (6.7) после преобразования приобретает вид:

$$\hat{Y}f = u. \tag{6.32}$$

Оценки f являются новыми, если после «возвращения» их в исходное пространство, которое производится умножением f слева на C, они не совпадут с оценками a, полученными в исходном пространстве, т.е. если $a \neq Cf$. Справедливость этого утверждения становится очевидной после следующего алгебраически эквивалентного преобразования исходного уравнения (6.7):

$$\underbrace{\hat{X}C}_{\hat{Y}} \underbrace{C^{-1}a}_{f} = e. \tag{6.33}$$

Понятно, что МНК-оценка f совсем не обязательно совпадет с $C^{-1}a$ — и тогда это будет новая оценка.

После преобразования меняется распределение ошибок в переменных регрессии. И именно для того, чтобы изменить это распределение в нужную сторону, осуществляются такие преобразования (см. гл. 8).

Результаты преобразований в пространстве переменных различны для простых и ортогональной регрессий.

В случае простой регрессии x_j по x_{-j} это преобразование не приводит к получению новых оценок, если j-я строка матрицы C является ортом, т.е. в объясняющие переменные правой части не «попадает» — после преобразования — объясняемая переменная.

Действительно, пусть для определенности $j = 1$ и $C = \begin{bmatrix} 1 & 0 \\ c_{-1} & C_{-1} \end{bmatrix}$ (первая строка является ортом), $C^{-1} = \begin{bmatrix} 1 & 0 \\ -C_{-1}^{-1}c_{-1} & C_{-1}^{-1} \end{bmatrix}$.

Уравнение (6.33) записывается следующим образом:

$$\underbrace{\begin{bmatrix} \hat{X}_1 + \hat{X}_{-1}c_{-1} & \hat{X}_{-1}C_{-1} \end{bmatrix}}_{\hat{Y}} \underbrace{\begin{bmatrix} 1 \\ -C_{-1}^{-1}c_{-1} - C_{-1}^{-1}a_{-1} \end{bmatrix}}_{f} = e_1$$

или, после переноса переменных в правую часть:

$$\underbrace{\left[\hat{X}_1 + \hat{X}_{-1}c_{-1}\right]}_{\hat{Y}_1} = \underbrace{\hat{X}_{-1}C_{-1}}_{\hat{Y}_{-1}} \underbrace{\left[C_{-1}^{-1}c_{-1} + C_{-1}^{-1}a_{-1}\right]}_{f_{-1}} + e_1.$$

Система нормальных уравнений для оценки f_{-1} имеет следующий вид:

$$\frac{1}{N}\underbrace{C'_{-1}\hat{X}'_{-1}}_{\hat{Y}'_{-1}}\underbrace{\left[\hat{X}_1 + \hat{X}_{-1}c_{-1}\right]}_{\hat{Y}_1} = \frac{1}{N}\underbrace{C'_{-1}\hat{X}'_{-1}}_{\hat{Y}'_{-1}}\underbrace{\hat{X}_{-1}C_{-1}}_{\hat{Y}_{-1}}\underbrace{\left[C_{-1}^{-1}c_{-1} + C_{-1}^{-1}a_{-1}\right]}_{f_{-1}}$$

или, раскрыв скобки:

$$C'_{-1}m_{-1} + C'_{-1}M_{-1}c_{-1} = C'_{-1}M_{-1}c_{-1} + C'_{-1}M_{-1}a_{-1}.$$

После взаимного сокращения одинаковых слагаемых в полученном матричном уравнении (2-го в левой части и 1-го в правой) и умножения обеих частей слева на C'^{-1}_{-1} получается система нормальных уравнений для оценки a_{-1}: $m_{-1} = M_{-1}a_{-1}$.

Это означает, что f_{-1} после «возвращения» в исходное пространство совпадает с a_{-1}, т.е. проведенное преобразование в пространстве переменных новых оценок регрессии не дает.

Верно и обратное утверждение: если j-я строка матрицы C не является ортом, то a и f совпадают с точностью до обратного преобразования только тогда, когда связь функциональна и $e = 0$.

Пусть теперь $C = \begin{bmatrix} 1 & c'_{-1} \\ 0 & I_{n-1} \end{bmatrix}$ (т.е. первая строка не является ортом), $C^{-1} = \begin{bmatrix} 1 & -c'_{-1} \\ 0 & I_{n-1} \end{bmatrix}$. Тогда уравнение (6.33) приобретает следующую форму:

$$\underbrace{\begin{bmatrix} \hat{X}_1 & \underbrace{\hat{X}_{-1} + \hat{X}_1 c'_{-1}}_{\hat{Y}_{-1}} \end{bmatrix}}_{} \underbrace{\begin{bmatrix} 1 + c'_{-1} a_{-1} \\ -a_{-1} \end{bmatrix}}_{f} = e_1, \tag{6.34}$$

или

$$\hat{X}_1 \left(1 + c'_{-1} a_{-1}\right) = \hat{Y}_{-1} a_{-1} + e_1,$$

и

$$\hat{X}_1 = \hat{Y}_{-1} \frac{a_{-1}}{\left(1 + c'_{-1} a_{-1}\right)} + \frac{1}{\left(1 + c'_{-1} a_{-1}\right)} e_1.$$

Таким образом, условием совпадения a и f с точностью до обратного преобразования является следующее:

$$f_{-1} = \frac{a_{-1}}{\left(1 + c'_{-1} a_{-1}\right)}. \tag{6.35}$$

Система нормальных уравнений для оценки f_{-1} имеет вид:

$$\frac{1}{N} \hat{Y}'_{-1} \hat{X}_1 = \frac{1}{N} \hat{Y}'_{-1} \hat{Y}_{-1} f_{-1},$$

или, учтя зависимость Y от X из (6.34) и раскрыв скобки:

$$m_{-1} + c_{-1} m_{11} = \left(M_{-1} + m_{-1} c'_{-1} + c_{-1} m'_{-1} + m_{11} c_{-1} c'_{-1}\right) f_{-1}.$$

Это равенство с учетом (6.35) и (6.11) принимает вид:

$$\left(m_{-1} + c_{-1} m_{11}\right) \left(1 + c'_{-1} M_{-1}^{-1} m_{-1}\right) = \\ = \left(M_{-1} + m_{-1} c'_{-1} + c_{-1} m'_{-1} + m_{11} c_{-1} c'_{-1}\right) M_{-1}^{-1} m_{-1}.$$

Раскрыв скобки и приведя подобные, можно получить следующее выражение:

$$c_{-1} m_{11} = c_{-1} m'_{-1} M_{-1}^{-1} m_{-1},$$

которое выполняется как равенство, только если

$$m_{11} = m'_{-1} M_{-1}^{-1} m_{-1},$$

т.е. если (в соответствии с (6.18))

$$m_{11} = s_{q1}^2.$$

Таким образом, a и f совпадают с точностью до обратного преобразования только тогда, когда полная дисперсия равна объясненной, т. е. связь функциональна и $e = 0$.

Что и требовалось доказать.

Итак, преобразования в пространстве переменных в простых регрессиях лишь в особых случаях приводят к получению новых оценок, обычно меняются только шкалы измерения. Некоторые из этих шкал находят применение в прикладном анализе. Такой пример дает **стандартизированная шкала**, которая возникает, если $C = S^{-1}$, где S — диагональная матрица среднеквадратических отклонений переменных.

Оценки параметров регрессии после преобразования оказываются измеренными в единицах среднеквадратических отклонений переменных от своих средних, и они становятся сопоставимыми между собой и с параметрами других регрессий.

В этом случае система нормальных уравнений формируется коэффициентами корреляции, а не ковариации, и $f_{-j} = R_{-j}^{-1} r_{-j}$, где R_{-j} — матрица коэффициентов корреляции объясняющих переменных между собой, r_{-j} — вектор столбец коэффициентов корреляции объясняющих переменных с объясняемой переменной.

Действительно (предполагается, что $j = 1$), соотношения (6.33) при указанной матрице C имеют следующую форму:

$$\begin{bmatrix} \underbrace{\hat{X}_1 \frac{1}{s_1}}_{\hat{Y}_1} & \underbrace{\hat{X}_{-1} S_{-1}^{-1}}_{\hat{Y}_{-1}} \end{bmatrix} \begin{bmatrix} s_1 \\ -S_{-1} a_{-1} \end{bmatrix} = e_1. \tag{6.36}$$

Для того чтобы вектор параметров приобрел необходимую для простой регрессии форму, его надо разделить на s_1. Тогда и e делится на s_1 (т.е. на s_1 делятся обе части уравнения (6.36)). После переноса объясняющих переменных в правую часть получается следующее уравнение регрессии:

$$\hat{Y}_1 = \hat{Y}_{-1} f_{-1} + \frac{1}{s_1} e_1, \quad \text{где} \quad f_{-1} = S_{-1} a_{-1} \frac{1}{s_1}.$$

Система нормальных уравнений для f_{-1} имеет следующий вид:

$$\frac{1}{N} \hat{Y}'_{-1} \hat{Y}_1 = \frac{1}{N} \hat{Y}'_{-1} \hat{Y}_{-1} f_{-1},$$

или, учитывая зависимость Y от X из (6.36),

$$\underbrace{S_{-1}^{-1} m_{-1} \frac{1}{s_1}}_{r_{-1}} = \underbrace{S_{-1}^{-1} M_{-1} S_{-1}^{-1}}_{R_{-1}} f_{-1}.$$

Что и требовалось доказать.

Преобразование в пространстве переменных в ортогональной регрессии при использовании любой квадратной и невырожденной матрицы $C \neq I_n$ приводит к получению новых оценок параметров.

В пункте 4.2 при $n = 2$ этот факт графически иллюстрировался в случае, когда

$$C = \begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix}.$$

В общем случае верно утверждение, состоящее в том, что в результате преобразования и «возвращения» в исходное пространство для получения оценок a надо решить следующую задачу:

$$(M - \lambda\Omega)\, a = 0, \quad a'\Omega a = 1, \tag{6.37}$$

где $\Omega = C'^{-1} C^{-1}$.

Действительно:

После преобразования в пространстве переменных задача ортогональной регрессии записывается следующим образом (6.24, 6.25):

$$(M_Y - \lambda I_n)\, f = 0, \quad f'f = 1, \tag{6.38}$$

где, учитывая (6.33), $M_Y = C'MC, \quad f = C^{-1}a$.

Выражение (6.37) получается в результате элементарных преобразований (6.38).

Понятно, что решение задачи (6.37) будет давать новую оценку параметрам a при любой квадратной и невырожденной матрице $C \neq I_n$. Такую регрессию иногда называют **регрессией в метрике** Ω^{-1}.

6.5. Упражнения и задачи

Упражнение 1

По наблюдениям из таблицы 6.1:

Таблица 6.1

X_1	X_{-1}	
	X_2	X_3
0.58	1.00	1.00
1.10	2.00	4.00
1.20	3.00	9.00
1.30	4.00	16.00
1.95	5.00	25.00
2.55	6.00	36.00
2.60	7.00	49.00
2.90	8.00	64.00
3.45	9.00	81.00
3.50	10.00	100.00
3.60	11.00	121.00
4.10	12.00	144.00
4.35	13.00	169.00
4.40	14.00	196.00
4.50	15.00	225.00

1.1. Вычислите
$M_{-1} = \frac{1}{N}\hat{X}'_{-1}\hat{X}_{-1}$, $m_{-1} = \frac{1}{N}\hat{X}'_{-1}\hat{X}_1$
и для регрессии $X_1 = X_{-1}a_{-1} + 1_N b_1 + e_1$ найдите оценки a_{-1} и b_1.

1.2. Рассчитайте вектор $X_1^c = X_{-1}a_{-1} + 1_N b_1$ и вектор $e_1 = X_1 - X_1^c$. Убедитесь, что $1'_N e_1 = 0$ и $\mathbf{cov}(X_{-1}, e) = \frac{1}{N}\hat{X}'_{-1}e_1 = 0$.

1.3. Вычислите объясненную дисперсию различными способами:

$$s^2_{q1} = \frac{1}{N}\hat{X}_1'^c\hat{X}_1^c;$$
$$s^2_{q1} = a'_{-1}m_{-1};$$
$$s^2_{q1} = m'_{-1}M^{-1}_{-1}m_{-1}.$$

1.4. Вычислите остаточную дисперсию различными способами:

$$s^2_{e1} = \frac{1}{N}e'_1e_1;$$
$$s^2_{e1} = s^2_1 - s^2_{q1} = \frac{1}{N}\hat{X}'_1\hat{X}_1 - s^2_{q1}.$$

1.5. Вычислите коэффициент детерминации различными способами:

$$R^2_1 = \frac{s^2_{q1}}{s^2_1};$$
$$R^2_1 = \left(\frac{\mathbf{cov}(x_1, x_1^c)}{s_1 s_{q1}}\right)^2.$$

1.6. Оцените параметры и коэффициент детерминации для ортогональной регрессии $x\alpha = \beta + \varepsilon$.

- сравните эти оценки с оценками линии регрессии, полученными в 1.1;
- рассчитайте расчетные значения переменных.

1.7. Оцените матрицу оценок и значений главных компонент (A^Q и Q), а также расчетное значение переменных.

1.8. Пусть единицы измерения x_1 увеличились в 100 раз. Как в этом случае должна выглядеть матрица преобразования D? Как изменятся оценки уравнения прямой и ортогональной регрессий?

Задачи

1. Может ли матрица

$$\text{а)}\begin{bmatrix} 9.2 & -3.8 & -2 \\ -3.8 & 2 & 0.6 \\ -2 & 0.5 & 2 \end{bmatrix} \quad \text{б)}\begin{bmatrix} 5.2 & -3.8 & -2 \\ -3.8 & 2 & 0.6 \\ -2 & 0.6 & 2 \end{bmatrix}$$

являться ковариационной матрицей переменных, для которых строится уравнение регрессии? Ответ обосновать.

2. Для $x = (x_1, x_2) = \begin{pmatrix} 1 & 1 \\ 2 & 2 \\ 6 & 3 \end{pmatrix}$ найдите оценки ковариаций переменных x, оценки параметров уравнения прямой $(x_1 = a_{12}x_2 + 1_N b_1 + e_1)$ и обратной регрессии $(x_2 = a_{21}x_1 + 1_N b_2 + e_2)$. Покажите, что $a_{12} \neq \dfrac{1}{a_{21}}$. Рассчитайте вектор-столбец остатков по прямой и обратной регрессии. Убедитесь, что сумма остатков равна нулю, вектора остатков и x_2 ортогональны при прямой регрессии, вектора остатков и x_1 ортогональны при обратной регрессии. Найдите объясненную и остаточную дисперсии различными способами, а также коэффициент детерминации.

3. Предположим, что мы, используя модель регрессии $x_1 = x_{-1}a_{-1} + 1_N b_1 + + e_1$, из условия минимизации $e_1' e_1$ получили следующую систему линейных уравнений:

$$\begin{cases} b_1 + 2a_{12} + a_{13} = 3, \\ 2b_1 + 5a_{12} + a_{13} = 9, \\ b_1 + a_{12} + 6a_{13} = -8. \end{cases}$$

Запишите условия задачи в матрично-векторной форме, решите ее, используя метод, указанный в приложении для обратных матриц, и найдите оценки параметров регрессии.

4. Оцените регрессию $x_1 = a_{12}x_2 + a_{13}x_3 + 1_N b_1 + e_1$ и рассчитайте:
   - оценку остаточной дисперсии,
   - объясненную дисперсию,
   - коэффициент детерминации,

   если

   а) матрица наблюдений имеет вид:

$$X = (X_1,\ X_2,\ X_3) = \begin{pmatrix} 5 & 1 & 3 \\ 1 & 2 & 1 \\ -2 & 3 & 5 \\ 0 & 4 & 2 \\ -4 & 5 & 4 \end{pmatrix},$$

   б) $X_1'X_1 = 96$, $X_2'X_2 = 55$, $X_3'X_3 = 129$, $X_1'X_2 = 72$, $X_1'X_3 = 107$, $X_2'X_3 = 81$, $X_1'1_N = 20$, $X_2'1_N = 15$, $X_3'1_N = 25$, $N = 5$.

5. Дисперсии двух переменных совпадают, корреляция отсутствует. Изобразить на графике — в пространстве переменных — линии прямой, обратной и ортогональной регрессий. Ответ обосновать.

6. Дисперсии выпуска продукции и количества занятых по предприятиям равны, соответственно, 10 и 20, их ковариация равна 12. Чему равен коэффициент детерминации в регрессии выпуска по занятым, коэффициент зависимости выпуска от занятых по прямой, обратной и ортогональной регрессии?

7. Дисперсии временных рядов индекса денежной массы и сводного индекса цен равны, соответственно, 150 и 200, их ковариация равна 100. Чему равен параметр влияния денежной массы на цены по модели прямой регрессии и доля объясненной дисперсии в дисперсии индекса цен?

8. По заданной матрице ковариации двух переменных $\begin{bmatrix} 14/3 & 5/3 \\ 5/3 & 2/3 \end{bmatrix}$ найти остаточную дисперсию уравнения регрессии первой переменной по второй.

9. В регрессии $x_1 = a_{12}x_2 + 1_N b_1 + e_1$, где $x_1' = (5,\ 3,\ 7,\ 1)$ коэффициент детерминации оказался равным 50%. Найдите сумму квадратов остатков.

10. Оцените модель $x_1 = a_{12}x_2 + 1_N b_1 + e_1$, используя следующие данные:

$$(x_1, x_2) = \begin{pmatrix} 3 & 3 \\ 1 & 1 \\ 8 & 5 \\ 3 & 2 \\ 5 & 5 \end{pmatrix}.$$

Вычислите остатки (e_i) и покажите, что $\sum\limits_{i=1}^{5} e_i = 0,\ \sum\limits_{i=1}^{5} x_{2i}e_i = 0$.

11. Две парные регрессии построены на одних и тех же данных: $x_1 = a_{12}x_2 + 1_N b_1 + e_1$ и $x_2 = a_{21}x_1 + 1_N b_2 + e_2$. R_1^2 — коэффициент детерминации в первой регрессии, R_2^2 — во второй. Запишите соотношение между R_1^2 и R_2^2. Ответ обосновать.

12. Возможна ли ситуация, когда угловые коэффициенты в уравнениях прямой и обратной регрессии построены на одних и тех же данных, соответственно равны 0.5 и 3.0. Почему?

13. Что геометрически означает $R^2 = 0$ и $R^2 = 1$?

14. Регрессия $x_1 = \alpha_{12}x_2 + \beta_1 + \varepsilon_1$ оценивается по двум наблюдениям. Чему равен коэффициент детерминации?

15. Для $x = (x_1,\ x_2) = \begin{pmatrix} 1 & 1 \\ 2 & 2 \\ 6 & 3 \end{pmatrix}$ оцените параметры ортогональной регрессии и коэффициент детерминации. Покажите, что линия ортогональной регрессии находится между линиями прямой и обратной регрессии.

16. Какая из двух оценок коэффициента зависимости выпуска продукции от количества занятых в производстве больше: по прямой или по ортогональной регрессии? Ответ обосновать.

17. Какая из двух оценок коэффициента зависимости спроса от цены больше: по прямой или по ортогональной регрессии? Ответ обосновать.

18. Какой вид имеет уравнение ортогональной регрессии для переменных x_1 и x_2 с одинаковыми значениями дисперсий и средних, а также имеющих положительную корреляцию равную ρ?

19. Покажите, что решение задачи

$$\left(\begin{pmatrix} m_{11} & m_{12} \\ m_{12} & m_{22} \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\right) \begin{pmatrix} 1 \\ -a_{12} \end{pmatrix} = 0, \quad \lambda \to \min!$$

эквивалентно решению задачи прямой регрессии $x_1 = a_{12}x_2 + 1_N b_1 + e_1$.

20. Пусть x_1 и x_2 — центрированные переменные. Уравнение ортогональной регрессии, построенные по множеству наблюдений над величинами x_1 и x_2, есть $x_1 - x_2 = 0$. Запишите вектор первой главной компоненты.

21. Оценка парной регрессии ведется в стандартизированной шкале. Как связан коэффициент детерминации и коэффициент регрессии (угловой)?

22. Была оценена регрессия $x_1 = \alpha_{12}x_2 + \beta_1 + \varepsilon_1$, где x_1 измеряется в рублях, а x_2 — в килограммах. Затем ту же регрессию оценили, изменив единицы измерения на тысячи рублей и тонны. Как при этом поменялись следующие величины: а) оценка коэффициента α_{12}; б) коэффициент детерминации? Как в этом случае должна выглядеть матрица преобразования D?

23. Пусть в ортогональной регрессии, построенной для переменных x_1 и x_2, из-за деноминации рубля единица измерения x_2 изменилась в 1000 раз. Как в этом случае должна выглядеть матрица преобразования D? Изменятся ли оценки? Ответ обосновать.

24. Пусть в наблюдениях задачи 2 единица измерения x_1 увеличилась в 10 раз. Как в этом случае должна выглядеть матрица преобразования D? Как изменятся оценки уравнения прямой и обратной регрессии?

25. В регрессии в метрике Ω^1 матрица Ω равна $\begin{bmatrix} 9 & 0 \\ 0 & 4 \end{bmatrix}$. Как преобразовать исходные переменные, чтобы свести эту регрессию к ортогональной?

Рекомендуемая литература

1. **Айвазян С.А.** Основы эконометрики. Т.2. — М.: «Юнити», 2001. (Гл. 2).

2. **Болч Б., Хуань К.Дж.** Многомерные статистические методы для экономики. — М.: «Статистика», 1979. (Гл. 7).

3. **Джонстон Дж.** Эконометрические методы. — М.: «Статистика», 1980. (Гл. 2, 11).

4. **Дрейпер Н., Смит Г.** Прикладной регрессионный анализ: В 2-х кн. Кн.1. — М.: «Финансы и статистика», 1986. (Гл. 1, 2).

5. **Езекиэл М., Фокс К.** Методы анализа корреляций и регрессий. — М.: «Статистика», 1966. (Гл. 5, 7).

6. **Кейн Э.** Экономическая статистика и эконометрия. Вып. 2. — М.: «Статистика», 1977. (Гл. 10, 11).

7. **Лизер С.** Эконометрические методы и задачи. — М.: «Статистика», 1971. (Гл. 2).

8. **(*) Маленво Э.** Статистические методы эконометрии. Вып. 1. — М.: «Статистика», 1975. (Гл. 1).

9. **Judge G.G., Hill R.C., Griffiths W.E., Luthepohl H., Lee T.** Introduction to the Theory and Practice of Econometric. John Wiley & Sons, Inc., 1993. (Ch. 5).

10. **William E., Griffiths R., Carter H., George G.** Judge Learning and Practicing econometrics, N 9 John Wiley & Sons, Inc., 1993. (Ch. 3).

Глава 7

Основная модель линейной регрессии

7.1. Различные формы уравнения регрессии

Основная модель линейной регрессии относится к классу простых регрессий, в левой части уравнения которых находится одна переменная: объясняемая, **моделируемая**, **эндогенная**, а в правой — несколько переменных: объясняющих, **факторных**, **независимых**, **экзогенных**. Объясняющие переменные называют также **факторами**, **регрессорами**.

Для объясняемой переменной сохраняется прежнее обозначение — x. А вектор-строка размерности n объясняющих переменных будет теперь обозначаться через z, поскольку свойства этих переменных в основной модели регрессии существенно отличаются от свойств объясняемой переменной. Через X и Z обозначаются, соответственно, вектор-столбец размерности N наблюдений за объясняемой переменной и матрица размерности $N \times n$ наблюдений за объясняющими переменными. Обозначения параметров регрессии и остатков по наблюдениям сохраняются прежними (отличие в том, что теперь вектор-столбцы α и a имеют размерность n).

Уравнения регрессии в исходной форме имеют следующий вид:

$$X = Z\alpha + 1_N\beta + \varepsilon, \tag{7.1}$$

или в оценках

$$X = Za + 1_N b + e. \tag{7.2}$$

В сокращенной форме:

$$\hat{X} = \hat{Z}\alpha + \varepsilon, \tag{7.3}$$

или

$$\hat{X} = \hat{Z}a + e. \tag{7.4}$$

Оператор МНК-оценивания ((6.11, 6.13) в п. 6.2) принимает теперь вид:

$$a = M^{-1}m, \quad b = \bar{x} - \bar{z}a, \tag{7.5}$$

где $M = {}^1\!/_N \hat{Z}'\hat{Z}$ — ковариационная матрица переменных z между собой, $m = {}^1\!/_N \hat{Z}'\hat{X}$ — вектор ковариаций переменных z с переменной x. Первую часть оператора (7.5) часто записывают в форме:

$$a = \left(\hat{Z}'\hat{Z}\right)^{-1} \hat{Z}'\hat{X}. \tag{7.6}$$

МНК-оценки e обладают следующими свойствами ((6.6, 6.12) в предыдущей главе):

$$\bar{e} = \frac{1}{N} 1'_N e = 0, \quad \mathbf{cov}\,(Z, e) = \frac{1}{N}\hat{Z}'e = 0. \tag{7.7}$$

Коэффициент детерминации рассчитывается следующим образом (см. (6.20)):

$$R^2 = \frac{s_q^2}{s_x^2} = 1 - \frac{s_e^2}{s_x^2}, \tag{7.8}$$

где s_x^2 — дисперсия объясняемой переменной, s_e^2 — остаточная дисперсия,

$$s_q^2 \overset{(6.2.6)}{=} a'Ma = a'm = m'a = m'M^{-1}m \tag{7.9}$$

— объясненная дисперсия.

Уравнение регрессии часто записывают в форме со скрытым свободным членом:

$$X = \widetilde{Z}\widetilde{\alpha} + \varepsilon, \tag{7.10}$$

$$X = \widetilde{Z}\widetilde{a} + e, \tag{7.11}$$

где $\widetilde{Z} = [Z \;\; 1_N]$, $\widetilde{\alpha} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$, $\widetilde{a} = \begin{bmatrix} a \\ b \end{bmatrix}$.

При таком представлении уравнения регрессии оператор МНК-оценивания записывается следующим, более компактным, чем (7.5), образом:

$$\widetilde{a} = \widetilde{M}^{-1}\widetilde{m}, \tag{7.12}$$

где $\widetilde{M} = \frac{1}{N}\widetilde{Z}'\widetilde{Z}$, $\widetilde{m} = \frac{1}{N}\widetilde{Z}'X$, или

$$\widetilde{a} = \left(\widetilde{Z}'\widetilde{Z}\right)^{-1}\widetilde{Z}'X. \tag{7.13}$$

$\widetilde{M}$ и $\widetilde{m}$ также, как M и m, являются матрицей и вектором вторых моментов, но не центральных, а начальных. Кроме того, их размерность на единицу больше.

Оператор (7.12) дает, естественно, такой же результат, что и оператор (7.5). Этот факт доказывался в п. 4.2 для случая одной переменной в правой части уравнения.

В общем случае этот факт доказывается следующим образом.

Учитывая, что

$$X = \hat{X} + 1_N\bar{x}, \tag{7.14}$$

$$\widetilde{Z} = \begin{bmatrix} \hat{Z} + 1_N\bar{z} & 1_N \end{bmatrix}, \tag{7.15}$$

можно установить, что

$$\widetilde{M} = \begin{bmatrix} M + \bar{z}'\bar{z} & \bar{z}' \\ \bar{z} & 1 \end{bmatrix}, \tag{7.16}$$

$$\widetilde{m} = \begin{bmatrix} m + \bar{z}'\bar{x} \\ \bar{x} \end{bmatrix},$$

и записать систему нормальных уравнений, решением которой является (7.12) в следующей форме:

$$\begin{bmatrix} M + \bar{z}'\bar{z} & \bar{z}' \\ \bar{z} & 1 \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} m + \bar{z}'\bar{x} \\ \bar{x} \end{bmatrix}.$$

Вторая (нижняя) часть этой матричной системы уравнений эквивалентна второй части оператора (7.5). После подстановки b, выраженного через a, в первую (верхнюю) часть данной матричной системы уравнений она приобретает следующий вид:

$$Ma + \bar{z}'\bar{z}a + \bar{z}'\bar{x} - \bar{z}'\bar{z}a = m + \bar{z}'\bar{x},$$

и после приведения подобных становится очевидной ее эквивалентность первой части оператора (7.5).

Что и требовалось доказать.

Кроме того, можно доказать, что

$$\widetilde{M}^{-1} = \begin{bmatrix} M^{-1} & -M^{-1}\bar{z}' \\ -\bar{z}M^{-1} & 1 + \bar{z}M^{-1}\bar{z}' \end{bmatrix}. \tag{7.17}$$

Этот факт потребуется ниже. (Правило обращения блочных матриц см. в Приложении А.1.2.)

Справедливость этого утверждения проверяется умножением $\widetilde{M}^{-1}$ из (7.17) на $\widetilde{M}$ из (7.16). В результате получается единичная матрица.

МНК-оценки вектора e ортогональны столбцам матрицы $\widetilde{Z}$:

$$\widetilde{Z}'e = 0. \tag{7.18}$$

Доказательство наличия этого свойства получается как побочный результат при выводе оператора оценивания (7.12) путем приравнивания нулю производных остаточной дисперсии по параметрам регрессии, как это делалось в п. 6.2.

Поскольку последним столбцом матрицы $\widetilde{Z}$ является 1_N, из (7.18) следует, что

$$1_N'e = 0, \tag{7.19}$$

т.е. $\bar{e} = 0$. Из остальной части (7.18):

$$Z'e = 0, \tag{7.20}$$

что в данном случае означает, что $\mathbf{cov}(Z, e) = 0$.

Действительно, раскрывая (7.20):

$$Z'e \overset{(7.15)}{=} \hat{Z}'e + \underbrace{\bar{z}'1_N'e}_{=0} = \hat{Z}'e = 0.$$

Таким образом, (7.18) эквивалентно (7.7).

Однако уравнения (7.10) допускают и иную интерпретацию. Если последним в $\widetilde{Z}$ является не 1_N, а столбец «обычной» переменной, то это — регрессия без свободного члена. В таком случае из (7.18) не следует (7.19), и свойства (7.7) не выполняются. Кроме того, для такой регрессии, очевидно, не возможна сокращенная запись уравнения. Этот случай в дальнейшем не рассматривается.

В дальнейшем будет применяться в основном форма записи уравнения со скрытым свободным членом, но чтобы не загромождать изложение материала, символ «$\sim$» будет опускаться, т.е. соотношения (7.10, 7.11, 7.12, 7.13, 7.18) будут использоваться в форме

$$X = Z\alpha + \varepsilon, \tag{7.21}$$

$$X = Za + e, \tag{7.22}$$

$$a = M^{-1}m, \tag{7.23}$$

$$a = \left(Z'Z\right)^{-1} Z'X, \tag{7.24}$$

$$Z'e = 0. \tag{7.25}$$

Случаи, когда a, Z, m, M означают не $\widetilde{a}$, $\widetilde{Z}$, $\widetilde{m}$, $\widetilde{M}$, а собственно a, Z, m, M, будут оговариваться специально.

7.2. Основные гипотезы, свойства оценок

Применение основной модели линейной регрессии корректно, если выполняются следующие гипотезы:

g1. Между переменными x и z существует линейная зависимость, и (7.10) является истинной моделью, т.е., в частности, правильно определен набор факторов z — модель верно **специфицирована**.

g2. Переменные z детерминированы, наблюдаются без ошибок и линейно независимы.

g3. $\mathbf{E}(\varepsilon) = 0$.

g4. $\mathbf{E}\left(\varepsilon\varepsilon'\right) = \sigma^2 I_N$.

Гипотеза **g2** является слишком жесткой и в экономике чаще всего нарушается. Возможности ослабления этого требования рассматриваются в следующей главе. Здесь можно заметить следующее: в тех разделах математической статистики, в которых рассматривается более общий случай, и z также случайны, предполагается, что ε не зависит от этих переменных-регрессоров.

В этих предположениях a относится к классу линейных оценок, поскольку

$$a = LX, \tag{7.26}$$

где $L \overset{(7.13)}{=} (Z'Z)^{-1} Z'$ — детерминированная матрица размерности $(n+1) \times N$, и доказывается ряд утверждений о свойствах этих МНК-оценок.

1) a — несмещенная оценка α.

Действительно:

$$a \overset{(7.26),\,\mathbf{g1}}{=} L(Z\alpha + \varepsilon) = LZ\alpha + L\varepsilon \overset{LZ=I_{n+1}}{=} \alpha + L\varepsilon \tag{7.27}$$

и

$$\mathbf{E}(a) \overset{\mathbf{g3}}{=} \alpha.$$

2) Ее матрица ковариации M_a удовлетворяет следующему соотношению:

$$M_a = \frac{1}{N}\sigma^2 M^{-1}, \tag{7.28}$$

в частности,

$$\sigma^2_{a_j} = \frac{\sigma^2}{N} m^{-1}_{jj}, \quad j = 1, \ldots, n+1 \quad (\sigma^2_{a_{n+1}} \equiv \sigma^2_b),$$

где m^{-1}_{jj} — j-й диагональный элемент матрицы M^{-1}.

Действительно:

$$M_a = \mathbf{E}((a-\alpha)(a-\alpha)') \overset{(7.27)}{=} \mathbf{E}(L\varepsilon\varepsilon'L') \overset{\mathbf{g4}}{=} \sigma^2 LL' = \sigma^2 (Z'Z)^{-1} = \frac{1}{N}\sigma^2 M^{-1}.$$

Этот результат при $n = 1$ означает, что $\sigma^2_a = \dfrac{\sigma^2}{N}\dfrac{1}{s^2_z}$, и его можно получить, используя формулу (5.17) распространения ошибок первичных измерений.

Действительно, $a = \sum d_i (x_i - \bar{x})$, где $d_i = \dfrac{z_i - \bar{z}}{\sum (z_i - \bar{z})^2}$. Тогда

$$\frac{\partial a}{\partial x_i} = \underbrace{-\frac{1}{N}\sum_{l=1}^{N} d_l}_{=0} + d_i = d_i$$

и в соответствии с указанной формулой:

$$\sigma^2_a = \sigma^2 \sum d_i^2 = \sigma^2 \frac{\sum (z_i - \bar{z})^2}{\left(\sum (z_i - \bar{z})^2\right)^2} = \frac{\sigma^2}{\sum (z_i - \bar{z})^2} = \frac{\sigma^2}{N}\frac{1}{s^2_z}.$$

Здесь важно отметить следующее.

Данная формула верна и в случае использования исходной или сокращенной записи уравнения регрессии, когда M — матрица ковариации регрессоров. Это следует из (7.17). Но в такой ситуации она (эта формула) определяет матрицу ковариации только оценок коэффициентов регрессии при объясняющих переменных, а дисперсию оценки свободного члена можно определить по формуле $\frac{\sigma^2}{N}\left(1+\bar{z}'M^{-1}\bar{z}\right)$, как это следует также из (7.17).

Следует также обратить внимание на то, что несмещенность оценок при учете только что полученной зависимости их дисперсий от N свидетельствует о состоятельности этих оценок.

Иногда формулу (7.28) используют в другой форме:

$$M_a = \sigma^2 \left(Z'Z\right)^{-1}. \tag{7.29}$$

3) Несмещенной оценкой остаточной дисперсии σ^2 является

$$\hat{s}_e^2 = \frac{N}{N-n-1}s_e^2 = \frac{1}{N-n-1}e'e. \tag{7.30}$$

Для доказательства этого факта сначала устанавливается зависимость МНК-оценок ошибок от их истинных значений, аналогично (5.10):

$$e = X - Za \overset{\mathbf{g1},\,(7.27)}{=} Z\alpha + \varepsilon - Z\left(\alpha + L\varepsilon\right) = \left(I_N - ZL\right)\varepsilon = B\varepsilon, \tag{7.31}$$

и устанавливаются свойства матрицы B (аналогично тому, как это делалось в п. 5.1)

$$B = I_N - ZL = I_N - Z\left(Z'Z\right)^{-1}Z' = I_N - \frac{1}{N}ZM^{-1}Z'. \tag{7.32}$$

Эта матрица:

а) вещественна и симметрична: $B' = B$,

б) вырождена и имеет ранг $N-n-1$, т.к. при любом $\xi \neq 0$ выполняется $BZ\xi = 0$ (поскольку $BZ \overset{(7.32)}{=} 0$), а в множестве $Z\xi$ в соответствии с **g2** имеется точно $n+1$ линейно независимых векторов,

в) идемпотентна: $B^2 = B$,

г) положительно полуопределена в силу симметричности и идемпотентности: $\xi'B\xi = \xi'B^2\xi = \xi'B'B\xi \geqslant 0$.

Теперь исследуется зависимость остаточной дисперсии от σ^2:

$$\begin{aligned} s_e^2 &= \frac{1}{N}e'e \overset{(7.31)}{=} \frac{1}{N}\varepsilon'B'B\varepsilon = \frac{1}{N}\varepsilon'B\varepsilon, \\ \mathbf{E}\left(s_e^2\right) &= \frac{1}{N}\mathbf{E}\left(\varepsilon'B\varepsilon\right) \overset{\mathbf{g4}}{=} \frac{\sigma^2}{N}\underbrace{\operatorname{tr}\left(B\right)}_{\sum b_{ii}}, \end{aligned} \tag{7.33}$$

где $\mathrm{tr}(\cdot)$ — операция следа матрицы, результатом которой является сумма ее диагональных элементов.

Далее, в силу коммутативности операции следа матрицы

$$\mathrm{tr}\,(B) = \mathrm{tr}\,(I_N) - \mathrm{tr}\,(ZL) = N - \mathrm{tr}\underbrace{(LZ)}_{I_{n+1}} = N - n - 1.$$

(См. Приложение А.1.2.)

Таким образом, $\mathbf{E}\left(s_e^2\right) = \dfrac{N-n-1}{N}\sigma^2$, и $\mathbf{E}\left(\dfrac{1}{N-n-1}e'e\right) = \sigma^2$.

Что и требовалось доказать.

Тогда оценкой матрицы ковариации M_a является (в разных вариантах расчета)

$$\frac{\hat{s}_e^2}{N}M^{-1} = \frac{e'e}{N\,(N-n-1)}M^{-1} = \frac{e'e}{N-n-1}\,(Z'Z)^{-1}, \tag{7.34}$$

и, соответственно, несмещенными оценками дисперсий (квадратов ошибок) оценок параметров регрессии:

$$\hat{s}_{a_j}^2 = \frac{e'e}{N\,(N-n-1)}m_{jj}^{-1}, \quad j = 1,\,\ldots,\,n+1 \quad (s_{a_{n+1}}^2 \equiv s_b^2). \tag{7.35}$$

4) Дисперсии a являются наименьшими в классе линейных несмещенных оценок, т.е. оценки a относятся к классу **BLUE** (см. п. 5.1). Это утверждение называется теоремой **Гаусса—Маркова**.

Доказательство этого факта будет проведено для оценки величины $c'\alpha$, где c — любой детерминированный вектор-столбец размерности $n+1$. Если в качестве c выбирать орты, данный факт будет относиться к отдельным параметрам регрессии.

МНК-оценка этой величины есть $c'a \overset{(7.26)}{=} c'LX$, она линейна, не смещена, т.к. $\mathbf{E}\,(c'a) = c'\alpha$, и ее дисперсия определяется следующим образом:

$$\mathbf{var}\,(c'a) \overset{(7.28)}{=} \frac{\sigma^2}{N}c'M^{-1}c. \tag{7.36}$$

Пусть $d'X$ — любая линейная оценка $c'\alpha$, где d — некоторый детерминированный вектор-столбец размерности N.

$$\mathbf{E}\,(d'X) \overset{\mathbf{g1}}{=} \mathbf{E}\,(d'Z\alpha + d'\varepsilon) \overset{\mathbf{g3}}{=} d'Z\alpha, \tag{7.37}$$

и для того, чтобы эта оценка была несмещенной, т.е. чтобы $d'Z\alpha = c'\alpha$, необходимо

$$d'Z = c'. \tag{7.38}$$

Из (7.37) следует, что $d'X = \mathbf{E}\,(d'X) + d'\varepsilon$, и тогда

$$\mathbf{var}\,(d'X) = \mathbf{E}((\underbrace{d'X - \mathbf{E}(d'X)}_{d'\varepsilon})^2) = \mathbf{E}\,(d'\varepsilon\varepsilon'd) \overset{\mathbf{g4}}{=} \sigma^2 d'd. \tag{7.39}$$

И, наконец, в силу положительной полуопределенности матрицы B (из (7.32)):

$$\mathbf{var}\,(d'X) - \mathbf{var}\,(c'a) \overset{(7.36,7.40)}{=} \sigma^2 d'd - \frac{\sigma^2}{N} c'M^{-1}c \overset{(7.38)}{=}$$
$$= \sigma^2 d' \left(I_N - \frac{1}{N} Z M^{-1} Z'\right) d \overset{(7.32)}{=} \sigma^2 d' B d \geqslant 0,$$

т.е. дисперсия МНК-оценки меньше либо равна дисперсии любой другой оценки в классе линейных несмещенных.

Что и требовалось доказать.

Теперь вводится еще одна гипотеза:

g5. Ошибки ε имеют многомерное нормальное распределение:

$$\varepsilon \sim N\left(0, \sigma^2 I_N\right).$$

(Поскольку по предположению **g4** они некоррелированы, то по свойству многомерного нормального распределения они независимы).

Тогда оценки a будут также иметь нормальное распределение:

$$a \sim N\,(\alpha, M_a)\,, \tag{7.40}$$

в частности,

$$a_j \sim N\left(\alpha_j, \sigma^2_{a_j}\right), \quad j = 1, \ldots, n+1 \quad (a_{n+1} \equiv b,\ \alpha_{n+1} \equiv \beta),$$

они совпадут с оценками максимального правдоподобия, что гарантирует их состоятельность и эффективность (а не только эффективность в классе линейных несмещенных оценок).

Применение метода максимального правдоподобия в линейной регрессии рассматривается в IV-й части книги. Здесь внимание сосредоточивается на других важных следствиях нормальности ошибок.

Поскольку

$$\frac{a_j - \alpha_j}{\sigma_{a_j}} \overset{N}{(0,1)}, \tag{7.41}$$

для α_j можно построить $(1-\theta)100$-процентный доверительный интервал:

$$\alpha_j \in \left[a_j \pm \sigma_{a_j} \hat{\varepsilon}_{1-\theta}\right]. \tag{7.42}$$

Чтобы воспользоваться этой формулой, необходимо знать истинное значение остаточной дисперсии σ^2, но известна только ее оценка. Для получения соответствующей формулы в операциональной форме, как и в п. 5.1, проводятся следующие действия.

Сначала доказывается, что

$$\frac{e'e}{\sigma^2} \sim \chi^2_{N-n-1}. \tag{7.43}$$

Это доказательство проводится так же, как и в пункте 5.1 для (5.9). Только теперь матрица B, связывающая в (7.31) оценки ошибок с их истинными значениями, имеет ранг $N-n-1$ (см. свойства матрицы B, следующие из (7.32)), а не $N-1$, как аналогичная матрица в (5.10).

Затем обращается внимание на то, что e и a не коррелированы, а значит, не коррелированы случайные величины в (7.41, 7.43).

Действительно (как и в 5.1):

$$a - \alpha \overset{(7.27)}{=} L\varepsilon$$

и

$$\mathbf{cov}\,(a,e) = \mathbf{E}\,((a-\alpha)e') \overset{(7.31)}{=} \mathbf{E}\,(L\varepsilon\varepsilon' B) \overset{\mathbf{g4}}{=} \sigma^2\,(Z'Z)^{-1}\underbrace{Z'B}_{=0} = 0.$$

Что и требовалось доказать.

Поэтому по определению случайной величины, имеющей t-распределение:

$$\frac{(a_j-\alpha_j)\sqrt{N}}{\sigma\sqrt{m_{jj}^{-1}}} \Big/ \sqrt{\frac{e'e}{\sigma^2}/(N-n-1)} \overset{(7.35)}{=} \frac{a_j-\alpha_j}{\hat{s}_{a_j}} \sim t_{N-n-1}. \tag{7.44}$$

Таким образом, для получения операциональной формы доверительного интервала в (7.42) необходимо заменить σ_{a_j} на $\hat{s}_{a_j}$ и $\hat{\varepsilon}_{1-\theta}$ на $\hat{t}_{N-n-1,1-\theta}$:

$$\alpha_j \in \left[a_j \pm \hat{s}_{a_j}\hat{t}_{N-n-1,1-\theta}\right]. \tag{7.45}$$

Полезно заметить, что данный в этом пункте материал обобщает результаты, полученные в п. 5.1. Так, многие приведенные здесь формулы при $n=0$ преобразуются в соответствующие формулы п. 5.1. Полученные результаты можно использовать также и для проверки гипотезы о том, что $\alpha_j = 0$ (нулевая гипотеза).

Рассчитывается t-статистика

$$t_j^c = \frac{a_j}{\hat{s}_{a_j}}, \tag{7.46}$$

которая в рамках нулевой гипотезы, как это следует из (7.44), имеет t-распределение.

Проверка нулевой гипотезы осуществляется по схеме, неоднократно применяемой в I части книги. В частности, если уровень значимости t-статистики sl (напоминание: sl таково, что $t_j^c = t_{N-n-1,sl}$) не превышает θ (обычно 0.05), то нулевая гипотеза отвергается с ошибкой (1-го рода) θ и принимается, что $\alpha_j \neq 0$. В противном случае, если нулевую гипотезу не удалось отвергнуть, считается, что j-й фактор не значим, и его не следует вводить в модель.

Операции построения доверительного интервала и проверки нулевой гипотезы в данном случае в определенном смысле эквивалентны. Так, если построенный доверительный интервал содержит нуль, то нулевая гипотеза не отвергается, и наоборот.

Гипотеза о нормальности ошибок позволяет проверить еще один тип нулевой гипотезы: $\alpha_j = 0,\ j = 1, \ldots, n$, т.е. гипотезы о том, что модель некорректна и все факторы введены в нее ошибочно.

При построении критерия проверки данной гипотезы уравнение регрессии используется в сокращенной форме, и условие (7.40) записывается в следующей форме:

$$a \sim N\left(\alpha, \frac{\sigma^2}{N} M^{-1}\right), \tag{7.47}$$

где a и α — вектора коэффициентов при факторных переменных размерности n, M — матрица ковариации факторных переменных. Тогда

$$\frac{N}{\sigma^2}\left(a' - \alpha'\right) M \left(a - \alpha\right) \sim \chi_n^2. \tag{7.48}$$

Действительно:

Матрица M^{-1} вслед за M является вещественной, симметричной и положительно полуопределенной, поэтому ее всегда можно представить в виде:

$$M^{-1} = CC', \tag{7.49}$$

где C — квадратная неособенная матрица.

Чтобы убедиться в этом, достаточно вспомнить (6.29) и записать аналогичные соотношения: $M^{-1}Y = Y\Lambda,\ Y'Y = YY' = I_n,\ \Lambda \geqslant 0$, где Y — матрица, столбцы

которой есть собственные вектора M^{-1}, Λ — диагональная матрица соответствующих собственных чисел. Тогда

$$M^{-1} = Y\Lambda Y' = \underbrace{Y\Lambda^{0.5}}_{C}\underbrace{\Lambda^{0.5}Y'}_{C'}$$

(см. Приложение А.1.2).

Вектор случайных величин $u = \dfrac{\sqrt{N}}{\sigma}\, C^{-1}(a-\alpha)$ обладает следующими свойствами: по построению $\mathbf{E}(u) = 0$, и в силу того, что

$$\mathbf{E}\left((a-\alpha)(a-\alpha)'\right) \overset{(7.47)}{=} \frac{\sigma^2}{N} M^{-1},$$

$$\mathbf{cov}(u) = \mathbf{E}\left(uu'\right) = \frac{N}{\sigma^2} C^{-1}\mathbf{E}\left((a-\alpha)(a-\alpha)'\right) C'^{-1} = C^{-1}M^{-1}C'^{-1} \overset{(7.49)}{=} I_n.$$

Следовательно, по определению χ^2 случайная величина

$$u'u = \frac{N}{\sigma^2}\left(a'-\alpha'\right) \underbrace{C'^{-1}C^{-1}}_{M}(a-\alpha)$$

имеет указанное распределение (см. Приложение А.3.2).

Как было показано выше, e и a не коррелированы, поэтому не коррелированы случайные величины, определенные в (7.43, 7.48), и в соответствии с определением случайной величины, имеющей F-распределение:

$$\frac{N}{\sigma^2}\left(a'-\alpha'\right) M \left(a-\alpha\right)(N-n-1) \Big/ \frac{e'e}{\sigma^2}\, n \sim F_{n,\,N-n-1}.$$

Отсюда следует, что при нулевой гипотезе $\alpha = 0$

$$\frac{a'Ma\,(N-n-1)}{(e'e)/_N\, n} \overset{(7.9)}{=} \frac{s_q^2\,(N-n-1)}{s_e^2 n} \sim F_{n,\,N-n-1},$$

или

$$\frac{R^2\,(N-n-1)}{(1-R^2)\,n} = F^c \sim F_{n,\,N-n-1}. \tag{7.50}$$

Сама проверка нулевой гипотезы проводится по обычной схеме. Так, если значение вероятности pv статистики F^c (величина, аналогичная sl для t-статистики) не превышает θ (например, 0.05), нулевая гипотеза отвергается с вероятностью ошибки θ, и модель считается корректной. В противном случае нулевая гипотеза не отвергается, и модель следует пересмотреть.

7.3. Независимые факторы: спецификация модели

В этом пункте используется модель линейной регрессии в сокращенной форме, поэтому переменные берутся в центрированной форме, а m и M — вектор и матрица соответствующих коэффициентов ковариации переменных.

Под спецификацией модели в данном случае понимается процесс и результат определения набора независимых факторов. При построении эконометрической модели этот набор должен обосновываться экономической теорией. Но это удается не во всех случаях. Во-первых, не все факторы, важные с теоретической точки зрения, удается количественно выразить. Во-вторых, эмпирический анализ часто предшествует попыткам построения теоретической модели, и этот набор просто неизвестен. Потому важную роль играют и методы формального отбора факторов, также рассматриваемые в этом пункте.

В соответствии с гипотезой **g2** факторные переменные не должны быть линейно зависимыми. Иначе матрица M в операторе МНК-оценивания будет необратима. Тогда оценки МНК по формуле $a = M^{-1}m$ невозможно будет рассчитать, но их можно найти, решая систему нормальных уравнений (6.14):

$$Ma = m.$$

Решений такой системы нормальных уравнений (в случае необратимости матрицы M) будет бесконечно много. Следовательно, оценки нельзя найти однозначно, т.е. уравнение регрессии невозможно идентифицировать. Действительно, пусть оценено уравнение

$$\hat{x} = \hat{z}_1 a_1 + e, \tag{7.51}$$

где $\hat{z}_1$ — вектор-строка факторных переменных размерности n_1, a_1 — вектор-столбец соответствующих коэффициентов регрессии, и пусть в это уравнение вводится дополнительный фактор $\hat{z}_2$, линейно зависимый от $\hat{z}_1$, т.е. $\hat{z}_2 = \hat{z}_1 c_{21}$.

Тогда оценка нового уравнения

$$\hat{x} = \hat{z}_1 a_1^* + \hat{z}_2 a_2 + e^* \tag{7.52}$$

(«звездочкой» помечены новые оценки «старых» величин) эквивалентна оценке уравнения $\hat{x} = \hat{z}_1 (a_1^* + a_2 c_{21}) + e^*$. Очевидно, что $a_1 = a_1^* + a_2 c_{21}$, $e = e^*$, и, произвольно задавая a_2, можно получать множество новых оценок $a_1^* = a_1 - a_2 c_{21}$. Логичнее всего положить $a_2 = 0$, т.е. не вводить фактор $\hat{z}_2$. Хотя, если из содержательных соображений этот фактор следует все-таки ввести, то тогда надо исключить из уравнения какой-либо ранее введенный фактор, входящий в $\hat{z}_1$. Таким образом, вводить в модель факторы, линейно зависимые от уже введенных, бессмысленно.

Случаи, когда на факторных переменных существуют точные линейные зависимости, встречаются редко. Гораздо более распространена ситуация, в которой зависимости между факторными переменными приближаются к линейным. Такая ситуация называется **мультиколлинеарностью**. Она чревата высокими ошибками получаемых оценок и высокой чувствительностью результатов оценивания к ошибкам в факторных переменных, которые, несмотря на гипотезу **g2**, обычно присутствуют в эмпирическом анализе.

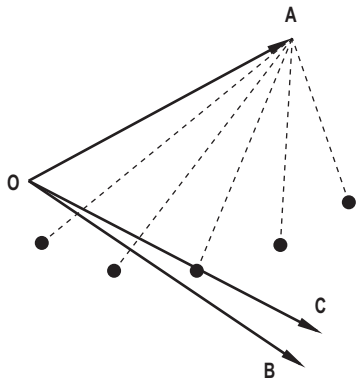

Рис. 7.1

Действительно, в такой ситуации матрица M плохо обусловлена и диагональные элементы M^{-1}, определяющие дисперсии оценок, могут принимать очень большие значения. Кроме того, даже небольшие изменения в M, связанные с ошибками в факторных переменных, могут повлечь существенные изменения в M^{-1} и, как следствие, — в оценках a.

Последнее наглядно иллюстрируется рисунком (рис. 7.1) в пространстве наблюдений при $n = 2$.

На этом рисунке: OA — $\hat{x}$, OB — $\hat{z}_1$, OC — $\hat{z}_2$.

Видно, что факторные переменные сильно коррелированы (угол между соответствующими векторами мал).

Поэтому даже небольшие колебания этих векторов, связанные с ошибками, значительно меняют положение плоскости, которую они определяют, и, соответственно, — нормали на эту плоскость.

Из рисунка видно, что оценки параметров регрессии «с легкостью» меняют не только свою величину, но и знак.

По этим причинам стараются избегать ситуации мультиколлинеарности. Для этого в уравнение регрессии не включают факторы, сильно коррелированные с другими.

Можно попытаться определить такие факторы, анализируя матрицу коэффициентов корреляции факторных переменных $S^{-1}MS^{-1}$, где S — диагональная матрица среднеквадратических отклонений. Если коэффициент $s_{jj'}$ этой матрицы достаточно большой, например, выше 0.75, то один из пары факторов j и j' не следует вводить в уравнение. Однако такого элементарного «парного» анализа может оказаться не достаточно. Надежнее построить все регрессии на множестве факторных переменных, последовательно оставляя в левой части уравнения эти переменные по отдельности. И не вводить в уравнение специфицируемой модели (с x в левой части) те факторы, уравнения регрессии для которых достаточно значимы по F-критерию (например, значение pv не превышает 0.05).

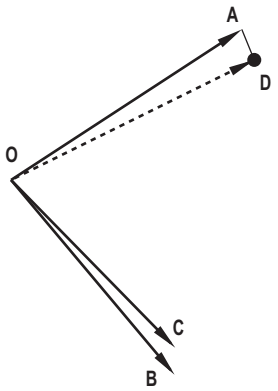

Рис. 7.2

Однако в эмпирических исследованиях могут возникать ситуации, когда только введение сильно коррелированных факторов может привести к построению значимой модели.

Это утверждение можно проиллюстрировать рисунком (рис. 7.2) в пространстве наблюдений при $n = 2$.

На этом рисунке: OA — $\hat{x}$, OB — $\hat{z}_1$, OC — $\hat{z}_2$, AD — нормаль на плоскость, определяемую векторами OB и OC, OD — проекция OA на эту плоскость.

Из рисунка видно, что $\hat{z}_1$ и $\hat{z}_2$ по отдельности не объясняют $\hat{x}$ (углы между соответствующими векторами близки к $90°$), но вместе они определяют плоскость, угол между которой и вектором OA очень мал, т.е. коэффициент детерминации в регрессии $\hat{x}$ на $\hat{z}_1$, $\hat{z}_2$ близок к единице.

Рисунок также показывает, что такая ситуация возможна только если факторы сильно коррелированы.

В таких случаях особое внимание должно уделяться точности измерения факторов.

Далее определяются последствия введения в уравнение дополнительного фактора. Для этого сравниваются оценки уравнений (7.51, 7.52) в предположении, что $\hat{z}_2$ линейно независим от $\hat{z}_1$.

В этом анализе доказываются два утверждения.

1) Введение дополнительного фактора не может привести к сокращению коэффициента детерминации, в большинстве случаев он растет (растет объясненная дисперсия). Коэффициент детерминации остается неизменным тогда и только тогда, когда вводимый фактор ортогонален остаткам в исходной регрессии (линейно независим от остатков), т.е. когда

$$m_{2e} = \frac{1}{N}\hat{Z}_2' e = 0 \tag{7.53}$$

(понятно, что коэффициент детерминации не меняется и в случае линейной зависимости $\hat{z}_2$ от $\hat{z}_1$, но такой случай исключен сделанным предположением о линейной независимости этих факторов; в дальнейшем это напоминание не делается).

Для доказательства этого факта проводятся следующие действия.

Записываются системы нормальных уравнений для оценки регрессий (7.51, 7.52):

$$m_1 = M_{11} a_1, \tag{7.54}$$

$$\begin{bmatrix} m_1 \\ m_2 \end{bmatrix} = \begin{bmatrix} M_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \begin{bmatrix} a_1^* \\ a_2 \end{bmatrix}, \tag{7.55}$$

где $m_1 = \frac{1}{N}\hat{Z}_1'\hat{X}$, $m_2 = \frac{1}{N}\hat{Z}_2'\hat{X}$, $M_{11} = \frac{1}{N}\hat{Z}_1'\hat{Z}_1$, $m_{12} = m_{21}' = \frac{1}{N}\hat{Z}_1'\hat{Z}_2$, $m_{22} = \frac{1}{N}\hat{Z}_2'\hat{Z}_2$.

Далее, с помощью умножения обеих частей уравнения (7.51), расписанного по наблюдениям, слева на $\frac{1}{N}\hat{Z}_2'$, устанавливается, что

$$m_2 - m_{21}a_1 \overset{(7.53)}{=} m_{2e}, \tag{7.56}$$

а из регрессии $\hat{Z}_2 = \hat{Z}_1 a_{21} + e_{21}$, в которой по предположению $e_{21} \neq 0$, находится остаточная дисперсия:

$$s_{e21}^2 = \frac{1}{N}e_{21}'e_{21} \overset{(7.9)}{=} m_{22} - m_{21}M_{11}^{-1}m_{12} > 0. \tag{7.57}$$

Из первой (верхней) части системы уравнений (7.55) определяется:

$$M_{11}a_1^* + m_{12}a_2 = m_1 \overset{(7.54)}{=} M_{11}a_1,$$

и далее

$$a_1^* = a_1 - M_{11}^{-1}m_{12}a_2. \tag{7.58}$$

Из второй (нижней) части системы уравнений (7.55) определяется:

$$m_{22}a_2 = m_2 - m_{21}a_1^* \overset{(7.58)}{=} m_2 - m_{21}\left(a_1 - M_{11}^{-1}m_{12}a_2\right).$$

Откуда

$$\left(m_{22} - m_{21}M_{11}^{-1}m_{12}\right)a_2 = m_2 - m_{21}a_1$$

и, учитывая (7.56, 7.57),

$$s_{e21}^2 a_2 = m_{2e}. \tag{7.59}$$

Наконец, определяется объясненная дисперсия после введения дополнительного фактора:

$$s_q^{2*} \overset{(7.9)}{=} m_1'a_1^* + m_2a_2 \overset{(7.58)}{=} \underbrace{m_1'a_1}_{s_q^2} + \left(m_2 - \underbrace{m_1'M_{11}^{-1}}_{a_1'}\, m_{12}\right)a_2 \overset{(7.56)}{=} s_q^2 + m_{2e}a_2, \tag{7.60}$$

т.е.

$$s_q^{2*} \overset{(7.59)}{=} s_q^2 + \frac{m_{2e}^2}{s_{e21}^2}.$$

Что и требовалось доказать.

Это утверждение легко проиллюстрировать рисунком 7.3 в пространстве наблюдений при $n_1 = 1$.

На этом рисунке: OA — $\hat{x}$, OB — $\hat{z}_1$, OC — $\hat{z}_2$, AD — нормаль $\hat{x}$ на $\hat{z}_1$ (DA — вектор e).

Рисунок показывает, что если $\hat{z}_2$ ортогонален e, то нормаль $\hat{x}$ на плоскость, определяемую $\hat{z}_1$ и $\hat{z}_2$, совпадает с AD, т.е. угол между этой плоскостью и $\hat{x}$ совпадает с углом между $\hat{x}$ и $\hat{z}_1$, введение в уравнение нового фактора $\hat{z}_2$ не меняет коэффициент детерминации. Понятно также и то, что во всех остальных случаях (когда $\hat{z}_2$ не ортогонален e) этот угол уменьшается и коэффициент детерминации растет.

После введения дополнительного фактора $\hat{z}_2$ в уравнение максимально коэффициент детерминации может увеличиться до единицы. Это произойдет, если $\hat{z}_2$ является линейной комбинацией $\hat{x}$ и $\hat{z}_1$.

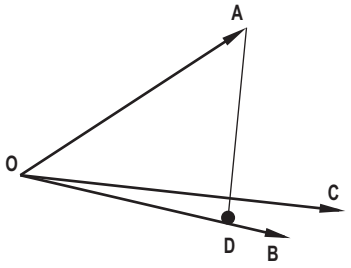

Рис. 7.3

Рост коэффициента детерминации с увеличением количества факторов — свойство коэффициента детерминации, существенно снижающее его содержательное (статистическое) значение. Введение дополнительных факторов, даже если они по существу не влияют на моделируемую переменную, приводит к росту этого коэффициента. И, если таких факторов введено достаточно много, то он начнет приближаться к единице. Он обязательно достигнет единицы при $n = N - 1$. Более приемлем в роли критерия качества коэффициент детерминации, скорректированный на число степеней свободы:

$$\tilde{R}^2 = 1 - \left(1 - R^2\right) \frac{N-1}{N-n-1}$$

($1 - R^2$ — отношение остаточной дисперсии к объясненной, которые имеют, соответственно, $N - n - 1$ и $N - 1$ степеней свободы), этот коэффициент может снизиться после введения дополнительного фактора. Однако наиболее правильно при оценке качества уравнения ориентироваться на показатель pv статистики F^c.

Скорректированный коэффициент детерминации построен так, что он, так сказать, штрафует за то, что в модели используется слишком большой набор факторов. На этом же принципе построено и большинство других критериев, используемых

для выбора модели: на них положительно отражается уменьшение остаточной дисперсии $s_e^2(z_1)$ (здесь имеется в виду смещенная оценка дисперсии из регрессии по z_1) и отрицательно — количество включенных факторов n_1 (без константы). Укажем только три наиболее известных критерия (из огромного числа предложенных в литературе):

Критерий Маллоуза:

$$C_p = s_e^2(z_1) + \frac{2(n_1+1)}{N}\hat{s}_e^2(z),$$

где $\hat{s}_e^2(z)$ — несмещенная оценка дисперсии в регрессии с полным набором факторов.

Информационный критерий Акаике:

$$AIC = \ln\left(2\pi s_e^2(z_1)\right) + \frac{2(n_1+1)}{N}.$$

Байесовский информационный критерий (критерий Шварца):

$$BIC = \ln\left(2\pi s_e^2(z_1)\right) + \frac{\ln(N)(n_1+1)}{N}.$$

В тех же обозначениях скорректированный коэффициент детерминации имеет вид

$$\tilde{R}^2 = 1 - \frac{s_e^2(z_1)}{s_e^2(\emptyset)}\frac{N-1}{N-n_1-1},$$

где $s_e^2(\emptyset)$ — остаточная дисперсия из регрессии с одной константой.

Регрессия тем лучше, чем ниже показатель C_p (AIC, BIC). Для $\tilde{R}^2$ используется противоположное правило — его следует максимизировать. Вместо $\tilde{R}^2$ при неизменном количестве наблюдений N можно использовать несмещенную остаточную дисперсию $\hat{s}_e^2 = \hat{s}_e^2(z_1)$, которую уже следует минимизировать.

В идеале выбор модели должен происходить при помощи полного перебора возможных регрессий. А именно, берутся все возможные подмножества факторов z_1, для каждого из них оценивается регрессия и вычисляется критерий, а затем выбирается набор z_1, дающий наилучшее значение используемого критерия.

Чем отличается поведение критериев $\tilde{R}^2$ ($\hat{s}_e^2$), C_p, AIC, BIC при выборе модели? Прежде всего, они отличаются по степени жесткости, то есть по тому, насколько велик штраф за большое количество факторов и насколько более «экономную» модель они имеют тенденцию предлагать. $\tilde{R}^2$ является наиболее мягким критерием. Критерии C_p и AIC занимают промежуточное положение; при больших N они ведут себя очень похоже, но C_p несколько жестче AIC, особенно при малых N. BIC является наиболее жестким критерием, причем, как можно увидеть из приведенной формулы, в отличие от остальных критериев его жесткость возрастает с ростом N.

Различие в жесткости проистекает из различия в целях. Критерии C_p и AIC направлены на достижение высокой точности прогноза: C_p направлен на минимизацию дисперсии ошибки прогноза (о ней речь пойдет в следующем параграфе),

а AIC — на минимизацию расхождения между плотностью распределения по истинной модели и по выбранной модели. В основе BIC лежит цель максимизации вероятности выбора истинной модели.

2) Оценки коэффициентов регрессии при факторах, ранее введенных в уравнение, как правило, меняются после введения дополнительного фактора. Они остаются прежними в двух и только двух случаях: а) если неизменным остается коэффициент детерминации и выполняется условие (7.53) (в этом случае уравнение в целом остается прежним, т.к. $a_2 = 0$); б) если новый фактор ортогонален старым ($\hat{z}_1$ и $\hat{z}_2$ линейно не зависят друг от друга), т.е.

$$m_{12} = \frac{1}{N}\hat{Z}_1'\hat{Z}_2 = 0 \qquad (7.61)$$

(в этом случае объясненная дисперсия равна сумме дисперсий, объясненных факторами $\hat{z}_1$ и $\hat{z}_2$ по отдельности).

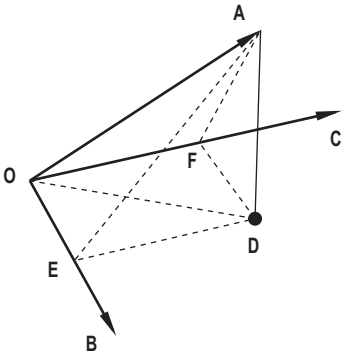

Рис. 7.4

Действительно, в соотношении (7.58) $M_{11}^{-1}m_{12}$ не может равняться нулю при $m_{12} \neq 0$, т.к. M_{11} невырожденная матрица. Поэтому из данного соотношения следует, что оценки a_1 не меняются, если $a_2 = 0$ (случай «а») или/и $m_{12} = 0$ (случай «б»).

Случай «а», как это следует из (7.59), возникает, когда выполняется (7.53).

В случае «б» соотношение (7.60) переписывается следующим образом:

$$s_q^{2*} \overset{(7.9)}{=} m_1'a_1^* + m_2a_2 \overset{a_1^*=a_1}{=} m_1'a_1 + m_2a_2,$$

т.к. вторая (нижняя) часть системы (7.55) означает в этом случае, что $m_{22}a_2 = m_2$, т.е. a_2 — оценка параметра в регрессии $\hat{x}$ по $\hat{z}_2$:

$$\hat{x} = \hat{z}_2a_2 + e_2 = s_q^2 + s_{q2}^2, \qquad (7.62)$$

где s_{q2}^2 — дисперсия $\hat{x}$, объясненная только $\hat{z}_2$.

Что и требовалось доказать.

Иллюстрация случая «а» при $n_1 = 1$ достаточно очевидна и дана выше. Рисунок 7.4 иллюстрирует случай «б». На этом рисунке: OA — $\hat{x}$, OB — $\hat{z}_1$, OC — $\hat{z}_2$, EA — e, нормаль $\hat{x}$ на $\hat{z}_1$, FA — e_2, нормаль $\hat{x}$ на $\hat{z}_2$, DA — e^*, нормаль $\hat{x}$ на плоскость, определенную $\hat{z}_1$ и $\hat{z}_2$, ED — нормаль к $\hat{z}_1$, FD — нормаль к $\hat{z}_2$.

Понятно (геометрически), что такая ситуация, когда точка E является одновременно началом нормалей EA и ED, а точка F — началом нормалей FA и FD, возможна только в случае, если угол COB равен $90°$.

Но именно этот случай означает (как это следует из рисунка) одновременное выполнение соотношений регрессий (7.51) ($OE + EA = OA$), (7.52) (при $a_1^* = a_1$) ($OE+OF+DA = OA$) и (7.62) ($OF+FA = OA$), т.е. что введение нового фактора не меняет оценку при «старом» факторе, а «новая» объясненная дисперсия равна сумме дисперсий, объясненных «старым» и «новым» факторами по отдельности (сумма квадратов длин векторов OE и OF равна квадрату длины вектора OD).

На основании сделанных утверждений можно сформулировать такое правило введения новых факторов в уравнение регрессии: вводить в регрессию следует такие факторы, которые имеют высокую корреляцию с остатками по уже введенным факторам и низкую корреляцию с этими уже введенными факторами. В этом процессе следует пользоваться F-критерием: вводить новые факторы до тех пор, пока уменьшается показатель pv F-статистики.

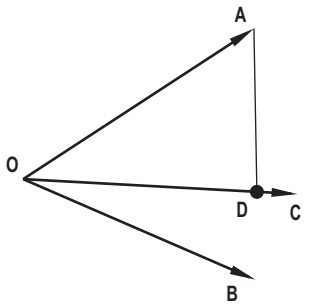

Рис. 7.5

В таком процессе добавления новых факторов в регрессионную модель некоторые из ранее введенных факторов могут перестать быть значимыми, и их следует выводить из уравнения.

Эту возможность иллюстрирует рисунок 7.5 в пространстве наблюдений при $n_1 = 1$.

На этом рисунке: OA — $\hat{x}$, OB— $\hat{z}_1$, OC — $\hat{z}_2$, AD — нормаль $\hat{x}$ на плоскость, определенную $\hat{z}_1$ и $\hat{z}_2$.

Рисунок показывает, что нормаль AD «легла» на вектор вновь введенного фактора. Следовательно, «старый» фактор входит в «новую» регрессию с нулевым коэффициентом.

Это — крайний случай, когда «старый» фактор автоматически выводится из уравнения. Чаще встречается ситуация, в которой коэффициенты при некоторых «старых» факторах оказываются слишком низкими и статистически незначимыми.

Процесс, в котором оценивается целесообразность введения новых факторов и выведения ранее введенных факторов, называется **шаговой регрессией**. В развитой форме этот процесс можно организовать следующим образом.

Пусть z — полный набор факторов, потенциально влияющих на x. Рассматривается процесс обращения матрицы ковариации переменных x, z, в начале которого рядом с этой матрицей записывается единичная матрица. С этой парой матриц производятся одновременные линейные преобразования. Известно, что если первую матрицу привести таким образом к единичной, то на месте второй будет получена матрица, обратная к матрице ковариации. Пусть этот процесс не завершен,

и только n_1 строк первой матрицы, начиная с ее второй строки (т.е. со строки первого фактора), преобразованы в орты; z_1 — множество факторов, строки которых преобразованы в орты, z_2 — остальные факторы. Это — ситуация на текущем шаге процесса.

В начале процесса пара преобразуемых матриц имеет вид (над матрицами показаны переменные, которые соответствуют их столбцам):

$$\begin{matrix} x & z_1 & z_2 \end{matrix}\\ \begin{bmatrix} m_{xx} & m_1' & m_2' \\ m_1 & M_{11} & M_{12} \\ m_2 & M_{12}' & M_{22} \end{bmatrix} \quad \text{и} \quad \begin{matrix} x & z_1 & z_2 \end{matrix}\\ \begin{bmatrix} 1 & 0 & 0 \\ 0 & I_1 & 0 \\ 0 & 0 & I_2 \end{bmatrix},$$

где

$m_{xx} = \frac{1}{N}\hat{X}'\hat{X}$ — дисперсия x,

$m_1 = \frac{1}{N}\hat{Z}_1\hat{X}$ — вектор-столбец коэффициентов ковариации z_1 и x,

$m_2 = \frac{1}{N}\hat{Z}_2\hat{X}$ — вектор-столбец коэффициентов ковариации z_2 и x,

$M_{11} = \frac{1}{N}\hat{Z}_1'\hat{Z}_1$ — матрица коэффициентов ковариации z_1 между собой,

$M_{12} = \frac{1}{N}\hat{Z}_1'\hat{Z}_2$ — матрица коэффициентов ковариации z_1 и z_2,

$M_{22} = \frac{1}{N}\hat{Z}_2'\hat{Z}_2$ — матрица коэффициентов ковариации z_2 между собой.

На текущем шаге эти матрицы преобразуются к виду:

$$\begin{matrix} x & z_1 & z_2 \end{matrix}\\ \begin{bmatrix} m_{xx} - m_1'M_1^{-1}m_1 & \underbrace{m_1'M_1^{-1}}_{a_1} & \underbrace{m_2' - m_1'M_1^{-1}M_{12}}_{c_{e2}} \\ 0 & I_1 & 0 \\ m_2 - M_{12}'M_1^{-1}m_1 & M_{12}'M_1^{-1} & M_2 - M_{12}'M_1^{-1}M_{12} \end{bmatrix}$$

$$\text{и} \quad \begin{matrix} x & z_1 & z_2 \end{matrix}\\ \begin{bmatrix} 1 & 0 & 0 \\ -M_1^{-1}m_1 & M_1^{-1} & -M_1^{-1}M_{12} \\ 0 & 0 & I_2 \end{bmatrix}.$$

Информация, используемая в шаговой регрессии, расположена в 1-й строке первой матрицы: остаточная дисперсия в текущей регрессии (в столбце x), коэффициенты a_1 текущей регрессии при переменных z_1 (в столбцах z_1), коэффициенты c_{e2} ковариации текущих остатков e с переменными z_2, не включенными в текущую регрессию (в столбцах z_2).

Для введения очередного фактора в регрессию (шаг вперед) следует его строку в первой матрице преобразовать в орт, для исключения фактора из регрессии (шаг назад) следует преобразовать в орт его строку во второй матрице. Шаг вперед увеличивает количество элементов в векторе z_1 на единицу и сокращает на единицу количество элементов в векторе z_2. Шаг назад приводит к обратным изменениям. Последствия любого из этих шагов можно оценить по F-критерию, рассчитав показатель pv F^c-статистики (информацию для такого расчета дает остаточная дисперсия — первый элемент первой строки первой матрицы).

На текущем шаге процесса проверяются последствия введения всех ранее не введенных факторов z_2 и исключения всех введенных факторов z_1. Выбирается тот вариант, который дает минимальное значение показателя pv. Процесс заканчивается, как только этот показатель перестает падать. В результате определяется наилучшая регрессия. Такой процесс не приводит, как правило, к включению в регрессию сильно коррелированных факторов, т.е. позволяет решить проблему мультиколлинеарности.

Если бы расчеты проводились в стандартизированной шкале (по коэффициентам корреляции, а не ковариации), «кандидатом» на введение был бы фактор с максимальным значением показателя в множестве c_{e2} (как было показано выше), а на исключение — фактор с минимальным значением показателя в множестве a_1. Но даже в этом случае для окончательного выбора (вводить-исключать) и решения вопроса о завершении процесса требуется использование F-критерия. При «работе» с коэффициентами ковариации использование F-критерия необходимо.

На последних шагах процесса, при приближении к минимуму критериального показателя pv, его величина меняется, как правило, весьма незначительно. Поэтому один из возможных подходов к использованию шаговой регрессии заключается в определении некоторого множества регрессий, получаемых на последних шагах процесса, которые практически одинаковы по своему качеству. И на этом множестве следует делать окончательный выбор, пользуясь содержательными критериями.

Иногда процесс шаговой регрессии предлагают строить на основе t-критерия: фактор вводится в уравнение, если его t-статистика больше некоторой заданной величины t_1, выводится из уравнения, если эта статистика меньше заданной величины t_2; как правило, $t_1 > t_2$. Такой процесс не гарантирует получение наилучшей

регрессии, его использовали в то время, когда вычислительные возможности были еще слабо развиты, и, в частности, точные значения показателя pv было трудно определить.

7.4. Прогнозирование

Пусть получены оценки параметров уравнения (7.11). Задача прогнозирования заключается в определении возможного значения (прогноза) переменной x, объясняемой этой моделью, при некоторых заданных значениях факторов z, которые не совпадают ни с одним из наблюдений в матрице Z. Более того, как правило, z лежит вне области, представляемой матрицей Z. При этом предполагается, что гипотезы **g1**–**g3** по-прежнему выполняются.

Обычно термин «прогнозирование» используется в случае, когда наблюдения $i = 1, \ldots, N$ в матрице Z даны по последовательным моментам (периодам) времени, и заданные значения факторов z, для которых требуется определить прогноз x, относятся к какому-то будущему моменту времени, большему N (т.е. z лежит вне области, представляемой матрицей Z).

Методы прогнозирования могут быть различными. Если применяются относительно простые статистические методы, как в данном случае, то часто используют термин «**экстраполирование**». Если аналогичная задача решается для z, лежащих внутри области, представляемой наблюдениями в матрице Z (например, для «пропущенных» по каким-то причинам наблюдений), то используют термин «**интерполирование**». Процедуры экстраполирования и интерполирования с использованием модели (7.11) с формальной точки зрения одинаковы.

Итак, задан некоторый $z_r = [z_{r1} \ \cdots \ z_{rn} \ \ 1]$, который отличается от всех z_i, $i = 1, \ldots, N$ (если i — обозначает момент времени, то $r > N$).

$x_r = z_r\alpha + \varepsilon_r$ — истинное значение искомой величины,

$x_r^0 = z_r\alpha$ — ожидаемое значение,

$x_r^p = z_r a$ — искомый (точечный) прогноз.

Предполагаем, что гипотезы **g1**–**g4** выполнены как для $i = 1, \ldots, N$, так и для $r > N$.

Это линейный (относительно случайных величин X) прогноз: $x_r^p \overset{(7.26)}{=} z_r LX$, он не смещен относительно ожидаемого значения вслед за несмещенностью a: $\mathbf{E}\,(x_r^p) = x_r^0$. Его ошибка $\varepsilon_r^p = x_r - x_r^p$ имеет нулевое математическое ожидание и дисперсию

$$\sigma_p^2 = \sigma^2 \left(1 + z_r \left(Z'Z\right)^{-1} z_r'\right), \tag{7.63}$$

которая минимальна на множестве всех возможных линейных несмещенных прогнозов.

Действительно:

$$\varepsilon_r^p = z_r(\alpha - a) + \varepsilon_r.$$

Поскольку случайные величины a и ε_r не зависят друг от друга,

$$\sigma_p^2 = \mathbf{E}\left((\varepsilon_r^p)^2\right) = \mathbf{E}\left(z_r(\alpha-a)(\alpha-a)'z_r'\right) + \mathbf{E}\left(\varepsilon_r^2\right) = \\ = z_r M_a z_r' + \sigma^2 \overset{(7.29)}{=} \sigma^2\left(1 + z_r\left(Z'Z\right)^{-1} z_r'\right).$$

Эта дисперсия минимальна среди всех возможных дисперсий линейных несмещенных прогнозов вслед за аналогичным свойством оценок a. Это является прямым следствием того, что оценки МНК относятся к классу **BLUE**. Для того чтобы в этом убедиться, достаточно в доказательстве данного свойства оценок a, которое приведено в п. 7.2, заменить c' на z_r.

Следует иметь в виду, что ошибка любого расчетного по модели значения x_i^c, являясь формально такой же: $\varepsilon_i^c = x_i - x_i^c$, имеет также нулевое математическое ожидание, но принципиально другую, существенно меньшую, дисперсию:

$$\sigma_i^2 = \sigma^2\left(1 - z_i\left(Z'Z\right)^{-1} z_i'\right).$$

Видно, что эта дисперсия даже меньше остаточной.

Действительно, как и прежде: $\varepsilon_i^c = z_i(\alpha - a) + \varepsilon_i$. Но теперь случайные величины a и ε_i коррелированы и поэтому:

$$\sigma_i^2 = \sigma^2\left(1 + z_i\left(Z'Z\right)^{-1} z_i'\right) + 2z_i\mathbf{E}(\underbrace{(\alpha-a)}_{\overset{(7.27)}{=} -L\varepsilon}\varepsilon_i) \overset{\substack{\mathbf{E}(\varepsilon\varepsilon_i)\overset{\mathbf{g4}}{=}\sigma^2 o_i,\\ \text{где } o_i \text{ — } i\text{-й орт}}}{=} \\ = \sigma^2\left(1 + z_i\left(Z'Z\right)^{-1} z_i'\right) - 2\sigma^2 z_i\left(Z'Z\right)^{-1} z_i' = \sigma^2\left(1 - z_i\left(Z'Z\right)^{-1} z_i'\right).$$

Величины $1 - z_i\left(Z'Z\right)^{-1} z_i'$ $(i = 1, \ldots, N)$, естественно, неотрицательны, поскольку они являются диагональными элементами матрицы B из (7.32), которая положительно полуопределена.

Структуру дисперсии ошибки прогноза (7.63) можно пояснить на примере $n = 1$. В этом случае (используются обозначения исходной формы уравнения регрессии, и все z — одномерные величины):

$$\sigma_p^2 = \sigma^2\left(1 + \frac{1}{N} + \frac{(z_r - \bar{z})^2}{\sum \hat{z}_i^2}\right). \tag{7.64}$$

В этом легко убедиться, если перейти к обозначениям исходной формы уравнения регрессии, подставить в (7.63) вместо z_r и Z, соответственно, $\begin{bmatrix} z_r & 1 \end{bmatrix}$ и $\begin{bmatrix} Z & 1_N \end{bmatrix}$ и сделать необходимые преобразования (правило обращения матрицы (2×2) см. в Приложении А.1.2), учитывая, что

$$\begin{bmatrix} \xi_1 & \xi_2 \\ \xi_3 & \xi_4 \end{bmatrix}^{-1} = \frac{1}{\xi_1\xi_4 - \xi_2\xi_3} \begin{bmatrix} \xi_4 & -\xi_2 \\ -\xi_3 & \xi_1 \end{bmatrix} \quad \text{и} \quad Z'Z = \sum \hat{z}_i^2 + N\bar{z}^2 :$$

$$\begin{aligned} \sigma_p^2 &= \sigma^2 \left(1 + \begin{bmatrix} z_r & 1 \end{bmatrix} \begin{bmatrix} Z'Z & N\bar{z} \\ N\bar{z} & N \end{bmatrix}^{-1} \begin{bmatrix} z_r \\ 1 \end{bmatrix} \right) = \\ &= \sigma^2 \left(1 + \frac{1}{Z'Z - N\bar{z}} \begin{bmatrix} z_r & 1 \end{bmatrix} \begin{bmatrix} 1 & -\bar{z} \\ -\bar{z} & \frac{1}{N}Z'Z \end{bmatrix} \begin{bmatrix} z_r \\ 1 \end{bmatrix} \right) = \\ &= \sigma^2 \left(1 + \frac{z_r^2 - 2\bar{z}z_r + \frac{1}{N}\left(\sum \hat{z}_i^2 + N\bar{z}^2\right)}{\sum \hat{z}_i^2} \right) = \sigma^2 \left(1 + \frac{1}{N} + \frac{(z_r - \bar{z})^2}{\sum \hat{z}_i^2} \right). \end{aligned}$$

Что и требовалось доказать.

Это выражение показывает «вклады» в дисперсию ошибки прогноза собственно остаточной дисперсии, ошибки оценки свободного члена и ошибки оценки углового коэффициента. Первые две составляющие постоянны и не зависят от горизонта прогнозирования, т.е. от того, насколько сильно условия прогноза (в частности, значение z_r) отличаются от условий, в которых построена модель (в частности, значение $\bar{z}$). Третья составляющая — ошибка оценки углового коэффициента — определяет расширяющийся конус ошибки прогноза.

Мы рассмотрели точечный прогноз. Если дополнительно к гипотезам **g1**–**g4** предположить выполнение гипотезы **g5** для $i = 1, \ldots, N$ и для $r > N$, то можно построить также интервальный прогноз.

По формуле (7.27) ошибка прогноза имеет вид:

$$\varepsilon_r^p = z_r(\alpha - a) + \varepsilon_r = z_r\mathbf{L}\varepsilon + \varepsilon_r.$$

Таким образом, она имеет нормальное распределение:

$$\varepsilon_r^p = x_r - x_r^p \sim N(0, \sigma_p^2).$$

Если бы дисперсия ошибки σ^2 была известна, то на основе того, что

$$\frac{x_r - x_r^p}{\sigma_p} \sim N(0, 1),$$

для x_r можно было бы построить $(1-\theta)100$-процентный прогнозный интервал:

$$x_r \in [x_r^p \pm \sigma_p \hat{\varepsilon}_{1-\theta}]\,.$$

Вместо неизвестной дисперсии $\sigma_p^2 = \sigma^2(1+z_r(Z'Z)^{-1}z_r')$ берется несмещенная оценка

$$s_p^2 = \hat{s}_e^2(1+z_r(Z'Z)^{-1}z_r').$$

По аналогии с (7.44) можно вывести, что

$$\frac{x_r - x_r^p}{s_p} \sim t_{N-n-1}.$$

Тогда в приведенной формуле прогнозного интервала необходимо заменить σ_p на s_p и $\hat{\varepsilon}_{1-\theta}$ на $\hat{t}_{N-n-1,\,1-\theta}$:

$$x_r \in \left[x_r^p \pm s_p \hat{t}_{N-n-1,\,1-\theta}\right].$$

Таблица 7.1

X	Z_1	Z_2
65.7	26.8	541
74.2	25.3	616
74	25.3	610
66.8	31.1	636
64.1	33.3	651
67.7	31.2	645
70.9	29.5	653
69.6	30.3	682
67	29.1	604
68.4	23.7	515
70.7	15.6	390
69.6	13.9	364
63.1	18.8	411
48.4	27.4	459
55.1	26.9	517
55.8	27.7	551
58.2	24.5	506
64.7	22.2	538
73.5	19.3	576
68.4	24.7	697

7.5. Упражнения и задачи

Упражнение 1

По наблюдениям за объясняемой переменной X и за объясняющими переменными $Z = (Z_1, Z_2)$ из таблицы 7.1:

1.1. Вычислите ковариационную матрицу переменных z ($M = \frac{1}{N}\hat{Z}'\hat{Z}$), вектор ковариаций переменных z с переменной x ($m = \frac{1}{N}\hat{Z}'\hat{X}$), дисперсию объясняемой переменной s_x^2. Для регрессии $X = Za + 1_N b + e$ найдите оценки a и b, объясненную дисперсию $s_q^2 = m'a$ и остаточную дисперсию $s_e^2 = s_x^2 - s_q^2$, а также коэффициент детерминации R^2.

1.2. Запишите для данной модели уравнение регрессии в форме со скрытым свободным членом $X = \widetilde{Z}\widetilde{a} + e$. Рассчитайте для переменных начальные моменты второго порядка двумя способами:

а) $\widetilde{M} = \frac{1}{N}\widetilde{Z}'\widetilde{Z}$ и $\widetilde{m} = \frac{1}{N}\widetilde{Z}'X$

б) $\widetilde{M} = \begin{bmatrix} M + \bar{z}'\bar{z} & \bar{z}' \\ \bar{z} & 1 \end{bmatrix}$ и $\widetilde{m} = \begin{bmatrix} m + \bar{z}'\bar{x} \\ \bar{x} \end{bmatrix}$.

1.3. Найдите оценку $\widetilde{a}$, рассчитайте $s_x^2 = \frac{1}{N}X'X - \bar{x}^2$ и $s_q^2 = \widetilde{m}'\widetilde{a} - \bar{x}^2$ и убедитесь, что результат совпадает с результатом пункта 1 упражнения 1.

1.4. Рассчитайте несмещенную оценку остаточной дисперсии

$$\hat{s}_e^2 = \frac{N}{N-n-1}s_e^2$$

и оцените матрицу ковариации параметров уравнения регрессии

$$M_a = \frac{\hat{s}_e^2}{N}\widetilde{M}^{-1}.$$

1.5. Используя уровень значимости $\theta = 0.05$, вычислите доверительные интервалы для коэффициентов уравнения регрессии и проверьте значимость факторов.

1.6. Рассчитайте статистику $F^c = \dfrac{R^2(N-n-1)}{(1-R^2)n}$ и, используя уровень значимости $\theta = 0.05$, проверьте гипотезу о том, что модель некорректна и все факторы введены в нее ошибочно.

1.7. Рассчитайте коэффициент детерминации, скорректированный на число степеней свободы $\tilde{R}^2$.

1.8. По найденному уравнению регрессии и значениям

а) $z = (\min Z_1, \min Z_2)$;

б) $z = (\bar{Z}_1, \bar{Z}_2)$;

в) $z = (\max Z_1, \max Z_2)$;

вычислите предсказанное значение для x и соответствующую интервальную оценку при $\theta = 0.05$.

Упражнение 2

Дано уравнение регрессии: $X = \widetilde{Z}\widetilde{\alpha} + \varepsilon = -1.410z_1 + 0.080z_2 + 56.962\ 1_{20} + \varepsilon$, где X — вектор-столбец 20 наблюдений за объясняемой переменной (20×1), ε — вектор-столбец случайных ошибок (20×1) с нулевым средним и ковариационной матрицей $\sigma^2 I_{20} = 21.611 I_{20}$ и $\widetilde{Z}$ — матрица размерности (20×3) наблюдений за объясняющими переменными. Используя нормальное распределение

с независимыми наблюдениями, со средним 0 и ковариационной матрицей $\sigma^2 I_{20} = 21.611 I_{20}$, получите 100 выборок вектора ε $(N \times 1)$, $k = 1, \ldots, 100$, где $N = 20$. Эти случайные векторы потом используйте вместе с известным вектором $\widetilde{\alpha} = (-1.410, 0.080, 56.962)$ и матрицей $\widetilde{Z} = (Z_1, Z_2, 1)$ из таблицы 7.1. Сначала получите ожидаемое значения $X^0 = \widetilde{Z}\widetilde{\alpha}$, затем, чтобы получить 100 выборок вектора X (20×1), добавьте случайные ошибки: $X^0 + \varepsilon = X$.

2.1. Используйте 10 из 100 выборок, чтобы получить выборочные оценки для α_1, α_2, β, σ и R^2.

2.2. Вычислите матрицу ковариаций параметров уравнения регрессии M_a для каждого элемента выборки и сравните с истинным значением ковариационной матрицы:

$$\sigma^2 \left(\widetilde{Z}'\widetilde{Z}\right)^{-1} = \begin{pmatrix} 0.099813 & -0.004112 & -0.233234 \\ -0.004112 & 0.000290 & -0.057857 \\ -0.233234 & -0.057857 & 39.278158 \end{pmatrix}.$$

Дайте интерпретацию диагональных элементов ковариационных матриц.

2.3. Вычислите среднее и дисперсию для 10 выборок для каждого из параметров, полученных в упражнении 2.1, и сравните эти средние значения с истинными параметрами. Обратите внимание, подтвердилась ли ожидаемые теоретические результаты.

2.4. Используя уровень значимости $\theta = 0.05$, вычислите и сравните интервальные оценки для α_1, α_2, β и σ для 10 выборок.

2.5. Объедините 10 выборок, по 20 наблюдений каждая, в 5 выборок по 40 наблюдений и повторите упражнения 2.1 и 2.2. Сделайте выводы о результатах увеличения объема выборки.

2.6. Повторите упражнения 2.1 и 2.5 для всех 100 и для 50 выборок и проанализируйте разницу в результатах.

2.7. Постройте распределения частот для оценок, полученных в упражнении 2.6, сравните и прокомментируйте результаты.

Задачи

1. В регрессии $X = Za + 1_N b + e$ матрица вторых начальных моментов регрессоров равна $\begin{pmatrix} 9 & 2 \\ 2 & 1 \end{pmatrix}$. Найдите дисперсию объясняющей переменной.

2. На основании ежегодных данных за 10 лет с помощью МНК была сделана оценка параметров производственной функции типа Кобба—Дугласа. Чему равна несмещенная оценка дисперсии ошибки, если сумма квадратов остатков равна 32?

3. В регрессии $X = Za + 1_N b + e$ с факторами $Z' = (1,\ 2,\ 3)$ сумма квадратов остатков равна 6. Найдите ковариационную матрицу оценок параметров регрессии.

4. Какие свойства МНК-оценок коэффициентов регрессии теряются, если ошибки по наблюдениям коррелированы и/или имеют разные дисперсии?

5. Что обеспечивает гипотеза о нормальности распределения ошибок при построения уравнения регрессии? Ответ обоснуйте.

6. Какие ограничения на параметры уравнения проверяются с помощью t-критерия (написать ограничения с расшифровкой обозначений)?

7. Четырехфакторное уравнение регрессии оценено по 20-ти наблюдениям. В каком случае отношение оценки коэффициента регрессии к ее стандартной ошибке имеет распределение t-Стьюдента? Сколько степенией свободы в этом случае имеет эта статистика?

8. Оценки МНК в регрессии по 20-ти наблюдениям равны $(2,\ -1)$, а ковариационная матрица этих оценок равна $\begin{pmatrix} 9 & 2 \\ 2 & 1 \end{pmatrix}$. Найти статистики t-Стьюдента для этих коэффициентов.

9. По 10 наблюдениям дана оценка 4 одному из коэффициентов двухфакторной регрессии. Дисперсия его ошибки равна 4. Построить 99%-ный доверительный интервал для этого коэффициента.

10. МНК-оценка параметра регрессии, полученная по 16 наблюдениям, равна 4, оценка его стандартной ошибки равна 1. Можно ли утверждать с вероятностью ошибки не более 5%, что истинное значение параметра равно 5.93? Объяснить почему.

11. Оценка углового коэффициента регрессии равна 4, а дисперсия этой оценки равна 4. Значим ли этот коэффициент, если табличные значения:

$$t_{N-n-1,\,0.95} = 2.4,\ t_{N-n-1,\,0.90} = 1.9?$$

12. В результате оценивания регрессии $x = z\alpha + 1_N\beta + \varepsilon$ на основе $N = 30$ наблюдений получены следующие результаты:

$x =$	$1.2z_1+$	$1.0z_2-$	$0.5z_3+$	25.1
Стандартные ошибки оценок	( )	(1.3)	(0.06)	(2.1)
t-статистика	(0.8)	( )	( )	( )
95% доверительные интервалы	$(-1.88;\ 4.28)$	( )	( )	( )

Заполните пропуски в скобках.

13. На основе годовых отчетов за 1973–1992 годы о затратах на продукты питания Q, располагаемом доходе Y, индексе цен на продукты питания PF и индексе цен на непродовольственные товары PNF, группа исследователей получила различные регрессионные уравнения для функции спроса на продукты питания:

$$\begin{aligned}
\ln Q &= 3.87 - 1.34\ln PF \\
&\quad\ (1.45)\quad (-4.54) \\
R^2 &= 0.56 \\
\ln Q &= 2.83 - 0.92\ln PF + 1.23\ln Y \\
&\quad\ (1.25)\quad (-2.70)\quad\ (2.99) \\
R^2 &= 0.76 \\
\ln Q &= 2.35 - 0.52\ln PF + 0.95\ln Y + 1.54\ln PNF \\
&\quad\ (1.54)\quad (-1.80)\quad\ (0.79)\quad\ (2.45) \\
R^2 &= 0.84
\end{aligned}$$

В скобках приведены значения t-статистики.

Прокомментируйте полученные оценки коэффициентов и t-статистики, объясните, почему значения могут различаться в трех уравнениях. Можете ли вы предложить решение проблемы статистической незначимости коэффициентов в последнем уравнении?

14. Используя приведенные ниже данные, оцените параметры модели $x_t = \beta + \alpha_1 z_{1t} + \alpha_2 z_{2t} + \varepsilon_t$ и, делая все необходимые предположения, проверьте статистическую значимость коэффициента α_1.

    а) $\sum \hat{z}_{1t}^2 = 10$, $\sum \hat{z}_{2t}^2 = 8$, $\sum \hat{z}_{1t}\hat{z}_{2t} = 8$, $\sum \hat{z}_{1t}\hat{x}_t = -10$, $\sum \hat{z}_{2t}\hat{x}_t = -8$, $\sum \hat{x}_t^2 = 20$, $t = 1, \ldots, 5$;

    б) $\sum z_{1t}^2 = 55$, $\sum z_{2t}^2 = 28$, $\sum z_{1t}z_{2t} = 38$, $\sum z_{1t}x_t = 35$, $\sum z_{2t}x_t = 22$, $\sum x_t = 15$, $\sum z_1 = 15$, $\sum z_2 = 10$, $N = 5$, $\sum x^2 = 65$.

15. Анализ годовых данных (21 наблюдение) о спросе на некоторый товар привел к следующим результатам:

Средние	Стандартные отклонения	Парные коэффициенты корреляции
$\bar{z} = 51.843$	$s_z = 9.205$	$r_{xz} = 0.9158$
$\bar{x} = 8.313$	$s_x = 1.780$	$r_{xt} = 0.8696$
$\bar{t} = 0$	$s_t = 6.055$	$r_{zt} = 0.9304$

    z — потребление на душу населения, x — цена с учетом дефлятора, t — время (годы).

    а) Найдите коэффициент при времени в оцененной регрессии x по z и t.

    б) Проверьте, будет ли этот коэффициент значимо отличен от нуля.

    в) Кратко объясните экономический смысл включения в регрессию времени в качестве объясняющей переменной.

16. Какие ограничения на параметры уравнения можно проверить с помощью F-критерия? Написать ограничения с расшифровкой обозначений.

17. Пяти-факторное уравнение линейной регрессии для переменной x оценено по 31 наблюдению. При этом объясненная и смещенная остаточная дисперсии соответственно равны 8 и 2. Вычислить коэффициент детерминации и расчетное значение F-статистики.

18. В регрессии $x = z_1\alpha_1 + z_2\alpha_2 + \beta + \varepsilon$ по 5-ти наблюдениям смещенная оценка остаточной дисперсии равна 1, а дисперсия зависимой переменной равна 2. Значима ли эта зависимость?

19. По 10 наблюдениям оценено двухфакторное уравнение линейной регрессии, коэффициент детерминации составляет 90%. При каком уровне доверия это уравнение статистически значимо? Записать уравнение для нахождения этого уровня значимости.

20. Используя следующие данные:

$X = (5,\ 1,\ -2,\ 5,\ -4)',\ Z = (1,\ 2,\ 3,\ 4,\ 5)',$

и делая все необходимые предположения

а) для $X = Z\alpha + 1_N\beta + \varepsilon$ оценить 95-процентные доверительные интервалы для параметров регрессии;

б) проверить значимость коэффициентов регрессии и оценить качество регрессии с вероятностью ошибки 5%.

21. Пусть $X = \alpha_1 Z_1 + \alpha_2 Z_2 + \varepsilon$, $X = (4,\ -2,\ 4,\ 0)'$, $Z_1 = (1,\ 1,\ 2,\ 2)'$ и $Z_2 = 2Z_1$. Постройте систему нормальных уравнений и покажите, что существует бесконечное множество решений для a_1 и a_2. Выберите любые два решения, покажите, что они дают одинаковые расчетные значения X и, таким образом, одинаковые значения суммы квадратов ошибок.

22. Для уравнения регрессии $X = Z\alpha + 1_5\beta + \varepsilon$ имеются следующие данные:

$$X = \begin{pmatrix} 4 \\ 8 \\ 5.5 \\ 5.8 \\ 7.0 \end{pmatrix}, \quad Z = (Z_1 \quad Z_2 \quad Z_3) = \begin{pmatrix} 1.03 & 2.08 & 0.41 \\ 1.46 & 2.80 & 2.03 \\ 1.14 & 2.30 & 0.98 \\ 1.71 & 3.05 & 0.81 \\ 1.06 & 2.17 & 1.17 \end{pmatrix}.$$

а) Являются ли факторы линейно зависимыми?

б) Найти матрицу коэффициентов корреляции факторных переменных, рассчитать определитель данной матрицы и сделать вывод о мультиколлинеарности факторов.

в) Рассчитать определитель матрицы коэффициентов корреляции факторных переменных в случае, если из уравнения выводится фактор Z_2.

г) Учесть дополнительную внешнюю информацию: $\alpha_1 = 1.5\alpha_2$ (с помощью подстановки в уравнение регрессии) и найти определитель матрицы коэффициентов корреляции факторных переменных.

д) Построить точечный прогноз x (x_r^p) для значений экзогенных переменных $z_r = (z_{1r}, z_{2r}, z_{3r}) = (0.8,\ 1.6,\ 0.6)$:

- при использовании исходного уравнения;
- при исключении из уравнения фактора Z_2;

– при использовании внешней информации из пункта (г).

23. Пусть цены сильно коррелируют с денежной массой и неплатежами. Коэффициент корреляции между денежной массой и неплатежами равен 0.975 $(R^2 = 0.95)$. Имеет ли смысл строить регрессию цен на эти два (сильно мультиколлинеарных) фактора?

24. Модель

$$x = \alpha_1 z_1 + \alpha_2 z_2 + \beta + \varepsilon \qquad (1)$$

была оценена по МНК, и был получен коэффициент детерминации R_1^2, а для преобразованной модели

$$x = \alpha_1 z_1 + \alpha_2 z_2 + \alpha_3 z_3 + \beta + \varepsilon \qquad (2)$$

был получен коэффициент детерминации R_2^2.

а) Объясните, почему R_1^2 не может быть больше, чем R_2^2. При каких условиях они равны?

б) Объясните последствия оценки модели (1), если верной является модель (2).

25. В регрессии $x = \alpha_1 z_1 + \beta + \varepsilon$ остатки равны $(-2,\ 1,\ 0,\ 1)$. Оценивается регрессия $x = \alpha_1 z_1 + \alpha_2 z_2 + \beta + \varepsilon$. Привести пример переменной z_2, чтобы коэффициенты детерминации в обеих регрессиях совпадали.

26. В регрессию $x = \alpha_1 z_1 + \beta + \varepsilon$ добавили переменную z_2. Переменная z_2 оказалась совершенно незначимой. Как изменились обычный и скорректированный коэффициенты детерминации?

27. Коэффициент детерминации в регрессии выпуска продукции по численности занятых в производстве, оцененной по 12 наблюдениям, равен 0.8. После введения в регрессию дополнительного фактора — основного капитала — он вырос до 0.819. Имело ли смысл вводить этот дополнительный фактор? Ответ обосновать без применения статистических критериев.

28. Дана модель регрессии $x_i = \alpha_1 z_i + \beta + \varepsilon_i$.

а) Как оценивается точечный прогноз x_{N+1}, если известно, что $\beta = 0$? Покажите, что дисперсия ошибок прогноза будет равна $\sigma^2\left(1 + \dfrac{z_{N+1}^2}{\sum\limits_{i=1}^{N} z_i^2}\right)$.

б) Как оценивается точечный прогноз x_{N+1}, если известно, что $\alpha = 0$? Покажите, что дисперсия ошибок прогноза будет равна $\sigma^2\left(1+\frac{1}{N}\right)$.

29. Почему ошибки прогнозирования по линейной регрессии увеличиваются с ростом горизонта прогноза?

30. Была оценена регрессия $x = \alpha_1 z + \beta + \varepsilon$ по 50 наблюдениям. Делается прогноз x в точке z_{51}. При каком значении z_{51} доверительный интервал прогноза будет самым узким?

31. Вычислите предсказанное значение для x и соответствующую интервальную оценку прогноза при $\theta = 0.05$ в точке $z_{26} = 14$, если регрессионная модель $x = 3z + 220 + e$ построена по 25 наблюдениям, остаточная дисперсия равна 25 и средняя по z равна 14.

Рекомендуемая литература

1. **Айвазян С.А.** Основы эконометрики. Т.2. — М.: Юнити, 2001. (Гл. 2).

2. **Демиденко Е.З.** Линейная и нелинейная регрессия. — М.: «Финансы и статистика», 1981. (Гл. 1, 2, 6).

3. **Джонстон Дж.** Эконометрические методы. — М.: «Статистика», 1980. (Гл. 2, 5).

4. **Дрейпер Н., Смит Г.** Прикладной регрессионный анализ: В 2-х книгах. Кн.1 — М.: «Финансы и статистика», 1986, (Гл. 1, 2).

5. **Кейн Э.** Экономическая статистика и эконометрия. Вып. 2. — М.: «Статистика», 1977. (Гл. 10, 11, 14).

6. **Магнус Я.Р., Катышев П.К., Пересецкий А.А.** Эконометрика — начальный курс. — М.: «Дело», 2000. (Гл. 3, 4, 8).

7. **(*) Маленво Э.** Статистические методы эконометрии. Вып. 1. — М.: «Статистика», 1975. (Гл. 3, 6).

8. **Себер Дж.** Линейный регрессионый анализ. — М.: Мир, 1980.

9. **Тинтер Г.** Введение в эконометрию. — М.: «Статистика», 1965. (Гл. 5).

10. **Davidson, Russel, Mackinnon, James.** Estimation and Inference in Econometrics, N 9, Oxford University Press, 1993. (Ch. 2).

11. **Greene W.H.** Econometric Analysis, Prentice-Hall, 2000. (Ch. 6, 7).

12. **Judge G.G., Hill R.C., Griffiths W.E., Luthepohl H., Lee T.** Introduction to the Theory and Practice of Econometric. John Wiley & Sons, Inc., 1993. (Ch. 5, 21).

13. **(*) William E., Griffiths R., Carter H., George G. Judge.** Learning and Practicing econometrics, N 9 John Wiley & Sons, Inc., 1993. (Ch. 8).

Глава 8

Нарушение гипотез основной линейной модели

8.1. Обобщенный метод наименьших квадратов (взвешенная регрессия)

Пусть нарушена гипотеза **g4** и матрица ковариации ошибок по наблюдениям равна не $\sigma^2 I_N$, а $\sigma^2\Omega$, где Ω — вещественная симметричная положительно полуопределенная матрица (см. Приложение А.1.2), т.е. ошибки могут быть коррелированы по наблюдениям и иметь разную дисперсию. В этом случае обычные МНК-оценки параметров регрессии (7.26) остаются несмещенными и состоятельными, но перестают быть эффективными в классе линейных несмещенных оценок.

Ковариационная матрица оценок МНК в этом случае приобретает вид

$$M_a = \sigma^2 \left(Z'Z\right)^{-1} Z'\Omega Z \left(Z'Z\right)^{-1}.$$

Действительно, $a - \mathbf{E}\,(a) = a - \alpha = (Z'Z)^{-1} Z'\varepsilon$, поэтому

$$\mathbf{E}\left[(a - \mathbf{E}(a))\,(a - \mathbf{E}(a))'\right] = (Z'Z)^{-1} Z'\mathbf{E}\,(\varepsilon\varepsilon')\,Z\,(Z'Z)^{-1} = \\ = \sigma^2 \left(Z'Z\right)^{-1} Z'\Omega Z \left(Z'Z\right)^{-1}.$$

(Ср. с выводом формулы (7.28), где $\Omega = \sigma^2 I$.)

Обычная оценка ковариационной матрицы $s_e^2 (Z'Z)^{-1}$ при этом является смещенной и несостоятельной. Как следствие, смещенными и несостоятельными оказываются оценки стандартных ошибок оценок параметров (7.35): чаще всего они преуменьшаются (т.к. ошибки по наблюдениям обычно коррелированы положительно), и заключения о качестве построенной регрессии оказываются неоправданно оптимистичными.

По этим причинам желательно применять **обобщенный МНК (ОМНК)**, заключающийся в минимизации обобщенной остаточной дисперсии

$$\frac{1}{N} e' \Omega^{-1} e.$$

В обобщенной остаточной дисперсии остатки взвешиваются в соответствии со структурой ковариационной матрицы ошибок. Минимизация приводит к получению следующего оператора ОМНК-оценивания (ср. с (7.13), где $\Omega = I_N$):

$$a = (Z'\Omega^{-1}Z)^{-1} Z'\Omega^{-1} X. \tag{8.1}$$

Для обоснования ОМНК проводится преобразование в пространстве наблюдений (см. параграф 6.4) с помощью невырожденной матрицы D размерности $N \times N$, такой, что $D^{-1}D'^{-1} = \Omega$ (такое представление допускает любая вещественная симметричная положительно определенная матрица, см. Приложение А.1.2):

$$DX = DZ\alpha + D\varepsilon. \tag{8.2}$$

Такое преобразование возвращает модель в «штатную» ситуацию, поскольку новые остатки удовлетворяют гипотезе **g4**:

$$\mathbf{E}(D\varepsilon\varepsilon' D') = D\sigma^2 \Omega D' = \sigma^2 D D^{-1} D'^{-1} D' = \sigma^2 I_N.$$

Остаточная дисперсия теперь записывается как $\frac{1}{N} e' D' D e$, а оператор оценивания — как $a = (Z'D'DZ)^{-1} Z'D'DX$.

Что и требовалось доказать, поскольку $D'D = \Omega^{-1}$.

Обычно ни дисперсии, ни тем более ковариации ошибок по наблюдениям не известны. В классической эконометрии рассматриваются два частных случая.

8.2. Гетероскедастичность ошибок

Пусть ошибки не коррелированы по наблюдениям, и матрица Ω (а вслед за ней и матрица D) диагональна. Если эта матрица единична, т.е. дисперсии ошибок

одинаковы по наблюдениям (гипотеза **g4** не нарушена), то имеет место **гомоскедастичность** или однородность ошибок по дисперсии — «штатная» ситуация. В противном случае констатируют **гетероскедастичность ошибок** или их неоднородность по дисперсии.

Пусть $\mathbf{var}(\varepsilon_i) = \sigma_i^2$ — дисперсия ошибки i-го наблюдения. Гомоскедастичность означает, что все числа σ_i^2 одинаковы, а гетероскедастичность — что среди них есть несовпадающие.

Факт неоднородности остатков по дисперсии мало сказывается на качестве оценок регрессии, если эти дисперсии не коррелированы с независимыми факторами. Это — случай гетероскедастичности «без негативных последствий».

Данное утверждение можно проиллюстрировать в случае, когда в матрице Z всего один столбец, т.е. $n = 1$ и свободный член отсутствует. Тогда формула (7.33) приобретает вид:

$$\mathbf{E}(s_e^2) = \frac{1}{N}\left(\sum_i \sigma_i^2 - \frac{\sum\limits_i \sigma_i^2 z_i^2}{\sum\limits_i z_i^2}\right).$$

Если ситуация штатная и $\sigma_i^2 = \sigma^2$, то правая часть этой формулы преобразуется к виду $\dfrac{N-1}{N}\sigma^2$, и $\dfrac{N}{N-1}s_e^2$ оказывается несмещенной оценкой σ^2, как и было показано в параграфе 7.2. Если σ_i и z_i не коррелированы, то, обозначив $\sigma^2 = \dfrac{1}{N}\sum\limits_i \sigma_i^2$, можно утверждать, что

$$\sum_i \sigma_i^2 z_i^2 \Big/ \sum_i z_i^2 \;\approx\; \sigma^2 \sum_i z_i^2 \Big/ \sum_i z_i^2 = \sigma^2,$$

т.е. ситуация остается прежней. И только если σ_i и z_i положительно (или отрицательно) коррелированы, факт гетероскедастичности имеет негативные последствия. Действительно, в случае положительной корреляции $\dfrac{\sum \sigma_i^2 z_i^2}{\sum z_i^2} > \sigma^2$ и, следовательно, $\mathbf{E}\left(\dfrac{N}{N-1}s_e^2\right) < \sigma^2$. Обычная «несмещенная» оценка остаточной дисперсии оказывается по математическому ожиданию меньше действительного значения остаточной дисперсии, т.е. она (оценка остаточной дисперсии) дает основания для неоправданно оптимистичных заключений о качестве полученной оценки модели.

Следует заметить, что факт зависимости дисперсий ошибок от независимых факторов в экономике весьма распространен. В экономике одинаковыми по дисперсии скорее являются относительные (ε/z), а не абсолютные (ε) ошибки. Поэтому, когда оценивается модель на основе данных по предприятиям, которые могут иметь

и, как правило, имеют различные масштабы, гетероскедастичности с негативными последствиями просто не может не быть.

Если имеет место гетероскедастичность, то, как правило, дисперсия ошибки связана с одной или несколькими переменными, в первую очередь — с факторами регрессии. Пусть, например, дисперсия может зависеть от некоторой переменной y_i, которая не является константой:

$$\sigma_i^2 = \sigma^2(y_i), \quad i = 1, \ldots, N.$$

Как правило, в качестве переменной y_i берется один из независимых факторов или математическое ожидание изучаемой переменной, т.е. $x^0 = Z\alpha$ (в качестве его оценки используют расчетные значения изучаемой переменной Za).

В этой ситуации желательно решить две задачи: во-первых, определить, имеет ли место предполагаемая зависимость, а во-вторых, если зависимость обнаружена, получить оценки с ее учетом. При этом могут использоваться три группы методов. Методы первой группы позволяют работать с гетероскедастичностью, которая задается произвольной непрерывной функцией $\sigma^2(\cdot)$. Для методов второй группы функция $\sigma^2(\cdot)$ должна быть монотонной. В методах третьей группы функция $\sigma^2(\cdot)$ предполагается известной с точностью до конечного числа параметров.

Примером метода из **первой** группы является **критерий Бартлетта**, который заключается в следующем.

Пусть модель оценена и найдены остатки e_i, $i = 1, \ldots, N$. Для расчета b^c — статистики, лежащей в основе применения этого критерия, все множество наблюдений делится по какому-либо принципу на k непересекающихся подмножеств. В частности, если требуется выявить, имеется ли зависимость от некоторой переменной y_i, то все наблюдения упорядочиваются по возрастанию y_i, а затем в соответствии с этим порядком делятся на подмножества. Пусть

N_l — количество элементов в l-м подмножестве, $\sum\limits_{l=1}^{k} N_l = N$;

s_l^2 — оценка дисперсии остатков в l-м подмножестве, найденная на основе остатков e_i;

$b_s = \dfrac{\frac{1}{N}\sum\limits_{l=1}^{k} N_l s_l^2}{\left(\prod\limits_{l=1}^{k} s_l^{2N_l}\right)^{1/N}}$ — отношение средней арифметической дисперсий к средней геометрической; это отношение в соответствии со свойством мажорантности средних (см. п. 2.2) больше или равно единице, и чем сильнее различаются дисперсии по подмножествам, тем оно выше.

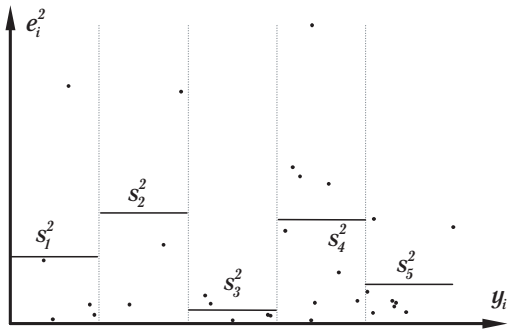

Рис. 8.1

Тогда **статистика Бартлетта** равна

$$b^c = \frac{N}{1 + \dfrac{\sum_{l=1}^{k} \frac{1}{N_l} - \frac{1}{N}}{3(k-1)}} \ln b_s.$$

При однородности наблюдений по дисперсии (нулевая гипотеза) эта статистика распределена как χ^2_{k-1}. Проверка нулевой гипотезы проводится по обычному алгоритму.

Если нулевую гипотезу отвергнуть не удалось, т.е. ситуация гомоскедастична, то исходная оценка модели удовлетворительна. Если же нулевая гипотеза отвергнута, то ситуация гетероскедастична.

Принцип построения статистики Бартлетта иллюстрирует рисунок 8.1.

Классический метод **второй** группы заключается в следующем. Все наблюдения упорядочиваются по возрастанию некоторой переменной y_i. Затем оцениваются две вспомогательные регрессии: по K «малым» и по K «большим» наблюдениям (с целью повышения мощности критерия средние $N - 2K$ наблюдения в расчете не участвуют, а K можно, например, выбрать равным приблизительно трети N). Пусть s_1^2 — остаточная дисперсия в первой из этих регрессий, а s_2^2 — во второй. В случае гомоскедастичности ошибок (нулевая гипотеза) отношение двух дисперсий распределено как

$$\frac{s_2^2}{s_1^2} \sim F_{K-n-1,\,K-n-1}.$$

Здесь следует применять обычный F-критерий. Нулевая гипотеза о гомоскедастичности принимается, если рассчитанная статистика превышает 95%-ный квантиль F-распределения.

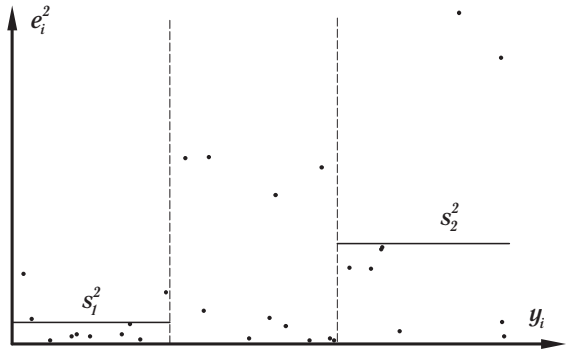

Рис. 8.2

Такой подход применяется, если ожидается, что дисперсия может быть только положительно коррелирована с переменной y_i. Если неизвестно, положительно или отрицательно коррелирована дисперсия с рассматриваемым фактором, то следует отклонять нулевую гипотезу как при больших, так и при малых значениях статистики s_2^2/s_1^2. Можно применить следующий прием: рассчитать статистику как отношение максимальной из дисперсий s_1^2 и s_2^2 к минимальной. Такая статистика будет иметь усеченное F-распределение, где усечение происходит на уровне медианы, и берется правая половина распределения. Отсюда следует, что для достижения, например, 5%-го уровня ошибки, следует взять табличную критическую границу, соответствующую, 2.5%-му правому хвосту обычного (не усеченного) F-распределения. Если указанная статистика превышает данную границу, то нулевая гипотеза о гомоскедастичности отвергается.

Данный метод известен под названием **метода Голдфельда—Квандта**.

Можно применять упрощенный вариант этого критерия, когда дисперсии s_2^2 и s_2^2 считаются на основе остатков из проверяемой регрессии. При этом s_1^2 и s_2^2 не будут независимы, и их отношение будет иметь F-распределение только приближенно. Этот метод иллюстрирует рисунок 8.2.

Для того чтобы можно было применять методы **третьей** группы, требуется обладать конкретной информацией о том, какой именно вид имеет гетероскедастичность.

Так, например, если остатки прямо пропорциональны значениям фактора ($n = 1$):

$$x = z\alpha + \beta + z\varepsilon,$$

и ε удовлетворяет необходимым гипотезам, то делением обеих частей уравнения на z ситуация возвращается в «штатную»:

$$\frac{x}{Z} = \alpha + \frac{1}{Z}\beta + \varepsilon,$$

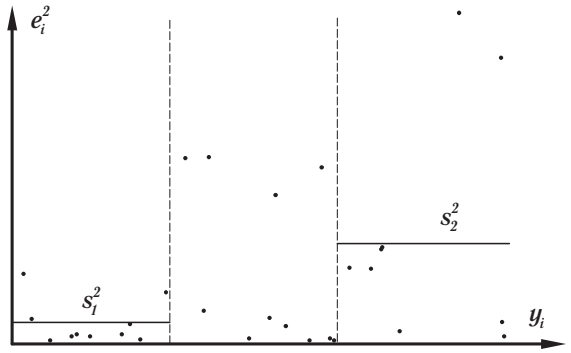

Рис. 8.3

в которой, правда, угловой коэффициент и свободный член меняются местами. Тем самым применяется преобразование в пространстве наблюдений такое, что диагональные элементы матрицы D равны $1/z_i$.

Если зависимость дисперсии от других переменных известна не точно, а только с точностью до некоторых неизвестных параметров, то для проверки гомоскедастичности следует использовать вспомогательные регрессии.

Так называемый **метод Глейзера** состоит в следующем. Строится регрессия модулей остатков $|e_i|$ на константу и те переменные, которые могут быть коррелированными с дисперсией (например, это может быть все множество независимых факторов или какое-то их подмножество). Если регрессия оказывается статистически значимой, то гипотеза гомоскедастичности отвергается.

Построение вспомогательной регрессии от некоторой переменной y_i показано на рисунке 8.3.

Другой метод (**критерий Годфрея**) использует аналогичную вспомогательную регрессию, в которой в качестве зависимой переменной используются квадраты остатков e_i^2.

Если с помощью какого-либо из перечисленных критериев (или других аналогичных критериев) проверены различные варианты возможной зависимости и нулевая гипотеза во всех случаях не была отвергнута, то делается вывод, что ситуация гомоскедастична или гетероскедастична без негативных последствий и что для оценки параметров модели можно использовать обычный МНК. Если же нулевая гипотеза отвергнута и поэтому, возможно, имеет место гетероскедастичность с негативными последствиями, то желательно получить более точные оценки, учитывающие гетероскедастичность.

Это можно сделать, используя для оценивания обобщенный МНК (см. уравнение (8.2)). Соответствующее преобразование в пространстве наблюдений состоит

в том, чтобы каждое наблюдение умножить на d_i, т.е. требуется оценить обычным методом наименьших квадратов преобразованную регрессию с переменными $d_i X_i$ и $d_i Z_i$. При этом не следует забывать, что если матрица факторов Z содержит свободный член, то его тоже нужно умножить на d_i, поэтому вместо свободного члена в регрессии появится переменная вида $(d_1, \ldots, d_N)$. Это приводит к тому, что стандартные статистические пакеты выдают неверные значения коэффициента детерминации и F-статистики. Чтобы этого не происходило, требуется пользоваться специализированными процедурами для расчета взвешенной регрессии. Описанный метод получил название **взвешенного МНК**, поскольку он равнозначен минимизации взвешенной суммы квадратов остатков $\sum\limits_{i=1}^{N} d_i^2 e_i^2$.

Чтобы это можно было осуществить, необходимо каким-то образом получить оценку матрицы D, используемой для преобразования в пространстве наблюдений. Перечисленные в этом параграфе методы дают возможность не только проверить гипотезу об отсутствии гетероскедастичности, но и получить определенные оценки матрицы D (возможно, не очень хорошие).

Если S^2 — оценка матрицы $\sigma^2\Omega$, где S^2 — диагональная матрица, составленная из оценок дисперсий, то S^{-1} (матрица, обратная к ее квадратному корню) — оценка матрицы σD.

Так, после проверки гомоскедастичности методом Глейзера в качестве диагональных элементов матрицы S^{-1} можно взять $1/|e_i|^c$, где $|e_i|^c$ — расчетные значения $|e_i|$. Если используются критерии Бартлетта или Голдфельда—Квандта, то наблюдения разбиваются на группы, для каждой из которых есть оценка дисперсии, s_l^2. Тогда для этой группы наблюдений в качестве диагональных элементов матрицы S^{-1} можно взять $1/s_l$.

В методе Голдфельда—Квандта требуется дополнительно получить оценку дисперсии для пропущенной средней части наблюдений. Эту оценку можно получить непосредственно по остаткам пропущенных налюдений или как среднее $(s_1^2+s_2^2)/2$.

Если точный вид гетероскедастичности неизвестен, и, как следствие, взвешенный МНК неприменим, то, по крайней мере, следует скорректировать оценку ковариационной матрицы оценок параметров, оцененных обычным МНК, прежде чем проверять гипотезы о значимости коэффициентов. (Хотя при использовании обычного МНК оценки будут менее точными, но как уже упоминалось, они будут несмещенными и состоятельными.) Простейший метод коррекции состоит в замене неизвестной ковариационной матрицы ошибок $\sigma^2\Omega$ на ее оценку S^2, где S^2 — диагональная матрица с типичным элементом e_i^2 (т.е. квадраты остатков используются как оценки дисперсий). Тогда получается следующая скорректированная оценка ковариационной матрицы a (**оценка Уайта** или **устойчивая к гетероскедастичности оценка**):

$$(Z'Z)^{-1} Z'S^2 Z (Z'Z)^{-1}.$$

8.3. Автокорреляция ошибок

Если матрица ковариаций ошибок не является диагональной, то говорят об **автокорреляции** ошибок. Обычно при этом предполагают, что наблюдения однородны по дисперсии, и их последовательность имеет определенный смысл и жестко фиксирована. Как правило, такая ситуация имеет место, если наблюдения проводятся в последовательные моменты времени. В этом случае можно говорить о зависимостях ошибок по наблюдениям, отстоящим друг от друга на 1, 2, 3 и т.д. момента времени. Обычно рассматривается частный случай автокорреляции, когда коэффициенты ковариации ошибок зависят только от расстояния во времени между наблюдениями; тогда возникает матрица ковариаций, в которой все элементы каждой диагонали (не только главной) одинаковы[1].

Поскольку действие причин, обуславливающих возникновение ошибок, достаточно устойчиво во времени, автокорреляции ошибок, как правило, положительны. Это ведет к тому, что значения остаточной дисперсии, полученные по стандартным («штатным») формулам, оказываются ниже их действительных значений. Что, как отмечалось и в предыдущем пункте, чревато ошибочными выводами о качестве получаемых моделей.

Это утверждение иллюстрируется рисунком 8.4 ($n = 1$).

На этом рисунке:

a — линия истинной регрессии. Если в первый момент времени истинная ошибка отрицательна, то в силу положительной автокорреляции ошибок все облако наблюдений сместится вниз, и линия оцененной регрессии займет положение b.

Если в первый момент времени истинная ошибка положительна, то по тем же причинам линия оцененной регрессии сместится вверх и займет положение c. Поскольку

[1] В теории временных рядов это называется слабой стационарностью.

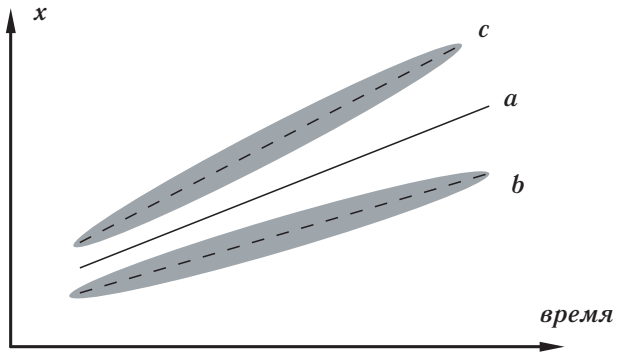

Рис. 8.4

ошибки случайны и в первый момент времени они примерно с равной вероятностью могут оказаться положительными или отрицательными, то становится ясно, насколько увеличивается разброс оценок регрессии вокруг истинных по сравнению с ситуацией без (положительной) автокорреляции ошибок.

Типичный случай автокорреляции ошибок, рассматриваемый в классической эконометрии, — это линейная авторегрессия ошибок первого порядка AR(1):

$$\varepsilon_i = \rho\varepsilon_{i-1} + \eta_i,$$

где η — остатки, удовлетворяющие обычным гипотезам;

ρ — коэффициент авторегрессии первого порядка.

Коэффициент ρ является также коэффициентом автокорреляции (первого порядка).

Действительно, по определению, коэффициент авторегрессии равен (как МНК-оценка):

$$\rho = \frac{\mathbf{cov}(\varepsilon_i, \varepsilon_{i-1})}{\mathbf{var}(\varepsilon_{i-1})},$$

но, в силу гомоскедастичности, $\mathbf{var}(\varepsilon_{i-1}) = \sqrt{\mathbf{var}(\varepsilon_i)\mathbf{var}(\varepsilon_{i-1})}$ и, следовательно, ρ, также по определению, является коэффициентом автокорреляции.

Если $\rho = 0$, то $\varepsilon_i = \eta_i$ и получаем «штатную» ситуацию. Таким образом, проверку того, что автокорреляция отсутствует, можно проводить как проверку нулевой гипотезы H_0: $\rho = 0$ для процесса авторегрессии 1-го порядка в ошибках.

Для проверки этой гипотезы можно использовать **критерий Дарбина—Уотсона** или **DW-критерий**. Проверяется нулевая гипотеза о том, что автокорреляция ошибок первого порядка отсутствует. (При автокорреляции второго и более высоких порядков его мощность может быть мала, и применение данного критерия становится ненадежным.)

Пусть была оценена модель регрессии и найдены остатки e_i, $i = 1, \ldots, N$. Значение статистики Дарбина—Уотсона (отношения фон Неймана), или **DW**-статистики, рассчитывается следующим образом:

$$d^c = \frac{\sum\limits_{i=2}^{N} (e_i - e_{i-1})^2}{\sum\limits_{i=1}^{N} e_i^2}. \tag{8.3}$$

Оно лежит в интервале от 0 до 4, в случае отсутствия автокорреляции ошибок приблизительно равно 2, при положительной автокорреляции смещается в мень-

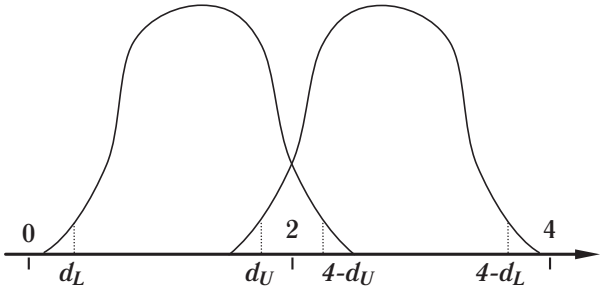

Рис. 8.5

шую сторону, при отрицательной — в большую сторону. Эти факты подтверждаются тем, что при больших N справедливо следующее соотношение:

$$d^c \approx 2(1-r), \tag{8.4}$$

где r — оценка коэффициента авторегрессии.

Минимального значения величина d^c достигает, если коэффициент авторегрессии равен $+1$. В этом случае $e_i = e,\ i = 1, \ldots, N$, и $d^c = 0$. Если коэффициент авторегрессии равен -1 и $e_i = (-1)^i e,\ i = 1, \ldots, N$, то величина d^c достигает значения $4\dfrac{N-1}{N}$ (можно достичь и более высокого значения подбором остатков), которое с ростом N стремится к 4. Формула (8.4) следует непосредственно из (8.3) после элементарных преобразований:

$$d^c = \frac{\sum\limits_{i=2}^{N} e_i^2}{\sum\limits_{i=1}^{N} e_i^2} - 2\frac{\sum\limits_{i=2}^{N} e_{i-1}e_i}{\sum\limits_{i=1}^{N} e_i^2} + \frac{\sum\limits_{i=2}^{N} e_{i-1}^2}{\sum\limits_{i=1}^{N} e_i^2},$$

поскольку первое и третье слагаемые при больших N близки к единице, а второе слагаемое является оценкой коэффициента автокорреляции (умноженной на -2).

Известно распределение величины d, если $\rho = 0$ (это распределение близко к нормальному), но параметры этого распределения зависят не только от N и n, как для t- и F-статистик при нулевых гипотезах. Положение «колокола» функции плотности распределения этой величины зависит от характера Z. Тем не менее, Дарбин и Уотсон показали, что это положение имеет две крайние позиции (рис. 8.5).

Поэтому существует по два значения для каждого (двустороннего) квантиля, соответствующего определенным N и n: его нижняя d_L и верхняя d_U границы. Нулевая гипотеза H_0: $\rho = 0$ принимается, если $d_U \leqslant d^c \leqslant 4 - d_U$; она отвергается в пользу гипотезы о положительной автокорреляции, если $d^c < d_L$, и в пользу

гипотезы об отрицательной автокорреляции, если $d^c > 4 - d_L$. Если $d_L \leqslant d^c < d_U$ или $4 - d_U < d^c \leqslant 4 - d_L$, вопрос остается открытым (это — зона неопределенности **DW**-критерия).

Пусть нулевая гипотеза отвергнута. Тогда необходимо дать оценку матрицы Ω.

Оценка r параметра авторегрессии ρ может определяться из приближенного равенства, следующего из (8.4):

$$r \approx 1 - \frac{d^c}{2},$$

или рассчитываться непосредственно из регрессии e на него самого со сдвигом на одно наблюдение с принятием «круговой» гипотезы, которая заключается в том, что $e_{N+1} = e_1$.

Оценкой матрицы Ω является

$$\frac{1}{1-r^2}\begin{bmatrix} 1 & r & r^2 & \cdots & r^{N-1} \\ r & 1 & r & \cdots & r^{N-2} \\ r^2 & r & 1 & \cdots & r^{N-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r^{N-1} & r^{N-2} & r^{N-3} & \cdots & 1 \end{bmatrix},$$

а матрица D преобразований в пространстве наблюдений равна

$$\begin{bmatrix} \sqrt{1-r^2} & 0 & 0 & \cdots & 0 \\ -r & 1 & 0 & \cdots & 0 \\ 0 & -r & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Для преобразования в пространстве наблюдений, называемом в данном случае авторегрессионным, используют обычно указанную матрицу без 1-й строки, что ведет к сокращению количества наблюдений на одно. В результате такого преобразования из каждого наблюдения, начиная со 2-го, вычитается предыдущее, умноженное на r, теоретическими остатками становятся η, которые, по предположению, удовлетворяют гипотезе $\mathbf{g4}$.

После этого преобразования снова оцениваются параметры регрессии. Если новое значение **DW**-статистики неудовлетворительно, то можно провести следующее авторегрессионное преобразование.

Обобщает процедуру последовательных авторегрессионных преобразований **метод Кочрена—Оркатта**, который заключается в следующем.

Для одновременной оценки r, a и b используется критерий ОМНК (в обозначениях исходной формы уравнения регрессии):

$$\frac{1}{N}\sum_{i=2}^{N}((x_i - rx_{i-1}) - (z_i - rz_{i-1})a - (1-r)b)^2 \to \min,$$

где z_i — n-вектор-строка значений независимых факторов в i-м наблюдении (i-строка матрицы Z).

Поскольку производные функционала по искомым величинам нелинейны относительно них, применяется итеративная процедура, на каждом шаге которой сначала оцениваются a и b при фиксированном значении r предыдущего шага (на первом шаге обычно $r = 0$), а затем — r при полученных значениях a и b. Процесс, как правило, сходится.

Как и в случае гетероскедастичности, можно не использовать модифицированные методы оценивания (тем более, что точный вид автокорреляции может быть неизвестен), а использовать обычный МНК и скорректировать оценку ковариационной матрицы параметров. Наиболее часто используемая **оценка Ньюи—Уэста** (устойчивая к гетероскедастичности и автокорреляции) имеет следующий вид:

$$(Z'Z)^{-1}\,Q\,(Z'Z)^{-1},$$

где

$$Q = \sum_{i=1}^{N} e_i^2 + \sum_{k=1}^{L}\sum_{i=k+1}^{N} \lambda_k e_i e_{i-k}(z_i z'_{i-k} + z_{i-k} z'_i),$$

а λ_k — понижающие коэффициенты, которые Ньюи и Уэст предложили рассчитывать по формуле $\lambda_k = 1 - \dfrac{k}{L+1}$. При $k > L$ понижающие коэффициенты становятся равными нулю, т.е. более дальние корреляции не учитываются.

Обоснование этой оценки достаточно сложно[2]. Заметим только, что если заменить попарные произведения остатков соответствующими ковариациями и убрать понижающие коэффициенты, то получится формула ковариационной матрицы оценок МНК.

Приведенная оценка зависит от выбора параметра отсечения L. В настоящее время не существует простых теоретически обоснованных методов для такого выбора. На практике можно ориентироваться на грубое правило $L = \left[4\left(T/100\right)^{2/9}\right]$.

[2] Оно связано с оценкой спектральной плотности для многомерного временного ряда.

8.4. Ошибки измерения факторов

Пусть теперь нарушается гипотеза **g2**, и независимые факторы наблюдаются с ошибками. Предполагается, что изучаемая переменная зависит от истинных значений факторов (далее в этом пункте используется сокращенная форма уравнения регрессии), $\hat{z}^0$, а именно:

$$\hat{x} = \hat{z}^0\alpha + \varepsilon,$$

но истинные значения неизвестны, а вместо этого имеются наблюдения над некоторыми связанными с $\hat{z}^0$ переменными $\hat{z}$:

$$\hat{z} = \hat{z}^0 + \varepsilon_z,$$

где ε_z — вектор-строка длиной n ошибок наблюдений.

В разрезе наблюдений:

$$\hat{X} = \hat{Z}^0\alpha + \varepsilon,$$
$$\hat{Z} = \hat{Z}^0 + \varepsilon_z,$$

где $\hat{Z}^0$ и ε_z — соответствующие $N \times n$-матрицы значений этих величин по наблюдениям (т.е., в зависимости от контекста, ε_z обозначает вектор или матрицу ошибок).

Предполагается, что ошибки факторов по математическому ожиданию равны нулю, истинные значения регрессоров и ошибки независимы друг от друга (по крайней мере не коррелированы друг с другом) и известны матрицы ковариации:

$$\begin{aligned} &\mathbf{E}(\varepsilon_z) = 0, && \mathbf{E}(\hat{z}^{0\prime}\varepsilon) = 0, && \mathbf{E}(\hat{z}^{0\prime}\varepsilon_z) = 0, \\ &\mathbf{E}(\hat{z}^{0\prime}\hat{z}^0) = M^0, && \mathbf{E}(\varepsilon_z'\varepsilon_z) = \Omega, && \mathbf{E}(\varepsilon_z'\varepsilon) = \omega. \end{aligned} \tag{8.5}$$

Важно отметить, что эти матрицы и вектора ковариации одинаковы во всех наблюдениях, а ошибки в разных наблюдениях не зависят друг от друга, т.е. речь, фактически, идет о «матричной» гомоскедастичности и отсутствии автокорреляции ошибок.

Через наблюдаемые переменные $\hat{x}$ и $\hat{z}$ уравнение регрессии записывается в следующей форме:

$$\hat{x} = \hat{z}\alpha + \varepsilon - \varepsilon_z\alpha. \tag{8.6}$$

В такой записи видно, что «новые» остатки не могут быть независимыми от факторов-регрессоров $\hat{z}$, т.е. гипотезы основной модели регрессии нарушены. В рамках

сделанных предположений можно доказать, что приближенно

$$\mathbf{E}(a) \approx (M^0 + \Omega)^{-1}(M^0\alpha + \omega) = \alpha + (M^0 + \Omega)^{-1}(\omega - \Omega\alpha), \qquad (8.7)$$

т.е. МНК-оценки теряют в такой ситуации свойства состоятельности и несмещенности[3], если $\omega \neq \Omega\alpha$ (в частности, когда ошибки регрессии и ошибки факторов не коррелированны, т.е. когда $\omega = 0$, а Ω и α отличны от нуля).

Для обоснования (8.7) перейдем к теоретическому аналогу системы нормальных уравнений, для чего обе части соотношения (8.6) умножаются на транспонированную матрицу факторов:

$$\mathbf{E}\left(\hat{z}'\hat{x}\right) = \mathbf{E}\left(\hat{z}'\hat{z}\right)\alpha + \mathbf{E}\left(\hat{z}'\varepsilon\right) - \mathbf{E}\left(\hat{z}'\varepsilon_z\right)\alpha.$$

Здесь, как несложно показать, пользуясь сделанными предположениями,

$$\begin{aligned}
&\mathbf{E}\left(\hat{z}'\hat{z}\right) = M^0 + \Omega,\\
&\mathbf{E}\left(\hat{z}'\varepsilon\right) = \omega,\\
&\mathbf{E}\left(\hat{z}'\varepsilon_z\right) = \Omega,
\end{aligned}$$

Поэтому

$$\mathbf{E}\left(\hat{z}'\hat{x}\right) = \mathbf{E}\left(\hat{z}'\hat{z}\right)\alpha + \omega - \Omega\alpha$$

или

$$\mathbf{E}\left(\hat{z}'\hat{z}\right)^{-1}\mathbf{E}\left(\hat{z}'\hat{x}\right) = \alpha + \left(M^0 + \Omega\right)^{-1}(\omega - \Omega\alpha).$$

Левая часть приближенно равна $\mathbf{E}(a)$.

Действительно, $a = M^{-1}m$, где $M = \frac{1}{N}\hat{Z}'\hat{Z}$ и $m = \frac{1}{N}\hat{Z}'\hat{x}$. Выборочные ковариационные матрицы M и m по закону больших чисел с ростом числа наблюдений сходятся по вероятности к своим теоретическим аналогам:

$$M \xrightarrow{p} \mathbf{E}\left(\hat{z}'\hat{z}\right) \quad \text{и} \quad m \xrightarrow{p} \mathbf{E}\left(\hat{z}'\hat{x}\right).$$

По свойствам сходимости по вероятности предел функции равен функции от предела, если функция непрерывна. Поэтому

$$a = M^{-1}m \xrightarrow{p} \mathbf{E}\left(\hat{z}'\hat{z}\right)^{-1}\mathbf{E}\left(\hat{z}'\hat{x}\right) = (M^0 + \Omega)^{-1}(M^0\alpha + \omega).$$

Существуют разные подходы к оценке параметров регрессии в случае наличия ошибок измерения независимых факторов. Здесь приводятся два из них.

[3] Они смещены даже асимптотически, т.е. при стремлении количества наблюдений к бесконечности смещение не стремится к нулю.

а) **Простая регрессия**. Если имеется оценка W ковариационной матрицы Ω и w — ковариационного вектора ω, то можно использовать следующий оператор оценивания:

$$a = (M - W)^{-1}(m - w),$$

который обеспечивает состоятельность оценок и делает их менее смещенными.

Это формула следует из

$$\mathbf{E}\,(\hat{z}'\hat{x}) = \mathbf{E}\,(\hat{z}'\hat{z})\,\alpha + \omega - \Omega\alpha$$

заменой теоретических моментов на их оценки.

Обычно предполагается, что W — диагональная матрица, а $w = 0$.

б) **Ортогональная регрессия**. Поскольку z теперь такие же случайные переменные, наблюдаемые с ошибками, как и x, имеет смысл вернуться к обозначениям 6-го раздела, где через x обозначался n-мерный вектор-строка всех переменных. Пусть ε — вектор их ошибок наблюдения, а x^0 — вектор их истинных значений, то есть

$$x = x^0 + \varepsilon,\ X = X^0 + \varepsilon.$$

Предположения (8.5) записываются следующим образом:

$$\mathbf{E}(\hat{x}^{0\prime},\varepsilon) = 0,\ \mathbf{E}(\hat{x}^{0\prime},\hat{x}^0) = M^0,\ \mathbf{E}(\varepsilon',\varepsilon) = \sigma^2\Omega.$$

Теперь через M^0 обозначается матрица, которую в обозначениях, используемых в этом пункте выше, можно записать следующим образом:

$$\begin{bmatrix} \sigma^2_{x^0} & m^0 \\ m^{0\prime} & M^0 \end{bmatrix},$$

а через $\sigma^2\Omega$ матрица

$$\begin{bmatrix} \sigma^2 & \omega \\ \omega' & \Omega \end{bmatrix}.$$

Поскольку речь идет о линейной регрессии, предполагается, что между истинными значениями переменных существует линейная зависимость:

$$x^0\alpha = 0.$$

Это означает, что

$$M^0\alpha = 0.$$

Рассуждая так же, как при доказательстве соотношения (8.7), легко установить, что

$$\mathbf{E}(M) = M^0 + \sigma^2\Omega,$$

(M — фактическая матрица ковариации X) т.е.

$$(\mathbf{E}(M) - \sigma^2\Omega)\alpha = 0.$$

Таким образом, если считать, что Ω известна, а σ^2 — минимизируемый параметр (в соответствии с логикой МНК), то решение задачи

$$(M - \sigma^2\Omega)a = 0,\ \sigma^2 \to \min!$$

даст несмещенную оценку вектора α. А это, как было показано в пункте 6.4, есть задача регрессии в метрике Ω^{-1} (см. (6.37)). Преобразованием в пространстве переменных она сводится к «обычной» ортогональной регрессии.

Т.е. если для устранения последствий нарушения гипотезы **g4** используется преобразование в пространстве наблюдений, то при нарушении гипотезы **g2** надо «работать» с преобразованием в пространстве переменных.

Несмотря на то, что методы ортогональной регрессии и регрессии в метрике Ω^{-1} в наибольшей степени соответствуют реалиям экономики (ошибки есть во всех переменных, стоящих как в левой, так и в правой частях уравнения регрессии), они мало используются в прикладных исследованиях. Основная причина этого заключается в том, что в большинстве случаев невозможно получить надежные оценки матрицы Ω. Кроме того, ортогональная регрессия гораздо сложнее простой с вычислительной точки зрения, и с теоретической точки зрения она существенно менее изящна и прозрачна.

В следующем параграфе излагается еще один метод, который позволяет решить проблему ошибок в переменных (и в целом может использоваться при любых нарушениях гипотезы **g2**).

8.5. Метод инструментальных переменных

Предполагаем, что в регрессии $x = z\alpha + \varepsilon$ переменные-факторы z являются случайными, и нарушена гипотеза **g2** в обобщенной формулировке: ошибка ε зависит от факторов z, так что корреляция между z и ошибкой ε не равна нулю. Такую

регрессию можно оценить, имея набор вспомогательных переменных y, называемых **инструментальными переменными**. Часто инструментальные переменные называют просто **инструментами**.

Для того, чтобы переменные y можно было использовать в качестве инструментальных, нужно, чтобы они удовлетворяли следующим требованиям:

1) Инструменты y некоррелированы с ошибкой ε. (В противном случае метод даст несостоятельные оценки, как и МНК.) Если это условие не выполнено, то такие переменные называют **негодными инструментами**[4].

2) Инструменты y достаточно сильно коррелированы с факторами z. Если данное условие не выполнено, то это так называемые **«слабые» инструменты**. Если инструменты слабые, то оценки по методу будут неточными и при малом количестве наблюдений сильно смещенными.

Обычно z и y содержат общие переменные, т.е. часть факторов используется в качестве инструментов. Например, типична ситуация, когда z содержит константу; тогда в y тоже следует включить константу.

Пусть имеются N наблюдений, и X, Z и Y — соответствующие данные в матричном виде. Оценки по **методу инструментальных переменных** (сокращенно IV от англ. instrumental variables) вычисляются по следующей формуле:

$$a_{IV} = \left(Z'Y\left(Y'Y\right)^{-1}Y'Z\right)^{-1} Z'Y\left(Y'Y\right)^{-1}Y'X. \tag{8.8}$$

В случае, если количество инструментальных переменных в точности равно количеству факторов, ($\operatorname{rank} Y = n+1$) получаем собственно классический метод инструментальных переменных. При этом матрица $Y'Z$ квадратная и оценки вычисляются как

$$a_{IV} = \left(Y'Z\right)^{-1}Y'Y\left(Z'Y\right)^{-1}Z'Y\left(Y'Y\right)^{-1}Y'X.$$

Средняя часть формулы сокращается, поэтому

$$a_{IV} = \left(Y'Z\right)^{-1}Y'X. \tag{8.9}$$

Рассмотрим вывод классического метода инструментальных переменных, т.е. случай точной идентификации (ср. с (6.15) в главе 6):

Умножим уравнение регрессии $x = z\alpha + \varepsilon$ слева на инструменты y (с транспонированием). Получим следующее уравнение:

$$y'x = y'z\alpha + y'\varepsilon.$$

[4] В модели ошибок в переменных ошибка регрессии имеет вид $\varepsilon - \varepsilon_z\alpha$, где ε — ошибка в исходном уравнении, а ε_z — ошибка измерения факторов z. Чтобы переменные y можно было использовать в качестве инструментов, достаточно, чтобы y были некоррелированы с ε и ε_z.

Если взять от обеих частей математическое ожидание, то получится

$$\mathbf{E}(y'x) = \mathbf{E}(y'z\alpha),$$

где мы учли, что инструменты некоррелированы с ошибкой, $\mathbf{E}(y'\varepsilon) = 0$.

Заменяя теоретические моменты на выборочные, получим следующие нормальные уравнения, задающие оценки a:

$$M_{yx} = M_{yz}a,$$

где $M_{yx} = \frac{1}{N}Y'X$ и $M_{yz} = \frac{1}{N}Y'Z$. Очевидно, что эти оценки совпадут с (8.9). Фактически, мы применяем здесь *метод моментов*.

Метод инструментальных переменных можно рассматривать как так называемый **двухшаговый метод наименьших квадратов**. (О нем речь еще пойдет ниже в пункте 10.3.)

1-й шаг. Строим регрессию каждого фактора Z_j на Y. Получим в этой регрессии расчетный значения Z_j^c. По формуле расчетных значений в регрессии $Z_j^c = Y\left(Y'Y\right)^{-1}Y'Z$. Заметим, что если Z_j входит в число инструментов, то по этой формуле получим $Z_j^c = Z_j$, т.е. эта переменная останется без изменений. Поэтому данную процедуру достаточно применять только к тем факторам, которые не являются инструментами (т.е. могут быть коррелированы с ошибкой). В целом для всей матрицы факторов можем записать $Z^c = Y\left(Y'Y\right)^{-1}Y'Z$.

2-й шаг. В исходной регрессии используются Z^c вместо Z. Смысл состоит в том, чтобы использовать факторы «очищенные от ошибок».

Получаем следующие оценки:

$$\begin{aligned} a_{2M} &= \left(Z^{c\prime}Z^c\right)^{-1}Z^{c\prime}x = \\ &= \left(Z'Y\left(Y'Y\right)^{-1}Y'Y\left(Y'Y\right)^{-1}Y'Z\right)^{-1}Z'Y\left(Y'Y\right)^{-1}Y'x = \\ &= \left(Z'Y\left(Y'Y\right)^{-1}Y'Z\right)^{-1}Z'Y\left(Y'Y\right)^{-1}Y'x = a_{IV}. \end{aligned}$$

Видим, что оценки совпадают.

Если записать оценки в виде $a_{IV} = \left(Z^{c\prime}Z\right)^{-1}Z^{c\prime}x$, то видно, что обобщенный метод инструментальных переменных можно рассматривать как простой метод инструментальных переменных с матрицей инструментов Z^c.

Такая запись позволяет обосновать обобщенный метод инструментальных переменных. Если исходных инструментов Y больше, чем факторов Z, и мы хотим построить на их основе меньшее количество инструментов, то имеет смысл сопоставить каждому фактору Z_j в качестве инструмента такую линейную комбинацию исходных инструментов, которая была бы *наиболее сильно коррелирована* с Z_j. Этому требованию как раз и удовлетворяют расчетные значения Z_j^c.

Другое обоснование обобщенного метода инструментальных переменных состоит, как и выше для классического метода, в использовании уравнений $\mathbf{E}(y'x) = \mathbf{E}(y'z\alpha)$. Заменой теоретических моментов выборочными получим уравнения $M_{yx} = M_{yz}a$, число которых больше числа неизвестных. Идея состоит в том, чтобы невязки $M_{yx} - M_{yz}a$ были как можно меньшими. Это достигается минимизацией следующей квадратичной формы от невязок:

$$(M_{yx} - M_{yz}a)' M_{yy}^{-1} (M_{yx} - M_{yz}a),$$

где $M_{yy} = \frac{1}{N} Y'Y$. Минимум достигается при

$$a = \left(M_{zy} M_{yy}^{-1} M_{yz} \right)^{-1} M_{zy} M_{yy}^{-1} M_{yx}.$$

Видим, что эта формула совпадает с (8.8). Эти рассуждения представляют собой применение так называемого **обобщенного метода моментов**, в котором количество условий на моменты может превышать количество неизвестных параметров.

Чтобы можно было использовать метод инструментальных переменных на практике, нужна оценка ковариационной матрицы, с помощью которой можно было бы вычислить стандартные ошибки коэффициентов и t-статистики. Такая оценка имеет вид

$$M_{a_{IV}} = s^2 \left(Z^{c\prime} Z^c \right)^{-1}.$$

Здесь s^2 — оценка дисперсии ошибок σ^2, например $s^2 = e'e/N$ или $s^2 = e'e/(N-1)$. Остатки рассчитываются по обычной формуле $e = x - Za_{IV}$. (Здесь следует помнить, что остатки, получаемые на втором шаге тут не годятся, поскольку они равны $x - Z^c a_{IV}$. Если использовать их для расчета оценки дисперсии, то получим заниженную оценку дисперсии и ковариационной матрицы. Отсюда следует, что из регрессии второго шага можно использовать только оценки коэффициентов. Стандартные ошибки и t-статистики требуется пересчитывать.)

Обсудим теперь более подробно проблему идентификации[5].

Чтобы можно было вычислить оценки (8.8), нужно, чтобы выполнялись следующие условия:

1) Матрица инструментов должна иметь полный ранг по столбцам, иначе $(Y'Y)^{-1}$ не существует.

2) $Z'Y\,(Y'Y)^{-1}\,Y'Z$ должна быть невырожденной.

В частности, матрица $Z'Y\,(Y'Y)^{-1}\,Y'Z$ необратима, когда $\operatorname{rank} Y < \operatorname{rank} Z$.

Предположим, что матрица факторов Z имеет полный ранг, т.е. $\operatorname{rank} Z = n+1$.

[5] См. также обсуждение идентификации в контексте систем уравнений ниже в пункте 10.2.

Т.е. если $\operatorname{rank} Y < n+1$, то уравнение **неидентифицируемо**, т.е. невозможно вычислить оценки (8.8). Таким образом, количество инструментов (включая константу) должно быть не меньше $n+1$ (количество регрессоров, включая константу). Если $\operatorname{rank} Y > n+1$, то говорят, что уравнение **сверхидентицировано**. Если количество инструментов равно $n+1$, то это **точная идентификация**.

Если возможен случай сверхидентификации, то это обобщенный метод инструментальных переменных. При точной идентификации ($\operatorname{rank} Y = n+1$) получаем собственно классический метод инструментальных переменных.

Таким образом, необходимое условие идентификации имеет следующий вид:

$$\operatorname{rank} Y \geqslant \operatorname{rank} Z (= n+1).$$

Это так называемое **порядковое условие** идентификации, условие на размерность матриц.

Словесная формулировка порядкового условия:

> Количество инструментов Y должно быть не меньше количества регрессоров Z (учитывая константу).

Заметим, что можно сначала «вычеркнуть» общие переменные в Z и Y и смотреть только на количество оставшихся. Количество оставшихся инструментов должно быть не меньше количества оставшихся регрессоров.

Почему это только необходимое условие? Пусть, например, некоторый фактор Z_j ортогонален Y. Тогда $Z_j^c = 0$, и невозможно получить оценки a_{IV}, т.е. данное условие не является достаточным.

Необходимое и достаточное условие идентификации формулируется следующим образом:

> Матрица Z^c имеет полный ранг по столбцам: $\operatorname{rank} Z^c = n+1$.

Это так называемое **ранговое условие** идентификации.

Встречаются случаи, когда ранговое условие идентификации соблюдается, но матрица Z^c близка к вырожденности, т.е. в Z^c наблюдается мультиколлинеарность. Например, если инструмент Z_j является слабым (Z_j и Y почти ортогональны), то Z^c близка к вырожденности. Один из способов проверки того, является ли инструмент слабым, состоит в анализе коэффициентов детерминации и F-статистик в регрессиях на первом шаге.

8.6. Упражнения и задачи

Упражнение 1

Дано уравнение регрессии $X = Z\alpha + \varepsilon = -1.410z_1 + 0.080z_2 + 56.962 + \varepsilon$, где ε — вектор-столбец нормальный случайных ошибок с нулевым средним и ковариационной матрицей

$$\mathbf{E}(\varepsilon\varepsilon') = \sigma^2\Omega = \frac{\sigma^2}{1-\rho^2}\begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{N-1} \\ \rho & 1 & \rho & \cdots & \rho^{N-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{N-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{N-1} & \rho^{N-2} & \rho^{N-3} & \cdots & 1 \end{bmatrix} \quad (8.10)$$

с $\rho = 0.9$ и $\sigma^2 = 21.611$.

Таблица 8.1

$\mathbf{Z}_1$	$\mathbf{Z}_2$	$\mathbf{1}$
26.8	541	1
25.3	616	1
25.3	610	1
31.1	636	1
33.3	651	1
31.2	645	1
29.5	653	1
30.3	682	1
29.1	604	1
23.7	515	1
15.6	390	1
13.9	364	1
18.8	411	1
27.4	459	1
26.9	517	1
27.7	551	1
24.5	506	1
22.2	538	1
19.3	576	1
24.7	697	1

Используя нормальное распределение с незасисимыми наблюдениями, средним 0 и ковариационной матрицей (8.10), получите 100 выборок вектора ε размерности $(N \times 1)$, $k = 1, \ldots, 100$, где $N = 20$. Эти случайные векторы потом используйте вместе с известным вектором $\alpha' = (-1.410,\ 0.080,\ 56.962)$ и матрицей регрессоров (табл. 8.1).

Сначала получите ожидаемое значения $X^0 = Z\alpha$, затем, чтобы получить 100 выборок вектора X размерности (20×1), добавьте случайные ошибки: $X^0 + \varepsilon = X$.

1.1. Рассчитайте невырожденную матрицу D такую, что $D^{-1}D'^{-1} = \Omega$.

1.2. Найдите истинную матрицу ковариации для МНК-оценки $(a = (Z'Z)^{-1}Z'X)$:

$$\mathbf{E}\left[(a-\alpha)(a-\alpha)'\right] = \mathbf{E}\left[(Z'Z)^{-1}Z'\varepsilon\varepsilon'Z(Z'Z)^{-1}\right] = \sigma^2(Z'Z)^{-1}Z'\Omega Z(Z'Z)^{-1}$$

и истинную матрицу ковариации для ОМНК-оценки
($a_{\text{омнк}} = \left(Z'\Omega^{-1}Z\right)^{-1} Z'\Omega^{-1}X$):

$$\mathbf{E}\left[\left(a_{\text{омнк}} - \alpha\right)\left(a_{\text{омнк}} - \alpha\right)'\right] = \sigma^2\left(Z'D'DZ\right) = \sigma^2\left(Z'\Omega^{-1}Z\right)^{-1}.$$

Результат поясните.

1.3. Используйте 10 из 100 выборок, чтобы посчитать по каждой выборке значения следующих оценок:

- МНК-оценки
  $a = \left(Z'Z\right)^{-1} Z'X$;
- ОМНК-оценки
  $a_{\text{омнк}} = \left(Z'\Omega^{-1}Z\right)^{-1} Z'\Omega^{-1}X$;
- МНК-оценки остаточной дисперсии
  $\hat{s}_e^2 = \dfrac{\left(x - Za\right)\left(x - Za\right)'}{N - n - 1}$;
- ОМНК-оценки остаточной дисперсии
  $\hat{s}_{ej\,\text{омнк}}^2 = \dfrac{\left(x - Za_{\text{омнк}}\right)\Omega^{-1}\left(x - Za_{\text{омнк}}\right)'}{N - n - 1}$.

Объясните результаты.

1.4. Вычислите среднее и дисперсию для 10 выборок для каждого из параметров, полученных в упражнении 1.3 и сравните эти средние значения с истинными параметрами.

1.5. На основе упражнения 1.3 рассчитайте $S_{a_1\,\text{омнк}}^2$, который является первым диагональным элементом матрицы $\hat{s}_{e\,\text{омнк}}^2\left(Z'\Omega^{-1}Z\right)^{-1}$ и $S_{a_1}^2$, который является первым диагональным элементом матрицы $\hat{s}_e^2\left(Z'Z\right)^{-1}$. Сравните различные оценки $S_{a_1}^2$ и $S_{a_1\,\text{омнк}}^2$ друг с другом и с соответствующими значениями из упражнения 1.2.

1.6. На основе результатов упражнений 1.3 и 1.5 рассчитайте значения t-статистики, которые могут быть использованы для проверки гипотез: $H_0:\ \alpha_1 = 0$.

1.7. Повторите упражнение 1.3 для всех 100 выборок, постройте распределения частот для оценок и прокомментируйте результаты.

Упражнение 2

Предположим, есть данные, состоящие из 100 выборок X, по 20 значений в каждой, сгенерированных при помощи модели $X = Z\alpha + \varepsilon = \alpha_1 z_1 + \alpha_2 z_2 + 1_N \beta + \varepsilon$, где ε_i — нормально и независимо распределенная случайная величина с $\mathbf{E}(\varepsilon_i) = 0$, $\mathbf{E}(\varepsilon_i^2) = \sigma_i^2$ и $\sigma_i^2 = e^{(\gamma_1 z_{i2} + \gamma_2)}$. Наблюдения за X были получены с использованием следующих значений параметров: $\alpha = (\alpha_1 \quad \alpha_2 \quad \beta)' = (-1.410, 0.080, 56.962)'$ и $\gamma = (\gamma_1 \quad \gamma_2)' = (0.25, -2)'$, а матрица значений факторов, упорядоченных в соответствии с величиной z_2, имеет следующий вид (табл. 8.2).

Таблица 8.2

z_1	z_2	1_N
13,9	364	1
15,6	390	1
18,8	411	1
27,4	459	1
24,5	506	1
23,7	515	1
26,9	517	1
22,2	538	1
26,8	541	1
27,7	551	1
19,3	576	1
29,1	604	1
25,3	610	1
25,3	616	1
31,1	636	1
31,2	645	1
33,3	651	1
29,5	653	1
30,3	682	1
24,7	697	1

2.1. Найдите матрицу ковариации для

- ОМНК-оценки $a_{\text{омнк}} = (Z'\Omega^{-1}Z)^{-1} Z'\Omega^{-1}X$;
- МНК-оценки $a = (Z'Z)^{-1} Z'X$.

Что вы можете сказать об относительной эффективности этих оценок?

2.2. Используйте 10 из 100 выборок, чтобы посчитать по каждой выборке значения следующих оценок:

- МНК-оценки $a = (Z'Z)^{-1} Z'X$;
- оценки $\gamma = \left(\sum_{i=1}^{N} y_i y_i'\right)^{-1} \sum_{i=1}^{N} y_i \ln(e_i^2)$, где $y_i = (z_{i2}, 1)$ и $e_i = x_i - z_i' a$;
- ОМНК-оценки a, используя найденную оценку γ.

Сравните все эти оценки друг с другом и с соответствующими истинными значениями.

2.3. На основе упражнения 2.2 рассчитайте $S^2_{a_1\ \text{омнк}}$, который является первым диагональным элементом матрицы $\hat{s}^2_{e\,\text{омнк}}(Z'\Omega^{-1}Z)^{-1}$, $S^2_{a_1}$, который является первым диагональным элементом матрицы $\hat{s}_e^2 (Z'Z)^{-1}$, а также $S^2_{a_1\ \text{Уайта}}$, который является первым диагональным элементом скорректированной оценки ковариационной матрицы (оценка Уайта или устойчивая к гетероскедастичности оценка). Сравните различные оценки $S^2_{a_1}$, $S^2_{a_1\ \text{омнк}}$ и $S^2_{a_1\ \text{Уайта}}$ друг с другом и с соответствующими значениями из упражнения 2.1.

2.4. На основе результатов упражнений 2.1 и 2.3 рассчитайте значения t-статистики, которые могут быть использованы для проверки гипотез $H_0: \alpha_1 = 0$.

2.5. Возьмите те же выборки, что и в упражнении 2.2, и проведите проверку на гетероскедастичность с помощью:

- **критерия Бартлета**;
- метода **второй** группы (метод Голдфельда—Квандта) с пропуском 4-х значений в середине выборки;
- метода **третьей** группы (метод Глейзера).

2.6. Выполните упражнение 2.2 для всех 100 выборок и, используя результаты, оцените математическое ожидание и матрицу среднеквадратических ошибок для каждой оценки. Есть ли среди оценок смещенные? Что можно сказать об относительной эффективности МНК-оценки и ОМНК-оценки?

Упражнение 3

Предположим, есть данные, состоящие из 100 выборок X, по 20 значений в каждой, сгенерированных при помощи модели $X = Z\alpha + \varepsilon = \alpha_1 z_1 + \alpha_2 z_2 + 1_N \beta + \varepsilon$, где $\varepsilon_i = \rho\varepsilon_{i-1} + \eta_i$, и η — нормально распределенная случайная величина с $\mathbf{E}(\eta_i) = 0$, $\mathbf{E}(\eta_i^2) = \sigma_\eta^2$. Наблюдения за X были получены с использованием следующих значений параметров: $\alpha' = (\alpha_1 \quad \alpha_2 \quad \beta) = (-1.410, 0.080, 56.962)$, $\rho = 0.8$ и $\sigma_\eta^2 = 6.4$, а матрица значений факторов взята из упражнения 1.

3.1. Найдите матрицу ковариации для:

- ОМНК-оценки $a_{\text{омнк}} = (Z'\Omega^{-1}Z)^{-1} Z'\Omega^{-1}X$;
- МНК-оценки $a = (Z'Z)^{-1} Z'X$.

Что вы можете сказать об относительной эффективности этих оценок?

3.2. Используйте 10 из 100 выборок, чтобы посчитать по каждой выборке значения следующих оценок:

- МНК-оценки $a = (Z'Z)^{-1} Z'X$;
- оценку $r = \dfrac{\sum\limits_{i=2}^{N} e_i e_{i-1}}{\sum\limits_{i=1}^{N} e_i^2}$;
- ОМНК-оценки, используя найденную оценку r.

Сравните все эти оценки друг с другом и с соответствующими истинными значениями.

3.3. Возьмите те же выборки, что и в упражнении 3.2, и проверьте гипотезу об автокорреляции ошибок.

3.4. Найдите скорректированную оценку ковариационной матрицы, устойчивую к гетероскедастичности и автокорреляции (оценку Ньюи—Уэста).

3.5. Выполните упражнение 3.2 для всех 100 выборок и, используя результаты, оцените математическое ожидание и матрицу среднеквадратических ошибок для каждой оценки. Есть ли среди оценок смещенные? Что можно сказать об относительной эффективности МНК-оценки и ОМНК-оценки?

Упражнение 4

Для уравнения $X = Z^o\alpha+\varepsilon = -1.410z_1^0+0.080z_2^0+1_N56.962+\varepsilon$, $z_1 = z_1^0+\varepsilon_{z_1}$, $z_2 = z_2^0 + \varepsilon_{z_2}$ и при предположении, что $\varepsilon_i \sim N(0, 21.611)$, $\varepsilon_{z_1} \sim N(0, 21.700)$ и $\varepsilon_{z_2} \sim N(0, 21.800)$, были генерированы 20 значений выборки. Результаты приведены в таблице 8.3.

Предполагая, что истинная матрица факторов Z^0 неизвестна, выполните следующие задания:

4.1. Найдите МНК-оценки $a = (Z'Z)^{-1}Z'X$ параметров уравнения регрессии $X = Z\alpha + \varepsilon = \alpha_1 z_1 + \alpha_2 z_2 + 1_N\beta + \varepsilon$.

4.2. Рассчитайте ковариационную матрицу ошибок измерения факторов — W и ковариационный вектор — w и оцените параметры регрессии как $a = (M - W)^{-1}(m - w)$.

4.3. Найдите оценку через ортогональную регрессию.

4.4. Сравните эти все оценки друг с другом и с соответствующими истинными значениями.

Задачи

1. Какие свойства МНК-оценок коэффициентов регрессии теряются, если ошибки по наблюдениям коррелированы и/или имеют разные дисперсии?

2. Как оцениваются параметры уравнения регрессии, если известна матрица ковариации ошибок и она не диагональна с равными элементами по диагонали?

Таблица 8.3

N	ε	ε_{z_1}	ε_{z_2}	z_1^0	z_2^0	z_1	z_2	X
1	26.19	1.96	37.94	13.9	364	15.86	401.94	92.67
2	6.94	–5.94	3.57	15.6	390	9.66	393.57	73.10
3	5.55	–13.85	–18.78	18.8	411	4.95	392.22	68.88
4	14.00	24.48	14.49	27.4	459	51.88	473.49	69.05
5	0.89	23.91	51.48	24.5	506	48.41	557.48	63.79
6	46.61	–32.80	10.99	23.7	515	–9.10	525.99	111.36
7	–20.52	13.27	11.07	26.9	517	40.17	528.07	39.87
8	10.15	–16.17	18.86	22.2	538	6.03	556.86	78.85
9	–13.95	–28.22	–18.57	26.8	541	–1.42	522.43	48.50
10	14.94	20.64	–10.89	27.7	551	48.34	540.11	76.92
11	19.38	–36.99	–0.91	19.3	576	–17.69	575.09	95.21
12	5.72	–32.44	–12.71	29.1	604	–3.34	591.29	69.97
13	1.08	25.91	7.70	25.3	610	51.21	617.70	71.17
14	11.07	10.90	9.24	25.3	616	36.20	625.24	81.64
15	5.81	–42.77	8.25	31.1	636	–11.67	644.25	69.80
16	27.21	25.63	–29.14	31.2	645	56.83	615.86	91.78
17	–11.63	–13.07	13.20	33.3	651	20.23	664.20	50.46
18	–4.24	10.27	–37.62	29.5	653	39.77	615.38	63.37
19	46.56	44.81	33.93	30.3	682	75.11	715.93	115.36
20	–7.57	–40.10	–6.34	24.7	697	–15.40	690.66	70.32

3. Рассматривается регрессионная модель $X = Z\alpha + \varepsilon$. Пусть $\alpha^* = AX$ — это любая несмещенная оценка параметра α. Полагая, что $\mathbf{E}(\varepsilon\varepsilon') = \sigma^2\Omega$, покажите, что матрица ковариации α^* превышает матрицу ковариации $\alpha_{\text{омнк}} = (Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}X$ на какую-то положительно полуопределенную матрицу.

4. Докажите, что $\sigma^2_{\text{омнк}} = \dfrac{(x - z\alpha)'\,\Omega^{-1}\,(x - z\alpha)}{N - n - 1}$ есть оценка σ^2.

5. Какое преобразование матрицы наблюдений перед оценкой регрессии полезно сделать, если среднеквадратические отклонения ошибок регрессии пропорциональны какому-либо фактору?

6. Оценивается регрессия по 10 наблюдениям. Известно, что дисперсия ошибок для первых 5 наблюдений в два раза больше, чем дисперсия ошибок остальных 5 наблюдений. Опишите процедуру оценивания этой регрессии.

7. Рассмотрите регрессию $x_t = \alpha_1 t + \beta + \varepsilon_t,\ t = 1, \ldots, 5,$ где $\mathbf{E}(\varepsilon_t) = 0,\ \mathbf{E}(\varepsilon_t^2) = \sigma^2 t^2,\ \mathbf{E}(\varepsilon_t\varepsilon_s) = 0,$ при $t \neq s$.
   Пусть $\varepsilon' = (\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5)$ и $\mathbf{E}(\varepsilon\varepsilon') = \sigma^2\Omega$.

   – определите Ω;
   – найдите Ω^{-1};
   – найдите матрицу ковариации МНК-оценки параметра $\alpha = \begin{pmatrix} \alpha_1 \\ \beta \end{pmatrix}$;
   – найдите матрицу ковариации ОМНК-оценки параметра $\alpha = \begin{pmatrix} \alpha_1 \\ \beta \end{pmatrix}$.

8. Рассмотрите регрессию $x_t = \alpha_1 t + \varepsilon_t,\ t = 1, \ldots, 5,$
   где $\mathbf{E}(\varepsilon_t) = 0,\ \mathbf{E}(\varepsilon_t^2) = \sigma^2 t^2\ ,\ \mathbf{E}(\varepsilon_t\varepsilon_s) = 0,\ t \neq s$. Если $x = (6,\ 4,\ 9,\ 8,\ 7)'$:

   – определите оценку МНК для α_1 и ее дисперсию;
   – определите оценку ОМНК для α_1 и ее дисперсию;
   – сравните эти оценки друг с другом и сделайте вывод.

9. Рассматривается модель $X = Z\alpha + \varepsilon$, где ε_i — нормально и независимо распределенная случайная величина с $\mathbf{E}(\varepsilon_i) = 0$ и $\mathbf{E}(\varepsilon_i^2) = \sigma_i^2 = e^{y_i\gamma}$.

В предположении, что $X = \begin{pmatrix} 4 \\ 8 \\ 6 \\ 2 \\ 9 \end{pmatrix}$, $Z = \begin{pmatrix} 2 & 1 \\ 5 & 1 \\ 2 & 1 \\ 1 & 1 \\ 10 & 1 \end{pmatrix}$, $Y = \begin{pmatrix} 2 & 1 \\ 3 & 1 \\ 1 & 1 \\ 0 & 1 \\ 2 & 1 \end{pmatrix}$,

- найдите МНК-оценки $a = (Z'Z)^{-1} ZX$;
- найдите ОМНК-оценки $a_{\text{омнк}} = \left(Z'\Omega^{-1}Z\right)^{-1} Z'\Omega^{-1}X$;
- постройте два 95%-х доверительных интервала для α_1: один неправильный, основанный на результатах МНК, а другой правильный, основанный на результатах ОМНК;
- проверьте гипотезу $\gamma_1 = 0$.

10. Параметры трехфакторного уравнения регрессии оценены по 20 наблюдениям. S_1 и S_2 — остаточные дисперсии по первой и второй половинам временного ряда. В каком случае гипотезы о гомоскедастичности следует отвергнуть?

11. Приведите примеры графиков зависимостей ошибки от времени в авторегресионной схеме первого порядка для случаев, когда модуль коэффициента авторегрессии превышает единицу. Что можно сказать об автокорреляции ошибок, если этот коэффициент равен нулю?

12. Ошибка в регрессии задана процессом $\varepsilon_i = 0.6\varepsilon_{i-1} + \eta_i$, и η — нормально распределенная случайная величина с $\mathbf{E}(\eta_i) = 0$, $\mathbf{E}(\eta_i^2) = \sigma_\eta^2$ и $i = 1, \ldots, 5$. Как выглядит матрица преобразования в пространстве переменных для ОМНК?

13. Проверьте, что $D'D = \Omega^{-1}$, где

$$D = \begin{bmatrix} \sqrt{1-r^2} & 0 & 0 & \cdots & 0 \\ -r & 1 & 0 & \cdots & 0 \\ 0 & -r & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix},$$

$$\Omega = \frac{1}{1-r^2}\begin{bmatrix} 1 & r & r^2 & \cdots & r^{N-1} \\ r & 1 & r & \cdots & r^{N-2} \\ r^2 & r & 1 & \cdots & r^{N-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r^{N-1} & r^{N-2} & r^{N-3} & \cdots & 1 \end{bmatrix}.$$

14. Найдите $D_0'D_0$, где D_0 — это матрица размерности $(N-1)\times N$, полученная из матрицы D путем удаления первой строки, и сравните ее с матрицей Ω^{-1}.

15. Какое преобразование матрицы наблюдений перед оценкой регрессии полезно сделать, если ошибки в каждом наблюдении имеют одинаковую дисперсию и коррелированы с ошибками в предыдущем наблюдении?

16. Почему при использовании критерия Дарбина—Уотсона требуется знать два критических значения для расчетной статистики?

17. Фактическое значение d^c статистики Дарбина—Уотсона равно 0.5. Что это означает? Какое преобразование следует применить к этой модели (запишите формулу)?

18. В регрессионной модели $X = Z\alpha + \varepsilon$ существует автокорреляции ошибок первого порядка и $\rho = 0.6$. Предположим, что

$$X = \begin{pmatrix} 4 \\ 8 \\ 6 \\ 2 \\ 9 \end{pmatrix}, \quad Z = \begin{pmatrix} 2 & 1 \\ 5 & 1 \\ 2 & 1 \\ 1 & 1 \\ 10 & 1 \end{pmatrix},$$

   - найдите преобразованные наблюдения Dx и Dz;
   - найдите ОМНК-оценки параметра α;
   - найдите фактическое значение d^c статистики Дарбина—Уотсона по остаткам после применения ОМНК.

19. Положим, построили регрессию для $N = 20$ и $n = 4$ и нашли оценку

$$z = \frac{\sum\limits_{i=2}^{N} e_i e_{i-1}}{\sum\limits_{i=1}^{N} e_i^2} = 0.5,\ e'e = 40,\ e_1^2 = 1,\ e_N^2 = 4.$$

Найдите фактическое значение d^c статистики Дарбина—Уотсона и с ее помощью проведите тест на автокорреляцию.

20. На основе годовых данных 1959–1983 годов были оценены следующие функции спроса на продовольственные товары.

$$\begin{array}{rcccccc} \ln Q_t & = & 2.83 & - & 0.47 \ln PF_t & + & 0.64 \ln Y_t, \\ & & (6.69) & & (-3.94) & & (24.48) \end{array}$$
$$R^2 = 0.987, \quad DW = d^c = 0.627,$$

$$\begin{array}{rcccccccc} \ln Q_t & = & 1.87 & - & 0.36 \ln PF_t & + & 0.38 \ln Y_t & + & 0.44 Q_{t-1}, \\ & & (3.24) & & (-2.79) & & (3.20) & & (24.10) \end{array}$$
$$R^2 = 0.990, \quad DW = d^c = 1.65,$$

где Q — спрос на продукты питания, PF — цены на продукты питания, Y — доход, в скобках приведены значения t-статистики.

Проверьте каждое уравнение на наличие автокорреляции первого порядка и дайте короткий комментарий результатов.

21. Пусть остатки в регрессии $x_i = \alpha + \beta z_i + \varepsilon_i$ равны $(1,\ 2,\ 0,\ -1,\ -2)'$. Опишите первый шаг метода Кочрена—Оркарта.

22. Денежная масса измеряется с ошибкой. Как смещен коэффициент зависимости динамики цен от динамики денежной массы относительно его истинного значения?

23. Пусть в парной линейной регрессии ошибки зависимой переменной и фактора независимы и имеют одинаковую дисперсию. Запишите задачу для нахождения оценок коэффициентов данной регрессии (с объяснением обозначений).

Рекомендуемая литература

1. **Айвазян С.А.** Основы эконометрики. Т.2. — М.: Юнити, 2001. (Гл. 2)

2. **Демиденко Е.З**. Линейная и нелинейная регрессия. — М.: «Финансы и статистика», 1981. (Гл. 1).

3. **Джонстон Дж.** Эконометрические методы. — М.: «Статистика», 1980. (Гл. 7, 8).

4. **Доугерти К.** Введение в эконометрику. — М.: «Инфра-М», 1997. (Гл. 7).

5. **Дрейпер Н., Смит Г.** Прикладной регрессионный анализ: В 2-х книгах. Кн.1 — М.: «Финансы и статистика», 1986. (Гл. 2, 3).

6. **Кейн Э.** Экономическая статистика и эконометрия. — М.: «Статистика», 1977. Вып. 2. (Гл. 15).

7. **Лизер С.** Эконометрические методы и задачи. — М.: «Статистика», 1971. (Гл. 2).

8. **Магнус Я.Р., Катышев П.К., Пересецкий А.А.** Эконометрика — начальный курс. — М.: Дело, 2000. (Гл. 6, 7, 9).

9. **Маленво Э.** Статистические методы эконометрии. Вып. 1. — М.: «Статистика», 1975. (Гл. 10).

10. **Тинтер Г.** Введение в эконометрию. — М.: «Статистика», 1965. (Гл. 6).

11. **Baltagi, Badi H.** Econometrics, 2nd edition, Springer, 1999. (Ch. 5).

12. **Davidson, Russel, Mackinnon, James.** Estimation and Inference in Econometrics, N 9, Oxford University Press, 1993. (Ch. 16).

13. **William E., Griffiths R., Carter H., George G. Judge** Learning and Practicing econometrics, N 9 John Wiley & Sons, Inc., 1993. (Ch. 9, 15, 16).

14. **Greene W.H.** Econometric Analysis, Prentice-Hall, 2000. (Ch. 9, 12, 13).

15. **Judge G.G., Hill R.C., Griffiths W.E., Luthepohl H., Lee T.** Introduction to the Theory and Practice of Econometric. John Wiley & Sons, Inc., 1993. (Ch 8, 9).

16. **Maddala G.S.** Introduction to Econometrics, 2nd ed., Prentice Hall, 1992. (Ch. 5, 6, 7).

Глава 9

Целочисленные переменные в регрессии

9.1. Фиктивные переменные

С помощью фиктивных или псевдопеременных, принимающих дискретные, обычно целые значения, в регрессию включают качественные факторы.

Уточнение обозначений:

Z — $N \times n$-матрица наблюдений за «обычными» независимыми факторами;

α — n-вектор-столбец параметров регрессии при этих факторах;

$$Z^0 = 1_N; \quad \beta^0 = \beta.$$

В этих обозначениях уравнение регрессии записывается следующим образом:

$$X = Z\alpha + Z^0\beta^0 + \varepsilon.$$

Пусть имеется один качественный фактор, принимающий два значения (например: «мужчина» и «женщина», если речь идет о модели некоторой характеристики отдельных людей, или «годы войны» и «годы мира» — в модели, построенной на временных рядах наблюдений, которые охватывают периоды войны и мира и т.д.). Ставится вопрос о том, влияет ли этот фактор на значение свободного члена регрессии.

$\widetilde{Z}^G = \{z_{ij}^G\}$ — $N\times 2$-матрица наблюдений за качественным фактором (матрица фиктивных переменных): z_{i1}^G равен единице, если фактор в i-м наблюдении принимает первое значение, и нулю в противном случае; z_{i2}^G равен единице, если фактор в i-м наблюдении принимает второе значение, и нулю в противном случае.

$$\widetilde{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

— двухкомпонентный вектор-столбец параметров при фиктивных переменных.

Исходная форма регрессии с фиктивными переменными:

$$X = Z\alpha + Z^0\beta^0 + \widetilde{Z}^G\widetilde{\beta} + \varepsilon.$$

Поскольку сумма столбцов матрицы равна Z^0, оценка параметоров непосредственно по этому уравнению невозможна.

Проводится преобразование фиктивных переменных одним из двух способов.

а) В исходной форме регрессии исключается один из столбцов матрицы фиктивных переменных, в данном случае — первый.

$\bar{Z}^G$ — матрица фиктивных переменных без первого столбца;

$$\bar{C} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix}.$$

Тогда эквивалентная исходной запись уравнения имеет вид:

$$X = Z\alpha + \left[Z^0, \bar{Z}^G\right]\bar{C}\begin{bmatrix} \beta^0 \\ \widetilde{\beta} \end{bmatrix} + \varepsilon,$$

и после умножения матрицы $\bar{C}$ справа на вектор параметров получается запись уравнения регресии, в которой отсутствует линейная зависимость между факторами-регрессорами:

$$X = Z\alpha + Z^0\bar{\beta}^0 + \bar{Z}^G\bar{\beta} + \varepsilon,$$

где $\bar{\beta}^0 = \beta^0 + \beta_1,\ \bar{\beta} = \beta_2 - \beta_1$.

После оценки этих параметров можно определить значения исходных параметров β^0 и $\widetilde{\beta}$, предполагая, что сумма параметров при фиктивных переменных

(в данном случае $\beta_1+\beta_2$) равна нулю, т.е. влияние качественного фактора приводит к колебаниям вокруг общего уровня свободного члена:

$$\beta_2 = \bar{\beta}/2,\ \beta_1 = -\beta_2,\ \beta^0 = \bar{\beta}^0 + \beta_2.$$

б) Предполагая, что сумма параметров при фиктивных переменных равна нулю, в исходной форме регрессии исключается один из этих параметров, в данном случае — первый.

β — вектор-стобец параметров при фиктивных переменных без первого элемента;

$$C = \begin{bmatrix} -1 \\ \\ 1 \end{bmatrix}.$$

Эквивалентная исходной запись уравнения принимает форму:

$$X = Z\alpha + Z^0\beta^0 + \widetilde{Z}^G C\beta + \varepsilon,$$

и после умножения матрицы C слева на матрицу наблюдений за фиктивными переменными получается запись уравнения регрессии, в которой также отсутствует линейная зависимость между регрессорами:

$$X = Z\alpha + Z^0\beta^0 + Z^G\beta + \varepsilon.$$

После оценки параметров этого уравнения недостающая оценка параметра β определяется из условия $\beta_1 = -\beta_2$.

Качественный фактор может принимать больше двух значений. Так, в классической модели выделения сезонных колебаний он принимает 4 значения, в случае поквартальных наблюдений, и 12 значений, если наблюдения проводились по месяцам. Матрица $\widetilde{Z}^G$ в этой модели имеет размерность, соответственно, $N\times 4$ или $N\times 12$.

Пусть в общем случае качественный фактор принимает k значений. Тогда: матрица $\widetilde{Z}^G$ имеет размерность $N\times k$, вектор-столбец $\widetilde{\beta}$ — размерность k, матрицы $\bar{Z}^G$ и Z^G — $N\times(k-1)$, вектор-столбцы $\bar{\beta}$ и β — $(k-1)$;

$k\times(k+1)$ матрица $\bar{C} = \begin{bmatrix} 1 & 1 & 0 \\ \\ 0 & -1_{k-1} & I_{k-1} \end{bmatrix}$;

$k\times(k-1)$ матрица $C = \begin{bmatrix} -1'_{k-1} \\ \\ I_{k-1} \end{bmatrix}$;

$$1'_k\widetilde{\beta} = 0, \quad \bar{C}\begin{bmatrix}\beta^0\\ \widetilde{\beta}\end{bmatrix} = \begin{bmatrix}\bar{\beta}^0\\ \bar{\beta}\end{bmatrix}, \quad \widetilde{Z}^G C = Z^G.$$

Можно показать, что

$$\begin{bmatrix}1 & -1'_{k-1}\\ 0 & I_{k-1} - 1^{k-1}\end{bmatrix}\begin{bmatrix}\beta^0\\ \beta\end{bmatrix} = \begin{bmatrix}\bar{\beta}^0\\ \bar{\beta}\end{bmatrix}, \text{ или}$$

$$\begin{bmatrix}1 & 1'_{k-1}(I_{k-1} - \frac{1}{k}1^{k-1})\\ 0 & I_{k-1} - \frac{1}{k}1^{k-1}\end{bmatrix}\begin{bmatrix}\bar{\beta}^0\\ \bar{\beta}\end{bmatrix} = \begin{bmatrix}\beta^0\\ \beta\end{bmatrix},$$

где $1^{k-1} = 1_{k-1}1'_{k-1}$ — $(k-1)\times(k-1)$-матрица, состоящая из единиц; и далее показать, что результаты оценки параметров уравнения с фиктивными переменными при использовании обоих указанных подходов к устранению линейной зависимости факторов-регрессоров одинаковы.

В дальнейшем для устранения линейной зависимости столбцов значений фиктивных переменных используется способ «б».

После оценки регрессии можно применить t-критерий для проверки значимости влияния качественного фактора на свободный член уравнения.

Если k слишком велико и приближается к N, то на параметры при фиктивных переменных накладываются более жесткие ограничения (чем равенство нулю их суммы). Так, например, если наблюдения проведены в последовательные моменты времени, и вводится качественный фактор «время», принимающий особое значение в каждый момент времени, то $\widetilde{Z}^G = I_N$, и обычно предполагается, что значение параметра в каждый момент времени (при фиктивной переменной каждого момента времени) больше, чем в предыдущий момент времени на одну и ту же величину. Тогда роль матрицы C играет N-вектор-столбец T, состоящий из чисел натурального ряда, начиная с 1, и $\widetilde{\beta} = T\beta_T$, где β_T — скаляр. Уравнение регрессии с фактором времени имеет вид (эквивалентная исходной форма уравнения при использовании способа «б» исключения линейной зависимости фиктивных переменных):

$$X = Z\alpha + Z^0\beta^0 + T\beta_T + \varepsilon.$$

Метод фиктивных переменных можно использовать для проверки влияния качественного фактора на коэффициент регрессии при любом обычном факторе. Исходная форма уравнения, в которое вводится качественный фактор для параметра α, имеет следующий вид:

$$X = Z\alpha + Z^0\beta^0 + Z_j\bar{\otimes}\widetilde{Z}^G\widetilde{\alpha}^j + \varepsilon,$$

где Z_j — j-й столбец матрицы Z; $\tilde{\alpha}^j$ — k-вектор-столбец параметров влияния качественного фактора на α_j; в векторе α j-я компонента теперь обозначается α_j^0 — средний уровень параметра α_j; $\bar{\otimes}$ — операция прямого произведения столбцов матриц.

Прямое произведение матриц $A \otimes B$ (произведение Кронекера, см. Приложение А.1.2), имеющих размерность, соответственно, $m_A \times n_A$ и $m_B \times n_B$, есть матрица размерности $(m_A m_B) \times (n_A n_B)$ следующей структуры:

$$\begin{bmatrix} a_{11}B & \cdots & a_{1n_A}B \\ \vdots & \ddots & \vdots \\ a_{m_A1}B & \cdots & a_{m_An_A}B \end{bmatrix}.$$

Прямое произведение матриц обладает следующими свойствами:

$$(A_1 \otimes \cdots \otimes A_m)(B_1 \otimes \cdots \otimes B_2) = (A_1B_1) \otimes \cdots \otimes (A_mB_m),$$

если, конечно, соответствующие матричные произведения имеют смысл:

$$(A_1 \otimes \cdots \otimes A_m)' = A_1' \otimes \cdots \otimes A_m',$$
$$(A_1 \otimes \cdots \otimes A_m)^{-1} = A_1^{-1} \otimes \cdots \otimes A_m^{-1},$$

если все матрицы A квадратны и неособенны.

Прямое произведение столбцов матриц применимо к матрицам, имеющим одинаковое число строк, и осуществляется путем проведения операции прямого произведения последовательно с векторами-строками матриц:

$$A \bar{\otimes} B = \begin{bmatrix} A_1 \\ \vdots \\ A_m \end{bmatrix} \bar{\otimes} \begin{bmatrix} B_1 \\ \vdots \\ B_m \end{bmatrix} = \begin{bmatrix} A_1 \otimes B_1 \\ \vdots \\ A_m \otimes B_m \end{bmatrix}.$$

Эта операция обладает следующим важным свойством:

$$(A_1 \bar{\otimes} \cdots \bar{\otimes} A_m)(B_1 \otimes \cdots \otimes B_2) = (A_1B_1) \bar{\otimes} \cdots \bar{\otimes} (A_mB_m).$$

Приоритет прямого произведения матриц выше, чем обычного матричного произведения.

При использовании способа «а» эквивалентная исходной форма уравнения имеет вид (форма «а»):

$$X = Z_{-j}\alpha_{-j} + Z^0\beta^0 + Z_j \bar{\otimes} \left[Z^0, \bar{Z}^G\right] \bar{C} \begin{bmatrix} \alpha_j^0 \\ \tilde{\alpha}^j \end{bmatrix} + \varepsilon,$$

где Z_{-j} — матрица Z без j-го столбца, α_{-j} — вектор α без j-го элемента, и после устранения линейной зависимости фиктивных переменных:

$$X = Z\alpha + Z^0\beta^0 + Z_j \bar{\otimes} \widetilde{Z}^G C\alpha^j + \varepsilon.$$

Все приведенные выше структуры матриц и соотношения между матрицами и векторами сохраняются.

В уравнение регрессии можно включать более одного качественного фактора. В случае двух факторов, принимающих, соответственно, k_1 и k_2 значения, форма «б» уравнения записывается следующим образом:

$$X = Z\alpha + Z^0\beta^0 + Z^1\beta^1 + Z^2\beta^2 + \varepsilon,$$

где вместо «^{G}» в качестве индекса качественного фактора используется его номер.

Это уравнение может включать фиктивные переменные совместного влияния качественных факторов (взаимодействия факторов). В исходной форме компонента совместного влияния записывается следующим образом:

$$\widetilde{Z}^1 \bar{\otimes} \widetilde{Z}^2 \widetilde{\beta}^{12},$$

где $\widetilde{\beta}^{12} = (\beta^{12}_{11}, \ldots, \beta^{12}_{1k_2}, \beta^{12}_{21}, \ldots, \beta^{12}_{2k_2}, \ldots, \beta^{12}_{k_1 1}, \ldots, \beta^{12}_{k_1 k_2})'$ — $k_1 \times k_2$-вектор-столбец, а $\beta^{12}_{i_1 i_2}$ — параметр при фиктивной переменной, которая равна 1, если первый фактор принимает i_1-е значение, а второй фактор — i_2-е значение, и равна 0 в остальных случаях (вектор-столбцом наблюдений за этой переменной является $(k_1(i_1-1)+i_2)$-й столбец матрицы $\widetilde{Z}^1 \bar{\otimes} \widetilde{Z}^2$).

Как и прежде, вектор параметров, из которого исключены все компоненты, линейно выражаемые через остальные, обозначается β^{12}. Он имеет размерность $(k_1-1)\times(k_2-1)$ и связан с исходным вектором параметров таким образом:

$$\widetilde{\beta}^{12} = C^1 \otimes C^2 \beta^{12},$$

где C^1 и C^2 — матрицы размерности $k_1 \times (k_1-1)$ и $k_2 \times (k_2-1)$, имеющие описанную выше структуру (матрица C).

Теперь компоненту совместного влияния можно записать следующим образом:

$$(\widetilde{Z}^1 \bar{\otimes} \widetilde{Z}^2)(C^1 \otimes C^2)\beta^{12} = (\widetilde{Z}^1 C^1) \bar{\otimes} (\widetilde{Z}^2 C^2)\beta^{12} = Z^1 \bar{\otimes} Z^2 \beta^{12} = Z^{12}\beta^{12},$$

а уравнение, включающее эту компоненту (форма «б») —

$$X = Z\alpha + Z^0\beta^0 + Z^1\beta^1 + Z^2\beta^2 + Z^{12}\beta^{12} + \varepsilon.$$

В общем случае имеется n качественных факторов, j-й фактор принимает k_j значений, см. пункт 1.9. Пусть упорядоченное множество $\{1, \ldots, n\}$ обозначается

G, а J — его подмножества. Общее их количество, включая пустое подмножество, равно 2^n. Каждому такому подмножеству взаимно-однозначно соответствует число, например, в системе исчисления с основанием $\max\limits_j k_j$, и их можно упорядочить по возрастанию этих чисел. Если пустое подмножество обозначить 0, то можно записать:

$$J = 0,\ 1,\ \ldots,\ n, \{1,2\},\ \ldots,\ \{1,n\},\ \{2,3\},\ \ldots,\ \{1,2,3\},\ \ldots,\ G.$$

Тогда уравнение регрессии записывается следующим образом:

$$X = Z\alpha + \sum_{J=0}^{G} \widetilde{Z}^J \widetilde{\beta}^J + \varepsilon = Z\alpha + \sum_{J=0}^{G} \widetilde{Z}^J C^J \beta^J + \varepsilon = Z\alpha + \sum_{J=0}^{G} Z^J \beta^J + \varepsilon,$$

где $\widetilde{Z}^J = \prod\limits_{j\in J} \bar{\otimes} \tilde{Z}^j$, $C^J = \prod\limits_{j\in J} \otimes C^j$ при $j > 0$, $C^0 = 1$. Выражение $j \in J$ под знаком произведения означает, что j принимает значения последовательно с первого по последний элемент подмножества J.

Очевидно, что приведенная выше запись уравнения для $n^* = 2$ является частным случаем данной записи.

Если $p(J)$ — количество элементов в подмножестве J, то $\widetilde{Z}^J \widetilde{\beta}^J$ или $Z^J \beta^J$ — J-е эффекты, **эффекты $p(J)$-го порядка**; при $p(J) = 1$ — **главные эффекты**, при $p(J) > 1$ — **эффекты взаимодействия, эффекты совместного влияния** или **совместные эффекты**.

$\widetilde{\beta}^J$ или β — параметры соответствующих J-х эффектов или также сами эти эффекты.

9.2. Модели с биномиальной зависимой переменной

Рассмотрим теперь модели, в которых **зависимая** переменная принимает только два значения, т.е. является фиктивной переменной. При этом придется отойти от модели линейной регрессии, о которой речь шла выше.

Если изучается спрос на рынке некоторого товара длительного пользования, например, на рынке холодильников определенной марки, то спрос в целом возможно предсказывать с помощью стандартной регрессии. Однако, если изучать спрос на холодильники отдельной семьи, то изучаемая переменная должна быть либо дискретной (0 или 1), либо качественной (не покупать холодильник, купить холодильник марки A, купить холодильник марки B и т.д.). Аналогично, разные методы приходится применять при изучении рынка труда и при изучении решения

отдельного человека по поводу занятости (работать/не работать). Данные о том, произошло какое-либо событие или нет, также можно представить дискретной переменной вида 0 или 1. При этом не обязательно наличие ситуации выбора. Например, можно исследовать данные об экономических кризисах, банкротствах (произошел или не произошел кризис или банкротство).

9.2.1. Линейная модель вероятности, логит и пробит

В биномиальную модель входит изучаемая переменная x, принимающая два значения, а также объясняющие переменные z, которые содержат факторы, определяющие выбор одного из значений. Без потери общности будем предполагать, что x принимает значения 0 и 1.

Предположим, что мы оценили на основе имеющихся наблюдений линейную регрессию

$$x = z\alpha + \varepsilon.$$

Очевидно, что для почти всех значений z построенная линейная регрессия будет предсказывать абсурдные значения изучаемой переменной x — дробные, отрицательные и большие единицы, что делает ее не очень полезной на практике.

Более того, линейная модель не может быть вполне корректной с формальной точки зрения. Поскольку у биномиальной зависимой переменной распределение будет распределением Бернулли (биномиальным распределением с одним испытанием Бернулли), то оно полностью задается вероятностью получения единицы. В свою очередь, вероятность того, что $x = 1$, совпадает с математическим ожиданием x, если эта переменная принимает значения 0 и 1:

$$\mathbf{E}(x) = \Pr(x = 1) \cdot 1 + \Pr(x = 0) \cdot 1 = \Pr(x = 1).$$

С другой стороны, ожидание x при данной величине z для линейной модели равно

$$\mathbf{E}(x) = z\alpha + \mathbf{E}(\varepsilon) = z\alpha.$$

Отсюда следует, что обычная линейная регрессионная модель не совсем подходит для описания рассматриваемой ситуации, поскольку величина za, вообще говоря, не ограничена, в то время как вероятность всегда ограничена нулем и единицей. Ожидаемое значение зависимой переменной, $\mathbf{E}(x)$, может описываться только нелинейной функцией.

Желательно каким-то образом модифицировать модель, чтобы она, с одной стороны, принимала во внимание тот факт, что вероятность не может выходить

за пределы отрезка $[0;1]$, и, с другой стороны, была почти такой же простой как линейная регрессия. Этим требованиям удовлетворяет модель, для которой

$$\Pr(x=1) = F(z\alpha),$$

где $F(\cdot)$ — некоторая достаточно простая функция, преобразующая $z\alpha$ в число от нуля до единицы. Естественно выбрать в качестве $F(\cdot)$ какую-либо дифференцируемую функцию распределения, определенную на всей действительной прямой. В дальнейшем мы рассмотрим несколько удобных функций распределения, которые удовлетворяют этим требованиям.

Заметим, что если выбрать $F(\cdot)$, соответствующую равномерному распределению на отрезке $[0;1]$, то окажется, что

$$\mathbf{E}(x) = \Pr(x=1) = \begin{cases} 0, & z\alpha \leqslant 0, \\ z\alpha, & 0 \leqslant z\alpha \leqslant 1, \\ 1, & z\alpha \geqslant 1. \end{cases}$$

Таким образом, при $z\alpha \in [0;1]$ получим «линейную регрессию». Это так называемая линейная модель вероятности. Однако, вообще говоря, такой выбор $F(\cdot)$ скорее не упрощает оценивание, а усложняет, поскольку в целом математическое ожидание зависимой переменной является здесь нелинейной функцией неизвестных параметров α (т.е. это нелинейная регрессия), причем эта функция недифференцируема.

В то же время, если данные таковы, что можно быть уверенным, что величина $z\alpha$ далека от границ 0 и 1, то линейную модель вероятности можно использовать, оценивая ее как обычную линейную регрессию. То, что величина $z\alpha$ далека от границ 0 и 1, означает, что z плохо предсказывает x. Таким образом, линейная модель вероятности применима в случае, когда изучаемая зависимость слаба, и в имеющихся данных доля как нулей, так и единиц не слишком мала. Ее можно рассматривать как приближение для нелинейных моделей.

Есть два удобных вида распределения, которые обычно используют для моделирования вероятности получения единицы в модели с биномиальной зависимой переменной. Оба распределения симметричны относительно нуля.

1) Логистическое распределение.

Плотность логистического распределения равна

$$\lambda(y) = \frac{e^y}{(1+e^y)^2},$$

а функция распределения равна

$$\Lambda(y) = \frac{e^y}{1+e^y} = \frac{1}{1+e^{-y}}.$$

Модель с биномиальной зависимой переменной с логистически распределенным отклонением называют **логит**. Для логита

$$\mathbf{E}(x) = \Pr(x=1) = \Lambda(z\alpha) = \frac{e^{z\alpha}}{1+e^{z\alpha}} = \frac{1}{1+e^{-z\alpha}}.$$

2) Нормальное распределение (см. Приложение А.3.2).

Модель с нормально распределенным отклонением ε называют **пробит**. При этом используется стандартное нормальное распределение, т.е. нормальное распределение с нулевым ожиданием и единичной дисперсией, $N(0,\ 1)$. Для пробита

$$\mathbf{E}(x) = \Pr(x=1) = \Phi(z\alpha) = \int\limits_{-\infty}^{z\alpha} \varphi(t)dt = \frac{1}{\sqrt{2\pi}} \int\limits_{-\infty}^{z\alpha} e^{-t^2/2}dt,$$

где $\Phi(\cdot)$ — функция распределения стандартного нормального распределения, $\varphi(\cdot)$ — его плотность.

Логистическое распределение похоже на нормальное с нулевым ожиданием и дисперсией $\pi^2/3$ (дисперсия логистического распределения). В связи с этим оценки коэффициентов в моделях различаются примерно на множитель $\pi/\sqrt{3} \approx 1.8$. Если вероятности далеки от границ 0 и 1 (около 0,5), то более точной оценкой множителя является величина $\varphi(0)/\lambda(0) = \sqrt{8/\pi} \approx 1.6$. При малом количестве наблюдений из-за схожести распределений сложно решить, когда следует применять логит, а когда — пробит. Различие наиболее сильно проявляется при вероятностях, близких к 0 и 1, поскольку логистическое распределение имеет более длинные хвосты, чем нормальное (оно характеризуется положительным коэффициентом эксцесса).

Можно использовать в модели и другие распределения, например, асимметричные.

9.2.2. Оценивание моделей с биномиальной зависимой переменной

Требуется по N наблюдениям $(x_i, z_i),\ i = 1,\ \ldots,\ N$, получить оценки коэффициентов α. Здесь наблюдения x_i независимы и имеют биномиальное распределение с одним испытанием (т.е. распределение Бернулли) и вероятностью

$$\Pr(x_i = 1) = F(z_i\alpha).$$

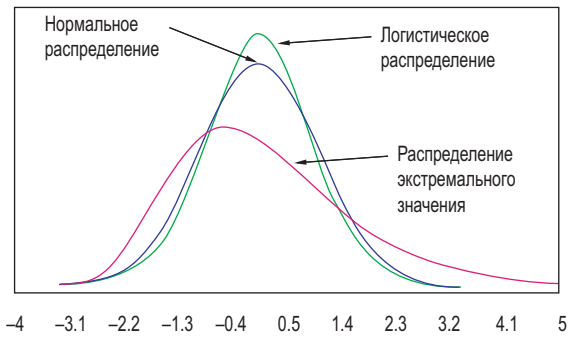

Рис. 9.1

Можно рассматривать модель с биномиальной зависимой переменной как модель регрессии:

$$x_i = F(z_i\alpha) + \xi_i,$$

где ошибки $\xi_i = x_i - F(z_i\alpha)$ имеют нулевое математическое ожидание и независимы. Каждая из ошибок ξ_i может принимать только два значения, и поэтому их распределение мало похоже на нормальное. Кроме того, имеет место гетероскедастичность. Обозначим

$$p_i = p_i(\alpha) = F(z_i\alpha).$$

В этих обозначениях дисперсия ошибки ξ_i равна

$$\mathbf{var}(\xi_i) = \mathbf{E}\Big[(x_i - p_i)^2\Big] = \mathbf{E}(x_i^2) - 2p_i\mathbf{E}(x_i) + p_i^2 = p_i(1 - p_i).$$

При выводе этой формулы мы воспользовались тем, что $x_i^2 = x_i$ и $\mathbf{E}(x_i) = p_i$.

Несмотря на эти нарушения стандартных предположений, данную модель, которая в общем случае представляет собой модель нелинейной регрессии, можно оценить нелинейным методом наименьших квадратов, минимизируя по α следующую сумму квадратов:

$$\sum_{i=1}^{N} (x_i - p_i(\alpha))^2.$$

Для минимизации такой суммы квадратов требуется использовать какой-либо алгоритм нелинейной оптимизации. Этот метод дает состоятельные оценки коэффициентов α. Гетероскедастичность приводит к двум важным последствиям. Во-первых, оценки параметров будут неэффективными (не самыми точными). Во-вторых, что более серьезно, ковариационная матрица коэффициентов, стандартные

ошибки коэффициентов и t-статистики будут вычисляться некорректно (если использовать стандартные процедуры оценивания нелинейной регрессии и получения в ней оценки ковариационной матрицы оценок параметров).

В частном случае модели линейной вероятности имеем линейную регрессию с гетероскедастичными ошибками:

$$x_i = z_i\alpha + \xi_i.$$

Для такой модели можно предложить следующую процедуру, делающую поправку на гетероскедастичность:

1) Оцениваем модель обычным МНК и получаем оценки a.

2) Находим оценки вероятностей:

$$p_i = z_i a.$$

3) Используем взвешенную регрессию и получаем оценки a^*.

Чтобы оценить взвешенную регрессию, следует разделить каждое наблюдение исходной модели на корень из оценки дисперсии ошибки, т.е. на величину $\sqrt{p_i(1-p_i)} = \sqrt{z_i a(1-z_i a)}$:

$$\frac{x_i}{\sqrt{p_i(1-p_i)}} = \frac{z_i}{\sqrt{p_i(1-p_i)}}\alpha + \frac{\xi_i}{\sqrt{p_i(1-p_i)}},$$

и далее применить к этой преобразованной регрессии обычный метод наименьших квадратов. При использовании данного метода получим асимптотически эффективные оценки a^* и корректную ковариационную матрицу этих оценок, на основе которой можно рассчитать t-статистики.

Те же идеи дают метод оценивания модели с произвольной гладкой функцией $F(\cdot)$. Для этого можно использовать линеаризацию в точке 0:

$$F(z_i\alpha) \approx F(0) + f(0)z_i\alpha,$$

где $f(\cdot)$ — производная функции $F(\cdot)$ (плотность распределения). Тогда получим следующую приближенную модель:

$$x_i \approx F(0) + f(0)\, z_i\alpha + \xi_i$$

или

$$x_i' \approx\ z_i\alpha + \xi_i',$$

где

$$x_i' = \frac{x_i - F(0)}{f(0)} \quad \text{и} \quad \xi' = \frac{\xi_i}{f(0)},$$

которую можно оценить с помощью только что описанной процедуры. Для симметричных относительно нуля распределений $F(0) = 0,5$. В случае логита, учитывая $\lambda(0) = 1/4$, получаем

$$x_i' = 4x_i - 2,$$

а в случае пробита, учитывая $\phi(0) = 1/\sqrt{2\pi}$, получаем

$$x_i' = \sqrt{2\pi}(x_i - 0,5).$$

Таким образом, можно получить приближенные оценки для коэффициентов пробита и логита, используя в качестве зависимой переменной регрессии вместо переменной, принимающей значения 0 и 1, переменную, которая принимает значения ± 2 для логита и $\pm\sqrt{\pi/2}$ для пробита ($\sqrt{\pi/2} \approx 1,25$). Ясно, что это хорошее приближение только когда величины $z_i\alpha$ близки к нулю, то есть когда модель плохо описывает данные.

Приближенные оценки можно получить также по группированным наблюдениям. Предположим, что все наблюдения разбиты на несколько непересекающихся подгрупп, в пределах каждой из которых значения факторов z_i примерно одинаковы. Введем обозначения:

$$\bar{p}_j = \frac{1}{N_j}\sum_{i \in I_j} x_i$$

и

$$\bar{z}_j = \frac{1}{N_j}\sum_{i \in I_j} z_i,$$

где I_j — множество наблюдений, принадлежащих j-й группе, N_j — количество наблюдений в j-й группе. Величина $\bar{p}_j$ является оценкой вероятности получения единицы в случае, когда факторы принимают значение $\bar{z}_j$, т.е.

$$\bar{p}_j \approx F(\bar{z}_j\alpha),$$

откуда

$$F^{-1}(\bar{p}_j) \approx \bar{z}_j\alpha.$$

Получаем модель регрессии, в которой в качестве зависимой переменной выступает $F^{-1}(\bar{p}_j)$, а в качестве факторов — $\bar{z}_j$. В частном случае логистического распределения имеем:

$$\Lambda^{-1}(\bar{p}_j) = \ln\left(\frac{\bar{p}_j}{1 - \bar{p}_j}\right),$$

т.е. для логита зависимая переменная представляет собой логарифм так называемого «соотношения шансов».

Чтобы такое приближение было хорошим, следует правильно сгруппировать наблюдения. При этом предъявляются два, вообще говоря, противоречивых требования:

- в пределах каждой группы значения факторов должны быть примерно одинаковы (идеальный случай — когда в пределах групп z_i совпадает, что вполне может случиться при анализе экспериментальных данных),
- в каждой группе должно быть достаточно много наблюдений.

Описанный метод лучше всего подходит тогда, когда в модели имеется один объясняющий фактор (и константа), поскольку в этом случае проще группировать наблюдения.

В настоящее время в связи с развитием компьютерной техники для оценивания моделей с биномиальной зависимой переменной, как правило, используется метод максимального правдоподобия, рассмотрение которого выходит за рамки данной главы.

9.2.3. Интерпретация результатов оценивания моделей с биномиальной зависимой переменной

Предположим, что каким-либо методом получен вектор оценок a. Как в этом случае можно интерпретировать результаты и судить о качестве модели?

Для логита коэффициенты a описывают влияние факторов на логарифм соотношения шансов. В общем случае по знаку коэффициентов можно судить о направлении зависимости, а по соответствующим t-статистикам — о наличии или отсутствии зависимости. Однако интерпретировать коэффициенты в содержательных терминах затруднительно. Поэтому помимо коэффициентов полезно рассмотреть, как влияют факторы на вероятность получения единицы:

$$\frac{\partial F(za)}{\partial z_j} = f(za)a_j.$$

Эти величины называют **маргинальными значениями**. Ясно, что маргинальные значения зависят от точки z, в которой они рассматриваются. Обычно берут z на среднем уровне по имеющимся наблюдениям: $z = \bar{z}$. Другой распространенный подход состоит в том, чтобы вычислить маргинальные значения во всех точках z_i, $i = 1, \ldots, N$, и по ним вычислить средние маргинальные значения:

$$\frac{1}{N}\left(\sum_{i=1}^{N} f(z_i a)\right) a_j.$$

Таблица 9.1

		Предсказано		
		0	1	Сумма
На самом деле	0	×	×	×
	1	×	×	×
	Сумма	×	×	×

Величину $x_i^c = z_i a$ можно назвать по аналогии с линейной регрессией расчетными значениями. При $za > 0$ для логита и пробита предсказанная вероятность единицы, $F(za)$, превосходит $1/2$, поэтому для такого наблюдения более вероятно наблюдать 1, чем 0. Таким образом, уравнение $za = 0$ задает ту гиперплоскость, которой разделяются две группы точек — те точки, для которых предсказано $x = 0$, и те точки, для которых предсказано $x = 1$. Поэтому наглядно о качестве модели можно судить по диаграмме x_i по x_i^c: чем лучше разделены две группы точек, тем более качественна модель. О качестве модели можно судить также по графику оценки $\mathbf{E}(x)$ по x^c. Этот график в случае «хорошей» модели должен быть «крутым» в нуле.

На этих двух графиках (рис. 9.2) слева внизу и справа вверху расположены правильно предсказанные точки, а слева вверху и справа внизу — неправильно. То же самое можно представить таблицей 9.1.

Понятно, что «хорошая» модель должна давать высокий процент правильных предсказаний (в таблице они лежат на диагонали).

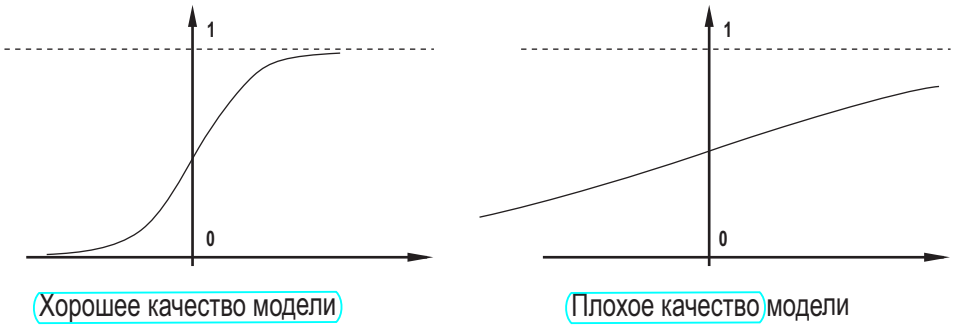

Рис. 9.2

9.3. Упражнения и задачи

Упражнение 1

1.1. Пусть $\widetilde{Z}^G = \{z_{i1}^G z_{i2}^G\}$ — фиктивная переменная, где z_{i1}^G равно единице, если фактор в i-м наблюдении относится к годам войны $(1941, \ldots, 1945)$, и нулю в противном случае. Как выглядит вектор z_{i2}^G? Оцените двумя способами модель $X = Z\alpha + Z^0\beta^0 + \tilde{Z}^G\tilde{\beta} + \varepsilon$ с помощью искусственно созданных данных из табл. 9.2, рассмотрев в качестве X столбец X_1:

а) исключив столбец z_1^G в исходной форме регрессии;

б) исключив в исходной форме регрессии параметр при переменной z_1^G.

Убедитесь, что значения коэффициентов исходной регрессии по способам а) и б) совпадают.

1.2. Запишите модель регрессии, в которой качественный фактор влияет не только на значение свободного члена регрессии, но и на коэффициент регрессии при факторе Z_1.

Посчитайте матрицы $Z_1 \bar{\otimes} \tilde{Z}^G$ и $Z_1 \bar{\otimes} [Z^0, \bar{Z}^G]$. Оцените данную модель регрессии на данных таблицы 9.2, рассмотрев в качестве X столбец X_2 способами а) и б).

Упражнение 2

Самостоятельно подберите ряды наблюдений и охарактеризуйте цены на российском вторичном рынке жилья в зависимости от жилой и нежилой площади, площади кухни, местоположения квартиры по районам города, расположения на этажах, количество комнат, наличия телефона, балкона, лифта и т.д.

Упражнение 3

В таблице 9.3 приводится данные о голосовании по поводу увеличения налогов на содержание школ в городе Троя штата Мичиган в 1973 г. Наблюдения относятся к 95 индивидуумам: результаты голосования и различные характеристики индивидов.

Pub = 1, если хотя бы один ребенок посещает государственную школу, иначе 0,

Priv = 1, если хотя бы один ребенок посещает частную школу, иначе 0,

Years = срок проживания в данном районе,

Teach = 1, если работает учителем, иначе 0,

Таблица 9.2

Годы	X_1	X_2	Z_1	Z_2	**Годы**	X_1	X_2	Z_1	Z_2
1935	2.81	2.81	117.10	9.70	**1945**	24.95	19.93	200.70	32.00
1936	10.66	10.66	201.60	10.40	**1946**	16.44	16.44	220.80	34.60
1937	4.16	4.16	280.30	11.80	**1947**	15.04	15.04	165.60	45.60
1938	8.30	8.30	204.00	15.60	**1948**	15.44	15.44	160.40	54.30
1939	16.94	16.94	225.60	17.20	**1949**	23.43	23.43	61.80	55.50
1940	5.01	5.01	213.20	18.60	**1950**	6.98	6.98	161.10	64.70
1941	35.49	30.90	183.40	22.10	**1951**	18.61	18.61	181.90	67.10
1942	26.76	22.79	158.80	28.80	**1952**	22.74	22.74	207.90	72.60
1943	34.88	30.50	174.90	32.00	**1953**	24.63	24.63	237.10	80.00
1944	35.27	31.06	168.70	32.10	**1954**	31.35	31.35	275.90	88.90

LnInc = логарифм годового дохода семьи в долларах,

PropTax = логарифм налогов на имущество в долларах за год (заменяет плату за обучение — плата зависит от имущественного положения),

Yes = 1, если человек проголосовал на референдуме «за», 0, если «против».

Зависимая переменная — **Yes**. В модель включаются все перечисленные факторы, а также квадрат **Years**.

3.1. Получите приближенные оценки для логита и пробита с помощью линейной регрессии

3.2. Вычислите коэффициенты логита через коэффициенты пробита и сравните.

3.3. Для логита найдите маргинальные значения для **Teach, LnInc** и **PropTax** при среднем уровне факторов.

3.4. Постройте график вероятности голосования «за» в зависимости от **Years** при среднем уровне остальных факторов.

3.5. Постройте аналогичный график маргинального значения **Years**.

Таблица 9.3. (Источник: R. Pindyck and D. Rubinfeld, Econometric Models and Economic Forecasts, 1998, Fourth Edition, Table 11.8, p. 332)

Номер	Pub	Priv	Years	Teach	LnInc	PropTax	Yes
1	1	0	10	1	9.77	7.0475	1
2	1	0	8	0	10.021	7.0475	0
3	1	0	4	0	10.021	7.0475	0
4	1	0	13	0	9.4335	6.3969	0
5	1	0	3	1	10.021	7.2792	1
6	1	0	5	0	10.463	7.0475	0
7	0	0	4	0	10.021	7.0475	0
8	1	0	5	0	10.021	7.2793	1
9	1	0	10	0	10.222	7.0475	0
10	1	0	5	0	9.4335	7.0475	1
11	1	0	3	0	10.021	7.0475	1
12	1	0	30	0	9.77	6.3969	0
13	1	0	1	0	9.77	6.7452	1
14	1	0	3	0	10.021	7.0475	1
15	1	0	3	0	10.82	6.7452	1
16	1	0	42	0	9.77	6.7452	1
17	1	0	5	1	10.222	7.0475	1
18	1	0	10	0	10.021	7.0475	0
19	1	0	4	0	10.222	7.0475	1
20	1	1	4	0	10.222	6.7452	1
21	1	0	11	1	10.463	7.0475	1
22	0	0	5	0	10.222	7.0475	1
23	1	0	35	0	9.77	6.7452	1
24	1	0	3	0	10.463	7.2793	1
25	1	0	16	0	10.021	6.7452	1
26	0	1	7	0	10.463	7.0475	0
27	1	0	5	1	9.77	6.7452	1
28	1	0	11	0	9.77	7.0475	0
29	1	0	3	0	9.77	6.7452	0
30	1	1	2	0	10.222	7.0475	1
31	1	0	2	0	10.021	6.7452	1
32	1	0	2	0	9.4335	6.7452	0
33	1	0	2	1	8.294	7.0475	0
34	0	1	4	0	10.463	7.0475	1

Таблица 9.3. (продолжение)

Номер	Pub	Priv	Years	Teach	LnInc	PropTax	Yes
35	1	0	2	0	10.021	7.0475	1
36	1	0	3	0	10.222	7.2793	0
37	1	0	3	0	10.222	7.0475	1
38	1	0	2	0	10.222	7.4955	1
39	1	0	10	0	10.021	7.0475	0
40	1	0	2	0	10.222	7.0475	1
41	1	0	2	0	10.021	7.0475	0
42	1	0	3	0	10.82	7.4955	0
43	1	0	3	0	10.021	7.0475	1
44	1	0	3	0	10.021	7.0475	1
45	1	0	6	0	10.021	6.7452	1
46	1	0	2	0	10.021	7.0475	1
47	1	0	26	0	9.77	6.7452	0
48	0	1	18	0	10.222	7.4955	0
49	0	0	4	0	9.77	6.7452	0
50	0	0	6	0	10.021	7.0475	0
51	0	0	12	0	10.021	6.7452	1
52	1	0	49	0	9.4335	6.7452	1
53	1	0	6	0	10.463	7.2793	1
54	0	1	18	0	9.77	7.0475	0
55	1	0	5	0	10.021	7.0475	1
56	1	0	6	0	9.77	5.9915	1
57	1	0	20	0	9.4335	7.0475	0
58	1	0	1	1	9.77	6.3969	1
59	1	0	3	0	10.021	6.7452	1
60	1	0	5	0	10.463	7.0475	0
61	1	0	2	0	10.021	7.0475	1
62	1	1	5	0	10.82	7.2793	0
63	1	0	18	0	9.4335	6.7452	0
64	1	0	20	1	9.77	5.9915	1
65	0	0	14	0	8.9227	6.3969	0
66	1	0	3	0	9.4335	7.4955	0
67	1	0	17	0	9.4335	6.7452	0
68	1	0	20	0	10.021	7.0475	0

Таблица 9.3. (продолжение)

Номер	Pub	Priv	Years	Teach	LnInc	PropTax	Yes
69	1	1	3	0	10.021	7.0475	1
70	1	0	2	0	10.021	7.0475	1
71	0	0	5	0	10.222	7.0475	1
72	1	0	35	0	9.77	7.0475	1
73	1	0	10	0	10.021	7.2793	0
74	1	0	8	0	9.77	7.0475	1
75	1	0	12	0	9.77	7.0475	0
76	1	0	7	0	10.222	6.7452	1
77	1	0	3	0	10.463	6.7452	1
78	1	0	25	0	10.222	6.7452	0
79	1	0	5	1	9.77	6.7452	1
80	1	0	4	0	10.222	7.0475	1
81	1	0	2	0	10.021	7.2793	1
82	1	0	5	0	10.463	6.7452	1
83	1	0	3	0	9.77	7.0475	0
84	1	0	2	0	10.82	7.4955	1
85	0	1	6	0	8.9227	5.9915	0
86	1	1	3	0	9.77	7.0475	1
87	1	0	12	0	9.4335	6.3969	1
88	0	0	3	0	9.77	6.7452	1
89	1	0	3	0	10.021	7.0475	1
90	0	0	3	0	10.021	6.7452	1
91	1	0	3	0	10.222	7.2793	1
92	1	0	3	1	10.021	7.0475	1
93	1	0	5	0	10.021	7.0475	1
94	0	0	35	1	8.9277	5.9915	1
95	1	0	3	0	10.463	7.4955	0

Задачи

1. Какие из перечисленных факторов учитываются в регрессии с помощью фиктивных переменных: а) профессия; б) курс доллара; в) численность населения; г) размер среднемесячных потребительских расходов?

2. В уравнение регрессии для доходов населения вводятся два качественных фактора: «пол» и «наличие судимости». Сколько фиктивных переменных (с учетом взаимодействия факторов) в исходной и преобразованной (после устранения линейных зависимостей) форме уравнения?

3. В уравнение регрессии для доходов населения вводятся три качественных фактора: пол («муж.», «жен.»), образование («нач.», «сред.», «высш.») и место проживания («гор.», «сел.»). Сколько фиктивных переменных (с учетом всех взаимодействий факторов) в исходной и преобразованной (после устранения линейных зависимостей) форме уравнения? Как выглядят матрицы преобразований $\bar{C}$ и C?

4. Известно, что котировки многих ценных бумаг зависят от того, в какой день рабочей недели (понедельник, вторник, среда, ...) проходят торги. Как учесть эту зависимость при построении регрессионной модели котировок?

5. Предположим, что оценивается зависимость спроса на лыжи от располагаемого личного дохода, используя наблюдения по месяцам. Как ввести фиктивную переменную для оценивания сезонных колебаний? Запишите соответствующие матрицы преобразований $\bar{C}$ и C для каждого фиктивного фактора.

6. Рассмотрим регрессионную модель $x_t = \alpha_1 z_{t1} + \alpha_2 z_{t2} + \beta^0 + \varepsilon_t$, $t = 1, \ldots, T$. Пусть для наблюдений $t = 1$ и 2 параметры α_1, α_2 и β^0 отличаются от остальных ($T - 2$) наблюдений. Запишите регрессионную модель с фиктивными переменными и опишите возникшие проблемы оценивания.

7. На основе данных о расходах на автомобили (X) и располагаемом личном доходе (Z) за период с 1963 по 1982 года получена модель: $X = 0.77 + + 0.035Z - 4.7Z_1^G$, где Z_1^F — фиктивная переменная, учитывающая нефтяной кризис 1974 года, равная 0 для периодов с 1963 по 1973 гг. и равной единице для периода с 1974 по 1982 гг.

   а) Схематично нарисуйте график регрессионной функции и дайте полную интерпретацию.

   б) Запишите модель, в которой качественный фактор z^G не влияет на свободный член, но влияет на наклон линии регрессии. Схематично нарисуйте график регрессионной функции.

8. Как меняется коэффициент детерминации при добавлении в регрессионную модель фиктивной объясняющей переменной?

9. На основе опроса населения США Current Population Survey за 1985 г. изучаются факторы, определяющие зарплату:

   WAGE: зарплата (долларов за час) — изучаемая переменная,

   EDU: образование (лет),

   SOUTH: индикаторная переменная для Юга (1 = человек живет на Юге, 0 = человек живет в другом месте),

   SEX: индикаторная переменная для пола (1 = жен, 0 = муж),

   EXPER: стаж работы (лет),

   UNION: индикаторная переменная для членства в профсоюзе (1 = член профсоюза, 0 = нет),

   AGE: возраст (лет),

   RACE: раса (1 = другое, 2 = «Hispanic», 3 = белый),

   OCCUP: профессиональная категория (1 = другое, 2 = Management, 3 = Sales, 4 = Clerical, 5 = Service, 6 = Professional),

   SECTOR: сектор экономики (0 = другое, 1 = промышленность, 2 = строительство),

   MARR: семейное положение (0 = неженатый/незамужняя, 1 = женатый/замужняя).

   а) Какие из перечисленных переменных можно назвать фиктивными? Объясните.

   б) Объясните, в каком виде следует учитывать переменные RACE, OCCUP и SECTOR в регрессии.

   в) Для каждого фиктивного фактора запишите соответствующую матрицу преобразований C.

   г) Объясните, как будут выглядеть фиктивные переменные, соответствующие эффектам второго порядка для пола и расы.

10. Модель регрессии с биномиальной зависимой переменной можно представить в виде: (зависимая переменная) = (математическое ожидание) + (ошибка). Какие предположения классической линейной регрессии при этом будут нарушены?

11. Предположим, что с помощью обычного линейного МНК с биномиальной зависимой переменной были получены оценки a. Как на их основе получить приближенные оценки для модели пробит?

12. Логит-оценивание модели $\Pr(x = 1) = F(z\alpha)$ дало результат $x^* = -5.89 + + 0.2z$. Чему равна вероятность $x = 1$ при $z = 50$?

13. Пробит-оценивание модели $\Pr(x = 1) = F(z\alpha)$ дало результат $x^* = = -2.85 + 0.092z$. Чему равна вероятность $x = 1$ при $z = 50$?

14. Логит-оценивание модели $\Pr(x = 1) = F(z\alpha)$ дало результат $x^* = -5.89 + + 0.2z$. Чему равно увеличение вероятности $\Pr(x = 1)$ при увеличении z на единицу, если $z = 50$?

15. Пробит-оценивание модели $\Pr(x = 1) = F(z\alpha)$ дало результат $x^* = = -2.85 + 0.092z$. Чему равно увеличение вероятности $\Pr(x = 1)$ при увеличении z на единицу, если $z = 50$?

16. Логит-модель применили к выборке, в которой $x = 1$, если производительность труда на предприятии выросла, и $x = 0$ в противном случае. z_1 — доход предприятия в млн. руб. в год, $z_1^G = 1$ если предприятие относится к области высоких технологий ($z_1^G = 0$ в противном случае). Получена следующая модель: $x = 0.5 + 0.1z_1 + 0.4z_1^G$. Определите оценку вероятности роста производительности труда для высокотехнологичного предприятия с доходом 100 млн. руб. в год и для предприятия, не относящегося к сфере высоких технологий, с доходом 150 млн. руб. в год.

17. Имеется выборка, состоящая из 600 наблюдений, в которой $x = 1$, если работник состоит в профсоюзе, и $x = 0$ в противном случае. Предполагается, что членство в профсоюзе зависит от образования, лет (z_1), стажа работы, лет (z_2) и пола (z_3). Выборочные средние равны $\bar{x} = 0.2$, $\bar{z}_1 = 14$, $\bar{z}_2 = 18$ и $\bar{z}_3 = 0.45$. На основе выборочных данных получена следующая пробит-модель: $x = -0.9 - 0.01z_1 + 0.4z_2 - 0.6z_3$. Определить, насколько снижается вероятность быть членом профсоюза в расчете на год дополнительного образования.

18. Пусть переменная x, принимающая значения 0 или 1, зависит от одного фактора z. Модель включает также константу. Данные приведены в таблице:

x	0	0	1	1	0	1	0	1	0	1
z	1	2	3	4	5	6	7	8	9	10

а) Получите приближенные оценки логита и пробита методом усреднения, разбив данные на две группы по 5 наблюдений. Каким будет процент правильных предсказаний по модели для этих данных?

б) Ответьте на вопросы предыдущего пункта для метода приближенного оценивания логита и пробита с помощью линейной регрессии.

в) Найдите маргинальное значение для фактора z в точке, соответствующей его среднему уровню.

19. Пусть переменная x, принимающая значения 0 или 1, зависит от фиктивной переменной z, принимающей значения 0 или 1. Модель включает также константу. Данные резюмируются следующей таблицей (в клетках стоят количества соответствующих наблюдений):

	$x=0$	$x=1$
$z=0$	N_{00}	N_{01}
$z=1$	N_{10}	N_{11}

а) При каких условиях можно на основе этих данных оценить логит и пробит?

б) Получите приближенные оценки логита и пробита методом усреднения. Чему они будут равны при $N_{00}=15,\ N_{01}=5,\ N_{10}=5,\ N_{11}=15$? Каким будет процент правильных предсказаний по модели для этих данных?

в) Ответьте на вопросы предыдущего пункта для метода приближенного оценивания логита и пробита с помощью линейной регрессии.

Рекомендуемая литература

1. **Айвазян С.А.** Основы эконометрики. Т.2. — М.: «Юнити», 2001. (Гл. 2)

2. **Доугерти К.** Введение в эконометрику. — М.: «Инфра-М», 1997. (Гл. 9).

3. **Дрейпер Н., Смит Г.** Прикладной регрессионный анализ. В 2-х книгах. Кн. 2. — М.: «Финансы и статистика», 1986. (Гл. 9).

4. **Магнус Я.Р., Катышев П.К., Пересецкий А.А.** Эконометрика — начальный курс. — М.: «Дело», 2000. (Гл. 4).

5. **Маленво Э.** Статистические методы эконометрии. — М.: «Статистика». Вып. 1, 1975. (Гл. 8).

6. **Baltagi, Badi H.** Econometrics, 2nd edition, Springer, 1999. (Ch. 13).

7. **Davidson, Russel, Mackinnon, James.** Estimation and Inference in Econometrics, No. 9, Oxford University Press, 1993. (Ch. 7).

8. **Greene W.H.** Econometric Analysis, Prentice-Hall, 2000. (Ch. 8).

9. **Judge G.G., Hill R.C., Griffiths W.E., Luthepohl H., Lee T.** Introduction to the Theory and Practice of Econometric. John Wiley & Sons, Inc., 1993. (Ch. 10).

10. **Maddala G.S.** Introduction to Econometrics, 2nd ed., Prentice Hall, 1992. (Ch. 8).

11. **Ruud Paul A.** An Introduction to Classical Econometric Theory, Oxford University Press, 2000. (Ch. 27).

12. **Wooldridge Jeffrey M.** Introductory Econometrics: A Modern Approach, 2nd ed., Thomson, 2003. (Ch. 7, 17).

Глава 10

Оценка параметров систем уравнений

Пусть теперь имеется несколько изучаемых переменных, для каждой из которых существует свое уравнение регрессии. В совокупности эти уравнения образуют систему, которая является **невзаимозависимой**, если одни изучаемые переменные не выступают факторами-регрессорами для других изучаемых переменных. Если изучаемые переменные возникают не только в левых, но и правых частях уравнений, то такие системы называются **одновременными** или **взаимозависимыми**.

10.1. Невзаимозависимые системы

В этом пункте используется сокращенная форма записи уравнений регрессии:

$$\hat{X} = \hat{Z}A + \varepsilon, \tag{10.1}$$

где $\hat{X}$ — $N \times k$-матрица центрированных наблюдений за изучаемыми переменными,

$\hat{Z}$ — $N \times n$-матрица центрированных наблюдений за факторными переменными,

A — $n \times k$-матрица параметров уравнений регрессии,

ε — $N \times n$-матрица ошибок изучаемых переменных (остатков по наблюдениям).

Относительно ошибок предполагается, что в каждом наблюдении их математическое ожидание равно нулю, матрица ковариации размерности $k \times k$ одинакова и равна Ω (Ω — вещественная, симметричная, положительно определенная матрица), и что они не коррелированы по наблюдениям.

Оценивать параметры этой системы можно отдельно по каждому уравнению:

$$A = M^{-1}\tilde{m}, \tag{10.2}$$

где $M = \frac{1}{N}\hat{Z}'\hat{Z}$, $\tilde{m} = \frac{1}{N}\hat{Z}'\hat{X}$, или через обычные операторы МНК-оценивания (8.1), записанные последовательно для всех уравнений системы

$$a_l = M^{-1}m_l, \quad l = 1, \ldots, k.$$

Т.е. факт коррелированности ошибок разных изучаемых переменных ($\Omega \neq I_k$) не создает дополнительных проблем.

Действительно, преобразованием в пространстве изучаемых переменных легко перейти в ситуацию, когда ошибки изучаемых переменных не коррелированы.

Пусть матрица C такая, что $\Omega = {C'}^{-1}C^{-1}$ (такое представление допускает любая вещественная симметричная положительно определенная матрица, см. Приложение А.1.2). Умножим обе части (10.1) справа на эту матрицу:

$$\hat{X}C = \hat{Z}AC + \varepsilon C. \tag{10.3}$$

Новые ошибки изучаемых переменных во всех наблюдениях оказываются не коррелированными:

$$\mathbf{E}(C'\varepsilon_i'\varepsilon_i C) \overset{E(\varepsilon_i'\varepsilon_i)=\Omega}{=} I_N,$$

где ε_i — вектор-строка ошибок в i-том наблюдении.

Теперь уравнения системы не связаны между собой, и их можно оценить обычным МНК по отдельности, что, очевидно, приводит к матричному оператору $AC = M^{-1}\tilde{m}C$, который эквивалентен (10.2).

Что и требовалось доказать.

Ситуация резко усложняется, если для коэффициентов матрицы A имеются априорные ограничения.

Пусть, например, эта матрица имеет следующую структуру:

$$\begin{bmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_k \end{bmatrix},$$

где a_l — n_l-вектор-столбец коэффициентов в l-м уравнении (для l-й изучаемой переменной), $\sum\limits_{l=1}^{k} n_l = n$, т.е. многие элементы матрицы A априорно приравнены нулю.

Фактически это означает, что для каждой изучаемой переменной имеется свой набор объясняющих факторов с $N \times n_l$-матрицей наблюдений $\hat{Z}_l$ $(\hat{Z} = [\hat{Z}_1 \cdots \hat{Z}_k])$, и система уравнений (10.1) представляется как совокупность внешне не связанных между собой уравнений:

$$\hat{X}_l = \hat{Z}_l a_l + \varepsilon_l, \quad l = 1, \ldots, k. \tag{10.4}$$

Сразу можно заметить, что теперь оператор (10.2) применить невозможно, т.к. система нормальных уравнений, решением которой является этот оператор, записывается следующим образом:

$$\begin{bmatrix} M_{11}a_1 & \cdots & M_{1k}a_k \\ \vdots & \ddots & \vdots \\ M_{k1}a_1 & \cdots & M_{kk}a_k \end{bmatrix} = \begin{bmatrix} m_{11} & \cdots & m_{1k} \\ \vdots & \ddots & \vdots \\ m_{k1} & \cdots & m_{kk} \end{bmatrix}, \tag{10.5}$$

где $M_{ll'} = \frac{1}{N}\hat{Z}_l'\hat{Z}_{l'}$, $m_{ll'} = \frac{1}{N}\hat{Z}_l'\hat{X}_{l'}$, т.е. вектор оценок параметров каждого уравнения должен удовлетворять k взаимоисключающим, в общем случае, системам уравнений.

Правильная оценка параметров регрессии дается решением следующих уравнений:

$$\sum_{l'=1}^{k} \omega_{ll'}^{-1} M_{ll'}\, a_{l'} = \sum_{l'=1}^{k} \omega_{ll'}^{-1} m_{ll'}, \quad l = 1, \ldots, k,$$

где $\omega_{ll'}^{-1}$ — элемент матрицы Ω^{-1}.

Или в матричной записи:

$$\begin{bmatrix} \omega_{11}^{-1}M_{11}a_1 + & \cdots & +\omega_{1k}^{-1}M_{1k}a_k \\ \vdots & \ddots & \vdots \\ \omega_{k1}^{-1}M_{k1}a_1 + & \cdots & +\omega_{kk}^{-1}M_{kk}a_k \end{bmatrix} = \begin{bmatrix} \omega_{11}^{-1}m_{11} + & \cdots & +\omega_{1k}^{-1}m_{1k} \\ \vdots & \ddots & \vdots \\ \omega_{k1}^{-1}m_{k1} + & \cdots & +\omega_{kk}^{-1}m_{kk} \end{bmatrix}, \tag{10.6}$$

которая при сравнении с (10.5) оказывается результатом умножения в (10.5) всех $M_{ll'}$ и $m_{ll'}$ на $\omega_{ll'}^{-1}$ и сложения столбцов в обеих частях этого выражения.

Для доказательства этого утверждения необходимо перегруппировать уравнения системы так, чтобы

$$\tilde{X} = \begin{bmatrix} \hat{X}_1 \\ \hat{X}_2 \\ \vdots \end{bmatrix}, \quad \tilde{Z} = \begin{bmatrix} \hat{Z}_1 & 0 & \cdots \\ 0 & \hat{Z}_2 & \ddots \\ \vdots & \ddots & \ddots \end{bmatrix}, \quad \tilde{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \end{bmatrix}, \quad \tilde{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \end{bmatrix},$$

т.е. если забыть об особой структуре матрицы $\tilde{Z}$, формально имеется одна изучаемая переменная, для которой имеется $N \cdot k$ «наблюдений».

Теперь система (10.4) записывается следующим образом:

$$\tilde{X} = \tilde{Z}\tilde{a} + \tilde{\varepsilon},$$

и применение простого МНК приводит к получению обычных оценок уравнений в отдельности:

$$a_l = M_{ll}^{-1} m_{ll}.$$

Однако такой подход неприемлем, надо применять ОМНК, поскольку остатки коррелированы по «наблюдениям», ибо в соответствии со сделанными предположениями

$$\mathbf{E}(\tilde{\varepsilon}\tilde{\varepsilon}') = \Omega \otimes I_N,$$

где $\otimes$ — операция прямого умножения матриц (см. Приложения А.1.1 и А.1.2).

Из (8.1) следует, что система нормальных уравнений ОМНК в данном случае выглядит так:

$$\tilde{Z}'\Omega^{-1} \otimes I_N \tilde{Z}\tilde{a} = \tilde{Z}'\Omega^{-1} \otimes I_N \tilde{X}. \tag{10.7}$$

Легко убедиться, что

$$\tilde{Z}'\Omega^{-1} \otimes I_N = \begin{bmatrix} \omega_{11}^{-1}\hat{Z}'_{11} & \omega_{12}^{-1}\hat{Z}'_{12} & \vdots \\ \omega_{21}^{-1}\hat{Z}'_{21} & \omega_{22}^{-1}\hat{Z}'_{22} & \ddots \\ \vdots & \ddots & \ddots \end{bmatrix}.$$

Умножение этой матричной конструкции справа на $\tilde{Z}$ и деление на N дает блочную матрицу $\{\omega_{ll'}^{-1} M_{ll'}\}$, которая является матрицей системы (10.6), а умножение ее справа на $\tilde{X}$ и деление на N — вектор $\left\{\sum_{l'} \omega_{ll'}^{-1} m_{ll'}\right\}$, являющийся правой частью системы (10.6).

Таким образом, (10.7) эквивалентна (10.6). Что и требовалось доказать.

Эта оценка совпадает с обычной МНК-оценкой $a_l = M_{ll}^{-1} m_{ll}$, если матрица Ω диагональна, т.е. ошибки изучаемых переменных не коррелированы.

10.2. Взаимозависимые или одновременные уравнения. Проблема идентификации

Далее в этом разделе уравнения регрессии записываются в форме со скрытым свободным членом.

X — $N \times k$-матрица наблюдений за изучаемыми переменными x;

Z — $N \times (n+1)$-матрица наблюдений за независимыми факторами z;

B — $k \times k$-матрица параметров регрессии при изучаемых переменных; $B \neq I_k$, иначе система была бы невзаимозависимой; $|B| \neq 0$ и $\beta_{ll} = 1$ — **условия нормализации**, т.е. предполагается, что, в конечном счете, в левой части l-го уравнения остается только l-я переменная, а остальные изучаемые переменные переносятся в правую часть;

A — $(n+1) \times k$-матрица параметров регрессии (последняя строка — свободные члены в уравнениях);

ε — $N \times k$-матрица значений случайных ошибок по наблюдениям;

$$XB = ZA + \varepsilon. \tag{10.8}$$

Такая запись одновременных уравнений называется **структурной формой**. Умножением справа обеих частей этой системы уравнений на B^{-1} она приводится к форме, описанной в предыдущем пункте. Это — **приведенная форма** системы:

$$X = ZAB^{-1} + \varepsilon B^{-1}.$$

$D = AB^{-1}$ — $(n+1) \times k$-матрица параметров регрессии приведенной формы. Как показано в пункте 10.1, для их оценки можно использовать МНК:

$$D = (Z'Z)^{-1}Z'X.$$

Таким образом, матрица D оценивается без проблем, и ее можно считать известной. Однако задача заключается в оценке параметров B и A системы в приведенной форме. Эти параметры, по определению, удовлетворяют следующим условиям:

$$DB - A = 0 \tag{10.9}$$

или $WH = 0$, где

W — $(n+1) \times (n+k+1)$-матрица $\begin{bmatrix} D & I_{n+1} \end{bmatrix}$,

H — $(n+k+1) \times k$-матрица $\begin{bmatrix} B \\ -A \end{bmatrix}$.

Это — условия для оценки параметров структурной формы. В общем случае эти условия достаточно бессмысленны, т.к. они одинаковы для параметров всех уравнений. Они описывают лишь множество допустимых значений параметров (одинаковое для всех уравнений), поскольку для $n+k+1$ параметров каждого уравнения структурной формы имеется только $n+1$ одинаковых уравнений. Необходимы дополнительные условия, специальные для каждого уравнения.

Пусть для параметров l-го уравнения кроме требования

$$WH_l = 0 \quad ((Z'Z)^{-1}Z'XB_l - A_l = 0) \tag{10.10}$$

имеется дополнительно r_l условий:

$$R_l H_l = 0, \tag{10.11}$$

где R_l — $r_l \times (n+k+1)$-матрица дополнительных условий,

H_l — $(n+k+1)$-вектор-столбец $\begin{bmatrix} B_l \\ -A_l \end{bmatrix}$ параметров l-го уравнения — l-й столбец матрицы H.

$\begin{bmatrix} W \\ R_l \end{bmatrix} H_l = W_l H_l = 0$ — общие условия для определения структурных параметров l-го уравнения, где W_l — $(n+r_l+1) \times (n+k+1)$-матрица.

Они позволяют определить искомые параметры с точностью до постоянного множителя (при выполнении условий нормализации $\beta_l = 1$ параметры определяются однозначно), если и только если ранг матрицы W_l равен $n+k$. Для этого необходимо, чтобы

$$r_l \geqslant k-1. \tag{10.12}$$

Однако, это условие не является достаточным. Имеется необходимое и достаточное условие для определения параметров l-го уравнения (более операциональное, чем требование равенства $n+k$ ранга матрицы W_l):

$$\operatorname{rank}(R_l H) = k-1. \tag{10.13}$$

Доказательство данного утверждения опускается по причине сложности.

Теперь вводятся определения, связанные с возможностью нахождения параметров уравнения структурной формы: l-е уравнение **не идентифицировано**, если $r_l < k-1$; оно **точно идентифицировано**, если $r_l = k-1$ и ранг W_l равен $n+k$; **сверхидентифицировано**, если $r_l > k-1$. В первом случае параметры не

могут быть оценены, и, хотя формально, например, используя МНК, оценки можно получить, они никакого смысла не имеют; во втором случае параметры уравнения оцениваются однозначно; в третьем — имеется несколько вариантов оценок.

Обычно строки матрицы R_l являются ортами, т.е. дополнительные ограничения исключают некоторые переменные из структурной формы. Тогда, если k_l и n_l — количества, соответственно, изучаемых переменных, включая l-ю, и независимых факторов в l-м уравнении, то для его идентификации необходимо, чтобы

$$k_l + n_l \leqslant n + 1. \tag{10.14}$$

По определению, $r_l = n - n_l + k - k_l \overset{(10.12)}{\geqslant} k - 1 \;\Rightarrow\; n_l + k_l \leqslant n + 1.$

В таком случае условие (10.13) означает, что матрица, составленная из коэффициентов во всех прочих уравнениях, кроме l-го, при переменных, которые исключены из l-го уравнения, должна быть не вырождена. При этом l-й столбец матрицы $R_l H$ из (10.13), равный нулю, как это следует из (10.11), исключается из рассмотрения.

Для иллюстрации введенных понятий используется элементарная модель равновесия спроса и предложения на рынке одного товара в предположении, что уравнения спроса и предложения линейны (в логарифмах):

$s = b_{21}p + c_1 + \varepsilon_1$ — предложение,
$d = -b_{22}p + c_2 + \varepsilon_2$ — спрос,

где p — цена, b_{21}, b_{22} — эластичности предложения и спроса по цене, s, d и p — логарифмы предложения, спроса и цены.

Наблюдаемой переменной является фактический объем продаж x, и, предположив, что в действительности рынок находится в равновесии: $x = s = d$, эту модель в структурной форме (10.8) можно записать следующим образом:

$$[\,x \;\; p\,] \begin{bmatrix} 1 & 1 \\ -b_{21} & b_{22} \end{bmatrix} = [\,c_1 \;\; c_2\,] + [\,\varepsilon_1 \;\; \varepsilon_2\,]. \tag{10.15}$$

В такой записи условия нормализации не выполнены, т.к. в левой части обоих уравнений находится одна и та же переменная x; понятно, что принципиального значения эта особенность модели не имеет.

Следует напомнить, что одной из главных гипотез применения статистических методов вообще и МНК в частности является **g1**: уравнения регрессии представляют истинные зависимости, и речь идет лишь об оценке параметров этих истинных зависимостей. В данном случае это означает, что на спрос и предложение влияет только

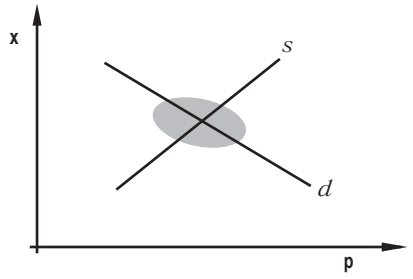

Рис. 10.1

цена, и линии спроса и предложения в плоскости, абсциссой которой является цена, не меняют своего положения. Поэтому наблюдаемые пары (p, x) сконцентрированы вокруг единственной точки равновесия, облако наблюдений не имеет вытянутостей, и зависимости x от p статистически выявить невозможно (рис. 10.1).

Статистически оба уравнения одинаковы, и нет оснований считать коэффициент регрессии, например, x по p, эластичностью спроса или предложения по цене. Более того, в данном случае эта регрессия будет не значима. Эти уравнения не идентифицированы. Действительно, $k = 2$, $n = 0$, $r_1 = r_2 = 0$, и необходимое условие идентификации (10.12) для обоих уравнений не выполнено.

Пусть речь идет о товаре, имеющем сельскохозяйственное происхождение. Тогда его предложение зависит от погодных условий, и в модель следует ввести переменную z_1 — некий индекс погоды в течение сельскохозяйственного цикла. В правую часть соотношения (10.15) вводится дополнительное слагаемое:

$$z_1 \,[\, a_{11} \;\; 0 \,]\,. \tag{10.16}$$

Если модель (10.15, 10.16) истинна (гипотеза **g3**), то подвижной становится линия предложения (погодные условия в разные сельскохозяйственные циклы различны), и облако фактических наблюдений вытягивается вдоль линии спроса. Регрессия x на p дает оценку эластичности спроса по цене (рис. 10.2). В этой ситуации уравнение предложения по-прежнему не идентифицировано, но для уравнения спроса условия идентификации (10.12) выполнены, и это уравнение идентифицировано.

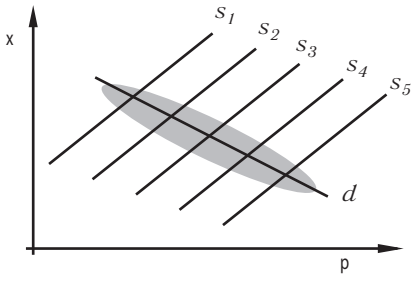

Рис. 10.2

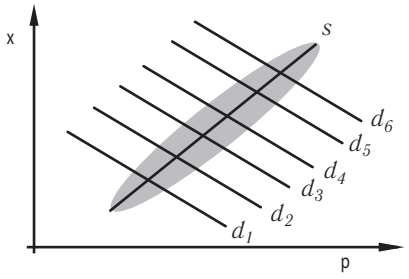

Рис. 10.3

Действительно: $k = 2,\ n = 1,\ r_1 = 0,\ r_2 = 1$ и $r_1 < k - 1,\ r_2 = k - 1$. Более убедительно этот результат можно получить, используя необходимые и достаточные условия идентификации (10.13).

Матрица H в этих условиях имеет следующий вид:

$$H = \begin{bmatrix} 1 & 1 \\ -b_{21} & b_{22} \\ -a_{11} & 0 \\ c_1 & c_2 \end{bmatrix}.$$

Матрица R_1 — пустая ($r_l = 0$), и условия (10.13) для первого уравнения не выполняются. Для второго уравнения $R_2 = [\,0\ \ 0\ \ 1\ \ 0\,]$, и матрица R_2H равна $[\,-a_{11}\ \ 0\,]$, т.е. ее ранг равен единице, и условие (10.13) выполнено. А матрица, составленная из коэффициентов во всех прочих уравнениях, кроме второго, при переменных, которые исключены из второго уравнения, есть $[-a_{11}]$, т.е. она не вырождена.

Теперь рассматривается другая возможность: изучаемый товар входит в потребительскую корзину, и спрос на него зависит от доходов домашних хозяйств. В модель вводится переменная z_2 доходов домашних хозяйств, т.е. в правую часть соотношений (10.15) добавляется слагаемое

$$z_2\,[\,0\ \ a_{22}\,]\,. \tag{10.17}$$

Если истинна модель (10.15, 10.17), то подвижной окажется линия спроса (разные домашние хозяйства имеют разные доходы), и регрессия x на p даст оценку эластичности предложения по цене (рис. 10.3). В такой ситуации не идентифицировано уравнение спроса. Уравнение предложения идентифицировано: $k = 2,\ n = 1,\ r_1 = 1,\ r_2 = 0$ и $r_1 = k - 1,\ r_2 < k - 1$.

Понятно, что можно говорить о модели, в которую входят обе отмеченные переменные: и z_1 и z_2. Это — модель (10.15, 10.16, 10.17). В правую часть (10.15)

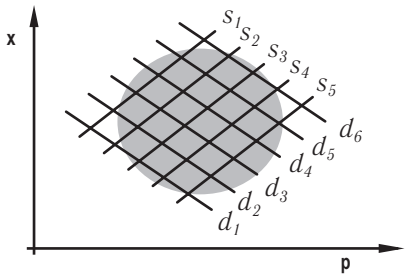

Рис. 10.4

добавляется слагаемое

$$[\,z_1 \;\; z_2\,]\begin{bmatrix} a_{11} & 0 \\ 0 & a_{22} \end{bmatrix}.$$

В этом случае идентифицированы оба уравнения: $k=2$, $n=1$, $r_1=r_2=1=k-1$. Но поскольку подвижны обе линии — и спроса, и предложения — облако наблюдений не имеет вытянутостей (рис. 10.4), и регрессия x на p опять оказывается не значимой. Для оценки параметров регрессии требуется использовать специальные методы, рассматриваемые ниже. Впрочем, и в двух предыдущих случаях необходимо использование специальных методов оценки параметров взаимозависимых систем, т.к. обычный МНК дает смещенные и несостоятельные оценки.

Пусть теперь на предложение товара влияет еще один фактор z_3, показывающий, например, количество удобрений на единицу площади, с которой собирается продукт, принимающий в дальнейшем форму товара. Тогда в правой части уравнения (10.15) возникает слагаемое

$$[\,z_1 \;\; z_3\,]\begin{bmatrix} a_{11} & 0 \\ a_{31} & 0 \end{bmatrix},$$

и первое уравнение по-прежнему остается не идентифицированным, а второе оказывается сверхидентифицированным.

Далее ряд утверждений будет иллюстрироваться на примере модели (10.15, 10.16). В иллюстрациях эту модель удобнее записывать в сокращенном виде:

$$[\,\hat{x} \;\; \hat{p}\,]\begin{bmatrix} 1 & 1 \\ -\beta_{21} & \beta_{22} \end{bmatrix} = \hat{z}_1\,[\,\alpha_{11} \;\; 0\,] + [\,\varepsilon_1 \;\; \varepsilon_2\,]. \qquad (10.18)$$

Поскольку

$$\begin{bmatrix} 1 & 1 \\ -\beta_{21} & \beta_{22} \end{bmatrix}^{-1} = \frac{1}{\beta_{21}+\beta_{22}}\begin{bmatrix} \beta_{22} & -1 \\ \beta_{21} & 1 \end{bmatrix},$$

приведенная форма модели имеет следующий вид:

$$[\hat{x}\ \ \hat{p}] = \hat{z}_1 [d_{11}\ \ d_{12}] + [\eta_1\ \ \eta_2] = \\ = \frac{1}{\beta_{21} + \beta_{22}} (\hat{z}_1 [\alpha_{11}\beta_{22} \quad -\alpha_{11}] + [\varepsilon_1\beta_{22} + \varepsilon_2\beta_{21} \quad \varepsilon_2 - \varepsilon_1]). \quad (10.19)$$

Из этого соотношения видно, как d и η связаны с β и ε.

Дальнейшее изложение ведется в предположении, что строки матрицы R_l — орты.

10.3. Оценка параметров отдельного уравнения

Вводятся дополнительные обозначения:

X^l — $N \times k_l$-матрица наблюдений за изучаемыми переменными x^l, входящими в l-е уравнение;

X_l — N-вектор-столбец наблюдений за l-й переменной x_l;

X^l_- — $N \times (k_l - 1)$-матрица X^l без столбца X_l наблюдений за x^l_-;

β^l — k_l-вектор-столбец параметров при изучаемых переменных в l-м уравнении;

β_l — $(k_l - 1)$-вектор-столбец β^l с обратным знаком и без l-го элемента $\beta_{ll} = 1$;

Z^l — $N \times (n_l + 1)$-матрица наблюдений за независимыми факторами z^l, входящими в l-е уравнение, включая единичный столбец, соответствующий свободному члену;

α_l — $(n_l + 1)$-вектор-столбец параметров при этих факторах вместе со свободным членом;

ε_l — N-вектор-столбец остатков в l-м уравнении по наблюдениям.

Тогда l-е уравнение регрессии можно записать следующим образом:

$$X^l\beta^l = Z^l\alpha_l + \varepsilon_l \quad (10.20)$$

или

$$X_l = X^l_-\beta_l + Z^l\alpha_l + \varepsilon_l. \quad (10.21)$$

Применение обычного МНК к этому уравнению дает в общем случае смещенные и несостоятельные оценки, прежде всего потому, что остатки ε_l скорее всего коррелированы с регрессорами X^l_-, которые к тому же недетерминированы и наблюдаются с ошибками (гипотеза **g2** нарушена).

Для иллюстрации справедливости этого утверждения используется модель (10.15, 10.16). Пусть эта модель истинна, и тогда регрессия x на p даст оценку $-\beta_{22}$:

$$-b_{22}^{\text{МНК}} = \frac{\sum \hat{x}_i \hat{p}_i}{\sum \hat{p}_i^2}. \tag{10.22}$$

Это выражение можно преобразовать, используя (10.18, 10.19) (чтобы не загромождать записи, $\frac{1}{\sum \hat{p}_i^2}$ обозначено через P):

$$-b_{22}^{\text{МНК}} = P\sum \hat{x}_i \hat{p}_i \overset{\hat{x}_i = -\beta_{22}\hat{p}_i + \varepsilon_{i2}}{=} -\beta_{22} + P\sum \varepsilon_{i2}\hat{p}_i \overset{\hat{p}_i = \hat{z}_{i1}d_{12} + \eta_{i2}}{=}$$

$$= -\beta_{22} + P\left(d_{12}\sum \hat{z}_{i1}\varepsilon_{i2} + \sum \eta_{i2}\varepsilon_{i2}\right) \overset{\eta_{i2} = \frac{\varepsilon_{i2} - \varepsilon_{i1}}{\beta_{21} + \beta_{22}}}{=} -\beta_{22} +$$

$$+ P\left(d_{12}\sum \hat{z}_{i1}\varepsilon_{i2} + \frac{1}{\beta_{21} + \beta_{22}}\left(\sum \varepsilon_{i2}^2 - \sum \varepsilon_{i1}\varepsilon_{i2}\right)\right).$$

Очевидно, что $-b_{22}^{\text{МНК}}$ по математическому ожиданию никак не может равняться $-\beta_{22}$, поскольку в правой части полученного выражения имеется $\sum \varepsilon_{i2}^2$, т.е. дисперсия (в математическом ожидании) остатка в уравнении по спросу, которая не равна нулю и к тому же не будет уменьшаться с ростом N. Эта оценка смещена и несостоятельна.

Если данное уравнение точно идентифицировано, то для оценки его параметров можно использовать **косвенный метод** (КМ) наименьших квадратов: с помощью МНК оцениваются параметры приведенной формы системы уравнений, через которые однозначно выражаются структурные параметры данного уравнения.

В качестве примера можно использовать оценку параметров второго уравнения модели (10.15, 10.16), которое точно идентифицировано. Действительно, параметры приведенной формы модели однозначно определяют оценку $-\beta_{22}$, как это следует из (10.19):

$$-b_{22}^{KM} = \frac{d_{11}}{d_{12}}. \tag{10.23}$$

Поскольку

$$d_{11} = \frac{\sum \hat{x}_i \hat{z}_{i1}}{\sum \hat{z}_{i1}^2}, \quad d_{12} = \frac{\sum \hat{p}_i \hat{z}_{i1}}{\sum \hat{z}_{i1}^2},$$

то соотношение (10.23) означает, что

$$-b_{22}^{KM} = \frac{\sum \hat{x}_i \hat{z}_{i1}}{\sum \hat{p}_i \hat{z}_{i1}},$$

т.е. что (ср. с (10.22)) используется метод инструментальных переменных с z_1 в качестве инструментальной переменной.

Можно записать уравнения для оценки косвенным методом в общем случае.

Сначала следует обратить внимание на то, что условия (10.11) эквивалентны требованиям

$$T_l^B \beta^l = B_l, \quad T_l^A \alpha_l = A_l, \tag{10.24}$$

где T_l^B — $k \times k_l$-матрица, полученная из I_k вычеркиванием столбцов, соответствующих тем изучаемым переменным, которые исключены из l-го уравнения;

T_l^A — аналогичная $(n+1) \times (n_l+1)$-матрица для A_l.

B_l и A_l имеют нулевые компоненты, соответствующие исключенным из l-го уравнения переменным.

Далее необходимо учесть, что параметры структурной формы, удовлетворяющие условиям (10.24), должны для своей идентификации еще удовлетворять соотношениям (10.10). Тем самым получается система уравнений для нахождения параметров структурной формы:

$$DT_l^B b^l - T_l^A a_l = 0,$$

или по определению матрицы T_l^B:

$$D^l b^l - T_l^A a_l = 0,$$

где D^l — оценки параметров приведенной формы уравнений для изучаемых переменных, вошедших в l-е уравнение, или, наконец,

$$D_l = D_-^l b_l + T_l^A a_l, \tag{10.25}$$

где D_l — оценки параметров l-го уравнения в приведенной форме,

D_-^l — оценки параметров приведенной формы уравнений для изучаемых переменных, вошедших в правую часть l-го уравнения.

Эти матрицы коэффициентов приведенной формы представляются следующим образом:

$$D^l = (Z'Z)^{-1} Z' X^l, \;\; D_l = (Z'Z)^{-1} Z' X_l, \;\; D_-^l = (Z'Z)^{-1} Z' X_-^l.$$

Система уравнений (10.25) может быть также получена умножением обеих частей системы (10.21) слева на $(Z'Z)^{-1}Z'$, т.к. третье слагаемое правой части отбрасывается (МНК-остатки должны быть ортогональны регрессорам), а во 2-м слагаемом $(Z'Z)^{-1}Z'Z^l$ заменяется на T_l^A (т.к. по определению этой матрицы $Z^l = ZT_l^A$).

В общем случае, матрица этой системы $\left[D_-^l \;\; T_l^A\right]$ имеет размерность $(n+1) \times$ $\times (k_l + n_l)$. Первый ее блок имеет размерность $(n+1) \times (k_l - 1)$, второй — $(n+1) \times (n_l+1)$.

В случае точной идентификации и строгого выполнения условий (10.14) эта матрица квадратна и не вырождена. Система (10.25) дает единственное решение — оценку параметров структурной формы l-го уравнения косвенным методом наименьших квадратов.

В структурной форме со скрытым свободным членом модель (10.15+10.16) записывается следующим образом:

$$\begin{bmatrix} X & P \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -b_{21} & b_{22} \end{bmatrix} = [\, Z_1 \;\; 1_N \,] \begin{bmatrix} a_{11} & 0 \\ c_1 & c_2 \end{bmatrix} + [\, e_1 \;\; e_2 \,],$$

а ее второе, точно идентифицированное уравнение в форме (10.21) —

$$X = P(-b_{22}) + [\, Z_1 \;\; 1_N \,] \begin{bmatrix} 0 \\ c_2 \end{bmatrix} + [\, e_1 \;\; e_2 \,]. \tag{10.26}$$

Как это было показано выше, обе части (10.26) умножаются на матрицу

$$\begin{bmatrix} \begin{bmatrix} Z_1' \\ 1_N' \end{bmatrix} & \begin{bmatrix} Z_1 & 1_N \end{bmatrix} \end{bmatrix}^{-1} \begin{bmatrix} Z_1' \\ 1_N' \end{bmatrix}:$$

$$\begin{bmatrix} d_{11} \\ d_{21} \end{bmatrix} = \begin{bmatrix} d_{12} \\ d_{22} \end{bmatrix} (-b_{22}) + \begin{bmatrix} 0 \\ c_2 \end{bmatrix},$$

или

$$D_1 = D_2(-b_{22}) + T_2^A c_2, \;\; \text{где} \;\; T_2^A = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Непосредственно в форме (10.25) при учете условий нормализации эта система записалась бы в виде:

$$D_2 b_{22} = -D_1 + T_2^A c_2.$$

Из решения этой системы $-b_{22}^{KM}$ получается таким же, как в (10.23), кроме того, получается оценка свободного члена:

$$c_2^{KM} = d_{21} - d_{22}\frac{d_{11}}{d_{12}}.$$

Если уравнение не идентифицировано, переменных в системе (10.21) оказывается больше, чем уравнений, и эта система представляет бесконечное множество значений параметров структурной формы. Чтобы выбрать из этого множество какое-то решение, часть параметров структурной формы надо зафиксировать, т.е. сделать уравнение идентифицированным.

Для сверхидентифицированного уравнения система (10.21) является переопределенной, и ее уравнения не могут выполняться как равенства. Различные методы оценки такого уравнения реализуют различные подходы к минимизации невязок по уравнениям этой системы.

Одним из таких методов является **двухшаговый метод** (2М) наименьших квадратов.

На первом шаге с помощью МНК оцениваются параметры приведенной формы для переменных X_-^l:

$$X_-^l = ZD_-^l + V^l,$$

где V^l — $N \times (k_l - 1)$-матрица остатков по уравнениям; и определяются расчетные значения этих переменных уже без ошибок:

$$X_-^{lc} = ZD_-^l.$$

На втором шаге с помощью МНК оцениваются искомые параметры структурной формы из уравнения:

$$X_l = X_-^{lc} b_l + Z^l a_l + e_l. \tag{10.27}$$

Для этого уравнения гипотеза **g2** выполняется, т.к. регрессоры не имеют ошибок, и поэтому применим обычный МНК.

Можно определить единый оператор 2М-оценивания. Поскольку

$$X_-^{lc} = FX_-^l,$$

где $F = Z(Z'Z)^{-1}Z'$, уравнение (10.24) записывается как:

$$X_l = \begin{bmatrix} FX_-^l & Z^l \end{bmatrix} \begin{bmatrix} b_l \\ a_l \end{bmatrix} + e_l, \tag{10.28}$$

а оператор, входящий в него, как:

$$\begin{bmatrix} b_l \\ a_l \end{bmatrix} = \begin{bmatrix} X_-^{l\prime} F X_-^l & X_-^{l\prime} Z^l \\ Z^{l\prime} X_-^l & Z^{l\prime} Z^l \end{bmatrix}^{-1} \begin{bmatrix} X_-^{l\prime} F X_l \\ Z^{l\prime} X_l \end{bmatrix}. \tag{10.29}$$

Оператор в такой форме получается как результат применения МНК к уравнению (10.25), т.е. результат умножения обеих частей этого уравнения слева на транспонированную матрицу регрессоров и отбрасывания компоненты остатков:

$$\begin{bmatrix} X_-^{l\prime}F \\ Z^{l\prime} \end{bmatrix} X_l = \begin{bmatrix} X_-^{l\prime}F \\ Z^{l\prime} \end{bmatrix} \begin{bmatrix} FX_-^l & Z^l \end{bmatrix} \begin{bmatrix} b_l \\ a_l \end{bmatrix}_l . \qquad (10.30)$$

Откуда следует оператор 2М-оценивания в указанной форме, т.к. F — симметричная идемпотентная матрица и

$$FZ^l = FZT_l^A = ZT_l^A = Z^l.$$

Такой оператор оценивания сверхидентифицированного уравнения можно получить, если МНК применить к системе (10.21) (в этом случае она переопределена и в ее уравнениях возникают невязки), умножив предварительно обе ее части слева на Z.

Система нормальных уравнений для оценки (10.21), умноженной на Z, записывается следующим образом:

$$\begin{bmatrix} D_-^{l\prime} \\ T_l^{A\prime} \end{bmatrix} Z'ZD_l = \begin{bmatrix} D_-^{l\prime} \\ T_l^{A\prime} \end{bmatrix} Z'Z \begin{bmatrix} D_-^l & T_l^A \end{bmatrix} \begin{bmatrix} b_l \\ a_l \end{bmatrix},$$

и, учитывая, что

$$D_-^{l\prime}Z'ZD_l = X_-^{l\prime}FX_l, \quad T_l^{A\prime}Z'ZD_l = Z^{l\prime}X_l \text{ и т.д.},$$

она преобразуется к виду (10.29).

Отсюда, в частности, следует, что для точно идентифированного уравнения 2М-оценка совпадает с КМ-оценкой, т.к. параметры структурной формы уравнения, однозначно определяемые соотношениями (10.21), удовлетворяют в этом случае и условиям (10.25).

Соотношения (10.29) — первая форма записи оператора 2М-оценивания. Если в (10.24) учесть, что $X_-^{l^c} = X_-^l - V^l$, этот оператор можно записать в более прозрачной второй форме:

$$\begin{bmatrix} b_l \\ a_l \end{bmatrix} = \begin{bmatrix} X_-^{l\prime}X_-^l - V^{l\prime}V^l & X_-^{l\prime}Z^l \\ Z^{l\prime}X_-^l & Z^{l\prime}Z^l \end{bmatrix}^{-1} \begin{bmatrix} (X_-^{l\prime} - V^{l\prime})X_l \\ Z^{l\prime}X_l \end{bmatrix}. \qquad (10.31)$$

Это доказывается аналогично с учетом того, что остатки V^l ортогональны регрессорам Z и, соответственно,

$$Z'V^l = 0, \quad X_-^{l'}V^l = V^{l'}V^l, \quad X_-^{l'c}V^l = 0.$$

Попытка применить оператор 2M-оценивания для не идентифицированного уравнения не имеет смысла, т.к. обращаемая матрица в данном операторе вырождена.

В этом легко убедиться, т.к.

$$\left[FX_-^l \;\; Z^l\right] = Z\left[D_-^l \;\; T_l^A\right],$$

т.е. матрица наблюдений за регрессорами в (10.25) получается умножением на Z слева матрицы системы (10.21). В последней, если уравнение не идентифицировано, — столбцов больше, чем строк. Следовательно, регрессоры в (10.25) линейно связаны между собой, а матрица системы нормальных уравнений (матрица оператора оценивания) вырождена.

Для сверхидентифицированного уравнения можно использовать также **метод наименьшего дисперсионного отношения** (МНДО). Строгое обоснование его применимости вытекает из метода максимального правдоподобия.

Пусть b^l в уравнении (10.20) оценено, и $X^l b^l$ рассматривается как единая эндогенная переменная. В результате применения МНК определяются:

$$\begin{aligned} &a_l = (Z^{l'}Z^l)^{-1}Z^{l'}X^l b^l, \\ &e_l = (I_N - F^l)X^l b^l, \qquad \text{где } F^l = Z^l(Z^{l'}Z^l)^{-1}Z^{l'}, \\ &e_l'e_l = b^{l'}W^l b^l, \qquad\quad \text{где } W^l = X^{l'}(I_N - F^l)X^l. \end{aligned} \tag{10.32}$$

Теперь находится остаточная сумма квадратов при условии, что все экзогенные переменные входят в l-е уравнение. Она равна $b^{l'}Wb^l$, где $W = X^{l'}(I_N - F)X^l$. Тогда b^l должны были бы быть оценены так, чтобы

$$\lambda = \frac{b^{l'}W^l b^l}{b^{l'}Wb^l} \;\to\; \min!$$

Иначе было бы трудно понять, почему в этом уравнении присутствуют не все экзогенные переменные.

Решение этой задачи приводит к следующим условиям:

$$(W^l - \lambda W)b^l = 0. \tag{10.33}$$

Действительно, из условия равенства нулю первой производной:

$$\frac{\partial\lambda}{\partial b^l} = \frac{2W^l b^l (b^{l'} W b^l) - 2W b^l (b^{l'} W^l b^l)}{(b^{l'} W b^l)^2} = \frac{2}{b^{l'} W b^l}(W^l b^l - \lambda W b^l) = 0,$$

сразу следует (10.33).

Следовательно, λ находится как минимальный корень характеристического уравнения (см. Приложение А.1.2)

$$\left|W^l - \lambda W\right| = 0,$$

а b^l определяется из 10.33 с точностью до постоянного множителя, т.е. с точностью до нормировки $b_{ll} = 1$.

В общем случае $\lambda_{\min} > 1$, но при правильной спецификации модели $\lambda_{\min} \underset{N\to\infty}{\longrightarrow} 1$.

Оператор

$$\begin{bmatrix} b_l \\ a_l \end{bmatrix} = \begin{bmatrix} X_-^{l'} X_-^l - kV^{l'} V^l & X_-^{l'} Z^l \\ Z^{l'} X_-^l & Z^{l'} Z^l \end{bmatrix}^{-1} \begin{bmatrix} (X_-^{l'} - kV^{l'}) X_l \\ Z^{l'} X_l \end{bmatrix}$$

позволяет получить так называемые **оценки k-класса** (не путать с k — количеством эндогенных переменных в системе).

При $k = 0$, они являются обычными МНК-оценками для l-го уравнения, что легко проверяется; при $k = 1$, это — 2М-оценки; при $k = \lambda_{min}$ — МНДО-оценки (принимается без доказательства). 2М-оценки занимают промежуточное положение между МНК- и МНДО-оценками (т.к. $\lambda_{min} > 1$). Исследования показывают, что эффективные оценки получаются при $k < 1$.

10.4. Оценка параметров системы идентифицированных уравнений

Из приведенной формы системы уравнений следует, что

$$x'\varepsilon = (B^{-1})' A' z' \varepsilon + (B^{-1})' \varepsilon' \varepsilon.$$

Как и прежде, в любом наблюдении $\mathbf{E}(\varepsilon) = 0$, $\mathbf{E}(\varepsilon'\varepsilon) = \sigma^2\Omega$, и ошибки не коррелированы по наблюдениям. Тогда

$$\mathbf{E}(x'\varepsilon) = (B^{-1})' \mathbf{E}(\varepsilon'\varepsilon) = \sigma^2 (B^{-1})' \Omega,$$

т.е. в общем случае все эндогенные переменные коррелированы с ошибками во всех уравнениях. Это является основным препятствием для применения обычного МНК ко всем уравнениям по отдельности.

Но в случае, если в матрице B все элементы, расположенные ниже главной диагонали, равны нулю, т.е. в правой части l-го уравнения могут появляться только более младшие эндогенные переменные $x_{l'}$, $l' < l$, и последней компонентой любого вектора x^l является x_l, а матрица Ω диагональна, то ε_l не коррелирует с переменными x^l_- при любом l. Это — **рекурсивная система**, и для оценки ее параметров можно применять МНК к отдельным уравнениям.

Для оценки параметров всех идентифицированных уравнений системы можно применить **трехшаговый метод** (3М) наименьших квадратов.

Первые два шага 3М совпадают с 2М, но представляются они по сравнению с предыдущим пунктом в несколько иной форме.

Предполагается, что идентифицированы все k уравнений:

$$X_l = X^l_-\beta_l + Z^l\alpha_l + \varepsilon_l = Q^l\gamma_l + \varepsilon_l, \quad l = 1, \ldots, k,$$

где $Q^l = [X^l_-, Z^l]$, $\gamma_l = [\,\beta_l \;\; \alpha_l\,]'$. Учитывая указанные выше свойства остатков:

$$\mathbf{E}(\varepsilon_l\varepsilon_l') = \sigma^2\omega_{ll}I_N, \quad \mathbf{E}(\varepsilon_{l'}\varepsilon_l') = \sigma^2\omega_{l'l}I_N.$$

Теперь обе части l-го уравнения умножаются слева на Z':

$$Z'X_l = Z'Q^l\gamma_l + Z'\varepsilon_l, \tag{10.34}$$

и $Z'X_l$ рассматривается как вектор $n+1$ наблюдений за одной эндогенной переменной, а $Z'Q^l$ — как матрица $n+1$ наблюдений за $n_l + k_l$ экзогенными переменными, включая свободный член. Так как все уравнения идентифицированы, и выполнено условие (10.14), во всех этих новых регрессиях количество наблюдений не меньше количества оцениваемых параметров. Для сверхидентифицированных уравнений количество наблюдений в новой регрессии будет превышать количество оцениваемых параметров. Это более естественный случай. Поэтому 3М-метод обычно применяют для всех сверхидентифицированных уравнений системы.

Матрица ковариации остатков по уравнению (10.34) равна $\sigma^2\omega_{ll}Z'Z$. Она отлична от $\sigma^2 I_N$, и для получения оценок c_l параметров γ_l этого уравнения нужно использовать ОМНК:

$$c_l = (Q^{l'}Z(Z'Z)^{-1}Z'Q^l)^{-1}Q^{l'}Z(Z'Z)^{-1}Z'X_l, \text{ или}$$
$$c_l = (Q^{l'}FQ^l)^{-1}Q^{l'}FX_l.$$

Сравнив полученное выражение с (10.29), легко убедится в том, что c_l — 2М-оценка.

Если 2М на этом заканчивается, то в 3М полученные оценки c_l используются для того, чтобы оценить e_l, и затем получить оценки W матрицы $\sigma^2\Omega$:

$$w_{ll} = \frac{1}{N} e_l' e_l, \quad w_{l'l} = \frac{1}{N} e_{l'}' e_l.$$

Теперь все уравнения (10.34) записываются в единой системе (подобная запись использовалась в п.10.1 при доказательстве одного из утверждений):

$$\begin{bmatrix} Z'X_1 \\ Z'X_2 \\ \vdots \\ Z'X_k \end{bmatrix} = \begin{bmatrix} Z'Q^1 & 0 & \cdots & 0 \\ 0 & Z'Q^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & Z'Q^k \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_k \end{bmatrix} + \begin{bmatrix} Z'\varepsilon_1 \\ Z'\varepsilon_2 \\ \vdots \\ Z'\varepsilon_k \end{bmatrix}, \tag{10.35}$$

или

$$Y = Q\gamma + \eta,$$

где Y — соответствующий $k \cdot (n+1)$-вектор-столбец наблюдений за изучаемой переменной;

Q — $k(n+1) \times \sum\limits_{l=1}^{k} (k_l + n_l)$-матрица наблюдений за экзогенными переменными;

γ — $\sum\limits_{l=1}^{k} (k_l + n_l)$-вектор-столбец параметров регрессии;

η — $k(n+1)$-вектор-столбец остатков по наблюдениям.

Легко проверить, что матрица ковариации остатков η удовлетворяет следующему соотношению:

$$\mathbf{E}(\eta\eta') = \sigma^2\Omega \otimes (Z'Z).$$

Для нее имеется оценка: $k(n+1) \times (n+1)$-матрица $\Sigma = W \otimes (Z'Z)$. Эта матрица отлична от $\sigma^2 I_{k(n+1)}$, поэтому на третьем шаге 3М-оценивания к единой системе (10.35) применяется ОМНК и получается окончательная оценка c параметров γ:

$$c = (Q'\Sigma^{-1}Q)^{-1}Q'\Sigma^{-1}Y.$$

10.5. Упражнения и задачи

Упражнение 1

Рассматривается простая Кейнсианская модель:

$$\begin{cases} c = \alpha 1_N + \beta y + \varepsilon, \\ y = c + i, \end{cases}$$

где c, i и y — объем потребления, инвестиции и доход соответственно, 1_N — столбец, состоящий из единиц. Пусть каждый вектор имеет размерность 20×1, $\mathbf{E}(\varepsilon) = 0$ и $\mathbf{E}(\varepsilon\varepsilon') = \sigma^2 I_N = 0.2^2 I_{20}$. Система уравнений приведенной формы следующая:

$$\begin{cases} c = \frac{\alpha}{1-\beta} 1_N + \frac{\beta}{1-\beta} i + \frac{1}{1-\beta}\varepsilon, \\ y = \frac{\alpha}{1-\beta} 1_N + \frac{1}{1-\beta} i + \frac{1}{1-\beta}\varepsilon, \end{cases}$$

Ошибки в приведенной форме для c и y таковы:

$$\eta_1 = \eta_2 = \frac{1}{1-\beta}\varepsilon = \frac{1}{1-0.8}\varepsilon = \frac{1}{0.2}\varepsilon,$$

т.е. в модели в приведенной форме ошибки η_1 и η_2 распределены как $N(0, I)$. В таблице 10.1 на основе заданных 20-ти гипотетических значений для i (первая колонка) и нормально распределенных ошибок (последняя колонка) получены данные для c и y из уравнений приведенной формы, используя значения параметров $\alpha = 2$ и $\beta = 0.8$.

В реальной ситуации существуют только значения i, c и y. Значения ошибки в модели и значения α и β неизвестны.

Таблица 10.1

i	c	y	$\eta_1 = \eta_2$
2.00	18.19	20.19	0.193
2.00	17.50	19.50	–0.504
2.20	16.48	18.68	–2.318
2.20	19.06	21.26	0.257
2.40	21.38	23.78	1.784
2.40	21.23	23.63	1.627
2.60	21.11	23.71	0.708
2.60	22.65	25.25	2.252
2.80	20.74	23.54	–0.462
2.80	19.85	22.65	–1.348
3.00	22.23	25.23	0.234
3.00	22.23	25.23	0.226
3.20	23.43	26.63	0.629
3.20	23.04	26.24	0.244
3.40	23.03	26.43	–0.569
3.40	24.45	27.85	0.853
3.60	26.63	30.23	2.227
3.60	24.47	28.07	0.074
3.80	24.67	28.47	–0.527
3.80	26.00	29.80	0.804

1.1. Используя данные таблицы 10.1, оцените уравнения приведенной формы для объема потреблении и дохода.

1.2. Используя данные таблицы 10.1, посчитайте косвенные МНК-оценки для α и β из

а) уравнения приведенной формы для объема потребления и

б) уравнения приведенной формы для дохода.

Идентичны ли косвенные МНК-оценки, полученные из обоих уравнений приведенной формы?

1.3. Используя данные таблицы 10.1, посчитайте простые МНК-оценки для α и β и сравните их с косвенными МНК-оценками из упражнениия 1.2.

1.4. Используя данные таблицы 10.1 для i и используя значения параметров $\alpha = 2$ и $\beta = 0.8$ составьте 100 выборок для c и y.

1.5. Примените простой МНК к каждому структурному уравнению системы для 100 выборок. Посчитайте среднее 100 оценок α и β. Проверьте степень эмпирического смещения.

1.6. Посчитайте косвенные МНК-оценки для α и β для 100 выборок. Посчитайте среднее 100 оценок α и β. Посчитайте степень смещения в маленьких выборках — размером по 20 наблюдений. Сравните смещение косвенных МНК-оценок со смещением обычных МНК-оценок.

1.7. Объедините пары выборок так, чтобы получились 50 выборок по 40 наблюдений. Посчитайте косвенные МНК-оценки для α и β для этих 50 выборок. Посчитайте среднее и проверьте смещение оценок. Будут ли эмпирические смещения в этом случае меньше, чем рассчитанные из 100 выборок по 20 наблюдений?

Упражнение 2

Таблица 10.2 содержит векторы наблюдений z_1, z_2, z_3, z_4, z_5 и x_1, x_2, x_3 которые представляют выборку, полученную из модели:

$$x_1 = \beta_{12}x_2 + \beta_{13}x_3 + \alpha_{11}z_1 + \varepsilon_1,$$
$$x_2 = \beta_{21}x_1 + \alpha_{21}z_1 + \alpha_{22}z_2 + \alpha_{23}z_3 + \alpha_{24}z_4 + \varepsilon_2,$$
$$x_3 = \beta_{32}x_2 + \alpha_{31}z_1 + \alpha_{32}z_2 + \alpha_{35}z_5 + \varepsilon_3,$$

Таблица 10.2

z_1	z_2	z_3	z_4	z_5	x_1	x_2	x_3
1	3.06	1.34	8.48	28.00	359.27	102.96	578.49
1	3.19	1.44	9.16	35.00	415.76	114.38	650.86
1	3.30	1.54	9.90	37.00	435.11	118.23	684.87
1	3.40	1.71	11.02	36.00	440.17	120.45	680.47
1	3.48	1.89	11.64	29.00	410.66	116.25	642.19
1	3.60	1.99	12.73	47.00	530.33	140.27	787.41
1	3.68	2.22	13.88	50.00	557.15	143.84	818.06
1	3.72	2.43	14.50	35.00	472.80	128.20	712.16
1	3.92	2.43	15.47	33.00	471.76	126.65	722.23
1	4.15	2.31	16.61	40.00	538.30	141.05	811.44
1	4.35	2.39	17.40	38.00	547.76	143.71	816.36
1	4.37	2.63	18.83	37.00	539.00	142.37	807.78
1	4.59	2.69	20.62	56.00	677.60	173.13	983.53
1	5.23	3.35	23.76	88.00	943.85	223.21	1292.99
1	6.04	5.81	26.52	62.00	893.42	198.64	1179.64
1	6.36	6.38	27.45	51.00	871.00	191.89	1134.78
1	7.04	6.14	30.28	29.00	793.93	181.27	1053.16
1	7.81	6.14	25.40	22.00	850.36	180.56	1085.91
1	8.09	6.19	28.84	38.00	967.42	208.24	1246.99
1	9.24	6.69	34.36	41.00	1102.61	235.43	1401.94

или в матричной форме: $XB = ZA + \varepsilon$, где ε_i — нормально распределенные векторы с $\mathbf{E}(\varepsilon_i) = 0$ и

$$\mathbf{E}\left(\varepsilon\varepsilon'\right) = \mathbf{E}\left(\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \varepsilon_3\end{bmatrix}\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \varepsilon_3\end{bmatrix}'\right) = \Sigma \otimes I_N.$$

Гипотетические структурные матрицы коэффициентов B, A и ковариационная матрица Σ следующие:

$$B = \begin{bmatrix} 0 & 0.2 & 0\\ -10 & -1 & 2\\ 2.5 & 0 & -1\end{bmatrix}, \quad A = \begin{bmatrix} 60 & -40 & 10\\ 0 & 4 & -80\\ 0 & 6 & 0\\ 0 & -1.5 & 0\\ 0 & 0 & -5\end{bmatrix},$$

$$\Sigma = \begin{bmatrix} 227.55 & 8.91 & -56.89\\ 8.91 & 0.66 & -1.88\\ -56.89 & -1.88 & 15.76\end{bmatrix}$$

Матрица коэффициентов в приведенной форме для гипотетической модели следующая:

$$D = AB^{-1} = \begin{bmatrix} -142.50 & 11.50 & 13.00\\ 110.00 & 18.00 & 116.00\\ 15.00 & -3.00 & -6.00\\ -3.75 & 0.75 & 1.50\\ 6.25 & 1.25 & 7.50\end{bmatrix}$$

В реальной ситуации B, A, Σ, D были бы неизвестны, доступны были бы только наблюдения в таблице 10.2.

2.1. Используя данные таблицы 10.2, проверьте каждое структурное уравнение системы на идентифицируемость.

2.2. Оцените матрицу параметров приведенной формы $D = (Z'Z)^{-1}Z'X$.

2.3. Примените простой МНК к каждому структурному уравнению системы и оцените матрицы B и A.

2.4. Рассчитайте

$$\begin{bmatrix} b_l \\ a_l \end{bmatrix} = \begin{bmatrix} X_-^{l'}X_-^l - kV^{l'}V^l & X_-^{l'}Z^l \\ Z^{l'}X_-^l & Z^{l'}Z^l \end{bmatrix}^{-1} \begin{bmatrix} (X_-^{l'} - kV^{l'})X_l \\ Z^{l'}X_l \end{bmatrix} \qquad (10.36)$$

при $k = 0$ и сравните с результатом упражнения 2.3.

2.5. Используя косвенный МНК, оцените параметры второго строго идентифицированого уравнения.

2.6. Найдите b_2 и a_2, решая систему $D_2 = D_-^2 b_2 + T_2^{A'}a_2$, и сравните с результатом упражнения 2.5.

2.7. Найдите минимальный корень λ из уравнения $\left|W^l - \lambda W\right| = 0$ и, используя формулу метода наименьшего дисперсионного отношения (10.36) при $k = \lambda$, оцените параметры в каждом из трех структурных уравнений.

2.8. Используя формулу двухшагового метода наименьших квадратов (10.36) при $k = 1$, сравните оценки матрицы D, полученные на основе оценок простым МНК, МНДО и 2МНК, с исходными гипотетическими матрицами параметров приведенной формы.

2.9. Используя формулу 3МНК, оцените параметры первого и третьего структурных уравнений совместно.

Упражнение 3

Имеем модель Клейна, в которой

$C = \alpha P + \beta(W + V) + \chi P_{-1} + \delta + \varepsilon_1$ — функция потребления,

$I = \varphi P + \gamma P_{-1} + \eta K_{-1} + \pi + \varepsilon_2$ — функция инвестиционного спроса,

$W = \mu(Y + T - V) + \theta(Y_{-1} + T_{-1} - V_{-1}) + \psi t + \zeta + \varepsilon_3$ — функция спроса на труд.

Выполняются следующие макроэкономические соотношения:

$$Y + T = C + I + G, \quad Y = W + V + P, \quad K = K_{-1} + I,$$

где C — потребительские расходы, I — инвестиционные расходы, G — государственные расходы, P — прибыль, W — спрос на труд негосударственного сектора, V — спрос на труд государственного сектора, K — капитал, T — налоги, t — время, Y — чистый доход от налогов.

На основе данных из таблицы 10.3 оценить параметры модели Клейна простым методом наименьших квадратов и двухшаговым методом наименьших квадратов. Показать величину смещения оценок.

Задачи

1. Эконометрическая модель описана следующими уравнениями:

$$\begin{cases} x_1 = \alpha_{10} + \alpha_{11} z_1 + \beta_{12} x_2 + \varepsilon_1, \\ x_2 = \alpha_{20} + \beta_{21} x_1 + \varepsilon_2, \end{cases}$$

   где x_1 и x_2 — эндогенные переменные, z_1 — экзогенная переменная, ε_1 и ε_2 — случайные ошибки. Определите направление смещения оценки для β_{21}, если для оценивания второго уравнения используется метод наименьших квадратов.

2. Дана следующая макроэкономическая модель:

   $y = c + i + g$ — макроэкономическое тождество;
   $c = \alpha_{10} + \beta_{11} y$ — функция потребления,
   $i = \alpha_{20} + \beta_{21} y - \beta_{22} r$ — функция инвестиций,
   $(m/p) = \beta_{31} y - \beta_{32} r$ — уравнение денежного рынка,

   где эндогенными переменными являются доход y, потребление c, инвестиции i и процентная ставка r. Переменные g (государственные расходы) и (m/p) (реальная денежная масса) — экзогенные. Проверьте, является ли данная система идентифицируемой, и перепишите модель в приведенной форме.

3. Дана следующая модель краткосрочного равновесия для малой открытой экономики (модель Манделла—Флеминга):

   $y = c + i + nx$ — макроэкономическое тождество,

   $c = \alpha_{11} + \beta_{11} y + \varepsilon_1$ — функция потребления,

   $i = \alpha_{21} - \alpha_{21} r + \beta_{21} y + \varepsilon_2$ — функция инвестиций,

   $nx = \alpha_{31} - \beta_{31} y - \beta_{32} ec + \varepsilon_3$ — функция чистого экспорта,

   $(m/p) = \beta_{41} y - \alpha_{41} r + \varepsilon_4$ — уравнение денежного рынка,

Таблица 10.3. (Источник: G.S. Maddala(1977), Econometrics, p. 237)

t	C	P	W	I	K_{-1}	V	G	T
1920	39.8	12.7	28.8	2.7	180.1	2.2	2.4	3.4
1921	41.9	12.4	25.5	-0.2	182.8	2.7	3.9	7.7
1922	45	16.9	29.3	1.9	182.6	2.9	3.2	3.9
1923	49.2	18.4	34.1	5.2	184.5	2.9	2.8	4.7
1924	50.6	19.4	33.9	3	189.7	3.1	3.5	3.8
1925	52.6	20.1	35.4	5.1	192.7	3.2	3.3	5.5
1926	55.1	19.6	37.4	5.6	197.8	3.3	3.3	7
1927	56.2	19.8	37.9	4.2	203.4	3.6	4	6.7
1928	57.3	21.1	39.2	3	207.6	3.7	4.2	4.2
1929	57.8	21.7	41.3	5.1	210.6	4	4.1	4
1930	55	15.6	37.9	1	215.7	4.2	5.2	7.7
1931	50.9	11.4	34.5	–3.4	216.7	4.8	5.9	7.5
1932	45.6	7	29	–6.2	213.3	5.3	4.9	8.3
1933	46.5	11.2	28.5	–5.1	207.1	5.6	3.7	5.4
1934	48.7	12.3	30.6	–3	202	6	4	6.8
1935	51.3	14	33.2	–1.3	199	6.1	4.4	7.2
1936	57.7	17.6	36.8	2.1	197.7	7.4	2.9	8.3
1937	58.7	17.3	41	2	199.8	6.7	4.3	6.7
1938	57.5	15.3	38.2	–1.9	201.8	7.7	5.3	7.4
1939	61.6	19	41.6	1.3	199.9	7.8	6.6	8.9
1940	65	21.1	45	3.3	201.2	8	7.4	9.6
1941	69.7	23.5	53.3	4.9	204.5	8.5	14	12

где эндогенными переменными являются доход y, потребление c, инвестиции i, чистый экспорт nx и валютный курс ec. Переменные r (процентная ставка, значение которой формируется на общемировом уровне) и (m/p) (реальная денежная масса) — экзогенные; $\varepsilon_1, \ldots, \varepsilon_4$ — случайные ошибки. Запишите общие условия для определения структурных параметров каждого уравнения модели. Какие уравнения модели точно идентифицируемы? Перепишите модель Манделла—Флеминга в приведенной форме.

4. Приведите пример системы одновременных уравнений, к которой можно применить косвенный МНК (с объяснением обозначений).

5. Приведите пример сверхидентифицированной системы одновременных уравнений (с объяснением обозначений).

6. Рассмотрите модель:

$$x_{1t} = \beta_{12}x_{2t} + \alpha_{11}z_{1t} + \alpha_{12}z_{2t} + \alpha_{13}z_{3t} + \alpha_{14}z_{4t} + \varepsilon_{1t},$$

$$x_{2t} = \beta_{21}x_{1t} + \alpha_{21}z_{1t} + \alpha_{22}z_{2t} + \alpha_{23}z_{3t} + \alpha_{24}z_{4t} + \varepsilon_{2t},$$

где вектор z — экзогенные переменные, а вектор ε — случайные последовательно некоррелированные ошибки с нулевыми средними. Используя исключающие ограничения (т.е. обращая в нуль некоторые коэффициенты), определите три альтернативные структуры, для которых простейшими состоятельными процедурами оценивания являются соответственно обыкновенный метод наименьших квадратов, косвенный метод наименьших квадратов и двухшаговый метод наименьших квадратов.

7. Имеется следующая макроэкономическая модель:

$$c = \alpha_{10} + \beta_{11}y + \varepsilon_1,$$

$$i = \alpha_{20} + \beta_{21}y + \beta_{22}y_{-1} + \varepsilon_2,$$

$$y = c + i + g,$$

где c, i и y — объем потребления, инвестиции и доход, соответственно, а y_{-1} — доход предыдущего периода, g — государственные расходы.

   а) Определите типы структурных уравнений;

   б) классифицируйте типы переменных;

   в) представьте структурные уравнения в матричной форме;

   г) запишите модель в приведенной форме;

   д) проверьте идентифицируемость и метод оценки параметров каждого уравнения в структурной форме модели;

8. Пусть дана простая Кейнсианская модель:

   $c = \beta y + \varepsilon,$

   $y = c + i,$

   где c, i и y — объем потребления, инвестиции и доход, соответственно. Пусть каждый вектор имеет размерность $N \times 1$, $\mathbf{E}(\varepsilon) = 0$ и $\mathbf{E}(\varepsilon\varepsilon') = \sigma^2 I_N$.

   а) Запишите модель в приведенной форме;

   б) найдите оценку для параметра дохода для приведенной формы;

   в) получите косвенную МНК-оценку для β из результатов (б);

   г) найдите оценку для параметра потребления для приведенной формы;

   д) получите косвенную МНК-оценку для β из результатов (г);

   е) покажите, что результаты (в) и (д) совпадают;

   ж) определите направление смещения МНК-оценки для β.

9. Известны МНК-оценки параметров регрессии (угловые коэффициенты) агрегированного объема продаж продовольственных товаров и цены на них от индекса погодных условий:

   а) 0.3 и -0.6; б) 0.3 и 0.6.

   Определить коэффициенты эластичности спроса и предложения от цены.

10. Пусть система одновременных уравнений имеет вид:

    $x_1 = \alpha_{10} + \beta_{12} x_2 + \alpha_{11} z_1 + \varepsilon_1,$

    $x_2 = \alpha_{20} + \beta_{21} x_1 + \alpha_{22} z_2 + \varepsilon_2.$

    Получены следующие оценки приведенной формы этой системы:

    $x_1 = 1 + 2z_1 + 3z_2,$

    $x_2 = -2 + 1z_1 + 4z_2.$

    Найдите оценки параметров исходной системы.

11. Рассматривается следующая модель краткосрочного равновесия типа IS-LM:

    $y_t = c_t + i_t + g_t + nx_t,$

    $c_t = \alpha_{11} + \beta_{11} y_t + \varepsilon_{1t},$

    $i_t = \alpha_{21} + \alpha_{21} r_t + \varepsilon_{2t},$

    $nx_t = \alpha_{31} + \beta_{31} y_t + \beta_{32} r_t + \varepsilon_{3t},$

    $m_t = \alpha_{40} + \beta_{41} y_t + \alpha_{41} r_t + \varepsilon_{4t},$

где эндогенными переменными являются валовой доход (выпуск) y, объем личных потребительских расходов c, объем инвестиций i, чистый экспорт nx и ставка процента r. Экзогенные переменные: g — совокупные государственные расходы и m — предложение денег. Опишите процедуру оценивания модели с помощью двухшагового метода наименьших квадратов.

12. Дано одно уравнение $x_{1t} = \beta_{12}x_{2t} + \beta_{13}x_{3t} + \alpha_{11}z_{1t} + \varepsilon_{1t}$ модели, состоящей из трех уравнений. В нее входят еще три экзогенные переменные z_1, z_2 и z_3. Наблюдения заданы в виде следующих матриц:

$$Z'Z = \begin{bmatrix} 20 & 15 & -5 \\ 15 & 60 & -45 \\ -5 & -45 & -70 \end{bmatrix}, \quad Z'X = \begin{bmatrix} 2 & 2 & 4 & 5 \\ 0 & 4 & 12 & -5 \\ 0 & -2 & -12 & 10 \end{bmatrix},$$

$$X'X = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 5 \end{bmatrix}.$$

Получите оценки двухшаговым методом наименьших квадратов для параметров этого уравнения и оцените их стандартные ошибки.

Рекомендуемая литература

1. **Айвазян С.А.** Основы эконометрики. Т.2. — М.: «Юнити», 2001. (Гл. 4).

2. **Бриллинджер Д.** Временные ряды. Обработка данных и теория. — М.: «Мир», 1980. (Гл. 10).

3. **Джонстон Дж.** Эконометрические методы. — М.: «Статистика», 1980. (Гл. 12).

4. **Доугерти К.** Введение в эконометрику. — М.: «Инфра-М», 1997. (Гл. 11).

5. **Кейн Э.** Экономическая статистика и эконометрия. Вып. 2. — М.: «Статистика», 1977. (Гл. 13).

6. **Магнус Я.Р., Катышев П.К., Пересецкий А.А.** Эконометрика — начальный курс. — М.: «Дело», 2000. (Гл. 10).

7. **(*) Маленво Э.** Статистические методы эконометрии. Вып. 2. — М., 1975. (Гл. 17–20).

8. **Тинтер Г.** Введение в эконометрию. — М.: «Статистика», 1965. (Гл. 6).

9. **Baltagi, Badi H.** Econometrics, 2nd edition, Springer, 1999. (Ch. 11).

10. **Davidson, Russel, Mackinnon, James.** Estimation and Inference in Econometrics, No. 9, Oxford University Press, 1993. (Ch. 7, 18).

11. **Greene W.H.** Econometric Analysis, Prentice-Hall, 2000. (Ch. 15, 16).

12. **Judge G.G., Hill R.C., Griffiths W.E., Luthepohl H., Lee T.** Introduction to the Theory and Practice of Econometric. John Wiley & Sons, Inc., 1993. (Ch. 14, 15).

13. **Maddala G.S.** Introduction to Econometrics, 2nd ed., Prentice Hall, 1992. (Ch. 9).

14. **Ruud Paul A.** An Introduction to Classical Econometric Theory, Oxford University Press, 2000. (Ch. 26).

15. **William E., Griffiths R., Carter H., George G.** Judge Learning and Practicing econometrics, N 9 John Wiley & Sons, Inc., 1993. (Ch. 17).

Часть III

Эконометрия — I: Анализ временных рядов

Это пустая страница

Глава 11

Основные понятия в анализе временных рядов

11.1. Введение

В каждой сфере экономики встречаются явления, которые интересно и важно изучать в их развитии, т.к. они изменяются во времени. С течением времени изменяются цены, экономические условия, режим протекания того или иного производственного процесса. Совокупность измерений подобного рода показателей в течение некоторого периода времени и представляет временной ряд.

Цели анализа временных рядов могут быть различными. Можно, например, стремиться предсказать будущее на основании знаний прошлого, пытаться выяснить механизм, лежащий в основе процесса, и управлять им. Необходимо уметь освобождать временной ряд от компонент, которые затемняют его динамику. Часто требуется сжато представлять характерные особенности ряда.

Временным рядом называют последовательность наблюдений, обычно упорядоченную во времени, хотя возможно упорядочение и по какому-либо другому параметру. Основной чертой, выделяющей анализ временных рядов среди других видов статистического анализа, является существенность порядка, в котором производятся наблюдения.

Различают два вида временных рядов. Измерение некоторых величин (температуры, напряжения и т.д.) производится непрерывно, по крайней мере теоретически. При этом наблюдения можно фиксировать в виде графика. Но даже в том случае,

когда изучаемые величины регистрируются непрерывно, практически при их обработке используются только те значения, которые соответствуют дискретному множеству моментов времени. Следовательно, если время измеряется непрерывно, временной ряд называется **непрерывным**, если же время фиксируется дискретно, т.е. через фиксированный интервал времени, то временной ряд **дискретен**. В дальнейшем мы будем иметь дело только с дискретными временными рядами. Дискретные временные ряды получаются двумя способами:

- выборкой из непрерывных временных рядов через регулярные промежутки времени (например, численность населения, величина собственного капитала фирмы, объем денежной массы, курс акции), — такие временные ряды называются **моментными**;
- накоплением переменной в течение некоторого периода времени (например, объем производства какого-либо вида продукции, количество осадков, объем импорта), — в этом случае временные ряды называются **интервальными**.

В эконометрии принято моделировать временной ряд как **случайный процесс**, называемый также **стохастическим процессом**, под которым понимается статистическое явление, развивающееся во времени согласно законам теории вероятностей. Случайный процесс — это случайная последовательность. Обычно предполагают, что эта последовательность идет от минус до плюс бесконечности: $\{X_t\}_{t=-\infty,\ldots,+\infty}$. Временной ряд — это лишь одна частная реализация такого теоретического стохастического процесса: $x = \{x_t\}_{t=1,\ldots,T} = (x_1, \ldots, x_T)'$, где T — длина временного ряда. Временной ряд $x = (x_1, \ldots, x_T)'$ также часто неформально называют выборкой[1]. Обычно стоит задача по данному ряду сделать какие-то заключения о свойствах лежащего в его основе случайного процесса, оценить параметры, сделать прогнозы и т.п. В литературе по временным рядам существует некоторая неоднозначность, и иногда временным рядом называют сам случайный процесс $\{X_t\}_{t=-\infty,\ldots,+\infty}$, либо его отрезок $\{x_t\}_{t=1,\ldots,T}$, а иногда статистическую модель, которая порождает данный случайный процесс. В дальнейшем мы не будем в явном виде посредством особых обозначений различать случайный процесс и его реализацию. Из контекста каждый раз будет ясно, о чем идет речь.

Возможные значения временного ряда в данный момент времени t описываются с помощью случайной величины x_t и связанного с ней распределения вероятностей $p(x_t)$. Тогда наблюдаемое значение x_t временного ряда в момент t рассматривается как одно из множества значений, которые могла бы принять случайная величина x_t в этот момент времени. Следует отметить, однако, что, как правило, наблюдения временного ряда взаимосвязаны, и для корректного его описания следует рассматривать совместную вероятность $p(x_1, \ldots, x_T)$.

[1] Хотя, по формальному определению, выборка должна состоять из независимых, одинаково распределенных случайных величин.

Для удобства можно провести классификацию случайных процессов и соответствующих им временных рядов на **детерминированные** и **случайные** процессы (временные ряды). Детерминированным называют процесс, который принимает заданное значение с вероятностью единица. Например, его значения могут точно определяться какой-либо математической функцией от момента времени t, как в следующем примере: $x_t = R\cos(2\omega t - \theta)$. Когда же мы будем говорить о случайном процессе и случайном временном ряде, то, как правило, будем подразумевать, что он не является детерминированным.

Стохастические процессы подразделяются на **стационарные** и **нестационарные**. Стохастический процесс является стационарным, если он находится в определенном смысле в статистическом равновесии, т.е. его свойства с вероятностной точки зрения не зависят от времени. Процесс **нестационарен**, если эти условия нарушаются.

Важное теоретическое значение имеют **гауссовские процессы** . Это такие процессы, в которых любой набор наблюдений имеет совместное нормальное распределение. Как правило, термин «временной ряд» сам по себе подразумевает, что этот ряд является **одномерным** (**скалярным**). Часто бывает важно рассмотреть совместную динамику набора временных рядов $x_t = (x_{1t}, \ldots, x_{kt})$, $t = 1, \ldots, T$. Такой набор называют **многомерным временным рядом**, или **векторным временным рядом**. Соответственно, говорят также о многомерных, или векторных, случайных процессах.

При анализе экономических временных рядов традиционно различают разные виды эволюции (динамики). Эти виды динамики могут, вообще говоря, комбинироваться. Тем самым задается **разложение временного ряда** на составляющие, которые с экономической точки зрения несут разную содержательную нагрузку. Перечислим наиболее важные:

- **тенденция** — соответствует медленному изменению, происходящему в некотором направлении, которое сохраняется в течение значительного промежутка времени. Тенденцию называют также **трендом** или долговременным движением.

- **циклические колебания** — это более быстрая, чем тенденция, квазипериодическая динамика, в которой есть фаза возрастания и фаза убывания. Наиболее часто цикл связан с флуктуациями экономической активности.

- **сезонные колебания** — соответствуют изменениям, которые происходят регулярно в течение года, недели или суток. Они связаны с сезонами и ритмами человеческой активности.

- **календарные эффекты** — это отклонения, связанные с определенными предсказуемыми календарными событиями — такими, как праздничные дни, количество рабочих дней за месяц, високосность года и т.п.
- **случайные флуктуации** — беспорядочные движения относительно большой частоты. Они порождаются влиянием разнородных событий на изучаемую величину (несистематический или случайный эффект). Часто такую составляющую называют **шумом** (этот термин пришел из технических приложений).
- **выбросы** — это аномальные движения временного ряда, связанные с редко происходящими событиями, которые резко, но лишь очень кратковременно отклоняют ряд от общего закона, по которому он движется.
- **структурные сдвиги** — это аномальные движения временного ряда, связанные с редко происходящими событиями, имеющие скачкообразный характер и меняющие тенденцию.

Некоторые экономические ряды можно считать представляющими те или иные виды таких движений почти в чистом виде. Но бóльшая часть их имеет очень сложный вид. В них могут проявляться, например, как общая тенденция возрастания, так и сезонные изменения, на которые могут накладываться случайные флуктуации. Часто для анализа временных рядов оказывается полезным изолированное рассмотрение отдельных компонент.

Для того чтобы можно было разложить конкретный ряд на эти составляющие, требуется сделать какие-то допущения о том, какими свойствами они должны обладать. Желательно построить сначала формальную статистическую модель, которая бы включала в себя в каком-то виде эти составляющие, затем оценить ее, а после этого на основании полученных оценок вычленить составляющие. Однако построение формальной модели является сложной задачей. В частности, из содержательного описания не всегда ясно, как моделировать те или иные компоненты. Например, тренд может быть детерминированным или стохастическим. Аналогично, сезонные колебания можно комбинировать с помощью детерминированных переменных или с помощью стохастического процесса определенного вида. Компоненты временного ряда могут входить в него аддитивно или мультипликативно. Более того, далеко не все временные ряды имеют достаточно простую структуру, чтобы можно было разложить их на указанные составляющие.

Существует два основных подхода к разложению временных рядов на компоненты. Первый подход основан на использовании множественных регрессий с факторами, являющимися функциями времени, второй основан на применении линейных фильтров.

11.2. Стационарность, автоковариации и автокорреляции

Статистический процесс называется **строго стационарным**, если взаимное распределение вероятностей m наблюдений инвариантно по отношению к общему сдвигу временного аргумента, т.е. совместная плотность распределения случайных величин $x_{t_1}, x_{t_2}, \ldots, x_{t_m}$ такая же, как для величин $x_{t_1+k}, x_{t_2+k}, \ldots, x_{t_m+k}$ при любых целых значениях сдвига k. Когда $m = 1$, из предположения стационарности следует, что безусловное распределение величины x_t, $p(x_t)$, одинаково для всех t и может быть записано как $p(x)$.

Требование стационарности, определенное этими условиями, является достаточно жестким. На практике при изучении случайных процессов ограничиваются моментами первого и второго порядка, и тогда говорят о **слабой стационарности** или **стационарности второго порядка**[2]. В этом случае процесс имеет постоянные для всех t моменты первого и второго порядков: среднее значение $\mu = \mathbf{E}(x_t)$, определяющее уровень, относительно которого он флуктуирует, дисперсию $\sigma^2 = \mathbf{E}(x_t - \mu)^2$ и **автоковариацию** $\gamma_k = \mathbf{E}(x_t - \mu)(x_{t+k} - \mu)$. Ковариация между x_t и x_{t+k} зависит только от величины сдвига k и не зависит от t. **Автокорреляция** k-го порядка стационарного процесса с ненулевой дисперсией

$$\rho_k = \frac{\mathbf{E}(x_t - \mu)(x_{t+k} - \mu)}{\sqrt{\mathbf{E}(x_t - \mu)^2 \mathbf{E}(x_{t+k} - \mu)^2}}$$

сводится к простой формуле

$$\rho_k = \frac{\gamma_k}{\gamma_0}.$$

Следует иметь в виду, что два процесса, имеющие одинаковые моменты первого и второго порядка, могут иметь разный характер распределения.

Автоковариационной функцией стационарного процесса называют последовательность автоковариаций $\{\gamma_k\}_{k=-\infty,\ldots,+\infty}$. Так как автоковариационная функция симметрична относительно нуля: $\gamma_k = \gamma_{-k}$, то достаточно рассматривать $k = 0,\ 1,\ 2,\ 3,\ \ldots$

Автокорреляционной функцией (АКФ) называют последовательность автокорреляций $\{\rho_k\}_{k=-\infty,\ldots,+\infty}$. Автокорреляционная функция также симметрична, причем $\rho_0 = 1$, поэтому рассматривают $k = 0,\ 1,\ 2,\ 3,\ \ldots$

[2] В русскоязычной литературе строгую стационарность также называют стационарностью в узком смысле, а слабую стационарность — стационарностью в широком смысле.

Автоковариационная матрица Γ_T для стационарного ряда $x_1, \ldots, x_T$ имеет вид:

$$\Gamma_T = \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{T-1} & \gamma_{T-2} & \cdots & \gamma_0 \end{pmatrix} = \gamma_0 \begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{T-1} \\ \rho_1 & 1 & \cdots & \rho_{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{T-1} & \rho_{T-2} & \cdots & 1 \end{pmatrix},$$

$$\Gamma_T = \gamma_0 P_T.$$

Особенность автоковариационной матрицы Γ_T и соответствующей автокорреляционной матрицы P_T в случае стационарности состоит в том, что они имеют одни и те же элементы на любой диагонали. Матрицы такого вида принято называть тёплицевыми матрицами.

Как известно, любая ковариационная матрица является симметричной и положительно полуопределенной. Кроме того, если компоненты рассматриваемого случайного вектора x линейно независимы в том смысле, что не существует ненулевой вектор коэффициентов λ, такой что $\lambda' x$ — детерминированная величина, то ковариационная матрица является положительно определенной. Напомним, что, по определению (см. Приложение А.1.1), симметричная $T \times T$ матрица A называется положительно полуопределенной, если для каждого вектора λ выполняется неравенство $\lambda' A \lambda \geqslant 0$; матрица A называется положительно определенной, если для каждого ненулевого вектора λ выполняется неравенство $\lambda' A \lambda > 0$. Автоковариационная и автокорреляционная матрица являются ковариационными матрицами, поэтому они обладают указанными свойствами. С другой стороны, если матрица обладает указанными свойствами, то она может быть автоковариационной матрицей некоторого временного ряда.

Из этих рассуждений следует, что условие слабой стационарности процесса, компоненты которого линейно независимы в указанном выше смысле, налагает ряд ограничений на вид автокорреляционной и автоковариационной функций. Они вытекают из того, что главные миноры положительно определенной матрицы, в том числе ее определитель, должны быть положительны.

В частности, положительная определенность главного минора второго порядка дает

$$\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix} = 1 - \rho_1^2 > 0, \quad \text{или} \quad -1 < \rho_1 < 1,$$

А для третьего порядка:

$$2\rho_1^2 - 1 < \rho_2 < 1.$$

Среди стационарных процессов в теории временных рядов особую роль играют процессы типа **белый шум**. Это неавтокоррелированные слабо стационарные процессы $\{\varepsilon_t\}$ с нулевым математическим ожиданием и постоянной дисперсией:

$$\begin{aligned} &\mu = \mathbf{E}(\varepsilon_t) = 0, \\ &\gamma_k = \begin{cases} \sigma^2 & , k = 0 \\ 0 & , k \neq 0 \end{cases} \end{aligned} \tag{11.1}$$

Следовательно, для белого шума $\Gamma_T = \sigma^2 I_T$, где I_T — единичная матрица порядка T.

Название «белый шум» связано с тем, что спектральная плотность такого процесса постоянна, то есть он содержит в одинаковом количестве все частоты, подобно тому, как белый цвет содержит в себе все остальные цвета. Если белый шум имеет нормальное распределение, то его называют **гауссовским белым шумом**.

Аналогичные определения стационарности можно дать и для векторного стохастического процесса $\{x_t\}$. Слабо стационарный векторный процесс будет характеризоваться уже не скалярными автоковариациями γ_k и автокорреляциями ρ_k, а аналогичными по смыслу матрицами. Вне главной диагонали таких матриц стоят, соответственно, **кросс-ковариации** и **кросс-корреляции**.

11.3. Основные описательные статистики для временных рядов

Предположим, у нас имеются некоторые данные в виде временного ряда $\{x_t\}_{t=1,\ldots,T}$. **Среднее** и **дисперсия** временного ряда рассчитываются по обычным формулам:

$$\overline{x} = \sum_{t=1}^{T} x_t \;\text{ и }\; s^2 = \frac{1}{T}\sum_{t=1}^{T}(x_t - \overline{x})^2.$$

Выборочная автоковариация k-го порядка вычисляется как

$$c_k = \frac{1}{T}\sum_{t=1}^{T-k}(x_t - \overline{x})(x_{t+k} - \overline{x}).$$

Если временной ряд слабо стационарен, то эти описательные статистики являются оценками соответствующих теоретических величин и при некоторых предположениях обладают свойством состоятельности.

Заметим, что в теории временных рядов при расчете дисперсии и ковариаций принято сумму квадратов и, соответственно, произведения делить на T. Вместо этого при расчете дисперсии, например, можно было бы делить на $T-1$, что дало бы несмещенную оценку, а при расчете ковариации k-го порядка — на $T-k$ по числу слагаемых. Оправданием данной формулы может служить простота расчетов и то, что в таком виде это выражение гарантирует положительную полуопределенность матрицы выборочных автоковариаций C_T:

$$C_T = \begin{pmatrix} c_0 & c_1 & \cdots & c_{T-1} \\ c_1 & c_0 & \cdots & c_{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ c_{T-1} & c_{T-2} & \cdots & c_0 \end{pmatrix}.$$

Это отражает важное свойство соответствующей матрицы Γ_T истинных автоковариаций.

Любую положительно определенную матрицу B можно представить в виде $B = A'A$, где A — некоторая матрица (см., например, Приложения А.1.2 и А.1.2). В нашем случае $A = \frac{1}{\sqrt{T}}\widehat{X}$, поскольку матрица C_T выражается в виде произведения:

$$C_T = \frac{1}{T}\widehat{X}'\widehat{X},$$

где $\widehat{X}$ — T-диагональная матрица, составленная из центрированных значений ряда $\widehat{x}_t = x_t - \overline{x}$:

$$\widehat{X} = \begin{pmatrix} \widehat{x}_1 & 0 & \cdots & 0 \\ \widehat{x}_2 & \widehat{x}_1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \widehat{x}_T & \widehat{x}_{T-1} & \ddots & \widehat{x}_1 \\ 0 & \widehat{x}_T & \cdots & \widehat{x}_2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \widehat{x}_T \end{pmatrix}.$$

Статистической оценкой автокорреляции k-го порядка для стационарных процессов является выборочный коэффициент автокорреляции: $r_k = c_k/c_0$. При анализе изменения величин c_k и r_k в зависимости от значения k обычно пользуются выборочными автоковариационной и автокорреляционной функциями, определяемыми как последовательности $\{c_k\}$ и $\{r_k\}$, соответственно. Выборочная автокорреляционная функция играет особую роль в анализе стационарных временных рядов, поскольку может быть использована в качестве инструмента для распознавания типа процесса. При этом обычно анализируют график автокорреляционной функции, называемый **коррелограммой**.

Заметим, что по ряду длиной T можно вычислить автокорреляции вплоть до r_{T-1}. Однако «дальние» автокорреляции вычисляются неточно. С ростом порядка k количество наблюдений, по которым вычисляется коэффициент автокорреляции r_k, уменьшается. Для расчета r_{T-1} используется два наблюдения. Таким образом, с ростом k выборочные автокорреляции r_k становятся все менее надежными оценками теоретических автокорреляций ρ_k. Таким образом, при анализе ряда следует принимать во внимание только самые «ближние» автокорреляции, например, первые $[T/5]$ автокорреляций.

По аналогии с автоковариациями и автокорреляциями для анализа совместной динамики *нескольких рядов* можно использовать выборочные кросс-ковариации и кросс-корреляции.

Выборочная **кросс-ковариация** двух временных рядов, $\{x_t\}$ и $\{y_t\}$, рассчитывается по формуле:

$$\delta_k = \frac{1}{T}\sum_{t=1}^{T-k}(x_{t+k} - \overline{x})(y_t - \overline{y}).$$

Она характеризует взаимосвязи двух рядов во времени с различной величиной сдвига k. Следует помнить, что, в отличие от автоковариации, кросс-ковариация не является симметричной по k, поэтому ее следует рассматривать и при положительных, и при отрицательных k.

Выборочная **кросс-корреляция** определяется как:

$$\frac{\sum_{t=1}^{T-k}(x_{t+k} - \overline{x})(y_t - \overline{y})}{\sqrt{\sum_{t=1}^{T}(x_t - \overline{x})^2 \sum_{t=1}^{T}(y_t - \overline{y})^2}}.$$

11.4. Использование линейной регрессии с детерминированными факторами для моделирования временного ряда

Сравнительно простой моделью временного ряда может служить модель вида:

$$x_t = \mu_t + \varepsilon_t, \quad t = 1, \ldots, T, \tag{11.2}$$

где μ_t — полностью детерминированная последовательность или систематическая составляющая, ε_t — последовательность случайных влеичин, являющаяся белым шумом. Если μ_t зависит от вектора неизвестных параметров θ: $\mu_t = \mu_t(\theta)$, то модель (11.2) является моделью регрессии, и ее параметры можно оценить с помощью МНК.

Детерминированная компонента μ_t, как правило, сама моделируется как состоящая из нескольких компонент. Например, можно рассмотреть аддитивную модель, в которой временной ряд содержит три компоненты: тренд τ_t, сезонные движения v_t и случайные флуктуации ε_t:

$$x_t = \tau_t + v_t + \varepsilon_t.$$

Зачастую изучаемый экономический ряд ведет себя так, что аддитивной схеме следует предпочесть мультипликативную схему:

$$x_t = \tau_t v_t \exp(\varepsilon_t).$$

Однако, если это выражение прологарифмировать, то получится аддитивный вариант:

$$\ln(x_t) = \ln(\tau_t) + \ln(v_t) + \varepsilon_t = \tau_t^* + v_t^* + \varepsilon_t,$$

что позволяет оставаться в рамках линейной регрессии и значительно упрощает моделирование.

11.4.1. Тренды

Существует три основных типа трендов.

Первым и самим очевидным типом тренда представляется **тренд среднего**, когда временной ряд выглядит как колебания около медленно возрастающей или убывающей величины.

Второй тип трендов — это **тренд дисперсии**. В этом случае во времени меняется амплитуда колебаний переменной. Иными словами, процесс гетероскедастичен.

Часто экономические процессы с возрастающим средним имеют и возрастающую дисперсию.

Третий и более тонкий тип тренда, визуально не всегда наблюдаемый, — изменение величины корреляции между текущим и предшествующим значениями ряда, т.е. **тренд автоковариации и автокорреляции**.

Проводя разложение ряда на компоненты, мы, как правило, подразумеваем под трендом изменение среднего уровня переменной, то есть тренд среднего.

В рамках анализа тренда среднего выделяют следующие основные способы аппроксимации временных рядов и соответствующие основные виды трендов среднего.

- **Полиномиальный тренд:**

$$\tau_t = a_0 + a_1 t + \ldots + a_p t^p. \tag{11.3}$$

  Для $p = 1$ имеем **линейный тренд**.

- **Экспоненциальный тренд:**

$$\tau_t = e^{a_0 + a_1 t + \ldots + a_p t^p}. \tag{11.4}$$

- **Гармонический тренд:**

$$\tau_t = R\cos(\omega t + \varphi), \tag{11.5}$$

  где R — амплитуда колебаний, ω — угловая частота, φ — фаза.

- **Тренд, выражаемый логистической функцией:**

$$\tau_t = \frac{k}{1 + be^{-at}}. \tag{11.6}$$

Оценивание параметров полиномиального и экспоненциального трендов (после введения обозначения $z_i = t^i$, $i = 1, \ldots, p$, — в первом случае и логарифмирования функции во втором случае) производится с помощью обычного МНК.

Гармонический тренд оправдан, когда в составе временного ряда отчетливо прослеживаются периодические колебания. При этом если частота ω известна (или ее можно оценить), то функцию (11.5) несложно представить в виде линейной комбинации синуса и косинуса:

$$\tau_t = \alpha\cos(\omega t) + \beta\sin(\omega t)$$

и, рассчитав векторы $\cos(\omega t)$ и $\sin(\omega t)$, также воспользоваться МНК для оценивания параметров α и β.

Логистическая кривая нуждается в особом рассмотрении.

11.4.2. Оценка логистической функции

Проанализируем логистическую функцию:

$$\tau_t = \frac{k}{1 + be^{-at}}, \tag{11.7}$$

где a, b, k — параметры, подлежащие оцениванию. Функция ограничена и имеет горизонтальную асимптоту (рис. 11.1):

$$\lim_{t \to \infty} \tau_t = k.$$

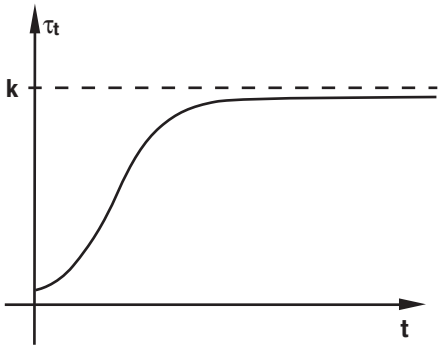

Рис. 11.1. Логистическая кривая

В этом преимущество логистической функции перед полиномиальной или экспоненциальной функциями, которые по мере роста t стремятся в бесконечность и, следовательно, не всегда годятся для прогнозирования.

Логистическая кривая наиболее часто используется при изучении социальных и, в частности, демографических процессов.

Особенностью логистической кривой является нелинейность по оцениваемым параметрам a, b, k, поэтому система уравнений, получаемая с помощью МНК, нелинейна относительно неизвестных параметров и для ее решения могут применяться только итеративные численные методы.

Гарольд Готтелинг (H. Hotteling) предложил интересный метод для оценки этих параметров, основанный на использовании дифференциального уравнения логистической функции. Дифференцирование функции τ_t по времени t дает первую производную:

$$\frac{d\tau_t}{dt} = \frac{kabe^{-at}}{(1 + be^{-at})^2}.$$

Поскольку

$$\frac{\tau_t^2}{k} = \frac{k}{(1 + be^{-at})^2} \quad \text{и} \quad be^{-at} = \frac{k}{\tau_t} - 1,$$

то, подставляя эти выражения в формулу первой производной, получаем дифференциальное уравнение, выражающее зависимость темпа прироста исследуемой

переменной от абсолютного уровня показателя в момент времени t:

$$\frac{d\tau_t/dt}{\tau_t} = a - \frac{a}{k}\tau_t. \tag{11.8}$$

Исходя из этого соотношения, можно предположить, что в реальности абсолютный прирост показателя Δx_t связан с фактическим его уровнем x_t следующей статистической зависимостью:

$$\Delta x_t = ax_t + \left(-\frac{a}{k}\right) x_t^2 + \eta_t,$$

где η_t — белый шум.

К этому уравнению теперь можно применить непосредственно метод наименьших квадратов, получить оценки параметров a и $-\frac{a}{k}$ и, следовательно, найти k.

Оценка параметра b методом моментов впервые предложена Родсом. Так как $be^{-at} = \frac{k}{\tau_t} - 1$, то $\ln b = at + \ln\left(\frac{k}{\tau_t} - 1\right)$ и с помощью метода моментов получаем:

$$\ln b = \frac{1}{T}\left(a \cdot \frac{T(T+1)}{2} + \sum_{t=1}^{T} \ln\left(\frac{k}{\tau_t} - 1\right)\right),$$

или фактически после замены τ_t на x_t имеем:

$$\ln b = \frac{a(T+1)}{2} + \frac{\sum\limits_{t=1}^{T} \ln\left(\frac{k}{x_t} - 1\right)}{T}. \tag{11.9}$$

Описанный выше метод Готтелинга имеет ограниченную сферу применения, его использование оправдано лишь в том случае, если наблюдения в исходном временном ряду представлены через равные промежутки времени (например, ежегодные или еженедельные данные).

11.4.3. Сезонные колебания

Для моделирования сезонной составляющей τ_t можно использовать формулу:

$$v_t = \lambda_1\delta_{1t} + \ldots + \lambda_h\delta_{ht},$$

где δ_{jt} — сезонные **фиктивные переменные**, соответствующие h сезонам: $\delta_{jt} = 1$, когда наблюдение относится к сезону j, и $\delta_{jt} = 0$ в противном случае.

Использование в линейной регрессии полного набора таких переменных связано с одной особенностью. В сумме они дают единицу:

$$\delta_{1t} + \ldots + \delta_{ht} = 1.$$

Поэтому, коль скоро в регрессии имеется константа, то будет иметь место линейная зависимость, и $\lambda_1, \ldots, \lambda_h$ нельзя будет оценить однозначно. Таким образом, требуется наложить на коэффициенты $\lambda_1, \ldots, \lambda_h$ какое-либо нормирующее ограничение. В частности, можно положить один из коэффициентов равным нулю, что эквивалентно неиспользованию соответствующей переменной при построении регрессии. Однако более удачная нормировка состоит в том, чтобы положить $\lambda_1 + \ldots + \lambda_h = 0$. При этом сезонная компонента центрируется, то есть в среднем влияние эффекта сезонности на уровень ряда оказывается равным нулю.

Подставим это ограничение в сезонную компоненту, исключив коэффициент λ_1:

$$v_t = -(\lambda_2 + \ldots + \lambda_h)\delta_{1t} + \lambda_2\delta_{2t} + \ldots + \lambda_h\delta_{ht} = \\ = \lambda_2(\delta_{2t} - \delta_{1t}) + \ldots + \lambda_h(\delta_{ht} - \delta_{1t}).$$

Новые переменные $\delta_{2t} - \delta_{1t}, \ldots, \delta_{ht} - \delta_{1t}$ будут уже линейно независимыми, и их можно использовать в линейной регрессии в качестве факторов, а также получить и оценку структуры сезонности $\lambda_1, \ldots, \lambda_h$. Трактовать ее следует так: в j-м сезоне сезонность приводит к отклонению от основной динамики ряда на величину λ_j.

Если для описания тренда взять полиномиальную функцию, то, используя аддитивную схему, можно представить временной ряд в виде следующей линейной регрессии:

$$x_t = a_0 + a_1 t + \ldots + a_p t^p + \lambda_1\delta_{1t} + \ldots + \lambda_h\delta_{ht} + \varepsilon_t,$$

где $\lambda_1 + \ldots + \lambda_h = 0$.

В этой регрессии a_i и λ_j являются неизвестными коэффициентами. Применение МНК дает оценки $p + h + 1$ неизвестных коэффициентов и приводит к выделению составляющих τ_t, v_t и ε_t.

11.4.4. Аномальные наблюдения

При моделировании временного ряда часто отбрасываются **аномальные** наблюдения, резко отклоняющиеся от направления эволюции ряда. Такого рода **выбросы**, вместо исключения, можно моделировать с помощью фиктивных переменных, соответствующих фиксированным моментам времени. Предположим,

что в момент t^* в экономике произошло какое-нибудь важное событие (например, отставка правительства). Тогда можно построить фиктивную переменную $\delta_t^{t^*}$, которая равна нулю всегда, кроме момента $t = t^*$, когда она равна единице: $\delta_t^{t^*} = (0, \ldots, 0, 1, 0, \ldots, 0)$.

Такая фиктивная переменная пригодна только для моделирования кратковременного отклонения временного ряда. Если же в экономике произошел **структурный сдвиг**, вызвавший скачок в динамике ряда, то следует использовать фиктивную переменную другого вида: $(0, \ldots, 0, 1, \ldots, 1)$. Эта переменная равна нулю до некоторого фиксированного момента t^*, а после этого момента становится равной единице.

Заметим, что последние два вида переменных нельзя использовать для прогнозирования, поскольку они относятся к единичным непрогнозируемым событиям.

11.5. Прогнозы по регрессии с детерминированными факторами. Экстраполирование тренда

Предположим, что данные описываются линейной регрессией с детерминированными регрессорами, являющимися функциями t, и получены оценки параметров регрессии на основе данных $x = (x_1, \ldots, x_T)'$ и соответствующей матрицы факторов Z. Это позволяет построить прогноз на будущее, например на период $T + k$. Вообще говоря, прогноз в такой регрессии строится так же, как в любой классической линейной регрессии. Отличие состоит только в том, что значения факторов z_{T+k}, необходимые для осуществления прогноза, в данном случае всегда известны.

Рассмотрим прогнозирование на примере, когда временной ряд моделируется по упрощенной схеме — тренд плюс шум: $x_t = \tau_t + \varepsilon_t$, где $\tau_t = z_t\alpha$, z_t — вектор-строка значения факторов регрессии в момент t, α — вектор-столбец коэффициентов регрессии.

Такое моделирование имеет смысл, если циклические и сезонные компоненты отсутствуют или мало значимы. Тогда выявленный тренд τ_t может служить основой для прогнозирования. Прогноз величины x_{T+k} строится по формуле условного математического ожидания $x_T(k) = z_{T+k}a$, где a — оценки параметров, полученные с помощью МНК, т.е. $a = (Z'Z)^{-1}Z'x$. Известно, что такой прогноз обладает свойством оптимальности.

Предположим, что для описания тренда выбран многочлен:

$$\tau_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \ldots + \alpha_p t^p, \quad t = 1, \ldots, T.$$

В такой модели матрица факторов имеет следующий вид:

$$Z = \begin{pmatrix} 1^0 & 1^1 & \cdots & 1^p \\ 2^0 & 2^1 & \cdots & 2^p \\ \vdots & \vdots & \ddots & \vdots \\ T^0 & T^1 & \cdots & T^p \end{pmatrix}.$$

Вектор значений факторов на момент $T+k$ известен определенно:

$$z_{T+k} = \left(1, (T+k), (T+k)^2, \ldots, (T+k)^p\right).$$

Точечный прогноз исследуемого показателя в момент времени T на k шагов вперед равен:

$$x_T(k) = z_{T+k}a = a_0 + a_1(T+k) + a_2(T+k)^2 + \ldots + a_p(T+k)^p.$$

Возвратимся к общей теории прогноза. Ошибка прогноза равна:

$$d = x_{T+k} - x_T(k) = x_{T+k} - z_{T+k}a.$$

Ее можно представить как сумму двух отдельных ошибок:

$$d = (x_{T+k} - z_{T+k}\alpha) + (z_{T+k}\alpha - z_{T+k}a) = \varepsilon_{T+k} + z_{T+k}(\alpha - a).$$

Первое слагаемое здесь — это будущая ошибка единичного наблюдения, а второе — ошибка, обусловленная выборкой и связанная с тем, что вместо неизвестных истинных параметров α используются оценки a.

Прогноз будет несмещенным, поскольку

$$\mathbf{E}(d) = \mathbf{E}(\varepsilon_{T+k}) + z_{T+k}\mathbf{E}(\alpha - a) = 0.$$

Величина $x_T(k)$ представляет собой **точечный прогноз**. Поскольку точечный прогноз всегда связан с ошибкой, то важно иметь оценку точности этого прогноза. Кроме того, вокруг точечного прогноза желательно построить доверительный интервал и, тем самым, получить **интервальный прогноз**.

Точность прогноза измеряется, как правило, средним квадратом ошибки прогноза, т.е. величиной $\mathbf{E}(d^2)$, или корнем из нее — среднеквадратической ошибкой прогноза. Поскольку $\mathbf{E}(d) = 0$, то средний квадрат ошибки прогноза равен **дисперсии ошибки прогноза**. Полезным показателем точности является корень из этой

дисперсии — стандартная ошибка прогноза. В предположении отсутствия автокорреляции ошибок ε_t дисперсия ошибки прогноза, подобно самой ошибке прогноза, является суммой двух дисперсий: дисперсии ε_{T+k} и дисперсии $z_{T+k}(\alpha - a)$, а именно:

$$\sigma_d^2 = \mathbf{var}(d) = \mathbf{var}(\varepsilon_{T+k}) + \mathbf{var}\left(z_{T+k}(\alpha - a)\right).$$

Найдем эту дисперсию, исходя из того, что ошибки гомоскедастичны:

$$\sigma_d^2 = \sigma^2 + z_{T+k}\mathbf{var}(\alpha - a)z'_{T+k} = \sigma^2 + z_{T+k}\mathbf{var}(a)z'_{T+k}.$$

Как известно, при отсутствии автокорреляции и гетероскедастичности, оценки МНК имеют дисперсию

$$\mathbf{var}(a) = \sigma^2 \left(Z'Z\right)^{-1}.$$

Поэтому

$$\sigma_d^2 = \sigma^2 \left(1 + z_{T+k} \left(Z'Z\right)^{-1} z'_{T+k}\right).$$

Для того чтобы построить доверительный интервал прогноза, следует предположить нормальность ошибок. Более конкретно, предполагаем, что ошибки регрессии, включая ошибку наблюдения, для которого делается прогноз, имеют многомерное нормальное распределение с нулевым математическим ожиданием и ковариационной матрицей $\sigma^2 I$. При таком предположении ошибка прогноза имеет нормальное распределение с нулевым математическим ожиданием и дисперсией σ_d^2:

$$d \sim N(0, \sigma_d^2).$$

Приводя к стандартному нормальному распределению, получим

$$\frac{d}{\sigma_d} \sim N(0, 1).$$

Однако, эта формула еще не дает возможности построить доверительный интервал, поскольку истинная дисперсия прогноза σ_d^2 неизвестна. Вместо нее следует использовать оценку

$$s_d^2 = \hat{s}_e^2 \left(1 + z_{T+k} \left(Z'Z\right)^{-1} z'_{T+k}\right),$$

где $\hat{s}_e^2$ — несмещенная оценка дисперсии ошибок регрессии, или остаточная дисперсия.

Оказывается, что получающаяся величина d/s_d имеет распределение Стьюдента с $(T-p-1)$ степенями свободы (см. Приложение А.3.2), где p — количество факторов в регрессии (без учета константы): $d/s_d \sim t_{T-p-1}$.

Построим на основе этого вокруг прогноза $x_T(k)$ доверительный интервал для x_{T+k}, учитывая, что $d = x_{T+k} - x_T(k)$:

$$\left[x_T(k) - s_d t_{T-p-1,1-q};\quad x_T(k) + s_d t_{T-p-1,1-q}\right],$$

где $t_{T-p-1,1-q}$ — $(1-q)$-квантиль t-распределения Стьюдента с $(T-p-1)$ степенями свободы.

Рассмотрим прогнозирование на примере линейного тренда. В этом случае

$$z_{T+k} = (1, T+k),$$

С учетом того, что

$$Z' = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 2 & \cdots & T \end{bmatrix},$$

произведение $Z'Z$ имеет вид:

$$Z'Z = \begin{pmatrix} T & \sum_{t=1}^{T} t \\ \sum_{t=1}^{T} t & \sum_{t=1}^{T} t^2 \end{pmatrix},$$

$$(Z'Z)^{-1} = \frac{1}{T\sum_{t=1}^{T} t^2 - \left(\sum_{t=1}^{T} t\right)^2} \begin{pmatrix} \sum_{t=1}^{T} t^2 & -\sum_{t=1}^{T} t \\ -\sum_{t=1}^{T} t & T \end{pmatrix},$$

$$z_{T+k}\,(Z'Z)^{-1}\,z'_{T+k} = \frac{\sum_{t=1}^{T} t^2 - 2(T+k)\sum_{t=1}^{T} t + T(T+k)^2}{T\sum_{t=1}^{T} t^2 - \left(\sum_{t=1}^{T} t\right)^2} =$$

$$= \frac{T\left((T+k)^2 - 2(T+k)\bar{t} + \bar{t}^2 + \sum_{t=1}^{T} \frac{t^2}{T} - \bar{t}^2\right)}{T\sum_{t=1}^{T}(t-\bar{t})^2} =$$

$$= \frac{T((T+k)-\bar{t})^2}{T\sum_{t=1}^{T}(t-\bar{t})^2} + \frac{\sum_{t=1}^{T} t^2 - T\bar{t}^2}{T\sum_{t=1}^{T}(t-\bar{t})^2} = \frac{((T+k)-\bar{t})^2}{\sum_{t=1}^{T}(t-\bar{t})^2} + \frac{1}{T}.$$

Тогда:

$$\sigma_d^2 = \sigma^2\left(1 + \frac{1}{T} + \frac{((T+k)-\bar{t})^2}{\sum_{t=1}^{T}(t-\bar{t})^2}\right).$$

Соответственно,

$$s_d = \hat{s}_e \sqrt{1 + \frac{1}{T} + \frac{((T+k) - \bar{t})^2}{\sum_{t=1}^{T}(t - \bar{t})^2}}.$$

Из этой формулы видно, что чем больше горизонт прогноза k, тем больше дисперсия прогноза и шире прогнозный интервал.

11.6. Критерии, используемые в анализе временных рядов

В анализе временных рядов наиболее разработанными критериями являются **критерии случайности**, которые призваны определить, является ли ряд чисто случайным, либо в его поведении проявляются определенные закономерности, которые позволяют делать предсказания. «Чисто случайный ряд» — это в данном случае неформальный термин, подчеркивающий отсутствие закономерностей. Здесь может, например, подразумеваться ряд, состоящий из независимых и одинаково распределенных наблюдений (что соответствует понятию выборки в обычной статистике), либо белый шум, в том смысле, который указан ранее.

Среди экономических временных рядов редко встречаются такие, которые подходят под это описание[3]. Типичный экономический ряд характеризуется сильной положительной корреляцией. Очень часто экономические ряды содержат тенденцию, сезонность и т.д. В связи с этим применение критериев случайности по прямому назначению не имеет особого смысла. Тем не менее, критерии случайности играют очень важную роль в анализе временных рядов, и существуют различные способы их использования:

1) Критерий может быть чувствительным к определенным отклонениям от «случайности». Тогда большое значение соответствующей статистики может указывать на наличие именно такого отклонения. Таким образом, статистика критерия может использоваться просто как описательная статистика. При этом формальная проверка гипотезы не производится.

Так, например, автокорреляционная функция, о которой речь пойдет ниже, очень чувствительна к наличию периодичностей и трендов. Кроме того, по автокорреляционной функции можно определить, насколько быстро затухает временна́я зависимость в рядах[4].

[3] Близки к этому, видимо, только темпы прироста курсов ценных бумаг.

[4] При интерпретации автокорреляционной функции возникают сложности, связанные с тем, что соседние значения автокорреляций коррелированы между собой.

2) Критерий можно применять к *остаткам* от модели, а не к самому исходному ряду. Пусть, например, была оценена модель вида «тренд плюс шум». После вычитания из ряда выявленного тренда получаются остатки, которые можно рассматривать как оценки случайной компоненты. Наличие в остатках каких-либо закономерностей свидетельствует о том, что модель неполна, либо в принципе некорректна. Поэтому критерии случайности можно использовать в качестве диагностических критериев при моделировании.

Следует помнить, однако, что распределение статистики, рассчитанной по остаткам, и распределение статистики, рассчитанной по исходному случайному шуму, вообще говоря, не совпадают. В некоторых случаях при большом количестве наблюдений это различие несущественно, но часто в результате критерий становится несостоятельным и критические значения в исходном виде применять нельзя[5].

Существует большое количество различных критериев случайности. По-видимому, наиболее популярными являются критерии, основанные на автокорреляционной функции.

11.6.1. Критерии, основанные на автокорреляционной функции

Для того чтобы сконструировать критерии, следует рассмотреть, какими статистическими свойствами характеризуется автокорреляционная функция стационарного процесса.

Известно, что выборочные автокорреляции имеют нормальное асимптотическое распределение. При большом количестве наблюдений математическое ожидание r_k приближенно равно ρ_k. Дисперсия автокорреляции приближенно равна

$$\mathbf{var}(r_k) \approx \frac{1}{T} \sum_{i=-\infty}^{+\infty} [\rho_i^2 + \rho_{i-k}\rho_{i+k} - 4\rho_k\rho_i\rho_{i+k} + 2\rho_k^2\rho_i^2]. \qquad (11.10)$$

Для ковариации двух коэффициентов автокорреляции верно приближение

$$\mathbf{cov}(r_k, r_l) \approx \qquad (11.11)$$
$$\approx \frac{1}{T} \sum_{i=-\infty}^{+\infty} [\rho_{i+k}\rho_{i+l} + \rho_{i-k}\rho_{i+l} - 2\rho_k\rho_i\rho_{i+l} - 2\rho_l\rho_i\rho_{i+k} + 2\rho_k\rho_l\rho_i^2]$$

Эти аппроксимации были выведены Бартлеттом.

[5] Так, Q-статистика, о которой идет речь ниже, в случае остатков модели ARMA(p, q) будет распределена не как χ^2_m, а как χ^2_{m-p-q}. Применение распределения χ^2_m приводит к тому, что нулевая гипотеза о «случайности» принимается слишком часто.

В частности, для белого шума (учитывая, что $\rho_k = 0$ при $k \neq 0$) получаем согласно формуле (11.10)

$$\mathbf{var}(r_k) \approx \frac{1}{T}. \tag{11.12}$$

Это только грубое приближение для дисперсии. Для гауссовского белого шума известна точная формула для дисперсии коэффициента автокорреляции:

$$\mathbf{var}(r_k) = \frac{T-k}{T(T+2)}. \tag{11.13}$$

Кроме того, из приближенной формулы (11.11) следует, что автокорреляции r_k и r_l, соответствующие разным порядкам ($k \neq l$), некоррелированы.

Эти формулы позволяют проверять гипотезы относительно автокорреляционных коэффициентов. Так, в предположении, что ряд представляет собой белый шум, можно использовать следующий доверительный интервал для отдельного коэффициента автокорреляции:

$$\left[r_k - \sqrt{\frac{T-k}{T(T+2)}}\hat{\varepsilon}_{1-\theta}, r_k + \sqrt{\frac{T-k}{T(T+2)}}\hat{\varepsilon}_{1-\theta}\right],$$

где $\hat{\varepsilon}_{1-\theta}$ — квантиль нормального распределения. При больших T и малых k оправдано использование более простой формулы

$$\left[r_k - \frac{\hat{\varepsilon}_{1-\theta}}{\sqrt{T}}, r_k + \frac{\hat{\varepsilon}_{1-\theta}}{\sqrt{T}}\right],$$

Вместо того чтобы проверять отсутствие автокорреляции для каждого отдельного коэффициента, имеет смысл использовать критерий случайности, основанный на нескольких ближних автокорреляциях. Рассмотрим m первых автокорреляций: $r_1, \ldots, r_m$. В предположении, что ряд является белым шумом, при большом количестве наблюдений их совместное распределение приближенно равно $N\left(0, \frac{1}{T}I_m\right)$. На основе этого приближения Бокс и Пирс предложили следующую статистику, называемую **Q-статистикой Бокса—Пирса**:

$$Q(r) = T\sum_{k=1}^{m} r_k^2.$$

Она имеет асимптотическое распределение χ^2_m.

При дальнейшем изучении выяснилось, что выборочные значения Q-статистики Бокса—Пирса могут сильно отклонятся от распределения χ^2_m. Для улучшения

аппроксимации Льюнг и Бокс предложили использовать точную формулу дисперсии (11.13) вместо (11.12). Полученная ими статистика, ***Q*-статистика Льюнга—Бокса**:

$$\tilde{Q}(r) = T(T+2)\sum_{k=1}^{m}\frac{r_k^2}{T-k},$$

тоже имеет асимптотическое распределение χ^2_m, однако при малом количестве наблюдений демонстрирует гораздо лучшее соответствие этому асимптотическому распределению, чем статистика Бокса—Пирса.

Было показано, что критерий не теряет своей состоятельности даже при невыполнении гипотезы о нормальности процесса. Требуется лишь, чтобы дисперсия была конечной.

Нулевая гипотеза в Q-критерии заключается в том, что ряд представляет собой белый шум, то есть является чисто случайным процессом. Используется стандартная процедура проверки: если расчетное значение Q-статистики больше заданного квантиля распределения χ^2_m, то нулевая гипотеза отвергается и признается наличие автокорреляции до m-го порядка в исследуемом ряду.

Кроме критериев случайности можно строить и другие критерии на основе автокорреляций. Пусть, например, $\rho_i = 0$ при $i \geqslant k$, т.е. процесс автокоррелирован, но автокорреляция пропадает после порядка k[6]. Тогда по формуле 11.10 получаем

$$\mathbf{var}(r_k) \approx 1 + 2\sum_{i=1}^{k-1}\rho_i^2.$$

Если в этой формуле заменить теоретические автокорреляции выборочными, то получим следующее приближение:

$$\mathbf{var}(r_k) \approx 1 + 2\sum_{i=1}^{k-1} r_i^2.$$

На основе этого приближения (**приближения Бартлетта**) с учетом асимптотической нормальности можно стандартным образом построить доверительный интервал для r_k:

$$\left[r_k - \hat{\varepsilon}_{1-\theta}\sqrt{\mathbf{var}(r_k)}, r_k + \hat{\varepsilon}_{1-\theta}\sqrt{\mathbf{var}(r_k)}\right].$$

[6]Это предположение выполнено для процессов скользящего среднего MA(q) при $q < k$ (см. п. 14.4).

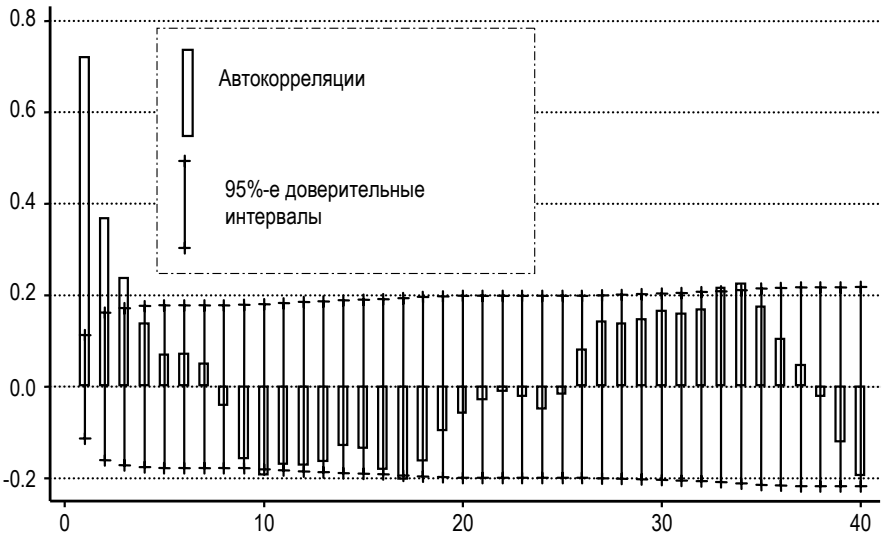

Рис. 11.2. Коррелограмма с доверительными интервалами, основанными на формуле Бартлетта.

На рисунке 11.2 представлена коррелограмма некоторого ряда с доверительными интервалами, основанными на формуле Бартлетта[7]. Для удобства доверительные интервалы построены вокруг нуля, а не вокруг r_k.

11.6.2. Критерий Спирмена

Критерий Спирмена принадлежит к числу непараметрических[8] критериев проверки случайности временного ряда и связан с использованием коэффициента ранговой корреляции Спирмена. Он позволяет уловить наличие или отсутствие тренда в последовательности наблюдений за исследуемой переменной.

Идея критерия состоит в следующем. Допустим, что имеется временной ряд, представленный в хронологической последовательности. Если ряд случайный, то распределение отдельного наблюдения не зависит от того, в каком месте ряда стоит это наблюдение, какой номер оно имеет. При расчете критерия Спирмена в соответствие исходному ряду ставится проранжированный ряд, т.е. полученный в результате сортировки изучаемой переменной по возрастанию или по убыванию. Новый порядок, или ранг θ_t, сравнивается с исходным номером t, соответству-

[7] При использовании нескольких доверительных интервалов следует отдавать себе отчет, что они не являются *совместными*. В связи с этим при одновременном использовании интервалов вероятность ошибки первого рода будет выше θ

[8] В отличие от параметрических, непараметрические критерии не имеют в своей основе априорных предположений о законах распределения временного ряда.

ющим хронологической последовательности. Эти порядки будут независимы для чисто случайного процесса и коррелированы при наличии тенденции.

В крайнем случае, если ряд всегда возрастает, то полученная ранжировка совпадает с исходным порядком наблюдений, т.е. $t = \theta_t$ для всех наблюдений $t = 1, \ldots, T$. В общем случае тесноту связи между двумя последовательностями $1, \ldots, T$ и $\theta_1, \ldots, \theta_T$ можно измерить с помощью обычного коэффициента корреляции:

$$\eta = \frac{\sum_{t=1}^{T} \hat{x}_t \hat{y}_t}{\sqrt{\sum_{t=1}^{T} \hat{x}_t^2 \sum_{t=1}^{T} \hat{y}_t^2}}, \tag{11.14}$$

заменяя x_t на t и y_t на θ_t. Такой показатель корреляции между рангами наблюдений (когда x_t и y_t представляют собой перестановки первых T натуральных чисел) в статистике называется коэффициентом ранговой корреляции Спирмена:

$$\eta = 1 - \frac{6}{T(T^2-1)} \sum_{t=1}^{T} (\theta_t - t)^2. \tag{11.15}$$

Для чисто случайных процессов η имеет нулевое математическое ожидание и дисперсию, равную $\dfrac{1}{T-1}$. В больших выборках величина η приближенно имеет нормальное распределение $N(0, \dfrac{1}{T-1})$. Для малых выборок предпочтительнее использовать в качестве статистики величину $\eta\sqrt{\dfrac{T-2}{1-\eta^2}}$, которая приближенно имеет распределение Стьюдента с $T-2$ степенями свободы. Если искомая расчетная величина по модулю меньше двусторонней критической границы распределения Стьюдента, то нулевая гипотеза о том, что процесс является случайным, принимается и утверждается, что тенденция отсутствует. И наоборот, если искомая величина по модулю превосходит табличное значение, т.е. значение коэффициента η существенно отлично от нуля, то нулевая гипотеза о случайности ряда отвергается. Как правило, это можно интерпретировать как наличие тенденции.

11.6.3. Сравнение средних

Кроме критериев случайности можно использовать различные способы проверки неизменности во времени моментов первого и второго порядков. Из всего многообразия подобных критериев рассмотрим лишь некоторые.

В статистике существует ряд критериев, оценивающих неоднородность выборки путем ранжирования наблюдений с последующим разбиением их на группы

и сравнением межгрупповых показателей. Эти критерии применимы и к временным рядам. При анализе временных рядов нет необходимости в ранжировании наблюдений и поиске адекватного способа сортировки — их порядок автоматически закреплен на временном интервале. Например, можно проверять, является ли математическое ожидание («среднее») постоянным или же в начале ряда оно иное, чем в конце.

Разобьем ряд длиной T на две части примерно равной длины: $x_1, \ldots, x_{T_1}$ и $x_{T_1+1}, \ldots, x_T$. Пусть $\bar{x}_1$ — среднее, s_1^2 — выборочная дисперсия (несмещенная оценка), T_1 — количество наблюдений по первой части ряда, а $\bar{x}_2$, s_2^2 и $T_2 = T - T_1$ — те же величины по второй части.

Статистика Стьюдента для проверки равенства средних в двух частях ряда равна[9]

$$t = (\bar{x}_1 - \bar{x}_2)\sqrt{\frac{T_1 + T_2 - 2}{(1/T_1 + 1/T_2)\left[(T_1 - 1)s_1^2 + (T_2 - 1)s_2^2\right]}}. \qquad (11.16)$$

В предположении, что ряд является гауссовским белым шумом, данная статистика имеет распределение Стьюдента с $T_1 + T_2 - 2$ степенями свободы. Если статистика t по модулю превосходит заданный двусторонний квантиль распределения Стьюдента, то нулевая гипотеза отвергается.

Данный критерий имеет хорошую мощность в случае, если альтернативой является ряд со структурным сдвигом. С помощью данной статистики также можно обнаружить наличие тенденции в изучаемом ряде. Для того чтобы увеличить мощность критерия в этом случае, можно среднюю часть ряда (например, треть наблюдений) не учитывать. При этом $T_1 + T_2 < T$.

Рассчитать статистику при $T_1 + T_2 = T$ можно с помощью вспомогательной регрессии следующего вида:

$$x_t = \alpha z_t + \beta + \varepsilon,$$

где z_t — фиктивная переменная, принимающая значение 0 в первой части ряда и 1 во второй части ряда. Статистика Стьюдента для переменной z_t совпадает со статистикой (11.16).

Критерий сравнения средних применим и в случае, когда ряд x_t не является гауссовским, а имеет какое-либо другое распределение. Однако его использование в случае автокоррелированного нестационарного ряда для проверки неизменности среднего неправомерно, поскольку критерий чувствителен не только

[9] Формула (11.16) намеренно записана без учета того, что $T_1 + T_2 = T$, чтобы она охватывала и вариант использования с $T_1 + T_2 < T$, о котором речь идет ниже.

к структурным сдвигам, но и к автокоррелированности ряда. Поэтому в исходном виде критерий сравнения средних следует считать одним из критериев случайности.

В какой-то степени проблему автокорреляции (а одновременно и гетероскедастичности) можно решить за счет использования устойчивой к автокорреляции и гетероскедастичности оценки Ньюи—Уэста (см. п. 8.3). При использовании этой модификации критерий сравнения средних перестает быть критерием случайности и его можно использовать как критерий стационарности ряда.

Легко распространить этот метод на случай, когда ряд разбивается более чем на две части. В этом случае во вспомогательной регрессии будет более одной фиктивной переменной и следует применять уже F-статистику, а не t-статистику. Так, разбиение на три части может помочь выявить U-образную динамику среднего (например, в первой и третьей части среднее велико, а во второй мало).

Ясно, что с помощью подобных регрессий можно также проверять отсутствие неслучайной зависящей от времени t компоненты другого вида. Например, переменная z_t может иметь вид линейного тренда $z_t = t$. Можно также дополнительно включить в регрессию t^2, t^3 и т.д. и тем самым «уловить» нелинейную тенденцию. Однако в таком виде по указанным выше причинам следует проявлять осторожность при анализе сильно коррелированных рядов.

11.6.4. Постоянство дисперсии

Сравнение дисперсий

Так же как при сравнении средних, при сравнении дисперсий последовательность x_t разбивается на две группы с числом наблюдений T_1 и $T_2 = T - T_1$, для каждой из них вычисляется несмещенная дисперсия s_i^2 и строится **дисперсионное отношение**:

$$F = \frac{s_2^2}{s_1^2}. \tag{11.17}$$

Этот критерий представляет собой частный случай критерия Голдфельда—Квандта (см. п. 8.2).

Если дисперсии однородны и выполнено предположение о нормальности распределения исходного временного ряда (более точно — ряд представляет собой гауссовский белый шум), то F-статистика имеет распределение Фишера $F_{T_2-1,\,T_1-1}$ (см. Приложение А.3.2).

Смысл данной статистики состоит в том, что, когда дисперсии сильно отличаются, статистика будет либо существенно больше единицы, либо существенно

меньше единицы. В данном случае естественно использовать двусторонний критерий (поскольку мы априорно не знаем, растет дисперсия или падает). Это, конечно, не совсем обычно для критериев, основанных на F-статистике. Для уровня θ можно взять в качестве критических границ такие величины, чтобы вероятность попадания и в левый, и в правый хвост была одной и той же — $\theta/2$.

Нулевая гипотеза состоит в том, что дисперсия однородна. Если дисперсионное отношение попадает в один из двух хвостов, то нулевая гипотеза отклоняется.

Мощность критерия можно увеличить, исключив часть центральных наблюдений. Этот подход оправдан в случае монотонного поведения дисперсии временного ряда, тогда дисперсионное отношение покажет больший разброс значений.

Если же временной ряд не монотонен, например имеет U-образную форму, то мощность теста в результате исключения центральных наблюдений существенно уменьшается.

Как и в случае сравнения средних, критерий применим только в случае, когда проверяемый процесс является белым шумом. Если же, например, ряд является стационарным, но автокоррелированным, то данный критерий применять не следует.

11.7. Лаговый оператор

Одним из основных понятий, употребляемых при моделировании временных рядов, является понятие **лага**. В буквальном смысле в переводе с английского лаг — запаздывание. Под лагом некоторой переменной понимают ее значение в предыдущие периоды времени. Например, для переменной x_t лагом в k периодов будет x_{t-k}.

При работе с временными рядами удобно использовать **лаговый оператор** $\mathbf{L}$, т.е. оператор запаздывания, сдвига назад во времени. Хотя часто использование этого оператора сопряжено с некоторой потерей математической строгости, однако это окупается значительным упрощением вычислений.

Если к переменной применить лаговый оператор, то в результате получится лаг этой переменной:

$$\mathbf{L}x_t = x_{t-1}.$$

Использование лагового оператора $\mathbf{L}$ обеспечивает сжатую запись разностных уравнений и помогает изучать свойства целого ряда процессов.

Удобство использования лагового оператора состоит в том, что с ним можно обращаться как с обычной переменной, т.е. операторы можно преобразовывать сами по себе, без учета тех временных рядов, к которым они применяются. Основное

отличие лагового оператора от обычной переменной состоит в том, что оператор должен стоять **перед** тем рядом, к которому применяется, т.е. нельзя переставлять местами лаговый оператор и временной ряд.

Как и для обычных переменных, существуют функции от лагового оператора, они, в свою очередь, тоже являются операторами. Простейшая функция — степенная.

По определению, для целых m

$$\mathbf{L}^m x_t = x_{t-m},$$

т.е. $\mathbf{L}^m$, действующий на x_t, означает запаздывание этой переменной на m периодов.

Продолжая ту же логику, можно определить многочлен от лагового оператора, или **лаговый многочлен**:

$$\alpha(\mathbf{L}) = \sum_{i=0}^{m} \alpha_i \mathbf{L}^{t-k} = \alpha_0 + \alpha_1 \mathbf{L} + \cdots + \alpha_m \mathbf{L}^m.$$

Если применить лаговый многочлен к переменной x_t, то получается

$$\alpha(\mathbf{L})x_t = (\alpha_0 + \alpha_1 \mathbf{L} + \cdots + \alpha_m \mathbf{L}^m)x_t = \alpha_0 x_t + \alpha_1 x_{t-1} + \cdots + \alpha_m x_{t-m}.$$

Нетрудно проверить, что лаговые многочлены можно перемножать как обычные многочлены. Например,

$$(\alpha_0 + \alpha_1 \mathbf{L})(\beta_0 + \beta_1 \mathbf{L}) = \alpha_0\beta_0 + (\alpha_1\beta_0 + \alpha_0\beta_1)\mathbf{L} + \alpha_1\beta_1\mathbf{L}^2.$$

При $m \to \infty$ получается бесконечный степенной ряд от лагового оператора:

$$\left(\sum_{i=0}^{\infty} \alpha_i \mathbf{L}^i\right) x_t = (\alpha_0 + \alpha_1 \mathbf{L} + \alpha_2 \mathbf{L}^2 + \cdots)x_t =$$

$$= \alpha_0 x_t + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots = \sum_{i=0}^{\infty} \alpha_i x_{t-i}.$$

Полезно помнить следующие свойства лаговых операторов:

1) Лаг константы есть константа: $\mathbf{L}C = C$.

2) Дистрибутивность: $(\mathbf{L}^i + \mathbf{L}^j)x_t = \mathbf{L}^i x_t + \mathbf{L}^j x_t = x_{t-i} + x_{t-j}$.

3) Ассоциативность: $\mathbf{L}^i\mathbf{L}^j x_t = \mathbf{L}^i(\mathbf{L}^j x_t) = \mathbf{L}^i x_{t-j} = x_{t-i-j}$. Заметим, что: $\mathbf{L}^0 x_t = x_t$, т.е. $\mathbf{L}^0 = I$.

4) $\mathbf{L}$, возведенный в отрицательную степень, — опережающий оператор: $\mathbf{L}^{-i}x_t = x_{t+i}$.

5) При $|\alpha| < 1$ бесконечная сумма
$(1 + \alpha\mathbf{L} + \alpha^2\mathbf{L}^2 + \alpha^3\mathbf{L}^3 + \dots)x_t = (1 - \alpha\mathbf{L})^{-1}x_t$.

Для доказательства умножим обе части уравнения на $(1 - \alpha\mathbf{L})$:

$(1 - \alpha\mathbf{L})(1 + \alpha\mathbf{L} + \alpha^2\mathbf{L}^2 + \alpha^3\mathbf{L}^3 + \dots)x_t = x_t$, поскольку при $|\alpha| < 1$ выражение $\alpha^n\mathbf{L}^n x_t \to 0$ при $n \to \infty$.

Кроме лагового оператора в теории временных рядов широко используют **разностный оператор** $\boldsymbol{\Delta}$, который определяется следующим образом:

$$\Delta = 1 - \mathbf{L},$$

так что $\boldsymbol{\Delta}x_t = (1 - \mathbf{L})x_t = x_t - x_{t-1}$.

Разностный оператор превращает исходный ряд в ряд **первых разностей**.

Ряд d-х разностей (разностей d-го порядка) получается как степень разностного оператора, то есть применением разностного оператора d раз.

При $d = 2$ получается $\boldsymbol{\Delta}^2 = (1 - \mathbf{L})^2 = 1 - 2\mathbf{L} + \mathbf{L}^2$, поэтому $\boldsymbol{\Delta}^2 x_t = (1 - 2\mathbf{L} + \mathbf{L}^2)x_t = x_t - 2x_{t-1} + x_{t-2}$.

Для произвольного порядка d следует использовать формулу бинома Ньютона:

$$\boldsymbol{\Delta}^d = (1 - \mathbf{L})^d = \sum_{k=0}^{d}(-1)^k C_d^k \mathbf{L}^k \quad \text{, где } C_d^k = \frac{d!}{k!(d-k)!},$$

так что $\boldsymbol{\Delta}^d x_t = (1 - \mathbf{L})^d x_t = \sum_{k=0}^{d}(-1)^k C_d^k x_{t-k}$.

11.8. Модели регрессии с распределенным лагом

Часто при моделировании экономических процессов на изучаемую переменную x_t влияют не только текущие значения объясняющего фактора z_t, но и его лаги. Типичным примером являются капиталовложения: они всегда дают результат с некоторым лагом.

Модель **распределенного лага** можно записать следующим образом:

$$x_t = \mu + \sum_{j=0}^{q}\alpha_j z_{t-j} + \varepsilon_t = \mu + \alpha(\mathbf{L})z_t + \varepsilon_t. \qquad (11.18)$$

где q — величина наибольшего лага, $\alpha(B) = \sum_{j=0}^{q} \alpha_j^j \mathbf{L}^j$ — лаговый многочлен, ε_t — случайное возмущение, ошибка. Коэффициенты α_j задают **структуру лага** и называются весами. Конструкцию $\sum_{j=0}^{q} \alpha_j z_{t-j}$ часто называют **«скользящим средним»** переменной z_t [10].

Рассмотрим практические проблемы получения оценок коэффициентов α_j в модели (11.18). Модель распределенного лага можно оценивать обычным методом наименьших квадратов, если выполнены стандартные предположения регрессионного анализа. В частности, количество лагов не должно быть слишком большим, чтобы количество регрессоров не превышало количество наблюдений, и все лаги переменной z_t, т.е. z_{t-j} $(j = 0, \ldots, q)$, не должны быть коррелированы с ошибкой ε_t.

Одна из проблем, возникающих при оценивании модели распределенного лага, найти величину наибольшего лага q. При этом приходится начать с некоторого предположения, то есть взять за основу число Q, выше которого q быть не может. Выбор такого числа осуществляется на основе некоторой дополнительной информации, например, опыта человека, который оценивает модель. Можно предложить следующие способы практического определения величины q.

1) Для каждого конкретного q оценивается модель (11.18), и из нее берется t-статистика для последнего коэффициента, т.е. α_q. Эти t-статистики рассматриваются в обратном порядке, начиная с $q = Q$ (и заканчивая $q = 0$). Как только t-статистика оказывается значимой при некотором наперед заданном уровне, то следует остановиться и выбрать соответствующую величину q.

2) Следует оценить модель (11.18) при $q = Q$. Из этой регрессии берутся F-статистики для проверки нулевой гипотезы о том, что коэффициенты при последних $Q - q + 1$ лагах, т.е. $\alpha_q, \ldots, \alpha_Q$, одновременно равны нулю:

$$H_0 : \alpha_j = 0, \quad \forall j = q, \ldots, Q.$$

Соответствующие F-статистики рассчитываются по формулам:

$$F_q = \frac{(RSS_Q - RSS_{q-1})/(Q - q + 1)}{RSS_Q/(T - Q - 2)},$$

где RSS_r — сумма квадратов остатков из модели распределенного лага при $q = r$, T — количество наблюдений. При этом при проведении расчетов для сопоставимости во всех моделях надо использовать одни и те же наблюдения — те, которые использовались при $q = Q$ (следовательно, при всех q используется одно и то же T). Эти F-статистики рассматриваются в обратном порядке от $q = Q$ до $q = 0$ (в последнем случае в модели переменная z отсутствует). Как только F-статистика

[10] Другое часто используемое название — «линейный фильтр».

оказывается значимой при некотором наперед заданном уровне, то следует остановиться и выбрать соответствующую величину q.

3) Для всех q от $q = 0$ до $q = Q$ рассчитывается величина **информационного критерия**, а затем выбирается модель с наименьшим значением этого информационного критерия. Приведем наиболее часто используемые информационные критерии.

Информационный критерий Акаике:

$$AIC = \ln(\frac{RSS}{T}) + \frac{2(n+1)}{T},$$

где RSS сумма квадратов остатков в модели, T — фактически использовавшееся количество наблюдений, n — количество факторов в регрессии (не считая константу). В рассматриваемом случае $n = q+1$, а $T = T_0 - q$, где T_0 — количество наблюдений при $q = 0$.

Байесовский информационный критерий (информационный критерий Шварца):

$$BIC = \ln(\frac{RSS}{T}) + \frac{(n+1)\ln T}{T}.$$

Как видно из формул, критерий Акаике благоприятствует выбору более короткого лага, чем критерий Шварца.

11.9. Условные распределения

Условные распределения играют важную роль в анализе временных рядов, особенно при прогнозировании. Мы не будем вдаваться в теорию условных распределений, это предмет теории вероятностей (определения и свойства условных распределений см. в Приложении А.3.1). Здесь мы рассмотрим лишь основные правила, по которым можно проводить преобразования. При этом будем использовать следующее стандартное обозначение: если речь идет о распределении случайной величины X, условном по случайной величине Y (условном относительно Y), то это записывается в виде $X|Y$.

Основное правило работы с условными распределениями, которое следует запомнить, состоит в том, что если рассматривается распределение, условное относительно случайной величины Y, то с Y и ее функциями следует поступать так же, как с детерминированными величинами. Например, для условных математических ожиданий и дисперсий выполняется

$$\mathbf{E}\,(\alpha(Y) + \beta(Y)X|Y) = \alpha(Y) + \beta(Y)\mathbf{E}(X|Y),$$
$$\mathbf{var}\,(\alpha(Y) + \beta(Y)X|Y) = \beta^2(Y)\mathbf{var}(X|Y).$$

Как и обычное безусловное математическое ожидание, условное ожидание представляет собой линейный оператор. В частности, ожидание суммы есть сумма ожиданий:

$$\mathbf{E}\left(X_1 + X_2|Y\right) = \mathbf{E}(X_1|Y) + \mathbf{E}(X_2|Y).$$

Условное математическое ожидание $\mathbf{E}(X|Y)$ в общем случае не является детерминированной величиной, т.е. оно является случайной величиной, которая может иметь свое математическое ожидание, характеризоваться положительной дисперсией и т.п.

Если от условного математического ожидания случайной величины X еще раз взять обычное (безусловное) математическое ожидание, то получится обычное (безусловное) математическое ожидание случайной величины X. Таким образом, действует следующее **правило повторного взятия ожидания**:

$$\mathbf{E}\left(\mathbf{E}(X|Y)\right) = \mathbf{E}(X).$$

В более общей форме это правило имеет следующий вид:

$$\mathbf{E}\left(\mathbf{E}(X|Y,Z)|Y\right) = \mathbf{E}(X|Y),$$

что позволяет применять его и тогда, когда второй раз ожидание берется не полностью, т.е. не безусловное, а лишь условное относительно информации, являющейся частью информации, относительно которой ожидание бралось первый раз.

Если случайные величины X и Y статистически независимы, то распределение X, условное по Y, совпадает с безусловным распределением X. Следовательно, для независимых случайных величин X и Y выполнено, в частности,

$$\mathbf{E}(X|Y) = \mathbf{E}(X), \quad \mathbf{var}(X|Y) = \mathbf{var}(X).$$

11.10. Оптимальное в среднеквадратическом смысле прогнозирование: общая теория

11.10.1. Условное математическое ожидание как оптимальный прогноз

Докажем в абстрактном виде, безотносительно к моделям временных рядов, общее свойство условного математического ожидания, заключающееся в том, что оно минимизирует средний квадрат ошибки прогноза.

Предположим, что строится прогноз некоторой случайной величины x на основе другой случайной величины, z, и что точность прогноза при этом оценивается

на основе среднего квадрата ошибки прогноза $\eta = x - x^p(z)$, где $x^p(z)$ — прогнозная функция. Таким образом, требуется получить прогноз, который бы минимизировал

$$\mathbf{E}\left(\eta^2\right) = \mathbf{E}\left[(x - x^p(z))^2\right].$$

Оказывается, что наилучший в указанном смысле прогноз дает математическое ожидание x, условное относительно z, т.е. $\mathbf{E}(x|z)$, которое мы будем обозначать $\bar{x}(z)$. Докажем это. Возьмем произвольный прогноз $x^p(z)$ и представим ошибку прогноза в виде:

$$x - x^p(z) = \eta = (x - \bar{x}(z)) + (\bar{x}(z) - x^p(z)).$$

Найдем сначала математическое ожидание квадрата ошибки, условное относительно z:

$$\mathbf{E}\left(\eta^2|z\right) = \mathbf{E}\left[(x - \bar{x}(z))^2|z\right] + \\ + 2\mathbf{E}\left[(x - \bar{x}(z))(\bar{x}(z) - x^p(z))|z\right] + \mathbf{E}\left[(\bar{x}(z) - x^p(z)t)^2|z\right].$$

При взятии условного математического ожидания с функциями z можно обращаться как с константами. Поэтому

$$\mathbf{E}\left[(\bar{x}(z) - x^p(z))^2|z\right] = (\bar{x}(z) - x^p(z))^2$$

и

$$\mathbf{E}\left[(x - \bar{x}(z))(\bar{x}(z) - x^p(z))|z\right] = \mathbf{E}\left(x - \bar{x}(z)|z\right)(\bar{x}(z) - x^p(z)) = \\ = (\bar{x}(z) - \bar{x}(z))(\bar{x}(z) - x^p(z)) = 0.$$

Используя эти соотношения, получим

$$\mathbf{E}\left(\eta^2|z\right) = \mathbf{E}\left((x - \bar{x}(z))^2|z\right) + (\bar{x}(z) - x^p(z))^2.$$

Если теперь взять от обеих частей безусловное математическое ожидание, то (по правилу повторного взятия ожидания) получится

$$\mathbf{E}\left(\eta^2\right) = \mathbf{E}\left[(x - \bar{x}(z))^2\right] + \mathbf{E}\left[(\bar{x}(z) - x^p(z))^2\right].$$

Поскольку второе слагаемое неотрицательно, то

$$\mathbf{E}\left[(x - x^p(z))^2\right] = \mathbf{E}\left(\eta^2\right) \leq \mathbf{E}\left[(x - \bar{x}(z))^2\right].$$

Другими словами, средний квадрат ошибки прогноза достигает минимума при $x^p(z) = \bar{x}(z) = \mathbf{E}(x|z)$.

Оптимальный прогноз $x^p(z) = \bar{x}(z) = \mathbf{E}(x|z)$ является несмещенным. Действительно, по правилу повторного взятия ожидания

$$\mathbf{E}(\mathbf{E}(x|z)) = \mathbf{E}(x).$$

Поэтому

$$\mathbf{E}\eta = \mathbf{E}(x - x^p(z)) = \mathbf{E}(x) - \mathbf{E}(\mathbf{E}(x|z)) = 0.$$

11.10.2. Оптимальное линейное прогнозирование

Получим теперь формулу для оптимального (в смысле минимума среднего квадрата ошибки) линейного прогноза. Пусть случайная переменная z, на основе которой делается прогноз x, представляет собой n-мерный вектор: $z = (z_1, \ldots, z_n)'$. Без потери общности можно предположить, что x и z имеют нулевое математическое ожидание. Будем искать прогноз x в виде линейной комбинации z_j:

$$x^p(z) = \alpha_1 z_1 + \ldots + \alpha_n z_n = z'\alpha,$$

где $\alpha = (\alpha_1, \ldots, \alpha_n)'$ — вектор коэффициентов. Любой прогноз такого вида является несмещенным, поскольку, как мы предположили, $\mathbf{E}x = 0$ и $\mathbf{E}z = 0$.

Требуется решить задачу минимизации среднего квадрата ошибки (в данном случае это эквивалентно минимизации дисперсии ошибки):

$$\mathbf{E}\left[(x - x^p(z))^2\right] \to \min_{\alpha}!$$

Средний квадрат ошибки можно представить в следующем виде:

$$\mathbf{E}\left[(x - x^p(z))^2\right] = \mathbf{E}\left[x^2 - 2\alpha' zx + \alpha' zz'\alpha\right] = \sigma_x^2 - 2\alpha' M_{zx} + \alpha' M_{zz}\alpha,$$

где $\sigma_x^2 = \mathbf{E}x^2$ — дисперсия x, $M_{zx} = \mathbf{E}[zx]$ — вектор, состоящий из ковариаций z_j и x, а $M_{zz} = \mathbf{E}[zz']$ — ковариационная матрица z. (Напомним, что мы рассматриваем процессы с нулевым математическим ожиданием.) Дифференцируя по α, получим следующие нормальные уравнения:

$$-2M_{zx} + 2M_{zz}\alpha = 0,$$

откуда

$$\alpha = M_{zz}^{-1} M_{zx}.$$

Очевидна аналогия этой формулы с оценками МНК, только матрицы вторых моментов здесь не выборочные, а теоретические.

Таким образом, оптимальный линейный прогноз имеет вид:

$$x^p(z) = z' M_{zz}^{-1} M_{zx}. \tag{11.19}$$

Ошибка оптимального линейного прогноза равна

$$\eta = x - x^p(z) = x - z' M_{zz}^{-1} M_{zx}.$$

Эта ошибка некоррелирована с z, то есть с теми переменными, по которым делается прогноз. Действительно, умножая на z и беря математическое ожидание, получим

$$\mathbf{E}(z\eta) = \mathbf{E}\left(zx - zz' M_{zz}^{-1} M_{zx}\right) = M_{zx} - M_{zz} M_{zz}^{-1} M_{zx},$$

т.е.

$$\mathbf{E}(z\eta) = 0.$$

Средний квадрат ошибки оптимального прогноза равен

$$\mathbf{E}\left(\eta^2\right) = \mathbf{E}\left[(x - x^p(z))^2\right] = \sigma_x^2 - 2M_{xz} M_{zz}^{-1} M_{zx} + M_{xz} M_{zz}^{-1} M_{zz} M_{zz}^{-1} M_{zx}.$$

После преобразований получаем

$$\mathbf{E}\left(\eta^2\right) = \sigma_x^2 - M_{xz} M_{zz}^{-1} M_{zx}. \tag{11.20}$$

Несложно увидеть аналогии между приведенными формулами и формулами МНК. Таким образом, данные рассуждения можно считать одним из возможных теоретических обоснований линейного МНК.

Для того чтобы применить приведенные формулы, требуется, чтобы матрица M_{zz} была обратимой. Если она вырождена, то это означает наличие мультиколлинеарности между переменными z.

Проблема вырожденности решается просто. Во-первых, можно часть «лишних» компонент z не использовать — оставить только такие, которые линейно независимы между собой. Во-вторых, в вырожденном случае прогноз можно получить по той же формуле $x^p(z) = z'\alpha$, взяв в качестве коэффициентов α любое решение системы линейных уравнений $M_{zz}\alpha = M_{zx}$ (таких решений будет бесконечно много). Средний квадрат ошибки прогноза рассчитывается по формуле:

$$\mathbf{E}\left(\eta^2\right) = \sigma_x^2 - M_{xz}\alpha.$$

В общем случае оптимальный линейный прогноз (11.19) не совпадает с условным математическим ожиданием $\mathbf{E}(x|z)$. Другими словами, он не является оптимальным среди всех возможных прогнозов. Пусть, например, z имеет стандартное нормальное распределение: $z \sim N(0,1)$, а x связан с z формулой $x = z^2 - 1$. Тогда, поскольку x и z некоррелированы, то $\alpha = 0$, и оптимальный линейный прогноз имеет вид $x^p(z) = 0$ при среднем квадрате ошибки прогноза равном $\mathbf{E}((z^2-1)^2) = 2$. В то же время прогноз по нелинейной формуле $x^p(z) = z^2 - 1$ будет безошибочным (средний квадрат ошибки прогноза равен 0).

В частном случае, когда совместное распределение x и z является многомерным нормальным распределением:

$$\begin{pmatrix} x \\ z \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0_n \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & M_{xz} \\ M_{zx} & M_{zz} \end{pmatrix}\right),$$

оптимальный линейный прогноз является просто оптимальным. Это связано с тем, что по свойствам многомерного нормального распределения (см. Приложение А.3.2) условное распределение x относительно z будет иметь следующий вид:

$$x|z \sim N\left(z' M_{zz}^{-1} M_{zx}, \sigma_x^2 - M_{xz} M_{zz}^{-1} M_{zx}\right).$$

Таким образом, $\mathbf{E}(x|z) = z' M_{zz}^{-1} M_{zx}$, что совпадает с формулой оптимального линейного прогноза (11.19).

11.10.3. Линейное прогнозирование стационарного временного ряда

Пусть x_t — слабо стационарный процесс с нулевым математическим ожиданием. Рассмотрим проблему построения оптимального линейного прогноза этого процесса, если в момент t известны значения ряда, начиная с момента 1, т.е. только конечный ряд $x = (x_1, \ldots, x_t)$. Предположим, что делается прогноз на τ шагов вперед, т.е. прогноз величины $x_{t+\tau}$. Для получения оптимального линейного (по x) прогноза можно воспользоваться формулой (11.19). В случае стационарного временного ряда ее можно переписать в виде:

$$x_t(\tau) = x' \Gamma_t^{-1} \gamma^{t,\tau}, \tag{11.21}$$

где

$$\Gamma_t = \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{t-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{t-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{t-1} & \gamma_{t-2} & \cdots & \gamma_0 \end{pmatrix}$$

— автоковариационная матрица ряда $(x_1, \ldots, x_t)$, а вектор $\gamma^{t,\tau}$ составлен из ковариаций $x_{t+\tau}$ с $(x_1, \ldots, x_t)$, т.е.

$$\gamma^{t,\tau} = (\gamma_{t+\tau-1}, \ldots, \gamma_\tau)'.$$

Можно заметить, что автоковариации здесь нужно знать только с точностью до множителя. Например, их можно заменить автокорреляциями.

Рассмотрим теперь прогнозирование на один шаг вперед. Обозначим через γ^t вектор, составленный из ковариаций x_{t+1} с $(x_1, \ldots, x_t)$, т.е. $\gamma^t = (\gamma_t, \ldots, \gamma_1)' = \gamma^{t,1}$. Прогноз задается формулой:

$$x_t(1) = x'\Gamma_t^{-1}\gamma^t = x'\alpha^t = \sum_{i=1}^{t} \alpha_i^t x_{t-i}.$$

Прогноз по этой формуле можно построить только если матрица Γ_t неособенная. Коэффициенты α_i^t, минимизирующие средний квадрат ошибки прогноза, задаются нормальными уравнениями $\Gamma_t\alpha = \gamma^t$ или, в развернутом виде,

$$\sum_{i=1}^{t} \alpha_i^t \gamma_{|k-i|} = \gamma_k,\ k = 1, \ldots, t.$$

Ошибка прогноза равна

$$\eta = x_{t+1} - x_t(1).$$

Применив (11.20), получим, что средний квадрат этой ошибки равен

$$\mathbf{E}\left(\eta^2\right) = \gamma_0 - \gamma^{t\prime}\Gamma_t^{-1}\gamma^t.$$

Заметим, что $\gamma_0 - \gamma^{t\prime}\Gamma_t^{-1}\gamma^t = |\Gamma_{t+1}| / |\Gamma_t|$, т.е предыдущую формулу можно переписать как

$$\mathbf{E}\left(\eta^2\right) = |\Gamma_{t+1}| / |\Gamma_t|. \qquad (11.22)$$

Действительно, матрицу Γ_{t+1} можно представить в следующей блочной форме:

$$\Gamma_{t+1} = \begin{pmatrix} \Gamma_t & \gamma^t \\ \gamma^{t\prime} & \gamma_0 \end{pmatrix}.$$

По правилу вычисления определителя блочной матрицы имеем:

$$|\Gamma_{t+1}| = \left(\gamma_0 - \gamma^{t\prime}\Gamma_t^{-1}\gamma^t\right)|\Gamma_t|.$$

Если $|\Gamma_{t+1}| = 0$, т.е. если матрица Γ_{t+1} вырождена, то средний квадрат ошибки прогноза окажется равным нулю, т.е. оптимальный линейный прогноз будет безошибочным. Процесс, для которого существует такой безошибочный линейный прогноз, называют линейно детерминированным.

Укажем без доказательства следующее свойство автоковариационных матриц: если матрица Γ_t является вырожденной, то матрица Γ_{t+1} также будет вырожденной.

Отсюда следует, что на основе конечного отрезка стационарного ряда $(x_1, \ldots, x_t)$ можно сделать безошибочный линейный прогноз на один шаг вперед в том и только в том случае, если автоковариационная матрица Γ_{t+1} является вырожденной ($|\Gamma_{t+1}| = 0$).

> Действительно, пусть существует безошибочный линейный прогноз. Возможны два случая: $|\Gamma_t| \neq 0$ и $|\Gamma_t| = 0$. Если $|\Gamma_t| \neq 0$, то средний квадрат ошибки прогноза равен $|\Gamma_{t+1}| / |\Gamma_t|$, откуда $|\Gamma_{t+1}| = 0$, если же $|\Gamma_t| = 0$, то из этого также следует $|\Gamma_{t+1}| = 0$.
>
> Наоборот, если $|\Gamma_{t+1}| = 0$, то найдется s ($s \leqslant t$) такое, что $|\Gamma_{s+1}| = 0$, но $|\Gamma_s| \neq 0$. Тогда можно сделать безошибочный прогноз для x_{t+1} на основе $(x_{1+t-s}, \ldots, x_t)$, а, значит, и на основе $(x_1, \ldots, x_t)$.

При использовании приведенных формул на практике возникает трудность, связанная с тем, что обычно теоретические автоковариации γ_k неизвестны. Требуется каким-то образом получить оценки автоковариаций. Обычные выборочные автоковариации c_k здесь не подойдут, поскольку при больших k (сопоставимых с длиной ряда) они являются очень неточными оценками γ_k. Можно предложить следующий подход[11]:

1) Взять за основу некоторую параметрическую модель временного ряда. Пусть β — соответствующий вектор параметров. Рассчитать теоретические автоковариации для данной модели в зависимости от параметров: $\gamma_k = \gamma_k(\beta)$.

[11] Этот подход, в частности, годится для стационарных процессов ARMA. В пункте 14.8 дается альтернативный способ прогнозирования в рамках модели ARMA.

2) Оценить параметры на основе имеющихся данных. Пусть b — соответствующие оценки.

3) Получить оценки автоковариаций, подставив b в формулы теоретических автоковариаций: $\gamma_k \approx \gamma_k(b)$.

4) Использовать для прогнозирования формулу (11.21), заменяя теоретические автоковариации полученными оценками автоковариаций.

11.10.4. Прогнозирование по полной предыстории. Разложение Вольда

Можно распространить представленную выше теорию на прогнозирование ряда в случае, когда в момент t известна полная предыстория $\Omega_t = (x_t, x_{t-1}, \dots)$. Можно определить соответствующий прогноз как *предел* прогнозов, полученных на основе конечных рядов $(x_t, x_{t-1}, \dots, x_{t-j})$, $j = t, t-1, \dots, -\infty$. Без доказательства отметим, что этот прогноз будет оптимальным в среднеквадратическом смысле. Если рассматривается процесс, для которого $|\Gamma_t| \neq 0\ \forall t$, то по аналогии с (11.22) средний квадрат ошибки такого прогноза равен

$$\mathbf{E}\left(\eta^2\right) = \lim_{t\to\infty} \frac{|\Gamma_{t+1}|}{|\Gamma_t|}.$$

Заметим, что всегда выполнено $0 < \frac{|\Gamma_{t+1}|}{|\Gamma_t|} \leq \frac{|\Gamma_t|}{|\Gamma_{t-1}|}$, т.е. средний квадрат ошибки не увеличивается с увеличением длины ряда, на основе которого делается прогноз, и ограничен снизу нулем, поэтому указанный предел существует всегда.

Существуют процессы, для которых $|\Gamma_t| \neq 0\ \forall t$, т.е. для них нельзя сделать безошибочный прогноз, имея только конечный отрезок ряда, однако

$$\lim_{t\to\infty} \frac{|\Gamma_{t+1}|}{|\Gamma_t|} = 0.$$

Такой процесс по аналогии можно назвать линейно детерминированным. Его фактически можно безошибочно предсказать, если имеется полная предыстория процесса $\Omega_t = (x_t,\ x_{t-1},\ \dots)$.

Если же данный предел положителен, то линейный прогноз связан с ошибкой: $\mathbf{E}\left(\eta^2\right) > 0$. Такой процесс можно назвать регулярным.

Выполнены следующие свойства стационарных рядов.

А. Пусть x_t — слабо стационарный временной ряд, и пусть η_t — ошибки одношагового оптимального линейного прогноза по полной предыстории процесса $(x_{t-1}, x_{t-2}, \dots)$. Тогда ошибки η_t являются белым шумом, т.е. имеют нулевое ма-

тематическое ожидание, не автокоррелированы и имеют одинаковую дисперсию:

$$\begin{aligned}
&\mathbf{E}\,(\eta_t) = 0,\ \forall t,\\
&\mathbf{E}\,(\eta_s \eta_t) = 0,\ \text{при } s \neq t,\\
&\mathbf{E}\left(\eta^2\right) = \sigma_\eta^2,\ \forall t.
\end{aligned}$$

B. Пусть, кроме того, x_t является регулярным, т.е. $\mathbf{E}\,(\eta^2) = \sigma_\eta^2 > 0$. Тогда он представим в следующем виде:

$$x_t = \sum_{i=0}^{\infty} \psi_i \eta_{t-i} + v_t, \tag{11.23}$$

где $\psi_0 = 1$, $\sum_{i=0}^{\infty} \psi_i^2 < \infty$; процесс v_t здесь является слабо стационарным, линейно детерминированным и не коррелирован с ошибками η_t: $\mathbf{E}\,(\eta_s v_t) = 0$ при $\forall s, t$. Такое представление единственно.

Утверждения **A** и **B** составляют **теорему Вольда**. Эта теорема является одним из самых фундаментальных результатов в теории временных рядов. Утверждение **B** говорит о том, что любой стационарный процесс можно представить в виде так называемого **линейного фильтра** от белого шума[12] плюс линейно детерминированная компонента. Это так называемое **разложение Вольда**.

Доказательство теоремы Вольда достаточно громоздко. Мы не делаем попытки его излагать и даже обсуждать; отсылаем заинтересованных читателей к гл. 7 книги Т. Андерсона [2].

Заметим, что коэффициенты разложения ψ_i удовлетворяют соотношению

$$\psi_i = \frac{\mathbf{E}\,(\eta_{t-i} x_t)}{\mathbf{E}\left(\eta_{t-i}^2\right)} = \frac{\mathbf{E}\,(\eta_{t-i} x_t)}{\sigma_\eta^2}.$$

Для того чтобы это показать, достаточно умножить (11.23) на η_{t-i} и взять математическое ожидание от обеих частей.

Разложение Вольда имеет в своей основе прогнозирование на один шаг вперед. С другой стороны, если мы знаем разложение Вольда для процесса, то с помощью него можно делать прогнозы. Предположим, что в момент t делается прогноз на τ шагов вперед, т.е. прогноз величины $x_{t+\tau}$ на основе предыстории $\Omega_t = (x_t, x_{t-1}, \dots)$. Сдвинем формулу разложения Вольда (11.23) на τ периодов вперед:

$$x_{t+\tau} = \sum_{i=0}^{\infty} \psi_i \eta_{t+\tau-i} + \tau_{t+\tau}.$$

[12] Можно назвать первое слагаемое в 11.23 также процессом скользящего среднего бесконечного порядка MA(∞). Процессы скользящего среднего обсуждаются в пункте 14.4.

Второе слагаемое, $v_{t+\tau}$, можно предсказать без ошибки, зная Ω_t. Из первой суммы первые τ слагаемых не предсказуемы на основе Ω_t. При прогнозировании их можно заменить ожидаемыми значениями — нулями. Из этих рассуждений следует следующая формула прогноза:

$$x_t(\tau) = \sum_{i=\tau}^{\infty} \psi_i \eta_{t+\tau-i} + \tau_{t+\tau}.$$

Без доказательства отметим, что $x_t(\tau)$ является оптимальным линейным прогнозом. Ошибка прогноза при этом будет равна

$$\sum_{i=0}^{\tau-1} \psi_i \eta_{t+\tau-i}.$$

Поскольку η_t — белый шум с дисперсией σ_η^2, то средний квадрат ошибки прогноза равен

$$\sigma_\eta^2 \sum_{i=0}^{\tau-1} \psi_i^2.$$

Напоследок обсудим природу компоненты v_t. Простейший пример линейно детерминированного ряда — это, говоря неформально, «случайная константа»:

$$v_t = \xi,$$

где ξ — наперед заданная случайная величина, $\mathbf{E}\xi = 0$.

Типичный случай линейно детерминированного ряда — это, говоря неформально, «случайная синусоида»:

$$v_t = \xi \cos(\omega t + \varphi),$$

где ω — фиксированная частота, ξ и φ — независимые случайные величины, причем φ имеет равномерное распределение на отрезке $[0; 2\pi]$.

Это два примера случайных слабо стационарных рядов, которые можно безошибочно предсказывать на основе предыстории. Первый процесс можно моделировать с помощью константы, а второй — с помощью линейной комбинации синуса и косинуса:

$$\alpha \cos(\omega t) + \beta \sin(\omega t).$$

С точки зрения практики неформальный вывод из теоремы Вольда состоит в том, что любые стационарные временные ряды можно моделировать при помощи моделей линейного фильтра с добавлением констант и гармонических трендов.

11.11. Упражнения и задачи

Упражнение 1

1.1. Дан временной ряд $x = (5,\ 1,\ 1,\ -3,\ 2,\ 9,\ 6,\ 2,\ 5,\ 2)'$.

Вычислите среднее, дисперсию (смещенную), автоковариационную и автокорреляционную матрицы.

1.2. Для временного ряда $y = (6,\ 6,\ 1,\ 6,\ 0,\ 6,\ 6,\ 4,\ 3,\ 2)'$ повторите упражнение 1.1.

1.3. Вычислите кросс-ковариации и кросс-корреляции для рядов x и y из предыдущих упражнений для сдвигов $-9, \ldots, 0, \ldots, 9$.

1.4. Для временного ряда $x = (7,\ -9,\ 10,\ -2,\ 21,\ 13,\ 40,\ 36,\ 67,\ 67)$ оцените параметры полиномиального тренда второго порядка. Постройте точечный и интервальный прогнозы по тренду на 2 шага вперед.

1.5. Сгенерируйте 20 рядов, задаваемых полиномиальным трендом третьего порядка $\tau_t = 5 + 4t - 0.07t^2 + 0.0005t^3$ длиной 100 наблюдений, с добавлением белого шума с нормальным распределением и дисперсией 20.

Допустим, истинные значения параметров тренда неизвестны.

а) Для 5 рядов из 20 оцените полиномиальный тренд первого, второго и третьего порядков и выберите модель, которая наиболее точно аппроксимирует сгенерированные данные.

б) Для 20 рядов оцените полиномиальный тренд третьего порядка по первым 50 наблюдениям. Вычислите оценки параметров тренда и их ошибки. Сравните оценки с истинными значениями параметров.

в) Проведите те же вычисления, что и в пункте (б), для 20 рядов, используя 100 наблюдений. Результаты сравните.

г) Используя предшествующие расчеты, найдите точечные и интервальные прогнозы на три шага вперед с уровнем доверия 95%.

1.6. Найдите данные о динамике денежного агрегата M0 в России за 10 последовательных лет и оцените параметры экспоненциального тренда.

1.7. Ряд $x = (0.02,\ 0.05,\ 0.06,\ 0.13,\ 0.15,\ 0.2,\ 0.31,\ 0.46,\ 0.58,\ 0.69,\ 0.78,\ 0.81,\ 0.95,\ 0.97,\ 0.98)'$ характеризует долю семей, имеющих телевизор. Оцените параметры логистического тренда.

1.8. По ряду x из упражнения 3 рассчитайте ранговый коэффициент корреляции Спирмена и сделайте вывод о наличии тенденции.

Таблица 11.1

Расходы на рекламу	10	100	50	200	20	70	100	50	300	80
Объем продаж	1011	1030	1193	1149	1398	1148	1141	1223	1151	1576

1.9. Дан ряд:

$x = (10, 9, 12, 11, 14, 12, 17, 14, 19, 16, 18, 21, 20, 23, 22, 26, 23, 28, 25, 30)'$.

а) Оцените модель линейного тренда. Остатки, полученные после исключения тренда, проверьте на стационарность с использованием рангового коэффициента корреляции Спирмена.

б) Рассчитайте для остатков статистику Бартлетта, разбив ряд на 4 интервала по 5 наблюдений. Проверьте однородность выборки по дисперсии.

в) Рассчитайте для остатков статистику Голдфельда—Квандта, исключив 6 наблюдений из середины ряда. Проверьте однородность выборки по дисперсии. Сравните с выводами, полученными на основе критерия Бартлетта.

1.10. По данным таблицы 11.1 оцените модель распределенного лага зависимости объема продаж от расходов на рекламу с лагом 2.

Определите величину максимального лага в модели распределенного лага, используя различные критерии (t-статистики, F-статистики, информационные критерии).

Задачи

1. Перечислить статистики, использующиеся в расчете коэффициента автокорреляции, и записать их формулы.

2. Чем различается расчет коэффициента автокорреляции для стационарных и нестационарных процессов? Записать формулы.

3. Вычислить значение коэффициента корреляции для двух рядов:
$x = (1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ \ldots)$ и $y = (2,\ 4,\ 6,\ 8,\ 10,\ 12,\ 14,\ \ldots)$.

4. Посчитать коэффициент автокорреляции первого порядка для ряда
$x = (2,\ 4,\ 6,\ 8)'$.

5. Есть ли разница между автокорреляционной функцией и трендом автокорреляции?

6. Записать уравнения экспоненциального и полиномиального трендов и привести формулы для оценивания их параметров.
7. Записать формулу для оценки темпа прироста экспоненциального тренда.
8. Привести формулу логистической кривой и указать особенности оценивания ее параметров.
9. Оценить параметры линейного тренда для временного ряда $x = (1,\ 2,\ 5,\ 6)$ и записать формулу доверительного интервала для прогноза на 1 шаг вперед.
10. Дан временной ряд: $x = (1,\ 0.5,\ 2,\ 5,\ 1.5)$. Проверить его на наличие тренда среднего.
11. Пусть $\mathbf{L}$ — лаговый оператор. Представьте в виде степенного ряда следующие выражения:
    а) $\dfrac{2}{1-0.8\mathbf{L}}$; б) $\dfrac{-1,5}{1-0.9\mathbf{L}}$; в) $\dfrac{2.8}{1+0.4\mathbf{L}}$; г) $\dfrac{-3}{1+0.5\mathbf{L}}$.

Рекомендуемая литература

1. **Айвазян С.А.** Основы эконометрики. Т.2. — М.: «Юнити», 2001.
2. **Андерсон Т.** Статистический анализ временных рядов. — М.: «Мир», 1976. (Гл. 1, 3, 7).
3. **Бокс Дж., Дженкинс Г.** Анализ временных рядов. Прогноз и управление. Вып. 1. — М.: «Мир», 1974. (Гл. 1).
4. **Кендалл М. Дж., Стьюарт А.** Многомерный статистический анализ и временные ряды. — М.: «Наука», 1976. (Гл. 45–47).

   Маленво Э. Статистические методы эконометрии. Вып. 2. — М.: «Статистика», 1976. (Гл. 12).
5. **Магнус Я.Р., Катышев П.К., Пересецкий А.А.** Эконометрика — начальный курс. — М.: «Дело», 2000. (Гл. 12).
6. **Enders Walter.** Applied Econometric Time Series. — Iowa State University, 1995. (Ch. 1).
7. **Mills Terence C.** Time Series Techniques for Economists, Cambridge University Press, 1990 (Ch. 5).
8. **Wooldridge Jeffrey M.** Introductory Econometrics: A Modern Approach, 2nd ed., Thomson, 2003 (Ch. 10).

Глава 12

Сглаживание временного ряда

12.1. Метод скользящих средних

Одним из альтернативных по отношению к функциональному описанию тренда вариантов сглаживания временного ряда является **метод скользящих** или, как еще говорят, **подвижных средних**.

Суть метода заключается в замене исходного временного ряда последовательностью средних, вычисляемых на отрезке, который перемещается вдоль временного ряда, как бы скользит по нему. Задается длина отрезка скольжения $(2m+1)$ по временной оси, т.е. берется нечетное число наблюдений. Подбирается полином

$$\tau_t = \sum_{k=0}^{p} a_k t^k \qquad (12.1)$$

к группе первых $(2m+1)$ членов ряда, и этот полином используется для определения значения тренда в средней $(m+1)$-й точке группы. Затем производится сдвиг на один уровень ряда вперед и подбирается полином того же порядка к группе точек, состоящей из 2-го, 3-го , ..., $(2m+2)$-го наблюдения. Находится значение тренда в $(m+2)$-й точке и т.д. тем же способом вдоль всего ряда до последней группы из $(2m+1)$ наблюдения. В действительности нет необходимости строить полином для каждого отрезка. Как будет показано, эта процедура эквивалентна нахождению линейной комбинации уровней временного ряда с коэффициентами,

которые могут быть определены раз и навсегда и зависят только от длины отрезка скольжения и степени полинома.

Для определения коэффициентов $a_0, a_1, \ldots, a_p$ полинома (12.1) с помощью МНК по первым $(2m+1)$ точкам минимизируется функционал:

$$\varepsilon = \sum_{t=-m}^{m} (x_t - a_0 - a_1 t - \ldots - a_p t^p)^2. \qquad (12.2)$$

Заметим, что t принимает условные значения от $-m$ до m. Это весьма удобный прием, существенно упрощающий расчеты. Дифференцирование функционала по $a_0, a_1, \ldots, a_p$ дает систему из $p+1$ уравнения типа:

$$a_0 \sum_{t=-m}^{m} t^j + a_1 \sum_{t=-m}^{m} t^{j+1} + a_2 \sum_{t=-m}^{m} t^{j+2} + \cdots + a_p \sum_{t=-m}^{m} t^{j+p} = \sum_{t=-m}^{m} x_t t^j,$$
$$j = 0, 1, \ldots, p. \qquad (12.3)$$

Решение этой системы уравнений относительно неизвестных параметров $a_0,\ a_1, \ldots, a_p\ (i = 0, 1, \ldots, 2p)$ облегчается тем, что все суммы $\sum_{t=-m}^{m} t^i$ при нечетных i равны нулю. Кроме того, т. к. полином, подобранный по $2m+1$ точкам, используется для определения значения тренда в средней точке, а в этой точке $t = 0$, то, положив в уравнении (12.1) $t = 0$, получаем значение тренда, равное a_0. Стало быть, задача сглаживания временного ряда сводится к поиску a_0.

Система нормальных уравнений (12.3), которую нужно разрешить относительно a_0, разбивается на две подсистемы: одну — содержащую коэффициенты с четными индексами $a_0, a_2, a_4, \ldots$, другую — включающую коэффициенты с нечетными индексами $a_1, a_3, a_5, \ldots$. Решение системы относительно a_0 зависит от численных значений $\sum_{t=-m}^{m} t^i$ и линейных функций от x типа $\sum_{t=-m}^{m} x_t t^j$.

В итоге, значением тренда в центральной точке отрезка будет средняя арифметическая, взвешенная из значений временного ряда от x_{-m} до x_m с весовыми коэффициентами β_t, которые зависят от значений m и p:

$$a_0 = \sum_{t=-m}^{m} \beta_t x_t.$$

Указанная формула применяется для всех последующих отрезков скольжения, с вычислением значений тренда в их средних точках.

Продемонстрируем рассматриваемый метод на примере полинома второй степени и длины отрезка скольжения, равной пяти точкам. Здесь надо свести к минимуму сумму:

$$\varepsilon = \sum_{t=-2}^{2} (x_t - a_0 - a_1 t - a_2 t^2)^2.$$

Получается система уравнений:

$$\begin{cases} \sum\limits_{t=-2}^{2} a_0 + a_1 \sum\limits_{t=-2}^{2} t + a_2 \sum\limits_{t=-2}^{2} t^2 = \sum\limits_{t=-2}^{2} x_t, \\ a_0 \sum\limits_{t=-2}^{2} t + a_1 \sum\limits_{t=-2}^{2} t^2 + a_2 \sum\limits_{t=-2}^{2} t^3 = \sum\limits_{t=-2}^{2} x_t t, \\ a_0 \sum\limits_{t=-2}^{2} t^2 + a_1 \sum\limits_{t=-2}^{2} t^3 + a_2 \sum\limits_{t=-2}^{2} t^4 = \sum\limits_{t=-2}^{2} x_t t^2. \end{cases}$$

Для конкретных значений сумм при a_p система уравнений приобретает вид:

$$\begin{cases} 5a_0 + 10a_2 = \sum\limits_{t=-2}^{2} x_t, \\ 10a_1 = \sum\limits_{t=-2}^{2} x_t t, \\ 10a_0 + 34a_2 = \sum\limits_{t=-2}^{2} x_t t^2. \end{cases}$$

Решение этой системы относительно a_0 дает следующий результат:

$$a_0 = \frac{1}{35}\left(17 \sum_{t=-2}^{2} x_t - 5 \sum_{t=-2}^{2} x_t t^2\right) = \frac{1}{35}\left(-3x_{-2} + 12x_{-1} + 17x_0 + 12x_1 - 3x_2\right).$$

Весовые коэффициенты для полиномов $2-5$ степени и длины отрезка скольжения от 5 до 9 представлены в таблице 12.1.

Таблица 12.1. Фрагмент таблицы Каудена для весов β_t

Длина отрезка скольжения		Степени полинома	
$2m+1$	m	$p=2, p=3$	$p=4, p=5$
5	2	$\frac{1}{35}(-3, 12, 17, 12, -3)$	
7	3	$\frac{1}{21}(-2, 3, 6, 7, 6, 3, -2)$	$\frac{1}{231}(5, -30, 75, 131, 75, -30, 5)$
9	4	$\frac{1}{231}(-21, 14, 39, 54, 59, 54, 39, 14, -21)$	$\frac{1}{429}(15, -55, 30, 135, 179, 135, 30, -55, 15)$

Метод скользящих средних в матричной форме

Введем следующие обозначения:

1. $c_j = \frac{1}{2} \sum\limits_{t=-m}^{m} x_t t^j$.

   Так как x_t и t^j известны, то c_j также известно для каждого $j = 0, 1, \ldots, p$.

2. $\omega_i = \dfrac{1}{2} \sum\limits_{t=-m}^{m} t^i, \quad i = 0, 1, \ldots, 2p$. Тогда

$$\omega_i = \begin{cases} 0, & \text{если } i \text{ — нечетно,} \\ \dfrac{2m+1}{2}, & \text{если } i = 0, \\ 1^i + 2^i + \ldots + m^i, & \text{если } i \text{ — четно.} \end{cases}$$

В таких обозначениях система (12.3) принимает вид:

$$\begin{pmatrix} \omega_0 & \omega_1 & \cdots & \omega_p \\ \omega_1 & \omega_2 & \cdots & \omega_{p+1} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_p & \omega_{p+1} & \cdots & \omega_{2p} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_p \end{pmatrix} = \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_p \end{pmatrix}.$$

В краткой записи эта система выглядит как

$$Ma = c,$$

где матрица M — известна, кроме того, ее элементы с нечетными индексами равны нулю, вектор c также известен.

Из полученной системы следует

$$a = M^{-1}c.$$

Теперь можно использовать формулу Крамера для нахождения a_k:

$$a_k = \frac{\det M_{k+1}}{\det M},$$

где матрица M_{k+1} получается из матрицы M заменой $(k+1)$-го столбца вектором c.

Таким образом,

$$a = \left(\frac{\det M_1}{\det M}, \frac{\det M_2}{\det M}, \ldots, \frac{\det M_{p+1}}{\det M} \right)'.$$

Рассмотрим частный случай, когда $m = 2$ и $p = 2$, т.е. временной ряд аппроксимируется полиномом второй степени:

$$\tau_t = a_0 + a_1 t + a_2 t^2.$$

Система уравнений, которую нужно решить относительно a_k, имеет вид:

$$a_0 \sum_{t=-2}^{2} t^j + a_1 \sum_{t=-2}^{2} t^{j+1} + a_2 \sum_{t=-2}^{2} t^{j+2} = \sum_{t=-2}^{2} x_t t^j,$$

где $x_{-2}, x_{-1}, x_0, x_1, x_2$ — известны, $j = 0, \ldots, p$. Находим ω_i:

$$\omega_i = \begin{cases} 0, & \text{если } i \text{ — нечетно,} \\ \dfrac{5}{2}, & \text{если } i = 0, \\ 1^i + 2^i, & \text{если } i \text{ — четно.} \end{cases}$$

Тогда

$$M = \begin{pmatrix} \omega_0 & \omega_1 & \omega_2 \\ \omega_1 & \omega_2 & \omega_3 \\ \omega_2 & \omega_3 & \omega_4 \end{pmatrix} = \begin{pmatrix} 5/2 & 0 & 5 \\ 0 & 5 & 0 \\ 5 & 0 & 17 \end{pmatrix}.$$

Находим определители:

$$\det M = \frac{25 \cdot 7}{2}$$

$$\det M_1 = \begin{vmatrix} c_0 & 0 & 5 \\ c_1 & 5 & 0 \\ c_2 & 0 & 17 \end{vmatrix} = 5(17c_0 - 5c_2) = \frac{5}{2}\left(17 \sum_{t=-2}^{2} x_t - 5 \sum_{t=-2}^{2} x_t t^2 \right),$$

$$\det M_2 = \begin{vmatrix} 5/2 & c_0 & 5 \\ 0 & c_1 & 0 \\ 5 & c_2 & 17 \end{vmatrix} = \frac{35}{2} c_1 = \frac{35}{4} \sum_{t=-2}^{2} x_t t,$$

$$\det M_3 = \begin{vmatrix} 5/2 & 0 & c_0 \\ 0 & 5 & c_1 \\ 5 & 0 & c_2 \end{vmatrix} = 25\left(\frac{c_2}{2} - c_0\right) = \frac{25}{2}\left(\frac{1}{2}\sum_{t=-2}^{2} x_t t^2 - \sum_{t=-2}^{2} x_t\right).$$

Отсюда:

$$a_0 = \frac{\det M_1}{\det M} = \frac{1}{35}\left(17\sum_{t=-2}^{2} x_t - 5\sum_{t=-2}^{2} x_t t^2\right),$$
$$a_1 = \frac{\det M_2}{\det M} = \frac{1}{10}\sum_{t=-2}^{2} x_t t,$$
$$a_2 = \frac{\det M_3}{\det M} = \frac{1}{14}\sum_{t=-2}^{2} x_t t^2 - \frac{1}{7}\sum_{t=-2}^{2} x_t.$$

Таким образом,

$$a_0 = -\frac{3}{35}x_{-2} + \frac{12}{35}x_{-1} + \frac{17}{35}x_0 + \frac{12}{35}x_1 - \frac{3}{35}x_2,$$
$$a_1 = -0,2x_{-2} - 0,1x_{-1} + 0,1x_1 + 0,2x_2,$$
$$a_2 = \frac{1}{7}x_{-2} - \frac{1}{14}x_{-1} - \frac{1}{7}x_0 - \frac{1}{14}x_1 + \frac{1}{7}x_2,$$

и каждый из этих коэффициентов получается как взвешенная средняя из уровней временного ряда, входящих в отрезок.

Оценки параметров $a_1, a_2, \ldots, a_p$ необходимы для вычисления значений тренда в первых m и последних m точках временного ряда, поскольку рассмотренный способ сглаживания ряда через a_0 сделать это не позволяет.

Размерность матрицы M определяется степенью полинома: $(p+1)\times(p+1)$, пределы суммирования во всех формулах задаются длиной отрезка скольжения. Следовательно, для выбранных значений p и m можно получить общее решение в виде вектора $(a_0, a_1, \ldots, a_p)'$.

Свойства скользящих средних

1. Сумма весов β_t в формуле $a_0 = \sum_{t=-m}^{m} \beta_t x_t$ равна единице.

Действительно, пусть все значения временного ряда равны одной и той же константе c. Тогда $\sum_{t=-m}^{m} \beta_t x_t = c\sum_{t=-m}^{m} \beta_t$ должна быть равна этой константе c, а это возможно только в том случае, если $\sum_{t=-m}^{m} \beta_t = 1$.

2. Веса симметричны относительно нулевого значения t, т.е. $\beta_t = \beta_{-t}$

Это следует из того, что весовые коэффициенты при каждом x_t зависят от t^j, а j принимает только четные значения.

3. Для полиномов четного порядка $p = 2k$ формулы расчета a_0 будут теми же самыми, что и для полиномов нечетного порядка $p = 2k + 1$.

Пусть $p = 2k + 1$, тогда матрица коэффициентов системы (12.3) при неизвестных параметрах $a_0, a_1, \ldots, a_p$ будет выглядеть следующим образом:

$$\begin{pmatrix} \sum_{t=-m}^{m} t^0 & \sum_{t=-m}^{m} t & \cdots & \sum_{t=-m}^{m} t^{2k} & \sum_{t=-m}^{m} t^{2k+1} \\ \sum_{t=-m}^{m} t & \sum_{t=-m}^{m} t^2 & \cdots & \sum_{t=-m}^{m} t^{2k+1} & \sum_{t=-m}^{m} t^{2k+2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \sum_{t=-m}^{m} t^{2k} & \sum_{t=-m}^{m} t^{2k+1} & \cdots & \sum_{t=-m}^{m} t^{4k} & \sum_{t=-m}^{m} t^{4k+1} \\ \sum_{t=-m}^{m} t^{2k+1} & \sum_{t=-m}^{m} t^{2k+2} & \cdots & \sum_{t=-m}^{m} t^{4k+1} & \sum_{t=-m}^{m} t^{4k+2} \end{pmatrix}.$$

Для нахождения a_0 используются уравнения с четными степенями t при a_0, следовательно, половина строк матрицы, включая последнюю, в расчетах участвовать не будет.

В этом блоке матрицы, содержащем коэффициенты при $a_0, a_2, a_4, \ldots$, последний столбец состоит из нулей, так как его элементы — суммы нечетных степеней t. Таким образом, уравнения для нахождения a_0 при нечетном значении $p = 2k + 1$ в точности совпадают с уравнениями, которые надо решить для нахождения a_0 при меньшем на единицу четном значении $p = 2k$:

$$\begin{pmatrix} \sum_{t=-m}^{m} t^0 & \sum_{t=-m}^{m} t^2 & \cdots & \sum_{t=-m}^{m} t^{2k} \\ \sum_{t=-m}^{m} t^2 & \sum_{t=-m}^{m} t^4 & \cdots & \sum_{t=-m}^{m} t^{2k+2} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{t=-m}^{m} t^{2k} & \sum_{t=-m}^{m} t^{2k+2} & \cdots & \sum_{t=-m}^{m} t^{4k} \end{pmatrix}.$$

4. Оценки параметров $a_1, \ldots, a_p$ тоже выражены в виде линейной комбинации уровней временного ряда, входящих в отрезок, но весовые коэффициенты в этих формулах в сумме равны нулю и не симметричны.

Естественным образом возникает вопрос, какой степени полином следует выбирать и какой должна быть длина отрезка скольжения. Закономерность такова: чем выше степень полинома и короче отрезок скольжения, тем ближе расчетные

значения к первоначальным данным. При этом, помимо тенденции могут воспроизводиться и случайные колебания, нарушающие ее смысл. И наоборот, чем ниже степень полинома и чем длиннее отрезок скольжения, тем более гладкой является сглаживающая кривая, тем в большей мере она отвечает свойствам тенденции, хотя ошибка аппроксимации будет при этом выше.

В принципе, если ставится задача выявления тренда, то, с учетом особенностей покомпонентного разложения временного ряда, следует ориентироваться не на минимальную остаточную дисперсию, а на стационарность остатков, получающихся после исключения тренда.

12.2. Экспоненциальное сглаживание

Кроме метода скользящей средней как способа фильтрации временного ряда известностью пользуется экспоненциальное сглаживание, в основе которого лежит расчет экспоненциальных средних.

Экспоненциальная средняя рассчитывается по рекуррентной формуле:

$$s_t = \alpha x_t + \beta s_{t-1}, \tag{12.4}$$

где s_t — значение экспоненциальной средней в момент t,

α — параметр сглаживания (вес последнего наблюдения), $0 < \alpha < 1$,

$\beta = 1 - \alpha$.

Экспоненциальную среднюю, используя рекуррентность формулы (12.4), можно выразить через значения временного ряда:

$$\begin{aligned} s_t &= \alpha x_t + \beta(\alpha x_{t-1} + \beta s_{t-2}) = \alpha x_t + \alpha\beta x_{t-1} + \beta^2 s_{t-2} = \ldots = \\ &= \alpha x_t + \alpha\beta x_{t-1} + \alpha\beta^2 x_{t-2} + \ldots + \alpha\beta^j x_{t-j} + \ldots + \alpha\beta^{t-1} x_1 + \beta^t s_0 = \\ &= \alpha \sum_{j=0}^{t-1} \beta^j x_{t-j} + \beta^t s_0, \end{aligned} \tag{12.5}$$

t — количество уровней ряда, s_0 — некоторая величина, характеризующая начальные условия для первого применения формулы (12.4) при $t = 1$. В качестве s_0 можно использовать первое значение временного ряда, т.е. x_1.

Так как $\beta < 1$, то при $t \to \infty$ величина $\beta^t \to 0$, а сумма коэффициентов $\alpha \sum_{j=0}^{t-1} \beta^j \to 1$.

Действительно,

$$\alpha \sum_{j=0}^{\infty} \beta^j = \alpha \frac{1}{1-\beta} = (1-\beta)\frac{1}{1-\beta} = 1.$$

Тогда последним слагаемым в формуле (12.5) можно пренебречь и

$$s_t = \alpha \sum_{j=0}^{\infty} \beta^j x_{t-j} = \alpha \sum_{j=0}^{\infty} (1-\alpha)^j x_{t-j}.$$

Таким образом, величина s_t оказывается взвешенной суммой всех уровней ряда, причем веса уменьшаются экспоненциально, по мере углубления в историю процесса, отсюда название — **экспоненциальная средняя**.

Несложно показать, что экспоненциальная средняя имеет то же математическое ожидание, что и исходный временной ряд, но меньшую дисперсию.

Что касается параметра сглаживания α, то чем ближе α к единице, тем менее ощутимо расхождение между сглаженным рядом и исходным. И наоборот, чем меньше α, тем в большей степени подавляются случайные колебания ряда и отчетливее вырисовывается его тенденция. Экспоненциальное сглаживание можно представить в виде фильтра, на вход которого поступают значения исходного временного рядя, а на выходе формируется экспоненциальная средняя.

Использование экспоненциальной средней в качестве инструмента выравнивания временного ряда оправдано в случае стационарных процессов с незначительным сезонным эффектом. Однако многие процессы содержат тенденцию, сочетающуюся с ярко выраженными сезонными колебаниями.

Довольно эффективный способ описания таких процессов — **адаптивные сезонные модели**, основанные на экспоненциальном сглаживании. Особенность адаптивных сезонных моделей заключается в том, что по мере поступления новой информации происходит корректировка параметров модели, их приспособление, адаптация к изменяющимся во времени условиям развития процесса.

Выделяют два вида моделей, которые можно изобразить схематично:

1. Модель с аддитивным сезонным эффектом, предложенная Тейлом и Вейджем (Theil H., Wage S.):

$$x_t = f_t + g_t + \varepsilon_t, \tag{12.6}$$

где f_t отражает тенденцию развития процесса, $g_t, g_{t-1}, \ldots, g_{t-k+1}$ — аддитивные коэффициенты сезонности; k — количество опорных временных интервалов (фаз) в полном сезонном цикле; ε_t — белый шум.

2. Модель с мультипликативным сезонным эффектом, разработанная Уинтерсом (Winters P.R.):

$$x_t = f_t \cdot m_t \cdot \varepsilon_t, \tag{12.7}$$

где $m_t, m_{t-1}, \ldots, m_{t-k+1}$ — мультипликативные коэффициенты сезонности.

В принципе, эта модель после логарифмирования может быть преобразована в модель с аддитивным сезонным эффектом.

Мультипликативные модели целесообразно использовать в тех ситуациях, когда наряду, допустим, с повышением среднего уровня увеличивается амплитуда колебаний, обусловленная сезонным фактором. Если в аддитивных моделях индексы сезонности измеряются в абсолютных величинах, то в мультипликативных — в относительных.

И в том, и в другом случае обновление параметров модели производится по схеме экспоненциального сглаживания. Оба варианта допускают как наличие тенденции (линейной или экспоненциальной), так и ее отсутствие.

Множество комбинаций различных типов тенденций с циклическими эффектами аддитивного и мультипликативного характера можно представить в виде обобщенной формулы:

$$f_t = \alpha_f d_1 + (1 - \alpha_f) d_2,$$

где f_t — некоторый усредненный уровень временного ряда в момент t после устранения сезонного эффекта,

α_f — параметр сглаживания, $0 < \alpha_f < 1$,

d_1 и d_2 — характеристики модели.

$$d_1 = \begin{cases} x_t, & \text{— если сезонный эффект отсутствует,} \\ x_t - g_{t-k}, & \text{— в случае аддитивного сезонного эффекта,} \\ \dfrac{x_t}{m_{t-k}}, & \text{— в случае мультипликативного сезонного эффекта.} \end{cases}$$

Таким образом, d_1 представляет собой текущую оценку процесса x_t, очищенную от сезонных колебаний с помощью коэффициентов сезонности g_{t-k} или m_{t-k}, рассчитанных для аналогичной фазы предшествующего цикла.

$$d_2 = \begin{cases} f_{t-1}, & \text{— при отсутствии тенденции,} \\ f_{t-1} + c_{t-1}, & \text{— в случае аддитивного роста,} \\ f_{t-1} \cdot r_{t-1}, & \text{— в случае экспоненциального роста.} \end{cases}$$

В этой формуле c_{t-1} — абсолютный прирост, характеризующий изменение среднего уровня процесса, или аддитивный коэффициент роста, r_{t-1} — коэффициент экспоненциального роста.

Например, для модели с аддитивным ростом и мультипликативным сезонным эффектом подойдет график, изображенный на рисунке 12.1а, а для модели с экспоненциальным ростом и аддитивным сезонным эффектом — график на рисунке 12.1б.

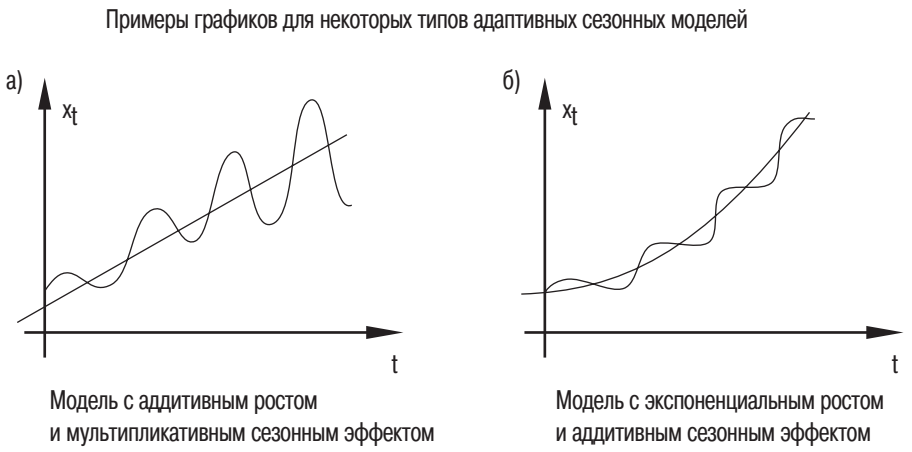

Рис. 12.1. Графики некоторых типов временных рядов

Адаптация всех перечисленных параметров осуществляется с помощью экспоненциального сглаживания:

$$
\begin{aligned}
g_t &= \alpha_g(x_t - f_t) + (1 - \alpha_g)g_{t-k},\\
m_t &= \alpha_m\frac{x_t}{f_t} + (1 - \alpha_m)m_{t-k},\\
c_t &= \alpha_c(f_t - f_{t-1}) + (1 - \alpha_c)c_{t-1},\\
r_t &= \alpha_r\frac{f_t}{f_{t-1}} + (1 - \alpha_r)r_{t-1},
\end{aligned}
$$

где $0 < \alpha_g, \alpha_m, \alpha_c, \alpha_r < 1$.

Первые две формулы представляют собой линейную комбинацию текущей оценки коэффициента сезонности, полученной путем устранения из исходного уровня процесса значения тренда ($x_t - f_t$ и x_t/f_t), и оценки этого параметра на аналогичной фазе предшествующего цикла (g_{t-k} и m_{t-k}). Аналогично, две последние формулы являются взвешенной суммой текущей оценки коэффициента роста (соответственно, аддитивного $f_t - f_{t-1}$ и экспоненциального f_t/f_{t-1}) и предыдущей его оценки (c_{t-1} и r_{t-1}).

Очевидно, что в случае отсутствия тенденции и сезонного эффекта получается простая экспоненциальная средняя:

$$f_t = \alpha_f x_t + (1 - \alpha_f)f_{t-1}.$$

Рассмотрим для иллюстрации модель Уинтерса с аддитивным ростом и мультипликативным сезонным эффектом:

$$\begin{aligned} f_t &= \alpha_f \frac{x_t}{m_{t-k}} + (1-\alpha_f)(f_{t-1} + c_{t-1}), \\ m_t &= \alpha_m \frac{x_t}{f_t} + (1-\alpha_m) m_{t-k}, \\ c_t &= \alpha_c (f_t - f_{t-1}) + (1-\alpha_c) c_{t-1}. \end{aligned} \tag{12.8}$$

Расчетные значения исследуемого показателя на каждом шаге, после обновления параметров f_t, m_t и c_t, получаются как произведение $f_t \cdot m_t$.

Прежде чем воспользоваться полной схемой экспоненциального сглаживания (12.8), а сделать это можно начиная с момента $t = k+1$, необходимо получить начальные, отправные значения перечисленных параметров.

Для этого с помощью МНК можно оценить коэффициенты f_1 и c_1 регрессии:

$$x_t = f_1 + c_1 t + \varepsilon_t,$$

и на первом сезонном цикле (для $t = 1, \ldots, k$) адаптацию параметров произвести по усеченному варианту:

$$\begin{aligned} f_t &= \alpha_f x_t + (1-\alpha_f) f_{t-1}, \\ m_t &= \frac{x_t}{f_t}, \quad t = 1, \ldots, k, \\ c_t &= \alpha_c (f_t - f_{t-1}) + (1-\alpha_c) c_{t-1}, \\ g_t &= x_t - f_t. \end{aligned}$$

Задача оптимизации модели сводится к поиску наилучших значений параметров α_f, α_m, α_c, выбор которых определяется целями исследования и характером моделируемого процесса. Уинтерс предлагает находить оптимальные уровни этих коэффициентов экспериментальным путем, с помощью сетки значений α_f, α_m, α_c (например, $(0,1; 0,1; 0,1)$, $(0,1; 0,1; 0,2)$, $\ldots$). В качестве критерия сравнения вариантов рекомендуется стандартное отклонение ошибки.

12.3. Упражнения и задачи

Упражнение 1

1.1. Сгенерируйте 20 рядов по 100 наблюдений на основе полиномиального тренда $\tau_t = 5 + 4t - 0,07t^2 + 0.0005t^3$ с добавлением белого шума с нормальным распределением и дисперсией 20 .

Таблица 12.2. Производство природного газа в СССР (миллиардов кубических футов)

	январь	февраль	март	апрель	май	июнь	июль	август	сентябрь	октябрь	ноябрь	декабрь
1971	653.1	589.5	653.1	610.7	610.7	583.2	600.1	614.2	600.1	642.5	642.5	670.7
1972	670.8	649.5	695.4	664.5	638.9	621.3	620.7	619.4	624.8	653.1	663.6	706.0
1973	720.1	656.6	734.2	691.9	688.3	688.4	691.2	701.2	653.1	673.2	720.1	673.2
1974	720.1	709.5	776.6	737.7	741.3	723.7	724.6	758.9	760.0	808.4	811.2	882.5
1975	864.9	871.9	868.4	861.2	864.8	833.3	833.1	829.6	829.6	840.2	900.1	953.1
1976	953.1	914.3	967.2	921.3	917.8	916.8	924.9	924.9	917.8	988.4	974.3	1009.6
1977	1048.4	960.2	960.2	1048.4	998.9	956.6	984.9	995.5	999.0	1175.5	1180.0	1190.0
1978	1129.6	1129.4	1126.1	1076.7	1080.2	1034.3	1062.5	1064.7	1023.7	1147.2	1136.7	1196.8
1979	1230.0	1220.0	1220.0	1175.5	1182.5	1140.2	1157.8	1161.4	1164.9	1249.6	1250.6	1306.1
1980	1309.6	1232.0	1306.1	1246.1	1256.7	1200.2	1246.1	1260.2	1270.8	1270.0	1323.8	1376.7
1981	1419.1	1299.0	1420.0	1345.0	1313.0	1271.0	1270.0	1334.0	1334.0	1430.0	1430.0	1460.0
1982	1504.0	1380.0	1528.5	1436.7	1457.9	1412.0	1419.1	1436.7	1447.3	1546.1	1528.5	1623.8
1983	1627.4	1486.1	1652.0	1528.5	1570.8	1517.9	1514.4	1539.1	1482.6	1648.5	1648.5	1747.3
1984	1747.4	1648.5	1757.9	1680.3	1697.9	1623.8	1669.7	1697.9	1694.4	1821.5	1803.8	1870.9
1985	1930.9	1775.6	1941.5	1853.2	1892.1	1765.0	1825.0	1846.2	1870.9	1990.9	1962.7	2047.4
1986	2075.6	1895.6	2118.0	1983.9	2005.0	1906.2	1959.2	1969.7	1976.8	2103.9	2089.8	2188.6
1987	2221.3	2030.6	2210.7	2083.6	2118.9	2012.9	2048.3	2048.3	2083.6	2223.9	2259.2	2330.8
1988	2369.6	2221.3	2366.1	2224.8	2275.0	2146.6	2118.9	2189.5	2189.5	2357.0	2394.8	2447.7
1989	2510.0	2300.0	2391.0	2333.0	2336.0	2187.7	2208.0	2279.0	2200.0	2500.0	2484.0	2495.0
1990	2630.0	2400.0	2420.0	2391.0	2430.0	2250.0	2340.0	2340.0	2250.0	2500.0	2450.0	2460.0

а) Проведите сглаживание сгенерированных рядов с помощью полинома первой степени с длиной отрезка скольжения 5 и 9.

б) Выполните то же задание, используя полином третьей степени.

в) Найдите отклонения исходных рядов от сглаженных рядов, полученных в пунктах (а) и (б). По каждому ряду отклонений вычислите среднеквадратическую ошибку. Сделайте вывод о том, какой метод дает наименьшую среднеквадратическую ошибку.

1.2. Имеются данные о производстве природного газа в СССР (табл. 12.2).

а) Постройте графики ряда и логарифмов этого ряда. Чем они различаются? Выделите основные компоненты временного ряда. Какой характер носит сезонность: аддитивный или мультипликативный? Сделайте вывод о целесообразности перехода к логарифмам.

б) Примените к исходному ряду метод экспоненциального сглаживания, подобрав параметр сглаживания.

в) Проведите сглаживание временного ряда с использованием адаптивной сезонной модели.

Задачи

1. Сгладить временной ряд $x = (3,\ 4,\ 5,\ 6,\ 7,\ 11)$, используя полином первого порядка с длиной отрезка скольжения, равной трем.

2. Записать формулу расчета вектора коэффициентов для полинома третьей степени с помощью метода скользящей средней в матричной форме с расшифровкой обозначений.

3. В чем специфика аппроксимации первых m и последних m точек временного ряда при использовании метода скользящих средних?

4. Найти параметры адаптивной сезонной модели для временного ряда
   $x = (1,\ 2,\ 3,\ 4,\ 1,\ 2,\ 3,\ 4,\ 1,\ 2,\ 3,\ 4,\ \ldots)$.

5. Изобразить график временного ряда с аддитивным ростом и мультипликативным сезонным эффектом.

6. Изобразить график временного ряда с экспоненциальным ростом и аддитивным сезонным эффектом.

7. Записать модель с экспоненциальным ростом и мультипликативным сезонным эффектом, а также формулу прогноза на 5 шагов вперед.

Рекомендуемая литература

1. **Андерсон Т.** Статистический анализ временных рядов. — М.: «Мир», 1976. (Гл. 3).

2. **Лукашин Ю.П.** Адаптивные методы краткосрочного прогнозирования. — М.: «Статистика», 1979. (Гл. 1, 2).

3. **Кендалл М. Дж., Стьюарт А.** Многомерный статистический анализ и временные ряды. — М.: «Наука», 1976. (Гл. 46).

4. **Маленво Э.** Статистические методы эконометрии. Вып. 2. — М.: «Статистика», 1976. (Гл. 11, 12).

5. **Mills Terence C.** Time Series Techniques for Economists, Cambridge University Press, 1990 (Ch. 9).

Глава 13

Спектральный и гармонический анализ

13.1. Ортогональность тригонометрических функций и преобразование Фурье временного ряда

Как известно, тригонометрические функции $\cos t$ и $\sin t$ являются периодическими с периодом 2π:

$$\cos(t + 2\pi) = \cos t,\ \sin(t + 2\pi) = \sin t.$$

Функции $\cos(\lambda t - \theta)$ и $\sin(\lambda t - \theta)$ периодичны с **периодом** $2\pi/\lambda$. Действительно,

$$\cos(\lambda t - \theta) = \cos(\lambda t + 2\pi - \theta) = \cos\left(\lambda(t + 2\pi/\lambda) - \theta\right),$$
$$\sin(\lambda t - \theta) = \sin(\lambda t + 2\pi - \theta) = \sin(\lambda(t + 2\pi/\lambda) - \theta).$$

Величина $\lambda/2\pi$, обратная периоду, называется **линейной частотой**, λ называют **угловой частотой**. Линейная частота равна числу периодов (не обязательно целому), содержащемуся в единичном интервале, то есть именно такое число раз функция повторяет свои значения в промежутке $[0, 1]$.

Рассмотрим функцию:

$$R\cos(\lambda t - \theta) = R(\cos\lambda t\cos\theta + \sin\lambda t\sin\theta) = \alpha\cos(\lambda t) + \beta\sin(\lambda t),$$

где $\alpha = R\cos\theta$, $\beta = R\sin\theta$ или, что эквивалентно, $R = \sqrt{\alpha^2 + \beta^2}$, $\operatorname{tg}\theta = {}^{\beta}/_{\alpha}$.

Коэффициент R, являющийся максимумом функции $R\cos(\lambda t - \theta)$ называется **амплитудой** этой функции, а угол θ называется **фазой**.

Особенность тригонометрических функций заключается в том, что на определенном диапазоне частот они обладают свойством ортогональности.

Две функции $\varphi(t)$ и $\psi(t)$, определенные на конечном множестве $\{1, \ldots, T\}$, называются **ортогональными**, если их скалярное произведение, определенное как сумма произведений значений $\varphi(t)$ и $\psi(t)$ в этих точках, равно нулю:

$$\sum_{t=1}^{T} \varphi(t) \cdot \psi(t) = 0.$$

Система T тригонометрических функций в точках $t \in \{1, \ldots, T\}$

$$\begin{cases} c_{jt} = \cos\dfrac{2\pi j}{T}t, \quad j = 0,\ 1,\ \ldots,\ \left[\dfrac{T}{2}\right], \\ s_{jt} = \sin\dfrac{2\pi j}{T}t, \quad j = 1,\ \ldots,\ \left[\dfrac{T-1}{2}\right] \end{cases} \tag{13.1}$$

ортогональна, т.е. скалярное произведение векторов

$$(c_j, c_k) = \sum_{t=1}^{T} c_{jt}c_{kt} = 0,\ j \neq k,\ 0 \leqslant j,\ k \leqslant \left[\frac{T}{2}\right], \tag{13.2}$$

$$(s_j, s_k) = \sum_{t=1}^{T} s_{jt}s_{kt} = 0,\ j \neq k,\ 0 < j,\ k \leqslant \left[\frac{T-1}{2}\right], \tag{13.3}$$

$$(c_j, s_k) = \sum_{t=1}^{T} c_{jt}s_{kt} = 0,\ 0 \leqslant j \leqslant \left[\frac{T}{2}\right],\ 0 < k \leqslant \left[\frac{T-1}{2}\right], \tag{13.4}$$

где операция $[\ldots]$ — это выделение целой части числа.

Для доказательства этого утверждения полезны следующие равенства

$$\sum_{t=1}^{T} \cos\frac{2\pi j}{T}t = \begin{cases} 0, \text{ при } j \neq 0, \\ T, \text{ при } j = 0,\ T, \end{cases} \tag{13.5}$$

$$\sum_{t=1}^{T} \sin\frac{2\pi j}{T}t = 0, \tag{13.6}$$

истинность которых легко установить, выразив тригонометрические функции через показательные с использованием формул Эйлера:

$$e^{\pm i\gamma} = \cos\gamma \pm i\sin\gamma, \tag{13.7}$$

$$\cos\gamma = \frac{1}{2}(e^{i\gamma} + e^{-i\gamma}), \tag{13.8}$$

$$\sin\gamma = \frac{1}{2i}(e^{i\gamma} - e^{-i\gamma}). \tag{13.9}$$

Итак, при $j \neq 0$

$$\sum_{t=1}^{T} \cos\frac{2\pi j}{T}t = \frac{1}{2}\sum_{t=1}^{T}\left(e^{i\frac{2\pi j}{T}t} + e^{-i\frac{2\pi j}{T}t}\right) = \\ = \frac{1}{2}e^{i\frac{2\pi j}{T}}\frac{1-e^{i2\pi j}}{1-e^{i\frac{2\pi j}{T}}} + \frac{1}{2}e^{-i\frac{2\pi j}{T}}\frac{1-e^{-i2\pi j}}{1-e^{-i\frac{2\pi j}{T}}} = 0,$$

где предпоследнее равенство получено из формулы суммы геометрической прогрессии, а последнее — из формулы (13.7), т.к.

$$e^{\pm i2\pi j} = \cos(2\pi j) \pm i\sin(2\pi j) = 1.$$

Очевидно, что при $j = 0, T$ $\sum_{t=1}^{T}\cos\frac{2\pi j}{T}t = T$.

Равенство (13.6) доказывается аналогично. При доказательстве соотношений (13.2−13.4) используются утверждения (13.5, 13.6).

Таким образом,

$$(c_j, c_k) = \sum_{t=1}^{T}\cos\frac{2\pi j}{T}t\cdot\cos\frac{2\pi k}{T}t = \frac{1}{2}\sum_{t=1}^{T}\cos\frac{2\pi(j-k)}{T}t + \frac{1}{2}\sum_{t=1}^{T}\cos\frac{2\pi(j+k)}{T}t = \\ = \begin{cases} 0, & j \neq k, \quad 0 \leqslant j,\, k \leqslant \left[\dfrac{T}{2}\right], \\ \dfrac{T}{2}, & j = k, \quad 0 < j,\, k < \dfrac{T}{2}, \\ T, & j = k = 0, \dfrac{T}{2} \quad (\text{для четных } T). \end{cases} \tag{13.10}$$

$$(s_j, s_k) = \sum_{t=1}^{T}\sin\frac{2\pi j}{T}t\cdot\sin\frac{2\pi k}{T}t = \frac{1}{2}\sum_{t=1}^{T}\cos\frac{2\pi(j-k)}{T}t - \frac{1}{2}\sum_{t=1}^{T}\cos\frac{2\pi(j+k)}{T}t = \\ = \begin{cases} 0, & j \neq k,\ 0 < j,\, k \leqslant \left[\dfrac{T-1}{2}\right], \\ \dfrac{T}{2}, & j = k,\ 0 < j,\, k \leqslant \left[\dfrac{T-1}{2}\right]. \end{cases} \tag{13.11}$$

$$(c_j, s_k) = \sum_{t=1}^{T} \cos\frac{2\pi j}{T}t \cdot \sin\frac{2\pi k}{T}t =$$

$$= \frac{1}{2}\sum_{t=1}^{T} \sin\frac{2\pi(j+k)}{T}t + \frac{1}{2}\sum_{t=1}^{T} \sin\frac{2\pi(j-k)}{T}t = 0. \quad (13.12)$$

Мы доказали выполнение (13.2−13.4) для указанного набора функций, получив одновременно некоторые количественные их характеристики. Таким образом, функции $\cos\frac{2\pi j}{T}t$ и $\sin\frac{2\pi j}{T}t$ образуют ортогональный базис и всякую функцию, в том числе и временной ряд $\{x_t\}$, определенный на множестве $\{1, \ldots, T\}$, можно разложить по этому базису, т.е. **представить в виде конечного ряда Фурье**:

$$x_t = \sum_{j=0}^{[T/2]} \left(\alpha_j \cos\frac{2\pi j}{T}t + \beta_j \sin\frac{2\pi j}{T}t\right), \quad (13.13)$$

или, вспоминая (13.1), кратко

$$x_t = \sum_{j=0}^{[T/2]} \left(\alpha_j c_{jt} + \beta_j s_{jt}\right),$$

где β_0 и $\beta_{[T/2]}$ при четном T отсутствуют (т.к. $\sin 0 = 0,\ \sin \pi t = 0$).

Величину $2\pi j/T = \lambda_j$ называют **частотой Фурье**, а набор скаляров α_j и β_j ($j = 0, 1, \ldots, [T/2]$) — **коэффициентами Фурье**.

Если c_{jt} и s_{jt} — элементы векторов c_j и s_j, стоящие на t-ом месте, то, переходя к векторным обозначениям, (13.13) можно переписать в матричном виде:

$$x = \begin{pmatrix} C & S \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \quad (13.14)$$

где

$$x = (x_1, \ldots, x_T)',$$
$$\alpha = (\alpha_0, \ldots, \alpha_{[T/2]})',$$
$$\beta = (\beta_1, \ldots, \beta_{[(T-1)/2]})',$$
$$C = \{c_{jt}\}, \quad j = 0, 1, \ldots, [T/2],\ t = 1, \ldots, T,$$
$$S = \{s_{jt}\}, \quad j = 1, \ldots, [(T-1)/2],\ t = 1, \ldots, T.$$

Перепишем в матричной форме свойства ортогональности тригонометрических функций, которые потребуются при вычислении коэффициентов Фурье:

$$\begin{aligned}
&c_j' s_k = 0, && \forall j,k,\\
&c_k' 1_T = 0, && \forall k \neq 0,\\
&s_k' 1_T = 0, && \forall k,\\
&c_j' c_k = s_j' s_k = 0, && j \neq k,\\
&c_k' c_k = s_k' s_k = T/2, && k \neq 0, T/2,\\
&c_0' c_0 = T,\\
&c_{T/2}' c_{T/2} = T, && \text{для четных } T,
\end{aligned} \tag{13.15}$$

где $1_T = (1, \ldots, 1)'$ — T-компонентный вектор.

Для нахождения коэффициентов Фурье скалярно умножим c_j' на вектор x и, воспользовавшись изложенными свойствами ортогональности (13.15), получим:

$$c_j' x = c_j' \begin{pmatrix} C & S \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = (c_j' c_0, \ldots, c_j' c_{[T/2]}, c_j' s_1, \ldots, c_j' s_{[(T-1)/2]}) \begin{pmatrix} \alpha \\ \beta \end{pmatrix} =$$
$$= \alpha_j c_j' c_j = \frac{T}{2} \alpha_j, \quad \text{для } j \neq 0, \frac{T}{2}.$$

Таким образом,

$$\begin{aligned}
\alpha_j &= \frac{2}{T} c_j' x = \frac{2}{T} \sum_{t=1}^{T} x_t \cos\left(\frac{2\pi j}{T} t\right), \text{ для } j \neq 0, \frac{T}{2},\\
\alpha_0 &= \frac{1}{T} c_0' x = \frac{1}{T} \sum_{t=1}^{T} x_t,\\
\alpha_{T/2} &= \frac{1}{T} c_{T/2}' x = \frac{1}{T} \sum_{t=1}^{T} (-1)^t x_t, \text{ для четных } T.
\end{aligned} \tag{13.16}$$

Аналогично находим коэффициенты β_j:

$$\beta_j = \frac{2}{T} s_j' x = \frac{2}{T} \sum_{t=1}^{T} x_t \sin\left(\frac{2\pi j}{T} t\right). \tag{13.17}$$

13.2. Теорема Парсеваля

Суть теоремы Парсеваля состоит в том, что дисперсия процесса x_t разлагается по частотам соответствующих гармоник следующим образом:

$$\mathbf{var}(x_t) = \frac{1}{2}\sum_{j=1}^{T/2-1} R_j^2 + R_{T/2}^2, \text{ для четных } T, \tag{13.18}$$

$$\mathbf{var}(x_t) = \frac{1}{2}\sum_{j=1}^{(T-1)/2} R_j^2, \text{ для нечетных } T. \tag{13.19}$$

Покажем, что это действительно так. Из (13.14) мы имеем:

$$\begin{aligned} x'x &= \begin{pmatrix} \alpha' & \beta' \end{pmatrix} \begin{pmatrix} C' \\ S' \end{pmatrix} \begin{pmatrix} C & S \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \\ &= \begin{pmatrix} \alpha' & \beta' \end{pmatrix} \begin{pmatrix} C'C & C'S \\ S'C & S'S \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \\ &= \begin{pmatrix} \alpha' & \beta' \end{pmatrix} \begin{pmatrix} \Lambda_C & 0 \\ 0 & \Lambda_S \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \\ &= \alpha'\Lambda_C\alpha + \beta'\Lambda_S\beta = \alpha_0^2 1_T' 1_T + \sum_{j=1}^{[T/2]} \alpha_j^2 c_j' c_j + \sum_{j=1}^{[(T-1)/2]} \beta_j^2 s_j' s_j, \end{aligned}$$

где Λ_c и Λ_s — диагональные матрицы. Таким образом, если T — четно, то

$$\begin{aligned} x'x &= \alpha_0^2 1_T' 1_T + \sum_{j=1}^{T/2} \alpha_j^2 c_j' c_j + \sum_{j=1}^{T/2-1} \beta_j^2 s_j' s_j = \\ &= \alpha_0^2 T + \frac{T}{2}\sum_{j=1}^{T/2-1} \alpha_j^2 + \frac{T}{2}\sum_{j=1}^{T/2-1} \beta_j^2 + \alpha_{T/2}^2 T = \alpha_0^2 T + \frac{T}{2}\sum_{j=1}^{T/2-1} (\alpha_j^2 + \beta_j^2) + \alpha_{T/2}^2 T = \\ &= \alpha_0^2 T + \frac{T}{2}\sum_{j=1}^{T/2-1} R_j^2 + \alpha_{T/2}^2 T = R_0^2 T + \frac{T}{2}\sum_{j=1}^{T/2-1} R_j^2 + R_{T/2}^2 T. \end{aligned} \tag{13.20}$$

Аналогично для нечетных T:

$$x'x = \alpha_0^2 1_T' 1_T + \sum_{j=1}^{(T-1)/2} \alpha_j^2 c_j' c_j + \sum_{j=1}^{(T-1)/2} \beta_j^2 s_j' s_j =$$

$$= \alpha_0^2 T + \frac{T}{2}\sum_{j=1}^{(T-1)/2}\alpha_j^2 + \frac{T}{2}\sum_{j=1}^{(T-1)/2}\beta_j^2 =$$

$$= \alpha_0^2 T + \frac{T}{2}\sum_{j=1}^{(T-1)/2}(\alpha_j^2+\beta_j^2) = R_0^2 T + \frac{T}{2}\sum_{j=1}^{(T-1)/2} R_j^2. \quad (13.21)$$

Разделим уравнения (13.20) и (13.21) на T и перенесем в левые части R_0^2. С учетом того, что $R_0^2 = \alpha_0^2 = \bar{x}^2$, получаем выражения для дисперсии процесса x_t.

$$\mathbf{var}(x_t) = \frac{x'x}{T} - R_0^2 = \frac{1}{2}\sum_{j=1}^{T/2-1} R_j^2 + R_{T/2}^2, \text{ для четных } T, \quad (13.22)$$

$$\mathbf{var}(x_t) = \frac{x'x}{T} - R_0^2 = \frac{1}{2}\sum_{j=1}^{(T-1)/2} R_j^2, \text{ для нечетных } T. \quad (13.23)$$

Таким образом, вклад в дисперсию процесса для $T/2$-й гармоники равен $R_{T/2}^2$, а для k-й гармоники, $k \neq T/2$, равен $\frac{1}{2}R_k^2$.

Следовательно, наряду с определением коэффициентов Фурье для k-й гармоники, можно определить долю этой же гармоники в дисперсии процесса.

13.3. Спектральный анализ

Введем понятия периодограммы и спектра.

Периодограммой называют последовательность значений $\{I_j\}$:

$$I_j = \frac{T}{2}(\alpha_j^2 + \beta_j^2), \quad j = 0, 1, \ldots, \left[\frac{T}{2}\right],$$

т.е. I_j равно квадрату амплитуды j-ой гармоники, умноженному на $\frac{T}{2}$, $I_j = \frac{T}{2}R_j^2$. Величина I_j называется **интенсивностью** на j-ой частоте.

На практике естественнее при вычислении периодограммы использовать центрированный ряд $\hat{x}_t = x_t - \bar{x}$. При этом меняется только I_0. Для центрированного ряда $\alpha_0 = 0$, поэтому $I_0 = \alpha_0^2 = 0$. Все остальные значения периодограммы не меняются, что следует из (13.5) и (13.6) — влияние константы на остальные значения обнуляется. В оставшейся части главы мы будем использовать только центрированный ряд.

В определении периодограммы принципиальным является то, что гармонические частоты $f_j = j/T$ $(j = 0, 1, \ldots, [T/2])$ изменяются дискретно, причем наиболее высокая частота составляет $0,5$ цикла за временной интервал.

Вводя понятие **спектра**, мы ослабляем это предположение и позволяем частоте изменяться непрерывно в диапазоне $0 - 0.5$ Гц (0.5 цикла в единицу времени).

Обозначим линейную частоту через f и введем следующие обозначения:

$$\alpha_f = \frac{2}{T}\sum_{t=1}^{T} \hat{x}_t \cos(2\pi f t), \quad \beta_f = \frac{2}{T}\sum_{t=1}^{T} \hat{x}_t \sin(2\pi f t)$$

и

$$R_f^2 = \alpha_f^2 + \beta_f^2.$$

Функция

$$\begin{aligned} \mathbf{p}^*(f) &= \frac{T}{2}R_f^2 = \frac{T}{2}\left(\alpha_f^2 + \beta_f^2\right) = \\ &= \frac{2}{T}\left(\left(\sum_{t=1}^{T} \hat{x}_t \cos(2\pi f t)\right)^2 + \left(\sum_{t=1}^{T} \hat{x}_t \sin(2\pi f t)\right)^2\right), \end{aligned} \quad (13.24)$$

где $0 \leqslant f \leqslant \frac{1}{2}$, называется **выборочным спектром**. Очевидно, что значения периодограммы совпадают со значениями выборочного спектра в точках f_j, то есть $\mathbf{p}^*(f_j) = I_j$.

Спектр показывает, как дисперсия стохастического процесса распределена в непрерывном диапазоне частот. Подобно периодограмме он может быть использован для обнаружения и оценки амплитуды гармонической компоненты неизвестной частоты f, скрытой в шуме.

И периодограмму, и спектр представляют для наглядности в виде графика, на оси ординат которого — интенсивность I_j или $\mathbf{p}^*(f)$, на оси абсцисс — частота $f_j = {}^{j}/_{T}$ или f, соответственно. График выборочного спектра часто называют **спектрограммой**.

Спектрограмма нужна для более наглядного изображения распределения дисперсии между отдельными частотами. Если частоте $f = {}^{k}/_{T}$ соответствует пик на спектрограмме, то в исследуемом ряду есть существенная гармоническая составляющая с периодом $1/f = {}^{T}/_{k}$.

Целью спектрального анализа является определение основных существенных гармонических составляющих случайного процесса путем разложения дисперсии процесса по различным частотам. Спектральный анализ позволяет исследовать смесь регулярных и нерегулярных спадов и подъемов, выделять существенные гармоники, получать оценку их периода и по значению спектра на соответствующих частотах судить о вкладе этих гармоник в дисперсию процесса.

Исследования показывают, что наличие непериодического тренда (тренда с бесконечным периодом) дает скачок на нулевой частоте, т.е. в начале координат спектральной функции. При наличии циклических составляющих в соответствующих частотах имеется всплеск; если ряд слишком «зазубрен», мощность спектра перемещается в высокие частоты.

Типичным для большинства экономических процессов является убывание спектральной плотности по мере того, как возрастает частота.

Процесс выделения существенных гармоник — итеративный. При изучении периодограммы выделяется две-три гармоники с максимальной интенсивностью. Находятся оценки параметров этих наиболее существенных гармоник, и они удаляются из временного ряда с соответствующими весами. Затем остатки временного ряда, получающиеся после исключения значимых гармоник, снова изучаются в той же последовательности, т.е. строится периодограмма для этих остатков, и проявляются те гармоники, которые на начальном этапе были незаметны, и т.д. Количество итераций определяется задаваемой точностью аппроксимации модели процесса, которая представляется в виде линейной комбинации основных гармоник.

Понятие спектра, являясь основополагающим в спектральном анализе, для экономистов играет важную роль еще и потому, что существует функциональная связь выборочного спектра и оценок автоковариационной функции.

13.4. Связь выборочного спектра с автоковариационной функцией

Покажем, что выборочный спектр представляет собой косинус-преобразование Фурье выборочной автоковариационной функции.

Теорема Винера—Хинчина:

$$\mathbf{p}^*(f) = 2(c_0 + 2\sum_{k=1}^{T-1} c_k \cos 2\pi fk) = 2s^2(1 + 2\sum_{k=1}^{T-1} r_k \cos 2\pi fk), \qquad (13.25)$$

где $r_k = c_k/c_0 = c_k/s^2$ — выборочные автокорреляции.

Доказательство.

Объединим коэффициенты Фурье α_f, β_f в комплексное число $d_f = \alpha_f - i\beta_f$, где i — мнимая единица. Тогда

$$\mathbf{p}^*(f) = \frac{T}{2}\left(\alpha_f^2 + \beta_f^2\right) = \frac{T}{2}\left(\alpha_f - i\beta_f\right)\left(\alpha_f + i\beta_f\right) = \frac{T}{2} d_f \overline{d}_f, \qquad (13.26)$$

где $\overline{d}_f$ комплексно сопряжено с d_f.

Используя формулы для α_f и β_f, получим:

$$d_f = \frac{2}{T}\sum_{t=1}^{T}\hat{x}_t\big(\cos 2\pi ft - i\sin 2\pi ft\big) = \frac{2}{T}\sum_{t=1}^{T}\hat{x}_t e^{-i2\pi ft}.$$

Точно так же

$$\overline{d}_f = \frac{2}{T}\sum_{t=1}^{T}\hat{x}_t e^{i2\pi ft}.$$

Подставляя полученные значения $\overline{d}_f$ и d_f в выражение (13.26), получаем:

$$\mathbf{p}^*(f) = \frac{T}{2}\cdot\frac{2}{T}\left(\sum_{t=1}^{T}\hat{x}_t e^{-i2\pi ft}\right)\cdot\frac{2}{T}\left(\sum_{t'=1}^{T}\hat{x}_t e^{i2\pi ft'}\right) = \\ = \frac{2}{T}\sum_{t=1}^{T}\sum_{t'=1}^{T}\hat{x}_t\hat{x}_{t'}e^{-i2\pi f(t-t')}. \quad (13.27)$$

Произведем замену переменных: пусть $k = t - t'$. Так как автоковариация равна

$$c_k = \frac{1}{T}\sum_{t=k+1}^{T}\hat{x}_t\hat{x}_{t-k},$$

что тождественно

$$c_k = \frac{1}{T}\sum_{t=1}^{T-k}\hat{x}_t\hat{x}_{t+k},$$

то выражение (13.27) преобразуется следующим образом:

$$\frac{2}{T}\sum_{t=1}^{T}\sum_{t'=1}^{T}\hat{x}_t\hat{x}_{t'}e^{-i2\pi f(t-t')} = \frac{2}{T}\sum_{k=-(T-1)}^{T-1}e^{-i2\pi fk}\sum_{t=k+1}^{T}\hat{x}_t\hat{x}_{t-k} = \\ = 2\sum_{k=-(T-1)}^{T-1}e^{-i2\pi fk}\cdot\frac{1}{T}\sum_{t=k+1}^{T}\hat{x}_t\hat{x}_{t-k} = 2\sum_{k=-(T-1)}^{T-1}e^{-i2\pi fk}c_k.$$

Тогда

$$\mathbf{p}^*(f) = 2\sum_{k=-(T-1)}^{T-1}e^{-i2\pi fk}c_k = 2\sum_{k=-(T-1)}^{T-1}c_k\big(\cos 2\pi fk - i\sin 2\pi fk\big) = \\ = 2\big(c_0 + 2\sum_{k=1}^{T-1}c_k\cos 2\pi fk\big),\ 0 \leqslant f \leqslant \frac{1}{2}.$$

Теперь допустим, что выборочный спектр, характеризующийся эмпирическими значениями частоты, амплитуды и фазы, вычислен для ряда из T наблюдений и мы можем неоднократно повторить этот эксперимент, соответственно собрав множество значений α_j, β_j и $\mathbf{p}^*(f)$ по повторным реализациям. Тогда среднее значение $\mathbf{p}^*(f)$ будет равно:

$$\mathbf{E}(\mathbf{p}^*(f)) = 2\big(\mathbf{E}(c_0) + 2\sum_{k=1}^{T-1}\mathbf{E}(c_k)\cos 2\pi f k\big), \tag{13.28}$$

где c_0 и c_k — эмпирические значения автоковариации. С учетом того, что $\mathbf{E}(c_k)$ при больших T стремится к теоретической автоковариации γ_k, получим, переходя к пределу, так называемую теоретическую **спектральную плотность**, или спектр мощности:

$$\mathbf{p}(f) = \lim_{T\to\infty}\mathbf{E}(\mathbf{p}^*(f)),\ 0 \leqslant f \leqslant \frac{1}{2}$$

или

$$\begin{aligned}\mathbf{p}(f) &= 2\big(\gamma_0 + 2\sum_{k=1}^{\infty}\gamma_k\cos(2\pi f k)\big) = \\ &= 2\sigma^2\big(1 + 2\sum_{k=1}^{\infty}\rho_k\cos(2\pi f k)\big),\end{aligned} \tag{13.29}$$

где $\rho_k = \gamma_k/\gamma_0 = \gamma_k/\sigma^2$ — теоретические автокорреляции.

Итак, это соотношение связывает функцию спектральной плотности с теоретическими автоковариациями.

Иногда более удобно использовать автокорреляции: разделим обе части $\mathbf{p}(f)$ на дисперсию процесса, σ^2, и получим **нормированный спектр** $\mathbf{g}(f)$:

$$\mathbf{g}(f) = 2\big(1 + 2\sum_{k=1}^{\infty}\rho_k\cos 2\pi f k\big),\ 0 \leqslant f \leqslant \frac{1}{2}.$$

Если процесс представляет собой белый шум, то, согласно приведенным формулам, $\mathbf{p}(f) = 2\sigma^2$, т.е. спектральная плотность белого шума постоянна. Этим объясняется термин «белый шум». Подобно тому, как в белом цвете смешаны в одинаковых объемах все цвета, так и белый шум содержит все частоты, и ни одна из них не выделяется.

13.5. Оценка функции спектральной плотности

На первый взгляд, выборочный спектр, определенный как

$$\mathbf{p}^*(f) = 2\sum_{k=-(T-1)}^{T-1} c_k e^{-i2\pi fk} = \tag{13.30}$$

$$= 2\big(c_0 + 2\sum_{k=1}^{T-1} c_k \cos 2\pi fk\big),\ 0 \leqslant f \leqslant \frac{1}{2}, \tag{13.31}$$

является естественной и правильной оценкой функции спектральной плотности:

$$\mathbf{p}(f) = 2\sum_{k=-\infty}^{+\infty} \gamma_k e^{-i2\pi fk} = 2\big(\gamma_0 + 2\sum_{k=1}^{\infty} \gamma_k \cos 2\pi fk\big),\ 0 \leqslant f \leqslant \frac{1}{2}. \tag{13.32}$$

Известно, что выборочная автоковариация c_k — это асимптотически несмещенная оценка параметра γ_k, так как

$$\lim_{T\to\infty} \mathbf{E}(c_k) = \gamma_k,$$

и поэтому выборочный спектр есть также асимптотически несмещенная оценка функции спектральной плотности.

Однако дисперсия оценки (выборочного спектра) не уменьшается по мере роста размера выборки. Это означает, что рассматриваемая оценка несостоятельна. В то время как график функции теоретической спектральной плотности стационарного стохастического процесса «гладкий», — график выборочного спектра, построенный на основе эмпирических данных, «неровный». Использование данной оценки в качестве оценки функции спектральной плотности может привести к ложным выводам.

Поведение выборочного спектра иллюстрируют спектрограммы на рис. 13.1 а), б), в). Гладкая жирная линия соответствует теоретической спектральной плотности случайного процесса, а неровная тонкая линия — оценке по формуле (13.24). Видно, что с увеличением длины ряда оценка не становится более точной, а только увеличивает частоту флуктуаций.

Существует два подхода к решению проблемы несостоятельности выборочного спектра как оценки теоретического спектра. Оба заключаются в том, что выборочный спектр *сглаживается*, так что за счет некоторого увеличения смещения этой оценки достигается существенное снижение дисперсии (см. рис. 13.1 г)). Один подход сглаживает оценку спектра в «частотной области», видоизменяя формулу (13.24), другой же подход видоизменяет формулу (13.25).

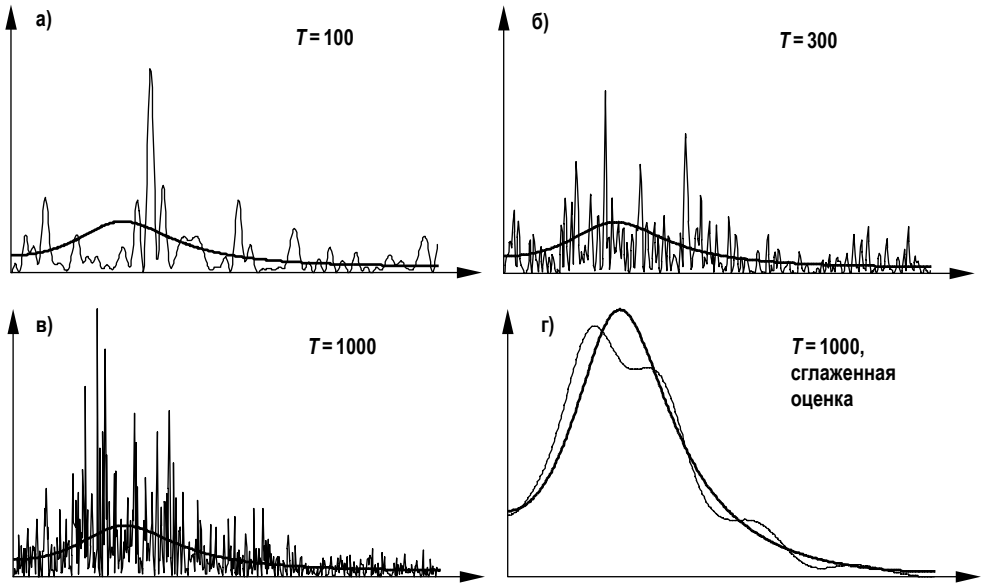

Рис. 13.1

1) Взвешивание ординат периодограммы.

Первый способ сглаживания выборочного спектра применяет метод скользящего среднего, рассмотренный в предыдущей главе, к значениям периодограммы. Сглаживающая оценка определяется в форме

$$\mathbf{p}^s\left(\frac{j}{T}\right) = \sum_{k=-M}^{M} \mu_k \, \mathbf{p}^*\left(\frac{j-k}{T}\right) = \sum_{k=-M}^{M} \mu_k I_{j-k}. \qquad (13.33)$$

Здесь $\{\mu_{-M}, \ldots, \mu_{-1}, \mu_0, \mu_1, \ldots, \mu_M\}$ — $2M+1$ коэффициентов скользящего среднего, которые в сумме должны давать единицу, а также должны быть симметричными, в смысле $\mu_{-k} = \mu_k$. Как правило, коэффициент μ_0 максимальный, а остальные коэффициенты снижаются в обе стороны. Таким образом, наибольший вес в этой оценке имеют значения выборочного спектра, ближайшие к данной частоте $\dfrac{j}{T}$, в связи с чем данный набор коэффициентов называют **спектральным окном**. Через это окно мы как бы «смотрим» на функцию спектральной плотности.

Формула сглаженной спектральной оценки определяется только для значений $j = M, \ldots, [T/2] - M$. Для гармонических частот с номерами $j = 0, \ldots, M-1$ и $j = [T/2] - M + 1, \ldots, [T/2]$ оценка не определена, поскольку при данных значениях j величина $(j-k)$ может принимать значения $-M, \ldots, -1$; $[T/2]+1, \ldots, [T/2]+M$. Однако проблема краевых эффектов легко решается,

если доопределить функцию выборочного спектра (и, соответственно, периодограмму), сделав ее периодической. Формула (13.24) дает такую возможность, поскольку основана на синусоидах и косинусоидах. Таким образом, будем считать, что выборочный спектр определен формулой (13.24) для всех частот $f \in (-\infty, +\infty)$. При этом ясно, что выборочный спектр будет периодической функцией с периодом 1, зеркально-симметричной относительно нуля (и относительно любой частоты вида $i/2$, где i — целое число):

$$\mathbf{p}^*(f - i) = \mathbf{p}^*(f),\ i = \ldots, -1, 0, 1, \ldots$$

и

$$\mathbf{p}^*(-f) = \mathbf{p}^*(f).$$

Формулу (13.33) несложно обобщить так, чтобы можно было производить сглаживание не только в гармонических частотах. Кроме того, в качестве расстояния между «усредняемыми» частотами можно брать не $1/T$, а произвольное $\Delta > 0$. Таким образом приходим к следующей более общей формуле:

$$\mathbf{p}^s(f) = \sum_{k=-M}^{M} \mu_k\, \mathbf{p}^*(f - k\Delta). \tag{13.34}$$

Просматривая поочередно значения выборочного спектра и придавая наибольший вес текущему значению, можно сгладить спектр в каждой интересующей нас точке.

2) Взвешивание автоковариационных функций.

Известно, что при увеличении лага k выборочные автоковариации c_k являются все более неточными оценками. Это связано с тем, что k-я автоковариация вычисляется как среднее по $T-k$ наблюдениям. Отсюда возникает идея в формуле (13.25) ослабить влияние дальних автоковариаций за счет применения понижающих весов так, чтобы с ростом k происходило уменьшение весового коэффициента при c_k.

Если ряд весов, связанных с автоковариациями $c_0,\ c_1, \ldots, c_{T-1}$, обозначить как $m_0, m_1, \ldots, m_{T-1}$, оценка спектра будет иметь вид:

$$\mathbf{p}^c(f) = 2\big(m_0 c_0 + 2\sum_{k=1}^{T-1} m_k c_k \cos 2\pi f k\big), \quad \text{где } 0 \leqslant f \leqslant \frac{1}{2}. \tag{13.35}$$

Набор весов m_k, используемый для получения такой оценки, получил название **корреляционное**, или **лаговое окно**.

При использовании корреляционного окна для уменьшения объема вычислений удобно брать такую систему весов, что $m_k = 0$ при $k \geqslant K$, где $K < T$. Тогда формула (13.35) приобретает вид

$$\mathbf{p}^c(f) = 2\big(m_0 c_0 + 2\sum_{k=1}^{K-1} m_k c_k \cos 2\pi f k\big). \tag{13.36}$$

Корреляционное окно удобно задавать с помощью функции $m(\cdot)$, заданной на интервале $[0;1]$, такой, что $m(0) = 1$, $m(1) = 0$. Обычно функцию выбирают так, чтобы между нулем и единицей эта функция плавно убывала. Тогда понижающие веса m_k при $k = 0, \ldots, K$ вычисляются по формуле

$$m_k = m(k/K).$$

Ясно, что при этом $m_0 = 1$ (это величина, с помощью которой мы взвешиваем дисперсию в формуле (13.35)) и $m_K = 0$.

Наиболее распространенные корреляционные окна, удовлетворяющие перечисленным свойствам — это окна Парзена и Тьюки—Хэннинга (см. рис. 13.2).

Окно Тьюки—Хэннинга:

$$m(z) = \frac{1}{2}\left(1 + \cos(\pi z)\right).$$

Окно Парзена:

$$m(z) = \begin{cases} 1 - 6z^2 + 6z^3, & \text{если } z \in [0; \frac{1}{2}]; \\ 2(1-z)^3, & \text{если } z \in [\frac{1}{2}; 1]. \end{cases}$$

Можно доказать, что сглаживание в частотной области эквивалентно понижающему взвешиванию автоковариаций. Чтобы это сделать, подставим в (13.34) выборочный спектр, выраженный через комплексные экспоненты (13.30):

$$\mathbf{p}^s(f) = 2\sum_{j=-M}^{M} \mu_j \sum_{k=-(T-1)}^{T-1} c_k e^{-i2\pi(f-j\Delta)k}.$$

Меняя здесь порядок суммирования, получим

$$\mathbf{p}^s(f) = 2\sum_{k=-(T-1)}^{T-1} c_k e^{-i2\pi f k} \sum_{j=-M}^{M} \mu_j e^{i2\pi j k \Delta}.$$

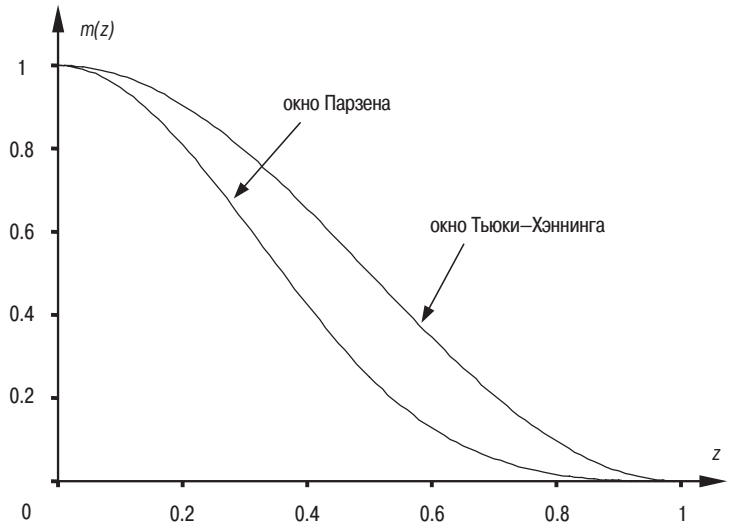

Рис. 13.2. Наиболее популярные корреляционные окна

Введя обозначение

$$\begin{aligned} m_k &= \sum_{j=-M}^{M} \mu_j e^{i2\pi jk\Delta} = \mu_0 + \sum_{j=1}^{M} \mu_j \left(e^{i2\pi jk\Delta} + e^{-i2\pi jk\Delta} \right) \\ &= \mu_0 + 2\sum_{j=1}^{M} \mu_j \cos(2\pi jk\Delta), \end{aligned}$$

придем к формуле

$$\begin{aligned} \mathbf{p}^s(f) &= 2\sum_{k=-(T-1)}^{T-1} m_k c_k e^{-i2\pi fk} = \\ &= 2\big(m_0 c_0 + 2\sum_{k=1}^{T-1} m_k c_k \cos 2\pi fk\big), \end{aligned}$$

которая, очевидно, совпадает с (13.35).

Кроме того, без доказательства отметим, что подбором шага Δ и весов μ_j мы всегда можем сымитировать применение корреляционного окна (13.35) с помощью (13.34). Существует бесконечно много способов это сделать.

В частности, как несложно проверить, частотное окно Тьюки—Хэннинга получим, если возьмем $M = 1$, $\mu_0 = \frac{1}{2}$, $\mu_1 = \mu_{-1} = \frac{1}{4}$ и $\Delta = \dfrac{1}{2K}$.

Особого внимания требует вопрос о том, насколько сильно нужно сглаживать спектральную плотность. Для корреляционных окон степень гладкости зависит от того, насколько быстро убывают понижающие веса. При фиксированной функции $m(\cdot)$ все будет зависеть от параметра K — чем меньше K, тем более гладкой будет оценка.

Для спектральных окон степень гладкости зависит от того, насколько близко «масса» коэффициентов μ_k лежит к той частоте, для которой вычисляется ошибка. При этом принято говорить о **ширине окна**, или **ширине полосы**.

Если ширина окна слишком большая, то произойдет «пересглаживание», оценка будет сильно смещенной, и пики спектральной плотности станут незаметными. (В предельном случае оценка будет ровной, похожей на спектр белого шума.) Если ширина окна слишком малая, то произойдет «недосглаживание», и оценка будет похожа на исходную несглаженную оценку и иметь слишком большую дисперсию. Таким образом, ширина окна выбирается на основе компромисса между смещением и дисперсией.

13.6. Упражнения и задачи

Упражнение 1

Сгенерируйте ряд длиной 200 по модели:

$$x_t = 10 + 0.1t + 4\sin(\pi t/2) - 3\cos(\pi t/2) + \varepsilon_t,$$

где ε_t — нормально распределенный белый шум с дисперсией 3. Предположим, что параметры модели неизвестны, а имеется только сгенерированный ряд x_t.

1.1. Оцените модель линейного тренда и найдите остатки. Постройте график ряда остатков, а также графики автокорреляционной функции и выборочного спектра.

1.2. Выделите тренд и гармоническую составляющую, сравните их параметры с истинными значениями.

Упражнение 2

Для временного ряда, представленного в таблице 12.2 (с. 403), выполните следующие задания.

2.1. Исключите из временного ряда тренд.

2.2. Остатки ряда, получившиеся после исключения тренда, разложите в ряд Фурье.

2.3. Найдите коэффициенты α_j и β_j разложения этого ряда по гармоникам.

2.4. Постройте периодограмму ряда остатков и выделите наиболее существенные гармоники.

2.5. Постройте модель исходного временного ряда как линейную комбинацию модели тренда и совокупности наиболее значимых гармоник.

2.6. Вычислите выборочные коэффициенты автоковариации и автокорреляции для ряда остатков после исключения тренда.

2.7. Найдите значение периодограммы для частоты 0.5 разными способами (в том числе через автоковариационную функцию).

Упражнение 3

Используя данные таблицы 12.2, выполните следующие задания.

3.1. С помощью периодограммы вычислите оценку спектра на частоте $f_j = \dfrac{j}{2K}$, $K = 4$. Получите сглаженную оценку спектра с помощью спектрального окна Тьюки—Хэннинга.

3.2. Рассчитайте автокорреляционную функцию r_j, $j = 1, \ldots, 4$. Оцените спектр с помощью корреляционного окна Тьюки—Хэннинга при $K = 4$. Сравните с предыдущим результатом.

3.3. Оцените спектр с помощью корреляционного окна Парзена при $K = 4$.

3.4. Постройте график оценки спектра для корреляционного окна Тьюки—Хэннинга в точках $f_j = \frac{j}{40}$, $j = 0, \ldots, 20$.

Задачи

1. Записать гармонику для ряда $x = (1, -1, 1, -1, 1, -1, \ldots)$.

2. Пусть временной ряд x_t имеет гармонический тренд: $\tau_t = 3\cos(\pi t) + 4\sin(\pi t)$. Найти значения амплитуды, фазы и периода.

3. Записать ортогональный базис, по которому разлагается исследуемый процесс x_t в ряд Фурье для $T = 6$, $T = 7$.

4. Записать ковариационную матрицу для гармонических переменных, составляющих ортогональный базис, если $T = 6,\ T = 7$.

5. Привести формулу расчета коэффициентов гармонических составляющих временного ряда.

6. Что описывает формула: $I(f) = \dfrac{T}{2}(\alpha_f^2 + \beta_f^2)$, где $0 \leqslant f \leqslant \dfrac{1}{2}$? Почему f меняется в указанном диапазоне значений?

7. Как соотносятся понятия интенсивности и амплитуды, периодограммы и спектра?

8. Вывести формулу для определения периодограммы на нулевой частоте.

9. Как связаны выборочный спектр и автокорреляционная функция для чисто случайного процесса? Записать формулу с расшифровкой обозначений.

10. Пусть для ряда из 4-х наблюдений выборочная автокорреляционная функция равна: $r_1 = {}^1\!/_{\sqrt{2}},\ r_2 = {}^1\!/_2,\ r_3 = {}^1\!/_{\sqrt{2}}$, дисперсия равна 1. Вычислить значение выборочного спектра на частоте ${}^1\!/_4$.

11. Как соотносится выборочный спектр с автоковариационной функцией, спектральными и корреляционными окнами?

12. Пусть для ряда из 4-х наблюдений выборочная автокорреляционная функция равна: $r_1 = {}^1\!/_2,\ r_2 = {}^1\!/_4,\ r_3 = -{}^1\!/_4$, дисперсия равна 1. Вычислить значение сглаженной оценки выборочного спектра на частоте ${}^1\!/_4$ с помощью окна Парзена с весами $m_k,\ k = 1,\ 2$.

13. По некоторому временному ряду рассчитана периодограмма:

    $I(0) = 2,\ I\left({}^1\!/_6\right) = 6,\ I\left({}^1\!/_3\right) = 1,\ I\left({}^1\!/_2\right) = 4.$

    Найти оценки спектральной плотности для тех же частот с использованием окна Тьюки—Хэннинга.

14. Записать уравнение процесса с одной периодической составляющей для частоты 0.33, амплитуды 2 и фазы 0.

15. Изобразить графики спектра для стационарных и нестационарных процессов.

Рекомендуемая литература

1. **Андерсон Т.** Статистический анализ временных рядов. — М.: «Мир», 1976. (Гл. 4, 9).

2. **Бокс Дж., Дженкинс Г.** Анализ временных рядов. Прогноз и управление. Вып. 1. — М.: «Мир», 1974. (Гл. 2).

3. **Гренджер К., Хатанака М.** Спектральный анализ временных рядов в экономике. — М.: «Статистика», 1972.

4. **Дженкинс Г., Ваттс Д.** Спектральный анализ и его приложения. — М.: «Мир», 1971.

5. **Кендалл М. Дж., Стьюарт А.** Многомерный статистический анализ и временные ряды. — М.: «Наука», 1976. (Гл. 49).

6. **Маленво Э.** Статистические методы эконометрии. Вып. 2. — М.: «Статистика», 1976. (Гл. 11, 12).

7. **Бриллинджер Д.** Временные ряды. Обработка данных и теория. — М.: Мир, 1980. (Гл. 5).

8. **Hamilton James D.,** Time Series Analysis. — Princeton University Press, 1994. (Ch. 6).

9. **Judge G.G., Griffiths W.E., Hill R.C., Luthepohl H., Lee T.** Theory and Practice of Econometrics. — New York: John Wiley & Sons, 1985. (Ch. 7).

Глава 14

Линейные стохастические модели ARIMA

14.1. Модель линейного фильтра

Стационарный стохастический процесс $\{x_t\}$ с нулевым математическим ожиданием иногда полезно представлять в виде линейной комбинации последовательности возмущений $\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots$, т.е.

$$x_t = \varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + \ldots = \sum_{i=0}^{\infty} \psi_i\varepsilon_{t-i}, \tag{14.1}$$

или с использованием лагового оператора:

$$x_t = (1 + \psi_1\mathbf{L} + \psi_2\mathbf{L}^2 + \cdots)\varepsilon_t,$$

где $\psi_0 = 1$ и выполняется

$$\sum_{i=0}^{\infty} |\psi_i| < \infty, \tag{14.2}$$

т.е. ряд абсолютных значений коэффициентов сходится.

Уравнение (14.1) называется **моделью линейного фильтра**, а линейный оператор:

$$\boldsymbol{\psi}(\mathbf{L}) = 1 + \psi_1\mathbf{L} + \psi_2\mathbf{L}^2 + \cdots = \sum_{i=0}^{\infty} \psi_i\mathbf{L}^i,$$

преобразующий ε_t в x_t, — **оператором линейного фильтра**.

Компактная запись модели линейного фильтра выглядит следующим образом:

$$x_t = \boldsymbol{\psi}(\mathbf{L})\varepsilon_t.$$

Предполагается, что последовательность $\{\,\varepsilon_t\,\}$ представляет собой чисто случайный процесс или, другими словами, *белый шум* (см. стр. 353). Напомним, что автоковариационная и автокорреляционная функции белого шума имеют очень простую форму:

$$\gamma_k^{\varepsilon} = \begin{cases} \sigma_{\varepsilon}^2, & k = 0, \\ 0, & k \neq 0, \end{cases} \qquad \rho_k^{\varepsilon} = \begin{cases} 1, & k = 0, \\ 0, & k \neq 0, \end{cases}$$

а его спектральная плотность имеет вид

$$\mathbf{p}^{\varepsilon}(f) = 2\gamma_0^{\varepsilon} = 2\sigma_{\varepsilon}^2 = \text{const}.$$

Таким образом, белый шум легко идентифицируется с помощью графиков автокорреляционной функции и спектра. Часто предполагается, что последовательность $\{\varepsilon_t\}$ состоит из независимых одинаково распределенных величин. Упростить анализ помогает дополнительное предположение о том, что $\{\varepsilon_t\}$ имеет нормальное распределение, т.е. представляет собой гауссовский белый шум.

Данная модель не является произвольной. Фактически, согласно теореме Вольда, любой слабо стационарный ряд допускает представление в виде модели линейного фильтра, а именно: **разложение Вольда** ряда x_t[1]. Следует помнить, однако, что разложение Вольда единственно, в то время как представление (14.1), вообще говоря, неоднозначно[2]. Таким образом, разложение Вольда представляет процесс в виде модели линейного фильтра, в то время как модель линейного фильтра не обязательно задает разложение Вольда.

Как мы увидим в дальнейшем, модель линейного фильтра (14.1) применима не только к стационарным процессам, таким что выполняется (14.2), — с соответствующими оговорками она упрощает анализ и многих нестационарных процессов.

Если процесс $\{x_t\}$ подчинен модели (14.1), то при выполнении условия (14.2) он имеет математическое ожидание, равное нулю:

$$\mathbf{E}(x_t) = \sum_{i=0}^{\infty} \psi_i \mathbf{E}(\varepsilon_{t-i}) = 0.$$

[1] В разложении Вольда произвольного стационарного процесса может присутствовать также полностью предсказуемая (линейно детерминированная) компонента. Однако такая компонента, если ее свойства известны, не создает больших дополнительных сложностей для анализа.

[2] См. ниже в этой главе анализ обратимости процесса скользящего среднего.

Если требуется, чтобы математическое ожидание x_t не было равно нулю, то уравнение модели линейного фильтра должно включать константу:

$$x_t = \mu + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} = \mu + \boldsymbol{\psi}(\mathbf{L})\varepsilon_t.$$

Выведем формулы для автоковариаций рассматриваемой модели:

$$\gamma_k = \mathbf{E}(x_t x_{t+k}) = \mathbf{E}\left(\sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} \sum_{j=0}^{\infty} \psi_j \varepsilon_{t+k-j}\right) = \\ = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} \psi_i \psi_j \mathbf{E}(\varepsilon_{t-i}\varepsilon_{t+k-j}) = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}. \quad (14.3)$$

Здесь учитывается, что для белого шума

$$\mathbf{E}(\varepsilon_{t-i}\varepsilon_{t+k-j}) = \begin{cases} \sigma_\varepsilon^2, \ j = i+k, \\ 0, \ j \neq i+k. \end{cases}$$

Заметим, что из (14.2) следует сходимость возникающих здесь рядов. Это говорит о том, что данное условие подразумевает стационарность.

Действительно, пусть (14.2) выполнено. Тогда существует индекс I, такой что $|\psi_i| \leqslant 1$ при $i > I$ (иначе бы ряд не сошелся). Тогда

$$\sum_{i=0}^{\infty} |\psi_i \psi_{i+k}| \leqslant \sum_{i=0}^{\infty} |\psi_i||\psi_{i+k}| \leqslant \sum_{i=0}^{I} |\psi_i||\psi_{i+k}| + \sum_{i=I+1}^{\infty} |\psi_{i+k}|.$$

Поскольку оба слагаемых здесь конечны, то

$$\sum_{i=0}^{\infty} |\psi_i \psi_{i+k}| < \infty.$$

Ясно, что модель линейного фильтра (14.1) в общем виде представляет в основном теоретический интерес, поскольку содержит бесконечное число параметров. Для прикладного моделирования желательно использовать уравнения с конечным числом параметров. В основе таких моделей может лежать так называемая **рациональная аппроксимация** для $\boldsymbol{\psi}(\mathbf{L})$, т.е. приближение в виде частного двух лаговых многочленов:

$$\boldsymbol{\psi}(\mathbf{L}) \approx \frac{\boldsymbol{\theta}(\mathbf{L})}{\boldsymbol{\varphi}(\mathbf{L})},$$

где лаговые многочлены $\boldsymbol{\theta}(\mathbf{L})$ и $\boldsymbol{\varphi}(\mathbf{L})$ имеют уже конечное число параметров. Как показывает практика, многие ряды можно достаточно хорошо аппроксимировать этим методом.

Частными случаями применения рациональной аппроксимации являются модели авторегрессии AR(p) и скользящего среднего MA(q). В общем случае получаем смешанные процессы авторегрессии — скользящего среднего ARMA(p, q). Прежде чем перейти к рассмотрению этих широко используемых линейных моделей временных рядов, рассмотрим общий вопрос о том, как изменяет применение линейного фильтра характеристики случайного процесса.

14.2. Влияние линейной фильтрации на автоковариации и спектральную плотность

Пусть два стационарных процесса с нулевым математическим ожиданием $\{z_t\}$ и $\{y_t\}$ связаны между собой соотношением:

$$z_t = \sum_j \alpha_j y_{t-j},$$

т.е. z_t получается применением к y_t линейного фильтра

$$\alpha(\mathbf{L}) = \sum_j \alpha_j \mathbf{L}^j.$$

Пределы суммирования не указываем, поскольку они могут быть произвольными[3].

Пусть γ_k^y — автоковариации процесса y_t, а $\mathbf{p}^y(f)$ — его спектральная плотность. Найдем те же величины для z_t. Автоковариации z_t равны

$$\begin{aligned}\gamma_k^z &= \mathbf{E}(z_t z_{t-k}) = \mathbf{E}\left(\sum_j \alpha_j y_{t-j} \sum_s \alpha_s y_{t-k-s}\right) = \\ &= \sum_j \sum_s \alpha_j \alpha_s \mathbf{E}(y_{t-j} y_{t-k-s}) = \sum_j \sum_s \alpha_j \alpha_s \gamma_{k+s-j}^y.\end{aligned}$$

Спектральную плотность z_t можно записать в виде

$$\mathbf{p}^z(f) = 2 \sum_{k=-\infty}^{\infty} \gamma_k^z e^{i2\pi fk}.$$

[3] В том числе, при соответствующих предположениях, пределы суммирования могут быть бесконечными. Кроме того, z_t здесь может зависеть от опережающих значений y_t .

Подставим сюда формулы для ковариаций:

$$\begin{aligned}\mathbf{p}^z(f) &= 2\sum_{k=-\infty}^{\infty}\sum_j\sum_s \alpha_j\alpha_s\gamma^y_{k+s-j}e^{i2\pi fk} = \\ &= 2\sum_j\sum_s\left(\alpha_j\alpha_s\sum_{k=-\infty}^{\infty}\gamma^y_{k+s-j}e^{i2\pi fk}\right).\end{aligned}$$

Произведем здесь замену $k' = k+s-j$:

$$\begin{aligned}\mathbf{p}^z(f) &= 2\sum_j\sum_s\left(\alpha_j\alpha_s\sum_{k'=-\infty}^{\infty}\gamma^y_{k'}e^{i2\pi f(k'+j-s)}\right) = \\ &= 2\sum_{k'=-\infty}^{\infty}\gamma^y_{k'}e^{i2\pi fk'}\cdot\sum_j\sum_s\alpha_je^{i2\pi fj}\alpha_se^{-i2\pi fs} = \\ &= 2\sum_{k'=-\infty}^{\infty}\gamma_{k'}e^{i2\pi fk'}\cdot\sum_j\alpha_je^{i2\pi fj}\cdot\sum_s\alpha_se^{-i2\pi fs}.\end{aligned}$$

Первый множитель — это спектральная плотность y_t. Два последних множителя представляют собой сопряженные комплексные числа, поэтому их произведение равно квадрату их модуля. Окончательно получим

$$\mathbf{p}^z(f) = \mathbf{p}^y(f)\Big|\sum_j\alpha_je^{-i2\pi fj}\Big|^2 \tag{14.4}$$

или

$$\mathbf{p}^z(f) = \mathbf{p}^y(f)\left(\left(\sum_j\alpha_j\cos 2\pi fj\right)^2+\left(\sum_j\alpha_j\sin 2\pi fj\right)^2\right). \tag{14.5}$$

Данную теорию несложно применить к модели линейного фильтра (14.1). Для этого заменяем z_t на x_t, а y_t на ε_t. Автоковариации x_t равны

$$\gamma_k = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\psi_i\psi_j\gamma^{\varepsilon}_{k+j-i}.$$

Поскольку $\gamma^{\varepsilon}_0 = \sigma^2_{\varepsilon}$, и $\gamma^{\varepsilon}_k = 0$ при $k\neq 0$, то

$$\gamma_k = \sigma^2_{\varepsilon}\sum_{i=0}^{\infty}\psi_i\psi_{i+k},$$

что совпадает с полученной ранее формулой (14.3).

Для спектральной плотности из (14.4) получаем

$$\mathbf{p}(f) = 2\sigma_\varepsilon^2 \Big| \sum_{j=0}^{\infty} \psi_j e^{-i2\pi f j} \Big|^2. \tag{14.6}$$

Поскольку $|e^{-i2\pi f j}| = 1$, то ряд здесь сходится и $|\mathbf{p}(f)| < \infty$.

14.3. Процессы авторегрессии

В модели авторегрессии текущее значение процесса x_t представляется в виде линейной комбинации конечного числа предыдущих значений процесса и белого шума ε_t:

$$x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots + \varphi_p x_{t-p} + \varepsilon_t, \tag{14.7}$$

при этом предполагается, что текущее значение ε_t не коррелировано с лагами x_t. Такая модель называется **авторегрессией p-го порядка** и обозначается AR(p) (от английского *autoregression*).

Используя лаговый оператор $\mathbf{L}$, представим уравнение авторегрессии в виде:

$$(1 - \varphi_1 \mathbf{L} - \varphi_2 \mathbf{L}^2 - \ldots - \varphi_p \mathbf{L}^p) x_t = \varepsilon_t,$$

или кратко, через лаговый многочлен $\boldsymbol{\varphi}(\mathbf{L}) = 1 - \varphi_1 \mathbf{L} - \varphi_2 \mathbf{L}^2 - \ldots - \varphi_p \mathbf{L}^p$:

$$\boldsymbol{\varphi}(\mathbf{L}) x_t = \varepsilon_t.$$

Нетрудно показать, что модель авторегрессии является частным случаем модели линейного фильтра:

$$x_t = \boldsymbol{\psi}(\mathbf{L}) \varepsilon_t,$$

где $\boldsymbol{\psi}(\mathbf{L}) = \boldsymbol{\varphi}^{-1}(\mathbf{L})$, т.е. $\boldsymbol{\psi}(\mathbf{L})$ — оператор, обратный оператору $\boldsymbol{\varphi}(\mathbf{L})$.

Удобным и полезным инструментом для изучения процессов авторегрессии является **характеристический многочлен** (характеристический полином)

$$\boldsymbol{\varphi}(z) = 1 - \varphi_1 z - \varphi_2 z^2 - \ldots - \varphi_p z^p = 1 - \sum_{j=1}^{p} \varphi_j z^j$$

и связанное с ним **характеристическое уравнение**

$$1 - \varphi_1 z - \varphi_2 z^2 - \ldots - \varphi_p z^p = 0.$$

Как мы увидим в дальнейшем, от того, какие корни имеет характеристическое уравнение, зависят свойства процесса авторегрессии, в частности, является ли процесс стационарным или нет.

Рассмотрим наиболее часто использующиеся частные случаи авторегрессионных процессов.

Процесс Маркова

Процессом Маркова (**марковским процессом**) называется авторегрессионный процесс первого порядка, AR(1):

$$x_t = \varphi x_{t-1} + \varepsilon_t, \tag{14.8}$$

где ε_t представляет собой белый шум, который не коррелирует с x_{t-1}. Здесь мы упростили обозначения, обозначив $\varphi = \varphi_1$.

Найдем необходимые условия стационарности марковского процесса. Предположим, что процесс $\{x_t\}$ слабо стационарен. Тогда его первые и вторые моменты неизменны. Находя дисперсии от обеих частей (14.8), получим, учитывая, что $\mathbf{cov}(x_{t-1}, \varepsilon_t) = 0$:

$$\mathbf{var}(x_t) = \varphi^2 \mathbf{var}(x_{t-1}) + \mathbf{var}(\varepsilon_t)$$

или

$$\sigma_x^2 = \varphi^2 \sigma_x^2 + \sigma_\varepsilon^2.$$

Ясно, что при $|\varphi| \geqslant 1$, с учетом $\sigma_\varepsilon^2 > 0$, правая часть этого равенства должна быть больше левой, что невозможно. Получаем, что у стационарного марковского процесса $|\varphi| < 1$.

Пусть, с другой стороны, $|\varphi| < 1$. Представим x_t через белый шум $\{\varepsilon_t\}$. Это можно осуществить с помощью последовательных подстановок по формуле (14.8):

$$x_t = \varphi x_{t-1} + \varepsilon_t = \varphi(\varphi x_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = \varphi^2 x_{t-2} + \varphi\varepsilon_{t-1} + \varepsilon_t,$$

потом

$$x_t = \varphi^2(\varphi x_{t-3} + \varepsilon_{t-3}) + \varphi\varepsilon_{t-1} + \varepsilon_t = \varphi^3 x_{t-3} + \varphi^2\varepsilon_{t-3} + \varphi\varepsilon_{t-1} + \varepsilon_t,$$

и т.д.

В пределе, поскольку множитель при лаге x_t стремится к нулю, получим следующее представление x_t в виде модели линейного фильтра:

$$x_t = \varepsilon_t + \varphi\varepsilon_{t-1} + \varphi^2\varepsilon_{t-2} + \dots .$$

Это же представление можно получить с использованием оператора лага. Уравнение (14.8) запишется в виде

$$(1 - \varphi\mathbf{L})x_t = \varepsilon_t.$$

Применив к обеим частям уравнения $(1 - \varphi\mathbf{L})^{-1}$, получим

$$x_t = (1 - \varphi\mathbf{L})^{-1}\varepsilon_t = (1 + \varphi\mathbf{L} + \varphi^2\mathbf{L}^2 + \dots)\varepsilon_t = \varepsilon_t + \varphi\varepsilon_{t-1} + \varphi^2\varepsilon_{t-2} + \dots .$$

В терминах модели линейного фильтра (14.1) для марковского процесса $\psi_i = \varphi^i$. Поэтому

$$\sum_{i=0}^{\infty} |\psi_i| = \sum_{i=0}^{\infty} |\varphi|^i = \frac{1}{1 - |\varphi|} < \infty, \tag{14.9}$$

т.е. условие стационарности модели линейного фильтра (14.2) выполняется при $|\varphi| < 1$.

Можно сделать вывод, что **условие стационарности процесса Маркова** имеет следующий вид:

$$|\varphi| < 1.$$

Свойства стационарного процесса AR(1):

1) Если процесс x_t слабо стационарен, то его математическое ожидание неизменно, поэтому, беря математическое ожидание от обеих частей (14.8), получим $\mathbf{E}(x_t) = \varphi\mathbf{E}(x_t)$, откуда

$$\mathbf{E}(x_t) = 0.$$

Если добавить в уравнение (14.8) константу:

$$x_t = \mu + \varphi x_{t-1} + \varepsilon_t,$$

то $\mathbf{E}(x_t) = \mu + \varphi\mathbf{E}(x_t)$ и

$$\mathbf{E}(x_t) = \frac{\mu}{1 - \varphi}.$$

2) Найдем дисперсию процесса Маркова, используя полученное выше уравнение $\sigma_x^2 = \varphi^2\sigma_x^2 + \sigma_\varepsilon^2$:

$$\mathbf{var}(x_t) = \gamma_0 = \sigma_x^2 = \frac{\sigma_\varepsilon^2}{1 - \varphi^2}. \tag{14.10}$$

Можно также применить общую формулу для автоковариаций в модели линейного фильтра (14.3) с $k = 0$:

$$\gamma_0 = \sigma_x^2 = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \psi_i^2 = \sigma_\varepsilon^2 (1 + \varphi^2 + \varphi^4 + \dots) = \frac{\sigma_\varepsilon^2}{1 - \varphi^2}.$$

При $|\varphi| \geqslant 1$ дисперсия процесса $\{x_t\}$, вычисляемая по этой формуле, неограниченно растет.

3) Коэффициент автоковариации k-го порядка по формуле (14.3) равен

$$\gamma_k = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k} = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \varphi^i \varphi^{i+k} = \\ = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \varphi^{2i+k} = \sigma_\varepsilon^2 \varphi^k \sum_{i=0}^{\infty} \varphi^{2i} = \frac{\sigma_\varepsilon^2}{1 - \varphi^2} \varphi^k.$$

Можно вывести ту же формулу иным способом. Ошибка ε_t некоррелирована не только с x_{t-1}, но и с x_{t-2}, x_{t-3} и т.д. Поэтому, умножая уравнение процесса (14.8) на x_{t-k} при $k > 0$ и беря математическое ожидание от обеих частей, получим

$$\mathbf{E}(x_t x_{t-k}) = \varphi \mathbf{E}(x_{t-1} x_{t-k})$$

или

$$\gamma_k = \varphi \gamma_{k-1}, \quad k > 0.$$

Таким образом, учитывая (14.10), получим

$$\gamma_k = \gamma_0 \varphi^k = \frac{\sigma_\varepsilon^2}{1 - \varphi^2} \varphi^k. \tag{14.11}$$

4) Коэффициент автокорреляции, исходя из (14.11), равен:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \varphi^k.$$

При $0 < \varphi < 1$ автокорреляционная функция имеет форму затухающей экспоненты (рис. 14.1а), при $-1 < \varphi < 0$ — форму затухающей знакопеременной экспоненты (рис. 14.1б).

Если $\varphi > 1$, процесс Маркова превращается во **«взрывной»** процесс.

В случае $\varphi = 1$ имеет место так называемый процесс **случайного блуждания**, который относится к разряду нестационарных.

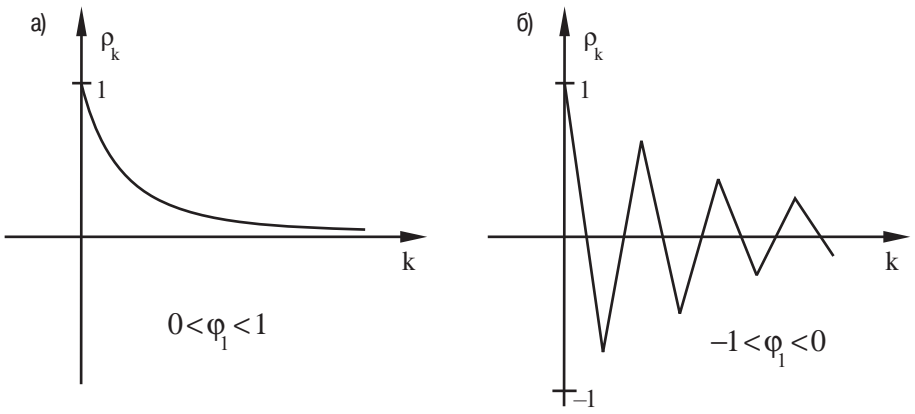

Рис. 14.1

Процесс Юла

Процессом Юла называют авторегрессию второго порядка AR(2):

$$x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \varepsilon_t, \tag{14.12}$$

или, через лаговый оператор:

$$(1 - \varphi_1 \mathbf{L} - \varphi_2 \mathbf{L}^2) x_t = \varepsilon_t.$$

Для стационарности процесса авторегрессии AR(2) необходимо, чтобы корни λ_1, λ_2 характеристического уравнения

$$1 - \varphi_1 z - \varphi_2 z^2 = 0,$$

которые, вообще говоря, могут быть комплексными, находились вне единичного круга на комплексной плоскости, т.е. $|\lambda_1| > 1, |\lambda_2| > 1$.

Неформально обоснуем условия стационарности AR(2), разложив характеристический полином $\boldsymbol{\varphi}(z)$ на множители (по теореме Виета $\lambda_1 \lambda_2 = -1/\varphi_2$):

$$\boldsymbol{\varphi}(z) = -\varphi_2(\lambda_1 - z)(\lambda_2 - z) = \Big(1 - \frac{z}{\lambda_1}\Big)\Big(1 - \frac{z}{\lambda_2}\Big) = (1 - G_1 z)(1 - G_2 z),$$

где мы ввели обозначения $G_1 = 1/\lambda_1$ и $G_2 = 1/\lambda_2$.

Такое разложение позволяет представить уравнение AR(2) в виде:

$$(1 - G_1 \mathbf{L})(1 - G_2 \mathbf{L}) x_t = \varepsilon_t. \tag{14.13}$$

Введем новую переменную:

$$v_t = (1 - G_2\mathbf{L})x_t, \tag{14.14}$$

тогда уравнение (14.13) примет вид:

$$(1 - G_1\mathbf{L})v_t = \varepsilon_t.$$

Видим, что ряд v_t является процессом Маркова:

$$v_t = G_1 v_{t-1} + \varepsilon_t.$$

Для того чтобы процесс v_t был стационарным, коэффициент G_1 по модулю должен быть меньше единицы, т.е. $|\lambda_1| > 1$.

С другой стороны, из (14.14) следует, что x_t имеет вид процесса Маркова с ошибкой v_t:

$$x_t = G_2 x_{t-1} + v_t.$$

Условие стационарности такого процесса имеет аналогичный вид: $|\lambda_2| > 1$.

Приведенные рассуждения не вполне корректны, поскольку v_t не является белым шумом. (Укажем без доказательства, что из стационарности v_t следует стационарность x_t.) Кроме того, корни λ_1, λ_2 могут быть, вообще говоря, комплексными, что требует изучения свойств комплексного марковского процесса. Более корректное обоснование приведенного условия стационарности будет получено ниже для общего случая AR(p).

Условия стационарности процесса AR(2), $|\lambda_1| > 1$, $|\lambda_2| > 1$, можно переписать в эквивалентном виде как ограничения на параметры уравнения авторегрессии:

$$\begin{cases} \varphi_2 + \varphi_1 < 1, \\ \varphi_2 - \varphi_1 < 1, \\ -1 < \varphi_2 < 1. \end{cases} \tag{14.15}$$

Проверим эти условия. Для этого рассмотрим два случая.

1) Пусть корни характеристического уравнения вещественные, то есть $\varphi_1^2 + 4\varphi_2 \geqslant 0$. Тогда для выполнения условий $|\lambda_1| > 1$, $|\lambda_2| > 1$ необходимым требованием является $|\lambda_1\lambda_2| > 1$ или $\left|-\frac{1}{\varphi_2}\right| > 1$. В таком случае один из корней обязательно лежит вне отрезка $[-1, 1]$. Для того чтобы и второй корень не попал в этот отрезок, необходимо и достаточно, чтобы значения характеристического полинома $\boldsymbol{\varphi}(\mathbf{L}) = 1 - \varphi_1\mathbf{L} - \varphi_2\mathbf{L}^2$ в точках -1 и 1 были одного знака. Это условие можно описать неравенством:

$$\boldsymbol{\varphi}(-1) \cdot \boldsymbol{\varphi}(1) > 0, \quad \text{или} \quad (1 + \varphi_1 - \varphi_2)(1 - \varphi_1 - \varphi_2) > 0.$$

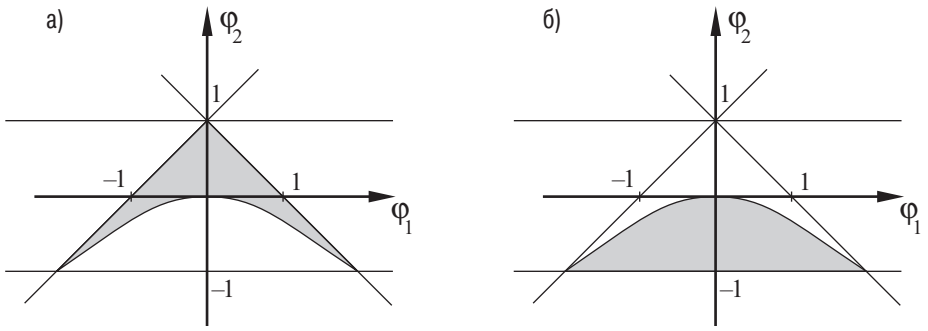

Рис. 14.2

Таким образом, случай вещественных корней описывается системой:

$$\begin{cases} \varphi_1^2 + 4\varphi_2 \geqslant 0, \\ |\varphi_2| < 1, \\ (1 + \varphi_1 - \varphi_2)(1 - \varphi_1 - \varphi_2) > 0. \end{cases}$$

Если связать φ_1 и φ_2 с координатными осями, то область, соответствующую данной системе, можно изобразить на рисунке (см. рис. 14.2а).

2) Если корни комплексные, то они имеют одинаковую абсолютную величину:

$$|\lambda_1| = |\lambda_2| = \sqrt{-\frac{1}{\varphi_2}}.$$

И тогда вместе с отрицательностью дискриминанта достаточно условия:

$$|\varphi_2| < 1 \quad \text{(область решений см. на рис. 14.2б).}$$

Если объединить случаи 1) и 2), то общее решение как раз описывается системой неравенств (14.15) и соответствующая область на координатной плоскости представляет собой треугольник, ограниченный прямыми:

$$\varphi_2 = \varphi_1 + 1, \quad \varphi_2 = -\varphi_1 + 1, \quad \varphi_2 = -1.$$

Автокорреляционная и автоковариационная функция AR(p)

Для стационарного процесса авторегрессии:

$$x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots + \varphi_p x_{t-p} + \varepsilon_t \tag{14.16}$$

можно вывести формулу автокорреляционной функции. Умножив обе части уравнения на x_{t-k}:

$$x_{t-k}x_t = \varphi_1 x_{t-k}x_{t-1} + \varphi_2 x_{t-k}x_{t-2} + \ldots + \varphi_p x_{t-k}x_{t-p} + x_{t-k}\varepsilon_t,$$

и перейдя к математическим ожиданиям, получим уравнение, связывающее коэффициенты автоковариации различного порядка:

$$\gamma_k = \varphi_1\gamma_{k-1} + \varphi_2\gamma_{k-2} + \ldots + \varphi_p\gamma_{k-p}, \quad k > 0. \tag{14.17}$$

Это выражение является следствием того, что соответствующие кроссковариации между процессом и ошибкой равны нулю: $\mathbf{E}(x_{t-k}\varepsilon_t) = 0$ при $k > 0$, т.к. x_{t-k} может включать лишь ошибки ε_j для $j \leqslant t-k$.

Делением уравнения (14.17) на γ_0 получаем важное рекуррентное соотношение для автокорреляционной функции:

$$\rho_k = \varphi_1\rho_{k-1} + \varphi_2\rho_{k-2} + \ldots + \varphi_p\rho_{k-p}, \quad k > 0. \tag{14.18}$$

Подставляя в выражение (14.18) $k = 1, \ldots, p$, получаем, с учетом симметричности автокорреляционной функции, так называемые **уравнения Юла—Уокера** (Yule—Walker) для AR(p):

$$\begin{cases} \rho_1 = \varphi_1 + \varphi_2\rho_1 + \ldots + \varphi_p\rho_{p-1}, \\ \rho_2 = \varphi_1\rho_1 + \varphi_2 + \ldots + \varphi_p\rho_{p-2}, \\ \quad \ldots \\ \rho_p = \varphi_1\rho_{p-1} + \varphi_2\rho_{p-2} + \ldots + \varphi_p. \end{cases} \tag{14.19}$$

Мы имеем здесь p линейных уравнений, связывающих p автокорреляций, $\rho_1, \ldots, \rho_p$. Из этой системы при данных параметрах можно найти автокорреляции. С другой стороны, при данных автокорреляциях из уравнений Юла—Уокера можно найти параметры $\varphi_1, \ldots, \varphi_p$. Замена теоретических автокорреляций выборочными дает метод оценивания параметров процесса AR(p).

В частности, для процесса Юла получим из (14.19)

$$\rho_1 = \frac{\varphi_1}{1-\varphi_2}, \quad \rho_2 = \frac{\varphi_1^2}{1-\varphi_2} + \varphi_2.$$

Зная $\rho_1, \ldots, \rho_p$, все последующие автокорреляции ρ_k ($k > p$) можем найти по рекуррентной формуле (14.18).

Для нахождения автоковариационной функции требуется знать γ_0, дисперсию процесса x_t. Если умножить обе части (14.16) на ε_t и взять математические ожидания, получим, что $\mathbf{E}(\varepsilon_t x_t) = \mathbf{E}(\varepsilon_t^2) = \sigma_\varepsilon^2$. Далее, умножая обе части (14.16) на x_t и беря математические ожидания, получим

$$\gamma_0 = \varphi_1\gamma_1 + \varphi_2\gamma_2 + \ldots + \varphi_p\gamma_p + \sigma_\varepsilon^2.$$

Значит, если известны автокорреляции, то дисперсию x_t можно вычислять по формуле:

$$\gamma_0 = \sigma_x^2 = \frac{\sigma_\varepsilon^2}{1 - \varphi_1\rho_1 - \varphi_2\rho_2 - \ldots - \varphi_p\rho_p}. \tag{14.20}$$

Автоковариации затем можно вычислить как $\gamma_j = \rho_j\sigma_x^2$.

С учетом полученного дополнительного уравнения можно записать вариант уравнений Юла—Уокера для автоковариаций:

$$\begin{cases} \gamma_0 = \varphi_1\gamma_1 + \varphi_2\gamma_2 + \ldots + \varphi_p\gamma_p + \sigma_\varepsilon^2, \\ \gamma_1 = \varphi_1\gamma_0 + \varphi_2\gamma_1 + \ldots + \varphi_p\gamma_{p-1}, \\ \gamma_2 = \varphi_1\gamma_1 + \varphi_2\gamma_0 + \ldots + \varphi_p\gamma_{p-2}, \\ \quad \ldots \\ \gamma_p = \varphi_1\gamma_{p-1} + \varphi_2\gamma_{p-2} + \ldots + \varphi_p\gamma_0. \end{cases} \tag{14.21}$$

В этой системе имеется $p+1$ уравнений, связывающих $p+1$ автоковариацию, что позволяет непосредственно вычислять автоковариации при данных параметрах.

Заметим, что 14.17 и 14.18 имеют вид линейных однородных конечно-разностных уравнений, а для подобных уравнений существует общий метод нахождения решения. Решив уравнение 14.18, можно получить общий вид автокорреляционной функции процесса авторегрессии. Проведем это рассуждение более подробно для процесса Юла, а затем рассмотрим общий случай AR(p).

Вывод формулы автокорреляционной функции процесса Юла

Из соотношения (14.18) для $p = 2$ получаем:

$$\rho_k = \varphi_1\rho_{k-1} + \varphi_2\rho_{k-2}, \quad k > 0. \tag{14.22}$$

или, используя лаговый оператор, который в данном случае действует на k, $\boldsymbol{\varphi}(\mathbf{L})\rho_k = 0$, где $\boldsymbol{\varphi}(z) = 1 - \varphi_1 z - \varphi_2 z^2$ — характеристический полином процесса Юла. Как мы видели, этот характеристический полином можно представить в виде:

$$\boldsymbol{\varphi}(z) = (1 - G_1 z)(1 - G_2 z),$$

где $G_1 = 1/\lambda_1$ и $G_2 = 1/\lambda_2$, а λ_1, λ_2 — корни характеристического уравнения.

Таким образом, ρ_k удовлетворяет уравнению:

$$(1 - G_1\mathbf{L})(1 - G_2\mathbf{L})\rho_k = 0. \tag{14.23}$$

Найдем общее решение этого уравнения. Введем обозначение:

$$\omega_k = (1 - G_2\mathbf{L})\rho_k. \tag{14.24}$$

Для ω_k имеем $(1 - G_1\mathbf{L})\omega_k = 0$ или $\omega_k = G_1\omega_{k-1}$. Поэтому

$$\omega_k = G_1\omega_{k-1} = \cdots = G_1^{k-1}\omega_1.$$

В свою очередь, из (14.24), с учетом того, что $\rho_0 = 1$ и $\rho_1 = \varphi_1(1 - \varphi_2)$, следует, что

$$\omega_1 = \rho_1 - G_2\rho_0 = \frac{\varphi_1}{1 - \varphi_2} - G_2.$$

Поскольку, по теореме Виета, $G_1 + G_2 = \varphi_1$, $G_1G_2 = -\varphi_2$, получаем выражение для ω_1 через корни характеристического уравнения:

$$\omega_1 = \frac{G_1 + G_2}{1 + G_1G_2} - G_2 = \frac{G_1(1 - G_2^2)}{1 + G_1G_2}. \tag{14.25}$$

Возвращаясь к формуле (14.24), имеем, исходя из рекуррентности соотношения:

$$\begin{aligned}\rho_k &= G_2\rho_{k-1} + \omega_k = G_2(G_2\rho_{k-2} + \omega_{k-1}) + \omega_k = \ldots = \\ &= G_2^k + G_2^{k-1}\omega_1 + G_2^{k-2}\omega_2 + \ldots + G_2\omega_{k-1} + \omega_k = G_2^k + \sum_{s=0}^{k-1} G_2^{k-1-s}\omega_{s+1}.\end{aligned}$$

Подставляя сюда $\omega_k = G_1^{k-1}\omega_1$, получим

$$\begin{aligned}\rho_k &= G_2^k + \omega_1 G_2^{k-1}\sum_{s=0}^{k-1}\left(\frac{G_1}{G_2}\right)^s = G_2^k + \omega_1 G_2^{k-1}\frac{(G_1/G_2)^k - 1}{(G_1/G_2) - 1} = \\ &= G_2^k + \omega_1\frac{G_1^k - G_2^k}{G_1 - G_2} = \left(\frac{\omega_1}{G_1 - G_2}\right)G_1^k + \left(1 - \frac{\omega_1}{G_1 - G_2}\right)G_2^k.\end{aligned}$$

Таким образом, общее решение уравнения (14.22) имеет вид:

$$\rho_k = A_1G_1^k + A_2G_2^k, \tag{14.26}$$

где коэффициенты A_1 и A_2 вычисляются по формулам:

$$A_1 = \frac{G_1(1 - G_2^2)}{(G_1 - G_2)(1 + G_1G_2)}, \quad A_2 = -\frac{G_2(1 - G_1^2)}{(G_1 - G_2)(1 + G_1G_2)}, \tag{14.27}$$

причем $A_1 + A_2 = 1$.

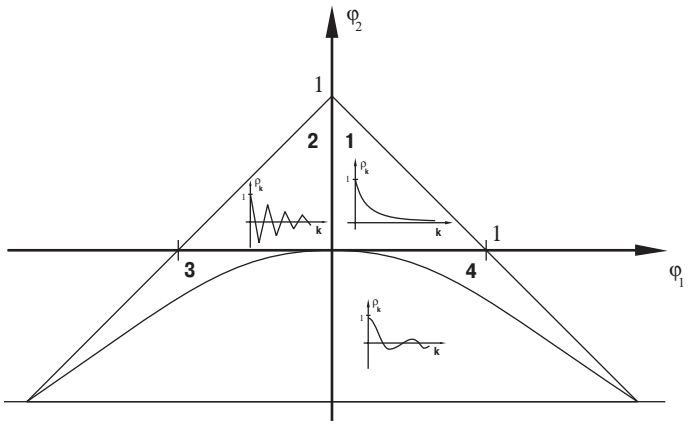

Рис. 14.3

В стационарных процессах корни характеристического уравнения лежат вне единичного круга. Следовательно, $|G_1| < 1$ и $|G_2| < 1$, и автокорреляционная функция состоит из совокупности затухающих экспонент, что на рисунке 14.3 соответствует областям 1, 2, 3 и 4, лежащим выше параболической границы $\varphi_1^2 + 4\varphi_2 \geqslant 0$.

При этом, если оба корня положительны или доминирует по модулю отрицательный корень (соответственно, положительное G), автокорреляционная функция затухает, асимптотически приближаясь к экспоненте (области 1 и 4 на рис. 14.3). Когда же оба корня отрицательны или доминирует по модулю положительный корень (или отрицательное G), автокорреляционная функция затухает по экспоненте знакопеременно (области 2 и 3 на рис. 14.3).

Если корни разного знака и совпадают по модулю, то затухание ρ_k происходит в области положительных значений, но имеет колебательный характер:

$G_1 = -G_2$, следовательно $A_1 = A_2 = 0.5$ и

$$\rho_k = \begin{cases} 0, & \text{если } k \text{ — нечетное,} \\ G_1^k, & \text{если } k \text{ — четное.} \end{cases}$$

Рассмотрим случай, когда корни $\lambda_1 = G_1^{-1}$ и $\lambda_2 = G_2^{-1}$ — комплексные. Тогда автокорреляционная функция процесса авторегрессии второго порядка будет представлять собой затухающую синусоиду:

$$\rho_k = \frac{d^k \cdot \sin(k\alpha + \beta)}{\sin \beta},$$

$$\text{где} \quad d = \sqrt{-\varphi_2}, \quad \alpha = \arccos \frac{\varphi_1}{2\sqrt{-\varphi_2}}, \quad \beta = \operatorname{arctg}\left(\frac{1+d^2}{1-d^2}\operatorname{tg}\alpha\right).$$

Подтвердим справедливость формулы. Поскольку коэффициенты характеристического многочлена — действительные числа, то λ_1 и λ_2 — комплексно-сопряженные, поэтому G_1 и G_2 тоже являются комплексно-сопряженными. Далее, G_1, как и любое комплексное число, можно представить в виде:

$$G_1 = de^{i\alpha} = d\cos\alpha + i\, d\sin\alpha,$$

где $d = |G_1|$ $(= |G_2|)$ — модуль, а α — аргумент комплексного числа G_1. Соответственно, для G_2 как для сопряженного числа верно представление:

$$G_2 = de^{-i\alpha} = d\cos\alpha - i\, d\sin\alpha.$$

Для нахождения d воспользуемся тем, что

$$-\varphi_2 = G_1 \cdot G_2 = de^{i\alpha} \cdot de^{-i\alpha} = d^2.$$

Значит, $d = \sqrt{-\varphi_2}$. Кроме того, $\cos\alpha = \dfrac{e^{i\alpha}+e^{-i\alpha}}{2} = \dfrac{G_1+G_2}{2d}$, поэтому, учитывая, что $G_1 + G_2 = \varphi_1$, получаем:

$$\cos\alpha = \frac{\varphi_1}{2\sqrt{-\varphi_2}}.$$

Подставим теперь $G_1 = de^{i\alpha}$ и $G_2 = de^{-i\alpha}$ в выражения (14.27) для A_1 и A_2:

$$A_1 = \frac{de^{i\alpha}(1-d^2e^{-2i\alpha})}{(de^{i\alpha}-de^{-i\alpha})(1+d^2e^{i\alpha}e^{-i\alpha})} = \frac{e^{i\alpha}-d^2e^{-i\alpha}}{(e^{i\alpha}-e^{-i\alpha})(1+d^2)} =$$

$$= \frac{(1-d^2)\cos\alpha + i(1+d^2)\sin\alpha}{2i(1+d^2)\sin\alpha} = \frac{1+i\dfrac{1+d^2}{1-d^2}\operatorname{tg}\alpha}{2i}.$$

Пусть β — такой угол, что

$$\operatorname{tg}\beta = \frac{1+d^2}{1-d^2}\operatorname{tg}\alpha.$$

Тогда

$$A_1 = \frac{1+i\operatorname{tg}\beta}{2i} = \frac{\cos\beta + i\sin\beta}{2i\sin\beta} = \frac{e^{i\beta}}{e^{i\beta}-e^{-i\beta}}.$$

Отсюда найдем A_2:

$$A_2 = 1 - A_1 = -\frac{e^{-i\beta}}{e^{i\beta}-e^{-i\beta}}.$$

Несложно увидеть, что A_1 и A_2 являются комплексно-сопряженными.

Подставим найденные A_1 и A_2 в (14.26):

$$\rho_k = A_1 G_1^k + A_2 G_2^k = \frac{e^{i\beta}}{e^{i\beta} - e^{-i\beta}} \cdot d^k e^{ik\alpha} - \frac{e^{-i\beta}}{e^{i\beta} - e^{-i\beta}} \cdot d^k e^{-ik\alpha} =$$
$$= d^k \frac{e^{i(k\alpha+\beta)} - e^{-i(k\alpha+\beta)}}{e^{i\beta} - e^{-i\beta}}.$$

Таким образом, подтверждается, что автокорреляционную функцию можно записать в форме

$$\rho_k = \frac{d^k \sin(k\alpha + \beta)}{\sin\beta}.$$

Нахождение автокорреляционной функции AR(p) с помощью решения конечно-разностного уравнения

Аналогичным образом можно изучать автокорреляционную функцию процесса AR(p) при произвольном p. Запишем уравнение (14.18) с помощью лагового оператора, действующего на k:

$$\boldsymbol{\varphi}(\mathbf{L})\rho_k = 0, \quad k > 0, \tag{14.28}$$

где $\boldsymbol{\varphi}(\mathbf{L}) = 1 - \varphi_1 \mathbf{L} - \varphi_2 \mathbf{L}^2 - \ldots - \varphi_p \mathbf{L}^p$.

Рассмотрим характеристическое уравнение:

$$\boldsymbol{\varphi}(z) = 1 - \varphi_1 z - \varphi_2 z^2 - \ldots - \varphi_p z^p = 0.$$

Пусть λ_i $(i = 1, \ldots, p)$ — корни этого уравнения. Мы будем предполагать, что все они различны. Характеристический многочлен $\boldsymbol{\varphi}(z)$ можно разложить следующим образом:

$$\boldsymbol{\varphi}(z) = -\varphi_p \prod_{i=1}^{p} (\lambda_i - z).$$

Обозначим через G_i значения, обратные корням характеристического уравнения: $G_i = 1/\lambda_i$. Тогда, учитывая, что $\lambda_1 \cdot \lambda_2 \cdot \ldots \cdot \lambda_p = -1/\varphi_p$ и соответственно $G_1 \cdot G_2 \cdot \ldots \cdot G_p = -\varphi_p$, имеем:

$$\boldsymbol{\varphi}(z) = -\varphi_p \prod_{i=1}^{p} \left(\frac{1}{G_i} - z\right) = -\varphi_p \frac{\prod\limits_{i=1}^{p} (1 - G_i z)}{\prod\limits_{i=1}^{p} G_i} = \prod_{i=1}^{p} (1 - G_i z).$$

Исходя из этого, перепишем уравнение (14.28) в виде:

$$(1 - G_1\mathbf{L})(1 - G_2\mathbf{L})\cdot\ldots\cdot(1 - G_p\mathbf{L})\rho_k = 0. \qquad (14.29)$$

Из теории конечно-разностных уравнений известно, что если все корни λ_i различны, то общее решение уравнения (14.18) имеет вид:

$$\rho_k = A_1G_1^k + A_2G_2^k + \ldots + A_pG_p^k, \quad k > -p, \qquad (14.30)$$

где A_i — некоторые константы, в общем случае комплексные. (Обсуждение решений линейных конечно-разностных уравнений см. в Приложении А.4.)

Проверим, что это действительно решение. Подставим ρ_k в (14.29):

$$\prod_{i=1}^{p}(1 - G_i\mathbf{L})\rho_k = \prod_{i=1}^{p}(1 - G_i\mathbf{L})(A_1G_1^k + A_2G_2^k + \ldots + A_pG_p^k) =$$

$$= \prod_{i=1}^{p}(1 - G_i\mathbf{L})A_1G_1^k + \prod_{i=1}^{p}(1 - G_i\mathbf{L})A_2G_2^k + \ldots + \prod_{i=1}^{p}(1 - G_i\mathbf{L})A_pG_p^k = 0.$$

Для доказательства использовался тот факт, что $\mathbf{L}G_j^k = G_j^{k-1}$ и поэтому

$$\prod_{i=1}^{p}(1 - G_i\mathbf{L})G_j^k = \prod_{i\neq j}(1 - G_i\mathbf{L})(1 - G_j\mathbf{L})G_j^k = 0.$$

Формулы для коэффициентов $A_1, \ldots, A_p$ можно получить из условий:

$$\rho_0 = 1,\ \rho_k = \rho_{-k},\ \text{откуда}$$
$$\sum_{i=1}^{p}A_i = 1 \quad \text{и} \quad \sum_{i=1}^{p}A_iG_i^k = \sum_{i=1}^{p}\frac{A_i}{G_i^k}, \quad k = 1, \ldots, p-1.$$

Другой способ состоит в том, чтобы вычислить $\rho_1, \ldots, \rho_{p-1}$ из уравнений Юла—Уокера (14.19), а затем составить на основе (14.30) при $k = 0, \ldots, p-1$ систему линейных уравнений, откуда и найти $A_1, \ldots, A_p$.

Если все корни характеристического уравнения удовлетворяют условию $|\lambda_i| > 1$, то $|G_i| < 1\ \forall i$ и все слагаемые в (14.30) затухают с ростом k. Если же для какого-то корня выполнено $|\lambda_i| < 1$, то (при условии, что $A_i \neq 0$) соответствующее слагаемое «уходит на бесконечность». Если $|\lambda_i| = 1$, то соответствующее слагаемое не затухает. Из этих рассуждений следует условие стационарности AR(p) — все корни соответствующего характеристического уравнения по модулю должны быть больше единицы.

Если корень $\lambda_i = G_i^{-1}$ действителен, элемент $A_i G_i^k$ в (14.30) убывает с ростом k экспоненциально, коль скоро $|\lambda_i| > 1$. Если же есть пара комплексно-сопряженных корней $\lambda_i = G_i^{-1}$, $\lambda_j = G_j^{-1}$, то соответствующие коэффициенты A_i, A_j также будут сопряженными и в составе автокорреляционной функции появится экспоненциально затухающая синусоида (см. вывод автокорреляционной функции процесса Юла).

Таким образом, из соотношения (14.30) следует, что в общем случае автокорреляционная функция стационарного процесса авторегрессии является комбинацией затухающих экспонент и затухающих синусоид.

Итак, мы вывели общий вид автокорреляционной функции стационарного процесса авторегрессии. Теоретически выборочная автокорреляционная функция может служить инструментом для распознавания авторегрессионого процесса. На практике же для коротких рядов различительная сила автокорреляционной функции не очень высока. Однако часто изучение автокорреляционной функции является хорошим заделом исследования системы.

Кроме автокорреляционной функции важным инструментом для распознавания типа процесса является его спектр.

Спектр стационарного процесса авторегрессии

В главе 13 мы определили спектральную плотность стационарного процесса как косинус-преобразование Фурье автоковариационной функции (13.29):

$$\mathbf{p}(f) = 2\Big(\gamma_0 + 2\sum_{k=1}^{\infty} \gamma_k \cos 2\pi f k\Big). \qquad (14.31)$$

Из этой общей формулы найдем спектральную плотность для **стационарного процесса Маркова** ($|\varphi_1| < 1$). Автоковариационная функция этого процесса имеет вид:

$$\gamma_k = \frac{\sigma_\varepsilon^2 \varphi_1^k}{1 - \varphi_1^2}.$$

Подставляя эти автоковариации в формулу (14.31), получим

$$\mathbf{p}(f) = \frac{2\sigma_\varepsilon^2}{1 - \varphi_1^2}\left(1 + 2\sum_{k=1}^{\infty} \varphi_1^k \cos 2\pi f k\right).$$

Воспользовавшись представлением косинуса через комплексную экспоненту (формулами Эйлера) — $2\cos 2\pi f k = e^{i2\pi f k} + e^{-i2\pi f k}$, после несложных преобра-

зований получим:

$$\mathbf{p}(f) = \frac{2\sigma_\varepsilon^2}{1-\varphi_1^2}\left(\sum_{k=0}^{\infty}\left(\varphi_1 e^{i2\pi f}\right)^k + \sum_{k=0}^{\infty}\left(\varphi_1 e^{-i2\pi f}\right)^k - 1\right),$$

т.е.

$$\mathbf{p}(f) = \frac{2\sigma_\varepsilon^2}{1-\varphi_1^2}\left(\sum_{k=0}^{\infty}(\varphi_1 z)^k + \sum_{k=0}^{\infty}(\varphi_1 z^{-1})^k - 1\right),$$

где мы ввели обозначение $z = e^{i2\pi f}$.

По формуле бесконечной геометрической прогрессии

$$\mathbf{p}(f) = \frac{2\sigma_\varepsilon^2}{1-\varphi_1^2}\left(\frac{1}{1-\varphi_1 z} + \frac{1}{1-\varphi_1 z^{-1}} - 1\right).$$

Произведение знаменателей двух дробей можно записать в разных формах:

$$(1-\varphi_1 z)(1-\varphi_1 z^{-1}) = 1 - \varphi_1(z+z^{-1}) + \varphi_1^2 = \\ = 1 - 2\varphi_1\cos 2\pi f + \varphi_1^2 = \left|1-\varphi_1 e^{-i2\pi f}\right|^2.$$

Приведя к общему знаменателю, получим

$$\mathbf{p}(f) = \frac{2\sigma_\varepsilon^2}{1-\varphi_1^2}\frac{1-\varphi_1 z^{-1}+1-\varphi_1 z-1+\varphi_1(z+z^{-1})-\varphi_1^2}{(1-\varphi_1 z)(1-\varphi_1 z^{-1})} = \\ = \frac{2\sigma_\varepsilon^2}{1-\varphi_1^2}\frac{1-\varphi_1^2}{(1-\varphi_1 z)(1-\varphi_1 z^{-1})}.$$

Таким образом, спектральная плотность марковского процесса равна

$$\mathbf{p}(f) = \frac{2\sigma_\varepsilon^2}{1-2\varphi_1\cos 2\pi f + \varphi_1^2},$$

или, в другой форме,

$$\mathbf{p}(f) = \frac{2\sigma_\varepsilon^2}{\left|1-\varphi_1 e^{-i2\pi f}\right|^2}.$$

Ошибку ε_t авторегрессии произвольного порядка AR(p) можно выразить в виде линейного фильтра от y_t:

$$\varepsilon_t = x_t - \sum_{j=1}^{p}\varphi_j x_{t-j},$$

поэтому для вычисления спектра авторегрессионного процесса можно воспользоваться общей формулой (14.4), характеризующей изменение спектра при применении линейного фильтра.

Спектральная плотность белого шума ε_t равна $2\sigma_\varepsilon^2$. Применение формулы (14.4) дает

$$2\sigma_\varepsilon^2 = \mathbf{p}(f)\Big|1 - \sum_{j=1}^{p} \varphi_j e^{-i2\pi fj}\Big|^2,$$

откуда

$$\mathbf{p}(f) = \frac{2\sigma_\varepsilon^2}{\left|1 - \varphi_1 e^{-i2\pi f} - \varphi_2 e^{-i4\pi f} - \cdots - \varphi_p e^{-i2p\pi f}\right|^2}.$$

В частном случае процесса Юла формула спектральной плотности имеет вид:

$$\begin{aligned}\mathbf{p}(f) &= \frac{2\sigma_\varepsilon^2}{\left|1 - \varphi_1 e^{-i2\pi f} - \varphi_2 e^{-i4\pi f}\right|^2} = \\ &= \frac{2\sigma_\varepsilon^2}{1 + \varphi_1^2 + \varphi_2^2 - 2\varphi_1(1 - \varphi_2)\cos 2\pi f - 2\varphi_2 \cos 4\pi f}.\end{aligned}$$

Разложение Вольда и условия стационарности процессов авторегрессии

Как уже говорилось, модель AR(p) можно записать в виде модели линейного фильтра:

$$x_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} = \boldsymbol{\psi}(\mathbf{L})\varepsilon_t,$$

где $\boldsymbol{\psi}(\mathbf{L}) = \boldsymbol{\varphi}^{-1}(\mathbf{L})$. Если процесс авторегрессии стационарен, то это разложение Вольда такого процесса.

Найдем коэффициенты модели линейного фильтра ψ_i процесса авторегрессии. Для этого в уравнении

$$\boldsymbol{\psi}(\mathbf{L})\boldsymbol{\varphi}(\mathbf{L}) = \Big(\sum_{i=0}^{\infty} \psi_i \mathbf{L}^i\Big)\Big(1 - \sum_{j=1}^{p} \varphi_j \mathbf{L}^j\Big) = 1$$

приравняем коэффициенты при одинаковых степенях $\mathbf{L}$. Получим следующие уравнения:

$$\begin{aligned}
\psi_0 &= 1,\\
\psi_1 - \psi_0\varphi_1 &= 0,\\
\psi_2 - \psi_1\varphi_1 - \psi_0\varphi_2 &= 0,\\
&\dots\\
\psi_p - \psi_{p-1}\varphi_1 - \dots - \psi_1\varphi_{p-1} - \psi_0\varphi_p &= 0,\\
\psi_{p+1} - \psi_p\varphi_1 - \dots - \psi_2\varphi_{p-1} - \psi_1\varphi_p &= 0,\\
&\dots
\end{aligned}$$

Общая рекуррентная формула имеет следующий вид:

$$\psi_i = \sum_{j=1}^{p} \varphi_j \psi_{i-j}, \quad i > 0, \tag{14.32}$$

где $\psi_0 = 1$ и $\psi_i = 0$ при $i < 0$.

Это разностное уравнение, которое фактически совпадает с уравнением для автокорреляций (14.18). Соответственно, если все корни характеристического уравнения различны, то общее решение такого уравнения такое же, как указано в (14.30), т.е.

$$\psi_i = B_1 G_1^i + B_2 G_2^i + \dots + B_p G_p^i, \quad i > -p,$$

где B_i — некоторые константы. Коэффициенты B_i можно вычислить, исходя из известных значений ψ_i при $i \leqslant 0$.

Очевидно, что если $|G_j| < 1 \ \forall j$, то все слагаемые здесь экспоненциально затухают, и поэтому ряд, составленный из коэффициентов ψ_i сходится абсолютно:

$$\sum_{i=0}^{\infty} |\psi_i| < \infty.$$

Таким образом, указанное условие гарантирует, что процесс авторегрессии является стационарным. Это дополняет вывод, полученный при анализе автокорреляционной функции.

Итак, условием стационарности процесса **AR(*p*)** является то, что корни λ_i характеристического уравнения лежат вне единичного круга на комплексной плоскости.

Для процесса **AR(2)** имеем два уравнения для коэффициентов B_1, B_2:

$$\begin{aligned}
\frac{B_1}{G_1} + \frac{B_2}{G_2} &= \psi_{-1} = 0,\\
B_1 + B_2 &= \psi_0 = 1,
\end{aligned}$$

откуда $B_1 = \frac{G_1}{G_1 - G_2}$ и $B_2 = -\frac{G_2}{G_1 - G_2}$. Таким образом, в случае процесса Юла имеем

$$\psi_i = \frac{G_1^{i+1} - G_2^{i+1}}{G_1 - G_2}, \quad i > -1,$$

и

$$x_t = \frac{1}{G_1 - G_2} \sum_{i=0}^{\infty} (G_1^{i+1} - G_2^{i+1}) \varepsilon_{t-i}.$$

Оценивание авторегрессий

Термин *авторегрессия* для обозначения модели (14.7) используется потому, что она фактически представляет собой модель регрессии, в которой регрессорами служат лаги изучаемого ряда x_t. По определению авторегрессии ошибки ε_t являются белым шумом и некоррелированы с лагами x_t. Таким образом, выполнены все основные предположения регрессионного анализа: ошибки имеют нулевое математическое ожидание, некоррелированы с регрессорами, не автокоррелированы и гомоскедастичны. Следовательно, модель (14.7) можно оценивать с помощью обычного метода наименьших квадратов.

Отметим, что при таком оценивании p начальных наблюдений теряются. Пусть имеется ряд $x_1, \ldots, x_T$. Тогда регрессия в матричной записи будет иметь следующий вид:

$$\begin{pmatrix} x_{p+1} \\ x_{p+2} \\ \vdots \\ x_T \end{pmatrix} = \begin{pmatrix} x_p & \cdots & x_1 \\ x_{p+1} & \cdots & x_2 \\ \vdots & & \vdots \\ x_{T-1} & \cdots & x_{T-p} \end{pmatrix} \begin{pmatrix} \varphi_1 \\ \vdots \\ \varphi_p \end{pmatrix} + \begin{pmatrix} \varepsilon_{p+1} \\ \varepsilon_{p+2} \\ \vdots \\ \varepsilon_T \end{pmatrix}.$$

Как видим, здесь используется $T - p$ наблюдений.

Оценки МНК параметров авторегрессии равны $\varphi = M^{-1}m$, где $\varphi = (\varphi_1, \varphi_2, \ldots, \varphi_p)'$, а матрицы M и m, как нетрудно увидеть, фактически состоят из выборочных автоковариаций ряда x_t. Отличие от стандартных выборочных автоковариаций состоит в том, что используются не все наблюдения.

Можно рассматривать данную регрессию как решение уравнений Юла—Уокера для автоковариаций (14.21) (или, что эквивалентно, уравнений для автокорреляций (14.19)), где теоретические автоковариации заменяются выборочными.

Действительно, уравнения Юла—Уокера (14.21) без первой строки записываются в виде

$$\gamma = \Gamma\varphi,$$

где

$$\gamma = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_p \end{pmatrix}, \ \Gamma = \begin{pmatrix} 1 & \gamma_1 & \gamma_2 & \cdots & \gamma_{p-1} \\ \gamma_1 & 1 & \gamma_1 & \cdots & \gamma_{p-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \gamma_{p-3} & \cdots & 1 \end{pmatrix},$$

откуда

$$\varphi = \Gamma^{-1}\gamma.$$

Замена теоретических значений автоковариаций γ_k выборочными автоковариациями c_k позволяет найти параметры процесса авторегрессии. Ясно, что при этом можно использовать и стандартные формулы выборочных автоковариаций. Тем самым, мы получаем еще один из возможных методов оценивания авторегрессий — метод моментов.

Частная автокорреляционная функция

Как мы видели, автокорреляционная функция процесса авторегрессии состоит из экспоненциально затухающих компонент. Такая характеристика не очень наглядна, поскольку соседние автокорреляции сильно связаны друг с другом, и, кроме того, для полного описания свойств ряда используется бесконечная последовательность автокорреляций.

Более наглядными характеристиками авторегрессии являются **частные автокорреляции**. Частная автокорреляция измеряет «чистую» корреляцию между уровнями временного ряда x_t и x_{t-k} при исключении опосредованного влияния промежуточных уровней ряда. Такой показатель корреляции между элементами ряда более информативен.

Пусть $\{x_t\}$ — произвольный стационарный ряд (не обязательно авторегрессия) и ρ_j — его автокорреляции. Применим к нему уравнения Юла—Уокера (14.19), как если бы процесс представлял собой авторегрессию k-го порядка, и найдем по автокорреляциям коэффициенты. Если обозначить j-й коэффициент уравнения авторегрессии порядка k через φ_{kj}, то уравнения Юла—Уокера

принимают вид:

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{k-2} \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{k-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \dots & 1 \end{bmatrix} \begin{bmatrix} \varphi_{k1} \\ \varphi_{k2} \\ \varphi_{k3} \\ \vdots \\ \varphi_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \vdots \\ \rho_k \end{bmatrix}.$$

Частная автокорреляция k-го порядка определяется как величина φ_{kk}, полученная из этих уравнений.

Решение этих уравнений соответственно для $k = 1,\ 2,\ 3$ дает следующие результаты (здесь используется правило Крамера):

$$\varphi_{11} = \rho_1, \qquad \varphi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}, \qquad \varphi_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}.$$

Частная автокорреляционная функция рассматривается как функция частной автокорреляции от задержки k, где $k = 1, 2, \dots$.

Для процесса авторегрессии порядка p частная автокорреляционная функция $\{\varphi_{kk}\}$ будет ненулевой для $k \leqslant p$ и равна нулю для $k > p$, то есть обрывается на задержке p.

Значение **выборочного** частного коэффициента автокорреляции φ_{kk} вычисляется как МНК-оценка последнего коэффициента в уравнении авторегрессии $\mathsf{AR}(k)$.

Частная автокорреляционная функция может оказаться полезной в решении задачи идентификации модели временного ряда: если она быстро затухает, то это авторегрессия, причем ее порядок следует выбрать по последнему большому значению частной автокорреляционной функции.

14.4. Процессы скользящего среднего

Другой частный случай модели линейного фильтра, широко распространенный в анализе временных рядов, — модель **скользящего среднего**, когда x_t линейно зависит от конечного числа q предыдущих значений ε:

$$x_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \ldots - \theta_q\varepsilon_{t-q}. \tag{14.33}$$

Модель скользящего среднего q-го порядка обозначают MA(q) (от английского *moving average*).

Данную модель можно записать и более сжато:

$$x_t = \boldsymbol{\theta}(\mathbf{L})\varepsilon_t,$$

через **оператор скользящего среднего**:

$$\boldsymbol{\theta}(L) = 1 - \theta_1\mathbf{L} - \theta_2\mathbf{L}^2 - \ldots - \theta_q\mathbf{L}^q. \tag{14.34}$$

Легко видеть, что процесс MA(q) является стационарным без каких-либо ограничений на параметры θ_j.

Действительно, математическое ожидание процесса

$$\mathbf{E}(x_t) = 0,$$

а дисперсия

$$\gamma_0 = (1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2)\sigma_\varepsilon^2,$$

т.е. равна дисперсии белого шума, умноженной на конечную величину $(1 + \theta_1^2 + + \theta_2^2 + \ldots + \theta_q^2)$.

Остальные моменты второго порядка (γ_k, ρ_k) также от времени не зависят.

Автоковариационная функция и спектр процесса MA(q)

Автоковариационная функция MA(q)

$$\gamma_k = \begin{cases} (-\theta_k + \theta_1\theta_{k+1} + \ldots + \theta_{q-k}\theta_q)\sigma_\varepsilon^2, & k = 1,\ 2,\ \ldots,\ q, \\ 0, & k > q. \end{cases} \tag{14.35}$$

В частном случае для MA(1) имеем:

$$\begin{aligned} \gamma_0 &= (1 + \theta_1^2)\sigma_\varepsilon^2, \\ \gamma_1 &= -\theta_1\sigma_\varepsilon^2, \\ \gamma_k &= 0, \quad k > 1, \end{aligned}$$

и автоковариационная матрица, соответствующая последовательности x_1, $x_2, \ldots, x_T$, будет иметь следующий трехдиагональный вид:

$$\Gamma = \sigma_\varepsilon^2 \begin{bmatrix} 1+\theta_1^2 & -\theta_1 & 0 & \cdots & 0 \\ -\theta_1 & 1+\theta_1^2 & -\theta_1 & \cdots & 0 \\ 0 & -\theta_1 & 1+\theta_1^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1+\theta_1^2 \end{bmatrix}.$$

В общем случае автоковариационная матрица процесса скользящего среднего порядка q имеет q ненулевых поддиагоналей и q ненулевых наддиагоналей, все же остальные элементы матрицы равны нулю.

Автокорреляционная функция имеет вид:

$$\rho_k = \begin{cases} \dfrac{-\theta_k + \theta_1\theta_{k+1} + \ldots + \theta_{q-k}\theta_q}{1+\theta_1^2+\ldots+\theta_q^2}, & k = 1, 2, \ldots, q, \\ 0, & k > q. \end{cases} \tag{14.36}$$

Таким образом, автокорреляционная функция процесса MA(q) обрывается на задержке q, и в этом отличительная особенность процессов скользящего среднего.

С другой стороны, частная автокорреляционная функция, в отличие от авторегрессий, не обрывается и затухает экспоненциально. Например, для MA(1) частная автокорреляционная функция имеет вид

$$-\theta_1^p \frac{1-\theta_1^2}{1-\theta_1^{2p+2}}.$$

Ясно, что модель скользящего среднего является частным случаем модели линейного фильтра (14.1), где $\psi_j = -\theta_j$ при $j = 1, \ldots, q$ и $\psi_j = 0$ при $j > q$. Фактически модель линейного фильтра является моделью MA(∞).

Формула спектра для процесса скользящего среднего следует из общей формулы для модели линейного фильтра (14.6):

$$\mathbf{p}(f) = 2\sigma_\varepsilon^2 \left| 1 - \theta_1 e^{-i2\pi f} - \theta_2 e^{-i4\pi f} - \ldots - \theta_q e^{-i2q\pi f} \right|^2.$$

Соответственно, для MA(1):

$$\mathbf{p}(f) = 2\sigma_\varepsilon^2 \left|1 - \theta_1 e^{-i2\pi f}\right|^2 = 2\sigma_\varepsilon^2 (1 + \theta_1^2 - 2\theta_1 \cos 2\pi f);$$

для MA(2):

$$\begin{aligned}\mathbf{p}(f) &= 2\sigma_\varepsilon^2 \left|1 - \theta_1 e^{-i2\pi f} - \theta_2 e^{-i4\pi f}\right|^2 = \\ &= 2\sigma_\varepsilon^2 \Big(1 + \theta_1^2 + \theta_2^2 - 2\theta_1(1 - \theta_2)\cos 2\pi f - 2\theta_2 \cos 4\pi f\Big).\end{aligned}$$

Обратимость процесса MA(q)

Авторегрессию, как мы видели выше, можно представить как MA(∞). С другой стороны, процесс скользящего среднего можно представить в виде AR(∞).

Рассмотрим, например, MA(1) (будем для упрощения писать θ вместо θ_1):

$$x_t = \varepsilon_t - \theta\varepsilon_{t-1}, \tag{14.37}$$

Сдвигом на один период назад получим $\varepsilon_{t-1} = x_{t-1} + \theta\varepsilon_{t-2}$ и подставим в (14.37):

$$x_t = \varepsilon_t - \theta x_{t-1} - \theta^2\varepsilon_{t-2}.$$

Далее, $\varepsilon_{t-2} = x_{t-2} + \theta\varepsilon_{t-3}$, поэтому

$$x_t = \varepsilon_t - \theta x_{t-1} - \theta^2 x_{t-2} - \theta^3\varepsilon_{t-3}.$$

Продолжая, получим на k-м шаге

$$x_t = \varepsilon_t - \theta x_{t-1} - \theta^2 x_{t-2} - \cdots - \theta^k x_{t-k} - \theta^{k+1}\varepsilon_{t-k-1}.$$

Если $|\theta| < 1$, то последнее слагаемое стремится к нулю при $k \to \infty$. Переходя к пределу, получаем представление AR(∞) для MA(1):

$$x_t = -\sum_{j=1}^{\infty} \theta^j x_{t-j} + \varepsilon_t. \tag{14.38}$$

С помощью лагового оператора можем записать это как

$$\pi(\mathbf{L})x_t = \varepsilon_t,$$

где

$$\pi(\mathbf{L}) = (1 - \theta\mathbf{L})^{-1} = \sum_{j=0}^{\infty} \theta^j \mathbf{L}^j.$$

В то время как процесс (14.37) стационарен при любом θ, процесс (14.38) стационарен только при $|\theta| < 1$. При $|\theta| \geqslant 1$ веса $-\theta^j$ в разложении (14.38) растут (при $|\theta| = 1$ не меняются) по абсолютной величине по мере увеличения j. Тем самым, нарушается разумная связь текущих событий с событиями в прошлом. Говорят, что при $|\theta| < 1$ процесс MA(1) является обратимым, а при $|\theta| \geqslant 1$ — необратимым.

В общем случае уравнение процесса MA(q) в обращенной форме можно записать как

$$\varepsilon_t = \boldsymbol{\theta}^{-1}(\mathbf{L})x_t = \pi(\mathbf{L})x_t = \sum_{j=0}^{\infty} \pi_j x_{t-j}.$$

Процесс MA(q) называется **обратимым**, если абсолютные значения весов π_j в обращенном разложении образуют сходящийся ряд. Стационарным процесс MA(q) является всегда, но для того, чтобы он обладал свойством обратимости, параметры процесса должны удовлетворять определенным ограничениям.

Выведем условия, которым должны удовлетворять параметры $\theta_1,\ \theta_2,\ \ldots,\ \theta_q$ процесса MA(q), чтобы этот процесс был обратимым.

Пусть $H_i^{-1},\ i = 1,\ \ldots,\ q$ — корни характеристического уравнения $\boldsymbol{\theta}(\mathbf{L}) = 0$ (будем предполагать, что они различны). Оператор скользящего среднего $\boldsymbol{\theta}(\mathbf{L})$ через обратные корни характеристического уравнения можно разложить на множители:

$$\boldsymbol{\theta}(\mathbf{L}) = \prod_{i=1}^{q}(1 - H_i\mathbf{L}).$$

Тогда обратный к $\boldsymbol{\theta}(\mathbf{L})$ оператор $\pi(\mathbf{L})$ можно представить в следующем виде:

$$\pi(\mathbf{L}) = \boldsymbol{\theta}^{-1}(\mathbf{L}) = \frac{1}{\prod\limits_{i=1}^{q}(1 - H_i\mathbf{L})} = \sum_{i=1}^{q} \frac{M_i}{1 - H_i\mathbf{L}}. \tag{14.39}$$

Каждое слагаемое (14.39) можно, по аналогии с MA(1), представить в виде бесконечного ряда:

$$\frac{M_i}{1 - H_i\mathbf{L}} = M_i \sum_{j=0}^{\infty} H_i^j \mathbf{L}^j, i = 1,\ \ldots,\ q,$$

который сходится, если $|H_i| < 1$.

Тогда процесс MA(q) в обращенном представлении выглядит как

$$\varepsilon_t = \sum_{i=1}^{q} M_i \sum_{j=0}^{\infty} H_i^j \mathbf{L}^j x_t,$$

и он стационарен, если корни характеристического уравнения $\boldsymbol{\theta}(\mathbf{L}) = 0$ лежат вне единичного круга. Иными словами, MA(q) обладает свойством обратимости, если для всех корней выполнено $|H_i^{-1}| > 1$, т.е. $|H_i| < 1\ \forall i$. Если же для одного из корней $|H_i| \geqslant 1,$ то ряд не будет сходиться, и процесс MA(q) будет необратимым.

Для каждого необратимого процесса MA(q), у которого корни характеристического уравнения не равны по модулю единице, существует неотличимый от него обратимый процесс того же порядка. Например, процесс MA(1) (14.37) с $|\theta| > 1$ можно записать в виде

$$x_t = \xi_t - \frac{1}{\theta}\xi_{t-1},$$

где $\xi_t = \dfrac{1 - \theta\mathbf{L}}{1 - 1/\theta \cdot \mathbf{L}}\varepsilon_t$ является белым шумом. Мы не будем доказывать, что ξ_t является белым шумом, поскольку это технически сложно. Вместо этого мы укажем на простой факт: пусть ξ_t — некоторый белый шум. Тогда процесс $\xi_t - \frac{1}{\theta}\xi_{t-1}$ имеет такую же автоковариационную функцию, как и процесс x_t, заданный уравнением (14.37), если дисперсии связаны соотношением $\sigma_\xi^2 = \theta^2\sigma_\varepsilon^2$. Для того чтобы в этом убедиться, достаточно проверить совпадение дисперсий и автоковариаций первого порядка (остальные автоковариации равны нулю).

В общем случае процесса MA(q), чтобы сделать его обратимым, требуется обратить все корни характеристического уравнения, которые по модулю меньше единицы. А именно, пусть $\boldsymbol{\theta}(\mathbf{L}) = \prod\limits_{i=1}^{q}(1 - H_i\mathbf{L}),$ — характеристический многочлен где $|H_i| < 1$ при $i = 1, \ldots, m$ и $|H_i| > 1$ при $i = m+1, \ldots, q$. Тогда

$$\tilde{\boldsymbol{\theta}}(\mathbf{L}) = \prod_{i=1}^{m}(1 - H_i\mathbf{L}) \prod_{i=m+1}^{q}(1 - H_i^{-1}\mathbf{L}) \tag{14.40}$$

— характеристический многочлен эквивалентного обратимого процесса.

Заметим, что хотя уравнение (14.33) по форме напоминает разложение Вольда процесса x_t, оно будет таким, только если все корни характеристического уравнения по модулю будут не меньше единицы. Для получения разложения Вольда произвольного процесса MA(q) требуется проделать описанную операцию обращения корней, которые по модулю меньше единицы.

14.5. Смешанные процессы авторегрессии — скользящего среднего ARMA (модель Бокса—Дженкинса)

На практике иногда бывает целесообразно ввести в модель как элементы авторегрессии, так и элементы скользящего среднего. Это делается для того, чтобы с использованием как можно меньшего числа параметров уловить характеристики исследуемого эмпирического ряда. Такой процесс называется **смешанным процессом авторегрессии — скользящего среднего** и обозначается ARMA(p, q):

$$x_t = \varphi_1 x_{t-1} + \ldots + \varphi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}, \qquad (14.41)$$

или, с использованием оператора лага,

$$(1 - \varphi_1 \mathbf{L} - \varphi_2 \mathbf{L}^2 - \ldots - \varphi_p \mathbf{L}^p) x_t = (1 - \theta_1 \mathbf{L} - \theta_2 \mathbf{L}^2 - \ldots - \theta_q \mathbf{L}^q) \varepsilon_t.$$

В операторной форме смешанная модель выглядит так:

$$\boldsymbol{\varphi}(\mathbf{L}) x_t = \boldsymbol{\theta}(\mathbf{L}) \varepsilon_t,$$

где $\boldsymbol{\varphi}(\mathbf{L})$ — оператор авторегрессии, $\boldsymbol{\theta}(\mathbf{L})$ — оператор скользящего среднего.

Модель (14.41) получила название **модели Бокса—Дженкинса**, поскольку была популяризирована Дж. Боксом и Г. Дженкинсом в их известной книге «Анализ временных рядов» [3]. Методология моделирования с помощью (14.41) получила название **методологии Бокса—Дженкинса**.

Автокорреляционная функция и спектр процесса ARMA(p, q)

Рассмотрим, как можно получить автоковариационную и автокорреляционную функции стационарного процесса ARMA(p, q), зная параметры этого процесса. Для этого умножим обе части уравнения (14.41) на x_{t-k}, где $k \geqslant 0$, и перейдем к математическим ожиданиям:

$$\begin{aligned}\mathbf{E}(x_{t-k} x_t) = \varphi_1 \mathbf{E}(x_{t-k} x_{t-1}) + \varphi_2 \mathbf{E}(x_{t-k} x_{t-2}) + \ldots + \varphi_p \mathbf{E}(x_{t-k} x_{t-p}) + \\ + \mathbf{E}(x_{t-k} \varepsilon_t) - \theta_1 \mathbf{E}(x_{t-k} \varepsilon_{t-1}) - \theta_2 \mathbf{E}(x_{t-k} \varepsilon_{t-2}) - \ldots - \theta_q \mathbf{E}(x_{t-k} \varepsilon_{t-q}).\end{aligned}$$

Обозначим через δ_s кросс-ковариацию изучаемого ряда x_t и ошибки ε_t с задержкой s, т.е.

$$\delta_s = \mathbf{E}(x_t \varepsilon_{t-s}).$$

Поскольку процесс стационарен, то эта кросс-ковариационная функция не зависит от момента времени t. В этих обозначениях

$$\mathbf{E}(x_{t-k}\varepsilon_{t-j}) = \delta_{j-k}.$$

Получаем выражение для автоковариационной функции:

$$\gamma_k = \varphi_1\gamma_{k-1} + \ldots + \varphi_p\gamma_{k-p} + \delta_{-k} - \theta_1\delta_{1-k} - \ldots - \theta_q\delta_{q-k}. \qquad (14.42)$$

Так как x_{t-k} зависит только от импульсов, которые произошли до момента $t-k$, то

$$\delta_{j-k} = \mathbf{E}(x_{t-k}\varepsilon_{t-j}) = 0 \quad \text{при } j < k.$$

Для того чтобы найти остальные нужные нам кросс-ковариации, $\delta_0, \ldots, \delta_q$, необходимо поочередно умножить все члены выражения (14.41) на ε_t, $\varepsilon_{t-1}, \ldots, \varepsilon_{t-q}$ и перейти к математическим ожиданиям. В итоге получится следующая система уравнений:

$$\begin{aligned}
\delta_0 &= \sigma_\varepsilon^2,\\
\delta_1 &= \varphi_1\delta_0 - \theta_1\sigma_\varepsilon^2,\\
\delta_2 &= \varphi_1\delta_1 + \varphi_2\delta_0 - \theta_2\sigma_\varepsilon^2,\\
&\ldots
\end{aligned}$$

Общая формула для всех $1 \leqslant s \leqslant p$ имеет вид:

$$\delta_s = \varphi_1\delta_{s-1} + \cdots + \varphi_s\delta_0 - \theta_s\sigma_\varepsilon^2.$$

При $s > p$ (такой случай может встретиться, если $p < q$)

$$\delta_s = \varphi_1\delta_{s-1} + \cdots + \varphi_p\delta_{s-p} - \theta_s\sigma_\varepsilon^2.$$

Отсюда рекуррентно, предполагая σ_ε^2 и параметры φ и θ известными, найдем δ_s.

Далее, зная δ_s, по аналогии с уравнениями Юла—Уокера (14.21) по формуле (14.42) при $k = 0, \ldots, p$ с учетом того, что $\gamma_{-k} = \gamma_k$ найдем автоковариации $\gamma_0, \ldots, \gamma_p$. Остальные автоковариации вычисляются рекуррентно по формуле (14.42).

Автокорреляции рассчитываются как $\rho_k = \gamma_k/\gamma_0$. Заметим, что если требуется найти только автокорреляции, то без потери общности можно взять ошибку ε_t с единичной дисперсией: $\sigma_\varepsilon^2 = 1$.

Если в уравнении (14.42) $k > q$, то все кросс-корреляции равны нулю, поэтому

$$\gamma_k = \varphi_1\gamma_{k-1} + \ldots + \varphi_p\gamma_{k-p}, \quad k > q. \qquad (14.43)$$

Поделив это выражение на γ_0, выводим уравнение автокорреляционной функции для $k > q$:

$$\rho_k = \varphi_1 \rho_{k-1} + \ldots + \varphi_p \rho_{k-p}, \tag{14.44}$$

или

$$\boldsymbol{\varphi}(\mathbf{L})\rho_k = 0, \quad k > q.$$

Таким образом, начиная с некоторой величины задержки, а точнее, когда $q < p$, поведение автокорреляционной функции стационарного процесса ARMA(p, q) определяется, как и в случае чистой авторегрессии AR(p), однородным конечно-разностным уравнением (14.44). В свою очередь, решение этого конечно-разностного уравнения определяется корнями характеристического уравнения $\boldsymbol{\varphi}(z) = 0$. То есть при $q < p$ автокорреляционная функция будет состоять из комбинации затухающих экспонент и экспоненциально затухающих синусоид[4].

По аналогии с AR(p) условия стационарности ARMA(p, q) определяются корнями характеристического уравнения $\boldsymbol{\varphi}(\mathbf{L}) = 0$: если эти корни лежат вне единичного круга, то процесс стационарен.

В качестве примера рассмотрим **процесс** ARMA(1, 1):

$$x_t = \varphi x_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}, \tag{14.45}$$

или через лаговый оператор:

$$(1 - \varphi \mathbf{L})x_t = (1 - \theta \mathbf{L})\varepsilon_t.$$

Процесс стационарен, если $-1 < \varphi < 1$, и обратим, если $-1 < \theta < 1$.

Для вывода формулы автокорреляционной функции умножим (14.45) на x_{t-k} и перейдем к математическим ожиданиям:

$$\mathbf{E}(x_{t-k}x_t) = \varphi \mathbf{E}(x_{t-k}x_{t-1}) + \mathbf{E}(x_{t-k}\varepsilon_t) - \theta \mathbf{E}(x_{t-k}\varepsilon_{t-1}),$$

или

$$\gamma_k = \varphi \gamma_{k-1} + \delta_{-k} - \theta \delta_{1-k}. \tag{14.46}$$

Исследуем поведение автоковариационной функции при различных значениях параметра k.

[4] Заметим, что граничные условия у AR(p) другие, поэтому автокорреляционные функции не будут совпадать.

При $k = 0$

$$\gamma_0 = \varphi\gamma_{-1} + \delta_0 - \theta\delta_1. \tag{14.47}$$

Чтобы найти второе слагаемое, умножим уравнение процесса (14.45) на ε_t и возьмем математическое ожидание:

$$\delta_0 = \mathbf{E}(x_t\varepsilon_t) = \varphi\mathbf{E}(x_{t-1}\varepsilon_t) + \mathbf{E}(\varepsilon_t\varepsilon_t) - \theta\mathbf{E}(\varepsilon_t\varepsilon_{t-1}) = \sigma_\varepsilon^2. \tag{14.48}$$

Аналогичным способом распишем $\mathbf{E}(x_t\varepsilon_{t-1})$:

$$\delta_1 = \mathbf{E}(x_t\varepsilon_{t-1}) = \varphi\mathbf{E}(x_{t-1}\varepsilon_{t-1}) + \mathbf{E}(\varepsilon_{t-1}\varepsilon_t) - \theta\mathbf{E}(\varepsilon_{t-1}\varepsilon_{t-1}) = (\varphi - \theta)\sigma_\varepsilon^2.$$

Равенство $\mathbf{E}(x_{t-1}\varepsilon_{t-1}) = \sigma_\varepsilon^2$ подтверждается так же, как (14.48).

Итак, принимая во внимание, что $\gamma_k = \gamma_{-k}$, выражение для дисперсии записывается как

$$\gamma_0 = \varphi\gamma_1 + \sigma_\varepsilon^2 - \theta(\varphi - \theta)\sigma_\varepsilon^2. \tag{14.49}$$

При $k = 1$ равенство (14.46) преобразуется в

$$\gamma_1 = \varphi\gamma_0 + \delta_{-1} - \theta\delta_0.$$

Используя ранее приведенные доводы относительно математических ожиданий, стоящих в этом уравнении, имеем:

$$\gamma_1 = \varphi\gamma_0 - \theta\sigma_\varepsilon^2. \tag{14.50}$$

При $k \geqslant 2$

$$\gamma_k = \varphi\gamma_{k-1}.$$

Выразим автоковариации в (14.49) и (14.50) через параметры модели φ и θ. Получим систему уравнений

$$\begin{cases} \gamma_0 = \varphi\gamma_1 + \sigma_\varepsilon^2 - \theta(\varphi - \theta)\sigma_\varepsilon^2; \\ \gamma_1 = \varphi\gamma_0 - \theta\sigma_\varepsilon^2; \end{cases}$$

и решим ее относительно γ_0 и γ_1.

Решение имеет вид:

$$\gamma_0 = \frac{1 - 2\varphi\theta + \theta^2}{1 - \varphi^2}\sigma_\varepsilon^2,$$
$$\gamma_1 = \frac{(\varphi - \theta)(1 - \varphi\theta)}{1 - \varphi^2}\sigma_\varepsilon^2.$$

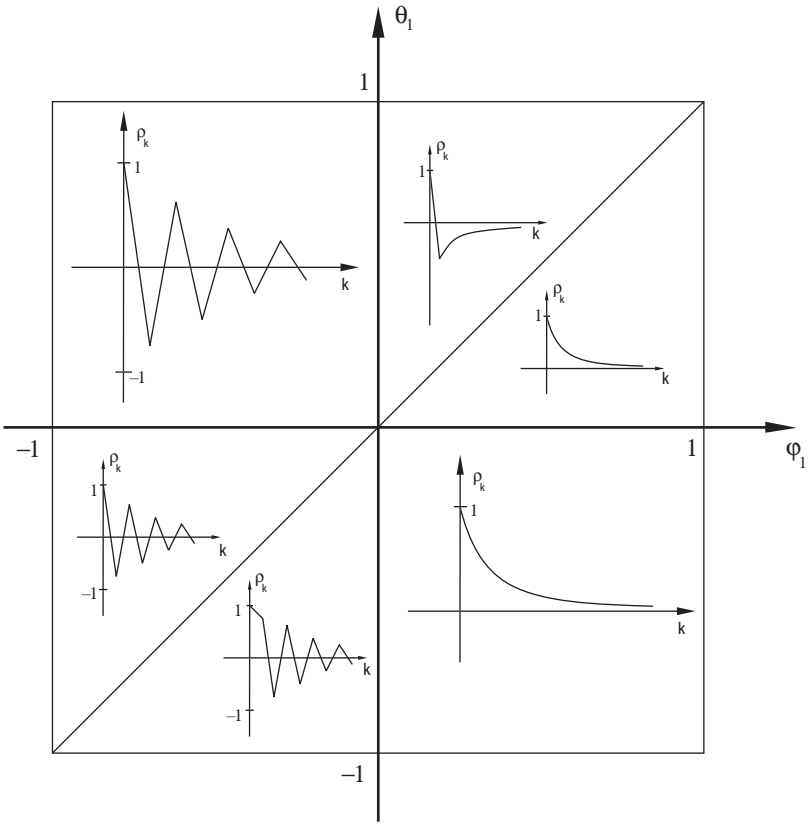

Рис. 14.4. График автокорреляционной функции процесса ARMA(1, 1)

С учетом того, что $\rho_k = \gamma_k/\gamma_0$, получаем выражения для автокорреляционной функции процесса ARMA(1, 1):

$$\begin{cases} \rho_1 = \dfrac{(\varphi-\theta)(1-\varphi\theta)}{1-2\varphi\theta+\theta^2}, \\ \rho_k = \varphi^{k-1}\rho_1, \quad k \geqslant 2. \end{cases}$$

На рисунке 14.4 изображены графики автокорреляционной функции процесса ARMA(1, 1) при различных сочетаниях значений параметров φ и θ.

По аналогии с процессами AR(p) и MA(q) выводится формула спектра процесса ARMA(p, q). Пусть $\{\eta_t\}$ — такой процесс, что

$$\eta_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \ldots - \theta_q\varepsilon_{t-q}.$$

Тогда x_t, описываемый уравнением (14.41), можно записать в виде

$$x_t = \varphi_1 x_{t-1} + \ldots + \varphi_p x_{t-p} + \eta_t.$$

По формуле (14.4), с одной стороны,

$$\mathbf{p}(f) = \mathbf{p}^x(f) = \mathbf{p}^\eta(f)\left|1 - \varphi_1 e^{-i2\pi f} - \varphi_2 e^{-i4\pi f} - \ldots - \varphi_p e^{-i2p\pi f}\right|^2,$$

а с другой стороны, для процесса скользящего среднего $\{\eta_t\}$ выполняется

$$\mathbf{p}^\eta(f) = \frac{2\sigma_\varepsilon^2}{\left|1 - \theta_1 e^{-i2\pi f} - \theta_2 e^{-i4\pi f} - \ldots - \theta_q e^{-i2q\pi f}\right|^2}.$$

Таким образом, получаем

$$\mathbf{p}(f) = 2\sigma_\varepsilon^2 \frac{\left|1 - \theta_1 e^{-i2\pi f} - \theta_2 e^{-i4\pi f} - \ldots - \theta_q e^{-i2q\pi f}\right|^2}{\left|1 - \varphi_1 e^{-i2\pi f} - \varphi_2 e^{-i4\pi f} - \ldots - \varphi_p e^{-i2p\pi f}\right|^2}. \qquad (14.51)$$

Представление процесса ARMA в виде MA(∞) и функция реакции на импульсы

Так же, как и в случае авторегрессии, стационарный процесс ARMA можно записать в виде модели линейного фильтра, или, другими словами, скользящего среднего бесконечного порядка MA(∞):

$$x_t = \frac{\boldsymbol{\theta}(\mathbf{L})}{\boldsymbol{\varphi}(\mathbf{L})}\varepsilon_t = \varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + \ldots = \sum_{i=0}^{\infty}\psi_i\varepsilon_{t-i} = \boldsymbol{\psi}(\mathbf{L})\varepsilon_t, \qquad (14.52)$$

где $\psi_0 = 1$.

Коэффициенты ψ_i представляют собой **функцию реакции на импульсы** для процесса ARMA, т.е. ψ_i является количественным измерителем того, как небольшое изменение («импульс») в ε_t влияет на x через i периодов, т.е. на x_{t+i}, что можно символически записать как

$$\psi_i = \frac{dx_{t+i}}{d\varepsilon_t}.$$

Один из способов вычисления функции реакции на импульсы сводится к использованию уравнения

$$\boldsymbol{\varphi}(z)\boldsymbol{\psi}(z) = \boldsymbol{\theta}(z),$$

или

$$(1 - \varphi_1 z - \varphi_2 z^2 - \ldots - \varphi_p z^p)(1 + \psi_1 z + \psi_2 z^2 + \ldots) = (1 - \theta_1 z - \ldots - \theta_q z^q),$$

из которого, приравнивая коэффициенты в левой и правой частях при одинаковых степенях z, можно получить выражения для ψ_i.

Более простой способ состоит в том, чтобы продифференцировать по ε_t уравнение ARMA-процесса, сдвинутое на i периодов вперед,

$$x_{t+i} = \sum_{j=1}^{p} \varphi_j x_{t+i-j} + \varepsilon_{t+i} - \sum_{j=1}^{q} \theta_j \varepsilon_{t+i-j},$$
$$\frac{dx_{t+i}}{d\varepsilon_t} = \sum_{j=1}^{p} \varphi_j \frac{dx_{t+i-j}}{d\varepsilon_t} - \theta_i,$$

где $\theta_0 = -1$ и $\theta_j = 0$ при $j > q$. Таким образом, получим рекуррентную формулу для $\psi_i = dx_{t+i}/d\varepsilon_t$:

$$\psi_i = \sum_{j=1}^{p} \varphi_j \psi_{i-j} - \theta_i.$$

При расчетах по этой формуле следует положить $\psi_0 = 1$ и $\psi_i = 0$ при $i < 0$.

Если процесс ARMA является обратимым[5], то полученное представление в виде MA(∞) является **разложением Вольда** этого процесса.

14.6. Модель авторегрессии — проинтегрированного скользящего среднего ARIMA

Характерной особенностью стационарных процессов типа ARMA(p, q) является то, что корни λ_i характеристического уравнения $\boldsymbol{\varphi}(\mathbf{L}) = 0$ находятся вне единичного круга. Если один или несколько корней лежат на единичной окружности или внутри нее, то процесс нестационарен.

Теоретически можно предложить много различных типов нестационарных моделей ARMA(p, q), однако, как показывает практика, наиболее распространенным типом нестационарных стохастических процессов являются **интегрированные процессы** или, как их еще называют, **процессы с единичным корнем**. Единичным называют корень характеристического уравнения, равный действительной единице: $\lambda_i = 1$.

[5] Разложение Вольда необратимого процесса, у которого некоторые корни характеристического уравнения по модулю больше единицы, такое же, как у эквивалентного обратимого процесса. Ошибки одноперiодных прогнозов, лежащие в основе разложения Вольда, при этом не будут совпадать с ошибками модели ε_t.

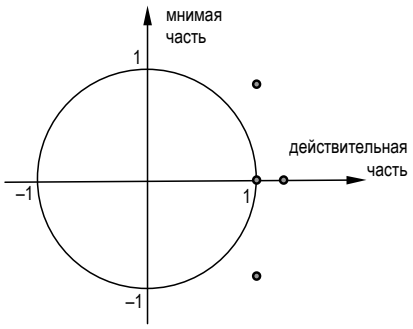

Рис. 14.5. Корни характеристического уравнения процесса
$x_t = 2.8x_{t-1} - 3.1x_{t-2} + 1.7x_{t-3} - 0.4x_{t-4} + \varepsilon_t + 0.5\varepsilon_{t-1} - 0.4\varepsilon_{t-2}$
на комплексной плоскости

Рассмотрим в качестве примера следующий процесс ARMA$(4,\ 2)$:

$$x_t = 2.8x_{t-1} - 3.1x_{t-2} + 1.7x_{t-3} - 0.4x_{t-4} + \varepsilon_t + 0.5\varepsilon_{t-1} - 0.4\varepsilon_{t-2}. \quad (14.53)$$

Характеристическое уравнение этого процесса имеет следующие корни: $\lambda_1 = 1 + \\ + i,\ \lambda_2 = 1 - i,\ \lambda_3 = 1.25,\ \lambda_4 = 1$. Все корни лежат за пределами единичного круга, кроме последнего, который является единичным. Эти корни изображены на рисунке 14.5.

Оператор авторегрессии этого процесса можно представить в следующем виде:

$$1 - 2.8\mathbf{L} + 3.1\mathbf{L}^2 - 1.7\mathbf{L}^3 + 0.4\mathbf{L}^4 = (1 - 1.8\mathbf{L} + 1.3\mathbf{L}^2 - 0.4\mathbf{L}^3)(1 - \mathbf{L}) = \\ = (1 - 1.8\mathbf{L} + 1.3\mathbf{L}^2 - 0.4\mathbf{L}^3)\boldsymbol{\Delta},$$

где $\boldsymbol{\Delta} = 1 - \mathbf{L}$ — оператор первой разности.

Введем обозначение $w_t = \boldsymbol{\Delta}x_t = x_t - x_{t-1}$. Полученный процесс $\{w_t\}$ является стационарным процессом ARMA$(3,\ 2)$, задаваемым уравнением:

$$w_t = 1.8w_{t-1} - 1.3w_{t-2} + 0.4w_{t-3} + \varepsilon_t + 0.5\varepsilon_{t-1} - 0.4\varepsilon_{t-2}. \quad (14.54)$$

В общем случае, если характеристическое уравнение процесса ARMA$(p + d,\ q)$ содержит d единичных корней, а все остальные корни по модулю больше единицы, то d-я разность этого временного ряда

$$w_t = \boldsymbol{\Delta}^d x_t = \boldsymbol{\varphi}(\mathbf{L})(1 - \mathbf{L})^d x_t$$

может быть представлена как стационарный процесс ARMA$(p,\ q)$:

$$\boldsymbol{\varphi}(\mathbf{L})\boldsymbol{\Delta}^d x_t = \boldsymbol{\theta}(\mathbf{L})\varepsilon_t \ \text{ или } \ \boldsymbol{\varphi}(\mathbf{L})w_t = \boldsymbol{\theta}(\mathbf{L})\varepsilon_t. \quad (14.55)$$

В развернутой форме модель 14.55 выглядит как

$$w_t = \varphi_1 w_{t-1} + \varphi_2 w_{t-2} + \ldots + \varphi_p w_{t-p} + \\ + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q}. \tag{14.56}$$

Из-за практического значения такую разновидность моделей ARMA выделяют в отдельный класс **моделей авторегрессии — проинтегрированного скользящего среднего** и обозначают ARIMA(p, d, q). При $d = 0$ модель описывает стационарный процесс. Как и исходную модель ARMA, модель ARIMA также называют моделью Бокса—Дженкинса.

Обозначив $\mathbf{f}(\mathbf{L}) = \boldsymbol{\theta}(\mathbf{L})(1 - \mathbf{L})^d$, представим процесс ARMA($p + d$, q) в виде:

$$\mathbf{f}(\mathbf{L})x_t = \boldsymbol{\theta}(\mathbf{L})\varepsilon_t.$$

$\mathbf{f}(\mathbf{L})$ называют обобщенным (нестационарным) оператором авторегрессии, таким, что d корней характеристического уравнения $\mathbf{f}(z) = 0$ равны единице, а остальные по модулю больше единицы. Такой процесс можно записать в виде модели ARIMA

$$\boldsymbol{\varphi}(\mathbf{L})(1 - \mathbf{L})^d x_t = \boldsymbol{\theta}(\mathbf{L})\varepsilon_t. \tag{14.57}$$

Ряд $\{x_t\}$ называют интегрированным, поскольку он является результатом применения к стационарному ряду $\{w_t\}$ операции кумулятивной (накопленной) суммы d раз. Так, если $d = 1$, то для $t > 0$

$$x_t = \sum_{i=1}^{t} w_i + x_0.$$

Этим объясняется название процесса авторегрессии — проинтегрированного скользящего среднего ARIMA(p, d, q).

Этот факт можно символически записать как

$$x_t = \boldsymbol{S}^d w_t,$$

где $\boldsymbol{S} = \boldsymbol{\Delta}^{-1} = (1 - \mathbf{L})^{-1}$ — оператор суммирования, обратный к оператору разности. Следует понимать, однако, что оператор $\boldsymbol{S}$ не определен однозначно, поскольку включает некоторую константу суммирования.

Простейшим процессом с единичным корнем является случайное блуждание:

$$x_t = \boldsymbol{S}\varepsilon_t,$$

где ε_t — белый шум.

14.7. Оценивание, распознавание и диагностика модели Бокса—Дженкинса

Для практического моделирования с использованием модели Бокса—Дженкинса требуется выбрать порядок модели (значения p, q и d), оценить ее параметры, а затем убедиться, правильно ли была выбрана модель и не нарушаются ли какие-либо предположения, лежащие в ее основе.

Заметим, что один и тот же процесс может быть описан разными моделями ARMA (14.41). Во-первых, неоднозначна компонента скользящего среднего $\boldsymbol{\theta}(\mathbf{L})\varepsilon_t$, о чем говорилось выше. Из разных возможных представлений MA здесь следует предпочесть обратимое. Во-вторых, характеристические многочлены авторегрессии и скользящего среднего могут содержать общие корни. Пусть $\boldsymbol{\varphi}(z)$ и $\boldsymbol{\theta}(z)$ содержат общий корень λ. Тогда характеристические многочлены можно представить в виде $\boldsymbol{\varphi}(z) = (1 - z/\lambda)\boldsymbol{\varphi}^*(z)$ и $\boldsymbol{\theta}(z) = (1 - z/\lambda)\boldsymbol{\theta}^*(z)$. Соответственно, один и тот же процесс можно записать как

$$\boldsymbol{\varphi}(\mathbf{L})x_t = \boldsymbol{\theta}(\mathbf{L})\varepsilon_t$$

или как

$$\boldsymbol{\varphi}^*(\mathbf{L})x_t = \boldsymbol{\theta}^*(\mathbf{L})\varepsilon_t.$$

Ясно, что вторая запись предпочтительнее, поскольку содержит меньше параметров. Указанные неоднозначности могут создавать проблемы при оценивании.

Прежде, чем рассмотреть оценивание, укажем, что уравнение (14.41) задает модель в довольно ограничительной форме. А именно, стационарный процесс, заданный уравнением (14.41), должен иметь нулевое математическое ожидание. Для того чтобы сделать математическое ожидание ненулевым, можно ввести в модель константу:

$$x_t = \mu + \varphi_1 x_{t-1} + \ldots + \varphi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}.$$

Если процесс $\{x_t\}$ стационарен, то

$$\mathbf{E}(x_t) = \frac{\mu}{1 - \varphi_1 - \cdots - \varphi_p}.$$

Альтернативно можно задать x_t как

$$x_t = \beta + w_t, \tag{14.58}$$

где ошибка $\{w_t\}$ является стационарным процессом ARMA:

$$w_t = \varphi_1 w_{t-1} + \ldots + \varphi_p w_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}. \tag{14.59}$$

При этом $\mathbf{E}(x_t) = \beta$. Ясно, что для стационарных процессов два подхода являются эквивалентными.

Последнюю модель можно развить, рассматривая регрессию

$$x_t = Z_t\alpha + w_t, \tag{14.60}$$

с ошибкой w_t в виде процесса (14.59). В этой регрессии Z_t не должны быть коррелированы с процессом w_t и его лагами. Составляющая Z_t может включать детерминированные тренды, сезонные переменные, фиктивные переменные для выбросов и т.п.

Метод моментов для оценивания параметров модели Бокса—Дженкинса

Опишем в общих чертах процедуру оценивания ARIMA(p, d, q). Предположим, что имеется ряд $x_1, \ldots, x_T$, по которому требуется оценить параметры процесса. Оценке подлежат три типа параметров: параметры детерминированной части модели (такие как β, α, о которых речь шла выше), авторегрессионные параметры φ и параметры скользящего среднего θ. При оценивании предполагается, что порядок разности d, порядок авторегрессии p и порядок скользящего среднего q заданы.

Если ряд $\{x_t\}$ описывается моделью (14.58) (которая предполагает $d = 0$), то параметр β этой модели можно оценить с помощью среднего $\bar{x}$, а далее действовать так, как если бы процесс сразу задавался моделью (14.59). В качестве w_t рассматриваются центрированные значения, полученные как отклонения исходных уровней временного ряда от их среднего значения: $w_t = x_t - \bar{x}$.

Если ряд $\{x_t\}$ описывается более общей моделью (14.60), которая тоже предполагает $d = 0$, то можно оценить параметры α с помощью обычного МНК, который дает здесь состоятельные, но не эффективные оценки a. Далее можно взять $w_t = x_t - Z_t a$ и действовать так, как если бы процесс задавался моделью (14.59).

При $d > 0$ от ряда x_t следует взять d-е разности: $w_t = \boldsymbol{\Delta}^d x_t$. Мы не будем рассматривать оценивание детерминированной составляющей в случае $d > 0$. Заметим только, что исходный ряд не нужно центрировать, поскольку уже первые разности исходных уровней ряда совпадают с первыми разностями центрированного ряда. Имеет смысл центрировать d-е разности $\boldsymbol{\Delta}^d x_t$.

Проведя предварительное преобразование ряда, мы сведем задачу к оцениванию стационарной модели ARMA (14.59), где моделируемая переменная w_t имеет нулевое математическое ожидание. Получив ряд $w_1, \ldots, w_T$ (при $d > 0$ ряд будет

на d элементов короче), можно приступить к оцениванию параметров авторегрессии и скользящего среднего.

Выше мы рассмотрели, как можно оценивать авторегрессии на основе уравнений Юла—Уокера (14.21). Прямое использование этого метода для модели ARMA(p, q) при $q > 0$ невозможно, поскольку в соответствующие уравнения будут входить кросс-ковариации между изучаемым процессом и ошибкой (см. 14.42). Однако можно избавиться от влияния элементов скользящего среднего, если сдвинуть уравнения на q значений вперед. Тогда уравнения для автокорреляций будут иметь вид (14.43). При $k = q + 1, \ldots, q + p$ получим следующую систему (т.е. используем здесь тот же подход, что и раньше: умножаем (14.59) на $w_{t-q-1}, \ldots, w_{t-q-p}$ и переходим к математическому ожиданию):

$$\begin{cases} \gamma_{q+1} = \varphi_1\gamma_q + \varphi_2\gamma_{q-1} + \ldots + \varphi_p\gamma_{q-p+1}, \\ \gamma_{q+2} = \varphi_1\gamma_{q+1} + \varphi_2\gamma_q + \ldots + \varphi_p\gamma_{q-p+2}, \\ \qquad \ldots \\ \gamma_{q+p} = \varphi_1\gamma_{q+p-1} + \varphi_2\gamma_{q+p-2} + \ldots + \varphi_p\gamma_q. \end{cases} \tag{14.61}$$

В итоге имеем систему, состоящую из p уравнений относительно p неизвестных параметров φ_j. Решение этих уравнений, в которых вместо γ_k берутся эмпирические значения автоковариаций c_k для последовательности значений $\{w_t\}$, т.е.

$$c_k = \frac{1}{T}\sum_{t=k+1}^{T} w_t w_{t-k},$$

дает нам оценки параметров $\varphi_1, \ldots, \varphi_p$ [6].

С помощью оценок авторегрессионных параметров можно, с учетом (14.59) построить новый временной ряд $\eta_{p+1}, \ldots, \eta_T$:

$$\eta_t = w_t - \varphi_1 w_{t-1} - \ldots - \varphi_p w_{t-p},$$

и для него рассчитать первые q выборочных автокорреляций $r_1^\eta, \ldots, r_q^\eta$. Полученные автокорреляции используются при расчете начальных оценок параметров скользящего среднего $\theta_1, \ldots, \theta_q$.

[6] На данный метод получения оценок параметров авторегрессии можно смотреть как на применение метода инструментальных переменных к уравнению регрессии:

$$w_t = \varphi_1 w_{t-1} + \ldots + \varphi_p w_{t-p} + \eta_t,$$

где ошибка η_t является MA(q) и поэтому коррелирована с лагами w_t только вплоть до q-го. В качестве инструментов здесь используются лаги $w_{t-q-1}, \ldots, w_{t-q-p}$.

Действительно, $\{\eta_t\}$ фактически представляет собой процесс скользящего среднего:

$$\eta_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \cdots - \theta_q\varepsilon_{t-q}, \tag{14.62}$$

для которого, как мы знаем, первые q автокорреляций могут быть выражены через параметры модели (см. (14.36)):

$$\rho_k^\eta = \frac{-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \ldots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \ldots + \theta_q^2}, \quad k = 1, \ldots, q.$$

Заменив в этих выражениях ρ_k^η на r_k^η, решаем полученную систему q нелинейных уравнений относительно q неизвестных параметров θ и получаем их оценки.

Поскольку система уравнений нелинейная, то могут возникнуть некоторые проблемы с ее решением. Во-первых, система может не иметь решений. Во-вторых, решение может быть не единственным.

Рассмотрим в качестве примера случай $q = 1$. При этом имеем одно уравнение с одним неизвестным:

$$r_1^\eta = \frac{-\theta_1}{1 + \theta_1^2}.$$

Максимальное по модулю значение правой части $1/2$ достигается при $\theta_1 = \pm 1$. Если $|r_1^\eta| > 1/2$, то уравнение не имеет действительного решения[7]. Если $|r_1^\eta| < 1/2$, то оценку θ_1 получим, решая квадратное уравнение. А оно будет иметь два корня:

$$\theta_1 = \frac{-1 \pm \sqrt{1 - 4(r_1^\eta)^2}}{2r_1^\eta}.$$

Один из корней по модулю больше единицы, а другой меньше, т.е. один соответствует обратимому процессу, а другой — необратимому.

Таким образом, из нескольких решений данных уравнений следует выбирать такие, которые соответствуют обратимому процессу скользящего среднего. Для этого, если некоторые из корней характеристического уравнения скользящего среднего по модулю окажутся больше единицы, то их следует обратить и получить коэффициенты, которые уже будут соответствовать обратимому процессу (см. 14.40).

Для $q > 1$ следует применить какую-либо итеративную процедуру решения нелинейных уравнений[8].

[7] Если решения не существует, то это может быть признаком того, что порядок разности d выбран неверно или порядок авторегрессии p выбран слишком низким.

[8] Например, метод Ньютона, состоящий в линеаризации нелинейных уравнений в точке текущих приближенных параметров (т.е. разложение в ряд Тейлора до линейных членов).

Описанный здесь метод моментов дает состоятельные, но не эффективные (не самые точные) оценки параметров. Существует ряд методов, позволяющих повысить эффективность оценок.

Методы уточнения оценок

Система (14.61) при $q > 1$ основана на уравнениях для автоковариаций, которые сдвинуты на q. Поскольку более дальние выборочные автоковариации вычисляются не очень точно, то это приводит к не очень точным оценкам параметров авторегрессии. Чтобы повысить точность, можно предложить следующий метод.

С помощью вычисленных оценок $\theta_1, \ldots, \theta_q$, на основе соотношения (14.62), находим последовательность значений $\{\varepsilon_t\}$ по рекуррентной формуле:

$$\varepsilon_t = \eta_t + \theta_1\varepsilon_{t-1} + \ldots + \theta_q\varepsilon_{t-q}.$$

В качестве ε_{t-j} при $t \leqslant j$ берем математическое ожидание ряда $\mathbf{E}(\varepsilon_t) = 0$.

Получив с помощью предварительных оценок φ и θ последовательность значений $\{\varepsilon_t\}$ и имея в наличии ряд $\{w_t\}$, методом наименьших квадратов находим уточненные оценки параметров модели (14.59), рассматривая ε_t в этом уравнении как ошибку.

Можно также получить уточненные оценки параметров детерминированной компоненты α в модели (14.60). Для этого можно использовать обобщенный метод наименьших квадратов (см. гл. 8), основанный на оценке ковариационной матрицы ошибок w_t, которую можно получить, имея некоторые состоятельные оценки параметров процесса ARMA.

Автоковариационную матрицу процесса ARMA можно представить в виде $\Gamma = \sigma_\varepsilon^2\Omega$. Оценку матрицы Ω можно получить, имея оценки параметров авторегрессии и скользящего среднего (см. выше вывод автоковариационной функции процесса ARMA). Имея оценку Ω, воспользуемся обобщенным МНК для оценивания параметров регрессии:

$$a_{\text{ОМНК}} = (Z'\Omega^{-1}Z)^{-1}Z'\Omega^{-1}X.$$

Можно использовать также автокорреляционную матрицу R:

$$a_{\text{ОМНК}} = (Z'R^{-1}Z)^{-1}Z'R^{-1}X.$$

В качестве примера приведем регрессию с процессом AR(1) в ошибке. Матрица автокорреляций для стационарного процесса AR(1), соответствующего последова-

тельности значений $w_1, \ldots, w_T$, имеет вид:

$$R = \begin{bmatrix} 1 & \varphi & \varphi^2 & \cdots & \varphi^{T-1} \\ \varphi & 1 & \varphi & \cdots & \varphi^{T-2} \\ \varphi^2 & \varphi & 1 & \cdots & \varphi^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \varphi^{T-1} & \varphi^{T-2} & \varphi^{T-3} & \cdots & 1 \end{bmatrix}.$$

Несложно убедиться, что обратная к R матрица имеет вид:

$$R^{-1} = \begin{bmatrix} 1 & -\varphi & 0 & \cdots & 0 & 0 \\ -\varphi & (1+\varphi^2) & -\varphi & \cdots & 0 & 0 \\ 0 & -\varphi & (1+\varphi^2) & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & (1+\varphi^2) & -\varphi \\ 0 & 0 & 0 & \cdots & -\varphi & 1 \end{bmatrix}.$$

Матрицу R^{-1} легко представить в виде произведения: $R^{-1} = D'D$, где

$$D = \begin{bmatrix} \sqrt{1-\varphi^2} & 0 & 0 & \cdots & 0 \\ -\varphi & 1 & 0 & \cdots & 0 \\ 0 & -\varphi & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Далее можем использовать полученную матрицу D для преобразования в пространстве наблюдений:

$$Z^* = DZ, \quad X^* = DX, \tag{14.63}$$

тогда полученные в преобразованной регрессии с помощью обычного МНК оценки будут оценками обобщенного МНК для исходной регрессии:

$$a_{\text{ОМНК}} = (Z^{*'}Z^*)^{-1}Z^{*'}X^*.$$

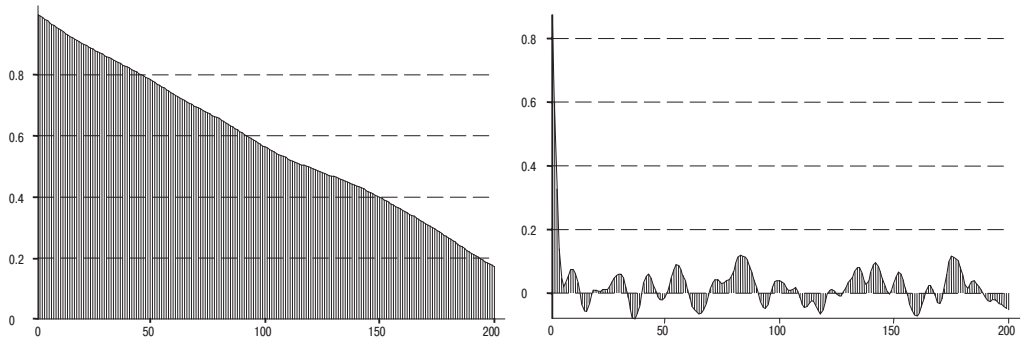

Рис. 14.6. Автокорреляционная функция интегрированного процесса (слева) и стационарного процесса (справа)

Они будут обладать не только свойством состоятельности, но и свойством эффективности.

К примеру, для парной регрессии $x_t = \alpha z_t + w_t$, где $w = \varphi w_{t-1} + \varepsilon_t$, преобразование (14.63) приводит к уравнениям

$$\sqrt{1-\varphi^2}x_1 = \alpha\sqrt{1-\varphi^2}z_1 + \varepsilon_1^*$$

и

$$x_t - \varphi x_{t-1} = \alpha(z_t - \varphi z_{t-1}) + \varepsilon_t^*, \quad t > 1,$$

для оценивания которых при данном φ применим обычный МНК.

После получения эффективных оценок параметров регрессии можно пересмотреть оценки параметров процесса ARMA. Можно продолжать такие итерации и далее до тех пор, пока не будет достигнута требуемая сходимость (см. метод Кочрена—Оркатта, описанный в п. 8.3).

Распознавание порядка модели

Сначала вычисляются разности исходного ряда до тех пор, пока они не окажутся стационарными относительно математического ожидания и дисперсии, и отсюда получают оценку d.

Если процесс является интегрированным, то его выборочная автокорреляционная функция затухает медленно, причем убывание почти линейное. Если же автокорреляционная функция затухает быстро, то это является признаком стационарности. В качестве примера на рисуке 14.6 слева изображена коррелограмма

ряда длиной в 1000 наблюдений, полученного по модели ARIMA(3, 1, 2), заданной уравнением (14.53). Справа на том же рисунке изображена коррелограмма первых разностей того же ряда, которые подчиняются стационарной модели ARMA(3, 2), заданной уравнением (14.54).

Формальные критерии выбора d рассматриваются в пункте 17.4.

Следует понимать, что достаточно сложно отличить процесс, который имеет единичный корень, от стационарного процесса, в котором есть корень близкий к единице.

Если d окажется меньше, чем требуется, то дальнейшее оценивание будет применяться к нестационарному процессу, что должно проявиться в оценках параметров авторегрессии — в сумме они будут близки к единице. Если d окажется больше, чем требуется, то возникнет эффект избыточного взятия разности (overdifferencing), который проявляется в том, что в характеристическом уравнении скользящего среднего появляется единичный корень. Это может создать трудности при оценивании скользящего среднего.

Для выбора порядка авторегрессии можно использовать выборочную частную автокорреляционную функцию. Как известно, теоретическая частная автокорреляционная функция процесса AR(p) обрывается на лаге p. Таким образом, p следует выбрать равным порядку, при котором наблюдается последнее достаточно большое (по модулю) значение выборочной частной автокорреляционной функции. Доверительные интервалы можно основывать на стандартной ошибке выборочного частного коэффициента автокорреляции, которая равна примерно $1/\sqrt{T}$ для порядков выше p (для которых теоретическая автокорреляционная функция равна нулю).

Аналогично, для выбора порядка скользящего среднего можно использовать выборочную автокорреляционную функцию, поскольку теоретическая автокорреляционная функция процесса MA(q) обрывается на лаге q. Таким образом, q следует выбрать равным порядку, при котором наблюдается последнее достаточно большое (по модулю) значение выборочной автокорреляционной функции. Стандартная ошибка выборочного коэффициента автокорреляции тоже примерно равна $1/\sqrt{T}$. Более точная формула стандартной ошибки для автокорреляции порядка k ($k > q$) имеет вид $\sqrt{\dfrac{T-k}{T(T+2)}}$ (см. стр. 367).

Если же нет уверенности, что процесс является чистой авторегрессией или чистым процессом скользящего среднего, то эти методы не подходят. Но, по крайней мере, по автокорреляционной функции и частной автокорреляционной функции можно проследить, насколько быстро угасает зависимость в ряде.

Порядок модели **ARMA**(p, q) можно выбирать на основе информационных критериев:

информационный критерий Акаике:

$$AIC = \ln(s_e^2) + \frac{2(p+q+n+1)}{T};$$

байесовский информационный критерий Шварца:

$$BIC = \ln(s_e^2) + \frac{(p+q+n+1)\ln T}{T}.$$

Здесь s_e^2 — остаточная дисперсия, рассчитанная по модели, $n+1$ относится к дополнительным оцениваемым параметрам — константе и коэффициентам при факторах регрессии (14.60). Порядок (p, q) выбирается посредством перебора из некоторого множества моделей так, чтобы информационный критерий достигал минимума. Критерий Акаике нацелен на повышение точности прогнозирования, а байесовский критерий — на максимизацию вероятности выбора истинного порядка модели.

Можно также выбирать порядок по тому принципу, что остатки должны быть похожи на белый шум, для чего использовать проверку остатков на автокорреляцию. Если остатки автокоррелированы, то следует увеличить p или q.

Диагностика

В основе модели **ARIMA** лежит предположение, что ошибки ε_t являются белым шумом. Это предполагает отсутствие автокорреляции и гомоскедастичность ошибок. Для проверки ошибок на гомоскедастичность могут использоваться те же критерии, которые были рассмотрены ранее в других главах. Здесь мы рассмотрим диагностику автокорреляции ошибок.

Простейший способ диагностики — графический, состоящий в изучении корелограммы и спектрограммы остатков. Кореллограмма должна показывать только малые, статистически незначимые значения автокорреляций. Спектрограмма должна быть достаточно «плоской», не иметь наклона и не содержать сильно выделяющихся пиков.

Для формальной проверки отсутствия автокорреляции ошибок можно использовать Q-статистику Бокса—Пирса и ее модификацию — статистику Льюнга—Бокса, которые основаны на квадратах нескольких (m) первых выборочных коэффициентов автокорреляции[9] (см. стр. 368).

[9] При выборе числа m следует помнить, что при малом m, если имеется автокорреляция высокого порядка, критерий может не показать автокорреляцию. При большом же m присутствие значитель-

Статистика Бокса—Пирса:

$$\tilde{Q} = n \sum_{k=1}^{m} r_k^2.$$

Статистика Льюнга—Бокса:

$$Q = n(n+2) \sum_{k=1}^{m} \frac{r_k^2}{n-k}.$$

Здесь в качестве r_k следует использовать выборочные коэффициенты автокорреляции, рассчитанные на основе остатков e_t модели ARIMA:

$$r_k = \frac{\sum\limits_{t=k+1}^{n} e_t e_{t-k}}{\sum\limits_{t=1}^{n} e_t^2}.$$

Поскольку для вычисления r_k используется не белый шум, а остатки, то асимптотическое распределение этих Q-статистик отличается от того, которое имеет место для истинного белого шума, на количество параметров авторегрессии и скользящего среднего, оцененных по модели, т.е. на величину $(p+q)$. Обе статистики асимптотически распределены как χ^2_{m-p-q}. Как показали Льюнг и Бокс, предложенная ими модифицированная Q-статистика, которая придает меньший вес дальним автокорреляциям, имеет распределение, которое ближе аппроксимирует свой асимптотический аналог, поэтому более предпочтительно использовать именно ее.

Нулевая гипотеза состоит в том, что ошибка представляет собой белый шум (автокорреляция отсутствует). Если Q-статистика превышает заданный квантиль распределения хи-квадрат, то нулевая гипотеза отвергается и делается вывод о том, что модель некорректна. Возможная причина некорректности — неудачный выбор порядка модели (слишком малые значения p и q).

14.8. Прогнозирование по модели Бокса—Дженкинса

Прогнозирование стационарного процесса ARMA

Пусть для стационарного обратимого процесса ARMA в момент t делается прогноз процесса x на τ шагов вперед, т.е. прогноз величины $x_{t+\tau}$. Для упрощения

ных автокорреляций может быть не замечено при наличии большого числа незначительных. То есть мощность критерия зависит от правильного выбора m.

рассуждений предположим, что при прогнозировании доступна вся информация о процессе x до момента t включительно, т.е. информация, на основе которой строится прогноз, совпадает с полной **предысторией** процесса

$$\Omega_t = (x_t, x_{t-1}, \dots).$$

Заметим, что на основе $(x_t, x_{t-1}, \dots)$ можно однозначно определить ошибки $(\varepsilon_t, \varepsilon_{t-1}, \dots)$ и наоборот, поэтому при сделанных предположениях ошибки $(\varepsilon_t, \varepsilon_{t-1}, \dots)$ фактически входят в информационное множество. Кроме того, имея полную предысторию, можно точно вычислить параметры процесса, поэтому будем далее исходить из того, что параметры процесса нам известны.

Из теории прогнозирования известно, что прогнозом, минимизирующим средний квадрат ошибки, будет математическое ожидание $x_{t+\tau}$, условное относительно Ω_t, т.е. $\mathbf{E}(x_{t+\tau}|\Omega_t)$. Убедимся в этом, воспользовавшись представлением модели ARMA в виде модели линейного фильтра (разложением Вольда) (14.52)

$$x_{t+\tau} = \varepsilon_{t+\tau} + \psi_1\varepsilon_{t+\tau-1} + \dots + \psi_{\tau-1}\varepsilon_{t+1} + \psi_\tau\varepsilon_t + \psi_{\tau+1}\varepsilon_{t-1} + \dots,$$

где во вторую строчку вынесены слагаемые, относящиеся к предыстории; получим таким образом следующее представление условного математического ожидания:

$$\begin{aligned}\mathbf{E}(x_{t+\tau}|\Omega_t) = \mathbf{E}(\varepsilon_{t+\tau}|\Omega_t) + \psi_1\mathbf{E}(\varepsilon_{t+\tau-1}|\Omega_t) + \dots + \\ + \psi_{\tau-1}\mathbf{E}(\varepsilon_{t+1}|\Omega_t) + \psi_\tau\varepsilon_t + \psi_{\tau+1}\varepsilon_{t-1} + \dots.\end{aligned}$$

Вторую строчку формулы пишем без оператора условного математического ожидания, поскольку соответствующие слагаемые входят в предысторию Ω_t.

Будем предполагать, что условное относительно предыстории математическое ожидание будущих ошибок равно нулю, т.е. $\mathbf{E}(\varepsilon_{t+k}|\Omega_t) = 0$ при $k > 0$. Это будет выполнено, например, если все ошибки ε_t независимы между собой. (Отсутствия автокорреляции тут недостаточно. В приложении приводится пример белого шума, для которого это неверно.) Тогда рассматриваемое выражение упрощается:

$$\mathbf{E}(x_{t+\tau}|\Omega_t) = \psi_\tau\varepsilon_t + \psi_{\tau+1}\varepsilon_{t-1} + \dots,$$

что дает нам линейную по ошибкам формулу для оптимального прогноза:

$$x_t(\tau) = \psi_\tau\varepsilon_t + \psi_{\tau+1}\varepsilon_{t-1} + \dots = \sum_{i=0}^{\infty}\psi_{\tau+i}\varepsilon_{t-i}, \qquad (14.64)$$

где мы обозначили через $x_t(\tau)$ прогноз на τ периодов, сделанный в момент t.

Проверим, что эта прогнозная функция будет оптимальной (в смысле минимума среднего квадрата ошибки) среди линейных прогнозных функций, т.е. среди прогнозных функций, представимых в виде линейной комбинации случайных ошибок, входящих в предысторию:

$$x_t(\tau) = \psi^*_\tau \varepsilon_t + \psi^*_{\tau+1}\varepsilon_{t-1} + \psi^*_{\tau+2}\varepsilon_{t-2} + \ldots = \sum_{i=0}^{\infty} \psi^*_{\tau+i}\varepsilon_{t-i}.$$

Для этого найдем веса $\psi^*_\tau, \psi^*_{\tau+1}, \psi^*_{\tau+2}, \ldots$, которые обеспечивают минимум среднего квадрата ошибки. С учетом того, что $x_{t+\tau} = \sum\limits_{i=0}^{\infty} \psi_i \varepsilon_{t+\tau-i}$, ошибка такого прогноза $\eta_t(\tau)$ равна

$$\begin{aligned}\eta_t(\tau) = x_{t+\tau} - x_t(\tau) &= \sum_{i=0}^{\infty} \psi_i \varepsilon_{t+\tau-i} - \sum_{i=0}^{\infty} \psi^*_{\tau+i}\varepsilon_{t-i} = \\ &= \sum_{i=0}^{\tau-1} \psi_i \varepsilon_{t+\tau-i} + \sum_{i=0}^{\infty} (\psi_{\tau+i} - \psi^*_{\tau+i})\varepsilon_{t-i},\end{aligned}$$

а средний квадрат ошибки прогноза (с учетом некоррелированности ошибок) равен

$$\begin{aligned}\mathbf{E}[\eta_t(\tau)^2] &= \mathbf{E}\left[\left(\sum_{i=0}^{\tau-1} \psi_i \varepsilon_{t+\tau-i} + \sum_{i=0}^{\infty} (\psi_{\tau+i} - \psi^*_{\tau+i})\varepsilon_{t-i}\right)^2\right] = \\ &= \mathbf{E}\left[\sum_{i=0}^{\tau-1} \psi_i^2 \varepsilon^2_{t+\tau-i} + \sum_{i=0}^{\infty} (\psi_{\tau+i} - \psi^*_{\tau+i})^2 \varepsilon^2_{t-i}\right] = \\ &= \sigma_\varepsilon^2 (1 + \psi_1^2 + \psi_2^2 + \ldots + \psi^2_{\tau-1}) + \sigma_\varepsilon^2 \sum_{i=0}^{\infty} (\psi_{\tau+i} - \psi^*_{\tau+i})^2.\end{aligned}$$

Очевидно, что средний квадрат ошибки достигает минимума при $\psi^*_{\tau+i} = \psi_{\tau+i}$ и равен

$$\mathbf{E}[\eta_t(\tau)^2] = \sigma_\varepsilon^2 \left[1 + \sum_{i=1}^{\tau-1} \psi_i^2\right].$$

Ошибка такого прогноза рассчитывается по формуле

$$\eta_t(\tau) = \sum_{i=0}^{\tau-1} \psi_i \varepsilon_{t+\tau-i}.$$

Из формулы видно, что эта ошибка проистекает из будущих ошибок ε_{t+k}, которые в момент t еще неизвестны. Беря математическое ожидание от обеих частей,

видим, что математическое ожидание ошибки прогноза равно нулю. Таким образом, прогноз, полученный по формуле (14.64), будет **несмещенным**.

Из несмещенности прогноза следует, что дисперсия ошибки прогноза равна среднему квадрату ошибки прогноза, т.е.

$$\sigma_p^2 = \mathbf{E}[\eta_t(\tau)^2] = \sigma_\varepsilon^2 \left[1 + \sum_{i=1}^{\tau-1} \psi_i^2\right].$$

или

$$\sigma_p^2 = \sigma_\varepsilon^2 \sum_{i=0}^{\tau-1} \psi_i^2,$$

где $\psi_0 = 1$.

Хотя представление в виде бесконечного скользящего среднего удобно для анализа прогнозирования, однако для вычисления прогноза предпочтительнее вернуться к исходному представлению модели ARMA в виде разностного уравнения (со сдвигом на τ периодов вперед):

$$\begin{aligned} x_{t+\tau} = \varphi_1 x_{t+\tau-1} + \ldots + \varphi_p x_{t+\tau-p} + \\ + \varepsilon_{t+\tau} - \theta_1 \varepsilon_{t+\tau-1} - \ldots - \theta_q \varepsilon_{t+\tau-q}. \end{aligned}$$

Возьмем от обеих частей уравнения условное относительно предыстории математическое ожидание:

$$\begin{aligned} x_t(\tau) = \mathbf{E}[x_{t+\tau}|\Omega_t] = \varphi_1 \mathbf{E}[x_{t+\tau-1}|\Omega_t] + \ldots + \varphi_p \mathbf{E}[x_{t+\tau-p}|\Omega_t] + \\ + \mathbf{E}[\varepsilon_{t+\tau}|\Omega_t] - \theta_1 \mathbf{E}[\varepsilon_{t+\tau-1}|\Omega_t] - \ldots - \theta_q \mathbf{E}[\varepsilon_{t+\tau-q}|\Omega_t]. \end{aligned}$$

Введем более компактные обозначения:

$$\mathbf{E}[x_{t+i}|\Omega_t] = \bar{x}_{t+i},\ \mathbf{E}[x_{t+\tau}|\Omega_i] = \bar{\varepsilon}_{t+i}.$$

В этих обозначениях

$$\begin{aligned} x_t(\tau) = \bar{x}_{t+\tau} = \varphi_1 \bar{x}_{t+\tau-1} + \ldots + \varphi_p \bar{x}_{t+\tau-p} + \\ + \bar{\varepsilon}_{t+\tau} - \theta_1 \bar{\varepsilon}_{t+\tau-1} - \ldots - \theta_q \bar{\varepsilon}_{t+\tau-q}. \end{aligned} \quad (14.65)$$

При вычислении входящих в эту формулу условных математических ожиданий используют следующие правила:

$$\bar{x}_{t+i} = \mathbf{E}[x_{t+i}|\Omega_t] = \begin{cases} x_{t+i}, & i \leqslant 0, \\ x_t(i), & i > 0, \end{cases}$$

$$\bar{\varepsilon}_{t+i} = \mathbf{E}[\varepsilon_{t+i}|\Omega_t] = \begin{cases} \varepsilon_{t+i}, & i \leqslant 0, \\ 0, & i > 0, \end{cases}$$

дающие удобную рекуррентную формулу для вычисления прогнозов.

Для вычисления показателя точности прогноза (дисперсии ошибки прогноза или, что в данном случае то же самое, поскольку прогноз несмещенный, среднего квадрата ошибки прогноза), удобно опять вернутся к представлению модели в виде бесконечного скользящего среднего. Как мы видели, дисперсия ошибки прогноза равна $\sigma_p^2 = \sigma_\varepsilon^2(1 + \sum_{i=1}^{\tau-1} \psi_i^2)$. Формулы для вычисления коэффициентов ψ_i скользящего среднего приведены на стр. 462.

Мы вывели формулы для расчета точечного прогноза по модели ARMA и дисперсии этого прогноза. Если дополнительно предположить, что ошибки ε_t подчиняются нормальному закону (т.е. представляют собой **гауссовский процесс**), то можно получить также интервальный прогноз. При этом предположении при известных значениях процесса до момента t распределение будущего значения процесса $x_{t+\tau}$ (т.е. условное распределение $x_{t+\tau}|\Omega_t$) также будет нормальным со средним значением $x_t(\tau)$ и дисперсией σ_p^2:

$$x_{t+\tau}|\Omega_t \sim N\left(x_t(\tau), \sigma_p^2\right).$$

Учитывая это, получаем доверительный интервал для $x_{t+\tau}$, т.е. интервальный прогноз:

$$\left[x_t(\tau) - \xi_{1-\alpha}\sigma_p,\ x_t(\tau) + \xi_{1-\alpha}\sigma_p\right],$$

или

$$\left[x_t(\tau) - \xi_{1-\alpha}\sigma_\varepsilon\sqrt{\sum_{i=0}^{\tau-1}\psi_i^2},\ x_t(\tau) + \xi_{1-\alpha}\sigma_\varepsilon\sqrt{\sum_{i=0}^{\tau-1}\psi_i^2}\right], \tag{14.66}$$

где $\xi_{1-\alpha}$ — двусторонний $(1-\alpha)$-квантиль стандартного нормального распределения. Это $(1-\alpha)\cdot 100$-процентный доверительный интервал.

Прогнозирование процесса ARIMA

Для прогнозирования процесса ARIMA(p, d, q) при $d > 0$ можно воспользоваться представлением его в виде ARMA($p + d$, q):

$$\mathbf{f}(\mathbf{L})x_t = \boldsymbol{\theta}(\mathbf{L})\varepsilon_t,$$

где

$$\mathbf{f}(\mathbf{L}) = 1 - f_1\mathbf{L} - f_2\mathbf{L}^2 - \ldots - f_{p+d}\mathbf{L}^{p+d} = \boldsymbol{\varphi}(\mathbf{L})(1 - \mathbf{L})^d,$$

а коэффициенты $f_1, f_2, \ldots, f_{p+d}$ выражаются через $\varphi_1, \varphi_2, \ldots, \varphi_p$. В развернутой записи

$$x_t = \sum_{j=1}^{p+d} f_j x_{t-j} + \varepsilon_t - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j}. \tag{14.67}$$

Например, модель ARIMA(1, 1, 1),

$$(1 - \varphi_1\mathbf{L})(1 - \mathbf{L})x_t = (1 - \theta_1\mathbf{L})\varepsilon_t,$$

можно записать в виде:

$$x_t = (1 + \varphi_1)x_{t-1} - \varphi_1 x_{t-2} + \varepsilon_t - \theta_1\varepsilon_{t-1} = f_1 x_{t-1} + f_2 x_{t-2} + \varepsilon_t - \theta_1\varepsilon_{t-1},$$

где $f_1 = 1 + \varphi_1,\ f_2 = -\varphi_1$.

При расчетах можно использовать те же приемы, что и выше для ARMA. Некоторую сложность представляет интерпретация ARIMA(p, d, q) в виде модели линейного фильтра, поскольку ряд модулей коэффициентов такого разложения является расходящимся. Однако это только технические сложности обоснования формул (в которые мы не будем вдаваться), а сами формулы фактически не меняются.

Таким образом, отвлекаясь от технических тонкостей, можем записать ARIMA(p, d, q) в виде MA(∞):

$$x_t = \frac{\boldsymbol{\theta}(\mathbf{L})}{\mathbf{f}(\mathbf{L})}\varepsilon_t = \varepsilon_t + \psi_1\varepsilon_{t-1} + \psi_2\varepsilon_{t-2} + \ldots = \boldsymbol{\psi}(\mathbf{L})\varepsilon_t. \tag{14.68}$$

Функцию реакции на импульсы можно рассчитать по рекуррентной формуле

$$\psi_i = \sum_{j=1}^{p+d} f_j \psi_{i-j} - \theta_i, \tag{14.69}$$

где $\psi_0 = 1$, $\psi_i = 0$ при $i < 0$ и $\theta_i = 0$ при $i > q$.

Кроме того, прогнозы $x_t(\tau)$ можно вычислять по рекуррентной формуле, которая получается из (14.67) по аналогии с (14.65):

$$x_t(\tau) = \bar{x}_{t+\tau} = \sum_{j=1}^{p+d} f_j \bar{x}_{t+\tau-j} + \bar{\varepsilon}_{t+\tau} - \sum_{j=1}^{q} \theta_j \bar{\varepsilon}_{t+\tau-j}, \tag{14.70}$$

где условные математические ожидания $\bar{x}_{t+i} = \mathbf{E}[x_{t+i}|\Omega_t]$ и $\bar{\varepsilon}_{t+i} = \mathbf{E}[\varepsilon_{t+i}|\Omega_t]$ рассчитываются по тому же принципу, что и в (14.65).

Таким образом, для прогнозирования в модели ARIMA можно использовать формулы (14.70), (14.8) и (14.66), где коэффициенты ψ_i рассчитываются в соответствии с (14.69).

Альтернативный подход к прогнозированию в модели ARIMA(p, d, q) состоит в том, чтобы сначала провести необходимые вычисления для $w_t = (1 - \mathbf{L})^d x_t$, т.е. процесса ARMA($p$, q), который лежит в основе прогнозируемого процесса ARIMA(p, d, q), а потом на их основе получить соответствующие показатели для x_t.

Так, (14.70) можно записать в виде

$$\mathbf{E}[\mathbf{f}(\mathbf{L})x_{t+\tau}|\Omega_t] = \mathbf{E}[\boldsymbol{\varphi}(\mathbf{L})(1 - \mathbf{L})^d x_{t+\tau}|\Omega_t] = \mathbf{E}[\boldsymbol{\theta}(\mathbf{L})\varepsilon_{t+\tau}|\Omega_t],$$

т.е.

$$\mathbf{E}[\boldsymbol{\varphi}(\mathbf{L})w_{t+\tau}|\Omega_t] = \mathbf{E}[\boldsymbol{\theta}(\mathbf{L})\varepsilon_{t+\tau}|\Omega_t]$$

или

$$\boldsymbol{\varphi}(\mathbf{L})\bar{w}_{t+\tau} = \boldsymbol{\theta}(\mathbf{L})\bar{\varepsilon}_{t+\tau}.$$

Здесь по аналогии $\bar{w}_{t+i} = \mathbf{E}[w_{t+i}|\Omega_t]$, причем $\bar{w}_{t+i} = w_t(i)$ (равно прогнозу) при $i > 0$ и $\bar{w}_{t+i} = w_{t+i}$ (равно значению самого ряда) при $i \leqslant 0$.

Отсюда видно, что можно получить сначала прогнозы для процесса w_t по формулам (14.65), заменив x_t на w_t, а затем применить к полученным прогнозам оператор $\boldsymbol{S}^d = (1 - \mathbf{L})^{-d}$, т.е. попросту говоря, просуммировать такой ряд d раз, добавляя каждый раз нужную константу суммирования. В частности, при $d = 1$ получаем

$$x_t(i) = x_t + \sum_{j=0}^{i} w_t(j).$$

Далее, $\boldsymbol{\psi}(\mathbf{L})$ можно записать в виде

$$\boldsymbol{\psi}(\mathbf{L}) = \frac{\boldsymbol{\theta}(\mathbf{L})}{\mathbf{f}(\mathbf{L})} = (1 - \mathbf{L})^{-d} \frac{\boldsymbol{\theta}(\mathbf{L})}{\boldsymbol{\varphi}(\mathbf{L})} = (1 - \mathbf{L})^{-d} \psi^w(\mathbf{L}) = \boldsymbol{S}^d \psi^w(\mathbf{L}).$$

Здесь $\psi^w(\mathbf{L}) = \boldsymbol{\theta}(\mathbf{L})/\boldsymbol{\varphi}(\mathbf{L})$ — оператор представления MA(∞) для w_t. Таким образом, коэффициенты ψ_i можно рассчитать из коэффициентов ψ_i^w. Например, при $d = 1$ получаем

$$\psi_i = \sum_{j=0}^{i} \psi_j^w.$$

Получение общих формул для прогнозирования в модели ARIMA с помощью решения разностных уравнений

Формула (14.70) представляет собой *разностное уравнение* для $\bar{x}_{t+\tau}$, решив которое получаем в явном виде общую формулу прогноза. Проиллюстрируем этот прием на примере процесса ARIMA(1, 1, 1), для которого $f_1 = 1 + \varphi_1,\ f_2 = -\varphi_1$:

$$x_{t+\tau} = (1 + \varphi_1)x_{t+\tau-1} - \varphi_1 x_{t+\tau-2} + \varepsilon_{t+\tau} - \theta_1\varepsilon_{t+\tau-1}. \qquad (14.71)$$

Берем условное математическое ожидание от обеих частей равенства (14.71), получаем точечные прогнозы на $1, 2, \ldots, \tau$ шагов вперед.

$$\begin{aligned}
x_t(1) &= (1 + \varphi_1)x_t - \varphi_1 x_{t-1} - \theta_1\varepsilon_t,\\
x_t(2) &= (1 + \varphi_1)x_t(1) - \varphi_1 x_t,\\
x_t(3) &= (1 + \varphi_1)x_t(2) - \varphi_1 x_t(1),\\
&\vdots\\
x_t(\tau) &= (1 + \varphi_1)x_t(\tau - 1) - \varphi_1 x_t(\tau - 2),\ \tau > 2.
\end{aligned}$$

Мы видим, что начиная с $\tau > q = 1$ природу прогнозирующей функции определяет только оператор авторегрессии:

$$\bar{x}_{t+\tau} = (1 + \varphi_1)\bar{x}_{t+\tau-1} - \varphi_1\bar{x}_{t+\tau-2},\ \tau > 1.$$

Общее решение этого разностного уравнения имеет следующий вид:

$$\bar{x}_{t+\tau} = A_0 + A_1\varphi_1^\tau.$$

Чтобы вычислить неизвестные коэффициенты, необходимо учесть, что $\bar{x}_t = x_t$ и $\bar{x}_{t+1} = x_t(1) = (1 + \varphi_1)x_t - \varphi_1 x_{t-1} - \theta_1\varepsilon_t$. Получается система уравнений:

$$\begin{aligned}
A_0 + A_1 &= x_t,\\
A_0 + A_1\varphi_1 &= (1 + \varphi_1)x_t - \varphi_1 x_{t-1} - \theta_1\varepsilon_t,
\end{aligned}$$

из которой находятся A_0 и A_1.

Точно так же можно рассматривать формулу (14.69) как разностное уравнение, решая которое относительно ψ_i, получим в явном виде общую формулу для функции реакции на импульсы. По формуле (14.69) получаем

$$\begin{aligned}
&\psi_0 = 1,\\
&\psi_1 = (1+\varphi_1)\psi_0 - \theta_1 = 1 + \varphi_1 - \theta_1,\\
&\psi_2 = (1+\varphi_1)\psi_1 - \varphi_1\psi_0,\\
&\quad\vdots\\
&\psi_i = (1+\varphi_1)\psi_{i-1} - \varphi_1\psi_{i-2},\ i > 1.
\end{aligned}$$

Легко показать, что решение этого разностного уравнения имеет следующий общий вид:

$$\psi_i = B_0 + B_1\varphi_1^i,$$

где $B_0 = \dfrac{1-\theta_1}{1-\varphi_1}$, $B_1 = 1 - B_0 = \dfrac{\theta_1 - \varphi_1}{1-\varphi_1}$. С учетом этого модель ARIMA(1, 1, 1) представляется в виде:

$$x_t = \sum_{i=0}^{\infty}\left(B_0 + B_1\varphi_1^i\right)\varepsilon_{t-i}.$$

Используя полученную формулу для коэффициентов ψ_i, найдем также дисперсию прогноза:

$$\sigma_p^2 = \sigma_\varepsilon^2\sum_{i=0}^{\tau-1}\psi_i^2 = \frac{\sigma_\varepsilon^2}{(1-\varphi_1)^2}\sum_{i=0}^{\tau-1}\left(1-\theta_1+(\theta_1-\varphi_1)\,\varphi_1^i\right)^2.$$

Отметим, что в пределе слагаемые стремятся к положительному числу $1-\theta_1$. Это означает, что с ростом горизонта прогноза τ дисперсия (а, следовательно, ширина прогнозного интервала) неограниченно возрастает. Такое поведение дисперсии связано с тем, что рассматриваемый процесс является нестационарным.

Прогнозирование по модели Бокса—Дженкинса в конечных выборках

Выше мы предполагали, что в момент t известна полная предыстория $\Omega_t = (x_t, x_{t-1}, \dots)$. Фактически, однако, человеку, производящему прогноз, известен только некоторый конечный ряд $(x_1, \dots, x_t)$. В связи с этим для практического использования приведенных формул, требуется внести в них определенные поправки.

В частности, параметры модели на практике не известны, и их требуется оценить. Это вносит дополнительную ошибку в прогноз.

Кроме того, ошибки ε_t, вообще говоря, неизвестны, и вместо них в выражении (14.65) следует использовать остатки e_t, полученные в результате оценивания модели. При наличии в модели скользящего среднего (т.е. при $q > 0$) ошибки не выражаются однозначно через наблюдаемый ряд $\{x_t\}$ и требуется использовать какое-то приближение. Наиболее простой метод состоит в том, чтобы положить остатки e_t при $t \leqslant 0$ равными нулю, а остальные остатки вычислять рекуррентно, пользуясь формулой

$$\varepsilon_t = x_t - \sum_{j=1}^{p+d} f_j x_{t-j} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j},$$

где вместо ошибок ε_t используются остатки e_t, а вместо неизвестных истинных параметров f_j и θ_j — их оценки.

Из-за того, что параметры не известны, а оцениваются, дисперсия ошибки прогноза будет выше, чем следует из (14.8). Имея некоторую оценку ковариационной матрицы оценок параметров, можно было бы внести приблизительную поправку, но эти расчеты являются достаточно громоздкими.

Далее, расчеты дисперсии прогноза с использованием (14.8) сами по себе являются приближенными, поскольку встречающиеся там величины приходится оценивать. Это относится и к функции реакции на импульсы ψ_i, и к дисперсии ошибки σ_ε^2.

Приложение. Неоптимальность линейных прогнозов в модели ARMA

То, что ошибка ε_t представляет собой белый шум (т.е. ошибки некоррелированы, имеют нулевое математическое ожидание и одинаковую дисперсию), не подразумевает, что $\mathbf{E}\,(\varepsilon_{t+k}|\Omega_t) = 0$ при $k > 0$. Поэтому в общем случае прогнозная функция, полученная по формуле (14.64) (или, эквивалентно, (14.65)) не будет оптимальной в среднеквадратическом смысле. Однако она, как было показано, является оптимальной среди *линейных* прогнозных функций. Кроме того, если ошибки независимы, то требуемое свойство выполнено. В частности, оно верно для гауссовского белого шума, т.к. для гауссовских процессов некоррелированность эквивалентна независимости.

Рассмотрим в качестве примера неоптимальности прогноза (14.64) следующий процесс $\{\varepsilon_t\}$. Значения, соответствующие четным t независимы и распределены как $N(0;1)$. При нечетных же t значения определяются по формуле

$$\varepsilon_t = \frac{\varepsilon_{t-1}^2 - 1}{\sqrt{2}}.$$

Таким образом, при нечетных t значения ряда полностью предопределены предысторией. Несложно проверить, что данный процесс представляет собой белый шум. Он имеет нулевое математическое ожидание, единичную дисперсию и не автокоррелирован. Если же построить на основе такого белого шума марковский процесс $x_t = \varphi_1 x_{t-1} + \varepsilon_t$, то при нечетных t оптимальным прогнозом на один шаг вперед будет не $\varphi_1 x_{t-1}$, а $\varphi_1 x_{t-1} + \left(\varepsilon_{t-1}^2 - 1\right)/\sqrt{2}$, причем прогноз будет безошибочным. При четных же t стандартная формула будет оптимальной.

Приведенный пример наводит на мысль о том, что во многих случаях можно подобрать нелинейную прогнозную функцию, которая позволяет сделать более точный прогноз, чем полученная нами оптимальная линейная прогнозная функция. В качестве менее экзотического примера можно привести модели с авторегрессионной условной гетероскедастичностью, о которых речь идет в одной из последующих глав.

14.9. Модели, содержащие стохастический тренд

Модели со стохастическим трендом можно отнести к классу линейных нестационарных моделей ARIMA(p, d, q), но они имеют свои особенности.

Рассмотрим эти модели.

1. **Модель случайного блуждания** (The Random Walk Model).

Эта модель является частным случаем модели AR(1) с единичным корнем:

$$x_t = x_{t-1} + \varepsilon_t. \tag{14.72}$$

Если начальное условие x_0 известно, общее решение может быть представлено в виде

$$x_t = x_0 + \sum_{i=1}^{t} \varepsilon_i.$$

Безусловное математическое ожидание: $\mathbf{E}(x_t) = \mathbf{E}(x_{t+k}) = x_0$.

Условное математическое ожидание:

$$\mathbf{E}(x_{t+k}|\Omega_t) = \mathbf{E}((x_t + \sum_{i=1}^{k} \varepsilon_{t+i})|\Omega_t) = x_t = x_0 + \sum_{i=1}^{t} \varepsilon_i.$$

Таким образом, условное математическое ожидание $\mathbf{E}(x_{t+k}|\Omega_t)$ обязательно включает в себя случайную компоненту, равную $\sum\limits_{i=1}^{t} \varepsilon_i$, которую называют **стохастическим трендом**.

Для любых значений k влияние каждой ошибки на последовательность $\{x_t\}$ со временем не исчезает.

Безусловная дисперсия: $\mathbf{var}(x_t) = t\sigma_\varepsilon^2,\ \mathbf{var}(x_{t+k}) = (t+k)\sigma_\varepsilon^2$.

Условная дисперсия: $\mathbf{var}(x_{t+k}|\Omega_t) = \mathbf{var}((x_t + \sum\limits_{i=1}^{k} \varepsilon_{t+i})|\Omega_t) = k\sigma_\varepsilon^2$.

Таким образом, и безусловная, и условная дисперсии зависят от времени, что свидетельствует о нестационарности процесса случайного блуждания.

Этот вывод подтверждается расчетом коэффициентов автоковариации и автокорреляции, которые также зависят от времени:

$$\gamma_k = \mathbf{cov}(x_t, x_{t+k}) = \mathbf{E}((x_t - x_0)(x_{t+k} - x_0)) = \\ = \mathbf{E}((\varepsilon_1 + \ldots + \varepsilon_t)(\varepsilon_1 + \ldots + \varepsilon_{t+k})) = \mathbf{E}(\varepsilon_1^2 + \ldots + \varepsilon_t^2) = t \cdot \sigma_\varepsilon^2.$$

Тогда

$$\rho_k = \frac{t\sigma_\varepsilon^2}{\sqrt{t\sigma_\varepsilon^2 (t+k)\sigma_\varepsilon^2}} = \sqrt{\frac{t}{t+k}}.$$

В практических ситуациях нередко модель случайного блуждания используется для описания динамики темпов роста.

В модели случайного блуждания первая разность $\Delta x_t = \varepsilon_t$ — чисто случайный процесс, следовательно, эта модель может быть интерпретирована как ARIMA(0, 1, 0).

2. **Модель случайного блуждания с дрейфом** (The Random Walk plus Drift Model).

Эта модель получается из модели случайного блуждания добавлением константы a_0:

$$x_t = x_{t-1} + a_0 + \varepsilon_t. \tag{14.73}$$

Общее решение для x_t при известном x_0:

$$x_t = x_0 + a_0 t + \sum_{i=1}^{t} \varepsilon_i.$$

Здесь поведение x_t определяется двумя нестационарными компонентами: линейным детерминированным трендом $a_0 t$ и стохастическим трендом $\sum\limits_{i=1}^{t} \varepsilon_i$.

Ясно, что динамику ряда определяет детерминированный тренд. Однако не следует думать, что всегда легко различить процесс случайного блуждания и процесс случайного блуждания с дрейфом.

На практике многие ряды, включая предложение денег и реальный ВНП, ведут себя как процесс случайного блуждания с дрейфом.

Заметим, что первая разность ряда стационарна, т.е. переход к первой разности создает стационарную последовательность: $\{\Delta x_t\} = \{a_0 + \varepsilon_t\}$ с математическим ожиданием, равным a_0, дисперсией σ_ε^2 и $\gamma_k = 0$ для всех t, следовательно, это тоже ARIMA(0, 1, 0).

3. **Модель случайного блуждания с шумом** (The Random Walk plus Noise Model).

Эта модель представляет собой совокупность стохастического тренда и компоненты белого шума. Формально модель описывается двумя уравнениями:

$$\begin{cases} x_t = \mu_t + \eta_t, \\ \mu_t = \mu_{t-1} + \varepsilon_t, \end{cases} \tag{14.74}$$

где $\{\eta_t\}$ — белый шум с распределением $N(0, \sigma_\eta^2)$, ε_t и η_t независимо распределены для всех t и k: $\mathbf{E}(\varepsilon_t, \eta_{t-k}) = 0$.

Общее решение системы (14.74) имеет вид:

$$x_t = \mu_0 + \sum_{i=1}^{t} \varepsilon_i + \eta_t.$$

Легко убедиться в том, что все моменты второго порядка зависят от времени:

$$\mathbf{var}(x_t) = t\sigma_\varepsilon^2 + \sigma_\eta^2,$$
$$\gamma_k = \mathbf{cov}(x_t, x_{t+k}) = \mathbf{E}((\varepsilon_1 + \ldots + \varepsilon_t + \eta_t)(\varepsilon_1 + \ldots + \varepsilon_{t-k} + \eta_t)) = t\sigma_\varepsilon^2,$$
$$\rho_k = \frac{t\,\sigma_\varepsilon^2}{\sqrt{(t\sigma_\varepsilon^2 + \sigma_\eta^2)\,((t+k)\,\sigma_\varepsilon^2 + \sigma_\eta^2)}}.$$

Следовательно, процесс 14.74 нестационарен. Но первая разность этого процесса $\Delta x_t = \varepsilon_t + \Delta\eta_t$ стационарна с параметрами:

$$\mathbf{E}(\Delta x_t) = \mathbf{E}(\varepsilon_t + \Delta\eta_t) = 0,$$
$$\mathbf{var}(\Delta x_t) = \mathbf{E}((\Delta x_t)^2) = \mathbf{E}((\varepsilon_t + \Delta\eta_t)^2) =$$
$$= \sigma_\varepsilon^2 + 2\mathbf{E}(\varepsilon_t\Delta\eta_t) + \mathbf{E}(\eta_t^2 - 2\eta_t\eta_{t-1} + \eta_{t-1}^2) = \sigma_\varepsilon^2 + 2\sigma_\eta^2,$$
$$\gamma_1 = \mathbf{cov}(\Delta x_t, \Delta x_{t-1}) = \mathbf{E}((\varepsilon_t + \eta_t - \eta_{t-1})(\varepsilon_{t-1} + \eta_{t-1} - \eta_{t-2})) = -\sigma_\eta^2,$$
$$\gamma_k = \mathbf{cov}(\Delta x_t, \Delta x_{t-k}) = \mathbf{E}((\varepsilon_t + \eta_t - \eta_{t-1})(\varepsilon_{t-k} + \eta_{t-k} - \eta_{t-k-1})) = 0, \quad k > 1.$$

Таким образом, первые разности ведут себя как MA(1)-процесс, а модель случайного блуждания с шумом можно квалифицировать как ARIMA(0, 1, 1).

4. **Модель общего тренда с нерегулярностью** (The General Trend plus Irregular Model).

Эта модель содержит детерменированный и стохастический тренды, а также $\mathsf{MA}(q)$-ошибку. Частный ее вариант:

$$\begin{cases} x_t = \mu_t + \eta_t, \\ \mu_t = \mu_{t-1} + a_0 + \varepsilon_t. \end{cases} \tag{14.75}$$

Решением (14.75) является модель общего тренда $a_0 t + \sum\limits_{i=1}^{t} \varepsilon_i$ с шумом:

$$x_t = \mu_0 + a_0 t + \sum_{i=1}^{t} \varepsilon_i + \eta_t.$$

Первая разность этой модели отличается от предыдущего варианта на константу a_0: $\Delta x_t = a_0 + \varepsilon_t + \Delta\eta_t$. Поэтому

$$\begin{aligned} &\mathbf{E}(\Delta x_t) = a_0, \\ &\mathbf{var}(\Delta x_t) = \sigma_\varepsilon^2 + 2\sigma_\eta^2, \\ &\gamma_1 = -\sigma_\eta^2, \\ &\gamma_k = 0, \quad k > 1. \end{aligned}$$

Следовательно, модель общего тренда с шумом — это также $\mathsf{ARIMA}(0,\,1,\,1)$.

В более общей постановке эта модель формулируется при помощи оператора $\boldsymbol{\theta}(\mathbf{L})$:

$$x_t = \mu_0 + a_0 t + \sum_{i=1}^{t} \varepsilon_i + \boldsymbol{\theta}(\mathbf{L})\eta_t.$$

5. **Модель локального линейного тренда** (The Local Linear Trend Model).

Пусть $\{\varepsilon_t\}$, $\{\eta_t\}$ и $\{\delta_t\}$ — три взаимно некоррелированных процесса белого шума. Тогда модель представляется следующими уравнениями:

$$\begin{cases} x_t = \mu_t + \eta_t, \\ \mu_t = \mu_{t-1} + a_t + \varepsilon_t, \\ a_t = a_{t-1} + \delta_t. \end{cases} \tag{14.76}$$

Легко показать, что рассмотренные ранее модели являются частными случаями данной модели.

Для нахождения решения выражаем a_t из последнего уравнения системы (14.76):

$$a_t = a_0 + \sum_{i=1}^{t} \delta_i.$$

Этот результат используется для преобразования μ_t:

$$\mu_t = \mu_{t-1} + a_0 + \sum_{i=1}^{t} \delta_i + \varepsilon_t.$$

Далее,

$$\mu_t = \mu_0 + \sum_{i=1}^{t} \varepsilon_i + a_0 t + \sum_{j=0}^{t-1} (t-j)\delta_{j+1}.$$

Наконец, находим решение для x_t:

$$x_t = \mu_0 + \sum_{i=1}^{t} \varepsilon_i + a_0 t + \sum_{j=0}^{t-1} (t-j)\delta_{j+1} + \eta_t.$$

Каждый элемент в последовательности $\{x_t\}$ содержит детерминированный тренд, причем весьма специфического вида, стохастический тренд и шум η_t.

Модель локального линейного тренда ведет себя как ARIMA(0, 2, 2). Действительно, первые разности процесса $\Delta x_t = a_t + \varepsilon_t + \Delta\eta_t$ нестационарны, поскольку a_t — процесс случайного блуждания. Однако, вторая разность

$$\Delta^2 x_t = \delta_t + \Delta\varepsilon_t + \Delta^2\eta_t$$

уже стационарна и имеет с параметры:

$$\begin{aligned}
&\mathbf{E}(\Delta^2 x_t) = 0,\\
&\mathbf{var}(\Delta^2 x_t) = \sigma_\delta^2 + \sigma_\varepsilon^2 + 6\sigma_\eta^2,\\
&\gamma_1 = -\sigma_\varepsilon^2 - 4\sigma_\eta^2,\\
&\gamma_2 = \sigma_\eta^2.
\end{aligned}$$

Все остальные коэффициенты автоковариации γ_k для $k > 2$ равны нулю.

14.10. Упражнения и задачи

Упражнение 1

Сгенерируйте ряд длиной 4000 наблюдений по модели AR(1) с параметром $\varphi = 0.5$ и нормально распределенной неавтокоррелированной ошибкой с единичной дисперсией, предполагая что значение ряда в момент $t = 0$ равно нулю. В действительности вид модели неизвестен, а задан только ряд.

1.1. Разбейте ряд на 200 непересекающихся интервалов по 20 наблюдений. По каждому из них с помощью МНК оцените модель AR(1). Проанализируйте распределение полученных оценок авторегрессионного параметра. Насколько велика дисперсия оценок и есть ли значимое смещение по сравнению с истинным параметром?

1.2. Рассмотрите построение прогноза на 1 шаг вперед с помощью трех моделей: AR(1), AR(2) и модели линейного тренда. Для этого ряд следует разбить на 200 непересекающихся интервалов по 20 наблюдений. По каждому из этих интервалов с помощью МНК необходимо оценить каждую из трех моделей, построить прогноз и найти ошибку прогноза. Сравните среднеквадратические ошибки прогноза по трем моделям и сделайте выводы.

Упражнение 2

Сгенерируйте 200 рядов длиной 20 наблюдений по модели AR(1) с авторегрессионным параметром $\varphi_1 = (k-1)/200,\ k = 1, \ldots, 200$, с нормально распределенной неавтокоррелированной ошибкой и единичной дисперсией. По каждому ряду с помощью МНК оцените модель AR(1). Постройте график отклонения оценки от истинного значения параметра в зависимости от истинного значения параметра φ_1. Что можно сказать по этому графику о поведении смещения оценок в зависимости от φ_1? Подтверждается ли, что оценки смещены в сторону нуля и смещение тем больше, чем φ_1 ближе к единице?

Упражнение 3

Сгенерируйте 200 рядов длиной 20 наблюдений по модели MA(2) с параметрами $\theta_1 = 0.5,\ \theta_2 = 0.3$ и нормально распределенной неавтокоррелированной ошибкой с единичной дисперсией.

3.1. Для каждого из рядов постройте выборочную автокорреляционную функцию для лагов 1, 2, 3. Рассмотрите распределение коэффициентов автокорреляции и сравните их с теоретическими значениями.

3.2. По каждому ряду на основе выборочных коэффициентов автокорреляции оцените модель MA(2), выбирая то решение квадратного уравнения, которое соответствует условиям обратимости процесса. Рассмотрите распределение оценок и сравните их с истинными значениями.

Упражнение 4

Имеется информация о реальных доходностях ценных бумаг для трех фирм Blaster, Mitre, и Celgene (дневные доходности, приведенные к годовым). (См. табл. 14.1 `http://oll.temple.edu/economics/hwkdata/univariate_TS/univariate.htm`.)

4.1. Изобразите график ряда для каждой из фирм и кратко охарактеризуйте свойства ряда.

4.2. Для каждой из фирм посчитайте среднее по двум разным непересекающимся подпериодам. Проверьте гипотезу равенства двух средних, используя простой t-критерий.

4.3. Для каждой из фирм посчитайте дисперсию по тем же двум подпериодам. Проверьте гипотезу равенства дисперсий, используя F-критерий. Какой вывод можно сделать относительно стационарности рядов?

4.4. Для каждой из фирм рассчитайте выборочные автоковариации, автокорреляции и выборочные частные автокорреляции для лагов $0, \ldots, 6$. Постройте соответствующие графики. Оцените значения параметров p и q модели ARMA(p, q).

4.5. Оцените параметры φ и θ модели ARMA(p, q) для каждой фирмы.

4.6. Постройте прогнозы доходности ценных бумаг на основе полученных моделей на 6 дней вперед.

Упражнение 5

Для данных по производству природного газа в СССР (табл. 12.2, с. 403) постройте модель ARIMA(p, d, q) и оцените доверительный интервал для прогноза на 5 шагов вперед.

Задачи

1. Записать с использованием лагового оператора случайный процесс:

   а) $x_t = \mu + \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \ldots + \varphi_p x_{t-p} + \varepsilon_t$;

Таблица 14.1

t	Blaster	Mitre	Celgene	t	Blaster	Mitre	Celgene	t	Blaster	Mitre	Celgene
1	0.41634	0.12014	0.37495	43	2.26050	0.04130	1.26855	85	1.86053	1.29274	2.52889
2	−0.23801	−0.05406	−0.18396	44	−2.67700	1.04649	−1.18533	86	−0.71307	0.01583	−0.79911
3	0.29481	−0.32615	0.15691	45	3.59910	1.50975	1.83224	87	1.43504	2.43129	0.86014
4	−0.61411	−0.29108	−0.40341	46	−1.91408	−0.78975	−0.08880	88	1.13397	0.92309	1.99122
5	0.27774	−0.02705	0.05401	47	1.41616	0.89511	−0.57678	89	0.06334	−1.09645	−0.48509
6	−0.40301	−0.78572	−0.11704	48	−0.13194	0.49100	1.61395	90	−0.42419	0.20130	−0.32999
7	−0.45766	−1.29175	−0.75328	49	0.30387	−0.62728	−0.74057	91	0.56689	−1.83314	0.76937
8	−0.89423	−2.42880	−0.73129	50	−0.40587	−0.53830	0.09482	92	−2.57400	−1.88158	−2.43071
9	−1.88832	−1.67366	−1.96585	51	−0.31838	1.17944	−0.52098	93	0.59467	1.17998	0.05017
10	−0.28478	2.05054	−0.33561	52	1.42911	0.42852	1.59066	94	−0.09736	−0.17633	0.85066
11	1.58336	0.13731	1.85731	53	−1.06378	−0.62532	−0.95433	95	−0.32876	−0.36017	−1.22773
12	−1.55354	−1.62557	−1.79765	54	0.83204	−1.31001	0.25028	96	0.31690	−0.82044	0.56133
13	0.47464	0.40286	−0.12857	55	−2.19379	−1.15293	−1.40754	97	−1.38083	−1.70314	−1.28727
14	−0.35276	−0.60609	0.58827	56	0.77155	−0.56983	−0.01986	98	−0.36896	1.23095	−0.68313
15	−0.54645	0.32509	−1.19108	57	−1.72097	0.12615	−0.77365	99	1.17509	0.94729	1.62377
16	1.00044	−0.02063	1.21914	58	1.49679	−0.37488	0.48386	100	−0.47758	1.29398	−0.64384
17	−1.26072	0.59998	−1.11213	59	−1.90651	0.63942	−0.99107	101	2.24964	0.14180	1.81862
18	2.05555	1.71220	1.51049	60	2.46331	0.47653	1.36935	102	−1.87715	−0.80889	−1.17413
19	−0.37029	1.48972	0.48379	61	−1.94601	0.52649	−0.69769	103	1.42413	0.46915	0.43877
20	2.20692	2.90700	1.21797	62	2.61428	2.10024	1.16649	104	−1.04233	1.22177	0.25152
21	1.26647	1.98764	2.19226	63	−0.29189	0.75862	1.24826	105	2.09249	1.09135	0.99786
22	1.25773	1.16628	0.60026	64	1.28001	0.44524	−0.08931	106	−0.65041	0.10892	0.29152
23	0.88074	−1.06153	1.25668	65	−0.12871	−1.52804	0.85085	107	1.02974	−1.06781	0.05320
24	−1.52961	−1.80661	−1.52435	66	−1.39518	−1.76352	−1.92776	108	−1.75288	−0.89211	−0.86576
25	−0.04273	1.19855	−0.15779	67	−0.25838	−0.16521	−0.01712	109	0.65799	0.73063	−0.05176
26	0.72479	0.76704	1.33383	68	−0.58997	0.26832	−0.30668	110	−0.17125	−0.07063	0.67682
27	−0.18366	0.18001	−0.56059	69	0.57022	−0.32595	0.19657	111	−0.08333	0.45717	−0.79937
28	0.79879	0.50357	0.61117	70	−0.90955	0.05597	−0.70997	112	0.81080	0.32701	1.08553
29	−0.21699	1.20339	0.12349	71	0.97552	0.40011	0.58198	113	−0.58788	−0.50112	−0.54259
30	1.55485	1.29821	1.23577	72	−0.64304	−0.27051	−0.11007	114	0.33261	1.75660	0.02756
31	−0.01895	0.57772	0.41794	73	0.42530	0.30712	−0.17654	115	1.37591	0.46400	1.79959
32	0.99008	0.30803	0.46562	74	−0.02964	−0.43812	0.48921	116	−0.84495	−2.36130	−0.98938
33	−0.32137	−0.55553	0.21106	75	−0.50316	−0.19260	−0.82995	117	−0.99043	−0.78306	−1.35391
34	−0.12852	−1.63234	−0.49739	76	0.41743	0.35469	0.50887	118	−0.04455	−0.13031	0.46131
35	−1.43837	−0.79997	−1.10015	77	−0.26740	−0.43595	−0.10447	119	−0.70314	−1.05764	−0.71196
36	0.28452	0.83711	0.05172	78	−0.10744	−0.69389	−0.39020	120	−0.43426	−0.38007	−0.74377
37	0.14292	1.01700	0.54283	79	−0.56917	0.77134	−0.37322	121	−0.13906	0.42198	0.04100
38	0.78978	0.93925	0.33560	80	1.14361	0.19675	1.03231	122	0.27676	−1.59661	0.20112
39	0.48850	−0.73179	0.65711	81	−0.99591	0.28117	−0.74445	123	−1.86932	−0.04115	−1.89508
40	−0.96077	0.31793	−1.09358	82	1.49172	0.62375	0.92118	124	1.74776	2.69062	1.46584
41	1.45039	−1.34070	1.42896	83	−0.83109	−0.96831	−0.06863	125	0.52476	1.04090	1.30779
42	−2.99638	−0.75896	−2.47727	84	−0.00917	1.77019	−0.81888	126	0.86322	0.23965	−0.09381

Таблица 14.1. (продолжение)

t	Blaster	Mitre	Celgene	t	Blaster	Mitre	Celgene	t	Blaster	Mitre	Celgene
127	0.12222	1.31876	0.60471	169	−2.20714	−0.41948	−0.27731	211	1.51485	−0.97684	0.86781
128	1.30642	0.45303	1.17204	170	1.83600	−1.04210	−0.00317	212	−2.50036	−1.13180	−1.74068
129	−0.57192	0.41682	−0.26719	171	−2.62281	0.90594	−0.89549	213	1.32618	1.48218	0.27384
130	1.30327	0.32936	0.87774	172	3.12742	−0.71167	1.64780	214	−0.16750	0.39183	1.11966
131	−0.84416	0.61935	−0.20724	173	−3.79724	−0.02254	−2.18906	215	0.42028	−0.52533	−0.69835
132	1.58641	0.83562	0.93209	174	3.80089	2.35129	1.77382	216	−0.42438	−1.01260	0.10685
133	−0.58774	0.34732	0.17484	175	−1.54768	−0.64010	0.83427	217	−0.74802	−0.83307	−0.98493
134	1.12133	−0.03793	0.34328	176	1.05327	−1.19450	−1.29993	218	−0.18508	−0.07511	−0.02618
135	−0.88277	−0.54302	−0.15044	177	−1.57945	0.55914	0.11559	219	−0.27858	0.72752	−0.21877
136	0.33158	−1.70839	−0.28336	178	1.54446	0.18847	0.55672	220	0.87498	−1.35822	0.70596
137	−2.03067	−1.22414	−1.43274	179	−1.28825	0.66119	−0.37065	221	−2.18781	−1.93248	−2.04429
138	0.43818	−0.54955	−0.13700	180	1.94968	0.46034	0.92503	222	0.29517	−0.01706	−0.23267
139	−1.44659	−0.53705	−0.71974	181	−1.29186	−0.82241	−0.26353	223	−0.87564	−2.02272	0.01847
140	0.56263	1.64035	−0.26807	182	0.58699	−1.93992	−0.48556	224	−1.55568	0.90288	−2.35370
141	0.96118	1.70204	1.61264	183	−2.43099	−0.43524	−1.43990	225	2.33976	1.99827	2.58159
142	0.71138	0.30461	0.25218	184	1.49486	−0.15310	0.67209	226	−0.86069	−0.33414	−0.42717
143	0.25957	−0.42989	0.26638	185	−2.02672	−0.86983	−1.00252	227	1.13389	−0.46117	−0.00198
144	−0.39911	1.64975	−0.31866	186	1.01176	0.74648	−0.28958	228	−1.28321	1.37416	−0.22666
145	2.09642	2.59804	2.14297	187	−0.32973	0.50541	0.83129	229	2.46195	0.98181	1.71017
146	0.55396	2.44417	0.91428	188	0.60712	−1.01920	−0.30054	230	−1.33630	0.18633	−0.40385
147	2.59219	0.87112	1.93566	189	−1.24396	−0.47570	−0.74958	231	1.88226	0.65077	0.67102
148	−0.92353	−1.59130	−0.23407	190	0.61988	0.73345	0.21236	232	−0.96236	0.01818	0.31244
149	−0.06896	−2.00737	−0.80695	191	−0.09182	1.10966	0.49188	233	0.99833	0.56680	−0.07017
150	−1.90109	−0.21639	−0.97295	192	1.14494	−0.35472	0.60138	234	−0.15850	0.72097	0.75775
151	1.04222	0.08106	0.52037	193	−1.15962	−0.25419	−0.80978	235	0.79860	0.01263	0.17952
152	−1.32252	−1.41365	−0.69199	194	1.05295	0.23071	0.57739	236	−0.44636	0.10951	−0.02671
153	−0.15447	0.10208	−1.09443	195	−0.92113	−0.21711	−0.17697	237	0.60990	−0.43066	0.26746
154	0.16722	0.11579	0.87557	196	0.66513	0.87219	−0.07976	238	−0.97024	−0.33060	−0.56602
155	−0.45657	−0.63291	−0.78207	197	0.28363	1.51583	0.92934	239	0.58530	1.47322	0.16698
156	0.00584	−0.89310	−0.08703	198	1.19450	0.91161	0.75373	240	0.78037	−0.52826	1.28774
157	−1.05101	0.07757	−0.86295	199	0.17008	0.90071	0.41868	241	−1.25522	−0.88925	−1.62637
158	0.93651	−0.18388	0.70155	200	1.08470	−0.51061	0.82681	242	0.72438	−1.90852	0.57099
159	−1.32255	−1.33106	−0.91757	201	−1.27643	−0.82689	−0.90190	243	−2.97849	−1.06136	−2.35532
160	0.01609	0.35961	−0.66348	202	0.57728	0.16717	0.14775	244	1.61219	1.26234	0.73721
161	0.15497	3.33824	0.82107	203	−0.61943	0.62127	0.07797	245	−0.92336	−0.22703	0.30842
162	2.87330	0.51674	2.49646	204	1.06910	1.37518	0.45461	246	0.60912	1.03409	−0.76861
163	−1.82484	0.42367	−1.53503	205	0.47147	−0.10956	0.94669	247	0.78275	−0.54114	1.64521
164	3.02300	0.22201	2.08618	206	−0.36583	1.29469	−0.80073	248	−1.45383	−0.91160	−1.83561
165	−2.73530	0.34709	−1.13867	207	2.02680	1.27776	2.17034	249	0.90928	2.84407	0.73683
166	3.23531	0.78318	1.51341	208	−0.75375	−0.42283	−0.35325	250	1.53899	1.62233	2.23506
167	−2.32656	0.76416	−0.36499	209	0.86800	−0.58282	0.10153				
168	3.15157	0.54297	1.16592	210	−1.27776	0.43351	−0.41637				

б) $x_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \ldots + \theta_q\varepsilon_{t-q}$;

в) $x_t = \mu + \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \ldots + \varphi_p x_{t-p} + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \ldots + \theta_q\varepsilon_{t-q}$.

2. Вывести формулы для вычисления математического ожидания, дисперсии и ковариаций случайного процесса $x_t = \mu + \varepsilon_t + \sum\limits_{i=1}^{\infty} \varphi_i\varepsilon_{t-i}$ при условии его слабой стационарности, если ε_t — белый шум с дисперсией σ^2 и математического ожидания $\mathbf{E}(\varepsilon_t x_{t-k}) = 0,\ \forall\ |k| \geqslant 1$.

3. Вывести формулы для вычисления математического ожидания, дисперсии и ковариаций случайного процесса $x_t = \mu + \varphi_1 x_{t-1} + \varepsilon_t$ при условии его слабой стационарности, если ε_t — белый шум с дисперсией σ^2.

4. Обосновать утверждение о том, что модель авторегрессии является частным случаем модели линейного фильтра.

5. Записать случайный процесс $x_t = 0.2 + 0.6x_{t-1} + \varepsilon_t$ с использованием лагового оператора и виде процесса скользящего среднего бесконечного порядка.

6. При каких условиях процесс AR(1) стационарен и обратим?

7. Задана модель: $x_t = 0.25x_{t-1} + \varepsilon_t$, где ε_t — белый шум. Дисперсия процесса x_t равна единице. Вычислить дисперсию белого шума.

8. Чему равна дисперсия Марковского процесса $x_t = 0.5x_{t-1} + \varepsilon_t$, если дисперсия белого шума равна 1? Изобразить график автокорреляционной функции данного процесса.

9. Для процесса $x_t = -0.7 - 0.7x_{t-1} + \varepsilon_t$, где ε_t — белый шум, рассчитать частную автокорреляционную функцию, вычислить первые 6 значений автокорреляционной функции и начертить ее график.

10. Для модели AR(1): $x_t = \mu + 0.5x_{t-1} + \varepsilon_t$ показать, что частная автокорреляционная функция $\varphi_{1,1} = 0.5,\ \varphi_{k,k} = 0$ при $k \geqslant 2$.

11. Даны два марковских процесса:

    $x_t = 0.5x_{t-1} + \varepsilon_t;\ y_t = 0.2y_{t-1} + \varepsilon_t$.

    Дисперсия какого процесса больше и во сколько раз?

12. Найти математическое ожидание, дисперсию и ковариации случайного процесса x_t, если ε_t — белый шум с единичной дисперсией.

    а) $x_t = 0.1 + 0.9x_{t-1} + \varepsilon_t$; б) $x_t = -0.2x_{t-1} + \varepsilon_t$.

13. Найти спектр процесса $x_t = \varepsilon_t + 0.1\varepsilon_{t-1} + 0.01\varepsilon_{t-2} + \ldots$.

14. Корень характеристического уравнения, соответствующего процессу Маркова, равен 2, остаточная дисперсия равна 1. Найти значение спектра на частоте 0.5.

15. Имеется ли разница в графиках спектра для процессов $x_t = 0.9x_{t-1} + \varepsilon_t$ и $x_t = 0.2x_{t-1} + \varepsilon_t$? Если да, то в чем она выражается?

16. Корни характеристических уравнений, соответствующих двум марковским процессам, равны $+1.25$ и -1.25. В чем отличие процессов и как это различие отражается на графиках спектра и автокорреляционной функции?

17. Пусть ε_t — белый шум с единичной дисперсией. Найти математическое ожидание, дисперсию и ковариации случайного процесса:

    а) $x_t = 1 + 0.5x_{t-1} + \varepsilon_t$; б) $x_t = 0.5x_{t-1} + \varepsilon_t$.

18. Проверить на стационарность следующие процессы:

    а) $x_t - 0.4x_{t-1} - 0.4x_{t-2} = \varepsilon_t$; б) $x_t + 0.4x_{t-1} - 0.4x_{t-2} = \varepsilon_t$;

    в) $x_t - 0.4x_{t-1} + 0.4x_{t-2} = \varepsilon_t$.

    Изобразить схематически графики автокорреляционной функции этих процессов. Проверить правильность выводов с помощью точного вычисления автокорреляционной функции для каждого из процессов.

19. Корни характеристического уравнения для процесса Юла равны, соответственно, 5 и -5. Изобразить график автокорреляционной функции. Дать обоснование.

20. Корни характеристического уравнения, соответствующего процессу Юла, равны 1.9 и -1.3. Изобразить график автокорреляционной функции этого процесса. Ответ обосновать.

21. Коэффициенты автокорреляции первого и второго порядка в процессе Юла равны, соответственно, 0.5 и 0.4. Оценить параметры процесса. Найти дисперсию белого шума, если дисперсия процесса равна 1.

22. При каких значениях φ_2 следующие случайные процессы являются стационарными в широком смысле?

    а) $x_t = x_{t-1} + \varphi_2 x_{t-2} + \varepsilon_t$; б) $x_t = -x_{t-1} + \varphi_2 x_{t-2} + \varepsilon_t$.

    Вывести автокорреляционные функции данных случайных процессов при $\varphi_2 = -0.5$.

23. Параметры φ_1 и φ_2 процесса AR(2) равны, соответственно, 0.6 и -0.2. Каковы первые три значения автокорреляционной функции?

24. Для процесса $x_t = 0.5x_{t-1} + 0.25x_{t-2} + \varepsilon_t$ коэффициенты автокорреляции первого и второго порядка равны, соответственно, $2/3$ и $7/12$. Найти коэффициент автокорреляции четвертого порядка.

25. Пусть процесс **AR(2)** $x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \varepsilon_t$ является стационарным в широком смысле, и ε_t — белый шум с дисперсией σ^2. Показать, что частная автокорреляционаня функция

$$\varphi_{1,1} = \frac{\varphi_1}{1-\varphi_2}, \quad \varphi_{2,2} = \frac{\varphi_1^2\varphi_2 + \varphi_2(1-\varphi_2)^2}{(1-\varphi_1-\varphi_2)(1+\varphi_1-\varphi_2)},$$
$$\varphi_{k,k} = 0, \text{ при } k \geqslant 3.$$

26. По известным значениям частной автокорреляционной функции $\varphi_{1,1} = 0.5$ и $\varphi_{2,2} = 2/3$ случайного процесса найти значения коэффициентов автокорреляции первого и второго порядка.

27. Пусть процесс **AR(p)** является стационарным в широком смысле. Показать, что частная автокорреляционаня функция $\varphi_{p+1,p+1} = 0$.

28. В каком случае процесс, описываемый моделью **MA(q)**, стационарен и обратим?

29. Коэффициент автокорреляции первого порядка для обратимого процесса скользящего среднего первого порядка равен -0.4. Записать уравнение процесса и изобразить график его автокорреляционной функции.

30. Показать, что обратимый процесс **MA(1)** можно представить в виде процесса авторегрессии.

31. Показать, что процесс $x_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1}$ эквивалентен процессу **AR(∞)**, если $|\theta_1| < 1$ и ε_t — белый шум.

32. Найти автокорреляционную функцию процесса:

$x_t = \varepsilon_t - 0.5x_{t-1} - 0.25x_{t-2} - 0.125x_{t-3} - 0.0625x_{t-4} + \dots$.

33. Вывести формулы для вычисления математического ожидания, дисперсии и ковариаций случайного процесса $x_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1}$ при условии его слабой стационарности, если ε_t — белый шум с дисперсией σ^2.

34. Пусть ε_t — белый шум с единичной дисперсией. Чему равна дисперсия процесса $x_t = \varepsilon_t + 0.2\varepsilon_{t-1}$? Изобразить график автокорреляционной функции.

35. Идентифицировать процесс, автокорреляционная функция которого имеет следующий вид:

а) $\rho_1 = 0.25,\ \rho_k = 0,\ \forall k \geqslant 2$; б) $\rho_1 = -0.4,\ \rho_k = 0,\ \forall k \geqslant 2$.

36. Является ли случайный процесс, автокорреляционная функция которого имеет следующий вид: $\rho_1 = 0.5,\ \rho_k = 0,\ \forall k \geqslant 2$, обратимым?

37. Для каждого из случайных процессов:

    а) $x_t = \varepsilon_t + 0.5\varepsilon_{t-1}$; б) $x_t = \varepsilon_t - 0.5\varepsilon_{t-1}$; в) $x_t = -1 + \varepsilon_t + 0.8\varepsilon_{t-1}$;

    рассчитать частную автокорреляционную функцию, вычислить первые 6 значений автокорреляционной функции и построить ее график.

38. Показать, что частные автокорреляционные функции следующих слабо стационарных случайных процессов совпадают:

    а) $x_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1}$ и $z_t = \upsilon_t + \theta_1\upsilon_{t-1}$,

    б) $x_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \ldots + \theta_q\varepsilon_{t-q}$

    и $z_t = \upsilon_t + \theta_1\upsilon_{t-1} + \theta_2\upsilon_{t-2} + \ldots + \theta_q\upsilon_{t-q}$,

    где ε_t и υ_t — процессы белого шума с дисперсиями σ_ε^2 и σ_υ^2, соответственно.

39. Имеется следующий обратимый процесс : $x_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}$, где ε_t — белый шум с дисперсией σ^2. Рассчитать коэффициенты автоковариации. Записать автокорреляционную функцию для этого процесса.

40. Построить график автокорреляционной функции процесса:

    а) $x_t = \varepsilon_t + 0.5\varepsilon_{t-1} - 0.3\varepsilon_{t-2}$; б) $x_t = 1 + \varepsilon_t - 0.4\varepsilon_{t-1} + 0.4\varepsilon_{t-2}$.

41. Переписать случайный процесс

    $x_t = 0.5x_{t-1} + 0.5x_{t-2} + \varepsilon_t - \varepsilon_{t-1} + 3\varepsilon_{t-2}$

    с использованием лагового оператора, где ε_t — белый шум. Проверить процесс на стационарность и обратимость.

42. Найти математическое ожидание, дисперсию и ковариации случайного процесса $x_t = 0.5x_{t-1} + \varepsilon_t - 0.7\varepsilon_{t-1}$, если ε_t — белый шум. Построить график автокорреляционной функции.

43. На примере процесса **ARMA**(1, 1) продемонстрировать алгоритм оценивания его параметров методом моментов.

44. Найти параметры модели **ARMA**(1, 1), если $\rho_1 = \dfrac{31}{41},\ \rho_2 = \dfrac{93}{205}$.

45. Проверить на стационарность и обратимость процесс

    $x_t = 0.6 + 0.3x_{t-1} + 0.4x_{t-2} + \varepsilon_t - 0.7\varepsilon_{t-1}$,

    где ε_t — белый шум с дисперсией σ^2. Представить процесс в виде **AR**(∞), если это возможно.

46. Определить порядок интегрирования процесса $x_t = 1.5x_{t-1} + 0.5x_{t-2} + \varepsilon_t - 0.5\varepsilon_{t-1}$. Ответ обосновать.

47. Для модели $(1 - \mathbf{L})(1 + 0.4\mathbf{L})x_t = (1 - 0.5\mathbf{L})\varepsilon_t$ определить параметры p, d, q. Является ли процесс стационарным?

48. Записать формулу расчета коэффициента автоковариации первого порядка для процесса ARIMA(2, 2, 2).

49. Какую роль выполняет оператор скользящего среднего в прогнозировании процессов ARMA(p, q)? Ответ обосновать.

50. Построить точечный прогноз на один шаг вперед, если известно, что процесс $x_t = 0.1x_{t-1} + \varepsilon_t + 0.2\varepsilon_{t-1}, \quad x_T = 10, \quad \varepsilon_T = 0.1.$

51. Построить доверительный интервал для прогноза на два шага вперед для случайного процесса $x_t = 0.5x_{t-1} + \varepsilon_t$, если известно, что $x_T = -1.6$ и ε_t — белый шум с единичной дисперсией.

52. Построить интервальный прогноз на 2 шага вперед для случайного процесса:

    а) $x_t = 1 + \varepsilon_t + 0.7\varepsilon_{t-1}$, если ε_t — белый шум с единичной дисперсией и $\varepsilon_T = -6.7$;

    б) $x_t = 1 + 1.3x_{t-1} + \varepsilon_t$, если ε_t — белый шум с единичной дисперсией и $x_T = 7.1,\ x_{T-1} = 6.7,\ \varepsilon_t = 0.5$.

53. Записать в компактной и развернутой формах уравнение процесса ARIMA(1, 2, 2), привести формулу доверительного интервала для прогноза на 4 шага вперед с выводом формул для параметров ψ_j и дисперсии белого шума.

54. Записать формулу доверительного интервала для прогноза по модели ARIMA(1, 1, 1), с выводом формул для ψ_j и дисперсии белого шума.

Рекомендуемая литература

1. **Айвазян С.А.** Основы эконометрики. Т. 2. — М.: «Юнити», 2001. (Гл. 3).

2. **Андерсон Т.** Статистический анализ временных рядов. — М.: «Мир», 1976. (Гл. 5).

3. **Бокс Дж., Дженкинс Г.** Анализ временных рядов. Прогноз и управление. Вып. 1. — М.: «Мир», 1974. (Гл. 3–6).

4. **Кендалл М. Дж., Стьюарт А.** Многомерный статистический анализ и временные ряды. — М.: «Наука», 1976. (Гл. 47).

5. **Магнус Я.Р., Катышев П.К., Пересецкий А.А.** Эконометрика — начальный курс. — М.: «Дело», 2000. (Гл. 12).

6. **Справочник по прикладной статистике.** В 2-х т. Т. 2. // Под ред. Э. Ллойда, У. Ледермана. — М.: «Финансы и статистика», 1990. (Гл. 18).

7. **Chatfield Chris.** The Analysis of Time Series: An Introduction... 5^{th} ed. — Chapman & Hall/CRC, 1996. (Ch. 3–5).

8. **Enders Walter.** Applied Econometric Time Series. — Iowa State University, 1995. (Ch. 5).

9. **Judge G.G., Griffiths W.E., Hill R.C., Luthepohl H., Lee T.** Theory and Practice of Econometrics. — New York: John Wiley & Sons, 1985. (Ch. 7).

10. **Hamilton James D.,** Time Series Analysis. — Princeton University Press, 1994. (Ch. 3, 4).

11. **Mills Terence C.** Time Series Techniques for Economists. — Cambridge University Press, 1990. (Ch. 5–8).

12. **Pollock D.S.G.** A handbook of time-series analysis, signal processing and dynamics. — «Academic Press», 1999. (Ch. 16–19).

Глава 15

Динамические модели регрессии

При моделировании экономических процессов с помощью регрессионного анализа часто приходится наряду с некоторым временным рядом вводить в модель также лаг этого ряда. В экономике практически нет примеров мгновенного реагирования на какое-либо экономическое воздействие — существуют задержки проявления эффектов от капиталовложений, внесения удобрений и т.д., иными словами, при моделировании необходимо учитывать воздействие факторов в предыдущие моменты времени. Выше были введены некоторые из таких моделей: регрессия с распределенным лагом и модели ARIMA. В этой главе рассматриваются различные аспекты подобного рода моделей.

15.1. Модель распределенного лага: общие характеристики и специальные формы структур лага

Напомним, что простейшая модель распределенного лага — это модель регрессии, в которой на динамику исследуемой переменной x_t влияет не только какой-то объясняющий фактор z_t, но и его лаги. Модель имеет следующий вид:

$$x_t = \mu + \sum_{j=0}^{q} \alpha_j z_{t-j} + \varepsilon_t = \mu + \alpha(\mathbf{L}) z_t + \varepsilon_t, \qquad (15.1)$$

где $\alpha(\mathbf{L}) = \sum\limits_{j=0}^{q} \alpha_j \mathbf{L}^j$, а q — величина максимального лага.

Данную модель можно охарактеризовать следующими показателями.

Функция реакции на импульс (*impulse response function*, IRF) показывает, насколько изменится x_t при изменении z_{t-j} на единицу для лагов $j = 0,\ 1,\ 2,\ \ldots$. Таким образом, можно считать, что речь идет о производной $\dfrac{dx_t}{dz_{t-j}}$ как функции запаздывания j. Ясно, что для модели распределенного лага этот показатель совпадает с коэффициентом α_j при $j \leqslant q$ и равен нулю при $j > q$. При $j < 0$ (влияние будущих значений переменной z на переменную x) реакцию на импульс можно положить равной нулю.

Накопленная реакция на импульс для лага k — это просуммированные значения простой функции реакции на импульс от $j = 0$ до $j = k$. Для модели распределенного лага это сумма коэффициентов:

$$\sum_{j=0}^{\min\{k,\,q\}} \alpha_j.$$

Долгосрочный мультипликатор является измерителем общего влияния переменной z на переменную x. Он равен

$$\alpha_\Sigma = \sum_{j=0}^{q} \alpha_j = \alpha(1).$$

Это предельное значение накопленной реакции на импульс. Если x и z — логарифмы исходных переменных, то α_Σ — **долгосрочная эластичность**.

Средняя длина лага показывает, на сколько периодов в среднем запаздывает влияние переменной z на переменную x. Она вычисляется по формуле

$$\bar{j} = \frac{\sum\limits_{j=0}^{q} j\alpha_j}{\sum\limits_{j=0}^{q} \alpha_j} = \frac{\sum\limits_{j=0}^{q} j\alpha_j}{\alpha_\Sigma}.$$

Заметим, что среднюю длину лага можно записать через производную логарифма многочлена $\alpha(\mathbf{L})$ в точке 1. Действительно,

$$\alpha'(v) = \left(\sum_{j=0}^{q} \alpha_j v^j\right)' = \sum_{j=0}^{q} j\alpha_j v^{j-1}$$

и $\alpha'(1) = \sum\limits_{j=0}^{q} j\alpha_j$. Поэтому $\bar{j} = \frac{\alpha'(1)}{\alpha(1)} = (\ln \alpha(v))' \Big|_{v=1}$.

Наряду со средней длиной лага можно рассматривать также **медианную длину лага**, то есть такую величину лага, при которой накопленная функция реакции на импульс равна половине долгосрочного мультипликатора. Ясно, что для большинства возможных структур лага такое равенство может выполняться только приближенно. Поэтому невозможно дать однозначное определение медианной длины лага.

Оценивание модели распределенного лага может быть затруднено проблемой мультиколлинеарности, если величина фактора z_t мало меняется со временем. Если z_t — случайный процесс, то такая ситуация возникает, когда данный процесс сильно положительно автокоррелирован. Например, это может быть авторегрессия первого порядка с коэффициентом авторегрессии, близким к единице. Если бы фактор z_t был линейным трендом, например, $z_t = t$, то модель невозможно было бы оценить. Действительно, несложно увидеть, что тогда z_t, $z_{t-1} = t - 1$ и константа связаны между собой линейной зависимостью. Если z_t — линейный тренд с добавлением небольшой стационарной случайной составляющей, то, хотя строгой линейной зависимости уже не будет, проблема мультиколлинеарности останется.

Если возникает подобная проблема мультиколлинеарности, то нельзя точно оценить структуру лага, хотя можно оценить сумму весов α_i — т.е. долгосрочный мультипликатор α_Σ. Эта сумма вычленяется из модели следующим образом:

$$x_t = \mu + \alpha_\Sigma z_t + \sum_{j=1}^{q} \alpha_j (z_{t-j} - z_t) + \varepsilon_t.$$

В случае мультиколлинеарности лаговых переменных обычно на лаговую структуру накладывают какое-нибудь ограничение, чтобы уменьшить количество оцениваемых коэффициентов. Ниже рассматриваются две наиболее важные модели этого типа.

Полиномиальный лаг

Одна из возможных структур лага — **полиномиальный лаг**[1], веса которого задаются многочленом от величины лага j:

$$\alpha_j = \sum_{s=0}^{p} \gamma_s j^s, \quad j = 0, \ldots, q, \tag{15.2}$$

[1] Эту модель предложила С. Алмон, поэтому часто используют термин «лаг Алмон» (*Almon lag*).

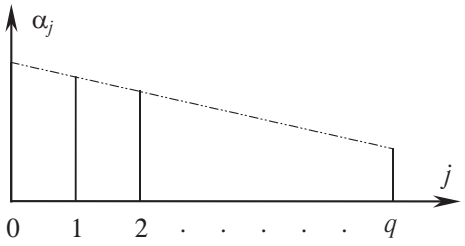

Рис. 15.1

где p — степень многочлена, $p < q$. Вводя такую зависимость, мы накладываем $q - p$ линейных ограничений на структуру лага.

Простейший полиномиальный лаг — линейный. Для него $\alpha_j = \gamma_0 + \gamma_1 j$. Как правило, здесь $\gamma_1 < 0$. Его структура изображена на диаграмме (рис. 15.1).

Поскольку исходная модель регрессии линейна и ограничения, которые полиномиальный лаг накладывает на ее коэффициенты, являются линейными, то полученная модель останется линейной. Рассмотрим, каким образом ее можно оценить.

С учетом выражений для α_j, проведем преобразование исходной модели:

$$\sum_{j=0}^{q} \alpha_j z_{t-j} = \sum_{j=0}^{q} \underbrace{\left(\sum_{s=0}^{p} \gamma_s j^s \right)}_{\alpha_j} z_{t-j} = \sum_{s=0}^{p} \gamma_s \sum_{j=0}^{q} j^s z_{t-j} = \sum_{s=0}^{p} \gamma_s y_{ts}.$$

Получим новую модель линейной регрессии:

$$x_t = \mu + \sum_{s=0}^{p} \gamma_s y_{ts} + \varepsilon_t$$

с преобразованными факторами

$$y_{ts} = \sum_{j=0}^{q} j^s z_{t-j}.$$

Оценив γ_s, можно вычислить веса α_j, воспользовавшись формулой (15.2).

При оценивании модели с ограничениями на структуру лага нужно проверить, правильно ли наложены ограничения. С помощью соответствующей F-статистики можно сравнить ее с исходной, неограниченной моделью, поскольку она является ее частным случаем. Модель

$$x_t = \mu + \sum_{s=0}^{q} \gamma_s y_{ts} + \varepsilon_t$$

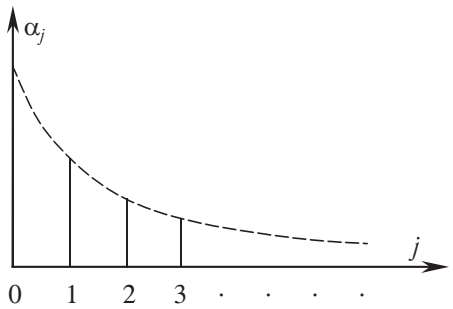

Рис. 15.2

эквивалентна исходной модели с точностью до линейных преобразований, поэтому достаточно проверить гипотезу о том, что последние $q - p$ коэффициентов $(\gamma_{p+1}, \ldots, \gamma_q)$ равны нулю.

Часто принимают, что веса на концах полиномиальной лаговой структуры (15.2) равны нулю. Это требование накладывает на коэффициенты модели дополнительные ограничения. Можно, например, потребовать, чтобы $\alpha_q = 0$, то есть

$$\sum_{s=0}^{p} \gamma_s q^s = 0.$$

Учесть такие ограничения несколько сложнее, но в целом не требуется выходить за рамки обычной линейной регрессии.

Геометрический лаг

Еще один популярный вид структуры лага — **геометрический лаг**. Его веса α_j задаются следующими соотношениями:

$$\alpha_j = \alpha_0 \delta^j, \quad j = 0, \ldots, \infty,$$

где $0 < \delta < 1$. Веса геометрического лага убывают экспоненциально с увеличением лага (рис. 15.2).

Модель распределенного лага с этими весами, **модель Койка**, имеет следующий вид:

$$x_t = \mu + \alpha_0 \sum_{j=0}^{\infty} \delta^j z_{t-j} + \varepsilon_t. \tag{15.3}$$

Используя формулу суммы бесконечной геометрической прогрессии, получим

$$\alpha(v) = \sum_{j=0}^{\infty} \alpha_j v^j = \alpha_0 \sum_{j=0}^{\infty} (\delta v)^j = \frac{\alpha_0}{1 - v\delta}.$$

Сумма весов в этой модели (долгосрочный мультипликатор) равна

$$\alpha_\Sigma = \sum_{j=0}^{\infty} \alpha_j = \alpha(1) = \frac{\alpha_0}{1-\delta}.$$

Кроме того,

$$\ln \alpha(v) = \ln \alpha_0 - \ln(1 - v\alpha) \quad \text{и} \quad (\ln \alpha(v))' = \frac{\delta}{1 - v\delta},$$

поэтому средняя длина геометрического лага равна

$$\bar{j} = (\ln \alpha(v))'\Big|_{v=1} = \frac{\delta}{1-\delta}.$$

Чтобы избавиться от бесконечного ряда, к модели с геометрическим лагом применяют **преобразование Койка** (*Koyck transformation*). Сдвинем исходное уравнение на один период назад:

$$x_{t-1} = \mu + \sum_{j=0}^{\infty} \alpha_0 \delta^j z_{t-j-1} + \varepsilon_{t-1},$$

затем умножим это выражение на δ и вычтем из исходного уравнения (15.3):

$$x_t - \delta x_{t-1} = (1-\delta)\mu + \alpha_0 z_t + \varepsilon_t - \delta\varepsilon_{t-1}. \tag{15.4}$$

Такой же результат можно получить, используя лаговые операторы:

$$x_t = \mu + \alpha_0 \sum_{j=0}^{\infty} \delta^j z_{t-j} + \varepsilon_t = \mu + \alpha_0 \Big(\sum_{j=0}^{\infty} (\delta \mathbf{L})^j\Big) z_t + \varepsilon_t.$$

Выражение в скобках упрощается с использованием формулы суммы бесконечной геометрической прогрессии:

$$x_t = \mu + \alpha_0 \frac{1}{1 - \delta\mathbf{L}} z_t + \varepsilon_t.$$

Умножим это уравнение на оператор $(1 - \delta\mathbf{L})$:

$$(1 - \delta\mathbf{L})\, x_t = (1 - \delta\mathbf{L})\, \mu + \alpha_0 z_t + (1 - \delta\mathbf{L})\, \varepsilon_t$$

или учитывая, что оператор сдвига, стоящий перед константой, ее сохраняет, получаем формулу (15.4). В результате имеем следующую модель:

$$x_t = \mu' + \delta x_{t-1} + \alpha_0 z_t + \varepsilon'_t,$$

где $\mu' = (1-\delta)\mu$ и $\varepsilon'_t = \varepsilon_t - \delta\varepsilon_{t-1}$. Это частный случай авторегрессионной модели с распределенным лагом, рассматриваемой в следующем пункте.

Заметим, что в полученной здесь модели ошибка ε'_t не является белым шумом, а представляет собой процесс скользящего среднего первого порядка. Модель является линейной регрессией, однако для нее не выполнено требование о некоррелированности регрессоров и ошибки. Действительно, ε_{t-1} входит как в x_{t-1}, так и в ε'_t. Следовательно, оценки метода наименьших квадратов не являются состоятельными и следует пользоваться другими методами.

Можно оценивать модель Койка в исходном виде (15.3). Сумму в этом уравнении можно разделить на две части: соответствующую имеющимся наблюдениям для переменной z_t и относящуюся к прошлым ненаблюдаемым значениям, т.е. z_0, z_{-1} и т.д.:

$$x_t = \mu + \alpha_0 \sum_{j=0}^{t-1} \delta^j z_{t-j} + \alpha_0 \sum_{j=t}^{\infty} \delta^j z_{t-j} + \varepsilon_t.$$

Далее, во второй сумме сделаем замену $j = s + t$:

$$x_t = \mu + \alpha_0 \sum_{j=0}^{t-1} \delta^j z_{t-j} + \alpha_0 \delta^t \sum_{s=0}^{\infty} \delta^s z_{-s} + \varepsilon_t.$$

Обозначив $\theta = \alpha_0 \sum\limits_{s=0}^{\infty} \delta^s z_{-s}$, получим модель нелинейной регрессии с четырьмя неизвестными параметрами:

$$x_t = \mu + \alpha_0 \sum_{j=0}^{t-1} \delta^j z_{t-j} + \theta\delta^t + \varepsilon_t.$$

В такой модели ошибка и регрессоры некоррелированы, поэтому нелинейный МНК дает состоятельные оценки.

15.2. Авторегрессионная модель с распределенным лагом

Авторегрессионная модель с распределенным лагом является примером динамической регрессии, в которой, помимо объясняющих переменных и их лагов, в качестве регрессоров используются лаги зависимой переменной.

Авторегрессионную модель с распределенным лагом, которая включает одну независимую переменную, можно представить в следующем виде:

$$x_t = \mu + \sum_{j=1}^{p} \varphi_j x_{t-j} + \sum_{j=0}^{q} \alpha_j z_{t-j} + \varepsilon_t, \tag{15.5}$$

где первая сумма представляет собой авторегрессионную компоненту — распределенный лаг изучаемой переменной, вторая сумма — распределенный лаг независимого фактора. Обычно предполагается, что в этой модели ошибки ε_t являются белым шумом и не коррелированны с фактором z_t, его лагами и с лагами изучаемой переменой x_t. При этих предположениях МНК дает состоятельные оценки параметров модели.

Сокращенно эту модель обозначают ADL(p, q) (от английского *autoregressive distributed lag*), также часто используется аббревиатура ARDL, где p — порядок авторегрессии, q — порядок распределенного лага. Более компактно можно записать модель в операторной форме:

$$\varphi(\mathbf{L})\, x_t = \mu + \alpha(\mathbf{L})\, z_t + \varepsilon_t,$$

где $\varphi(\mathbf{L}) = 1 - \sum\limits_{j=1}^{p} \varphi_j \mathbf{L}^j$ и $\alpha(\mathbf{L}) = \sum\limits_{j=0}^{q} \alpha_j \mathbf{L}^j$ — лаговые многочлены.

Модель ADL(1, 1) имеет следующий вид:

$$x_t = \mu + \varphi_1 x_{t-1} + \alpha_0 z_t + \alpha_1 z_{t-1} + \varepsilon_t.$$

Некоторые частные случаи модели ADL уже были рассмотрены ранее.

Модель ADL(0, q) — это модель распределенного лага, рассмотренная в предыдущем пункте (в правой части нет лагов зависимой переменной).

Модель геометрического распределенного лага после преобразования Койка можно интерпретировать как ADL(1, 0) с процессом MA(1) в ошибке и ограничением на коэффициент при x_{t-1}, который равен параметру MA-процесса (δ):

$$x_t = \mu' + \delta x_{t-1} + \alpha_0 z_t + (\varepsilon_t - \delta \varepsilon_{t-1}).$$

Авторегрессионную модель AR(p) можно считать ADL(p, -1). В этой модели переменная в левой части зависит только от своих собственных лагов:

$$x_t = \mu + \sum_{j=1}^{p} \varphi_j x_{t-j} + \varepsilon_t.$$

Как и в случае модели распределенного лага, можно ввести ряд показателей, характеризующих модель ADL. Если обратить лаговый многочлен $\varphi(\mathbf{L})$ и умножить

на него исходное уравнение модели, то получим

$$x_t = \varphi^{-1}(\mathbf{L})\varphi(\mathbf{L})x_t = \frac{\mu}{\varphi(\mathbf{L})} + \frac{\alpha(\mathbf{L})}{\varphi(\mathbf{L})}z_t + \frac{\varepsilon_t}{\varphi(\mathbf{L})}$$

или

$$x_t = \mu^* + \sum_{i=0}^{\infty} \pi_i z_{t-i} + \varepsilon_t^*,$$

где

$$\mu^* = \frac{\mu}{\varphi(1)}, \quad \varepsilon_t^* = \frac{\varepsilon_t}{\varphi(\mathbf{L})} \quad \text{и} \quad \frac{\alpha(\mathbf{L})}{\varphi(\mathbf{L})} = \pi(\mathbf{L}) = \sum_{i=0}^{\infty} \pi_i \mathbf{L}^i.$$

Как и в модели ARMA, такое преобразование корректно, если все корни многочлена $\varphi(\cdot)$ лежат за пределами единичной окружности.

Коэффициенты π_i показывают влияние лагов переменной z на переменную x, то есть они представляют собой функцию реакции на импульс. Символически эти коэффициенты можно записать в виде:

$$\pi_i = \frac{dx_t}{dz_{t-i}}.$$

Рекуррентная формула для расчета коэффициентов π_i получается дифференциацией по z_{t-i} исходного уравнения модели (15.5):

$$\pi_i = \frac{d(\mu + \sum\limits_{j=1}^{p} \varphi_j x_{t-j} + \sum\limits_{j=0}^{q} \alpha_j z_{t-j} + \varepsilon_t)}{dz_{t-i}} = \sum_{j=1}^{p} \varphi_j \pi_{i-j} + \alpha_i.$$

Здесь принимается во внимание, что

$$\frac{dx_{t-j}}{dz_{t-i}} = \pi_{i-j}, \quad \frac{dz_{t-j}}{dz_{t-i}} = \begin{cases} 0, & j \neq i, \\ 1, & j = i, \end{cases} \quad \text{и} \quad \frac{d\varepsilon_t}{dz_{t-i}} = 0.$$

При использовании этой рекуррентной формулы следует взять $\pi_i = 0$ для $i < 0$. В частном случае модели распределенного лага (когда $p = 0$) эта формула дает $\pi_i = \alpha_i$, то есть влияние z_{t-i} на π_i количественно выражается коэффициентом при z_{t-i} (весом лага).

Сумма коэффициентов π_i показывает долгосрочное влияние z на x (долгосрочный мультипликатор). Она равна

$$\pi_\Sigma = \sum_{i=0}^{\infty} \pi_i = \pi(1) = \frac{\alpha(1)}{\varphi(1)} = \frac{\sum\limits_{j=0}^{q} \alpha_j}{1 - \sum\limits_{j=1}^{p} \varphi_j}. \tag{15.6}$$

По аналогии с моделью распределенного лага можно ввести показатель средней длины лага влияния z на x. Он равен

$$\frac{\sum\limits_{i=0}^{\infty} i\pi_i}{\sum\limits_{i=0}^{\infty} \pi_i} = (\ln \pi(v))' \Big|_{v=1} = (\ln \alpha(v) - \ln \varphi(v))' \Big|_{v=1} = \frac{\sum\limits_{j=0}^{q} j\alpha_j}{\sum\limits_{j=0}^{q} \alpha_j} + \frac{\sum\limits_{j=1}^{p} j\varphi_j}{1 - \sum\limits_{j=1}^{p} \varphi_j}.$$

15.3. Модели частичного приспособления, адаптивных ожиданий и исправления ошибок

Рассмотрим некоторые прикладные динамические модели, сводящиеся к модели авторегрессионного распределенного лага.

Модель частичного приспособления

В экономике субъекты не сразу могут приспособиться к меняющимся условиям — это происходит постепенно. Нужно время на изменение запасов, обучение, переход на новые технологии, изменение условий долгосрочных контрактов и т.д. Эти процессы можно моделировать с помощью модели **частичного приспособления**.

Для иллюстрации приведем следующий пример: инфляция зависит от денежной массы, меняя денежную массу, мы можем получить какой-то желаемый уровень инфляции. Но реальность несколько запаздывает.

Пусть x_t^D — желаемый уровень величины x_t, z_t — независимый фактор, определяющий x_t^D. Тогда модель частичного приспособления задается следующими двумя уравнениями:

$$\begin{aligned} x_t^D &= \beta + \alpha z_t + \xi_t, \\ x_t - x_{t-1} &= \gamma(x_t^D - x_{t-1}) + \varepsilon_t. \end{aligned} \tag{15.7}$$

Здесь $\gamma \in [0; 1]$ — скорость приспособления. Если $\gamma = 0$, то $x_t = x_{t-1}$, то есть x_t не меняется, если же $\gamma = 1$, то приспособление происходит мгновенно, и в этом случае сразу $x_t = x_t^D$.

Предположим, что переменная x_t^D ненаблюдаема. Исключим из этих двух выражений ненаблюдаемую переменную:

$$x_t = \gamma\beta + (1 - \gamma)x_{t-1} + \gamma\alpha z_t + \varepsilon_t + \gamma\xi_t.$$

Ясно, что это модель ADL(1, 0), где $\gamma\beta = \mu$, $1-\gamma = \varphi_1$ и $\gamma\alpha = \alpha_0$. Оценив параметры μ, φ_1 и α_0, мы можем с помощью обратного преобразования вычислить оценки параметров исходной модели.

Модель адаптивных ожиданий

Очень часто экономические решения, принимаемые людьми, зависят от прогнозов того, что будет в будущем. При этом уровень экономических величин, на которые воздействуют такие решения, зависит не от текущего значения показателя, а от ожидаемого значения (например, если ожидается высокий уровень инфляции, то следует скупать доллары, курс доллара в результате вырастет). В теории рассматриваются 2 вида ожиданий — **рациональные** и **адаптивные**. В соответствии с одним из определений, ожидания называют рациональными, если математическое ожидание прогноза равно фактическому значению, которое будет в будущем. Модели рациональных ожиданий часто оказываются довольно сложными. Адаптивные ожидания — это ожидания, которые зависят только от предыдущих значений величины. По мере того, как наблюдаются процессы движения реальной величины, мы адаптируем наши ожидания к тому, что наблюдаем на самом деле.

Чтобы ввести в экономические модели ожидания экономических субъектов, в простейшем случае используют модель **адаптивных ожиданий**. Адаптивные ожидания некоторой величины формируются только на основе прошлых значений этой величины. Например, пусть x_t зависит от ожиданий (z_t^E) величины z_t, z_t — величина, от прогноза которой должен зависеть x_t (например, инфляция), z_t^E — ожидание (прогноз) этой величины в момент времени t.

$$x_t = \beta + \alpha z_t^E + \varepsilon_t.$$

В целом x_t выгодно выбирать в зависимости от того, какой величина z_t будет в будущем: $z_{t+1}, z_{t+2}, \ldots$, однако в момент выбора t известны только текущее и прошлые значения ($\ldots, z_{t-1}, z_t$).

Ошибка в ожиданиях z_t^E приводит к их корректировке. Модель адаптации ожиданий к фактическому значению z_t записывается так:

$$z_t^E - z_{t-1}^E = \theta(z_t - z_{t-1}^E),$$

где θ — скорость приспособления ожиданий. Если $\theta = 0$, то ожидания никак не адаптируются к действительности и прогнозы не сбываются (скорость адаптации нулевая); если $\theta = 1$, скорость адаптации мгновенная, наши ожидания сбываются (полностью адаптировались): $z_t^E = z_t$. Обычно $0 < \theta < 1$.

Легко видеть, что модель адаптации ожиданий основывается на формуле экспоненциальной средней:

$$z_t^E = \theta z_t + (1-\theta) z_{t-1}^E.$$

Для оценки параметров модели надо исключить ненаблюдаемые ожидания z_t^E. Используя лаговый оператор, получаем:

$$z_t^E - (1-\theta) z_{t-1}^E = (1-(1-\theta)\mathbf{L}) z_t^E = \theta z_t,$$

откуда

$$z_t^E = \frac{\theta z_t}{1-(1-\theta)\mathbf{L}} = \theta \sum_{i=0}^{\infty} (1-\theta)^i z_{t-i}.$$

Таким образом, ожидания в рассматриваемой модели описываются бесконечным геометрическим распределенным лагом с параметром затухания $\delta = 1-\theta$.

Если в уравнение для x_t вместо z_t^E подставить данный бесконечный ряд, то получится модель регрессии с геометрическим распределенным лагом:

$$x_t = \beta + \frac{\alpha\theta z_t}{1-(1-\theta)\mathbf{L}} + \varepsilon_t. \tag{15.8}$$

Как было показано ранее, модель геометрического лага с помощью преобразования Койка приводится к модели ADL. Умножим обе части уравнения 15.8 на $1-(1-\theta)\mathbf{L}$ и получим:

$$(1-(1-\theta)\mathbf{L}) x_t = (1-(1-\theta)\mathbf{L})\beta + \alpha\theta z_t + (1-(1-\theta)\mathbf{L})\varepsilon_t.$$

После соответствующего переобозначения параметров модель адаптивных ожиданий приобретает новую форму — ADL(1, 0) с MA(1)-ошибкой:

$$x_t = \theta\beta + (1-\theta) x_{t-1} + \alpha\theta z_t + \varepsilon_t - (1-\theta)\varepsilon_{t-1}.$$

Оценивать модель адаптивных ожиданий можно теми же методами, что и модель Койка.

Модель исправления ошибок

В динамических регрессионных моделях важно различие между долгосрочной и краткосрочной динамикой. Это различие можно анализировать в рамках модели исправления ошибок. Рассмотрим в долгосрочном аспекте модель ADL(1, 1):

$$x_t = \mu + \varphi_1 x_{t-1} + \alpha_0 z_t + \alpha_1 z_{t-1} + \varepsilon_t.$$

Предположим, что фактор z_t и ошибка ε_t являются стационарными процессами. Тогда при $|\varphi_1| < 1$ изучаемая переменная x_t также стационарна. Возьмем математические ожидания от обеих частей уравнения модели:

$$\bar{x} = \mu + \varphi_1 \bar{x} + \alpha_0 \bar{z} + \alpha_1 \bar{z}.$$

В этой формуле $\bar{x} = \mathbf{E}(x_t)$, $\bar{z} = \mathbf{E}(z_t)$ (стационарные уровни x и z) и учитывается, что $\mathbf{E}(\varepsilon_t) = 0$. Получаем уравнение

$$\bar{x} = \frac{\mu}{1-\varphi_1} + \frac{\alpha_0 + \alpha_1}{1-\varphi_1}\bar{z} = \mu' + \lambda \bar{z},$$

которое описывает **долгосрочное стационарное состояние** экономического процесса. Коэффициент

$$\lambda = \frac{\alpha_0 + \alpha_1}{1-\varphi_1} \tag{15.9}$$

отражает долгосрочное влияние z на x. Он совпадает с долгосрочным мультипликатором (15.6).

Модель ADL(1, 1) можно привести к виду, который описывает краткосрочную динамику экономической системы. В этом виде модель называется **моделью исправления ошибок**, сокращенно ECM (*error-correction model*):

$$\Delta x_t = \mu - (1-\varphi_1)x_{t-1} + \alpha_0 \Delta z_t + (\alpha_0 + \alpha_1) z_{t-1} + \varepsilon_t$$

или

$$\Delta x_t = \alpha_0 \Delta z_t - \theta\left(x_{t-1} - (\mu' + \lambda z_{t-1})\right) + \varepsilon_t, \tag{15.10}$$

где $\theta = 1 - \varphi_1$, $\Delta x_t = x_t - x_{t-1}$, $\Delta z_t = z_t - z_{t-1}$.

Предполагается, что если в предыдущий период переменная x отклонилась от своего «долгосрочного значения» $\mu' + \lambda z$, то элемент $x_{t-1} - (\mu' + \lambda z_{t-1})$ корректирует динамику в нужном направлении. Для того чтобы это происходило, необходимо выполнение условия $|\varphi_1| < 1$.

Иногда из теории, описывающей явление, следует, что $\lambda = 1$, тогда $\varphi_1 + \alpha_0 + \alpha_1 = 1$. Часто именно такую модель называют ECM.

Модели частичного приспособления и адаптивных ожиданий являются частными случаями модели исправления ошибок — не только формально математически, но и по экономическому содержанию. Например, модель частичного приспособления (15.7) в форме ECM выглядит как

$$\Delta x_t = \alpha\gamma\Delta z_t - \gamma(x_{t-1} - \beta - \alpha z_{t-1}) + \varepsilon_t + \gamma\xi_t.$$

Рассмотрим теперь авторегрессионную модель с распределенным лагом общего вида (15.5) и покажем, что ее можно представить в виде модели исправления ошибок. При предположениях о стационарности x_t и ε_t математические ожидания от обеих частей уравнения (15.5) приводят к выражению:

$$\bar{x} = \mu + \sum_{j=1}^{p} \varphi_j \bar{x} + \sum_{j=0}^{q} \alpha_j \bar{z}$$

или

$$\bar{x} = \frac{\mu}{1 - \sum_{j=1}^{p} \varphi_j} + \frac{\sum_{j=0}^{q} \alpha_j}{1 - \sum_{j=1}^{p} \varphi_j} \bar{z} = \mu' + \lambda\bar{z},$$

где коэффициент долгосрочного влияния z на x:

$$\lambda = \frac{\sum_{j=0}^{q} \alpha_j}{1 - \sum_{j=1}^{p} \varphi_j},$$

как и в случае ADL(1, 1), совпадает с долгосрочным мультипликатором π_Σ.

В этих обозначениях можно представить модель ADL(p, q) в виде модели исправления ошибок:

$$\Delta x_t = -\theta\left(x_{t-1} - (\mu' + \lambda z_{t-1})\right) + \sum_{j=1}^{p-1} \gamma_j \Delta x_{t-j} + \sum_{j=0}^{q-1} \beta_j \Delta z_{t-j} + \varepsilon_t, \qquad (15.11)$$

где

$$\theta = \Big(1 - \sum_{j=1}^{p} \varphi_j\Big), \quad \gamma_j = -\sum_{i=j+1}^{p} \varphi_i, \quad \beta_j = -\sum_{i=j+1}^{q} \alpha_i \text{ при } j > 0, \text{ и } \beta_0 = \alpha_0.$$

15.4. Упражнения и задачи

Упражнение 1

Сгенерируйте нормально распределенный некоррелированный ряд x_t длиной 24 со средним 2000 и дисперсией 900.

1.1. На основе этого ряда по модели $y_t = 5 + 5x_t + 8x_{t-1} + 9x_{t-2} + 8x_{t-3} + + 5x_{t-4} + \varepsilon_t$, где ε_t — нормально распределенный белый шум с дисперсией 100, сгенерируйте 100 рядов y_t, $t = 1, \ldots, 20$.

1.2. Используя 100 сгенерированных наборов данных, оцените модель распределенного лага с максимальной длиной лага $q = 4$. Найдите среднее и дисперсию оценок коэффициентов. Сравните с истинными значениями.

1.3. Для первых 5 наборов данных выберите наиболее подходящую длину лага на основе информационных критериев Акаике (AIC) и Шварца (BIC), оценив модель для $q = 2, \ldots, 6$.

1.4. Используя все 100 наборов, оцените модель полиномиального лага с максимальной длиной лага $q = 4$ и степенью полинома $p = 2$. Рассчитайте средние и дисперсии оценок весов распределенного лага. Сравните с истинными значениями. Сравните с результатами из упражнения 1.2 и сделайте вывод о том, какие оценки точнее.

1.5. Повторите упражнение 1.4 для а) $q = 4$, $p = 3$; б) $q = 6$, $p = 2$; в) $q = 3$, $p = 2$.

Упражнение 2

В таблице 15.1 приведены данные из известной статьи С. Алмон по промышленным предприятиям США (EXPEND — capital expenditures, капитальные расходы, APPROP — appropriations).

2.1. Постройте графики двух рядов. Что можно сказать по ним о рядах? Видна ли зависимость между рядами (в тот же период или с запаздыванием)?

2.2. Постройте кросс-корреляционную функцию для сдвигов
$-12, \ldots, 0, \ldots, 12$.
Сделайте выводы.

2.3. Используя данные, оцените неограниченную модель распределенного лага с максимальной длиной лага $q = 6, \ldots, 12$ для зависимости EXPEND от APPROP. Используя известные вам методы, выберите длину лага.

2.4. Оцените модель полиномиального лага с максимальным лагом 8 и степенью многочлена $p = 2, \ldots, 6$. С помощью статистики Стьюдента проверьте гипотезы $p = 5$ против $p = 6, \ldots, p = 3$ против $p = 2$ и выберите наиболее подходящую степень p.

Таблица 15.1. (Источник: Almon Shirley. «The Distributed Lag between Capital Appropriations and Expenditures», Econometrica 33, January1965, pp. 178–196)

Квартал	EXPEND	APPROP	Квартал	EXPEND	APPROP
1953.1	2072.0	1660.0	1960.3	2721.0	2131.0
1953.2	2077.0	1926.0	1960.4	2640.0	2552.0
1953.3	2078.0	2181.0	1961.1	2513.0	2234.0
1953.4	2043.0	1897.0	1961.2	2448.0	2282.0
1954.1	2062.0	1695.0	1961.3	2429.0	2533.0
1954.2	2067.0	1705.0	1961.4	2516.0	2517.0
1954.3	1964.0	1731.0	1962.1	2534.0	2772.0
1954.4	1981.0	2151.0	1962.2	2494.0	2380.0
1955.1	1914.0	2556.0	1962.3	2596.0	2568.0
1955.2	1991.0	3152.0	1962.4	2572.0	2944.0
1955.3	2129.0	3763.0	1963.1	2601.0	2629.0
1955.4	2309.0	3903.0	1963.2	2648.0	3133.0
1956.1	2614.0	3912.0	1963.3	2840.0	3449.0
1956.2	2896.0	3571.0	1963.4	2937.0	3764.0
1956.3	3058.0	3199.0	1964.1	3136.0	3983.0
1956.4	3309.0	3262.0	1964.2	3299.0	4381.0
1957.1	3446.0	3476.0	1964.3	3514.0	4786.0
1957.2	3466.0	2993.0	1964.4	3815.0	4094.0
1957.3	3435.0	2262.0	1965.1	4093.0	4870.0
1957.4	3183.0	2011.0	1965.2	4262.0	5344.0
1958.1	2697.0	1511.0	1965.3	4531.0	5433.0
1958.2	2338.0	1631.0	1965.4	4825.0	5911.0
1958.3	2140.0	1990.0	1966.1	5160.0	6109.0
1958.4	2012.0	1993.0	1966.2	5319.0	6542.0
1959.1	2071.0	2520.0	1966.3	5574.0	5785.0
1959.2	2192.0	2804.0	1966.4	5749.0	5707.0
1959.3	2240.0	2919.0	1967.1	5715.0	5412.0
1959.4	2421.0	3024.0	1967.2	5637.0	5465.0
1960.1	2639.0	2725.0	1967.3	5383.0	5550.0
1960.2	2733.0	2321.0	1967.4	5467.0	5465.0

Упражнение 3

В таблице 15.2 имеются следующие данные: gas — логарифм среднедушевых реальных расходов на бензин и нефть, price — логарифм реальной цены на бензин и нефть, income — логарифм среднедушевого реального располагаемого дохода, miles — логарифм расхода топлива (в галлонах на милю).

3.1. Оцените авторегрессионную модель с распределенным лагом. В качестве зависимой переменной возьмите gas, а в качестве факторов — income и price. Лаг для всех переменных равен 5.

3.2. Найдите коэффициенты долгосрочного влияния дохода на потребление топлива и цены на потребление топлива (долгосрочные эластичности):

$$\frac{d(\text{gas})}{d(\text{income})} \text{ — долгосрочная эластичность по доходу,}$$
$$\frac{d(\text{gas})}{d(\text{price})} \text{ — долгосрочная эластичность по цене.}$$

3.3. Вычислите функцию реакции на импульсы (для сдвигов $0,\ 1,\ \ldots,\ 40$) для влияния дохода на потребление топлива и цены на потребление топлива. Найдите среднюю длину лага для этих двух зависимостей.

3.4. Оцените ту же модель в виде модели исправления ошибок.

Упражнение 4

В таблице 15.3 приводятся данные о потреблении и располагаемом доходе в США за 1953–1984 гг.

4.1. Оцените следующую модель адаптивных ожиданий. Потребление C_t зависит от перманентного дохода Y_t^E: $C_t = \beta + \alpha Y_t^E + \varepsilon_t$, где Y_t^E задается уравнением $Y_t^E - Y_{t-1}^E = \theta\left(Y_t - Y_{t-1}^E\right)$. Найдите долгосрочную предельную склонность к потреблению.

Таблица 15.2. (Источник: Johnston and DiNardo's Econometric Methods (1997, 4th ed))

квартал	gas	price	income	miles
1959.1	−8.015248	4.67575	−4.50524	2.647592
1959.2	−8.01106	4.691292	−4.492739	2.647592
1959.3	−8.019878	4.689134	−4.498873	2.647592
1959.4	−8.012581	4.722338	−4.491904	2.647592
1960.1	−8.016769	4.70747	−4.490103	2.647415
1960.2	−7.976376	4.699136	−4.489107	2.647238
1960.3	−7.997135	4.72129	−4.492301	2.647061
1960.4	−8.005725	4.722736	−4.496271	2.646884
1961.1	−8.009368	4.706207	−4.489013	2.648654
1961.2	−7.989948	4.675196	−4.477735	2.650421
1961.3	−8.003017	4.694643	−4.469735	2.652185
1961.4	−7.999592	4.68604	−4.453429	2.653946
1962.1	−7.974048	4.671727	−4.446543	2.65377
1962.2	−7.972878	4.679437	−4.441757	2.653594
1962.3	−7.970209	4.668647	−4.439475	2.653418
1962.4	−7.963876	4.688853	−4.439333	2.653242
1963.1	−7.959317	4.675881	−4.437187	2.65148
1963.2	−7.951941	4.652961	−4.434306	2.649715
1963.3	−7.965396	4.658816	−4.427777	2.647946
1963.4	−7.960272	4.653714	−4.414002	2.646175
1964.1	−7.936954	4.645538	−4.39974	2.645998
1964.2	−7.92301	4.635629	−4.379863	2.64582
1964.3	−7.911883	4.631469	−4.371019	2.645643
1964.4	−7.918471	4.637514	−4.362371	2.645465
1965.1	−7.916095	4.652727	−4.360062	2.64582
1965.2	−7.894338	4.658141	−4.349969	2.646175
1965.3	−7.889203	4.656464	−4.327738	2.646529
1965.4	−7.871711	4.655432	−4.313128	2.646884
1966.1	−7.860066	4.644139	−4.309959	2.64511
1966.2	−7.840754	4.644212	−4.308563	2.643334
1966.3	−7.837658	4.647137	−4.29928	2.641554
1966.4	−7.838668	4.653788	−4.289879	2.639771
1967.1	−7.836081	4.659936	−4.278718	2.639057
1967.2	−7.830063	4.661554	−4.274659	2.638343
1967.3	−7.82532	4.654018	−4.270247	2.637628
1967.4	−7.812695	4.644671	−4.266353	2.636912
1968.1	−7.792308	4.641605	−4.256872	2.633506
1968.2	−7.781042	4.625596	−4.24436	2.630089
1968.3	−7.763533	4.627113	−4.24817	2.626659
1968.4	−7.766507	4.621535	−4.243286	2.623218
1969.1	−7.748152	4.620802	−4.247602	2.618672
1969.2	−7.728209	4.634679	−4.24156	2.614106
1969.3	−7.724049	4.61451	−4.225354	2.609518

Таблица 15.2. (продолжение)

квартал	gas	price	income	miles
1969.4	–7.707266	4.603419	–4.221179	2.604909
1970.1	–7.691294	4.589637	–4.223849	2.60343
1970.2	–7.696625	4.593972	–4.214468	2.601949
1970.3	–7.683176	4.57555	–4.207761	2.600465
1970.4	–7.68114	4.576285	–4.213643	2.598979
1971.1	–7.671161	4.563018	–4.20528	2.599722
1971.2	–7.660023	4.528258	–4.199421	2.600465
1971.3	–7.659501	4.5376	–4.20086	2.601207
1971.4	–7.659155	4.54487	–4.19967	2.601949
1972.1	–7.655547	4.51949	–4.206896	2.599351
1972.2	–7.657851	4.500299	–4.203082	2.596746
1972.3	–7.651443	4.515703	–4.187756	2.594135
1972.4	–7.634623	4.531717	–4.157555	2.591516
1973.1	–7.606151	4.531126	–4.150146	2.589642
1973.2	–7.625179	4.54356	–4.149023	2.587764
1973.3	–7.620612	4.540858	–4.144428	2.585882
1973.4	–7.628435	4.601014	–4.130499	2.583997
1974.1	–7.737227	4.73435	–4.155976	2.586259
1974.2	–7.703093	4.796311	–4.174081	2.588516
1974.3	–7.68182	4.773173	–4.174542	2.590767
1974.4	–7.647267	4.730467	–4.180104	2.593013
1975.1	–7.67028	4.721725	–4.198524	2.594882
1975.2	–7.673598	4.726068	–4.157104	2.596746
1975.3	–7.694903	4.766653	–4.175806	2.598607
1975.4	–7.689864	4.767389	–4.168096	2.600465
1976.1	–7.673608	4.74403	–4.158282	2.600836
1976.2	–7.66296	4.723825	–4.158494	2.601207
1976.3	–7.660979	4.723013	–4.159304	2.601578
1976.4	–7.651936	4.728776	–4.157702	2.601949
1977.1	–7.657642	4.731314	–4.160647	2.60694
1977.2	–7.64901	4.725569	–4.154308	2.611906
1977.3	–7.646001	4.705353	–4.139049	2.616848
1977.4	–7.649135	4.709094	–4.137531	2.621766
1978.1	–7.657363	4.699367	–4.129727	2.626298
1978.2	–7.642133	4.673999	–4.11695	2.630809
1978.3	–7.637718	4.678699	–4.113466	2.6353
1978.4	–7.644944	4.711411	–4.107745	2.639771
1979.1	–7.634155	4.737268	–4.10542	2.646352
1979.2	–7.684886	4.848168	–4.110224	2.65289
1979.3	–7.690778	4.965428	–4.108253	2.659385
1979.4	–7.699923	5.018858	–4.108714	2.665838
1980.1	–7.727037	5.130968	–4.107552	2.683928
1980.2	–7.747409	5.149823	–4.130193	2.701697

Таблица 15.2. (продолжение)

квартал	gas	price	income	miles
1980.3	–7.768291	5.121161	–4.123072	2.719155
1980.4	–7.770349	5.108043	–4.106715	2.736314
1981.1	–7.746104	5.176053	–4.107763	2.743739
1981.2	–7.749401	5.162098	–4.113578	2.75111
1981.3	–7.752783	5.128757	–4.10358	2.758426
1981.4	–7.758903	5.126579	–4.10885	2.76569
1982.1	–7.748258	5.089216	–4.116735	2.776643
1982.2	–7.740548	5.016708	–4.109173	2.787477
1982.3	–7.760993	5.046284	–4.1106	2.798196
1982.4	–7.765389	5.018583	–4.112362	2.8088
1983.1	–7.734547	4.951649	–4.111137	2.816157
1983.2	–7.761149	4.973684	–4.105544	2.82346
1983.3	–7.738656	4.978372	–4.09589	2.83071
1983.4	–7.72975	4.947877	–4.079108	2.837908
1984.1	–7.743051	4.937271	–4.059039	2.847957
1984.2	–7.723525	4.926296	–4.050855	2.857906
1984.3	–7.719794	4.885598	–4.04159	2.867757
1984.4	–7.715204	4.886159	–4.039545	2.877512
1985.1	–7.721207	4.876889	–4.040108	2.882564
1985.2	–7.717261	4.896123	–4.021586	2.88759
1985.3	–7.71862	4.877967	–4.034639	2.892591
1985.4	–7.71905	4.865803	–4.03058	2.897568
1986.1	–7.702826	4.79969	–4.020278	2.898395
1986.2	–7.689694	4.598733	–4.006467	2.899221
1986.3	–7.685297	4.52283	–4.012002	2.900047
1986.4	–7.670334	4.488154	–4.016886	2.900872
1987.1	–7.685728	4.581026	–4.011549	2.913573
1987.2	–7.664967	4.594879	–4.030798	2.926114
1987.3	–7.682947	4.624788	–4.019994	2.9385
1987.4	–7.672442	4.617012	–4.00732	2.950735
1988.1	–7.678686	4.583605	–3.996853	2.959328
1988.2	–7.669295	4.572124	–3.997311	2.967847
1988.3	–7.675229	4.578095	–3.993513	2.976295
1988.4	–7.657843	4.557322	–3.985354	2.984671
1989.1	–7.670982	4.561117	–3.979249	2.990091
1989.2	–7.713211	4.683528	–3.988068	2.995482
1989.3	–7.675435	4.627944	–3.985719	3.000844
1989.4	–7.636929	4.584531	–3.980442	3.006177
1990.1	–7.672857	4.628345	–3.971611	3.013695
1990.2	–7.706511	4.617825	–3.969884	3.021156
1990.3	–7.710044	4.693574	–3.971754	3.028562
1990.4	–7.717076	4.829451	–3.978981	3.035914

Таблица 15.3. (Источник: Greene W., Econometric Analysis, 3rd. edition, Macmillan, 1997)

Год, квартал	C	Y
1953.1	362.8	395.5
1953.2	364.6	401.0
1953.3	363.6	399.7
1953.4	362.6	400.2
1954.1	363.5	399.7
1954.2	366.2	397.3
1954.3	371.8	403.8
1954.4	378.6	411.8
1955.1	385.2	414.7
1955.2	392.2	423.8
1955.3	396.4	430.8
1955.4	402.6	437.6
1956.1	403.2	441.2
1956.2	403.9	444.7
1956.3	405.1	446.6
1956.4	409.3	452.7
1957.1	411.7	452.6
1957.2	412.4	455.4
1957.3	415.2	457.9
1957.4	416.0	456.0
1958.1	411.0	452.1
1958.2	414.7	455.1
1958.3	420.9	464.6
1958.4	425.2	471.3
1959.1	424.1	474.5
1959.2	439.7	482.2

Год, квартал	C	Y
1959.3	443.3	479.0
1959.4	444.6	483.1
1960.1	448.1	487.8
1960.2	454.1	490.7
1960.3	452.7	491.0
1960.4	453.2	488.8
1961.1	454.0	493.4
1961.2	459.9	500.7
1961.3	461.4	505.5
1961.4	470.3	514.8
1962.1	474.5	519.5
1962.2	479.8	523.9
1962.3	483.7	526.7
1962.4	490.0	529.0
1963.1	493.1	533.3
1963.2	497.4	538.9
1963.3	503.9	544.4
1963.4	507.5	552.5
1964.1	516.6	563.6
1964.2	525.6	579.4
1964.3	534.3	586.4
1964.4	535.3	593.0
1965.1	546.0	599.7
1965.2	550.7	607.8
1965.3	559.2	623.6
1965.4	573.9	634.6

Год, квартал	C	Y
1966.1	581.2	639.7
1966.2	582.3	642.0
1966.3	588.6	649.2
1966.4	590.5	700.7
1967.1	594.8	665.0
1967.2	602.4	671.3
1967.3	605.2	676.5
1967.4	608.2	682.0
1968.1	620.7	690.4
1968.2	629.9	701.9
1968.3	642.3	703.6
1968.4	644.7	708.7
1969.1	651.9	710.4
1969.2	656.2	717.0
1969.3	659.6	730.1
1969.4	663.9	733.2
1970.1	667.4	737.1
1970.2	670.5	752.6
1970.3	676.5	759.7
1970.4	673.9	756.1
1971.1	687.0	771.3
1971.2	693.3	779.7
1971.3	698.2	781.0
1971.4	708.6	785.5
1972.1	718.6	791.7
1972.2	731.1	798.5

Год, квартал	C	Y
1972.3	741.3	842.2
1972.4	757.1	838.1
1973.1	768.8	855.0
1973.2	766.3	862.1
1973.3	769.7	868.0
1973.4	766.7	873.4
1974.1	761.2	859.9
1974.2	764.1	859.7
1974.3	769.4	859.7
1974.4	756.5	851.1
1975.1	763.3	845.1
1975.2	775.6	891.3
1975.3	785.4	878.4
1975.4	793.3	884.9
1976.1	809.9	899.3
1976.2	817.1	904.1
1976.3	826.5	908.8
1976.4	838.9	914.9
1977.1	851.7	919.6
1977.2	858.0	934.1
1977.3	867.3	951.9
1977.4	880.4	965.9
1978.1	883.8	973.5
1978.2	901.1	982.6
1978.3	908.6	994.2
1978.4	919.2	1005.0

Год, квартал	C	Y
1979.1	921.2	1011.1
1979.2	919.5	1011.8
1979.3	930.9	1019.7
1979.4	938.6	1020.2
1980.1	938.3	1025.9
1980.2	919.6	1011.8
1980.3	929.4	1019.3
1980.4	940.0	1030.2
1981.1	950.2	1044.0
1981.2	949.1	1041.0
1981.3	955.7	1058.4
1981.4	946.8	1056.0
1982.1	953.7	1052.8
1982.2	958.9	1054.7
1982.3	964.2	1057.7
1982.4	976.3	1067.5
1983.1	982.5	1073.3
1983.2	1006.2	1082.2
1983.3	1015.6	1102.1
1983.4	1032.4	1124.4
1984.1	1044.1	1147.8
1984.2	1064.2	1165.3
1984.3	1065.9	1176.7
1984.4	1075.4	1186.9

4.2. Оцените по тем же данным модель частичного приспособления, преобразовав ее в ADL(1, 0). Желаемый уровень потребления C_t^D зависит от текущего дохода Y_t: $C_t^D = \beta + \alpha Y_t + \xi_t$, где C_t^D — ненаблюдаемая переменная, задаваемая уравнением частичного приспособления $C_t - C_{t-1} = \gamma\left(C_t^D - C_{t-1}\right) + \varepsilon_t$. Найдите долгосрочную предельную склонность к потреблению.

Задачи

1. С помощью какого метода можно оценить параметры модели распределенного лага (конечного)?

2. В чем основная причина перехода от обычной модели распределенного лага к модели полиномиального лага?

3. В чем основная причина перехода от обычной модели распределенного лага к модели с экспоненциальным лагом?

4. Пусть веса в модели распределенного лага экспоненциально убывают и сам лаг бесконечный. Каким преобразованием можно перейти к авторегрессионной модели с распределенным лагом? В чем особенность ошибок в этой модели?

5. Процесс с геометрическим лагом задан формулой

$$x_t = 1 + \frac{1}{2}z_t + \frac{1}{4}z_{t-1} + \frac{1}{8}z_{t-2} + \ldots + \varepsilon.$$

   Примените к нему преобразование Койка.

6. Примените обратное преобразование Койка к модели ADL(1, 0).

7. Представьте ожидания переменной X в модели адаптивных ожиданий в виде распределенного лага для фактически наблюдаемых значений переменной X.

8. Для процесса $x_t = 0.1 + 0.5x_{t-1} + 0.1z_t + 0.2z_{t-1} + \varepsilon_t$ запишите долгосрочную зависимость между x и z.

9. Запишите следующие процессы в виде модели исправления ошибок:

   а) $x_t = 0.5x_{-1} + 0.7z_{t-1} + \varepsilon_t$;

   б) $x_t = 2 + 0.4x_{t-1} + 2z_t + 3z_{t-1} - 3z_{t-2} + \varepsilon_t$.

Рекомендуемая литература

1. **Доугерти К.** Введение в эконометрику. — М.: «Инфра-М», 1997. (Гл. 10).
2. **Драймз Ф.** Распределенные лаги. Проблемы выбора и оценивания модели. — М.: «Финансы и статистика», 1982.
3. **Магнус Я.Р., Катышев П.К., Пересецкий А.А.** Эконометрика — начальный курс. — М.: «Дело», 2004. (Гл. 12).
4. **Маленво Э.** Статистические методы эконометрии. Вып. 2. — М.: «Статистика», 1976. (Гл. 15).
5. **Песаран М., Слейтер Л.** Динамическая регрессия: теория и алгоритмы. — М: «Финансы и статистика», 1984. (Гл. 5, стр. 67–91).
6. **Baltagi, Badi H.** Econometrics, 2nd edition. — Springer, 1999. (Ch. 6).
7. **Enders W.** Applied Econometric Time Series. — New York: John Wiley & Sons, 1992.
8. **Greene W.H.** Econometric Analysis. — Prentice-Hall, 2000. (гл.17)
9. **Judge G.G., Griffiths W.E., Hill R.C., Luthepohl H., Lee T.** Theory and Practice of Econometrics. — New York: John Wiley & Sons, 1985. (Ch. 9, 10).

Глава 16

Модели с авторегрессионной условной гетероскедастичностью

Традиционные модели временных рядов, такие как модель ARMA, не могут адекватно учесть все характеристики, которыми обладают финансовые временные ряды, и требуют расширения. Одна из характерных особенностей финансовых рынков состоит в том, что присущая рынку неопределенность изменяется во времени. Как следствие, наблюдается «кластеризация волатильности». Имеется в виду чередование периодов, когда финансовый показатель ведет себя непостоянно и относительно спокойно. На рисунке 16.1 для иллюстрации этого явления показаны темпы прироста индекса РТС[1] за несколько лет. На графике период **1** — сравнительно спокойный, период **2** — более бурный, период **3** — опять спокойный. Термин **волатильность** (*volatility* — англ. изменчивость, непостоянство) используется, как правило, для неформального обозначения степени вариабельности, разброса переменной. Формальной мерой волатильности служит дисперсия (или среднеквадратическое отклонение). Эффект кластеризации волатильности отмечен в таких рядах, как изменение цен акций, валютных курсов, доходности спекулятивных активов.

[1] Фондовый индекс Российской Торговой Системы. См. http://www.rts.ru.

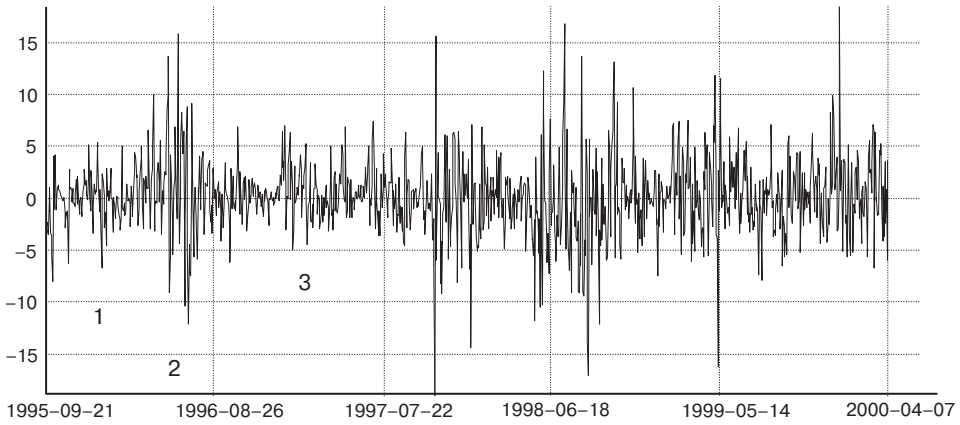

Рис. 16.1. Темпы прироста индекса РТС с 21 сентября 1995 г. по 7 апреля 2000 г., в процентах.

16.1. Модель ARCH

Модель ARCH, т.е. модель с авторегрессионной условной гетероскедастичностью (autoregressive conditional heteroskedasticity), предложена **Р. Энглом** в 1982 г. для моделирования кластеризации волатильности. Процесс ARCH q-го порядка, $\{\varepsilon_t\}_{t=-\infty}^{+\infty}$, задается следующими соотношениями:

$$\begin{aligned} &\varepsilon_t|\Omega_{t-1} \sim N(0,\sigma_t^2),\\ &\sigma_t^2 = \omega + \gamma_1\varepsilon_{t-1}^2 + \ldots + \gamma_q\varepsilon_{t-q}^2. \end{aligned} \tag{16.1}$$

Здесь $\Omega_{t-1} = (\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots)$ — предыстория процесса $\{\varepsilon_t\}$, а σ_t^2 — условная по предыстории дисперсия ε_t, т.е. $\sigma_t^2 = \mathbf{var}(\varepsilon_t|\Omega_{t-1}) = \mathbf{E}(\varepsilon_t^2|\Omega_{t-1})$. Условную дисперсию часто называют волатильностью процесса. Для того чтобы условная дисперсия оставалась положительной, требуется выполнение соотношений $\omega > 0$ и $\gamma_1, \ldots, \gamma_q \geqslant 0$.

Данный процесс можно записать несколько иначе:

$$\begin{aligned} &\xi_t \sim NID(0,1),\\ &\varepsilon_t = \xi_t\sigma_t,\\ &\sigma_t^2 = \omega + \gamma_1\varepsilon_{t-1}^2 + \ldots + \gamma_q\varepsilon_{t-q}^2. \end{aligned}$$

Аббревиатура *NID* означает, что ξ_t нормально распределены и независимы. Такая запись удобна тем, что нормированный случайный процесс ξ_t не зависит от предыстории.

Смысл модели ARCH состоит в том, что если абсолютная величина ε_t оказывается большой, то это приводит к повышению условной дисперсии в последующие

периоды. В свою очередь, при высокой условной дисперсии более вероятно появление больших (по абсолютной величине) значений ε_t. Наоборот, если значения ε_t в течение нескольких периодов близки к 0, то это приводит к понижению условной дисперсии в последующие периоды практически до уровня ω. В свою очередь, при низкой условной дисперсии более вероятно появление малых (по абсолютной величине) значений ε_t. Таким образом, ARCH-процесс характеризуется инерционностью условной дисперсии (кластеризацией волатильности).

Несложно показать, что процесс ARCH не автокоррелирован:

$$\mathbf{E}(\varepsilon_t\varepsilon_{t-j}) = \mathbf{E}\left(\mathbf{E}(\varepsilon_t\varepsilon_{t-j}|\Omega_{t-1})\right) = \mathbf{E}\left(\varepsilon_{t-j}\mathbf{E}(\varepsilon_t|\Omega_{t-1})\right) = 0.$$

Поскольку процесс имеет постоянное (нулевое) математическое ожидание и не автокоррелирован, то он является слабо стационарным в случае, если у него есть дисперсия.

Если обозначить разницу между величиной ε_t^2 и ее условным математическим ожиданием, σ_t^2, через η_t, то получится следующая эквивалентная запись процесса ARCH:

$$\varepsilon_t^2 = \omega + \gamma_1\varepsilon_{t-1}^2 + \ldots + \gamma_q\varepsilon_{t-q}^2 + \eta_t. \tag{16.2}$$

Поскольку условное математическое ожидание η_t равно 0, то безусловное математическое ожидание также равно 0. Кроме того, как можно показать, $\{\eta_t\}$ не автокоррелирован. Следовательно, квадраты процесса ARCH(q) следуют авторегрессионному процессу q-го порядка.

Если все корни характеристического уравнения

$$1 - \gamma_1 z - \ldots - \gamma_q z^q = 0$$

лежат за пределами единичного круга, то у процесса ARCH(q) существует безусловная дисперсия, и он является слабо стационарным. Поскольку коэффициенты γ_j неотрицательны, то это условие эквивалентно условию $\sum_{j=1}^{q}\gamma_j < 1$.

Действительно, вычислим безусловную дисперсию стационарного ARCH-процесса, которую мы обозначим через σ^2. Для этого возьмем математическое ожидание от обеих частей уравнения условной дисперсии (16.1):

$$\mathbf{E}(\sigma_t^2) = \omega + \gamma_1\mathbf{E}(\varepsilon_{t-1}^2) + \ldots + \gamma_q\mathbf{E}(\varepsilon_{t-q}^2).$$

Заметим, что $\mathbf{E}\left(\sigma_t^2\right) = \mathbf{E}\left(\mathbf{E}(\varepsilon_t^2|\Omega_{t-1})\right) = \mathbf{E}\left(\varepsilon_t^2\right) = \mathbf{var}(\varepsilon_t^2) = \sigma^2$, т.е. математическое ожидание условной дисперсии равно безусловной дисперсии. Следовательно,

$$\sigma^2 = \omega + \gamma_1\sigma^2 + \ldots + \gamma_q\sigma^2,$$

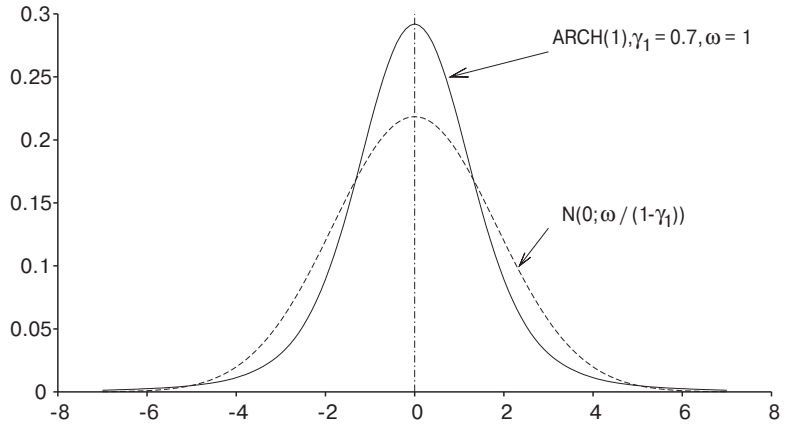

Рис. 16.2. Плотность ARCH(1) и плотность нормального распределения с той же дисперсией

или

$$\sigma^2 = \frac{\omega}{1 - \gamma_1 - \ldots - \gamma_q}.$$

Таким образом, для всех ε_t безусловная дисперсия одинакова, т.е. имеет место гомоскедастичность. Однако условная дисперсия меняется, поэтому одновременно имеет место **условная гетероскедастичность**[2].

Если не все корни приведенного выше характеристического уравнения лежат за пределами единичного круга, т.е. если $\sum\limits_{j=1}^{q} \gamma_j \geqslant 1$, то безусловная дисперсия не существует, и поэтому ARCH-процесс не будет слабо стационарным[3].

Еще одно свойство ARCH-процессов состоит в том, что безусловное распределение ε_t имеет более высокий куртозис (т.е. более толстые хвосты и острую вершину), чем нормальное распределение (определение куртозиса и эксцесса см. в Приложении А.3.1). У ARCH(1) эксцесс равен

$$\frac{\mathbf{E}(\varepsilon_t^4)}{\sigma^4} - 3 = \frac{6\gamma_1^2}{1 - 3\gamma_1^2},$$

[2]Она называется авторегрессионной, поскольку динамика квадратов ARCH-процесса описывается авторегрессией.

[3]При этом у ARCH-процессов есть интересная особенность: они могут быть строго стационарны, не будучи слабо стационарны. Дело в том, что определение слабой стационарности требует существования конечных первых и вторых моментов ряда. Строгая же стационарность этого не требует, поэтому даже если условная дисперсия бесконечна (и, следовательно, ряд не является слабо стационарным), ряд все же может быть строго стационарным.

причем при $3\gamma_1^2 \geqslant 1$ четвертый момент распределения не существует (эксцесс равен бесконечности). Это свойство ARCH-процессов хорошо соответствует финансовым временным рядам, которые обычно характеризуются толстыми хвостами. На рисунке 16.2 изображен график плотности безусловного распределения ARCH(1). Для сравнения на графике приведена плотность нормального распределения с той же дисперсией.

Получить состоятельные оценки коэффициентов ARCH-процесса можно, используя вышеприведенное представление его квадратов в виде авторегрессии (16.2). Более эффективные оценки получаются при использовании метода максимального правдоподобия.

При применении ARCH-моделей к реальным данным было замечено, что модель ARCH(1) не дает достаточно длительных кластеров волатильности, а только порождает большое число выбросов (выделяющихся наблюдений). Для корректного описания данных требуется довольно большая длина лага q, что создает трудности при оценивании. В частности, зачастую нарушается условие неотрицательности оценок коэффициентов γ_j. Поэтому Энгл предложил использовать модель со следующими ограничениями на коэффициенты лага: они задаются с помощью весов вида:

$$w_j = \frac{q+1-j}{0.5q(q+1)},$$

сумма которых равна 1 и которые линейно убывают до нуля. Сами коэффициенты берутся равными $\gamma_j = \gamma w_j$. Получается модель с двумя параметрами, ω и γ:

$$\sigma_t^2 = \omega + \gamma\left(w_1\varepsilon_{t-1}^2 + \ldots + w_q\varepsilon_{t-q}^2\right).$$

16.2. Модель GARCH

Модель GARCH (*generalized ARCH* — обобщенная модель ARCH), предложенная **Т. Боллерслевом**, является альтернативной модификацией модели ARCH (16.2), позволяющей получить более длинные кластеры при малом числе параметров. Модель ARMA зачастую позволяет получить более сжатое описание временных зависимостей для условного математического ожидания, чем модель AR. Подобным же образом модель GARCH дает возможность обойтись меньшим количеством параметров по сравнению с моделью ARCH, если речь идет об условной дисперсии. В дальнейшем мы проведем прямую аналогию между моделями GARCH и ARMA.

Для того чтобы вывести модель GARCH, используем в модели ARCH бесконечный геометрический лаг:

$$\sigma_t^2 = \omega + \gamma \sum_{j=1}^{\infty} \delta^{j-1} \varepsilon_{t-j}^2 = \omega + \frac{\gamma}{1 - \delta \mathbf{L}} \varepsilon_{t-1}^2.$$

Применяя преобразование Койка, получим

$$\sigma_t^2 = (1 - \delta)\omega + \delta \sigma_{t-1}^2 + \gamma \varepsilon_{t-1}^2.$$

Поменяв очевидным образом обозначения, получим модель GARCH(1, 1):

$$\sigma_t^2 = \omega + \delta \sigma_{t-1}^2 + \gamma \varepsilon_{t-1}^2.$$

Модель GARCH(p, q) обобщает эту формулу:

$$\sigma_t^2 = \omega + \delta_1 \sigma_{t-1}^2 + \ldots + \delta_p \sigma_{t-p}^2 + \gamma_1 \varepsilon_{t-1}^2 + \ldots + \gamma_q \varepsilon_{t-q}^2 = \\ = \omega + \sum_{j=1}^{p} \delta_j \sigma_{t-j}^2 + \sum_{j=1}^{q} \gamma_j \varepsilon_{t-j}^2. \quad (16.3)$$

При этом предполагается, что $\omega > 0$, $\delta_1, \ldots, \delta_p \geqslant 0$ и $\gamma_1, \ldots, \gamma_q \geqslant 0$. На практике, как правило, достаточно взять $p = 1$ и $q = 1$. Изредка используют GARCH(1, 2) или GARCH(2, 1).

Как и в модели ARCH, σ_t^2 служит условной дисперсией процесса:

$$\varepsilon_t | \Omega_t \sim N(0, \sigma_t^2).$$

Рассчитаем безусловную дисперсию GARCH-процесса, предполагая, что он стационарен. Для этого возьмем математические ожидания от обеих частей уравнения 16.3 для условной дисперсии:

$$\mathbf{E}(\sigma_t^2) = \sum_{j=1}^{p} \delta_j \mathbf{E}(\sigma_{t-j}^2) + \sum_{j=1}^{q} \gamma_j \mathbf{E}(\varepsilon_{t-j}^2),$$

откуда

$$\sigma^2 = \sum_{j=1}^{p} \delta_j \sigma^2 + \sum_{j=1}^{q} \gamma_j \sigma^2$$

и

$$\sigma^2 = \frac{1}{1 - \sum\limits_{j=1}^{p} \delta_j - \sum\limits_{j=1}^{q} \gamma_j}.$$

Таким образом, с точки зрения безусловной дисперсии GARCH-процесс гомоскедастичен.

Для того чтобы дисперсия была конечной, необходимо выполнение условия $\sum\limits_{j=1}^{p} \delta_j + \sum\limits_{j=1}^{q} \gamma_j < 1$. В частности, для модели GARCH(1, 1) требуется $\delta_1 + \gamma_1 < 1$.

Процесс GARCH можно записать в эквивалентной форме, если, как и выше, в уравнении 16.2 для модели ARCH, обозначить $\eta_t = \varepsilon_t^2 - \sigma_t^2$:

$$\varepsilon_t^2 = \omega + \sum_{j=1}^{m} (\delta_j + \gamma_j)\varepsilon_{t-j}^2 + \eta_t - \sum_{j=1}^{p} \delta_j \eta_{t-j},$$

где $m = \max(p, q)$. (В этой записи подразумевается $\delta_j = 0$ при $j > p$ и $\gamma_j = 0$ при $j > q$.) Такая форма записи позволяет увидеть, что квадраты GARCH-процесса подчиняются модели ARMA(m, p).

Этот факт дает возможность получить автокорреляционную функцию квадратов GARCH-процесса. В частности, для GARCH(1, 1) автокорреляционная функция квадратов имеет вид

$$\rho_1 = \frac{\gamma_1(1 - \delta_1^2 - \delta_1\gamma_1)}{1 - \delta_1^2 - 2\delta_1\gamma_1},$$
$$\rho_\tau = (\delta_1 + \gamma_1)^{\tau-1}\rho_{\tau-1}, \quad \tau > 1.$$

Условие существования безусловного четвертого момента у отдельного наблюдения процесса GARCH(1, 1) состоит в том, что $3\gamma_1^2 + 2\gamma_1\delta_1 + \delta_1^2 < 1$. Если это условие выполняется, то эксцесс равен

$$\frac{\mathbf{E}(\varepsilon_t^4)}{\sigma^4} - 3 = \frac{6\gamma_1^2}{1 - \delta_1^2 - 2\delta_1\gamma_1 - 3\gamma_1^2}$$

и является положительным. То есть GARCH-процесс (как и его частный случай — ARCH-процесс) имеет более высокий куртозис, чем нормальное распределение. В то же время, безусловное распределение отдельного наблюдения GARCH-процесса является симметричным, поэтому все нечетные моменты, начиная с третьего, равны нулю.

Стандартным методом оценивания для моделей GARCH является метод максимального правдоподобия. Условно по предыстории Ω_{t-1} отдельное наблюдение GARCH-процесса распределено нормально: $\varepsilon_t|\Omega_{t-1} \sim N(0, \sigma_t^2)$. Функция правдоподобия для ряда $\varepsilon_1, \ldots, \varepsilon_T$, подчиняющегося GARCH-процессу, вычисляется как произведение плотностей этих условных нормальных распределений:

$$L = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left(-\frac{\varepsilon_t^2}{2\sigma_t^2}\right).$$

Максимизируя эту функцию правдоподобия по неизвестным параметрам, получим оценки максимального правдоподобия для GARCH-процесса.

При оценивании условную дисперсию σ_t^2 следует считать функцией параметров модели и вычислять по рекуррентной формуле 16.3. Для этих вычислений требуются «довыборочные» значения самого процесса и его условной дисперсии, а они неизвестны. Для решения этой проблемы можно использовать различные приемы. Самый простой, по-видимому, состоит в том, чтобы заменить условные дисперсии в начале ряда ($t = 1, \ldots, m$) оценкой безусловной дисперсии, т.е. величиной

$$s^2 = \frac{1}{T}\sum_{t=1}^{T} \varepsilon_t^2.$$

Оценки максимального правдоподобия являются состоятельными и асимптотически эффективными.

На практике модель GARCH дополняют какой-либо моделью, описывающей поведение условного или безусловного математического ожидания наблюдаемого ряда. Например, можно предположить, что наблюдается не ε_t, а ε_t плюс константа, т.е. что наблюдаемый ряд $\{x_t\}$ имеет постоянное безусловное математическое ожидание β, к которому добавляется ошибка ε_t в виде процесса GARCH:

$$x_t = \beta + \varepsilon_t.$$

Можно моделировать безусловное математическое ожидание с помощью линейной регрессии, т.е.

$$x_t = Z_t\alpha + \varepsilon_t.$$

Это позволяет учитывать линейный тренд, детерминированные сезонные переменные и т.п. При оценивании в функции правдоподобия вместо ε_t используют $x_t - Z_t\alpha$.

С точки зрения прогнозирования перспективной является модель, сочетающая ARIMA с GARCH. Модель ARIMA в этом случае используется для моделирования поведения условного математического ожидания ряда, а GARCH — для моделирования условной дисперсии.

Важнейший вывод, который следует из анализа модели ARCH, состоит в том, что наблюдаемые изменения в дисперсии (волатильности) временного ряда могут иметь эндогенный характер, то есть порождаться определенной нелинейной моделью, а не какими-то внешними структурными сдвигами.

16.3. Прогнозы и доверительные интервалы для модели GARCH

Одна из важнейших целей эконометрических моделей временных рядов — построение прогнозов. Какие преимущества дают модели с авторегрессионной условной гетероскедастичностью с точки зрения прогнозирования временных рядов по сравнению с моделями линейной регрессии или авторегрессии — скользящего среднего? Оказывается, что прямых преимуществ нет, но есть ряд опосредованных преимуществ, которые в отдельных случаях могут иметь большое значение.

Рассмотрим модель линейной регрессии

$$x_t = Z_t\alpha + \varepsilon_t, \quad t = 1, \dots, T,$$

в которой ошибка представляет собой GARCH-процесс. Поскольку ошибки не автокоррелированы и гомоскедастичны, то, как известно, оценки наименьших квадратов являются наилучшими в классе линейных по x несмещенных оценок. Однако наличие условной гетероскедастичности позволяет найти более эффективные (т.е. более точные) оценки среди нелинейных и смещенных оценок. Действительно, метод максимального правдоподобия дает асимптотически эффективные оценки, более точные, чем оценки МНК. В ошибку прогноза вносит свой вклад, во-первых, ошибка ε_{T+1}, а во-вторых, разница между оценками параметров и истинными значениями параметров. Использование более точных оценок позволяет уменьшить в некоторой степени вторую составляющую ошибки прогноза.

В обычных моделях временного ряда с неизменными условными дисперсиями (например, ARMA) неопределенность ошибки прогноза — это возрастающая функция горизонта прогноза, которая (если не учитывать разницу между оценками параметров и истинными значениями параметров, отмеченную в предыдущем абзаце) не зависит от момента прогноза. Однако в присутствии ARCH-ошибок точность прогноза будет нетривиально зависеть от текущей информации и, следовательно, от момента прогноза. Поэтому для корректного построения интервальных прогнозов требуется иметь оценки будущих условных дисперсий ошибки.

Кроме того, в некоторых случаях полезно иметь прогнозы не только (условного) математического ожидания изучаемой переменной, но и ее (условной) дисперсии. Это важно, например, при принятии решений об инвестициях в финансовые активы. В этом случае дисперсию (волатильность) доходности естественно рассматривать как меру рискованности финансового актива. Таким образом, сами по себе прогнозы условной дисперсии могут иметь практическое применение.

Покажем, что доверительный интервал прогноза зависит от предыстории

$$\Omega_T = (x_T,\ x_{T-1},\ \dots,\ x_1,\ \dots).$$

Реально прогноз делается на основе имеющегося ряда $(x_1, \ldots, x_T)$, а не всей предыстории, однако различие это не столь существенно. При этом мы будем исходить из того, что нам известны истинные параметры процесса. Прогноз на τ периодов — это математическое ожидание прогнозируемой величины $x_{T+\tau}$, условное относительно имеющейся на момент t информации Ω_T. Он равен

$$x_T(\tau) = \mathbf{E}(x_{T+\tau}|\Omega_T) = \mathbf{E}(Z_{T+\tau}\alpha + \varepsilon_{T+\tau}|\Omega_T) = Z_{T+\tau}\alpha.$$

Здесь учитывается, что поскольку информация Ω_T содержится в информации $\Omega_{T+\tau-1}$ при $\tau \geqslant 1$, то по правилу повторного взятия математического ожидания выполнено

$$\mathbf{E}(\varepsilon_{T+\tau}|\Omega_T) = \mathbf{E}\left(\mathbf{E}\left(\varepsilon_{T+\tau}|\Omega_{T+\tau-1}\right)|\Omega_T\right) = \mathbf{E}(0|\Omega_T) = 0.$$

Таким образом, если известны истинные параметры, присутствие GARCH-ошибок не отражается на том, как строится точечный прогноз, — он оказывается таким же, как для обычной линейной регрессии. Ошибка предсказания в момент времени T на τ шагов вперед

$$d_T(\tau) = x_{T+\tau} - x_T(\tau) = \varepsilon_{T+\tau}.$$

Условная дисперсия ошибки предсказания равна

$$\sigma_p^2 = \mathbf{E}\left(d_T^2(\tau)|\Omega_T\right) = \mathbf{E}\left(\varepsilon_{T+\tau}^2|\Omega_T\right).$$

Из этого следует, что она зависит как от горизонта прогноза τ, так и от предыстории Ω_T.

Заметим, что при $t > T$ выполнено $\mathbf{E}\left(\varepsilon_t^2|\Omega_T\right) = \mathbf{E}\left(\sigma_t^2|\Omega_T\right)$, поскольку

$$\mathbf{E}\left(\varepsilon_t^2 - \sigma_t^2|\Omega_T\right) = \mathbf{E}\left(\mathbf{E}\left(\varepsilon_t^2 - \sigma_t^2|\Omega_{t-1}\right)|\Omega_T\right) = \mathbf{E}(0|\Omega_T) = 0.$$

Учитывая, что $\mathbf{E}\left(\varepsilon_t^2 - \sigma_t^2|\Omega_{t-1}\right) = 0$ и информация Ω_T содержится в информации Ω_{t-1} при $t > T$, применяем правило повторного взятия математического ожидания:

$$\sigma_p^2 = \mathbf{E}\left(\varepsilon_{T+\tau}^2|\Omega_T\right) = \mathbf{E}\left(\sigma_{T+\tau}^2|\Omega_T\right).$$

Таким образом, фактически дисперсия прогноза $x_{T+\tau}$ — это прогноз волатильности на τ шагов вперед.

Возьмем от обеих частей рекуррентного уравнения (16.3) для GARCH-процесса математическое ожидание, условное относительно Ω_T. Получим

$$\mathbf{E}\left(\sigma_t^2|\Omega_T\right) = \omega + \sum_{j=1}^{p}\delta_j\mathbf{E}\left(\sigma_{t-j}^2|\Omega_T\right) + \sum_{j=1}^{q}\gamma_j\mathbf{E}\left(\varepsilon_{t-j}^2|\Omega_T\right). \qquad (16.4)$$

Можно использовать эту рекуррентную формулу для расчета $\mathbf{E}\left(\sigma_t^2|\Omega_T\right)$ при $t > T$. При этом следует учесть, что $\mathbf{E}\left(\varepsilon_t^2|\Omega_T\right) = \varepsilon_t^2$ при $t \leqslant T$, поскольку информация об ε_t содержится в информационном множестве Ω_T, и по той же причине $\mathbf{E}\left(\sigma_t^2|\Omega_T\right) = \sigma_t^2$ при $t \leqslant T+1$. Кроме того, как только что было доказано, $\mathbf{E}\left(\varepsilon_t^2|\Omega_T\right) = \mathbf{E}\left(\sigma_t^2|\Omega_T\right)$ при $t > T$.

Таким образом, имеются все данные для того, чтобы с помощью формулы (16.4) рассчитать дисперсию ошибки прогноза для $x_{T+\tau}$ в модели GARCH. При $\tau = 1$ можно сразу без применения (16.4) записать

$$\sigma_{d_T(1)}^2 = \mathbf{E}\left(\sigma_{T+1}^2|\Omega_T\right) = \sigma_{T+1}^2,$$

где σ_{T+1}^2 рассчитывается по обычному правилу. В модели GARCH(1, 1) при $\tau > 1$ по формуле (16.4)

$$\mathbf{E}\left(\sigma_{T+\tau}^2|\Omega_T\right) = \omega + (\delta_1 + \gamma_1)\mathbf{E}\left(\sigma_{T+\tau-1}^2|\Omega_T\right),$$

т.е.

$$\sigma_{d_T(\tau)}^2 = \omega + (\delta_1 + \gamma_1)\sigma_{d_T(\tau-1)}^2.$$

Отсюда следует, что общее выражение для GARCH(1, 1), не подходящее только для случая $\delta_1 + \gamma_1 = 1$, имеет вид

$$\sigma_{d_T(\tau)}^2 = \omega\frac{1 - (\delta_1 + \gamma_1)^{\tau-1}}{1 - \delta_1 - \gamma_1} + (\delta_1 + \gamma_1)^{\tau-1}\sigma_{T+1}^2.$$

В пределе при условии стационарности $\delta_1 + \gamma_1 < 1$ условная дисперсия ошибки прогноза сходится к безусловной дисперсии процесса GARCH(1, 1):

$$\lim_{\tau\to\infty}\sigma_{d_T(\tau)}^2 = \frac{\omega}{1 - \delta_1 - \gamma_1}.$$

Хотя получено общее выражение для дисперсии ошибки прогноза, этого, вообще говоря, недостаточно для корректного построения доверительных интервалов, поскольку условное относительно Ω_T распределение $\varepsilon_{T+\tau}$, а следовательно, и распределение ошибки прогноза $d_T(\tau)$, имеет более толстые хвосты, чем нормальное распределение. Чтобы обойти эту проблему, можно использовать, например, прогнозные интервалы в виде плюс/минус двух среднеквадратических ошибок прогноза без выяснения того, какой именно доверительной вероятности это соответствует[4], т.е. $x_T(\tau) \pm 2\sigma_{d_T(\tau)}$.

[4] Ясно, что для нормального распределения это примерно 95 %-й двусторонний квантиль.

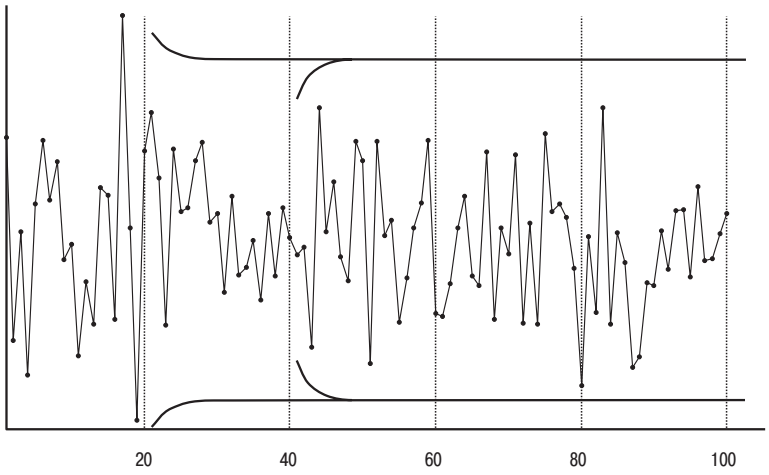

Рис. 16.3. Иллюстрация интервальных прогнозов для процесса GARCH

Чтобы проиллюстрировать зависимость доверительных интервалов прогнозов от предыстории, мы сгенерировали ряд GARCH(1, 1) длиной 100 наблюдений с параметрами $\delta_1 = 0.3$ и $\gamma_1 = 0.3$ и построили теоретические доверительные интервалы при $T = 20$ и $T = 40$. Прогноз везде равен нулю. Рисунок 16.3 показывает условные доверительные интервалы прогнозов для процесса GARCH(1, 1), а также сам ряд. Интервал для $T = 20$ постепенно сужается, а для $T = 40$ расширяется до уровня, соответствующего безусловной дисперсии. Такое поведение объясняется тем, что при $T = 21$ волатильность (условная дисперсия) была относительно высокой, а при $T = 41$ — относительно низкой. Очевидна способность условных прогнозных интервалов приспосабливаться к изменениям в волатильности. Примечательно то, что интервалы прогнозов могут сужаться с ростом горизонта прогнозов, если прогноз делается в момент, соответствующий высокому уровню волатильности. Это объясняется тем, что в будущем следует ожидать снижения (ожидаемого) уровня волатильности.

На практике следует внести изменения в приведенные выше формулы, которые выведены в предположении, что истинные параметры процесса известны. Если параметры неизвестны, они заменяются соответствующими оценками. Можно также добавить к дисперсии прогноза поправку, связанную с тем, что при прогнозировании используются оценки a, а не истинные коэффициенты регрессии α, которая примерно равна

$$Z_{T+k}\mathbf{var}(a)^{-1}Z'_{T+k}.$$

Вместо неизвестной ковариационной матрицы оценок коэффициентов $\mathbf{var}(a)$ следует взять ее оценку, получаемую в методе максимального правдоподобия.

16.4. Разновидности моделей ARCH

Существует огромное количество модификаций классической модели GARCH. Дадим обзор только важнейших направлений, в которых возможна модификация модели. Все эти модели включают в себя какие-либо авторегрессионно условно гетероскедастичные процессы. Формально процесс $\{\varepsilon_t\}$ с нулевым условным математическим ожиданием $\mathbf{E}(\varepsilon_t|\Omega_t) = 0$ является **авторегрессионно условно гетероскедастичным**, если его условная относительно предыстории дисперсия

$$\sigma_t^2 = \mathbf{E}(\varepsilon_t^2|\Omega_t) = \mathbf{var}(\varepsilon_t|\Omega_t)$$

нетривиальным образом зависит от предыстории Ω_t.

16.4.1. Функциональная форма динамики условной дисперсии

Модели авторегрессионно условно гетероскедастичных процессов могут различаться тем, какой именно функцией задается зависимость условной дисперсии от своих лагов и лагов ε_t. Например, в логарифмической GARCH-модели условная дисперсия задается уравнением

$$\ln \sigma_t^2 = \omega + \sum_{j=1}^{p} \delta_j \ln \sigma_{t-j}^2 + \sum_{j=1}^{q} \gamma_j \ln \varepsilon_{t-j}^2.$$

В такой модели условная дисперсия всегда положительна вне зависимости от значений коэффициентов.

Следующая нелинейная GARCH-модель включает в себя как частный случай обычную GARCH-модель:

$$\sigma_t^\lambda = \omega + \sum_{j=1}^{p} \delta_j \sigma_{t-j}^\lambda + \sum_{j=1}^{q} \gamma_j |\varepsilon_{t-j}|^\lambda.$$

Кроме того, логарифмическая GARCH-модель является предельным частным случаем этой модели (после небольших изменений) при $\lambda \to 0$.

В приведенных моделях условная дисперсия не зависит от знаков лагов ε_t, а зависит только от их абсолютной величины. Это может быть серьезным ограничением, поскольку в реальных финансовых данных часто наблюдается **эффект левереджа**. Снижение рыночной стоимости акционерного капитала увеличивает отношение заемных средств к собственным и, следовательно, повышает рискованность вложений в фирму. Последнее проявляется в увеличении волатильности.

В результате, будущие значения волатильности отрицательно коррелируют с текущей доходностью. Это дало толчок к развитию разного рода асимметричных по ε_t моделей. Самой известной является экспоненциальная модель GARCH (EGARCH), предложенная Д. Нельсоном. Она имеет следующий вид:

$$\begin{aligned}
&\xi_t \sim NID(0,1),\\
&\varepsilon_t = \xi_t \sigma_t,\\
&\ln \sigma_t^2 = \omega + \sum_{j=1}^{p} \delta_j \ln \sigma_{t-j}^2 + \sum_{j=1}^{q} \alpha_j g(\xi_{t-j}),\\
&g(\xi_t) = \theta \xi_t + \gamma(|\xi_t| - \mathbf{E}|\xi_t|).
\end{aligned}$$

В модели EGARCH логарифм условной дисперсии представляет собой процесс ARMA. Первая сумма в уравнении соответствует авторегрессии, а вторая — скользящему среднему. Функция $g(\cdot)$ построена так, что $\mathbf{E}(g(\xi_t)) = 0$. Таким образом, в EGARCH σ_t^2 зависит и от величины, и от знака лагов ε_t и ξ_t. Логарифм условной дисперсии $\ln \sigma_t^2$ описывается процессом ARMA(p, q) с обычными для ARMA условиями стационарности.

Эффект левереджа можно также учесть в нелинейной GARCH-модели, введя дополнительный параметр сдвига κ:

$$\sigma_t^{\lambda} = \omega + \sum_{j=1}^{p} \delta_j \sigma_{t-j}^{\lambda} + \sum_{j=1}^{q} \gamma_j |\varepsilon_{t-j} - \kappa|^{\lambda}.$$

16.4.2. Отказ от нормальности

Как уже говорилось, финансовые ряды обычно характеризуются большой величиной куртозиса. Модель GARCH частично учитывает это, поскольку в ней безусловное распределение GARCH-процесса имеет толстые хвосты. Это является результатом стохастического характера условной дисперсии. Однако, как показывает опыт, этот эффект не полностью улавливается моделью GARCH. Это проявляется в том, что нормированные остатки модели, соответствующие $\xi_t = \varepsilon_t / \sigma_t$, все еще характеризуются большой величиной куртозиса. Таким образом, не выполняется одно из предположений модели GARCH — о том, что ε_t условно по предыстории имеет нормальное распределение.

Это создает трудности при использовании метода максимального правдоподобия для оценивания модели. Допустим, на самом деле ошибки распределены не нормально, но мы максимизируем функцию правдоподобия, основывающуюся на нормальности, т.е. используем так называемый метод квазимаксимального правдоподобия. Что при этом произойдет? Во-первых, при нарушении предположения

о нормальности оценки хотя и будут состоятельными, но не будут асимптотически эффективными (т.е. наиболее точными в пределе). Во-вторых, стандартные методы оценивания ковариационной матрицы оценок максимального правдоподобия не годятся, — требуется скорректированная оценка ковариационной матрицы.

Альтернативой методу квазимаксимального правдоподобия служат модели, в которых в явном виде делается предположение о том, что $\xi_t = \varepsilon_t/\sigma_t$ имеет распределение, отличающееся от нормального. Наиболее часто используется t-распределение Стьюдента, поскольку это распределение при малых степенях свободы имеет большой куртозис (см. Приложение А.3.2). При этом количество степеней свободы рассматривается как неизвестный параметр, причем непрерывный (формула плотности t-распределения подходит и в случае, когда берется нецелое количество степеней свободы). Можно использовать и другие распределения, например, так называемое обобщенное распределение ошибки (GED).

Часто распределение ξ_t является скошенным вправо. Для учета этой ситуации следует использовать асимметричные распределения с толстыми хвостами. Например, можно использовать нецентральное t-распределение, известное из статистики. Другой вариант, более простой в использовании, — так называемое скошенное t-распределение, которое «склеивается» из двух половинок t-распределений, поразному масштабированных.

16.4.3. GARCH-M

В модели GARCH-M непосредственно в уравнение регрессии добавляется условная дисперсия:

$$x_t = Z_t\alpha + \pi g(\sigma_t^2) + \varepsilon_t,$$

где $g(\cdot)$ — некоторая возрастающая функция. Эта новая компонента вводится для отражения влияния волатильности временного ряда на зависимую переменную, поскольку из многих финансовых моделей следует, что доходность актива должна быть положительно связана с рискованностью этого актива.

В качестве $g(\cdot)$ обычно используют $g(\sigma_t^2) = \sigma_t^2$, $g(\sigma_t^2) = \sqrt{\sigma_t^2} = \sigma_t$ или $g(\sigma_t^2) = \ln\sigma_t^2$.

16.4.4. Стохастическая волатильность

В рассмотренных моделях с авторегрессионной гетероскедастичностью условная дисперсия однозначно определяется предысторией. Это не оставляет места для случайных влияний на волатильность, помимо влияний лагов самого процесса. Однако авторегрессионная гетероскедастичность может возникнуть по-другому. При-

мером является модель авторегрессионной стохастической волатильности, в которой логарифм условной дисперсии описывается авторегрессионным процессом. Модель авторегрессионной стохастической волатильности первого порядка имеет следующий вид:

$$\begin{aligned}
&\xi_t \sim NID(0,1),\\
&\eta_t \sim NID(0,\sigma_\eta^2),\\
&\varepsilon_t = \xi_t\sigma_t,\\
&\ln\sigma_t^2 = \omega + \delta\ln\sigma_{t-1}^2 + \eta_t.
\end{aligned}$$

Эта модель по структуре проще, чем модель GARCH, и лучше обоснована теоретически, с точки зрения финансовых моделей, однако ее широкому использованию мешает сложность эффективного оценивания. Проблема состоит в том, что для нее, в отличие от моделей типа GARCH, невозможно в явном виде выписать функцию правдоподобия. Таким образом, в случае применения модели стохастической волатильности возникает дилемма: либо использовать алгоритмы, которые дают состоятельные, но неэффективные оценки, например, метод моментов, либо применять алгоритмы, требующие сложных расчетов, например, алгоритмы, использующие метод Монте-Карло для интегрирования многомерной плотности.

Несложно придумать модели, которые бы объединяли черты моделей типа GARCH и моделей стохастической волатильности. Однако подобные модели наследуют описанные выше проблемы оценивания.

16.4.5. ARCH-процессы с долгосрочной памятью

Для многих финансовых данных оценка $\sum\limits_{j=1}^{p}\delta_j + \sum\limits_{j=1}^{q}\gamma_j$, оказывается очень близкой к единице. Это дает эмпирическое обоснование для так называемой **интегрированной** модели GARCH, сокращенно IGARCH. Это обычные модели GARCH, в которых характеристическое уравнение для условной дисперсии имеет корень равный единице, и, следовательно, $\sum\limits_{j=1}^{p}\delta_j + \sum\limits_{j=1}^{q}\gamma_j = 1$. В частности, процесс IGARCH$(1,\,1)$ можно записать следующим образом:

$$\sigma_t^2 = \omega + (1-\gamma)\sigma_{t-1}^2 + \gamma\varepsilon_{t-1}^2.$$

IGARCH-процессы могут быть строго стационарны, однако не имеют ограниченной безусловной дисперсии и поэтому не являются слабо стационарными.

В модели IGARCH$(1,\,1)$ прогноз волатильности на τ шагов вперед (или, что то же самое, дисперсия прогноза самого процесса на τ шагов вперед) равен

$$\mathbf{E}(\sigma_{T+\tau}^2|\Omega_T) = \sigma_{d_T^2(\tau)} = \omega(k-1) + \sigma_{T+1}^2.$$

Следовательно, шок условной дисперсии инерционен в том смысле, что он влияет на будущие прогнозы всех горизонтов.

В последние годы получило распространение понятие так называемой *дробной интегрированности*. **Дробно-интегрированный процесс** (ARFIMA) с параметром интегрированности $d \in (0,1)$ занимает промежуточное положение между стационарными процессами ARMA ($d = 0$) и интегрированными ($d = 1$). Такие процессы имеют автокорреляционную функцию, которая затухает гиперболически, в то время как автокорреляционная функция стационарного процесса ARMA затухает экспоненциально, т.е. более быстро. В связи с этим принято говорить, что дробно-интегрированные процессы характеризуются долгосрочной памятью. Это явление было обнаружено как в уровнях, так и в дисперсиях многих финансовых рядов. В связи с этим появились модели дробно-интегрированных ARCH-процессов, такие как FIGARCH, HYGARCH.

16.4.6. Многомерные модели волатильности

Часто из экономической теории следует, что финансовые временные ряды должны быть взаимосвязаны, в том числе и через волатильность: краткосрочные и долгосрочные процентные ставки; валютные курсы двух валют, выраженные в одной и той же третьей валюте; курсы акций фирм, зависящих от одного и того же рынка, и т.п. Кроме того, условные взаимные ковариации таких финансовых показателей могут меняться со временем. Ковариация между финансовыми активами играет существенную роль в моделях поиска оптимального инвестиционного портфеля. С этой точки зрения многомерные модели авторегрессионной условной гетероскедастичности являются естественным расширением одномерных моделей.

Общее определение многомерного ARCH-процесса не представляет никакой теоретической сложности: рассматривается m-мерный наблюдаемый случайный вектор x_t, m-мерный вектор его условного математического ожидания, условная ковариационная матрица размерностью $m \times m$. В современной литературе предложено множество подобных моделей разной степени сложности. Оценивание многомерной ARCH-модели, однако, сопряжено со значительными трудностями. В частности, эти трудности связаны с необходимостью максимизации по большому количеству неизвестных параметров. Поэтому в прикладных исследованиях отдается предпочтение таким многомерным моделям волатильности, в которых количество параметров мало. В то же время для таких компактных моделей (например, для факторных моделей волатильности) может не существовать явной формулы для функции правдоподобия, что создает дополнительные трудности при оценивании.

16.5. Упражнения и задачи

Упражнение 1

Сгенерируйте ряд длиной 1000 наблюдений в соответствии с моделью ARCH(4) по уравнению: $\sigma_t^2 = 1 + 0.3\varepsilon_{t-1}^2 + 0.25\varepsilon_{t-2}^2 + 0.15\varepsilon_{t-2}^2 + 0.1\varepsilon_{t-2}^2$. (Вместо начальных значений квадратов ошибок возьмите безусловную дисперсию.)

В действительности мы имеем только ряд наблюдений, а вид и параметры модели неизвестны.

1.1. Оцените модель ARCH(4). Сравните оценки с истинными параметрами модели. Сравните динамику оценки условной дисперсии и ее истинных значений.

1.2. Проделайте то же самое для модели GARCH(1, 1).

1.3. Рассчитайте описательные статистики ряда: среднее, дисперсию, автокорреляцию первого порядка, асимметрию и эксцесс. Соответствуют ли полученные статистики теории?

Упражнение 2

В Таблице 16.1 дана цена p_t акций IBM за период с 25 июня 2002 г. по 9 апреля 2003 г.

2.1. Получите ряд логарифмических доходностей акций по формуле $y_t = 100 \ln(p_t/p_{t-1})$.

2.2. Постройте график ряда y_t, график автокорреляционной функции y_t, график автокорреляционной функции квадратов доходностей y_t^2. Сделайте выводы об автокоррелированности самих доходностей и их квадратов. Наблюдается ли авторегрессионная условная гетероскедастичность?

2.3. Оцените для доходностей модель GARCH(1, 1) на основе первых 180 наблюдений. Значимы ли параметры модели? Постройте график условной дисперсии и укажите периоды высокой и низкой волатильности.

2.4. Найдите эксцесс изучаемого ряда доходностей y_t и эксцесс нормированных остатков из модели GARCH. Сравните и сделайте выводы.

2.5. Постройте прогноз условной дисперсии для 19 оставшихся наблюдений. Постройте интервальный прогноз изучаемого ряда для тех же 19 наблюдений. Какая доля наблюдений попадает в прогнозные интервалы?

Таблица 16.1

25.06.02	68.19	6.08.02	67.49	17.09.02	71.48	28.10.02	76.27	9.12.02	79.44	22.01.03	79.55	5.03.03	77.73
26.06.02	69.63	7.08.02	68.91	18.09.02	69.29	29.10.02	76.45	10.12.02	80.64	23.01.03	80.89	6.03.03	77.07
27.06.02	71.47	8.08.02	71.34	19.09.02	64.56	30.10.02	78.37	11.12.02	81.28	24.01.03	78.84	7.03.03	77.90
28.06.02	71.57	9.08.02	71.56	20.09.02	63.68	31.10.02	78.64	12.12.02	80.01	27.01.03	78.27	10.03.03	75.70
1.07.02	67.20	12.08.02	71.50	23.09.02	63.13	1.11.02	80.10	13.12.02	79.84	28.01.03	79.95	11.03.03	75.35
2.07.02	68.17	13.08.02	71.63	24.09.02	59.52	4.11.02	82.19	16.12.02	81.46	29.01.03	80.16	12.03.03	75.18
3.07.02	70.09	14.08.02	74.64	25.09.02	62.77	5.11.02	81.37	17.12.02	80.15	30.01.03	78.15	13.03.03	78.45
5.07.02	73.06	15.08.02	76.21	26.09.02	61.79	6.11.02	81.38	18.12.02	78.98	31.01.03	78.05	14.03.03	79.00
8.07.02	70.87	16.08.02	79.05	27.09.02	60.13	7.11.02	78.80	19.12.02	78.51	3.02.03	78.03	17.03.03	82.46
9.07.02	69.25	19.08.02	82.18	30.09.02	58.09	8.11.02	77.44	20.12.02	79.64	4.02.03	76.94	18.03.03	82.47
10.07.02	68.35	20.08.02	80.96	1.10.02	60.94	11.11.02	77.14	23.12.02	80.10	5.02.03	77.11	19.03.03	82.00
11.07.02	69.00	21.08.02	80.69	2.10.02	59.40	12.11.02	79.00	24.12.02	79.61	6.02.03	77.51	20.03.03	82.20
12.07.02	68.80	22.08.02	81.68	3.10.02	59.77	13.11.02	79.20	26.12.02	78.35	7.02.03	77.10	21.03.03	84.90
15.07.02	70.58	23.08.02	80.10	4.10.02	56.39	14.11.02	80.56	27.12.02	77.21	10.02.03	77.91	24.03.03	82.25
16.07.02	68.60	26.08.02	79.12	7.10.02	56.65	15.11.02	79.85	30.12.02	76.10	11.02.03	77.39	25.03.03	83.45
17.07.02	70.27	27.08.02	77.67	8.10.02	56.83	18.11.02	79.03	31.12.02	77.35	12.02.03	76.50	26.03.03	81.55
18.07.02	71.62	28.08.02	75.77	9.10.02	54.86	19.11.02	78.22	2.01.03	80.41	13.02.03	75.86	27.03.03	81.45
19.07.02	71.57	29.08.02	76.33	10.10.02	57.36	20.11.02	81.45	3.01.03	81.49	14.02.03	77.45	28.03.03	80.85
22.07.02	68.09	30.08.02	75.10	11.10.02	63.68	21.11.02	84.74	6.01.03	83.43	18.02.03	79.33	31.03.03	78.43
23.07.02	66.65	3.09.02	72.08	14.10.02	63.18	22.11.02	84.27	7.01.03	85.83	19.02.03	79.51	1.04.03	78.73
24.07.02	69.12	4.09.02	73.45	15.10.02	68.22	25.11.02	86.03	8.01.03	84.03	20.02.03	79.15	2.04.03	81.46
25.07.02	68.94	5.09.02	71.91	16.10.02	64.65	26.11.02	84.89	9.01.03	86.83	21.02.03	79.95	3.04.03	81.91
26.07.02	66.00	6.09.02	72.92	17.10.02	71.93	27.11.02	87.53	10.01.03	87.51	24.02.03	78.56	4.04.03	80.79
29.07.02	70.75	9.09.02	74.22	18.10.02	73.97	29.11.02	86.75	13.01.03	87.34	25.02.03	79.07	7.04.03	80.47
30.07.02	71.36	10.09.02	75.31	21.10.02	75.26	2.12.02	87.13	14.01.03	88.41	26.02.03	77.40	8.04.03	80.07
31.07.02	69.98	11.09.02	73.92	22.10.02	74.21	3.12.02	85.04	15.01.03	87.42	27.02.03	77.28	9.04.03	78.71
1.08.02	67.84	12.09.02	71.60	23.10.02	74.32	4.12.02	83.53	16.01.03	85.88	28.02.03	77.95		
2.08.02	67.47	13.09.02	72.23	24.10.02	71.83	5.12.02	82.90	17.01.03	81.14	3.03.03	77.33		
5.08.02	65.60	16.09.02	72.05	25.10.02	74.28	6.12.02	82.16	21.01.03	80.38	4.03.03	76.70		

Задачи

1. Объясните значение термина «волатильность».
2. Рассмотрите следующие два утверждения:
   а) «GARCH является слабо стационарным процессом»;
   б) «GARCH является процессом с изменяющейся во времени дисперсией».

   Поясните смысл каждого из них. Объясните, почему между ними нет противоречия.
3. Почему процесс GARCH представляет особый интерес для финансовой эконометрии?
4. В каком отношении находятся между собой модели ARCH и GARCH?
5. Коэффициент процесса ARCH(1) равен $\gamma = 0.8$, безусловная дисперсия равна 2. Запишите уравнение процесса.
6. Приведите конкретный пример процесса ARCH(1), который имел бы конечную безусловную дисперсию. Ответ обоснуйте.
7. Приведите конкретный пример процесса ARCH(1), куртозис которого не определен («бесконечен»). Ответ обоснуйте.
8. Вычислите безусловную дисперсию следующего процесса GARCH:

$$\sigma_t^2 = 0.4 + 0.1\sigma_{t-1}^2 + 0.7\varepsilon_t^2.$$

   Докажите формально, что ваш ответ верен.
9. Докажите, что следующий процесс GARCH не может иметь конечную безусловную дисперсию:

$$\sigma_t^2 = 1 + 0.6\sigma_{t-1}^2 + 0.6\varepsilon_{t-1}^2.$$

   (Подсказка: можно использовать доказательство от противного.)
10. Процесс GARCH(1, 1) задан уравнением $\sigma_t^2 = 1 + 0.8\sigma_{t-1}^2 + 0.1\varepsilon_{t-1}^2$. Известно, что $\sigma_T^2 = 9$, $\varepsilon_T = -2$. Чему равен прогноз на следующий период? Чему равна дисперсия этого прогноза?
11. Покажите, что модель GARCH(1, 1) можно записать в виде модели ARCH бесконечного порядка.

12. Утверждается, что «модель GARCH более компактна, чем модель ARCH». Что при этом имеется в виду? Почему важна «компактность» модели?

13. Докажите, что процесс GARCH(p, q) не автокоррелирован.

14. Докажите, что процесс GARCH(p, q) является белым шумом, если существует его безусловная дисперсия.

15. Пользуясь представлением квадратов процесса GARCH(1, 1) в виде ARMA(1, 1) выведите их автокорреляционную функцию.

16. Запишите автокорреляционную функцию квадратов процесса GARCH(1, 1). Покажите, что при значении суммы коэффициентов $\delta_1 + \gamma_1$, приближающемся к 1 (но меньшем 1) автокорреляционная функция затухает медленно. Покажите, что при фиксированном значении суммы коэффициентов $\delta_1 + \gamma_1 = \varphi \quad (0 < \varphi < 1)$, автокорреляция тем слабее, чем меньше γ_1, и стремится к нулю при $\gamma_1 \to 0$.

17. Рассмотрите следующую модель с авторегрессионной условной гетероскедастичностью:

$$\sigma_t^2 = \omega + \delta\sigma_{t-1}^2 + \gamma[f(\varepsilon_{t-1})]^2,$$

где $f(z) = |z| - \varphi z$.

    а) Укажите значения параметров, при которых эта модель сводится к обычной модели GARCH.

    б) Утверждается, что в этой модели имеет место асимметричность влияния «шоков» ε_t на условную дисперсию. Что при этом имеется в виду? При каких значениях параметров влияние будет симметричным?

18. Запишите модель GARCH(1, 1)-M с квадратным корнем условной дисперсии в уравнении регрессии (с расшифровкой обозначений).

19. Рассмотрите модель AR(1) с независимыми одинаково распределенными ошибками и модель AR(1) с ошибками, подчиняющимися процессу GARCH.

    а) Объясните, почему точечные прогнозы по этим двум моделям не будут отличаться.

    б) Как будут отличаться интервальные прогнозы?

20. Пусть имеется некоторый процесс с авторегрессионной условной гетероскедастичностью ε_t, задаваемый моделью

$$\sigma_t^2 = \mathbf{var}(\varepsilon_t|\Omega_{t-1}) = h(\sigma_{t-1}^2, \ldots, \sigma_{t-p}^2, \varepsilon_{t-1}^2, \ldots, \varepsilon_{t-q}^2),$$

где Ω_{t-1} — предыстория процесса, $h(\cdot)$ — некоторая функция и $\varepsilon_t = \xi_t \sigma_t$. Предполагается, что инновации ξ_t имеют стандартное нормальное распределение и не зависят от предыстории Ω_{t-1}. Найдите куртозис ε_t, если известно, что $\mathbf{E}(\sigma_t^2) = 5$, $\mathbf{E}(\sigma_t^4) = 100$. О чем говорит величина куртозиса?

21. Докажите, что эксцесс распределения отдельного наблюдения ε_t процесса ARCH(1)

$$\varepsilon_t = \xi_t \sigma_t, \quad \sigma_t^2 = \omega + \gamma \varepsilon_{t-1}^2, \quad \xi_t \sim NID(0,1)$$

равен $\dfrac{6\gamma^2}{1-3\gamma^2}$.

22. Какие сложности возникают при построении прогнозных интервалов процесса GARCH с заданным уровнем доверия (например, 95%)? Каким способом можно обойти эту проблему?

23. Почему модель GARCH не подходит для прогнозирования автокоррелированных временных рядов?

Рекомендуемая литература

1. **Магнус Я.Р., Катышев П.К., Пересецкий А.А.** Эконометрика — начальный курс. — М.: «Дело», 2000. (Гл. 12).

2. **Предтеченский А.Г.** Построение моделей авторегрессионной условной гетероскедастичности (ARCH) некоторых индикаторов российского финансового рынка (дипломная работа), ЭФ НГУ, 2000.

   (`http://www.nsu.ru/ef/tsy/ecmr/garch/predtech/index.htm`).

3. **Шепард Н.** Статистические аспекты моделей типа ARCH и стохастическая волатильность. Обозрение прикладной и промышленной математики. Т. 3, вып. 6. — 1996.

4. **Baillie Richard T. and Tim Bollerslev.** Prediction in Dynamic Models with Time Dependent Conditional Variances // Journal of Econometrics, No. 52, 1992.

5. **Bera A.K. and Higgins M.L.** ARCH Models: Properties, Estimation and Testing // Journal of Economic Surveys, No. 7, 1993.

6. **Bollerslev T., Engle R.F. and Nelson D.B.** ARCH Models // Handbook of Econometrics. Vol. IV. Ch. 49. — Elsevier Science, 1994.

7. **Bollerslev Tim**. Generalized Autoregressive Conditional Heteroskedasticity // Journal of Econometrics, No. 31, 1993.

8. **Bollerslev Tim, Ray Y. Chou and Kenneth F. Kroner.** ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence // Journal of Econometrics, No. 52, 1992.

9. **Campbell John Y., Lo Andrew W., MacKinlay A. Craig.** The Econometrics of Financial Markets. — Princeton University Press, 1997. (Ch. 12).

10. **Diebold, Francis X. and Jose A. Lopez** Modeling Volatility Dynamics, Macroeconometrics: Developments, Tensions and Prospects. — Kluwer Academic Press, 1995.

11. **Engle, Robert F.** Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation // Econometrica, No. 50, 1982.

12. **Greene W.H.** Econometric Analysis. — Prentice-Hall, 2000. (Ch. 18).

13. **Hamilton James D.** Time Series Analysis. Ch. 21. — Princeton University Press, 1994.

14. **Mills Terence C.** The Econometric Financial Modelling Time Series. — Cambridge University Press, 1999. (Ch. 4).

Глава 17

Интегрированные процессы, ложная регрессия и коинтеграция

17.1. Стационарность и интегрированные процессы

Для иллюстрации различия между стационарными и нестационарными случайными процессами рассмотрим марковский процесс, т.е. авторегрессию первого порядка:

$$x_t = \mu + \varphi x_{t-1} + \varepsilon_t,$$

или

$$(1 - \varphi \mathbf{L}) x_t = \mu + \varepsilon_t.$$

В данной модели x_t — не центрированы.

Будем предполагать, что ошибки ε_t — независимые одинаково распределенные случайные величины с нулевым математическим ожиданием и дисперсией σ_ε^2. Как известно, при $|\varphi| < 1$ процесс авторегрессии первого порядка слабо стационарен и его можно представить в виде бесконечного скользящего среднего:

$$x_t = \frac{\mu + \varepsilon_t}{1 - \varphi \mathbf{L}} = \frac{\mu}{1 - \varphi} + \sum_{i=0}^{\infty} \varphi^i \varepsilon_{t-i}.$$

Условие $|\varphi| < 1$ гарантирует, что коэффициенты ряда затухают. Математическое ожидание переменной x_t постоянно: $\mathbf{E}(x_t) = \dfrac{\mu}{1-\varphi}$. Дисперсия равна

$$\mathbf{var}(x_t) = \sum_{i=0}^{\infty} \varphi^{2i} \mathbf{var}(\varepsilon_{t-i}) = \frac{\sigma_\varepsilon^2}{1-\varphi^2}.$$

Найдем также автоковариации процесса:

$$\gamma_k = \mathbf{cov}(x_t, x_{t-k}) = \sum_{i=0}^{\infty} \varphi^{i+k}\varphi^i \sigma_\varepsilon^2 = \varphi^k \sum_{i=0}^{\infty} \varphi^{2i}\sigma_\varepsilon^2 = \frac{\varphi^k}{1-\varphi^2}\sigma_\varepsilon^2.$$

Таким образом, рассматриваемый процесс слабо стационарен, поскольку слабое определение стационарности требует, чтобы математическое ожидание x_t было постоянным, а ковариации не зависели от времени, но только от лага. На самом деле, поскольку ошибки ε_t одинаково распределены, то он стационарен и в строгом смысле.

При $|\varphi| > 1$ это будет взрывной процесс. Такие процессы рассматриваться не будут.

Как известно (см. гл. 14), авторегрессионный процесс первого порядка при $\varphi = 1$ называют случайным блужданием. Если $\mu = 0$, то это просто случайное блуждание, а при $\mu \neq 0$ это случайное блуждание с дрейфом.

У процесса случайного блуждания, начавшегося бесконечно давно, не существует безусловного математического ожидания и дисперсии. За бесконечное время процесс «уходит в бесконечность», его дисперсия становится бесконечной. В связи с этим будем рассматривать все моменты процесса случайного блуждания как условные, т.е. будем действовать так, как если бы x_0 была детерминированной величиной. Выразим x_t через x_0:

$$x_t = x_0 + \mu t + \sum_{i=1}^{t} \varepsilon_i.$$

Таким образом, константа (дрейф) в авторегрессионной записи процесса приводит к появлению линейного тренда в x_t. Мы получили разложение процесса x_t на две составляющие: детерминированный линейный тренд μt и случайное блуждание $\varepsilon_t^* = x_0 + \sum_{i=1}^{t} \varepsilon_i$, такое что ошибка ε_t представляет собой его приросты: $\varepsilon_t = \Delta\varepsilon_t^*$. Вторую составляющую, как мы помним, называют стохастическим трендом, поскольку влияние каждой ошибки не исчезает со временем.

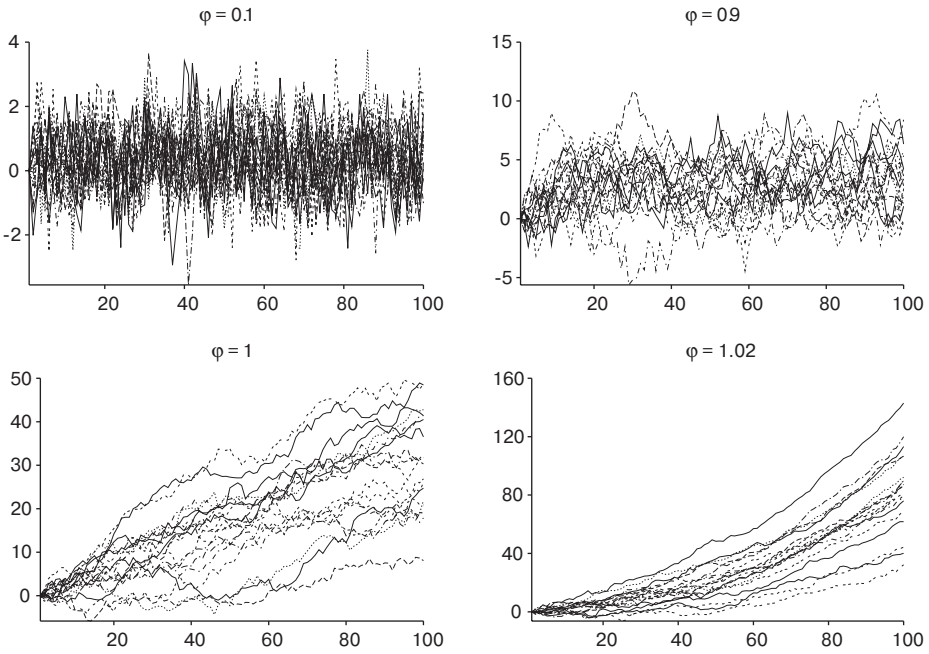

Рис. 17.1. Поведение процесса AR(1) в зависимости от значения φ

Используя данное представление, найдем математическое ожидание и дисперсию:

$$\mathbf{E}(x_t|x_0) = x_0 + \mu t.$$

$$\mathbf{var}(x_t|x_0) = \mathbf{var}\Big(\sum_{i=1}^{t} \varepsilon_i\Big) = \sum_{i=1}^{t} \mathbf{var}\,(\varepsilon_i) = t\sigma_\varepsilon^2. \qquad (17.1)$$

Дисперсия со временем растет линейно до бесконечности.

Случайное блуждание является примером авторегрессионого процесса с **единичным корнем**. Это название следует из того, что при $\varphi = 1$ корень характеристического многочлена $1 - \varphi\mathbf{L}$, соответствующего процессу AR(1), равен единице.

Рисунок 17.1 иллюстрирует поведение марковских процессов при различных коэффициентах авторегрессии. На каждом из графиков изображены 20 рядов длиной $T = 100$, случайно сгенерированных по формуле $x_t = 0.3 + \varphi x_{t-1} + \varepsilon_t$ с разными значениями φ: 1) $\varphi = 0.1$; 2) $\varphi = 0.9$; 3) $\varphi = 1$; 4) $\varphi = 1.02$. Во всех случаях использовалось стандартное нормальное распределение для ε_t и $x_0 = 0$.

Добавим к стационарному процессу AR(1) детерминированный тренд $\mu_1 t$:

$$x_t = \mu_0 + \mu_1 t + \varphi x_{t-1} + \varepsilon_t.$$

Тогда

$$x_t = \frac{\mu_0 + \mu_1 t + \varepsilon_t}{1 - \varphi \mathbf{L}} = \frac{\mu_0}{1-\varphi} + \mu_1 \sum_{i=0}^{\infty} \varphi^i (t-i) + \sum_{i=0}^{\infty} \varphi^i \varepsilon_{t-i} =$$
$$= \frac{\mu_0}{1-\varphi} - \mu_1 \sum_{i=0}^{\infty} i\varphi^i + \frac{\mu_1}{1-\varphi} t + \sum_{i=0}^{\infty} \varphi^i \varepsilon_{t-i}.$$

Ряд $\sum\limits_{i=0}^{\infty} i\varphi^i$ сходится, поскольку i возрастает линейно, а φ^i убывает экспоненциально при $|\varphi| < 1$, т.е. значительно быстрее. Его сумма равна $\dfrac{\varphi}{(1-\varphi)^2}$. Используя это, получаем

$$x_t = \frac{\mu_0}{1-\varphi} - \frac{\varphi\mu_1}{(1-\varphi)^2} + \frac{\mu_1}{1-\varphi} t + \sum_{i=0}^{\infty} \varphi^i \varepsilon_{t-i} = \gamma_0 + \gamma_1 t + \sum_{i=0}^{\infty} \varphi^i \varepsilon_{t-i}, \qquad (17.2)$$

где

$$\gamma_0 = \frac{\mu_0}{1-\varphi} - \frac{\varphi\mu_1}{(1-\varphi)^2} \quad \text{и} \quad \gamma_1 = \frac{\mu_1}{1-\varphi}.$$

Можно также записать уравнение процесса в виде:

$$(x_t - \gamma_0 - \gamma_1 t) = \varphi(x_{t-1} - \gamma_0 - \gamma_1 (t-1)) + \varepsilon_t.$$

Ясно, что если вычесть из x_t тренд $\gamma_1 t$, то получится стационарный процесс. Подобного рода процессы называют **стационарными относительно тренда**.

Рассмотрим теперь процесс ARMA(p, q):

$$x_t = \sum_{i=1}^{p} \varphi_i x_{t-i} + \varepsilon_t - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}.$$

Если все корни характеристического многочлена

$$\varphi(z) = 1 - \sum_{i=1}^{p} \varphi_i z^i$$

по абсолютной величине больше 1, т.е. лежат за пределами единичного круга на комплексной плоскости, то процесс стационарен. Если один из корней лежит в пределах единичного круга, то процесс взрывной. Если же d корней равны единице, а остальные лежат за пределами единичной окружности, то процесс нестационарный, но не взрывной и о нем говорят, что он имеет d единичных корней.

Нестационарный процесс, первые разности которого стационарны, называют **интегрированным первого порядка** и обозначают I(1). Стационарный процесс обозначают I(0). Если d-е разности случайного процесса стационарны, то его называют **интегрированным** d-го порядка и обозначают I(d).

Рассмотрим, например, процесс

$$y_t = \sum_{i=1}^{t} x_i, \quad \text{где } x_t = x_{t-1} + \varepsilon_t.$$

Он будет I(2), то есть его вторые разности, $\Delta^2 y_t$, стационарны.

Для процессов ARIMA можно дать более удачное определение интегрированности. Процессом I(0) называется стационарный процесс с обратимым скользящим средним. Процесс I(d) — такой процесс, d-е разности которого являются I(0). Соответственно, процесс, являющийся d-ой разностью процесса I(0), будет I($-d$). Такое уточнение нужно для того, чтобы необратимые процессы, такие как $\varepsilon_t - \varepsilon_{t-1}$, где ε_t — белый шум, по определению были I(-1), но не I(0). По этому уточненному определению процесс I(d) при $d > 0$ будет иметь в точности d единичных корней.

17.2. Разложение Бевериджа—Нельсона для процесса I(1)

Рассмотрим ARIMA-процесс I(1), интегрированный первого порядка. Пусть его исходная форма, записанная через лаговый оператор, имеет вид

$$\boldsymbol{\varphi}(\mathbf{L})x_t = \mu + \boldsymbol{\theta}(\mathbf{L})\varepsilon_t.$$

Поскольку это процесс I(1), то многочлен $\boldsymbol{\varphi}(\mathbf{L})$ имеет единичный корень и уравнение процесса можно представить в виде

$$(1 - \mathbf{L})\boldsymbol{\varphi}^*(\mathbf{L})x_t = \boldsymbol{\varphi}^*(\mathbf{L})\Delta x_t = \mu + \boldsymbol{\theta}(\mathbf{L})\varepsilon_t,$$

где у многочлена $\boldsymbol{\varphi}^*(\mathbf{L})$ все корни находятся за пределами единичного круга. Отсюда следует разложение Вольда для приростов Δx_t, которые являются стационарными:

$$\Delta x_t = \frac{\mu + \boldsymbol{\theta}(\mathbf{L})}{\boldsymbol{\varphi}^*(\mathbf{L})}\varepsilon_t = \frac{\mu}{\boldsymbol{\varphi}^*(\mathbf{L})} + \sum_{i=0}^{\infty} c_i \varepsilon_{t-i} = \frac{\mu}{1 - \sum\limits_{j=1}^{p} \varphi_j^*} + c(\mathbf{L})\varepsilon_t = \gamma + c(\mathbf{L})\varepsilon_t.$$

Ряд $c(z)$ можно представить следующим образом:

$$c(z) = c(1) + c^*(z)(1 - z),$$

где

$$c^*(z) = \sum_{i=0}^{\infty} c_i^* z^i, \quad \text{с коэффициентами} \quad c_i^* = -\sum_{j=i+1}^{\infty} c_j.$$

Действительно,

$$c(1) + c^*(z)(1-z) = \sum_{i=0}^{\infty} c_i + \sum_{i=0}^{\infty} c_i^* z^i - \sum_{i=0}^{\infty} c_i^* z^{i+1} =$$
$$= \sum_{i=0}^{\infty} c_i + c_0^* + \sum_{i=1}^{\infty} (c_i^* - c_{i-1}^*) z^i = c_0 + \sum_{i=1}^{\infty} c_i z^i = \sum_{i=0}^{\infty} c_i z^i.$$

Таким образом, можно представить Δx_t в виде

$$\Delta x_t = \gamma + (c(1) + c^*(\mathbf{L})(1-\mathbf{L}))\,\varepsilon_t = \gamma + c(1)\varepsilon_t + c^*(\mathbf{L})\Delta\varepsilon_t.$$

Суммируя Δx_t, получим

$$x_t = \gamma t + c(1)\varepsilon_t^* + c^*(\mathbf{L})\varepsilon_t,$$

где ε_t^* — случайное блуждание, такое что $\Delta\varepsilon_t^* = \varepsilon_t$. Без доказательства отметим, что ряд $c^*(\mathbf{L})$ сходится абсолютно[1]: $\sum\limits_{i=0}^{\infty} |c_i^*| < \infty$. Следовательно, он соответствует разложению Вольда стационарного процесса.

Мы получили так называемое разложение Бевериджа—Нельсона. Процесс x_t вида I(1) мы представили как комбинацию детерминированного тренда γt, стохастического тренда $c(1)\varepsilon_t^*$ и стационарного процесса $c^*(\mathbf{L})\varepsilon_t$, который здесь обычно интерпретируется как циклическая компонента.

17.3. Ложная регрессия

Одним из важнейших условий получения корректных оценок в регрессионных моделях является требование стационарности переменных. В экономике довольно часто встречаются стационарные ряды, например, уровень безработицы. Однако, как правило, экономические процессы описываются нестационарными рядами: объем производства, уровень цен и т.д.

[1] Это можно понять из того, что $\sum\limits_{i=0}^{\infty} |c_i^*| = \sum\limits_{i=0}^{\infty} \left| \sum\limits_{j=i+1}^{\infty} c_i \right| \leqslant \sum\limits_{i=0}^{\infty} \sum\limits_{j=i+1}^{\infty} |c_i| = \sum\limits_{i=0}^{\infty} i|c_i|$. Поскольку коэффициенты c_i у стационарного процесса ARMA сходятся экспоненциально, то ряд должен сойтись (экспоненциальное убывание превосходит рост i).

Очень важным условием корректного оценивания регрессионных моделей является условие стационарности регрессоров. Если зависимая переменная является I(1), и, кроме того, модель неверно специфицирована, т.е. некоторые из факторов, введенные ошибочно, являются I(1), то полученные оценки будут очень «плохими». Они не будут обладать свойством состоятельности, т.е. не будут сходиться к истинным значениям параметров по мере увеличения размеров выборки. Привычные показатели, такие как коэффициент детерминации R^2, t-статистики, F-статистики, будут указывать на наличие связи там, где на самом деле ее нет. Такой эффект называют **эффектом ложной регрессии**.

Показать эффект ложной регрессии для переменных I(1) можно с помощью метода Монте-Карло. Сгенерируем достаточно большое число пар независимых процессов случайного блуждания с нормально распределенными ошибками:

$$x_t = x_{t-1} + \varepsilon_t \quad \text{и} \quad z_t = z_{t-1} + \xi_t,$$

где $\varepsilon_t \sim N(0,1)$ и $\xi_t \sim N(0,1)$. Оценив для каждой пары рядов x_t и z_t достаточно много раз регрессию вида

$$x_t = az_t + b + u_t,$$

получим экспериментальные распределения стандартных статистик.

Проведенные экспериментальные расчеты для рядов длиной 50 наблюдений показывают, что t-статистика для a при номинальной вероятности 0.05 (т.е. 5%) в действительности отвергает верную гипотезу об отсутствии связи примерно в 75% случаев. Для того чтобы нулевая гипотеза об отсутствии связи отклонялась с вероятностью 5%, вместо обычного 5%-го квантиля распределения Стьюдента, равного примерно 2, нужно использовать критическую границу $t_{0.05} = 11.2$.

Из экспериментов также следует, что регрессии с независимыми процессами случайного блуждания с большой вероятностью имеют высокий коэффициент детерминации R^2 из-за нестационарности. Более чем в половине случаев коэффициент детерминации превышает 20%, и несколько менее чем в 5% случаев превышает 70%. Для сравнения можно построить аналогичные регрессии для двух независимых нормально распределенных процессов типа белый шум. Оказывается, что в таких регрессиях R^2 чрезвычайно редко превышает 20% (вероятность этого порядка 0.1%)[2].

То же самое, хотя и в меньшей степени, можно наблюдать и в случае двух стационарных AR(1)-процессов с коэффициентом автокорреляции φ, близким к единице. Отличие заключается в том, что здесь ложная связь асимптотически (при стремлении длины рядов к бесконечности) исчезает, а в случае I(1)-процессов — нет.

[2] Для двух независимых I(2)-процессов, построенных как проинтегрированные процессы случайного блуждания, примерно в половине случаев коэффициент детерминации превышает 80%!

Все же проблема остается серьезной, поскольку на практике экономист имеет дело с конечными и часто довольно короткими рядами.

Таким образом, наличие в двух независимых процессах стохастических трендов может с высокой вероятностью привести к получению ложного вывода об их взаимосвязанности, если пользоваться стандартными методами.

Стандартные методы проверки гипотез, применяемые в регрессионном анализе, в данном случае не работают. Это происходит по причине нарушения некоторых предположений, лежащих в основе модели регрессии. Какие же предположения нарушаются? Приведем одну из возможных точек зрения.

Предположим, как и выше, что x_t и z_t — два независимых процесса случайного блуждания, и оценивается регрессия

$$x_t = az_t + b + u_t.$$

Поскольку в этой регрессии истинное значение параметра a равно нулю, то $u_t = x_t - b$, т.е. ошибка в регрессии является процессом случайного блуждания. Выше получено выражение (17.1) для дисперсии процесса случайного блуждания (условной по начальному наблюдению):

$$\mathbf{var}(u_t) = t\sigma_\varepsilon^2,$$

где σ_ε^2 — дисперсия ε_t (приростов x_t). Таким образом, здесь наблюдается сильнейшая гетероскедастичность. С ростом номера наблюдения дисперсия ошибки растет до бесконечности. Вследствие этого t-статистика регрессии имеет нестандартное распределение, и обычные таблицы t-распределения использовать нельзя.

Отметим, что наличие в переменных регрессии обычного детерминированного тренда также может приводить к появлению ложной регрессии. Пусть, например, x_t и z_t заданы формулами

$$x_t = \mu_0 + \mu_1 t + \varepsilon_t \quad \text{и} \quad z_t = \nu_0 + \nu_1 t + \xi_t,$$

где ε_t и ξ_t — два независимых процесса типа белый шум. Регрессия x_t по константе и z_t может иметь высокий коэффициент детерминации, и этот эффект только усиливается с ростом размера выборки. С «детерминированным» вариантом ложной регрессии достаточно легко бороться. В рассматриваемом случае достаточно добавить в уравнение в качестве регрессора тренд, т.е. оценить регрессию $x_t = az_t + b + ct + u_t$, и эффект ложной регрессии исчезает.

17.4. Проверка на наличие единичных корней

С осознанием опасности применения ОМНК к нестационарным рядам появилась необходимость в критериях, которые позволили бы отличить стационарный процесс от нестационарного.

К неформальным методам проверки стационарности можно отнести визуальный анализ графиков спектральной плотности и автокорреляционной функции.

В настоящее время самым популярным из формальных критериев является критерий, разработанный **Дики** и **Фуллером** (**DF**-тест).

Предположим, нужно выяснить, какой из двух процессов лучше подходит для описания временного ряда:

$$x_t = \mu_0 + \mu_1 t + \varepsilon_t \;\text{ или }\; x_t = \mu_0 + x_{t-1} + \varepsilon_t,$$

где ε_t — стационарный ARMA-процесс. Первый из процессов является стационарным относительно тренда, а второй содержит единичный корень и дрейф. Каждый из вариантов может рассматриваться как правдоподобная модель экономического процесса.

Внешне два указанных процесса сильно различаются, однако можно показать, что оба являются частными случаями одного и того же процесса:

$$x_t = \gamma_0 + \gamma_1 t + v_t,$$
$$v_t = \varphi v_{t-1} + \varepsilon_t,$$

что можно переписать также в виде

$$(x_t - \gamma_0 - \gamma_1 t) = \varphi(x_{t-1} - \gamma_0 - \gamma_1(t-1)) + \varepsilon_t. \qquad (17.3)$$

Как было показано ранее (17.2) для марковского процесса, при $|\varphi| < 1$ данный процесс эквивалентен процессу $x_t = \mu_0 + \mu_1 t + \varepsilon_t$, где коэффициенты связаны соотношениями:

$$\gamma_0 = \frac{\mu_0}{1-\varphi} - \frac{\varphi\mu_1}{(1-\varphi)^2} \quad \text{и} \quad \gamma_1 = \frac{\mu_1}{1-\varphi}.$$

При $\varphi = 1$ получаем

$$x_t - \gamma_0 - \gamma_1 t = x_{t-1} - \gamma_0 - \gamma_1 t + \gamma_1 + \varepsilon_t,$$

т.е.

$$x_t = \gamma_1 + x_{t-1} + \varepsilon_t.$$

Таким образом, как и утверждалось, обе модели являются частными случаями одной и той же модели (17.3) и соответствуют случаям $|\varphi| < 1$ и $\varphi = 1$.

Модель (17.3) можно записать следующим образом:

$$x_t = \gamma_0 + \gamma_1 t + \varphi(x_{t-1} - \gamma_0 - \gamma_1(t-1)) + \varepsilon_t.$$

Это модель регрессии, нелинейная по параметрам. Заменой переменных мы можем свести ее к линейной модели:

$$x_t = \mu_0 + \mu_1 t + \varphi x_{t-1} + \varepsilon_t,$$

где $\mu_0 = (1-\varphi)\gamma_0 + \varphi\gamma_1$, $\mu_1 = (1-\varphi)\gamma_1$.

Эта новая модель фактически эквивалентна (17.3), и метод наименьших квадратов даст ту же самую оценку параметра φ. Следует, однако, иметь в виду, что линейная модель скрывает тот факт, что при $\varphi = 1$ будет выполнено $\mu_1 = 0$.

Базовая модель, которую использовали Дики и Фуллер, — авторегрессионный процесс первого порядка:

$$x_t = \varphi x_{t-1} + \varepsilon_t. \tag{17.4}$$

При $\varphi = 1$ это случайное блуждание. Конечно, вряд ли экономическая переменная может быть описана процессом (17.4). Более реалистично было бы предположить наличие в этом процессе константы и тренда (линейного или квадратичного):

$$x_t = \mu_0 + \varphi x_{t-1} + \varepsilon_t, \tag{17.5}$$

$$x_t = \mu_0 + \mu_1 t + \varphi x_{t-1} + \varepsilon_t, \tag{17.6}$$

$$x_t = \mu_0 + \mu_1 t + \mu_2 t^2 + \varphi x_{t-1} + \varepsilon_t. \tag{17.7}$$

Нулевая гипотеза в критерии Дики—Фуллера состоит в том, что ряд нестационарен и имеет один единичный корень $\varphi = 1$, при этом $\mu_i = 0$, альтернативная — в том, что ряд стационарен $|\varphi| < 1$:

$$H_0 : \varphi = 1, \quad \mu_i = 0,$$
$$H_A : |\varphi| < 1.$$

Здесь $i = 0$, если оценивается (17.5), $i = 1$, если оценивается (17.6), и $i = 2$, если оценивается (17.7).

Предполагается, что ошибки ε_t некоррелированы. Это предположение очень важно, без него критерий работать не будет.

Для получения статистики, с помощью которой можно было бы проверить нулевую гипотезу, Дики и Фуллер предложили оценить данную авторегрессионную модель и взять из нее **обычную t-статистику для гипотезы о том, что $\varphi = 1$**. Эту статистику называют статистикой Дики—Фуллера и обозначают DF. При этом критерий является **односторонним**, поскольку альтернатива $\varphi > 1$, соответствующая взрывному процессу, не рассматривается.

DF заключается в том, что с помощью одной t-статистики как бы проверяется гипотеза сразу о двух коэффициентах. На самом деле, фактически подразумевается модель вида (17.3), в которой проверяется гипотеза об одном коэффициенте φ.

Если в регрессии (17.6) нулевая гипотеза отвергается, то принимается альтернативная гипотеза — о том, что процесс описывается уравнением (17.6) с $\varphi < 1$, то есть это стационарный относительно линейного тренда процесс. В противном случае имеем нестационарный процесс, описываемый уравнением (17.5), где $\varphi = 1$, то есть случайное блуждание с дрейфом, но без временного тренда в уравнении авторегрессии.

Часто встречается несколько иная интерпретация этой особенности данного критерия: проверяется гипотеза $H_0 : \varphi = 1$ против гипотезы $H_A : \varphi < 1$, и **оцениваемая регрессия не совпадает с порождающим данные процессом**, каким он предполагается согласно альтернативной гипотезе, а именно, в оцениваемой регрессии имеется «лишний» детерминированный регрессор. Так, чтобы проверить нулевую гипотезу для процесса вида (17.5), нужно построить регрессию (17.6) или (17.7). Аналогично для проверки нулевой гипотезы о процессе вида (17.6) нужно оценить регрессию (17.7). Однако приведенная ранее интерпретация более удачная.

Поскольку статистика Дики—Фуллера имеет нестандартное распределение, для ее использования требуются специальные таблицы. Эти таблицы были составлены эмпирически методом Монте-Карло. Все эти статистики получены на основе одного и того же процесса вида (17.4) с $\varphi = 1$, но с асимптотической точки зрения годятся и для других процессов, несмотря на наличие мешающих параметров, которые приходится оценивать.

Чтобы удобно было использовать стандартные регрессионные пакеты, уравнения регрессии преобразуются так, чтобы зависимой переменной была первая разность. Например, в случае (17.4) имеем уравнение

$$\Delta x_t = \delta x_{t-1} + \varepsilon_t,$$

где $\delta = \varphi - 1$. Тогда нулевая гипотеза примет вид $\delta = 0$.

Предположение о том, что переменная следует авторегрессионному процессу первого порядка и ошибки некоррелированы, является, конечно, слишком ограничительным. Критерий Дики—Фуллера был модифицирован для авторегрессионных процессов более высоких порядков и получил название **дополненного критерия Дики—Фуллера** (augmented Dickey—Fuller test, **ADF**).

Базовые уравнения для него приобретают следующий вид:

$$\Delta x_t = (\varphi - 1)x_{t-1} + \sum_{j=1}^{k} \gamma_j \Delta x_{t-j} + \varepsilon_t, \qquad (17.8)$$

$$\Delta x_t = \mu_0 + (\varphi - 1)x_{t-1} + \sum_{j=1}^{k} \gamma_j \Delta x_{t-j} + \varepsilon_t, \tag{17.9}$$

$$\Delta x_t = \mu_0 + \mu_1 t + (\varphi - 1)x_{t-1} + \sum_{j=1}^{k} \gamma_j \Delta x_{t-j} + \varepsilon_t, \tag{17.10}$$

$$\Delta x_t = \mu_0 + \mu_1 t + \mu_2 t^2 + (\varphi - 1)x_{t-1} + \sum_{j=1}^{k} \gamma_j \Delta x_{t-j} + \varepsilon_t. \tag{17.11}$$

Распределения этих критериев асимптотически совпадают с соответствующими обычными распределениями Дики—Фуллера и используют те же таблицы. Грубо говоря, роль дополнительной авторегрессионной компоненты сводится к тому, чтобы убрать автокорреляцию из остатков. Процедура проверки гипотез не отличается от описанной выше.

Как показали эксперименты Монте-Карло, критерий Дики—Фуллера чувствителен к наличию процесса типа скользящего среднего в ошибке. Добавление в число регрессоров распределенного лага первой разности (с достаточно большим значением k) частично снимает эту проблему.

На практике решающим при использовании критерия ADF является вопрос о том, как выбирать k — порядок AR-процесса в оцениваемой регрессии. Можно предложить следующие подходы.

1) Выбирать k на основе обычных t- и F-статистик. Процедура состоит в том, чтобы начать с некоторой максимальной длины лага и проверять вниз, используя t- или F-статистики для значимости самого дальнего лага (лагов). Процесс останавливается, когда t-статистика или F-статистика значимы.

2) Использовать информационные критерии Акаике и Шварца. Длина лага с минимальным значением информационного критерия предпочтительна.

3) Сделать остатки регрессии ADF-критерия как можно более похожими на белый шум. Это можно проверить при помощи критерия на автокорреляцию. Если соответствующая статистика значима, то лаг выбран неверно и следует увеличить k.

Поскольку дополнительные лаги не меняют асимптотические результаты, то лучше взять больше лагов, чем меньше. Однако этот последний аргумент верен только с асимптотической точки зрения. ADF может давать разные результаты в зависимости от того, каким выбрано количество лагов. Даже добавление лага, который «не нужен» согласно только что приведенным соображениям, может резко изменить результат проверки гипотезы.

Особую проблему создает наличие сезонной компоненты в переменной. Если сезонность имеет детерминированный характер, то достаточно добавить в регрес-

сию фиктивные сезонные переменные — это не изменяет асимптотического распределения ADF-статистики. Для случая стохастической сезонности также есть специальные модификации критерия.

До сих пор рассматривались критерии I(1) против I(0). Временной ряд может быть интегрированным и более высокого порядка. Несложно понять, что критерии I(2) против I(1) сводятся к рассмотренным, если взять не уровень исследуемого ряда, а первую разность. Аналогичный подход рекомендуется для более высоких порядков интегрирования.

Имитационные эксперименты, проведенные Сэдом и Дики, показали, что следует проверять гипотезы последовательно, начиная с наиболее высокого порядка интегрирования, который можно ожидать априорно. То есть сначала следует проверить гипотезу о том, что ряд является I(2), и лишь после этого, если гипотеза отвергнута, что он является I(1).

17.5. Коинтеграция. Регрессии с интегрированными переменными

Как уже говорилось выше, привычные методы регрессионного анализа не подходят, если переменные нестационарны. Однако не всегда при применении МНК имеет место эффект ложной регрессии.

Говорят, что I(1)-процессы $\{x_{1t}\}$ и $\{x_{2t}\}$ являются **коинтегрированными** порядка 1 и (0, коротко CI(1, 0), если существует их линейная комбинация, которая является I(0), т.е. стационарна. Таким образом, процессы $\{x_{1t}\}$ и $\{x_{2t}\}$, интегрированные первого порядка I(1), — коинтегрированы, если существует коэффициент λ такой, что $x_{1t} - \lambda x_{2t} \sim$ I(0). Понятие коинтеграции введено **Грейнджером** в 1981 г. и использует модель исправления ошибок. Коинтегрированные процессы $\{x_{1t}\}$ и $\{x_{2t}\}$ связаны между собой долгосрочным стационарным соотношением, и предполагается, что существует некий корректирующий механизм, который при отклонениях возвращает x_{1t} и x_{2t} к их долгосрочному отношению.

Если $\lambda = 1$, то разность $x_{1t} - x_{2t}$ будет стационарной и, грубо говоря, x_{1t} и x_{2t} будут двигаться «параллельно» во времени. На рисунке 17.2 изображены две такие коинтегрированные переменные, динамика которых задана моделью исправления ошибок:

$$\Delta x_{1t} = -0.2(x_{1,t-1} - x_{2,t-1} + 2) + \varepsilon_{1t}, \tag{17.12}$$

$$\Delta x_{2t} = 0.5(x_{1,t-1} - x_{2,t-1} + 2) + \varepsilon_{2t}, \tag{17.13}$$

где ε_{1t} и ε_{2t} — независимые случайные величины, имеющие стандартное нормальное распределение.

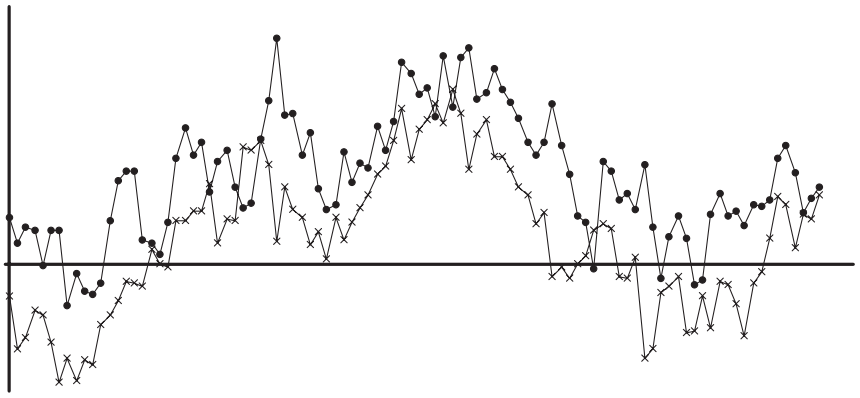

Рис. 17.2. Два коинтегрированных процесса при $\lambda = 1$

Если переменные в регрессии не стационарны, но действительно связаны друг с другом стационарной линейной комбинацией (модель специфицирована верно), то полученные оценки коэффициентов этой линейной комбинации будут на самом деле сверхсостоятельными, т.е. будут сходиться по вероятности к истинным коэффициентам со скоростью, пропорциональной не квадратному корню количества наблюдений $\sqrt{T}$, как в регрессии со стационарными переменными, а со скоростью, пропорциональной просто количеству наблюдений T. Другими словами, в обычной регрессии $\sqrt{T}(\lambda^* - \lambda)$ имеет невырожденное асимптотическое распределение, где λ^* — полученная из регрессии оценка λ, а в регрессии с $\mathsf{I}(1)$-переменными $T(\lambda^* - \lambda)$ имеет невырожденное асимптотическое распределение.

Обычные асимптотические аргументы сохраняют свою силу, если речь идет об оценках параметров краткосрочной динамики в модели исправления ошибок. Таким образом, можно использовать t-статистики, получаемые обычным методом наименьших квадратов, для проверки гипотез о значимости отдельных переменных. Важно помнить, что это относится к оценкам краткосрочных параметров. Этот подход не годится для проверки гипотез о коэффициентах коинтеграционной комбинации.

Определение коинтеграции естественным образом распространяется на случай нескольких переменных произвольного порядка интегрирования. Компоненты k-мерного векторного процесса $x_t = (x_{1t}, \ldots, x_{kt})$ называют **коинтегрированными** порядка d и b, что обозначается $x_t \sim \mathsf{CI}(d, b)$, если

1) каждая компонента x_{it} является $\mathsf{I}(d)$, $i = 1, \ldots, k$;

2) существует отличный от нуля вектор β, такой что $x_t\beta \sim \mathsf{I}(d-b)$, $d \geqslant b > 0$.

Вектор β называют **коинтегрирующим вектором**.

Коинтегрирующий вектор определен с точностью до множителя. В рассмотренном ранее примере коинтегрирующий вектор имеет вид $\beta = (-1, \lambda)$. Его можно пронормировать: $\beta = (-1/\lambda, 1)$, или так, чтобы сумма квадратов элементов была равна 1, т.е. $\beta = \left(-\frac{1}{\sqrt{1+\lambda^2}}, \frac{\lambda}{\sqrt{1+\lambda^2}}\right)$.

17.6. Оценивание коинтеграционной регрессии: подход Энгла—Грейнджера

Если бы коэффициент λ был известен, то выяснение того, коинтегрированы ли переменные x_{1t} и x_{2t}, было бы эквивалентно выяснению того, стационарна ли комбинация $x_{1t} - \lambda x_{2t}$ (например, с помощью критерия Дики—Фуллера). Но в практических ситуациях обычно стационарная линейная комбинация неизвестна. Значит, необходимо оценить коинтегрирующий вектор. После этого следует выяснить, действительно ли этот вектор дает стационарную линейную комбинацию.

Простейшим методом отыскания стационарной линейной комбинации является **метод Энгла—Грейнджера**. Энгл и Грейнджер предложили использовать оценки, полученные из обычной регрессии с помощью метода наименьших квадратов. Одна из переменных должна стоять в левой части регрессии, другая — в правой:

$$x_{1t} = \lambda x_{2t} + \varepsilon_t.$$

Для того чтобы выяснить, стационарна ли полученная линейная комбинация, предлагается применить метод Дики—Фуллера к остаткам из коинтеграционной регрессии. Нулевая гипотеза состоит в том, что ε_t содержит единичный корень, т.е. x_{1t} и x_{2t} не коинтегрированы. Пусть e_t — остатки из этой регрессии. Проверка нулевой гипотезы об отсутствии коинтеграции в методе Энгла—Грейнджера проводится с помощью регрессии:

$$e_t = \varphi e_{t-1} + u_t. \tag{17.14}$$

Статистика Энгла—Грейнджера представляет собой обычную t-статистику для проверки гипотезы $\varphi = 1$ в этой вспомогательной регрессии. Распределение статистики Энгла—Грейнджера будет отличаться (даже асимптотически), от распределения DF-статистики, но имеются соответствующие таблицы. Если мы отклоняем гипотезу об отсутствии коинтеграции, то это дает уверенность в том, что полученные результаты не являются ложной регрессией.

Игнорирование детерминированных компонент ведет к неверным выводам о коинтеграции. Чтобы этого избежать, в коинтеграционную регрессию следует добавить соответствующие переменные — константу, тренд, квадрат тренда, сезонные

фиктивные переменные. Например, добавляя константу и тренд, получим регрессию $x_{1t} = \mu_0 + \mu_1 t + \lambda x_{2t} + \varepsilon_t$. Такое добавление, как и в случае критерия DF, меняет асимптотическое распределение критерия Энгла—Грейнджера. Следует помнить, что, в отличие от критерия Дики—Фуллера, регрессия (17.14), из которой берется t-статистика, остается неизменной, то есть в нее не нужно добавлять детерминированные регрессоры.

В МНК-регрессии с коинтегрированными переменными оценки должны быть смещенными из-за того, что в правой части стоит эндогенная переменная x_{2t}, коррелированная с ошибкой ε_t. Кроме того, ошибка содержит пропущенные переменные. Коинтеграционная регрессия Энгла—Грейнджера является статической по форме, т.е. не содержит лагов переменных. С асимптотической точки зрения это не приводит к смещенности оценок, поскольку ошибка является величиной меньшего порядка, чем регрессор x_{2t}, дисперсия которого стремится к бесконечности. Как уже говорилось, оценки на самом деле сверхсостоятельны. Однако в малых выборках смещение может быть существенным.

После того как найдена стационарная линейная комбинация, можно оценить модель исправления ошибок (15.11), которая делает переменные коинтегрированными. В этой регрессии используются первые разности исходных переменных и остатки из коинтеграционной регрессии, которые будут представлять корректирующий элемент модели исправления ошибок:

$$\Delta x_{1t} = -\theta l_t + \sum_{j=1}^{p-1} \gamma_j \Delta x_{1,t-j} + \sum_{j=0}^{q-1} \beta_j \Delta x_{2,t-j} + v_t. \qquad (17.15)$$

Подчеркнем роль корректирующего элемента θl_t. До появления метода Энгла—Грейнджера исследователи часто оценивали регрессии в первых разностях, что хотя и приводило к стационарности переменных, но не учитывался стационарный корректирующий член, т.е. регрессионная модель была неверно специфицирована (проблема пропущенной переменной).

Несмотря на то, что в модели исправления ошибок (17.15) используется оценка коинтегрирующего вектора ($1 - \lambda$), оценки коэффициентов такой модели будут иметь такие же асимптотические свойства, как если бы коинтегрирующий вектор был точно известен. В частности, можно использовать t-статистики из этой регрессии (кроме t-статистики для θ), поскольку оценки стандартных ошибок являются состоятельными. Это является следствием сверхсостоятельности оценок коинтегрирующего вектора.

17.7. Коинтеграция и общие тренды

Можно предположить, что коинтеграция между двумя интегрированными переменными, скорее всего, проистекает из того факта, что обе они содержат одну и ту

же нестационарную компоненту, называемую **общим трендом**. Выше мы получили для интегрированной переменной разложение Бевериджа—Нельсона на детерминированный тренд, стохастический тренд и стационарную составляющую. Следует показать, что стохастические тренды в двух коинтегрированных переменных должны быть одними и теми же.

Проведем сначала подобный анализ для детерминированных трендов. Пусть процессы $\{x_t\}$ и $\{z_t\}$ стационарны относительно некоторого тренда $f(t)$, не обязательно линейного:

$$x_t = \mu_0 + \mu_1 f(t) + \varepsilon_t \quad \text{и} \quad z_t = \nu_0 + \nu_1 f(t) + \xi_t,$$

где ε_t и ξ_t — два стационарных процесса. В каком случае линейная комбинация этих двух процессов будет стационарной в обычном смысле (не относительно тренда)? Найдем линейную комбинацию $x_t - \lambda z_t$:

$$x_t - \lambda z_t = \mu_0 - \lambda\nu_0 + (\mu_1 - \lambda\nu_1) f(t) + \varepsilon_t - \lambda\xi_t.$$

Для ее стационарности требуется, чтобы $\mu_1 = \lambda\nu_1$.

С другой стороны, если бы $\{x_t\}$ содержал тренд $f(t)$, а $\{z_t\}$ — отличный от него тренд $g(t)$, то, в общем случае, не нашлось бы линейной комбинации, такой что $\mu_1 f(t) - \lambda\nu_1 g(t)$ оказалась бы постоянной величиной. Следовательно, для $\{x_t\}$ и $\{z_t\}$ не нашлось бы коинтегрирующего вектора. Коинтегрирующий вектор можно найти только в том случае, если $f(t) = \dfrac{\lambda\nu_1}{\mu_1} g(t)$ для некоторого λ, т.е. если $f(t)$ и $g(t)$ линейно зависимы.

Пусть теперь $\{x_t\}$ и $\{z_t\}$ — два процесса I(1), для которых существуют разложения Бевериджа—Нельсона:

$$x_t = \gamma t + v_t + \varepsilon_t,$$
$$z_t = \delta t + w_t + \xi_t,$$

где v_t и w_t — случайные блуждания, а ε_t и ξ_t — стационарные процессы.

Найдем условия, при которых линейная комбинация x_t и z_t,

$$x_t - \lambda z_t = \gamma t + v_t + \varepsilon_t - \lambda(\delta t + w_t + \xi_t) = (\gamma - \lambda\delta)t + v_t - \lambda w_t + \varepsilon_t - \lambda\xi_t,$$

может быть стационарной. Для стационарности требуется, чтобы в получившейся переменной отсутствовал как детерминированный, так и стохастический тренд. Это достигается при $\gamma = \lambda\delta$ и $v_t = \lambda w_t$. При этом x_t можно записать как

$$x_t = \lambda(\delta t + w_t) + \varepsilon_t,$$

т.е. $\{x_t\}$ и $\{z_t\}$ содержат общий тренд $\delta t + w_t$.

Этот взгляд на коинтеграцию развили в 1988 г. **Сток** и **Уотсон**: пусть есть k интегрированных переменных, которые коинтегрированы. Тогда каждая из этих переменных может быть записана как стационарная компонента плюс линейная комбинация меньшего количества общих трендов.

17.8. Упражнения и задачи

Упражнение 1

Сгенерируйте 20 марковских процессов $x_t = \varphi x_{t-1} + \varepsilon_t$ при различных коэффициентах авторегрессии: а) $\varphi = 0.1$; б) $\varphi = 0.9$; в) $\varphi = 1$; г) $\varphi = 1.02$. В качестве ошибки используйте нормально распределенный белый шум с единичной дисперсией и возьмите $x_0 = 0$.

1.1. Изобразите графики для всех 20 рядов для каждого из четырех случаев. Какой вывод можно сделать?

1.2. Рассчитайте для каждого из четырех случаев дисперсию значений соответствующих 20 рядов, рассматривая их как выборку для $t = 1, \ldots, 100$. Нарисуйте график дисперсии по времени для всех четырех случаев. Сделайте выводы.

1.3. Оцените для всех рядов авторегрессию первого порядка и сравните распределения оценок для всех 4 случаев с истинными значениями. Сделайте выводы.

Упражнение 2

Покажите эффект ложной регрессии для переменных I(1) с помощью метода Монте-Карло. Сгенерируйте по 100 рядов x_t и z_t по модели случайного блуждания:

$$x_t = x_{t-1} + \varepsilon_t \quad \text{и} \quad z_t = z_{t-1} + \xi_t,$$

где ошибки $\varepsilon_t \sim N(0,1)$ и $\xi_t \sim N(0,1)$ неавтокоррелированы и некоррелированы друг с другом.

2.1. Оцените для всех 100 наборов данных регрессию x_t по константе и z_t:

$$x_t = az_t + b + u_t.$$

Рассчитайте соответствующие статистики Стьюдента для коэффициента a и проанализируйте выборочное распределение этих статистик. В скольки

процентах случаев нулевая гипотеза (гипотеза о том, что коэффициент равен нулю) отвергается, если использовать стандартную границу t-распределения с уровнем доверия 95%? Найдите оценку фактической границы уровня доверия 95%.

2.2. Проанализируйте для тех же 100 регрессий выборочное распределение коэффициента детерминации.

2.3. Возьмите пять первых наборов данных и проверьте ряды на наличие единичных корней с помощью теста Дики—Фуллера.

2.4. Повторите упражнения 2.1, 2.2 и 2.3, сгенерировав данные x_t и z_t по стационарной модели AR(1) с авторегрессионным коэффициентом 0.5 и сравните с полученными ранее результатами. Сделайте выводы.

Упражнение 3

Сгенерируйте 100 рядов по модели случайного блуждания с нормально распределенным белым шумом, имеющим единичную дисперсию. Проверьте с помощью сгенерированных данных одно из значений в таблице статистики Дики—Фуллера.

Упражнение 4

Рассмотрите данные из таблицы 15.3 на стр. 520.

4.1. Преобразуйте ряды, перейдя к логарифмам, и постройте их графики. Можно ли сказать по графику, что ряды содержат единичный корень?

4.2. Проверьте формально ряды на наличие единичных корней с помощью дополненного теста Дики—Фуллера, включив в регрессии линейный тренд и 4 лага разностей.

4.3. Используя МНК, оцените параметры модели $C_t = \alpha Y_t + \beta + \varepsilon_t$, вычислите остатки из модели. К остатком примените тест Энгла—Грэйнджера.

4.4. Используя коэффициенты из упражнения 4.3, оцените модель исправления ошибок с четырьмя лагами разностей.

Задачи

1. Задан следующий процесс: $x_t = 0.8x_{t-1} + 0.2x_{t-2} + \varepsilon_t - 0.9\varepsilon_{t-1}$. При каком d процесс $\Delta^d x_t$ будет стационарным?

2. Изобразите графики автокорреляционной функции и спектра для второй разности стохастического процесса, содержащего квадратический тренд.

3. Дан процесс $x_t = f_t + \varepsilon_t$, $f_t = f_{t-1} + \nu_t$, где $\varepsilon_t \sim \mathrm{N}(0, 2)$ и $\nu_t \sim \mathrm{N}(0, 1)$. Определить тип процесса, перечислить входящие в его состав компоненты и вычислить условное математическое ожидание и условную дисперсию процесса x_t.

4. Чем отличается стохастический тренд от обычного линейного тренда с точки зрения устранения проблемы ложной регрессии?

5. Процессы x_t и y_t заданы уравнениями $x_t = \alpha x_{t-2} + \varepsilon_t$ и $y_t = \beta x_t + \xi_t + \gamma\xi_{t-1}$, где ε_t и ξ_t — стационарные процессы. При каких условиях на параметры α, β и γ можно было бы сказать, что x_t и y_t коинтегрированы как $\mathsf{CI}(1, 0)$?

6. Известно, что $z_t = \sum_{i=1}^{t} x_i$, где x_i — процесс случайного блуждания, а $y_t = \varepsilon_t + 0.5\varepsilon_{t-1} + 0.25\varepsilon_{t-2} + 0.125\varepsilon_{t-3} + \ldots$. Можно ли построить регрессию z_t от y_t? Ответ обосновать.

7. Пусть $x_t = \varepsilon + bx_{t-1} + \varepsilon_t$, где $\varepsilon_t \sim \mathrm{N}(0, 1)$ и $b \leqslant 1$, а z_t — стохастический процесс, содержащий линейный тренд. Можно ли установить регрессионную связь между x_t и z_t? Если да, то как?

8. Получены оценки МНК в регрессии $x_t = \varphi x_{t-1} + \varepsilon_t$. Приведите вид статистики, используемой в тесте Дики—Фуллера.

9. Правомерно ли построение долгосрочной зависимости y_t по x_t, если y_t — процесс случайного блуждания с шумом, x_t — процесс случайного блуждания с дрейфом? Если нет,— ответ обосновать. Если да, — изложить этапы построения регрессии с приведением формул, соответствующих каждому этапу.

10. В чем преимущества дополненного теста Дики—Фуллера по сравнению с обычным DF-тестом? Привести формулы.

Рекомендуемая литература

1. **Магнус Я.Р., Катышев П.К., Пересецкий А.А.** Эконометрика — начальный курс.— М.: «Дело», 2000 (гл. 12 — стр. 240–249).

2. **Banerjee A., Dolado J.J., Galbraith J.W. and Hendry D.F.**, Co-integration, Error Correction and the Econometric Analysis of Non-stationary Data, Oxford University Press, 1993 (гл. 3–5)

3. **Davidson, R., and J.G. MacKinnon.** Estimation and Inference in Econometrics. Oxford University Press, 1993 (Гл. 20.)

4. **Dickey, D.A. and Fuller W.A.**, «Distributions of the Estimators for Autoregressive Time Series With a Unit Root», Journal of American Statistical Association, 75 (1979), 427–431.

5. **Enders, W.** Applied Econometric Time Series. John Wiley & Sons, 1995.

6. **Engle, R.F. and Granger C.W.J.**, «Co-integration and Error Correction: Representation, Estimation and Testing», Econometrica, 55 (1987), 251–276.

7. **Granger, C.W.J., and Newbold P.** «Spurious Regressions in Econometrics», Journal of Econometrics, 21 (1974), 111–120.

8. **Greene W.H.** Econometric Analysis, Prentice-Hall, 2000. (гл.18 — стр. 776–784)

9. **Said, E.S. and Dickey D.A.**, «Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order», Biometrica, 71 (1984), 599–607.

10. **Stock, J.H. and Watson M.W.**, «Testing for Common Trends», Journal of the American Statistical Association, **83** (1988), 1097–1107.

11. **Stock, J.H**., «Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors», Econometrica, 55 (1987), 1035–1056.

12. **Wooldridge Jeffrey M.** Introductory Econometrics: A Modern Approach, 2nd ed., Thomson, 2003. (Ch. 18).

Часть IV

Эконометрия — II

Это пустая страница

Глава 18

Классические критерии проверки гипотез

18.1. Оценка параметров регрессии при линейных ограничениях

Рассмотрим модель линейной регрессии $X = Z\alpha + \varepsilon$ в случае, когда известно, что коэффициенты α удовлетворяют набору линейных ограничений

$$R\alpha = r,$$

где R — матрица размерности $k \times (n+1)$, а r — вектор свободных частей ограничений длины k. Метод наименьших квадратов для получения более точных оценок в этом случае следует модифицировать, приняв во внимание ограничения. Оценки наименьших квадратов при ограничениях получаются из задачи условной минимизации:

$$\begin{cases} (X - Z\alpha)'(X - Z\alpha) \to \min\limits_{\alpha}! \\ R\alpha = r. \end{cases}$$

Запишем для этой задачи функцию Лагранжа:

$$L = (X - Z\alpha)'(X - Z\alpha) + 2\lambda'(R\alpha - r),$$

где λ — вектор-столбец множителей Лагранжа. Возьмем производные (см. Приложение А.2.2):

$$\frac{\partial L}{\partial \lambda} = -2(R\alpha - r),$$
$$\frac{\partial L}{\partial \alpha} = -2(Z'(X - Z\alpha) - R'\lambda).$$

Следовательно, оценки наименьших квадратов при ограничениях, a_1, находятся из уравнений:

$$Ra_1 - r = 0,$$
$$Z'(X - Za_1) - R'\lambda = 0.$$

Умножим второе уравнение слева на $(Z'Z)^{-1}$. Получим

$$a_1 = (Z'Z)^{-1} Z'X - (Z'Z)^{-1} R'\lambda.$$

Пусть a_0 — оценки α без учета ограничений, то есть $a_0 = (Z'Z)^{-1} Z'X$. Тогда

$$a_1 = a_0 - (Z'Z)^{-1} R'\lambda. \tag{18.1}$$

Если умножим это уравнение слева на R, то получим

$$R(Z'Z)^{-1} R'\lambda = Ra_0 - r.$$

Отсюда

$$\lambda = A^{-1}(Ra_0 - r),$$

где мы ввели обозначение $A = R(Z'Z)^{-1} R'$. Это можно проинтерпретировать следующим образом: чем сильнее нарушается ограничение на оценках из регрессии без ограничений, тем более значимы эти ограничения. Подставим множители Лагранжа обратно в уравнение (18.1):

$$a_1 = a_0 - (Z'Z)^{-1} R'A^{-1}(Ra_0 - r). \tag{18.2}$$

Таким образом, оценки a_1 и a_0 различаются тем сильнее, чем сильнее нарушается ограничение $R\alpha = r$ в точке a_0, т.е. чем больше невязки $Ra_0 - r$.

Докажем **несмещенность** оценок. Для этого в выражении для a везде заменим $\bar{a}$ на

$$(Z'Z)^{-1} Z'X = (Z'Z)^{-1} Z'(Z\alpha + \varepsilon) = (Z'Z)^{-1} Z'(Z\alpha + \varepsilon) = \alpha + (Z'Z)^{-1} Z'\varepsilon.$$

Получим

$$a_1 = \alpha + (Z'Z)^{-1} Z'\varepsilon - (Z'Z)^{-1} R' A^{-1} \left(R\alpha + R (Z'Z)^{-1} Z'\varepsilon - r \right).$$

При выполнении нулевой гипотезы $R\alpha = r$ можем упростить это выражение:

$$a_1 = \alpha + (Z'Z)^{-1} Z'\varepsilon - (Z'Z)^{-1} R' A^{-1} R (Z'Z)^{-1} Z'\varepsilon = \\ = \alpha + \left(I - (Z'Z)^{-1} R' A^{-1} R \right) (Z'Z)^{-1} Z'\varepsilon.$$

Но по предположению классической модели регрессии $\mathbf{E}(\varepsilon) = 0$, следовательно, математическое ожидание второго слагаемого здесь равно нулю. А значит, $\mathbf{E}(a_1) = \alpha$. Несмещенность оценки a_1 доказана.

Величина $\left(I - (Z'Z)^{-1} R' A^{-1} R \right) (Z'Z)^{-1} Z'\varepsilon$ представляет собой отличие a_1 от α.

Найдем теперь ковариационную матрицу оценок a. После преобразований получаем

$$\mathbf{cov}(a_1) = \mathbf{E}\left[(a_1 - \alpha)(a_1 - \alpha)' \right] = \sigma^2 \left(I - (Z'Z)^{-1} R' A^{-1} R \right) (Z'Z)^{-1},$$

где используется еще одно предположение модели регрессии — отсутствие автокорреляции и гетероскедастичности в ошибках, т.е. $\mathbf{E}(\varepsilon\varepsilon') = \sigma^2 I$.

В то же время $\mathbf{cov}(a_0) = \sigma^2 (Z'Z)^{-1}$. Таким образом, ковариационные матрицы оценок различаются на положительно (полу-) определенную матрицу $\sigma^2 (Z'Z)^{-1} R' A^{-1} R (Z'Z)^{-1}$, что можно интерпретировать в том смысле, что оценки a являются более точными, чем $\bar{a}$, поскольку учитывают дополнительную информацию о параметрах.

Насколько отличаются между собой суммы квадратов остатков в двух рассматриваемых моделях? Ясно, что в регрессии с ограничениями сумма квадратов не может быть ниже (так как минимизируется та же функция при дополнительных ограничениях). Вычислим разность между двумя суммами квадратов остатков.

Пусть $e_0 = X - Za_0$ — вектор остатков при оценке параметров регрессии без ограничений, а $e_1 = X - Za_1$ — вектор остатков при оценке параметров с ограничениями, и пусть $RSS(a_0) = e_1' e_1$, $RSS(a_0) = e_0' e_0$ — соответствующие суммы квадратов остатков. Из (18.2) получаем

$$X - Za_1 = X - Za_0 + Z (Z'Z)^{-1} R' A^{-1} (Ra_0 - r),$$

или

$$e_1 = e_0 + Z (Z'Z)^{-1} R' A^{-1} (Ra_0 - r).$$

Введем обозначение: $\delta = Z\left(Z'Z\right)^{-1}R'A^{-1}\left(Ra_0 - r\right)$, тогда

$$e_1 = e_0 + \delta.$$

Поскольку $Z'e_0 = 0$, то $\delta'e_0 = 0$. Следовательно,

$$e_1'e_1 = e_0'e_0 + \delta'e_0 + e_0'\delta + \delta'\delta = e_0'e_0 + \delta'\delta.$$

Здесь

$$\delta'\delta = \left(Z\left(Z'Z\right)^{-1}R'A^{-1}\left(Ra_0 - r\right)\right)'\left(Z\left(Z'Z\right)^{-1}R'A^{-1}\left(Ra_0 - r\right)\right) = \\ = \left(Ra_0 - r\right)'A^{-1}R\left(Z'Z\right)^{-1}R'A^{-1}\left(Ra_0 - r\right),$$

или, учитывая определение матрицы A,

$$\delta'\delta = \left(Ra_0 - r\right)'A^{-1}\left(Ra_0 - r\right).$$

Итак,

$$RSS(a_1) - RSS(a_0) = e_1'e_1 - e_0'e_0 = \left(Ra_0 - r\right)'A^{-1}\left(Ra_0 - r\right).$$

Как и следовало ожидать, эта разность оказывается неотрицательной.

18.2. Тест на существенность ограничения

Пусть, как и раньше,

e_0 — вектор остатков в регрессии без ограничений,

e_1 — вектор остатков в регрессии с ограничениями,

N — число наблюдений,

n — количество факторов,

k — общее количество ограничений на параметры регрессии.

В случае, если ограничения $R\alpha = r$ выполнены, то

$$F^c = \frac{(e_1'e_1 - e_0'e_0)/k}{e_0'e_0/(N-n-1)} \sim F_{k,\,N-n-1}. \tag{18.3}$$

Иными словами, статистика, равная относительному приросту суммы квадратов остатков в регрессии с ограничениями по сравнению с регрессией без ограничений, скорректированному на степени свободы, имеет распределение Фишера (см. Приложение А.3.2). Чем больше $RSS(a_1)$ будет превышать $RSS(a_0)$, тем более существенно ограничение и тем менее правдоподобно, что ограничение выполнено.

Докажем, что распределение статистики будет именно таким, если предположить, что ошибки имеют нормальное распределение:

$$\varepsilon \sim N(0, \sigma^2 I).$$

Мы видели, что $e_1 = e_0 + \delta$. Выразим δ через ε с учетом нулевой гипотезы $R\alpha = r$:

$$\delta = Z\left(Z'Z\right)^{-1} R' A^{-1}\left(Ra_0 - r\right) = Z\left(Z'Z\right)^{-1} R' A^{-1} R\left(Z'Z\right)^{-1} Z'\varepsilon =$$
$$= Z\left(Z'Z\right)^{-1} R' \left(R\left(Z'Z\right)^{-1} R'\right)^{-1} R\left(Z'Z\right)^{-1} Z'\varepsilon.$$

Для вектора остатков e_0 имеем

$$e_0 = I - Z\left(Z'Z\right)^{-1} Z'\varepsilon.$$

Совместное распределение векторов δ и e_0 является нормальным (так как они линейно выражаются через ошибки ε). Эти вектора некоррелированы:

$$E\left(\delta e_0'\right) =$$
$$= Z\left(Z'Z\right)^{-1} R' \left(R\left(Z'Z\right)^{-1} R'\right)^{-1} R\left(Z'Z\right)^{-1} Z' E\left(\varepsilon\varepsilon'\right) \left(I - Z\left(Z'Z\right)^{-1} Z'\right) =$$
$$= \sigma^2 Z\left(Z'Z\right)^{-1} R' \left(R\left(Z'Z\right)^{-1} R'\right)^{-1} R\left(Z'Z\right)^{-1} Z' \left(I - Z\left(Z'Z\right)^{-1} Z'\right) = 0.$$

Последнее равенство здесь следует из того, что

$$Z'\left(I - Z\left(Z'Z\right)^{-1} Z'\right) = Z' - Z'Z\left(Z'Z\right)^{-1} Z' = Z' - Z' = 0.$$

По свойствам многомерного нормального распределения это означает, что δ и e_0 независимы (см. Приложение А.3.2). Кроме того δ имеет форму $Q\left(Q'Q\right)^{-1} Q'\varepsilon$, где $Q = Z\left(Z'Z\right)^{-1} R'$.

Для вектора же e_0 матрица преобразования равна $I - Z\left(Z'Z\right)^{-1} Z'$. Обе матрицы преобразования являются симметричными и идемпотентными. Ранг матрицы преобразования (другими словами, количество степеней свободы) равен k для δ и $N - n - 1$ для e_0. С учетом того, что $\varepsilon \sim N(0, \sigma^2 I)$, это означает:

$$\frac{1}{\sigma^2}\delta'\delta = \frac{1}{\sigma^2}\left(e_1'e_1 - e_0'e_0\right) \sim \chi^2_k,$$
$$\frac{1}{\sigma^2}e_0'e_0 \sim \chi^2_{N-n-1},$$

причем эти две величины независимы. Частное этих величин, деленных на количество степеней свободы, распределено как $F_{k,N-n-1}$, что и доказывает утверждение.

Проверяемая нулевая гипотеза имеет вид:

$$H_0 : R\alpha = r,$$

то есть ограничения выполнены. Критерий заключается в следующем: если $F^c > F_{k,\,N-n-1,\,1-\theta}$, то нулевая гипотеза отвергается, если $F^c < F_{k,\,N-n-1,\,1-\theta}$, то она принимается.

Распишем статистику подробнее:

$$F^c = \frac{(Ra_0 - r)' \left(R\left(Z'Z\right)^{-1} R'\right)^{-1} (Ra_0 - r)/k}{(X - Za_0)'(X - Za_0)/(N - n - 1)} \tag{18.4}$$

Покажем, что F-критерий базового курса — это частный случай данного критерия. С помощью F-критерия для регрессии в целом проверяется нулевая гипотеза о том, что все коэффициенты α, кроме свободного члена (последнего, $(n+1)$-го коэффициента), равны нулю:

$$H_0 : \alpha_1 = 0,\ \ldots,\ \alpha_n = 0.$$

Запишем эти ограничения на коэффициенты в матричном виде ($R\alpha = r$). Соответствующие матрицы будут равны

$$R = [I_n; 0_n] \quad \text{и} \quad r = 0_n.$$

Обозначим матрицу Z без последнего столбца (константы) через Z_-. В этих обозначениях $Z = [Z_-; 1_N]$. Тогда

$$Z'Z = N \begin{bmatrix} M + \bar{Z}'\bar{Z} & \bar{Z}' \\ \bar{Z} & 1 \end{bmatrix}$$

и

$$\left(Z'Z\right)^{-1} = \frac{1}{N} \begin{bmatrix} M^{-1} & -M^{-1}\bar{Z}' \\ -\bar{Z}M^{-1} & 1 + \bar{Z}M^{-1}\bar{Z}' \end{bmatrix},$$

где $\bar{Z} = \dfrac{1}{N} 1'_N Z_-$ — вектор средних, а $M = \dfrac{1}{N}\hat{Z}'\hat{Z}$ — ковариационная матрица факторов Z_-. Умножение $\left(Z'Z\right)^{-1}$ слева и справа на R выделяет из этой матрицы левый верхний блок:

$$R\left(Z'Z\right)^{-1} R' = \frac{1}{N} M^{-1}.$$

Подставим это выражение в (18.4), обозначая через $a_- = Ra_0$ вектор из первых n компонент оценок a_0 (все коэффициенты за исключением константы):

$$F^c = \frac{(a'_- M a_-)/n}{(X - Za_0)'(X - Za_0)/(N(N - n - 1))}.$$

Здесь $a'_- Ma_-$ — объясненная дисперсия в регрессии, а $(X-Za_0)'(X-Za_0)/N = s_e^2$ — остаточная дисперсия (смещенная оценка). Видим, что

$$F^c = \frac{R^2(N-n-1)}{(1-R^2)n},$$

а это стандартная форма F-статистики.

Более простой способ получения этого результата состоит в том, чтобы возвратиться к исходной задаче условной минимизации. Ограничение, что все коэффициенты, кроме свободного члена, равны нулю, можно подставить непосредственно в целевую функцию (сумму квадратов остатков). Очевидно, что решением задачи будет $a_{n+1} = \dfrac{1}{N}X'1_N = \bar{x}$ и $a_j = 0,\ j \neq n+1$, т.е.

$$a_1 = \begin{pmatrix} 0_n \\ \\ \bar{x} \end{pmatrix}.$$

Отсюда получаем, что остатки равны центрированным значениям зависимой переменной: $e_1 = \hat{X}$. Следовательно, из (18.3) получим

$$F^c = \frac{(\hat{X}'\hat{X} - e_0'e_0)/n}{e_0'e_0/(N-n-1)}.$$

В числителе стоит объясненная сумма квадратов, а в знаменателе — сумма квадратов остатков.

Покажем, что t-критерий также является частным случаем данного критерия. Нулевая гипотеза заключается в том, что j-й параметр регрессии равен нулю. Таким образом, в этом случае $k = 1$,

$$R = (0, \ldots, 0, \underset{j}{1}, 0, \ldots, 0) \quad \text{и} \quad r = 0.$$

При этом $R(Z'Z)^{-1}R' = m_{jj}^{-1}$, где m_{jj}^{-1} — j-й диагональный элемент матрицы M^{-1}, а $M = \dfrac{1}{N}Z'Z$ — матрица вторых начальных моментов факторов Z. В числителе F-статистики стоит несмещенная оценка остаточной дисперсии

$$\hat{s}_e^2 = (X - Za_0)'(X - Za_0)/(N-n-1).$$

Кроме того, $Ra_0 - r = a_{0j}$. Окончательно получаем

$$F^c = \frac{a_{0j}^2}{m_{jj}^{-1}\hat{s}_e^2/N} \sim F_{1,\,N-n-1}.$$

Здесь $m_{jj}^{-1}\hat{s}_e^2/N = s_{a_j}^2$ — оценка дисперсии коэффициента a_{0j}. Видим, что F-статистика имеет вид:

$$F^c = \frac{a_{0j}^2}{s_{a_j}^2} = \left(\frac{a_{0j}}{s_{a_j}}\right)^2 = t_j^2,$$

т.е. это квадрат t-статистики для коэффициента a_{0j}. Для квантилей F-распределения выполнено $F_{1,\,N-n-1,\,1-\theta} = t^2_{N-n-1,\,1-\theta}$. Нулевая гипотеза H_0 : $\alpha_j = 0$ отвергается, если $F^c > F_{1,\,N-n-1,\,1-\theta}$, т.е. если $|t_j| = \sqrt{F^c} > t_{N-n-1,\,1-\theta}$. Видим, что два критерия полностью эквивалентны.

Рассмотрим регрессию, в которой факторы разбиты на два блока:

$$X = Z_1\alpha_1 + Z_2\alpha_2 + \varepsilon.$$

В качестве промежуточного случая между F-критерием для регрессии в целом и t-критерием для одного фактора рассмотрим F-критерий для группы факторов. Требуется получить ответ на вопрос о том, нужна ли в регрессии группа факторов Z_2. Для проверки гипотезы $H_0 : \alpha_2 = 0$ мы можем воспользоваться полученными выше результатами. В этом случае $k = n_2$, где n_2 — количество факторов во второй группе, а матрицы, задающие ограничение, равны

$$R = \underbrace{\left[\begin{array}{ccc|ccc} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ \vdots & & \vdots & 0 & 1 & \cdots & 0 \\ & & & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & 1 \end{array}\right]}_{n_1 \qquad\qquad\qquad n_2} = [0_{n_2\times n_1}; I_{n_2}] \quad \text{и} \quad r = 0_{n_2}.$$

Очевидно, что оценки с этими ограничениями будут равны

$$a = \begin{pmatrix} (Z'_1Z_1)^{-1}\,Z'_1X \\ 0_{n_2} \end{pmatrix},$$

а остатки можно найти по формуле $e_1 = \left(I_N - Z_1\,(Z'_1Z_1)^{-1}\,Z'_1\right)X$.

F-статистику можно рассчитать по общей формуле (18.3):

$$F^c = \frac{(e_1'e_1 - e_0'e_0)/n_2}{e_0'e_0/(N-n-1)} \sim F_{n_2,N-n-1}.$$

Мы неявно рассматривали здесь тест на исключение факторов. Можно рассматривать его с другой точки зрения — как тест на включение факторов, при этом формулы не поменяются. То есть мы, оценив регрессию $X = Z_1\alpha_1 + \varepsilon$, можем проверить, следует ли включать в нее дополнительные факторы Z_2.

Тест на включение факторов особенно полезен для проверки того, не нарушаются ли предположения модели регрессии. Это так называемые диагностические тесты. Все они строятся по одному и тому же принципу: если модель $X = Z_1\alpha_1 + \varepsilon$ специфицирована корректно, то любые дополнительные факторы Z_2 скорее всего будут незначимы в тесте на их включение (т.е. с большой вероятностью будет принята нулевая гипотеза $\alpha_2 = 0$). Факторы Z_2 конструируются таким образом, чтобы тест имел бóльшую мощность против определенного класса нарушений предположений модели регрессии. Из сказанного следует, что во всех диагностических тестах нулевой гипотезой является то, что базовая модель корректно специфицирована. Если нулевая гипотеза будет отвергнута, то естественно искать другую модель, которая бы лучше описывала имеющиеся данные. Следует понимать, что регрессия $X = Z_1\alpha_1 + Z_2\alpha_2 + \varepsilon$ в этом случае будет носить обычно вспомогательный характер, то есть оценки коэффициентов в ней, вообще говоря, не призваны нести смысловой нагрузки. Она нужна только для проверки базовой модели $X = Z_1\alpha_1 + \varepsilon$. (Хотя здесь есть, конечно, и исключения.)

18.2.1. Тест Годфрея (на автокорреляцию ошибок)

Оценим регрессию $X = Z_1\alpha_1 + \varepsilon$. Для проверки отсутствия автокорреляции p-го порядка попытаемся ввести такой набор факторов:

$$Z_2 = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ e_1 & 0 & \cdots & 0 \\ e_2 & e_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & e_1 \\ \vdots & \vdots & \vdots & \vdots \\ e_{N-1} & \cdots & \cdots & e_{N-p} \end{bmatrix}.$$

Столбцы матрицы Z_2 состоят из лагов остатков, дополненных нулями. Если нулевая гипотеза ($\alpha_2 = 0$) отвергается, то делается вывод о наличии автокорреляции. При $p = 1$ тест Годфрея представляет собой близкий аналог теста

Дарбина—Уотсона, однако при $p > 1$ помогает обнаруживать и автокорреляцию более высоких порядков, в чем и состоит его преимущество.

18.2.2. Тест RESET Рамсея (Ramsey RESET test) на функциональную форму уравнения

Рассмотрим модель $X = Z_1\alpha_1 + \varepsilon$, которую можно записать в виде $X = X^0 + \varepsilon$, где $X^0 = Z_1\alpha_1$. Можно рассмотреть возможность наличия между X и X^0 более сложной нелинейной зависимости, например, квадратичной. Для проверки линейности модели против подобной альтернативы служит тест Рамсея.

Оценим регрессию $X = Z_1\alpha_1 + \varepsilon$. Попытаемся ввести фактор

$$Z_2 = \begin{pmatrix} (x_1^c)^2 \\ \vdots \\ (x_N^c)^2 \end{pmatrix},$$

где $X^c = Z_1 a_1$ — расчетные значения из проверяемой регрессии. Если $\alpha_2 = 0$, то зависимость линейная, если $\alpha_2 \neq 0$, то зависимость квадратичная.

Можно тем же способом проводить тест на 3-ю, 4-ю степени и т.д. Матрица добавляемых факторов имеет следующую общую форму:

$$Z_2 = \begin{pmatrix} (x_1^c)^2 & (x_1^c)^3 & \cdots & (x_N^c)^m \\ \vdots & \vdots & & \vdots \\ (x_N^c)^2 & (x_N^c)^3 & \cdots & (x_N^c)^m \end{pmatrix}.$$

18.2.3. Тест Чоу (Chow-test) на постоянство модели

Часто возникает сомнение в том, что для всех наблюдений $1, \ldots, N$ модель неизменна, в частности, что параметры неизменны.

Пусть все наблюдения в регрессии разбиты на две группы. В первой из них — N_1 наблюдений, а во второй — N_2 наблюдений, так что $N_1 + N_2 = N$. Без ограничения общности можно считать, что сначала идут наблюдения из первой группы, а потом из второй. Базовую регрессию $X = Z\alpha + \varepsilon$ можно представить в следующем блочном виде:

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix} \alpha + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}.$$

Требуется проверить, действительно ли наблюдения в обеих группах подчиняются одной и той же модели $x = z\alpha + \varepsilon$.

1-я форма теста Чоу

В качестве альтернативы базовой модели рассмотрим регрессию

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}.$$

Фактически, это две разные регрессии:

$$X_1 = Z_1\alpha_1 + \varepsilon_1 \quad \text{и} \quad X_2 = Z_2\alpha_2 + \varepsilon_2,$$

но предполагается, что дисперсия в них одинакова.

Пусть e_0 — вектор остатков в регрессии на всей выборке, e_1 — вектор остатков в регрессии с наблюдениями $1, \ldots, N_1$, e_2 — вектор остатков в регрессии с $N_1 + 1, \ldots, N$ наблюдениями. Требуется проверить нулевую гипотезу о равенстве коэффициентов по двум частям выборки:

$$H_0 : \alpha_1 = \alpha_2.$$

Для того чтобы применить здесь тест добавления переменных, обозначим $\alpha_2 - \alpha_1 = \delta$ и подставим $\alpha_2 = \alpha_1 + \delta$ в рассматриваемую модель:

$$X_2 = Z_2\alpha_1 + Z_2\delta + \varepsilon_2,$$

Таким образом, можно рассматривать следующую регрессию (для упрощения обозначений пишем α вместо α_1):

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{bmatrix} Z_1 & 0 \\ Z_2 & Z_2 \end{bmatrix} \begin{pmatrix} \alpha \\ \delta \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}. \tag{18.5}$$

Мы хотим проверить гипотезу $H_0 : \delta = 0$.

Если нулевая гипотеза принимается, то это означает, что $\alpha_1 = \alpha_2$, т.е. коэффициенты постоянны.

Заметим, что в регрессии (18.5) с ограничениями остатки окажутся равными e_0, а в регрессии (18.5) без ограничений — $\begin{pmatrix} e_1 \\ e_2 \end{pmatrix}$. Суммы квадратов остатков равны

$e_0'e_0$ и $e_1'e_1 + e_2'e_2$, соответственно. В регрессии с ограничениями оценивается $n+1$ параметров, без ограничений — $2(n+1)$. Всего проверяется $k = n+1$ ограничений. Используя общую формулу (18.2), получим

$$F^c = \frac{(e_0'e_0 - e_1'e_1 - e_2'e_2)/(n+1)}{(e_1'e_1 + e_2'e_2)/(N - 2(n+1))} \sim F_{n+1,\, N-2(n+1)}.$$

Для того чтобы применить этот тест, нужно оценить модель по двум частям выборки. Это можно сделать, когда количество наблюдений превышает количество параметров, т.е. $N_1 \geqslant n+1$ и $N_2 \geqslant n+1$. Кроме того, если $N_j = n+1$, то остатки $e_j = 0$ $(j = 1,\ 2)$. Таким образом, для применимости теста требуется, чтобы хотя бы в одной части количество наблюдений превышало количество параметров.

2-я форма теста Чоу

То, что модель в двух частях одна и та же, можно проверить и по-другому. Пусть сначала рассчитывается регрессия по $N_1 < N$ наблюдениям, а затем по всем N наблюдениям. Если полученные результаты существенно отличаются, то это должно означать, что во второй части модель каким-то образом поменялась.

Реализуем эту идею с помощью вспомогательной регрессии, к которой можно применить тест добавления переменных:

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{bmatrix} Z_1 & 0_{N_1 \times N_2} \\ Z_2 & I_{N_2} \end{bmatrix} \begin{pmatrix} \alpha \\ \delta \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}. \tag{18.6}$$

Эта регрессия с ограничением $\delta = 0$ совпадает с регрессией по всем N наблюдениям по первоначальной модели, и остатки равны e_0.

Если же оценить (18.6, 18.5) без ограничений, то, как можно показать, оценки α совпадут с оценками по $N_1 < N$ наблюдениям, т.е. по регрессии $X_1 = Z_1\alpha_1 + \varepsilon_1$. Остатки будут иметь вид $\begin{pmatrix} e_1 \\ 0_{n_2} \end{pmatrix}$.[1]

[1] Это связано с тем, что добавочные переменные являются фиктивными переменными, каждая из которых всюду равна нулю, за исключением одного из наблюдений второй части выборки. Такие фиктивные переменные приводят к «обнулению» остатков. Следовательно, задача минимизации суммы квадратов остатков по N наблюдениям здесь сводится к задаче минимизации суммы квадратов остатков по первым N_1 наблюдениям.

Действительно, запишем модель в матричном виде. Пусть

$$Z = \begin{bmatrix} Z_1 & 0 \\ Z_2 & I \end{bmatrix}, \quad X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}.$$

Тогда оценки коэффициентов модели $\begin{pmatrix} a \\ d \end{pmatrix}$ удовлетворяют нормальным уравнениям:

$$Z'Z \begin{pmatrix} a \\ d \end{pmatrix} = Z'X,$$

или

$$\begin{pmatrix} Z_1'Z_1 + Z_2'Z_2 & Z_2' \\ Z_2 & I \end{pmatrix} \begin{pmatrix} a \\ d \end{pmatrix} = \begin{pmatrix} Z_1'X_1 + Z_2'X_2 \\ X_2 \end{pmatrix}.$$

Получаем следующую систему уравнений для оценок коэффициентов:

$$\begin{cases} Z_1'Z_1a + Z_2'Z_2a + Z_2'd = Z_1'X_1 + Z_2'X_2, \\ Z_2a + d = X_2. \end{cases} \tag{18.7}$$

Умножив второе уравнение системы (18.7) слева на Z_2', получим

$$Z_2'Z_2a + Z_2'd = Z_2'X_2.$$

После вычитания из первого уравнения системы (18.7) получим

$$Z_1'Z_1a = Z_1'X_1.$$

Таким образом, оценки коэффициентов a являются оценками МНК по первой части выборки.

Далее, остатки второй части равны

$$e_2 = X_2 - Z_2a - d = 0.$$

Последнее равенство следует из второго уравнения системы (18.7).

Суммы квадратов остатков в двух моделях равны $e_0'e_0$ и $e_1'e_1$, соответственно. Количество ограничений $k = N_2$. Таким образом, получаем следующую статистику:

$$F^c = \frac{(e_0'e_0 - e_1'e_1)/N_2}{e_1'e_1/(N_1 - n - 1)} \sim F_{N_2,\, N_1-n-1}.$$

Статистика имеет указанное распределение, если выполнена нулевая гипотеза

$$H_0 : \delta = 0.$$

Если нулевая гипотеза принимается, это означает, что модель не менялась.

Заметим, что в случае, когда наблюдений в одной из частей выборки не хватает, чтобы оценить параметры, либо их столько же, сколько параметров, например, $N_2 \leqslant n + 1$, второй тест Чоу можно рассматривать как распространение на этот вырожденный случай первого теста Чоу.

Второй тест Чоу можно интерпретировать также как тест на точность прогноза. Поскольку Z_2a — прогнозы, полученные для второй части выборки на основе оценок первой части (a), то из второго уравнения системы (18.7) следует, что оценки d равны ошибкам такого прогноза:

$$d = X_2 - Z_2a.$$

Таким образом, проверяя гипотезу $\delta = 0$, мы проверяем, насколько точны прогнозы. Если модель по второй части выборки отличается от модели по первой части, то ошибки прогноза будут большими и мы отклоним нулевую гипотезу.

18.3. Метод максимального правдоподобия в эконометрии

18.3.1. Оценки максимального правдоподобия

Метод максимального правдоподобия — это один из классических методов оценивания, получивший широкое распространение в эконометрии благодаря своей универсальности и концептуальной простоте.

Для получения оценок максимального правдоподобия следует записать функцию правдоподобия, а затем максимизировать ее по неизвестным параметрам модели. Предположим, что изучаемая переменная x имеет распределение с плотностью $f_x(x)$, причем эта плотность зависит от вектора неизвестных параметров θ, что можно записать как $f_x(x|\theta)$. Тогда для N независимых наблюдений за переменной x, т.е. $x_1, \ldots, x_N$, **функция правдоподобия**, по определению, есть

плотность их совместного распределения, рассматриваемая как функция от θ при данном наборе наблюдений $x_1, \ldots, x_N$:

$$L(\theta) = \prod_{i=1}^{N} f_x(x_i|\theta).$$

Если изучаемая переменная имеет дискретное распределение, то $f_x(x|\theta)$ следует понимать как вероятность, а не как плотность. Наряду с функцией $L(\theta)$ из соображений удобства рассматривают также ее логарифм, называемый логарифмической функцией правдоподобия.

Оценки максимального правдоподобия θ^* для параметров θ являются, по определению, аргмаксимумом функции правдоподобия (или, что то же самое, логарифмической функции правдоподобия). Они являются решением уравнения правдоподобия:

$$\frac{\partial \ln L}{\partial \theta} = 0.$$

В более общем случае нельзя считать наблюдения за изучаемой переменной, $x_1, \ldots, x_N$, независимыми и одинаково распределенными. В этом случае задается закон совместного распределения всех наблюдений, $f_x(x_1, \ldots, x_N|\theta) = f_x(x|\theta)$, и функция правдоподобия для данного вектора наблюдений x полагается равной $f_x(x|\theta)$.

Известно, что оценки максимального правдоподобия обладают свойствами состоятельности, асимптотической нормальности и асимптотической эффективности.

Оценку ковариационной матрицы оценок θ^* можно получить на основе матрицы вторых производных (матрицы Гессе) логарифмической функции правдоподобия:

$$\left(-\frac{\partial^2 \ln L(\theta^*)}{\partial \theta \partial \theta'} \right)^{-1}.$$

Другая классическая оценка ковариационной матрицы имеет вид

$$(I(\theta^*))^{-1},$$

где

$$I(\theta) = E\left(-\frac{\partial^2 \ln L(\theta)}{\partial \theta \partial \theta'} \right)$$

— так называемая информационная матрица.

18.3.2. Оценки максимального правдоподобия для модели линейной регрессии

Рассмотрим модель линейной регрессии $x_i = z_i\alpha + \varepsilon_i$, где вектор коэффициентов имеет размерность $n+1$, ошибки ε_i независимы и распределены нормально: $\varepsilon_i \sim N(0, \sigma^2)$, а факторы z_i являются детерминированными. При этом изучаемая переменная тоже имеет нормальное распределение: $x_i \sim N(z_i\alpha, \sigma^2)$. Плотность этого распределения равна

$$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x_i - z_i\alpha)^2}.$$

Перемножая плотности для всех наблюдений (с учетом их независимости), получим функцию правдоподобия:

$$L(\alpha, \sigma) = \frac{1}{(2\pi)^{N/2}\,\sigma^N}\, e^{-\frac{1}{2\sigma^2}\sum\limits_{i=1}^{N}(x_i - z_i\alpha)^2}.$$

Соответствующая логарифмическая функция правдоподобия равна

$$\ln L(\alpha; \sigma) = -\frac{N}{2}\ln(2\pi) - N\ln\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i - z_i\alpha)^2,$$

или в матричных обозначениях

$$\ln L(\alpha; \sigma) = -\frac{N}{2}\ln(2\pi) - N\ln\sigma - \frac{1}{2\sigma^2}(X - Z\alpha)'(X - Z\alpha).$$

Берем производные:

$$\frac{\partial \ln L}{\partial \alpha} = \frac{1}{\sigma^2} Z'(X - Z\alpha) = 0,$$
$$\frac{\partial \ln L}{\partial \sigma} = -\frac{N}{\sigma} + \frac{1}{\sigma^3}\Big((X - Z\alpha)'(X - Z\alpha)\Big) = 0.$$

Из первого уравнения получим оценки максимального правдоподобия для коэффициентов α:

$$a = (Z'Z)^{-1} Z'X.$$

Видим, что оценки наименьших квадратов и оценки максимального правдоподобия совпадают. Из второго уравнения, подставляя в него оценки a вместо α, получим оценку дисперсии σ^2:

$$s^2 = \frac{1}{N} e'e,$$

где $e'e = (X - Za)'(X - Za)$ — сумма квадратов остатков. Оценка максимального правдоподобия для дисперсии ошибки смещена. Несмещенная оценка, используемая в МНК, равна

$$\hat{s}^2 = \frac{1}{N - n - 1} e'e.$$

Тем не менее, оценки (a, s) асимптотически несмещены, состоятельны, асимптотически эффективны в классе любых оценок (а не только линейных, как при МНК).

Чтобы проверить, на самом ли деле мы нашли точку максимума правдоподобия, исследуем матрицу вторых производных:

$$\begin{aligned}
&\frac{\partial^2 \ln L}{\partial \alpha \partial \alpha'} = -\frac{1}{\sigma^2} Z'Z, \\
&\frac{\partial^2 \ln L}{\partial \sigma^2} = \frac{N}{\sigma^2} - \frac{3}{\sigma^4} (X - Z\alpha)' (X - Z\alpha), \\
&\frac{\partial^2 \ln L}{\partial \alpha \partial \sigma} = \left(\frac{\partial^2 \ln L}{\partial \sigma \partial \alpha'} \right)' = -\frac{2}{\sigma^3} Z' (X - Z\alpha).
\end{aligned}$$

Таким образом,

$$\frac{\partial^2 \ln L}{\partial (\alpha; \sigma) \partial (\alpha; \sigma)'} = -\begin{pmatrix} \frac{1}{\sigma^2} Z'Z & \frac{2}{\sigma^3} (X - Z\alpha)' Z \\ \frac{2}{\sigma^3} Z' (X - Z\alpha) & \frac{3}{\sigma^4} (X - Z\alpha)' (X - Z\alpha) - \frac{N}{\sigma^2} \end{pmatrix}.$$

Значение матрицы вторых производных в точке оценок (a, s) равно

$$\left. \frac{\partial^2 \ln L}{\partial (\alpha; \sigma) \partial (\alpha; \sigma)'} \right|_{a,s} = -\frac{N}{e'e} \begin{pmatrix} Z'Z & 0' \\ 0 & 2N \end{pmatrix}.$$

Видно, что матрица вторых производных отрицательно определена, то есть найдена точка максимума. Это дает оценку ковариационной матрицы оценок (a, s):

$$\frac{e'e}{N} \begin{pmatrix} (Z'Z)^{-1} & 0' \\ 0 & \frac{1}{2N} \end{pmatrix}.$$

Таким образом, оценка ковариационной матрицы для a является смещенной (поскольку основана на смещенной оценке дисперсии):

$$M_a = \frac{e'e}{N} (Z'Z)^{-1}.$$

В методе наименьших квадратов в качестве оценки берут

$$M_a = \frac{e'e}{N-n-1}\left(Z'Z\right)^{-1}.$$

При $N \to \infty$ эти две оценки сходятся.

Метод максимального правдоподобия дает также оценку дисперсии для s:

$$\mathbf{var}(s) = \frac{e'e}{2N^2}.$$

Рассчитаем также информационную матрицу. Для этого возьмем математическое ожидание от матрицы вторых производных со знаком минус:

$$I = \mathbf{E}\begin{pmatrix} \dfrac{1}{\sigma^2}Z'Z & \dfrac{2}{\sigma^3}\left(X-Z\alpha\right)'Z \\ \dfrac{2}{\sigma^3}Z'\left(X-Z\alpha\right) & \dfrac{3}{\sigma^4}\left(X-Z\alpha\right)'\left(X-Z\alpha\right)-\dfrac{N}{\sigma^2} \end{pmatrix} = \begin{pmatrix} \dfrac{1}{\sigma^2}Z'Z & 0' \\ 0 & \dfrac{2N}{\sigma^2} \end{pmatrix},$$

где мы воспользовались тем, что $X - Z\alpha$ представляет собой вектор ошибок модели ε и выполнено $\mathbf{E}\left(\varepsilon\right) = 0$, $\mathbf{E}\left(\varepsilon'\varepsilon\right) = N\sigma^2$. Обращая информационную матрицу в точке (a, s), получим ту же оценку ковариационной матрицы, что и раньше. Таким образом, оба метода дают одинаковый результат.

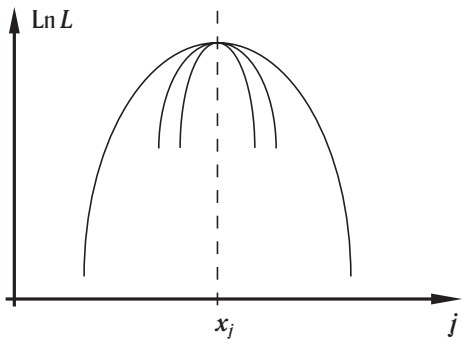

Рис. 18.1

Рассмотрим логарифмическую функцию правдоподобия как функцию одного из коэффициентов, α_j, при остальных коэффициентах зафиксированных на уровне оценок максимального правдоподобия, т.е. срез $(n+2)$-мерного пространства (см. рис. 18.1). Видим, что оценка a_j тем точнее, чем острее пик функции правдоподобия. А степень остроты пика показывает вторая производная (по абсолютному значению). Поэтому математическое ожидание матрицы вторых производных со знаком минус называется информационной матрицей. Эта матрица удовлетворяет естественным требованиям: чем больше имеем информации, тем точнее оценка.

Если в логарифмическую функцию правдоподобия $\ln L\left(\alpha;\sigma\right)$ подставить оценку s^2 для σ^2, которая найдена из условия $\partial \ln L/\partial\sigma = 0$:

$$s^2 = \frac{e'e}{N},$$

то получится так называемая концентрированная функция правдоподобия, которая зависит уже только от α:

$$\ln L^c(\alpha) = -\frac{N}{2}\ln(2\pi) - \frac{N}{2}\ln\left(\frac{1}{N}e'e\right) - \frac{N}{2}.$$

Очевидно, что максимизация концентрированной функции правдоподобия эквивалентна методу наименьших квадратов (минимизации суммы квадратов остатков).

18.3.3. Три классических теста для метода максимального правдоподобия

Рассмотрим линейную регрессию с нормальными ошибками. Требуется проверить гипотезу о том, что коэффициенты этой регрессии удовлетворяют некоторым линейным ограничениям. Пусть a_0 — оценки, полученные методом максимального правдоподобия без учета ограничений, а a_1 — оценки, полученные тем же методом с учетом ограничений, и пусть $\ln L_0$ — значение логарифмической функции правдоподобия в точке a_0, а $\ln L_1$ — значение логарифмической функции правдоподобия в точке a_1. Статистику для проверки такой гипотезы естественно строить как показатель, измеряющий существенность различий между двумя моделями — с ограничениями и без них. Если различия не очень велики (ограничения существенны), то гипотезу о том, что ограничения выполнены, следует принять, а если достаточно велики — то отвергнуть. Рассмотрим три возможных способа измерения этих различий, проиллюстрировав их графически.

Критерий отношения правдоподобия (*Likelihood ratio test* — LR) основан на различии значений логарифмической функции правдоподобия в точках a_0 и a_1 (см. рис. 18.2), или, что то же самое, на логарифме отношения правдоподобия, т.е. величине

$$\ln L_0 - \ln L_1 = \ln\left(\frac{L_0}{L_1}\right).$$

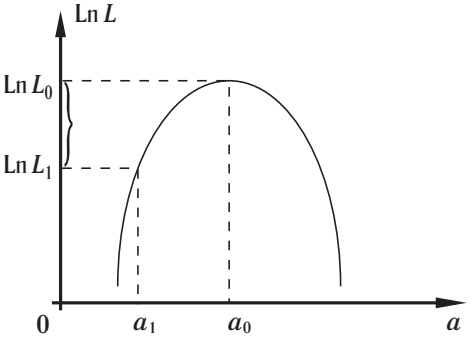

Рис. 18.2

Критерий множителей Лагранжа (*Lagrange multiplier test* — LM) основан на различии тангенса угла наклона касательной к логарифмической функции правдоподобия в точках a_0 и a_1. Поскольку в точке a_0 он равен нулю, то следует рассмотреть, насколько тангенс угла наклона касательной в точке a_1 отличен от нуля (см. рис. 18.3).

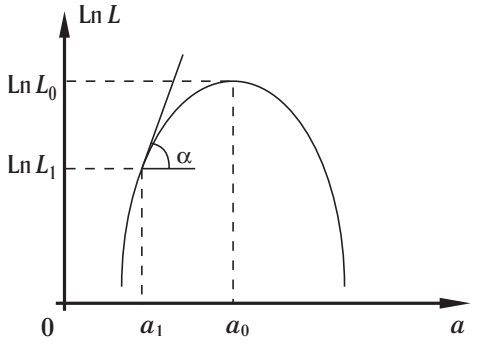

Рис. 18.3

Критерий Вальда (Wald test — W) основан на невязках рассматриваемых ограничений. В точке a_1, по определению, невязки равны нулю. Таким образом, следует рассмотреть, насколько невязки в точке a_0 отличны от нуля. В случае одного параметра точка a_1 однозначно задается ограничениями, и невязка в точке a_0 при линейных ограничениях будет некоторой линейной функцией разности оценок a_0 и a_1 (см. рис. 18.4).

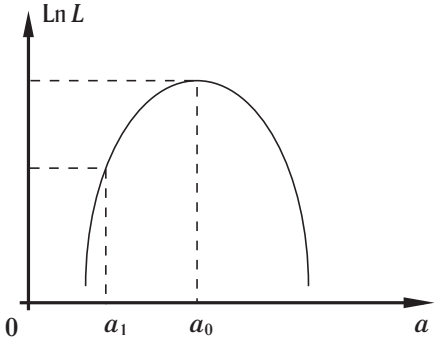

Рис. 18.4

Покажем, как соответствующие критерии выводятся в рассматриваемом нами случае линейной регрессии с нормальными ошибками, когда требуется проверить линейные ограничения на коэффициенты. (В общем случае построение критериев происходит аналогичным образом.) При выводе критериев нам понадобится следующая лемма (см. Приложение А.3.2).

Лемма: Пусть χ — вектор $(\chi \in R^k)$ случайных величин, подчиненных многомерному нормальному распределению: $\chi \sim N\left(0, \sigma^2\Omega\right)$, где матрица Ω неособенная. Тогда

$$\frac{1}{\sigma^2}\chi'\Omega^{-1}\chi \sim \chi^2_k.$$

Доказательство:

Так как Ω положительно определена (см. Приложения А.1.2 и А.1.2), то существует неособенная квадратная матрица C, такая, что $\Omega^{-1} = CC'$. Рассмотрим вектор $\frac{1}{\sigma}C\chi$. Ясно, что $\mathbf{E}\left(\frac{1}{\sigma}C\chi\right) = 0$, а ковариационная матрица этого вектора равна

$$\frac{1}{\sigma^2}E\left(C\chi\chi'C'\right) = C\Omega C' = I_k.$$

Таким образом, вектор $\frac{1}{\sigma}C\chi$ состоит из k некоррелированных и, как следствие (по свойству многомерного нормального распределения), независимых случайных

величин, имеющих стандартное нормальное распределение. Тогда (по определению распределения χ-квадрат) сумма квадратов вектора $\frac{1}{\sigma}C\chi$ распределена как χ^2_k.

Тест Вальда (W-тест)

Для оценки коэффициентов регрессии без ограничений выполнено

$$a_0 = (Z'Z)^{-1} Z'X \sim N\left(\alpha, \sigma^2 (Z'Z)^{-1}\right).$$

Рассмотрим невязки ограничений $Ra_0 - r$. Чем они больше, тем более правдоподобно, что ограничения не выполнены. Ясно, что (см. Приложение А.3.2)

$$Ra_0 - r \sim N\left(R\alpha - r; \sigma^2 A\right),$$

где, как и раньше, используется обозначение $A = R(Z'Z)^{-1} R'$. Матрица A имеет размерность $k \times k$, где k — количество ограничений. Пусть выполнена нулевая гипотеза

$$H_0: \quad R\alpha = r.$$

Тогда $Ra_0 - r \sim N(0; \sigma^2 A)$. По лемме

$$\frac{1}{\sigma^2} (Ra_0 - r)' A^{-1} (Ra_0 - r) \sim \chi^2_k.$$

Поскольку известны лишь a_0 — оценки без ограничений, то в качестве оценки неизвестной величины σ^2 берем $\frac{1}{N} e_0' e_0$, где $e_0 = X - Za_0$ — остатки из модели без ограничений. Отсюда получаем статистику Вальда:

$$W = \frac{N}{e_0' e_0} (Ra_0 - r)' \left(R (Z'Z)^{-1} R'\right)^{-1} (Ra_0 - r).$$

Эта статистика распределена примерно как χ^2_k. Тогда, если $W < \chi^2_{k,\gamma}$, то следует принять H_0, что ограничения выполнены. При $W > \chi^2_{k,\gamma}$ ограничения существенны и следует отвергнуть H_0.

Можно увидеть, что статистика Вальда имеет следующую структуру:

$$W = (Ra_0 - r)' (RM_{a_0} R')^{-1} (Ra_0 - r),$$

где $M_{a_0} = \frac{e_0' e_0}{N} (Z'Z)^{-1}$ — оценка ковариационной матрицы оценок a_0. Фактически это общая формула для статистики Вальда, применимая в случае произвольной модели, а не только линейной регрессии с нормальными ошибками.

Тест отношения правдоподобия (LR-тест)

Рассмотрим статистику $LR = -2\left(\ln L_1 - \ln L_0\right) = -2\ln\dfrac{L_1}{L_0}$, называемую статистикой отношения правдоподобия. Здесь L_1 и L_0 — значения логарифмической функции правдоподобия в точках a_0 и a_1:

$$\ln L_0 = -\frac{N}{2}\left(1 + \ln 2\pi\right) - \frac{N}{2}\ln\left(\frac{e_0' e_0}{N}\right),$$

$$\ln L_1 = -\frac{N}{2}\left(1 + \ln 2\pi\right) - \frac{N}{2}\ln\left(\frac{e_1' e_1}{N}\right).$$

Суммы квадратов остатков здесь равны

$$e_0' e_0 = \left(X - Za_0\right)'\left(X - Za_0\right)$$

и

$$\begin{aligned} e_1' e_1 = \left(X - Za_1\right)'\left(X - Za_1\right) = \\ = \left(X - Za_0\right)'\left(X - Za_0\right) + \left(Ra_0 - r\right)' A^{-1}\left(Ra_0 - r\right). \end{aligned}$$

Покажем, что если верна нулевая гипотеза $R\alpha = r$, то приближенно выполнено $-2\ln\left(L_1/L_0\right) \sim \chi^2_k$.

Действительно,

$$-2\ln\left(\frac{L_1}{L_0}\right) = N\ln\left(\frac{e_1' e_1}{e_0' e_0}\right) = N\ln\left[1 + \frac{\left(Ra_0 - r\right)' A^{-1}\left(Ra_0 - r\right)}{\left(X - Za_0\right)'\left(X - Za_0\right)}\right].$$

Для натурального логарифма при малых x выполнено $\ln\left(1 + x\right) \approx x$. Рассмотрим последнюю дробь. При большом количестве наблюдений оценки a_0 стремятся к вектору α, для которого выполнено $H_0 : R\alpha = r$. Отсюда следует, что при большом количестве наблюдений дробь — малая величина, и получаем приближенно

$$LR = -2\ln\left(\frac{L_1}{L_0}\right) \approx N\frac{\left(Ra_0 - r\right)' A^{-1}\left(Ra_0 - r\right)}{\left(X - Za_0\right)'\left(X - Za_0\right)} = W.$$

Таким образом, статистика отношения правдоподобия приближенно равна статистике Вальда, которая приближенно распределена как χ^2_k. Получили LR-тест: если $LR > \chi^2_{k,\gamma}$, то H_0 неверна, ограничения не выполнены, а если $LR < \chi^2_{k,\gamma}$, то наоборот.

Тест множителей Лагранжа (LM-тест)

Ранее мы получили выражение для множителей Лагранжа, соответствующих ограничению $R\alpha = r$:

$$\lambda = A^{-1}\left(Ra_0 - r\right).$$

Из того, что $Ra_0 - r \sim N\left(R\alpha - r; \sigma^2 A\right)$, следует, что $\lambda \sim N\left(A^{-1}(R\alpha - r); \sigma^2 A^{-1}\right)$.

Отсюда при $H_0 : R\alpha = r$ выполнено $\lambda \sim N\left(0; \sigma^2 A^{-1}\right)$, поэтому в силу леммы имеем $\dfrac{1}{\sigma^2}\lambda' A\lambda \sim \chi^2_k$. Поскольку известны только оценки с ограничением, a_1, то в качестве оценки σ^2 берем $\dfrac{1}{N}e_1'e_1$.

Получили статистику

$$LM = \frac{N}{e_1'e_1}\lambda' A\lambda = \frac{N}{e_1'e_1}\lambda' R\left(Z'Z\right)^{-1}R'\lambda.$$

Если $LM > \chi^2_{k,\gamma}$, то H_0 отвергается, ограничения не выполнены. Если $LM < \chi^2_{k,\gamma}$, то H_0 принимается.

Вспомним, что из нормальных уравнений для оценок при ограничениях

$$R'\lambda = Z'(X - Za_1).$$

В то же время

$$\frac{\partial \ln L(a_1, \sqrt{e_1'e_1/N})}{\partial \alpha} = \frac{N}{e_1'e_1}Z'\left(X - Za_1\right) \quad —$$

производная логарифмической функции правдоподобия (это функция без учета ограничений) по параметрам в точке оценок при ограничениях a_1 и $s_1 = \sqrt{\dfrac{e_1'e_1}{N}}$.

Статистика множителей Лагранжа, таким образом, имеет следующую структуру:

$$LM = \frac{e_1'e_1}{N}\frac{\partial \ln L(a_1, \sqrt{e_1'e_1/N})}{\partial \alpha'}\left(Z'Z\right)^{-1}\frac{\partial \ln L(a_1, \sqrt{e_1'e_1/N})}{\partial \alpha} =$$
$$= \frac{\partial \ln L(a_1, \sqrt{e_1'e_1/N})}{\partial \alpha'}M_{a_0}\left(a_1\right)\frac{\partial \ln L(a_1, \sqrt{e_1'e_1/N})}{\partial \alpha},$$

где $M_{a_0}\left(a_1\right) = \dfrac{e_1'e_1}{N}\left(Z'Z\right)^{-1}$ — оценка ковариационной матрицы оценок a_0, вычисленная на основе информации, доступной в точке a_1. Это общая формула для статистики множителей Лагранжа, применимая в случае произвольной модели, а не только линейной регрессии с нормальными ошибками. В таком виде тест называется скор-тестом (score test) или тестом Рао.

18.3.4. Сопоставление классических тестов

Величину $(Ra_0 - r)' \left(R\left(Z'Z\right)^{-1} R'\right)^{-1} (Ra_0 - r)$, которая фигурирует в формулах для рассматриваемых статистик, можно записать также в виде $e_1'e_1 - e_0'e_0$. Таким образом, получаем следующие формулы для трех статистик через суммы квадратов остатков:

$$W = N\frac{e_1'e_1 - e_0'e_0}{e_0'e_0},$$
$$LM = N\frac{e_1'e_1 - e_0'e_0}{e_1'e_1},$$
$$LR = N\ln\left(\frac{e_1'e_1}{e_0'e_0}\right).$$

F-статистику для проверки линейных ограничений можно записать аналогичным образом:

$$F = \frac{N-n-1}{k}\frac{e_1'e_1 - e_0'e_0}{e_0'e_0}.$$

Нетрудно увидеть, что все три статистики можно записать через F-статистику:

$$LR = N\ln\left(1 + \frac{k}{N-n-1}F\right),$$
$$W = \frac{N}{N-n-1}kF,$$
$$LM = \frac{N}{kF+N-n-1}kF.$$

Заметим, что по свойству F-распределения kF в пределе при $N \to \infty$ сходится к χ_k^2, чем можно доказать сходимость распределения всех трех статистик к этому распределению.

Так как $e_1'e_1 \geqslant e_0'e_0$, то $W \geqslant LM$. Следовательно, тест Вальда более жесткий, он чаще отвергает ограничения. Статистика отношения правдоподобия лежит всегда между W и LM. Чтобы это показать, обозначим

$$x = \frac{k}{N-n-1}F = \frac{e_1'e_1 - e_0'e_0}{e_0'e_0}.$$

Доказываемое свойство следует из того, что при $x > -1$ выполнено неравенство $\dfrac{x}{1+x} \leqslant \ln(1+x) \leqslant x$.

18.4. Упражнения и задачи

Упражнение 1

В Таблице 18.1 приведены данные о продаже лыж в США: SA — продажа лыж в США, млн. долл., PDI — личный располагаемый доход, млрд. долл.

1.1. Оценить регрессию SA по константе и PDI. Построить и проанализировать автокорреляционную функцию остатков.

1.2. Оценить регрессию с добавлением квартальных сезонных переменных Q_1, Q_2, Q_3, Q_4. Константу не включать (почему?). Оценить ту же регрессию, заменив Q_4 на константу. Построить и проанализировать автокорреляционную функцию остатков в регрессии с сезонными переменными.

1.3. Проверить гипотезу о том, что коэффициенты при сезонных переменных равны одновременно нулю. Есть ли сезонная составляющая в данных?

Таблица 18.1. (Источник: Chatterjee, Price, Regression Analysis by Example, 1991, p.138)

Квартал	SA	PDI	Квартал	SA	PDI	Квартал	SA	PDI
1965.1	37.4	118	1968.1	44.2	143	1971.1	52	180
1965.2	31.6	120	1968.2	40.4	147	1971.2	46.2	184
1965.3	34	122	1968.3	38.4	148	1971.3	47.1	187
1965.4	38.1	124	1968.4	45.4	151	1971.4	52.7	189
1966.1	40	126	1969.1	44.9	153	1972.1	52.2	191
1966.2	35	128	1969.2	41.6	156	1972.2	47	193
1966.3	34.9	130	1969.3	44	160	1972.3	47.8	194
1966.4	40.2	132	1969.4	48.1	163	1972.4	52.8	196
1967.1	41.9	133	1970.1	49.7	166	1973.1	54.1	199
1967.2	34.7	135	1970.2	43.9	171	1973.2	49.5	201
1967.3	38.8	138	1970.3	41.6	174	1973.3	49.5	202
1967.4	43.7	140	1970.4	51	175	1973.4	54.3	204

Таблица 18.2. (Источник: M.Pokorny, An Introduction to Econometrics. Basil Blackwell, 1987, p.230)

Год	X	L	K	D	Год	X	L	K	D
1965	190.6	565	4.1	0.413	1973	130.2	315	4.2	0.091
1966	177.4	518	4.3	0.118	1974	109.3	300	4.2	5.628
1967	174.9	496	4.3	0.108	1975	127.8	303	4.2	0.056
1968	166.7	446	4.3	0.057	1976	122.2	297	4.3	0.078
1969	153	407	4.3	1.041	1977	120.6	299	4.4	0.097
1970	144.6	382	4.3	1.092	1978	121.7	295	4.6	0.201
1971	147.1	368	4.3	0.065	1979	120.7	288	4.9	0.128
1972	119.5	330	4.3	10.8	1980	128.2	286	5.2	0.166

1.4. Оценить ту же регрессию, считая, что коэффициент при Q_1 равен коэффициенту при Q_4 («зимние» кварталы) и коэффициент при Q_2 равен коэффициенту при Q_3 («летние» кварталы).

1.5. Мы предполагали выше, что константа меняется в зависимости от квартала. Теперь предположим, что в коэффициенте при PDI также имеется сезонность. Создать необходимые переменные и включить их в регрессию. Меняется ли коэффициент при PDI в зависимости от квартала? Проверить соответствующую гипотезу.

Упражнение 2

В Таблице 18.2 приведены данные о добыче угля в Великобритании: X — общая добыча угля (млн. тонн), L — общая занятость в добыче угля (тыс. чел.), K — основные фонды в угледобывающей отрасли (восстановительная стоимость в ценах 1975 г., млн. фунтов), D — потери рабочих дней в угледобывающей отрасли из-за забастовок (млн. дней).

2.1. Оценить уравнение регрессии $X = \text{const} + bK + cD$. На основе графиков остатков по времени и по расчетным значениям сделать выводы относительно гетероскедастичности, автокорреляции и функциональной формы.

2.2. Провести вручную с помощью соответствующих регрессий тесты: а) на гетероскедастичность, б) тест Годфрея на автокорреляцию остатков, в) тест Рамсея на функциональную форму.

2.3. Провести Чоу-тест (тест на постоянство коэффициентов регрессии) с помощью умножения на фиктивную переменную. Данные разбить на 2 части: с 1965 по 1972 и с 1973 по 1980 гг. Сделать выводы.

2.4. Провести тест на добавление фактора L. Оценить уравнение регрессии $X = \text{const} + bK + cD + aL$.

2.5. Правильно ли выбрана функциональная форма регрессии? В случае, если она выбрана неправильно, попробовать исправить ее путем добавления квадратов переменных K и L.

Упражнение 3

В Таблице 15.3 на стр. 520 приведены данные о совокупном доходе и потреблении в США в 1953–1984 гг.

3.1. Оценить потребительскую функцию (в логарифмах) $\ln C_t = \beta + \alpha \ln Y_t + \varepsilon_t$.

3.2. Проверить гипотезу $\alpha = 1$ с помощью: а) теста Вальда, б) преобразованной модели $\ln Y_t = \beta + (\alpha - 1) \ln Y_t + \varepsilon_t$, в) теста отношения правдоподобия.

3.3. Проверить гипотезу об отсутствии автокорреляции первого порядка.

3.4. Предположим, что ошибка подчинена авторегрессионному процессу первого порядка $\varepsilon_t = \rho \varepsilon_{t-1} + u_t$. Оценить соответствующую модель.

3.5. Модель с авторегрессией в ошибке можно записать в следующем виде:

$$\ln C_t = \beta(1 - \rho) + \alpha \ln Y_t - \alpha\rho \ln Y_{t-1} + \rho \ln C_{t-1} + u_t.$$

Эта же нелинейная модель представляется в виде линейной регрессии:

$$\ln C_t = \beta' + \alpha' \ln Y_t + \gamma' \ln Y_{t-1} + \delta' \ln C_{t-1} + u_t,$$

где $\alpha' = \alpha$, $\beta' = \beta(1 - \rho)$, $\gamma' = -\alpha\rho$, $\delta' = \rho$. Оцените модель как линейную регрессию.

3.6. Для той же нелинейной модели должно выполняться соотношение между коэффициентами линейной регрессии: $\alpha'\delta' = -\gamma'$. Проверить данную гипотезу.

3.7. Оценить нелинейную регрессию. Использовать тест отношения правдоподобия для проверки той же гипотезы, т.е. $\alpha'\delta' = -\gamma'$.

Задачи

1. Оценивается функция Кобба—Дугласа (в логарифмическом виде) с ограничением однородности первой степени. Запишите матрицы (R и r) ограничений на параметры регрессии.

2. Запишите матрицы (R и r) ограничений на параметры регрессии в случае проверки того, что 1-й и 3-й коэффициенты регрессии $X = Z\alpha + \varepsilon$ совпадают, где матрица наблюдений

$$Z = \begin{pmatrix} 1 & 1 & 1 & 3 \\ 1 & 0 & 2 & 1 \\ 1 & 3 & 3 & -1 \\ 1 & 0 & 4 & 3 \\ 1 & 2 & 5 & 2 \end{pmatrix}.$$

3. Запишите матрицы (R и r) ограничений на параметры регрессии в случае проверки значимости j-го коэффициента регрессии.

4. Запишите матрицы (R и r) ограничений на параметры регрессии в случае проверки значимости уравнения регрессии в целом.

5. Исходные данные для модели линейной регрессии $x = \alpha_1 z_1 + \alpha_2 z_2 + \varepsilon$ неизвестны, известно только, что количество наблюдений $N = 100$, сумма квадратов остатков равна 196,

$$Z'Z = \begin{pmatrix} 2 & -3 \\ -3 & 5 \end{pmatrix} \quad \text{и} \quad Z'X = \begin{pmatrix} 3 \\ 6 \end{pmatrix}.$$

   а) Используя эту информацию, рассчитайте оценки МНК.

   б) Рассчитайте оценки МНК, учитывая ограничение $2\alpha_1 - 3\alpha_2 = 10$. Найдите сумму квадратов остатков.

   в) Рассчитайте статистику, с помощью которой можно проверить гипотезу $2\alpha_1 - 3\alpha_2 = 10$. Какое распределение имеет эта статистика?

6. Регрессию $x_i = a_0 + a_1 z_{i1} + a_2 z_{i2} + e_i$ оценили без ограничений на параметры и получили остатки $(3,\ -2,\ -4,\ 3)$, а затем оценили с ограничением $a_1 + a_2 = 1$ и получили остатки $(2,\ -1,\ -4,\ 3)$. Найдите F-статистику для

проверки ограничений. С чем ее следует сравнить? В каком случае гипотеза принимается?

7. В регрессии с одним фактором и свободным членом остатки равны $(1; -1; 0; -1; 1; 0)$. Если не включать в регрессию свободный член, то остатки равны $(1; -2; 1; -1; 1; 0)$. Проверьте гипотезу о том, что свободный член равен нулю, если 5%-ные границы F-распределения равны $F_{1,1} = 161.5$; $F_{1,2} = 18.51$; $F_{1,3} = 10.13$; $F_{1,4} = 7.71$; $F_{1,5} = 6.61$.

8. Как с помощью критерия Стьюдента проверить автокорреляцию первого порядка в остатках регрессии $X = Z_1\alpha_1 + \varepsilon$?

9. Пусть в простой линейной регрессии остатки равны $(0; 2; -2; 1; -2; 1)$. После добавления в исходную регрессию лага остатков $(0; 0; 2; -2; 1; -2)$ текущие остатки оказались равны $(0; 0; 0; 1; -1; 0)$. Проверить гипотезу об отсутствии автокорреляции ошибок, если 5%-ные границы F-распределения равны $F_{1,1} = 161.5$, $F_{1,2} = 18.51$, $F_{1,3} = 10.13$, $F_{1,4} = 7.71$, $F_{1,5} = 6.61$.

10. С помощью какой регрессии можно проверить правильность функциональной формы уравнения регрессии?

11. Уравнение регрессии с двумя факторами и константой оценено по временным рядам длиной 10. Сумма квадратов остатков, полученная в регрессии по всем наблюдениям, равна 100, сумма квадратов остатков, полученная в регрессии по первым 5-ти наблюдениям, равна 40, а сумма квадратов остатков, полученная в регрессии по последним 5-ти наблюдениям, равна 20. Найдите F-статистики для гипотезы о постоянстве коэффициентов регрессии. С чем ее следует сравнить? В каком случае гипотеза принимается?

12. В исходной регрессии было 12 наблюдений, 2 фактора и константа. Сумма квадратов остатков была равна 120. Затем выборку разбили на две части, в первой из которых 6 наблюдений. В регрессии по первой части выборки сумма квадратов оказалась равной 25, а по второй части 15. Проверить гипотезу о постоянстве коэффициентов в регрессии, если 5%-ные границы F-распределения равны: $F_{2,1} = 199.5$, $F_{2,2} = 19$, $F_{2,3} = 9.55$, $F_{2,4} = 6.94$, $F_{2,5} = 5.79$, $F_{2,10} = 4.10$, $F_{3,1} = 215.7$, $F_{3,2} = 19.16$, $F_{3,3} = 9.28$, $F_{3,4} = 6.59$, $F_{3,5} = 5.41$, $F_{3,10} = 3.71$.

13. Тест Чоу применили к регрессии, разбив выборку на N_1 и N_2 наблюдений. Количество факторов равно n. Приведите условия на N_1, N_2 и n, при которых невозможно использовать первую форму теста Чоу.

14. Тест Чоу применили к регрессии, разбив выборку на N_1 и N_2 наблюдений. Количество факторов равно n. Приведите условия на N_1, N_2 и n, при которых невозможно использовать вторую форму теста Чоу.

15. Для чего можно использовать информационную матрицу в методе максимального правдоподобия?

16. В регрессии $X = Z\alpha + \varepsilon$ матрица

$$Z = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix},$$

а остатки ε равны $(1,\ 1,\ 2,\ -2)'$. Запишите информационную матрицу.

17. В регрессии $X = Z\alpha + \varepsilon$ матрица

$$Z = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 1 \\ 1 & 2 \end{pmatrix},$$

а оценка дисперсии, найденная методом максимального правдоподобия, равна $\frac{5}{8}$. Запишите информационную матрицу.

18. Регрессию $x = a_0 + a_1 z_1 + a_2 z_2 + e$ оценили без ограничений на параметры и получили остатки $(0,\ -1,\ 1,\ 0)'$, а затем оценили с ограничением $a_1 + a_2 = 1$ и получили остатки $(2,\ -1,\ -4,\ 3)'$. Найдите статистику отношения правдоподобия для проверки ограничений. С чем ее следует сравнить? В каком случае гипотеза принимается?

19. Регрессию $x = a_0 + a_1 z_1 + a_2 z_2 + e$ оценили без ограничений на параметры и получили остатки $(2,\ -1,\ -4,\ 3)'$, а затем оценили с ограничением $a_1 + a_2 = 1$ и получили остатки $(3,\ -2,\ -4,\ 3)'$. Найдите статистику множителя Лагранжа для проверки ограничений. С чем ее следует сравнить? В каком случае гипотеза принимается?

20. Регрессию $x = a_0 + a_1 z_1 + a_2 z_2 + e$ оценили без ограничений на параметры и получили остатки $(0,\ -1,\ -1,\ 2)$, а затем оценили с ограничением $a_1 + a_2 = 1$ и получили остатки $(1,\ -2,\ -2,\ 3)$. Найдите статистику Вальда для проверки ограничений. С чем ее следует сравнить? В каком случае гипотеза принимается?

21. В модели линейной регрессии $x = \alpha_1 z_1 + \alpha_2 z_2 + e$ по некоторому набору данных ($N = 100$ наблюдений) получены следующие оценки МНК: $a = (0.4, -0.7)'$. Оценка ковариационной матрицы этих оценок равна

$$M_a = \begin{pmatrix} 0.01 & -0.02 \\ -0.02 & 0.08 \end{pmatrix}.$$

Используя общую формулу для статистики Вальда, проверьте следующие гипотезы на уровне $5\,\%$:

а) $H_0\colon\ \alpha_1 = 0.5,\ \alpha_2 = -0.5$,

б) $H_0\colon\ \alpha_1 - \alpha_2 = 1$,

в) $H_0\colon\ 3\alpha_1 + \alpha_2 = 0$.

22. В регрессии $x = a_0 + a_1 z_1 + a_2 z_2 + a_3 z_3 + a_4 z_4 + e$ по 40 наблюдениям с помощью теста Вальда проверяют гипотезы $a_1 = a_4 + 1,\ a_3 + a_2 = 1$. Как распределена статистика W (Вальда)? В каком случае гипотеза принимается?

23. Методом наименьших квадратов была оценена производственная функция:

$$\ln Y = 1.5 + \underset{(0.3)}{0.6}\ln K + \underset{(0.2)}{0.45}\ln L,$$

где Y — объем производства, K — капитал, L — труд. В скобках указаны стандартные ошибки коэффициентов. Ковариация оценок коэффициентов при $\ln K$ и $\ln L$ равна 0.05. Коэффициент детерминации равен $R^2 = 0.9$.

Проверьте следующие гипотезы:

а) как труд, так и капитал не влияют на объем производства;

б) эластичности объема производства по труду и капиталу совпадают;

в) производственная функция характеризуется постоянной отдачей от масштаба (сумма эластичностей равна единице).

24. При каких условиях можно применить критерии Вальда (W), отношения правдоподобия (LR), множителей Лагранжа (LM)?

а) без ограничений;

б) известны оценки параметров при ограничениях;

в) и те и другие оценки.

Для каждого из пунктов (а), (б) и (в) указажите имена тестов, которые можно применить.

Рекомендуемая литература

1. **Магнус Я.Р., Катышев П.К., Пересецкий А.А.** Эконометрика — начальный курс. — М.: «Дело», 2000. (Гл. 3, 11).

2. **Себер Дж.** Линейный регрессионый анализ. — М.: «Мир», 1980.

3. Статистические методы в экспериментальной физике. — М: Атомиздат, 1976. (Гл. 5, 8–10).

4. **Цыплаков А.А.** Некоторые эконометрические методы. Метод максимального правдоподобия в эконометрии. — Новосибирск: НГУ, 1997.

5. **Baltagi, Badi H.** Econometrics, 2nd edition, Springer, 1999. (Ch. 7).

6. **Davidson, R., and J.G. MacKinnon.** Estimation and Inference in Econometrics. Oxford University Press, 1993. (Ch. 1, 3, 8, 13).

7. **Engle R.** Wald, Likelihood Ratio and Lagrange Multiplier Tests in Econometrics, in Handbook of Econometrics, vol. II, Amsterdam: North Holland, 1984.

8. **Greene W.H.** Econometric Analysis, Prentice-Hall, 2000. (Ch. 4, 7).

9. **Judge G.G., Griffiths W.E., Hill R.C., Luthepohl H., Lee T.** Theory and Practice of Econometrics. — New York: John Wiley & Sons, 1985. (Ch. 2, 5).

10. **Ruud Paul A.** An Introduction to Classical Econometric Theory, Oxford University Press, 2000. (Ch. 4, 11, 17).

Глава 19

Байесовская регрессия

Прежде чем переходить к регрессии, полезно напомнить, в чем заключается байесовский подход. Он основан на теореме Байеса (см. Приложение А.3.1)

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}, \tag{19.1}$$

которая следует из определения вероятности совместного события:

$$p(A \cap B) = p(A|B)p(B) = p(B|A)p(A).$$

Пусть теперь

M_i, $i = 1, \ldots, k$ — гипотезы, модели, теории, суждения об изучаемом предмете; они являются взаимоисключающими и образуют исчерпывающее множество возможных объяснений изучаемого феномена;

$p(M_i)$ — априорные (доопытные, субъективные) вероятности, выражающие совокупность априорных (доопытных, субъективных) знаний об изучаемом предмете; $\sum p(M_i) = 1$;

D — результат наблюдения, опыта;

$p(D|M_i)$ — правдоподобия, вероятности того, насколько правдоподобен результат, если правильна i-я теория изучаемого предмета, считаются известными.

Тогда в соответствии с (19.1) записывается следующее соотношение:

$$p(M_i|D) = \frac{p(D|M_i)p(M_i)}{p(D)}, \tag{19.2}$$

где $p(D) = \sum p(D|M_i)p(M_i)$,

$p(M_i|D)$ — апостериорные (послеопытные) вероятности.

Это соотношение показывает, как априорные знания о предмете меняются в результате получения опытных данных, т.е. как накапливаются знания.

Пример трансформации представлений преподавателя об уровне знаний студента.

M_1 — студент знает предмет,

M_2 — студент не знает предмет.

Преподаватель имеет априорные оценки вероятностей этих состояний:

$p(M_1) = 0.2$,

$p(M_2) = 0.8$.

Наблюдение, опыт — в данном случае это экзамен. Результат опыта:

D_1 — студент сдал экзамен,

D_2 — студент не сдал экзамен.

Правдоподобия преподавателя:

$p(D_1|M_1) = 0.9$

$p(D_2|M_1) = 0.1$

$p(D_1|M_2) = 0.4$

$p(D_2|M_1) = 0.6$

Пусть студент сдал экзамен. Тогда априорные оценки преподавателя корректируются следующим образом:

$$p\left(M_1|D_1\right) = \frac{0.9 \cdot 0.2}{0.9 \cdot 0.2 + 0.4 \cdot 0.8} = 0.36,$$

$$p\left(M_2|D_1\right) = \frac{0.4 \cdot 0.8}{0.9 \cdot 0.2 + 0.4 \cdot 0.8} = 0.64\,.$$

Если студент не сдал экзамен, то апостериорные вероятности будут такими:

$$p\left(M_1|D_2\right) = 0.04\,,$$
$$p\left(M_2|D_2\right) = 0.96\,.$$

19.1. Оценка параметров байесовской регрессии

Для уравнения регрессии

$$X = Z\alpha + \varepsilon$$

имеются априорные представления об α и σ, которые выражаются плотностью вероятности совместного распределения (α, σ).

После эксперимента, результатами которого является выборка в виде вектора X и матрицы Z, эти представления корректируются. Аналогом (19.2) в данном случае выступает следующее выражение:

$$p(\alpha, \sigma | X, Z) = \frac{L(X, Z|\alpha, \sigma)\, p(\alpha, \sigma)}{p(X, Z)}, \tag{19.3}$$

где

$$p(X, Z) = \int\limits_{\alpha,\sigma} L(X, Z|\alpha, \sigma)\, d\alpha\, d\sigma.$$

Поскольку Z не зависит от α и σ, его можно «вынести за скобки»:

$$L(X, Z|\alpha, \sigma) = P_N(X|Z\alpha, \sigma^2 I) p(Z),$$
$$p(X, Z) = p(X|Z)p(Z),$$

и записать (19.3) в следующем виде:

$$p(\alpha, \sigma | X, Z) = \frac{P_N(X|Z\alpha, \sigma^2 I) p(\alpha, \sigma)}{p(X|Z)}.$$

Поскольку $p(X|Z)$ не зависит от α и σ, эту формулу можно записать, используя знак $\propto$, который выражает отношение «пропорционально», «равно с точностью до константы»:

$$p(\alpha, \sigma | X, Z) \propto P_N(X|Z\alpha, \sigma^2 I) p(\alpha, \sigma). \tag{19.4}$$

Пусть выполнены все гипотезы основной модели линейной регрессии, включая гипотезу о нормальности. Тогда

$$P_N(X|Z\alpha, \sigma^2 I) \propto \sigma^{-N} e^{-\frac{1}{2\sigma^2}(X - Z\alpha)'(X - Z\alpha)} =$$
$$= \sigma^{-N} e^{-\frac{1}{2\sigma^2}[e'e + (\alpha - a)Z'Z(\alpha - a)]} \propto \sigma^{-N} e^{-\frac{1}{2}(\alpha - a)\Omega_a^{-1}(\alpha - a)},$$

где $a = (Z'Z)^{-1}Z'X$ — МНК-оценка α (см. 7.13),

$\Omega_a = \sigma^2 (Z'Z)^{-1}$ — матрица ковариации (см. 7.29).

Действительно,

$$(X - Z\alpha)'(X - Z\alpha) = (\underbrace{X - Za}_{e} - Z(\alpha - a))'(X - Za - Z(\alpha - a)) =$$

$$= e'e - \underbrace{2e'Z(\alpha - a)}_{=0} + (\alpha - a)'Z'Z(\alpha - a),$$

т.к. e и Z ортогональны (см. 7.18).

Теперь предполагается, что σ известна. Тогда

$$P_N(X|Z\alpha, \sigma^2 I) \propto e^{-\frac{1}{2}(\alpha-a)\Omega_a^{-1}(\alpha-a)},$$

а соотношение (19.4) записывается в более простой форме:

$$p(\alpha|X, Z) \propto P_N(X|Z\alpha, \sigma^2 I)p(\alpha).$$

Пусть α априорно распределен нормально с математическим ожиданием $\hat{\alpha}$ и ковариацией Ω:

$$p(\alpha) \propto e^{-\frac{1}{2}\left(\alpha-\widehat{\alpha}\right)'\Omega^{-1}\left(\alpha-\widehat{\alpha}\right)}.$$

Тогда

$$p(\alpha|X, Z) \propto e^{-\frac{1}{2}\left[\left(\alpha-\widehat{\alpha}\right)'\Omega^{-1}\left(\alpha-\widehat{\alpha}\right)+(\alpha-a)'\Omega_a^{-1}(\alpha-a)\right]}. \tag{19.5}$$

Утверждается, что α апостериорно распределен также нормально с математическим ожиданием

$$\bar{a} = \bar{\Omega}(\Omega_a^{-1}a + \Omega^{-1}\hat{\alpha}) \tag{19.6}$$

и ковариацией

$$\bar{\Omega} = \left(\Omega_a^{-1} + \Omega^{-1}\right)^{-1}, \tag{19.7}$$

т.е.

$$p(\alpha|X, Z) \propto e^{-\frac{1}{2}\left[(\alpha-\bar{a})'\bar{\Omega}^{-1}(\alpha-\bar{a})\right]}. \tag{19.8}$$

Для доказательства этого утверждения необходимо и достаточно показать, что разность показателей экспонент в (19.5) и (19.8) не зависит от α.

Вводятся новые обозначения: $x = \alpha - a$; $y = \alpha - \hat{\alpha}$; $A = \Omega_a^{-1}$; $B = \Omega^{-1}$.

В этих обозначениях показатель степени в (19.5) записывается следующим образом (множитель $-1/2$ отбрасывается):

$$x'Ax + y'By. \tag{19.9}$$

В этих обозначениях

$$\bar{\Omega} = (A+B)^{-1},$$
$$\bar{a} = (A+B)^{-1}(A(\alpha - x) + B(\alpha - y)) =$$
$$= (A+B)^{-1}((A+B)\alpha - (Ax+By)) = \alpha - (A+B)^{-1}(Ax+By)$$

и, следовательно, показатель степени в (19.8) выглядит так (множитель $-1/2$ также отбрасывается):

$$(Ax - By)'(A+B)^{-1}(Ax - By). \tag{19.10}$$

Искомая разность (19.9) и (19.10) записывается следующим образом:

$$(1) - (2) - (3) + (4),$$

где

$$(1) = A - A(A+B)^{-1}A,$$
$$(2) = A(A+B)^{-1}B,$$
$$(3) = B(A+B)^{-1}A,$$
$$(4) = B - B(A+B)^{-1}B.$$

Легко показать, что все эти матрицы одинаковы и равны некоторой матрице C:

$$(1) = A(A+B)^{-1}A\left(A^{-1}(A+B)A^{-1}A - I\right) = A(A+B)^{-1}B = (2),$$
$$(2) = A\left[B\left(A^{-1}+B^{-1}\right)A\right]^{-1}B = AA^{-1}\left(A^{-1}+B^{-1}\right)^{-1}BB^{-1} =$$
$$= \left(A^{-1}+B^{-1}\right)^{-1} = C,$$
$$(3) = B\left[A\left(A^{-1}+B^{-1}\right)B\right]^{-1}A = BB^{-1}\left(A^{-1}+B^{-1}\right)A^{-1}A = C = (2),$$
$$(4) = B(A+B)^{-1}B\left[B^{-1}(A+B)B^{-1}B - I\right] = B(A+B)^{-1}A = (3),$$

и, следовательно, искомая разность представима в следующей форме:

$$x'Cx - x'Cy - y'Cx + y'Cy = (x-y)'C(x-y) = \left(\widehat{\alpha} - a\right)C\left(\widehat{\alpha} - a\right).$$

Что и требовалось доказать.

Как видно из (19.5, 19.6), апостериорная ковариация ($\bar{\Omega}$) является результатом гармонического сложения опытной (Ω_a) и априорной (Ω) ковариаций, апостериорные оценки регрессии ($\bar{a}$) — средневзвешенными (матричными) опытных (a) и априорных ($\hat{\alpha}$) оценок. Если априорные оценки имеют невысокую точность, и Ω велика, то влияние их на апостериорные оценки невелико, и последние определяются в большой степени опытными оценками. В предельном случае, когда $\Omega \to \infty$, т.е. априорная информация совершенно не надежна, $\bar{a} \to a$, $\bar{\Omega} \to \Omega_a$.

19.2. Объединение двух выборок

В действительности априорная информация может быть также опытной, но полученной в предшествующем опыте. Тогда формулы, полученные в предыдущем пункте показывают, как информация нового опыта — по новой выборке — корректирует оценки, полученные в предыдущем опыте — по старой выборке. В данном пункте показывается, что в результате применения этих формул получаемая апостериорная оценка в точности равна оценке, которую можно получить по объединенной выборке, включающей старую и новую.

Пусть имеется две выборки:

старая — Z_1, X_1: $X_1 = Z_1\alpha_1 + \varepsilon_1$,

и новая — Z_2, X_2: $X_2 = Z_2\alpha_2 + \varepsilon_2$.

Считается, что σ_1 и σ_2 известны (как и в предыдущем пункте).

Даются оценки параметров по этим двум выборкам:

$$a_1 = (Z'_1 Z_1)^{-1} Z'_1 X_1,\ \Omega_{a_1} = \sigma_1^2 (Z'_1 Z_1)^{-1},$$
$$a_2 = (Z'_2 Z_2)^{-1} Z'_2 X_2,\ \Omega_{a_2} = \sigma_2^2 (Z'_2 Z_2)^{-1}.$$

В предыдущем пункте первой выборке соответствовала априорная оценка, второй — опытная.

Теперь дается оценка параметров по объединенной выборке. При этом наблюдения должны быть приведены к одинаковой дисперсии:

$$\begin{bmatrix} X_1/\sigma_1 \\ X_2/\sigma_2 \end{bmatrix} = \begin{bmatrix} Z_1/\sigma_1 \\ Z_2/\sigma_2 \end{bmatrix} \alpha + \begin{bmatrix} \varepsilon_1/\sigma_1 \\ \varepsilon_2/\sigma_2 \end{bmatrix}.$$

В этой объединенной регрессии остатки имеют дисперсию, равную единице.

Оценки параметров рассчитываются следующим образом:

$$a = \left[\begin{bmatrix} Z_1'/\sigma_1 & Z_2'/\sigma_2 \end{bmatrix} \begin{bmatrix} Z_1/\sigma \\ Z_2/\sigma_2 \end{bmatrix}\right]^{-1} \begin{bmatrix} Z_1'/\sigma_1 & Z_2'/\sigma_2 \end{bmatrix} \begin{bmatrix} X_1/\sigma \\ X_2/\sigma_2 \end{bmatrix} =$$

$$= \left(\frac{1}{\sigma_1^2} Z_1'Z_1 + \frac{1}{\sigma_2^2} Z_2'Z_2\right)^{-1} \left(\frac{1}{\sigma_1^2} Z_1'X_1 + \frac{1}{\sigma_2^2} Z_2'X_2\right) =$$

$$= \left(\Omega_1^{-1} + \Omega_2^{-1}\right)^{-1} \left(\frac{1}{\sigma_1^2}(Z_1'Z_1)(Z_1'Z_1)^{-1} Z_1'X_1 + \frac{1}{\sigma_2^2}(Z_2'Z_2)(Z_2'Z_2)^{-1} Z_2'X_2\right) =$$

$$= \left(\Omega_1^{-1} + \Omega_2^{-1}\right)^{-1} \left(\Omega_1^{-1} a_1 + \Omega_2^{-1}\, a_2\right).$$

Ковариационная матрица (учитывая, что $\sigma^2 = 1$):

$$\Omega = \left(\Omega_1^{-1} + \Omega_2^{-1}\right)^{-1}.$$

Таким образом, оценки по объединенной выборке в терминах предыдущего пункта являются апостериорными.

19.3. Упражнения и задачи

Упражнение 1

По данным таблицы 19.1:

1.1. Оцените регрессию X по Z и константе, учитывая априорную информацию, что математические ожидания всех коэффициентов регрессии равны 2, а их ковариационная матрица — единичная. Считать, что дисперсия ошибки равна 2.

1.2. Разделите выборку на две части. Одна часть — 20 первых наблюдений, другая часть — 20 остальных наблюдений. Считать, что дисперсия ошибки в первой части равна 1, а во второй части — 4.

   а) Оцените обычную регрессию, воспользовавшись первой частью выборки. Найдите матрицу ковариаций полученных оценок.

   б) Используя информацию, полученную на шаге (а), как априорную информацию о математическом ожидании и ковариационной матрице коэффициентов, оцените байесовскую регрессию для второй части выборки.

Таблица 19.1

№	X	Z	№	X	Z	№	X	Z	№	X	Z
1	6.7	2.2	11	2.4	1.2	21	4.8	1.8	31	–1.4	1.7
2	5.5	1.8	12	5.8	0.8	22	3.3	0.8	32	–0.9	2
3	4.8	1.5	13	5.7	2.5	23	5.2	2.5	33	4.2	0.3
4	3	0.3	14	–0.9	1.7	24	5.4	2.1	34	–1.4	2
5	4.9	1.9	15	9.3	2.7	25	4.5	2.8	35	–2.6	1.3
6	2.8	0.7	16	3	2.2	26	3.8	1	36	3.1	0.1
7	2.7	0.8	17	–2.9	2.8	27	3.9	1.4	37	2.5	1.3
8	7	2.1	18	–1.5	1.8	28	6.4	2.4	38	–0.8	2.4
9	5.8	1.4	19	1.8	0.7	29	2.7	0.8	39	1.7	0.7
10	6.3	2.3	20	8.3	2.9	30	4.2	0.1	40	–0.1	1.6

в) Оцените регрессию, используя все наблюдения. Регрессия должна быть взвешенной, т.е. наблюдения каждой из частей нужно разделить на корень из соответствующей дисперсии. Найдите ковариационную матрицу оценок. Сравните с результатом, полученным на шаге (б). Совпадают ли коэффициенты и ковариационные матрицы?

Задачи

1. Чем отличается байесовская регрессия от обычной регрессии с точки зрения информации о коэффициентах? Приведите формулы для оценки параметров по этим двум регрессиям.

2. Налоговая инспекция считает, что предприятия в среднем недоплачивают налог на прибыль в 80% случаев. Вероятность того, что в ходе проверки некоторого предприятия будет выявлено такое нарушение, равна 40% для предприятия, которое недоплачивает налог, и 10% для предприятия, которое полностью выплачивает налог (ошибочно). Вычислите апостериорную вероятность того, что данное предприятие недоплачивает налог на прибыль, если в ходе проверки не было выявлено нарушений.

3. Студент может либо знать, либо не знать предмет и либо сдать, либо не сдать экзамен по этому предмету. Вероятность того, что студент знает предмет

равна 0.3. Если студент знает предмет, то вероятность того, что он сдаст экзамен, равна 0.9, а если не знает, то 0.6. Какова вероятность, что студент не знает предмет, если он сдал экзамен?

4. Предположим, что исследователь исходит из априорной информации, что коэффициенты регрессии распределены нормально с некоторым математическим ожиданием и ковариационной матрицей, а дисперсия ошибки равна некоторой известной величине. Исследователь получил какие-то данные и вычислил по ним апостериорное распределение. Затем он получил дополнительные данные и использовал прежнее апостериорное распределение как априорное. Можно ли утверждать, что новое апостериорное распределение будет нормальным? Ответ обоснуйте.

5. Случайная величина ξ имеет нормальное распределение с математическим ожиданием μ и дисперсией 16. Априорно известно, что μ имеет распределение $N(2, 9)$. Выборочное среднее по выборке длиной N равно 1. Найдите апостериорное распределение μ в зависимости от N.

6. Чему равна апостериорная оценка параметра, если его априорная оценка имеет нормальное распределение с математическим ожиданием 2 и дисперсией 0.25, а выборочная оценка равна 8 по выборке длиной 10?

7. Априорная оценка параметра имеет нормальное распределение с математическим ожиданием 2 и дисперсией 0.5, а выборочная оценка по выборке длиной 20 равна 2. Запишите плотность распределения апостериорных оценок.

8. Оценка параметра по первой части выборки равна 0 при дисперсии оценки 1, а по второй части выборки она равна 1 при дисперсии 2. Найдите оценку параметра по всей выборке.

9. Оценки регрессии по первой выборке совпадают с оценками по объединению двух выборок. Что можно сказать об оценках по второй выборке? Докажите свое утверждение.

Рекомендуемая литература

1. **Зельнер А.** Байесовские методы в эконометрии. — М.: «Статистика», 1980. (Гл. 2, 3).

2. **Лимер Э.** Сатистический анализ неэксперементальных данных. — М.: «Финансы и статистика», 1983.

3. Справочник по прикладной статистике. В 2-х т. Т 2. / Под ред. Э. Ллойда, У. Ледермана. — М.: «Финансы и статистика», 1990. (Гл. 15).

4. **Judge G.G., Griffiths W.E., Hill R.C., Luthepohl H., Lee T.** Theory and Practice of Econometrics. — New York: John Wiley & Sons, 1985. (Ch. 4).

Глава 20

Дисперсионный анализ

В этой главе продолжается рассмотрение темы, начатой в пункте 4.3. Здесь анализируются модели дисперсионного анализа в общем виде и доказываются некоторые из сделанных ранее утверждений.

Как и прежде, исходная совокупность $x_i,\ i = 1, \ldots, N$ сгруппирована по n факторам; j-й фактор может находиться на одном из k_j уровней. Регрессионная модель дисперсионного анализа общего вида получается исключением из модели регрессии с фиктивными переменными, полученной в конце пункта 9.1, «обычных» регрессоров:

$$X = \sum_{J=0}^{G} \widetilde{Z}^J \widetilde{\beta}^J + \varepsilon, \tag{20.1}$$

где $\widetilde{Z}^J = \bar{\otimes}_{j \in J} \widetilde{Z}^j$ (матрица $\widetilde{Z}^j$ имеет размерность $N \times k_j$, и в ее i_j-м столбце единицы стоят в строках тех наблюдений, в которых j-й фактор находится на i_j-м уровне, остальные элементы равны 0), или, как это следует из структуры $\widetilde{Z}$ и $\widetilde{\beta}$, представленной в пункте 9.1, в покомпонентной записи:

$$x_{I,\, i_I} = \beta_0 + \sum_{J=1}^{G} \beta^J_{I(J)} + \varepsilon_{I, i_I}, \tag{20.2}$$

где I — мультииндекс конечной группы, $I = I_1, \ldots, I_K$ (см. обозначения в п. 1.9);

i_I — линейный индекс элемента в конечной группе, $i_I = 1, \ldots, N_I$, N_I — численность конечной группы;

$\beta^J_{I(J)}$ (по сравнению с обозначениями, используемыми в п. 4.3, добавлен верхний индекс J, необходимый в данной главе для более точной идентификации параметра) — параметр эффекта сочетания (совместного влияния) факторов J на данный элемент совокупности (на значение изучаемой переменной в данном наблюдении).

Так, например, если $n = 3,\ I = \{2,3,1\},\ J = \{1,3\}$, то $\beta^J_{I(J)} = \beta^{1,3}_{2,1}$.

В пункте 9.1 отмечено, что в модели (20.1) на регрессорах существует много линейных зависимостей и поэтому непосредственно оценить ее нельзя. Для исключения линейных зависимостей регрессоров проводится следующее преобразование. Предполагая, что суммы компонент вектора $\widetilde{\beta}^J$ по всем значениям каждого элемента нижнего мультииндекса $I(J)$ равны нулю (в принятых ниже обозначениях: $\widetilde{Z}^{jJ'}\widetilde{b}^J = 0$ для всех $j \in J$), переходят к вектору β^J путем исключения из $\widetilde{\beta}^J$ всех тех его компонент, для которых хотя бы один элемент нижнего мультииндекса равен единице (благодаря сделанному предположению их всегда можно восстановить, поскольку они линейно выражаются через оставшиеся компоненты). Теперь модель можно записать в форме без линейных зависимостей регрессоров:

$$X = \sum_{J=0}^{G} Z^J \beta^J + \varepsilon, \tag{20.3}$$

где $Z^J = \widetilde{Z}^J C^J$, а $C^J = \underset{j\in J}{\otimes} C^j$, матрица C^j имеет следующую структуру:

$$\begin{bmatrix} -1_{k_j-1} \\ I_{k_j-1} \end{bmatrix}.$$

При этом, как и для модели (20.1), остается справедливым соотношение

$$Z^J = \underset{j\in J}{\bar{\otimes}} Z^j.$$

Эквивалентность моделей (20.1) и (20.3) очевидна, т.к. $\widetilde{\beta}^J = C^J \beta^J$.

В этой главе сначала рассматривается частный случай, когда численности всех конечных групп N_I равны единице, т.е. для каждого сочетания уровней факторов имеется строго одно наблюдение.

20.1. Дисперсионный анализ без повторений

В этом случае $N = K = \prod\limits_{G} k_j = \prod\limits_{j=1}^{n} k_j$, регрессионные модели (20.1) и (20.3) записываются без случайной ошибки, т.к. изучаемая переменная в точности раз-

лагается по эффектам всех возможных взаимодействий факторов (здесь и далее модели записываются в оценках параметров, т.е. β меняются на b):

$$X = \sum_{J=0}^{G} \widetilde{Z}^J \widetilde{b}^J, \tag{20.4}$$

$$X = \sum_{J=0}^{G} Z^J b^J, \tag{20.5}$$

а модель в покомпонентном представлении (20.2) еще и без линейного внутригруппового индекса:

$$x_I = b^0 + \sum_{J=1}^{G} b^J_{I(J)}. \tag{20.6}$$

Модель (20.5) можно переписать более компактно:

$$X = Zb. \tag{20.7}$$

Поскольку матрицы Z^J имеют размерности $N * K^J_-$ ($K^J_- = \prod_J (k_j - 1)$, $K^0_- = 1$), а $\sum_{J=0}^{G} K^J_- = K = N$ (как это было показано в п. 4.3), то матрица Z квадратна, и $b = Z^{-1}X$. Но для получения общих результатов, имеющих значение и для частных моделей, в которых эффекты высоких порядков принимаются за случайную ошибку, используется техника регрессионного анализа:

$$b = M^{-1}m = (\frac{1}{N}Z'Z)^{-1}\frac{1}{N}Z'X.$$

В этом параграфе сделанные утверждения будут иллюстрироваться примером, в котором $n = 2$, $k_1 = k_2 = 2$ и модели (20.4) и (20.5) записываются следующим образом:

$$\begin{bmatrix} x_{11} \\ x_{12} \\ x_{21} \\ x_{22} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} b^0 + \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} b^1_1 \\ b^1_2 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} b^2_1 \\ b^2_2 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} b^{12}_{11} \\ b^{12}_{12} \\ b^{12}_{21} \\ b^{12}_{22} \end{bmatrix},$$

$$\begin{bmatrix} x_{11} \\ x_{12} \\ x_{21} \\ x_{22} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} b^0 + \begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \end{bmatrix} b_2^1 + \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix} b_2^2 + \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix} b_{22}^{12}.$$

Каждая из матриц $\widetilde{Z}^J$ является прямым произведением ряда матриц и векторов:

$$\widetilde{Z}^J = \bigotimes_G \left\{ \begin{array}{l} I_{k_j}, \text{ если } j \in J \\ 1_{k_j}, \text{ если } j \notin J \end{array} \right\}.$$

В этом легко убедиться, рассуждая по индукции. Так, в рассматриваемом примере:

$$\widetilde{Z}^0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \; \widetilde{Z}^1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

$$\widetilde{Z}^2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \; \widetilde{Z}^{12} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Матрицы C^J можно представить следующим образом:

$$C^J = \bigotimes_J C^j = \bigotimes_G \left\{ \begin{array}{l} C^j, \text{ если } j \in J \\ 1, \text{ если } j \notin J \end{array} \right\}.$$

Тогда, используя свойство коммутативности прямого и «обычного» умножения матриц (см. п. 9.1), можно показать следующее:

$$Z^J = \widetilde{Z}^J C^J = \bigotimes_G \left\{ \begin{array}{l} I_{k_j}, \text{ если } j \in J \\ 1_{k_j}, \text{ если } j \notin J \end{array} \right\} \left\{ \begin{array}{l} C^j, \text{ если } j \in J \\ 1, \text{ если } j \notin J \end{array} \right\} =$$

$$= \bigotimes_G \left\{ \begin{array}{l} C^j, \text{ если } j \in J \\ 1_{k_j}, \text{ если } j \notin J \end{array} \right\}. \quad (20.8)$$

Теперь можно уточнить структуру матрицы M. Она состоит из блоков

$$M^{\bar{J}J} = \frac{1}{N} Z^{\bar{J}'} Z^J,$$

и все внедиагональные блоки (при $\bar{J} \neq J$), благодаря (20.8), равны 0.

Действительно,

$$M^{\bar{J}J} = \frac{1}{N} \underset{G}{\otimes} \left\{ \begin{array}{l} C^{j'}, \text{ если } j \in \bar{J} \\ 1'_{k_j}, \text{ если } j \notin \bar{J} \end{array} \right\} \left\{ \begin{array}{l} C^j, \text{ если } j \in J \\ 1_{k_j}, \text{ если } j \notin J \end{array} \right\}$$

и, если $j \in \bar{J}$, $\notin J$, то в ряду прямых произведений матриц возникает матрица (точнее, вектор-столбец) $C^{j'} 1_{k_j}$; если $j \notin \bar{J}$, $\in J$, то появляется матрица (вектор-строка) $1'_{k_j} C^j$. И та, и другая матрица (вектор-столбец или вектор-строка) по построению матриц C^j равны нулю. Следовательно, $M^{\bar{J}J} = 0$ при $\bar{J} \neq J$.

Для диагональных блоков выполняются следующие соотношения:

$$M^{JJ} = M^J = \frac{1}{N} \prod_{G-J} k_j \underset{J}{\otimes} C^{j'} C^j = \frac{1}{K^J} \underset{J}{\otimes} C^{j'} C^j = \underset{J}{\otimes} M^j,$$

где $M^j = \frac{1}{k_j} C^{j'} C^j = \frac{1}{k_j} (1_{k_j - 1} 1'_{k_j - 1} + I_{k_j - 1})$.

В рассматриваемом примере $M = I_4$.

Вектор m состоит из блоков m^J:

$$m^J = \frac{1}{N} Z^{J'} X = \frac{1}{N} C^{J'} \widetilde{Z}^{J'} X = \frac{1}{K^J} C^{J'} X^J,$$

где $X^J = \frac{K^J}{N} \widetilde{Z}^{J'} X$ — вектор-столбец средних по сочетаниям значений факторов J. Его компоненты в пункте 4.3 обозначались $x_{I(J)}$ ($x^J_{I(J)}$ — добавлен верхний индекс J — является средним значением x по тем наблюдениям, в которых 1-й фактор из множества J находится на i_{j_1}-м уровне, 2-й — на i_{j_2}-м уровне и т.д.); $X^0 = \bar{x}$, $X^G = X$. Это следует из структуры матрицы $\widetilde{Z}^{J'}$.

После решения системы нормальных уравнений

$$m^J = M^J b^J, \; J = 1, \ldots, G$$

и перехода к «полным» векторам параметров эффектов получается следующее:

$$\widetilde{b}^J = C^J (C^{J'} C^J)^{-1} C^{J'} X^J = B^J X^J = \underset{J}{\otimes} B^j X^J,$$

где $B^j = C^j (C^{j'} C^j)^{-1} C^{j'} = I_{k_j} - \frac{1}{k_j} 1^{k_j}$ ($1^{k_j} = 1_{k_j} 1'_{k_j}$), $B^0 = 1$.

В рассматриваемом примере

$$B^0 = 1,\ B^1 = B^2 = \frac{1}{2}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix},\ B^{12} = \begin{bmatrix} 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}.$$

В силу блочной диагональности матрицы B, параметры разных эффектов $\widetilde{b}^J$ (разных по J) не зависят друг от друга, и исключение из уравнения некоторых из них не повлияет на значения параметров оставшихся эффектов. Кроме того, это доказывает справедливость приведенного в пункте 4.3 дисперсионного тождества (4.41).

Действительно, воспользовавшись одной из формул (6.18) для объясненной дисперсии, которая в данном случае равна полной дисперсии, можно получить следующее:

$$s^2 = \sum_{J=1}^{G} b^{J'} M^J b^J = \sum_{J=1}^{G} \frac{1}{K^J} b^{J'} C^{J'} C^J b^J = \sum_{J=1}^{G} \frac{1}{K^J} \widetilde{b}^{J'} \widetilde{b}^J = \sum_{J=1}^{G} s_J^2,$$

т.е. то, что и требуется.

Введенное в пункте 4.3 рекуррентное правило расчета параметров эффектов, когда параметры более младших эффектов рассчитываются по значениям параметров более старших эффектов, действует, поскольку наряду с соотношениями (20.4) и (20.6) выполняются аналогичные соотношения для всех средних:

$$X^J = \sum_{0,\bar{J}\in J} \widetilde{Z}^{\bar{J}J} \widetilde{b}^{\bar{J}}, \tag{20.9}$$

где суммирование ведется от нуля и по всем подмножествам J ($\bar{J} \subset J$), а $\widetilde{Z}^{\bar{J}J}$ — матрица фиктивных переменных для сочетания факторов $\bar{J}$ в модели, для которой полным набором факторов является J, т.е.

$$\widetilde{Z}^{\bar{J}J} = \bigotimes_J \left\{ \begin{array}{l} I_{k_j},\ \text{если } j \in \bar{J} \\ 1_{k_j},\ \text{если } j \notin \bar{J} \end{array} \right\} \quad (X^G = X,\ \widetilde{Z}^{JG} = \widetilde{Z}^J),$$

$$x^J_{I(J)} = b^0 + \sum_{\bar{J}\in J} b^{\bar{J}}_{I(\bar{J})}. \tag{20.10}$$

Для доказательства этого факта обе части соотношения (20.5) умножаются слева на $\dfrac{K^J}{N}\widetilde{Z}^{J'}$ (текущим множеством в сумме становится $\bar{J}$):

$$\frac{K^J}{N}\widetilde{Z}^{J'}X = \sum_{\bar{J}=0}^{G}\frac{K^J}{N}\widetilde{Z}^{J'}Z^{\bar{J}}b^{\bar{J}}, \tag{20.11}$$

и рассматривается произведение $\widetilde{Z}^{J'}Z^{\bar{J}}$ из правой части полученного соотношения, которое представляется следующим образом:

$$\bigotimes_{G}\left\{\begin{array}{l} I_{k_j},\ \text{если } j\in J \\ 1'_{k_j},\ \text{если } j\notin J \end{array}\right\}\left\{\begin{array}{l} C^j,\ \text{если } j\in \bar{J} \\ 1_{k_j},\ \text{если } j\notin \bar{J} \end{array}\right\}. \tag{20.12}$$

Возможны четыре случая.

1) $j\notin J,\ j\in\bar{J}$, тогда в этом произведении возникает сомножитель $1'_{k_j}C^j$, который равен нулю, т.е. в правой части соотношения (20.11) остаются только такие слагаемые, для которых $\bar{J}\in J$.

2) $j\notin J,\ j\notin\bar{J}$, тогда возникает сомножитель k_j, и, следовательно, каждое слагаемое в правой части (20.11) получает сомножитель $\dfrac{N}{K^J}$, который сокращается с уже имеющимся сомножителем $\dfrac{K^J}{N}$.

3) $j\in J,\ j\in\bar{J}$, тогда возникает сомножитель C^j.

4) $j\in J,\ j\notin\bar{J}$, тогда возникает сомножитель 1_{k_j}.

Таким образом, рассматриваемое произведение в точности равно $Z^{\bar{J}J}$. Поскольку левая часть соотношения есть X^J по определению, доказательство завершено.

Соотношение (20.9) дает правило расчета $\widetilde{b}^J$, если все параметры более старших эффектов известны. При $J=0$ это соотношение означает

$$X^0=\bar{x}=b^0.$$

Далее последовательно рассчитываются параметры все более младших эффектов.

Техника применения F-критерия для проверки степени значимости отдельных факторов и их сочетаний приведена в пункте 4.3. Здесь важно отметить, что она применима только в рамках гипотезы о нормальности распределения x.

20.2. Дисперсионный анализ с повторениями

Переходя к более общему и более сложному случаю модели дисперсионного анализа с повторениями (20.1), полезно воспользоваться следующим подходом. Если в модели регрессионного анализа

$$X = Z\alpha + \varepsilon$$

несколько строк матрицы Z одинаковы, то можно перейти к сокращенной модели, в которой из всех этих строк оставлена одна, а в качестве соответствующей компоненты вектора X взято среднее по этим наблюдениям с одинаковыми значениями независимых факторов. Это агрегированное наблюдение в соответствии с требованием ОМНК должно быть взято с весом $\sqrt{N_g}$, где N_g — количество одинаковых строк в исходной модели, поскольку, как известно, дисперсия средней ошибки в этом наблюдении в N_g раз меньше дисперсии исходных ошибок. Значения оценок параметров в исходной и сокращенной моделях будут одинаковыми, но полная и остаточная суммы квадратов в исходной модели будут больше, чем в сокращенной, на сумму квадратов отклонений переменных x по исключенным наблюдениям от своей средней.

При доказательстве этого утверждения считается, что одинаковы первые N_1 строк в матрице Z:

$$\begin{bmatrix} x_1 \\ X \end{bmatrix} = \begin{bmatrix} 1_{N_1} \otimes z_1 \\ Z \end{bmatrix} a + \begin{bmatrix} e_1 \\ e \end{bmatrix}.$$

Система нормальных уравнений для оценки a записывается следующим образом:

$$\begin{bmatrix} 1'_{N_1} \otimes z'_1 & Z' \end{bmatrix} \begin{bmatrix} x_1 \\ X \end{bmatrix} = \begin{bmatrix} 1'_{N_1} \otimes z'_1 & Z' \end{bmatrix} \begin{bmatrix} 1_{N_1} \otimes z_1 \\ Z \end{bmatrix} a$$

или, после умножения векторов и матриц,

$$1'_{N_1} \otimes z'_1 \; x_1 + Z'X = (1'_{N_1} \otimes z'_1 \; 1_{N_1} \otimes z_1 + Z'Z)a \overset{1'_{N_1} \otimes z'_1 \; x_1 = 1'_{N_1} \otimes z'_1 \; x_1 \otimes 1}{\Rightarrow}$$
$$\Rightarrow \; N_1 z'_1 \; \bar{x}_1 + Z'X = (N_1 z'_1 z_1 + Z'Z)a.$$

Сокращенная модель записывается следующим образом:

$$\begin{bmatrix} \sqrt{N_1}\bar{x}_1 \\ X \end{bmatrix} = \begin{bmatrix} \sqrt{N_1} z_1 \\ Z \end{bmatrix} a + \begin{bmatrix} \sqrt{N_1}\bar{e}_1 \\ e \end{bmatrix}.$$

Видно, что система нормальных уравнений для оценки параметров этой модели в точности совпадает с системой нормальных уравнений для исходной модели, т.е. оценки параметров в исходной и сокращенной моделях одинаковы.

Остаточная сумма квадратов в исходной модели равна

$$e_1'e_1 + e'e, \tag{20.13}$$

в сокращенной модели —

$$N_1\bar{e}_1^2 + e'e. \tag{20.14}$$

Пусть первые N_1 наблюдений в исходной модели имеют нижний индекс $1i$, где $i = 1, \ldots, N_1$. Тогда

$$e_{1i} = x_{1i} - z_1 a = \bar{x}_1 + x_{1i} - \bar{x}_1 - z_1 a = \bar{e}_1 + (x_{1i} - \bar{x}_1)$$

и

$$e_1'e_1 = \sum e_{1i}^2 = \\ = \sum (\bar{e}_1 - (x_{1i} - \bar{x}_1))^2 = N_1\bar{e}_1^2 + \underbrace{2\bar{e}_1 \sum (x_{1i} - \bar{x}_1)}_{=0} + \sum (x_{1i} - \bar{x}_1)^2.$$

Сравнение (20.13) и (20.14) с учетом полученного результата завершает доказательство.

В исходной модели (20.1) строки матрицы Z, относящиеся к одной конечной группе, одинаковы, что позволяет в конечном счете перейти к сокращенной модели, существенно меньшей размерности. В исходной модели $N = \sum\limits_{I=I_1}^{I_K} N_I$, и пусть

$x_I,\ s_I^2$ — средняя и дисперсия в I-й конечной группе,

$s_e^2 = \frac{1}{N}\sum N_I s_I^2$ — внутригрупповая дисперсия,

$\bar{x} = \frac{1}{N}\sum N_I x_I$ — общая средняя,

$s_q^2 = \frac{1}{N}\sum N_I (x_I - \bar{x})^2$ — общая межгрупповая дисперсия.

Еще в пункте 4.3 было доказано, что

$$s^2 = s_e^2 + s_q^2.$$

На основании этого тождества, учитывая, что количество степеней свободы внутригрупповой дисперсии равно $N - K - 1$, а межгрупповой — K, можно проверить статистическую гипотезу о значимости влияния всех факторов сразу на изучаемую переменную. Но в данном случае можно провести более детальный анализ

влияния отдельных факторов и их сочетаний, аналогичный тому, который проводился в случае модели без повторений. В таком анализе используется сокращенная модель, дающая (как это было показано выше) такие же оценки параметров регрессии, что и исходная модель, но представляющая не всю дисперсию, а только межгрупповую:

$$\sqrt{\widetilde{N}}X^G = \sqrt{\widetilde{N}}\sum_{J=0}^{G}\widetilde{Z}^J\widetilde{b}^J = \sqrt{\widetilde{N}}\sum Z^J b^J, \tag{20.15}$$

где X^G — вектор средних по конечным группам x_I, $\widetilde{N}$ — диагональная матрица численностей конечных групп N_I.

Эта модель отличается от моделей (20.4) и (20.5) только наличием матричного множителя $\sqrt{\widetilde{N}}$. Но это отличие принципиальное. Оно влечет потерю всех тех «хороших» свойств, которыми обладала модель без повторений. В частности, матрица M в общем случае перестает быть блочно-диагональной, эффекты разных сочетаний факторов становятся зависимыми, а дисперсионное тождество теряет простую структуру.

С моделью (20.15) можно работать как с обычной регрессионной моделью, используя известные критерии проверки разных статистических гипотез (понимая при этом, что результаты проверки будут неоднозначны, в силу взаимозависимостей регрессоров). Но следует иметь в виду, что оценки параметров в этой модели смещены (что, впрочем, не влияет на результаты проверки гипотез). В частности, $b^0 \neq \bar{x}$.

Для того чтобы исключить смещенность оценок, необходимо правильно строить матрицы C, используемые при устранении линейных зависимостей в матрице $\widetilde{Z}$. Это связано с тем, что теперь должны равняться нулю не простые, а взвешенные суммы компонент векторов $\widetilde{\beta}^J$ по каждому элементу нижнего мультииндекса $I(J)$.

В частности, если $N^j_{i_j}$ — численность группы, в которой j-й фактор находится на i_j-м уровне, то

$$C^j = \begin{bmatrix} -\frac{1}{N^j_1}\begin{pmatrix} N^j_2 & \cdots & N^j_{k_j} \end{pmatrix} \\ I_{k_j-1} \end{bmatrix}$$

(понятно, что когда численности всех конечных групп равны единице, эта матрица приобретает обычную структуру).

Можно показать, что специальный выбор структуры матриц C^J может обеспечить максимальную «разреженность» матрицы M, т.е. обеспечить равенство нулю блоков $M^{0G}(G \neq 0)$, $M^{\bar{J}J}$ $(\bar{J} \subset J)$. Работая со структурой матриц C^J, можно обнаружить частный случай, когда модель с повторениями обладает теми

же свойствами, что и модель без повторений. Этот случай имеет место, если каждый последующий (более младший) фактор делит все полученные ранее группы в одинаковой пропорции. Однако усилия, которые необходимы для доказательства этих фактов, далеко не соответствуют их практической значимости. Так, вряд ли можно ожидать, что ряд групп, имеющих разную численность, можно разбить на подгруппы в одинаковой пропорции — хотя бы в силу целочисленности образуемых подгрупп.

В принципе, с моделью межгрупповой дисперсии (20.15) можно работать и без сомножителя $\sqrt{\widetilde{N}}$, т.е. в рамках «хороших» свойств модели без повторений. Для этого достаточно предположить, что исходная модель (20.1) неоднородна по дисперсии ошибок в разных наблюдениях. А именно: считать, что дисперсия ошибки наблюдения обратно пропорциональна численности конечной группы, в которую оно входит (чем больше наблюдений — повторений — в конечной группе, тем меньше дисперсия ошибки в отдельном наблюдении). Тогда сокращенная модель будет однородной по дисперсии и для ее оценки окажется применим простой МНК.

20.3. Упражнения и задачи

Упражнение 1

Провести дисперсионный анализ (без повторений) данных, приведенных в таблице 20.1:

Имеются 2 фактора по 3 уровня каждый (I, II, III и A, B, C, соответственно). Рассчитать коэффициенты $\widetilde{b}$, а также Z, $\widetilde{Z}$, b, C_1, C_2, C_{12}, B_1, B_2, B_{12}, M, m.

Таблица 20.1

	A	B	C
I	3	0	4
II	0	7	0
III	2	8	3

Упражнение 2

В Таблице 20.2 приведены данные о зарплатах 52-х преподавателей американского колледжа: SX — пол (жен. — 1, муж. — 0); ученое звание: RK1 — assistant professor, RK2 — associate professor, RK3 — full professor; DG — ученая степень (доктор — 1, магистр — 0); SL — средний заработок за академический год, долл.

2.1. Провести дисперсионный анализ с помощью обычной регрессии.

2.2. Провести дисперсионный анализ с помощью взвешенной регрессии, когда совокупность наблюдений с одинаковыми значениями независимых факторов заменяется одним групповым наблюдением.

Таблица 20.2. (Источник: S. Weisberg (1985), Applied Linear Regression, 2nd Ed, New York: Wiley, page 194)

SX	RK1	RK2	RK3	DG	SL	SX	RK1	RK2	RK3	DG	SL
0	0	0	1	1	36350	0	0	1	0	1	24800
0	0	0	1	1	35350	1	0	0	1	1	25500
0	0	0	1	1	28200	0	0	1	0	0	26182
1	0	0	1	1	26775	0	0	1	0	0	23725
0	0	0	1	0	33696	1	1	0	0	0	21600
0	0	0	1	1	28516	0	0	1	0	0	23300
1	0	0	1	0	24900	0	1	0	0	0	23713
0	0	0	1	1	31909	1	0	1	0	0	20690
0	0	0	1	0	31850	1	0	1	0	0	22450
0	0	0	1	0	32850	0	0	1	0	1	20850
0	0	0	1	1	27025	1	1	0	0	1	18304
0	0	1	0	1	24750	0	1	0	0	1	17095
0	0	0	1	1	28200	0	1	0	0	1	16700
0	0	1	0	0	23712	0	1	0	0	1	17600
0	0	0	1	1	25748	0	1	0	0	1	18075
0	0	0	1	1	29342	0	1	0	0	0	18000
0	0	0	1	1	31114	0	0	1	0	1	20999
0	0	1	0	0	24742	1	1	0	0	1	17250
0	0	1	0	0	22906	0	1	0	0	1	16500
0	0	0	1	0	24450	0	1	0	0	1	16094
0	1	0	0	0	19175	1	1	0	0	1	16150
0	0	1	0	0	20525	1	1	0	0	1	15350
0	0	0	1	1	27959	0	1	0	0	1	16244
1	0	0	1	1	38045	1	1	0	0	1	16686
0	0	1	0	1	24832	1	1	0	0	1	15000
0	0	0	1	1	25400	1	1	0	0	1	20300

2.3. Учесть эффекты второго порядка: добавить в регрессию попарные произведения исходных фиктивных переменных. Значимы ли они?

Задачи

Таблица 20.3

	A	B
I	43	2
II	4	53
III	8	1

1. Что является отличительной особенностью модели дисперсионного анализа по сравнению с «обычными» моделями регрессионного анализа?
2. С помощью таблицы 20.3 задана классификация по двум факторам.

   Запишите матрицы фиктивных переменных для главных эффектов.
3. Какую структуру имеет матрица ковариаций оценок в дисперсионном анализе без повторений?
4. Как называется в дисперсионном анализе то, что в регрессионном анализе называется объясненной и остаточной дисперсией?
5. При проведении дисперсионного анализа с повторениями по усредненным наблюдениям используется взвешенная регрессия. С какой целью это делается?
6. Если в дисперсионном анализе без повторений отбросить эффекты высшего порядка, то как изменятся значения параметров оставшихся эффектов?
7. В модели полного дисперсионного анализа без повторений с одним фактором, имеющим три уровня, запишите матрицу нецентральных вторых моментов для матрицы регрессоров Z.
8. Сколько наблюдений нужно иметь для применения модели дисперсионного анализа без повторений в случае четырех факторов, каждый из которых может принимать три уровня, если учитывать только эффекты первого порядка?
9. Сколько наблюдений нужно иметь для применения модели полного дисперсионного анализа без повторений в случае двух факторов, каждый из которых может принимать три уровня?
10. Для модели дисперсионного анализа с двумя факторами, первый из которых имеет три уровня, а второй — два, рассчитать матрицу C^{12}.
11. Рассмотрим модель дисперсионного анализа с двумя факторами, первый из которых принимает два уровня, а второй — три уровня. Рассчитайте матрицы Z^1, Z^2.

12. В первой группе 20 человек, а во второй — 30 человек. Дисперсия оценок по «Эконометрии» в первой группе равна 1.5, а во второй — 1. Вычислите остаточную дисперсию в модели дисперсионного анализа.

13. В первой группе 20 человек, а во второй — 30 человек. Средняя оценка по «Эконометрии» в первой группе равна 3.5, а во второй — 4. Вычислите объясненную дисперсию в модели дисперсионного анализа.

14. В первой группе 20 человек, а во второй — 30 человек. Средняя оценка по «Философии» в первой группе равна 4.5, а во второй — 3. Вычислите коэффициенты в модели дисперсионного анализа.

15. В первой группе 20 человек, а во второй — 30 человек. Средняя оценка по «Эконометрии» в первой группе равна 3.5, а во второй — 4. Дисперсия оценок в первой группе равна 1.5, а во второй — 1. Вычислите общую дисперсию оценок двум группам.

16. Проводится дисперсионный анализ без повторений с двумя факторами, один из которых принимает три уровня, а другой — четыре. Как вычисляется статистика для проверки значимости эффектов второго порядка? Какое она имеет распределение (сколько степеней свободы)?

Рекомендуемая литература

1. **Болч Б., Хуань К.Дж.** Многомерные статистические методы для экономики. — М.: «Статистика», 1979. (Гл. 5)

2. **Себер Дж.** Линейный регрессионый анализ. — М.: «Мир», 1980.

3. **Шеффе Г**. Дисперсионный анализ. — М.: «Наука», 1980.

Глава 21

Модели с качественными зависимыми переменными

При изучении экономических явлений на дезагрегированном уровне (уровне отдельных экономических субъектов) возникает потребность в новых методах. Дело в том, что стандартные эконометрические методы, такие как классическая модель регрессии, предназначены для анализа переменных, которые могут принимать любое значение на числовой прямой, причем предполагается фактически, что распределение изучаемой переменной похоже на нормальное. Модели, в которых диапазон значений зависимой переменной ограничен, называют **моделями с ограниченной зависимой переменной**. Среди них важную роль играют модели, в которых изучаемая переменная дискретна и может принимать только некоторые значения (конечное число), либо даже имеет нечисловую природу (так называемые **модели с качественной зависимой переменной**). Модели такого рода помогают, в частности, моделировать выбор экономических субъектов. В качестве примера можно привести выбор предприятия: внедрять какую-то новую технологию или нет. Если индивидуальный выбор исследовать методами, предназначенными для непрерывных переменных, то будет неправомерно проигнорирована информация о поведенческой структуре ситуации.

21.1. Модель дискретного выбора для двух альтернатив

Анализ дискретного выбора основывается на микроэкономической теории, которая моделирует поведение индивидуума как выбор из данного множества аль-

тернатив такой альтернативы, которая бы максимизировала его полезность. Этот выбор с точки зрения стороннего наблюдателя, однако, не полностью предопределен. Исследователь не может наблюдать все факторы, определяющие результат выбора конкретного индивидуума. Коль скоро ненаблюдаемые факторы случайны, то выбор двух индивидуумов может быть разным при том, что наблюдаемые факторы совпадают. С его точки зрения это выглядит как случайный разброс среди индивидуумов с одними и теми же наблюдаемыми характеристиками.

Предполагается, что выбор осуществляется на основе ненаблюдаемой полезности альтернатив $u(x)$. Если $u(1) > u(0)$, то индивидуум выбирает $x = 1$, если $u(1) < u(0)$, то индивидуум выбирает $x = 0$. В простейшем случае полезность является линейной функцией факторов: $u(1) = z\alpha^1$ и $u(0) = z\alpha^0$. Чтобы модель была вероятностной, ее дополняют отклоняющими факторами, так что

$$u(1) = z\alpha^1 + \varepsilon^1,$$
$$u(0) = z\alpha^0 + \varepsilon^0.$$

Предполагается, что распределение отклонений ε^1 и ε^0 непрерывно.

Заметим, что для описания выбора вполне достаточно знать разность между полезностями вместо самих полезностей:

$$\tilde{x} = u(1) - u(0) = z(\alpha^1 - \alpha^0) + \varepsilon^1 - \varepsilon^0 = z\alpha + \varepsilon,$$

при этом оказывается, что в основе выбора лежит переменная $\tilde{x}$, которая представляет собой сумму линейной комбинации набора факторов z и случайного отклонения ε, имеющего некоторое непрерывное распределение:

$$\tilde{x} = z\alpha + \varepsilon.$$

Эта переменная является ненаблюдаемой. Наблюдается только дискретная величина x, которая связана с $\tilde{x}$ следующим образом: если $\tilde{x}$ больше нуля, то $x = 1$, если меньше, то $x = 0$.

Ясно, что по наблюдениям за x и z мы могли бы оценить коэффициенты α только с точностью до множителя. Умножение ненаблюдаемых величин $\tilde{x}$, α и ε на один и тот же коэффициент не окажет влияния на наблюдаемые величины x и z. Таким образом, можно произвольным образом нормировать модель, например, положить дисперсию ошибки равной единице.

Кроме того, в этой модели есть дополнительный источник неоднозначности: одним и тем же коэффициентам α могут соответствовать разные пары α^0 и α^1. Таким образом, можно сделать вывод, что исходная модель выбора принципиально неидентифицируема. Однако это не мешает ее использованию для предсказания результата выбора, что мы продемонстрируем в дальнейшем.

Без доказательства отметим, что если в модели выбора ε^1 и ε^0 имеют распределение $F(y) = e^{-e^{-y}}$ (распределение экстремального значения) и независимы, то $\varepsilon = \varepsilon^1 - \varepsilon^0$ имеет логистическое распределение. При этом получается модель, называемая *логит*.

Если ε^1 и ε^0 имеют нормальное распределение с параметрами 0 и $1/2$ и независимы, то $\varepsilon = \varepsilon^1 - \varepsilon^0$ имеет стандартное нормальное распределение. При этом получается модель, называемая *пробит*.

Модели логит и пробит рассматривались в главе 9.

21.2. Оценивание модели с биномиальной зависимой переменной методом максимального правдоподобия

Предыдущие рассуждения приводят к следующей модели:

$$\tilde{x} = z\alpha + \varepsilon,$$

$$x = \begin{cases} 0, & \tilde{x} < 0, \\ 1, & \tilde{x} > 0. \end{cases}$$

Пусть $F_\varepsilon(\cdot)$ — функция распределения отклонения ε. Выведем из распределения ε распределение $\tilde{x}$, а из распределения $\tilde{x}$ — распределение x:

$$\Pr(x = 1) = \Pr(\tilde{x} > 0) = \Pr(z\alpha + \varepsilon > 0) = \Pr(\varepsilon > -z\alpha) = 1 - F_\varepsilon(-z\alpha).$$

Для удобства обозначим $F(y) = 1 - F_\varepsilon(-y)$. (При симметричности относительно нуля распределения ε будет выполнено $F(y) = 1 - F_\varepsilon(-y) = F_\varepsilon(y)$.) Таким образом,

$$\Pr(x = 1) = F(z\alpha).$$

Пусть имеются N наблюдений, (x_i, z_i), $i = 1, \ldots, N$, которые соответствуют этой модели, так что x_i имеют в основе ненаблюдаемую величину $\tilde{x}_i = z_i\alpha + \varepsilon_i$. Предполагаем, что ошибки ε_i имеют нулевое математическое ожидание, одинаково распределены и независимы. Рассмотрим, как получить оценки коэффициентов α методом максимального правдоподобия.

Обозначим через $p_i = p_i(\alpha) = F(z_i\alpha)$. Также пусть $I_0 = \{i|\ x_i = 0\}$, $I_1 = \{i|\ x_i = 1\}$. Функция правдоподобия, то есть вероятность получения наблюдений x_i при данных z_i, имеет вид:

$$L(\alpha) = \prod_{i \in I_1} p_i(\alpha) \prod_{i \in I_0} (1 - p_i(\alpha)).$$

Вместо самой функции правдоподобия удобно использовать логарифмическую функцию правдоподобия:

$$\ln L(\alpha) = \sum_{i \in I_1} \ln p_i(\alpha) + \sum_{i \in I_0} \ln(1 - p_i(\alpha)),$$

которую можно записать как

$$\ln L(\alpha) = \sum_{i=1}^{N} \left(x_i \ln p_i(\alpha) + (1 - x_i) \ln(1 - p_i(\alpha)) \right). \tag{21.1}$$

В результате максимизации этой функции по α получаем оценки максимального правдоподобия. Условия первого порядка максимума (уравнения правдоподобия), т.е.

$$\frac{\partial \ln L(\alpha)}{\partial \alpha} = 0,$$

имеют простой вид:

$$\sum_{i=1}^{N} (x_i - p_i) \frac{f(z_i \alpha)}{p_i(1 - p_i)} z_i = 0,$$

где мы учли, что

$$\frac{\partial p_i(\alpha)}{\partial \alpha} = \frac{dF(z_i \alpha)}{d\alpha} = f(z_i \alpha) z_i,$$

где f — производная функции $F(\cdot)$. Поскольку $F(\cdot)$ представляет собой функцию распределения, то $f(\cdot)$ — плотность распределения.

Можно использовать следующий метод, который дает те же оценки, что и метод максимального правдоподобия. Пусть a^0 — некоторая приближенная оценка коэффициентов модели. Аппроксимируем функцию $F(\cdot)$ ее касательной в точке $z_i \alpha$ (т.е. применим линеаризацию):

$$F(z_i \alpha) \approx F(z_i a^0) + f(z_i a^0) z_i \left(\alpha - a^0 \right).$$

Подставим затем эту аппроксимацию в исходную модель:

$$x_i - p_i(a) \approx f(z_i a) z_i (\alpha - a) + \xi_i,$$

или

$$x_i - p_i(a^0) + f(z_i a^0) z_i a^0 \approx f(z_i a^0) z_i \alpha + \xi_i.$$

При данном a^0 это линейная регрессия. Как несложно проверить, дисперсия ошибки ξ_i равна $p_i(\alpha)(1-p_i(\alpha))$, т.е. ошибки гетероскедастичны. К этой модели можно применить взвешенную регрессию. Следует разделить левую и правую части на корень из оценки дисперсии ошибки ξ_i, т.е. на $\sqrt{p_i(a^0)(1-p_i(a^0))}$:

$$\frac{x_i - p_i(a^0) + f(z_i a^0) z_i a^0}{\sqrt{p_i(a^0)(1-p_i(a^0))}} \approx \frac{f(z_i a^0) z_i}{\sqrt{p_i(a^0)(1-p_i(a^0))}} \alpha + \frac{\xi_i}{\sqrt{p_i(a^0)(1-p_i(a^0))}}.$$

Оценивая эту вспомогательную регрессию, мы на основе оценок a^0 получим новые оценки, скажем a^1. Повторяя эту процедуру, получим последовательность оценок $\{a^k\}$. Если процедура сойдется, т.е. $a^k \to a$ при $k \to \infty$, то a будут оценками максимального правдоподобия.

В качестве оценки ковариационной матрицы оценок a можно использовать

$$-\left(\frac{\partial^2 \ln L(a)}{\partial \alpha \partial \alpha'}\right)^{-1}.$$

По диагонали этой матрицы стоят оценки дисперсий коэффициентов. На их основе обычным способом можно получить аналоги t-статистик для проверки гипотезы о равенстве отдельного коэффициента нулю. Такой тест будет разновидностью теста Вальда.

Для проверки набора ограничений удобно использовать статистику отношения правдоподобия $LR = 2(\ln L(a) - \ln L(a_R))$, где $\ln L(a)$ — логарифмическая функция правдоподобия из 21.1, a — оценка методом максимума правдоподобия без ограничений, a_R — оценка при ограничениях.

Эту же статистику можно использовать для построения показателя качества модели, аналогичного F-статистике для линейной регрессии. Она позволяет проверить гипотезу о равенстве нулю коэффициентов при всех регрессорах, кроме константы. Соответствующая статистика отношения правдоподобия равна $LR_0 = 2(\ln L(a) - \ln L_0)$, где $\ln L_0$ — максимум логарифмической функции правдоподобия для модели с одной константой. Она распределена асимптотически как χ^2 с n степенями свободы, где n — количество параметров в исходной модели, не включая константу. Величина $\ln L_0$ получается следующим образом. Пусть N — общее количество наблюдений, N_0 — количество наблюдений, для которых $x_i = 0$, N_1 — количество наблюдений, для которых $x_i = 1$. Тогда предсказанная вероятность появления $x_i = 1$ в модели с одной константой будет равна для всех наблюдений N_1/N. Отсюда

$$\ln L_0 = N_0 \ln N_0 + N_1 \ln N_1 - N \ln N.$$

21.2.1. Регрессия с упорядоченной зависимой переменной

Регрессия с упорядоченной зависимой переменной имеет дело с альтернативами, которые можно расположить в определенном порядке. Например, это могут быть оценки, полученные на экзамене, или качество товара, которое может характеризоваться сортом от «высшего» до «третьего». Будем предполагать, что альтернативы пронумерованы от 0 до S. Переменная x принимает значение s, если выбрана альтернатива s. Предполагается, что в основе выбора лежит ненаблюдаемая величина $\tilde{x} = z\alpha + \varepsilon$. При этом $x = 0$ выбирается, если $\tilde{x}$ меньше нижнего (нулевого) порогового значения, $x = 1$, если $\tilde{x}$ попадает в промежуток от нулевого до первого порогового значения и т. д.; $x = S$ выбирается, если $\tilde{x}$ превышает верхнее пороговое значение:

$$x = \begin{cases} 0, & \tilde{x} < \gamma_0, \\ 1, & \gamma_0 < \tilde{x} < \gamma_1, \\ \dots & \\ S, & \tilde{x} > \gamma_{S-1}. \end{cases}$$

Если среди регрессоров z есть константа, то невозможно однозначно идентифицировать γ. В связи с этим следует использовать какую-либо нормировку. Можно, например, положить $\gamma_0 = 0$. Это оставляет $S-1$ неизвестных пороговых параметров.

Пусть $F_\varepsilon(\cdot)$ — функция распределения ошибки ε. Тогда вероятность того, что $x = s$, где $s = 1, \dots, S-1$, равна

$$\begin{aligned} \Pr(x = s) &= \Pr(\gamma_{s-1} < z\alpha + \varepsilon < \gamma_s) = \\ &= \Pr(\gamma_{s-1} - z\alpha < \varepsilon < \gamma_s - z\alpha) = F_\varepsilon(\gamma_s - z\alpha) - F_\varepsilon(\gamma_{s-1} - z\alpha). \end{aligned}$$

Аналогично для $s = 0$ и $s = S$ получаем

$$\begin{aligned} \Pr(x = 0) &= \Pr(\varepsilon < \gamma_0 - z\alpha) = F_\varepsilon(\gamma_0 - z\alpha), \\ \Pr(x = S) &= \Pr(\gamma_{S-1} - z\alpha < \varepsilon) = 1 - F_\varepsilon(\gamma_{S-1} - z\alpha). \end{aligned}$$

Пусть (x_i, z_i), $i = 1, \dots, N$ — имеющиеся наблюдения. По этим наблюдениям можно получить оценки максимального правдоподобия. Обозначим

$$p_{is}(\alpha, \gamma) = \Pr(x_i = s).$$

Соответствующая логарифмическая функция правдоподобия равна

$$\ln L(\alpha,\gamma)=\sum_{s=0}^{S}\sum_{i\in I_s}\ln p_{is}(\alpha,\gamma),\text{ где } I_s=\{i|\,x_i=s\}.$$

Максимизируя эту функцию по α и γ, получим требуемые оценки.

На практике обычно используют одну из двух моделей: **упорядоченный пробит**, то есть модель с нормально распределенным отклонением ε, или **упорядоченный логит**, то есть модель, основанную на логистическом распределении.

21.2.2. Мультиномиальный логит

Предположим, что принимающий решение индивидуум стоит перед выбором из S альтернатив, $s=0,\ldots,S-1$. Предполагается, что выбор делается на основе функции полезности $u(s)$. В линейной модели $u(s)=z_s\alpha_s$, где z_s — матрица факторов, α_s — неизвестные параметры. Обычно есть также факторы, не отраженные в z_s из-за их ненаблюдаемости, которые тоже влияют на полезность. Такие характеристики представлены случайной ошибкой $u(s)=z_s\alpha_s+\varepsilon_s$. При этом x выбирается равным s, если $u(s)>u(t),\ \forall s\neq t$.

В самой простой модели принимается, что ошибки ε_s подчинены распределению экстремального значения и независимы между собой. Распределение экстремального значения[1] в стандартной форме имеет функцию распределения $F(y)=e^{-e^{-y}}$. Распределение экстремального значения обладает следующими важными для рассматриваемой модели свойствами: максимум нескольких величин, имеющих распределение экстремального значения, также имеет распределение экстремального значения, а разность двух величин, имеющих распределение экстремального значения, имеет логистическое распределение. Используя эти свойства, можно вывести, что в данной модели

$$\Pr(x=s)=\frac{e^{z_s\alpha_s}}{\sum_{t=0}^{S-1}e^{z_t\alpha_t}}.$$

Эта модель называется **мультиномиальным логитом**.

Относительно функций $z_s\alpha_s$ обычно делаются какие-либо упрощающие допущения, например, что факторы для всех альтернатив одни и те же, то есть

[1] Это «распределение экстремального значения первого рода» (согласно теореме Гнеденко есть еще два распределения экстремального значения) или, как его еще называют, распределение Гумбеля. Данное распределение также изредка называют распределением Вейбулла. Кроме того, именем Вейбулла называют и другие распределения (в частности, «распределение экстремального значения третьего рода»), поэтому может возникнуть путаница.

$u(s) = z\alpha_s + \varepsilon_s$, или что функция имеет один и тот же вид, коэффициенты в зависимости от s не меняются, а меняются только факторы, определяющие выбор, то есть $u(s) = z_s\alpha + \varepsilon_s$. В первом случае z можно интерпретировать как характеристики индивидуума, принимающего решение. Это собственно мультиномиальный логит. Во втором случае z_s можно интерпретировать как характеристики s-й альтернативы. Этот второй вариант называют **условным логитом**.

Можно предложить модель, которая включает оба указанных варианта. Обозначим через w характеристики индивидуума, а через z_s характеристики s-ой альтернативы (в том числе те, которые специфичны для конкретных индивидуумов). Например, при изучении выбора покупателями супермаркета альтернативами являются имеющиеся супермаркеты, w мог бы включать информацию о доходах и т.п., а в z_s следует включить информацию о супермаркетах (уровень цен, широта ассортимента и т.п.) и характеристики пары покупатель—супермаркет, такие как расстояние до супермаркета от места жительства потребителя.

В такой модели $u(s) = z_s\alpha + w\delta_s + \varepsilon_s$ и вероятности вычисляются по формуле

$$\Pr(x = s) = \frac{e^{z_s\alpha + w\delta_s}}{\sum_{t=0}^{S-1} e^{z_t\alpha + w\delta_t}}.$$

Заметим, что в этой модели есть неоднозначность. В частности, если прибавить к коэффициентам δ_s один и тот же вектор Δ — это все равно, что умножить числитель и знаменатель на $e^w\Delta$. Таким образом, для идентификации модели требуется какая-либо нормировка векторов δ_s. Можно, например, положить $\delta_0 = 0$.

Для оценивания модели используется метод максимального правдоподобия. Пусть $(x_i, z_{i0}, \ldots, z_{i,S-1}, w_i)$, $i = 1, \ldots, N$ — имеющиеся наблюдения. Обозначим

$$p_{is}(\alpha, \delta) = \Pr(x_i = s) = \frac{e^{z_{is}\alpha + w_i\delta_s}}{\sum_{t=0}^{S-1} e^{z_{it}\alpha + w_i\delta_t}}.$$

Тогда логарифмическая функция правдоподобия имеет вид

$$\ln L(\alpha, \delta) = \sum_{s=0}^{S-1} \sum_{i \in I_s} \ln p_{is}(\alpha, \delta), \text{ где } I_s = \{i |\, x_i = s\}.$$

На основе x_i можно ввести набор фиктивных переменных d_{is}, таких что

$$d_{is} = \begin{cases} 1, & x_i = s, \\ 0, & x_i \neq s. \end{cases}$$

В этих обозначениях функция правдоподобия приобретет вид

$$\ln L(\alpha, \delta) = \sum_{i=1}^{N} \sum_{s=0}^{S-1} d_{is} \ln p_{is}(\alpha, \delta).$$

21.2.3. Моделирование зависимости от посторонних альтернатив в мультиномиальных моделях

Для мультиномиального логита отношение вероятностей двух альтернатив («соотношение шансов») равно

$$\frac{\Pr(x = s)}{\Pr(x = t)} = \frac{e^{(z_s\alpha + w\delta_s)}}{e^{(z_t\alpha + w\delta_t)}} = e^{((z_s - z_t)\alpha + w(\delta_s - \delta_t))}.$$

Оно зависит только от характеристик этих двух альтернатив, но не от характеристик остальных альтернатив. Это свойство называется **независимостью от посторонних альтернатив**. Оно позволяет оценивать мультиномиальные модели на подмножестве полного множества альтернатив и получать корректные (состоятельные) оценки. Однако это свойство мультиномиального логита во многих ситуациях выбора не очень реалистично.

Рассмотрим, например, выбор между передвижением на поезде, на самолете авиакомпании A и на самолете авиакомпании B. Известно, что 50% пассажиров выбирает поезд, 25% — авиакомпанию A и 25% — авиакомпанию B. Допустим, авиакомпании предоставляют примерно одинаковые услуги по схожей цене, и пассажиры предпочитают одну из двух авиакомпаний по каким-то чисто субъективным причинам. Если авиакомпании объединятся, то естественно ожидать, что соотношение шансов для поезда и самолета будет равно один к одному. Однако с точки зрения мультиномиального логита соотношение шансов должно остаться два к одному, поскольку характеристики передвижения поездом и передвижения самолетом остались теми же.

Предложено несколько модификаций этой модели, которые уже не демонстрируют независимость от посторонних альтернатив, и, следовательно, более реалистичны.

В модели **вложенного логита** используется иерархическая структура альтернатив. В двухуровневой модели сначала делается выбор между группами альтернатив, а затем делается выбор внутри выбранной группы. В приведенном примере есть две группы альтернатив: «самолет» и «поезд». Внутри группы «самолет» делается выбор между авиакомпаниями A и B. Группа «поезд» содержит только одну альтернативу, поэтому выбор внутри нее тривиален.

Пусть имеется l групп альтернатив. Обозначим через S_k множество альтернатив, принадлежащих k-й группе. Безусловная вероятность того, что будет выбрана

альтернатива s из группы k в модели вложенного логита, определяется формулой (запишем ее только для условного логита, т.е. модели, где от альтернативы зависят факторы, но не коэффициенты)

$$\Pr(x=s)=\frac{e^{(\dot{z}_k\dot{\alpha}+z_s\alpha)}}{\sum_{m=1}^{l}\sum_{t\in S_m}e^{(\dot{z}_m\dot{\alpha}+z_t\alpha)}}=\frac{e^{\dot{z}_k\dot{\alpha}}e^{z_s\alpha}}{\sum_{m=1}^{l}e^{\dot{z}_m\dot{\alpha}}\sum_{t\in S_m}e^{z_t\alpha}}.$$

Если альтернативы s и t принадлежат одной и той же группе k, то отношение вероятностей равно

$$\frac{\Pr(x=s)}{\Pr(x=t)}=\frac{e^{\dot{z}_k\dot{\alpha}}e^{z_s\alpha}}{e^{\dot{z}_k\dot{\alpha}}e^{z_t\alpha}}=\frac{e^{z_s\alpha}}{e^{z_t\alpha}}.$$

Это отношение, как и в обычном мультиномиальном логите, зависит только от характеристик этих альтернатив. В то же время, если альтернативы s и t принадлежат разным группам, k и m соответственно, то отношение вероятностей равно

$$\frac{\Pr(x=s)}{\Pr(x=t)}=\frac{e^{\dot{z}_k\dot{\alpha}}e^{z_s\alpha}}{e^{\dot{z}_m\dot{\alpha}}e^{z_t\alpha}}=\frac{e^{\dot{z}_k\dot{\alpha}+z_s\alpha}}{e^{\dot{z}_m\dot{\alpha}+z_t\alpha}}.$$

Это отношение зависит, кроме характеристик самих альтернатив, также от характеристик групп, к которым они принадлежат.

Другое направление модификации модели мультиномиального логита исходит из того, что независимость от посторонних альтернатив является следствием двух предположений, лежащих в основе модели: то, что ошибки ε_s одинаково распределены и, следовательно, имеют одинаковую дисперсию, и то, что они независимы.

Во-первых, можно предположить, что имеет место гетероскедастичность. (Имеется в виду не гетероскедастичность по наблюдениям, а гетероскедастичность по альтернативам.) Для того чтобы ввести гетероскедастичность в модель, достаточно дополнить распределения ошибок масштабирующими коэффициентами. При этом ошибка ε_s имеет функцию распределения

$$F_s(y)=e^{-e^{-y/\sigma_s}}.$$

Поскольку одновременно все σ_s идентифицировать нельзя, то требуется нормировка. Например, можно принять, что $\sigma_0=1$. С помощью такой модификации мы получим гетероскедастичную модель с распределением экстремального значения.

Во-вторых, можно предположить, что ошибки ε_s могут быть коррелированными друг с другом. Обычно в таком случае используют многомерное нормальное

распределение ошибок:

$$\varepsilon = \begin{pmatrix} \varepsilon_0 \\ \vdots \\ \varepsilon_{S-1} \end{pmatrix} \sim N(0, \Sigma_\varepsilon).$$

Здесь Σ_ε — ковариационная матрица ошибок, которая обычно предполагается неизвестной. С помощью такой модификации мы получим модель **мультиномиального пробита**.

Ковариационная матрица Σ_ε не полностью идентифицирована. Дело в том, что, во-первых, важны разности между ошибками, а не сами ошибки, а во-вторых, ковариационная матрица разностей между ошибками идентифицируется только с точностью до множителя. Можно предложить различные варианты нормировки. Как следствие нормировки, количество неизвестных параметров в матрице Σ_ε существенно уменьшается. Если в исходной матрице их $S(S+1)/2$, то после нормировки остается $S(S-1)/2-1$ неизвестных параметров.

К сожалению, не существует аналитических формул для расчета вероятностей альтернатив в мультиномиальном пробите. Вероятности имеют вид многомерных интегралов. Обозначим через B_s множество таких ошибок ε, которые приводят к выбору s-й альтернативы, т.е.

$$B_s = \{\varepsilon | u(s) > u(t),\ \forall s \neq t\} = \{\varepsilon | z_s\alpha_s + \varepsilon_s > z_t\alpha_t + \varepsilon_t,\ \forall s \neq t\},$$

а через $\varphi(\varepsilon)$ — многомерную плотность распределения ε. Тогда вероятность того, что будет выбрана альтернатива s, равна[2]

$$\Pr(x = s) = \int\limits_{\varepsilon \in B_s} \varphi(\varepsilon) d\varepsilon.$$

Для вычисления таких интегралов, как правило, используется метод Монте-Карло.

21.3. Упражнения и задачи

Упражнение 1

В Таблице 9.3 на стр. 306 приведены данные о голосовании по поводу увеличения налогов на содержание школ в городе Троя штата Мичиган в 1973 г. Наблюдения относятся к 95-ти индивидуумам. Приводятся различные их характеристики:

[2]Реально требуется вычислить не S-мерный интеграл, а $(S-1)$-мерный, поскольку важны не сами ошибки, а разности между ними.

Pub = 1, если хотя бы один ребенок посещает государственную школу, иначе 0; Priv = 1, если хотя бы один ребенок посещает частную школу, иначе 0; Years — срок проживания в данном районе; Teach = 1, если человек работает учителем, иначе 0; LnInc — логарифм годового дохода семьи в долл.; PropTax — логарифм налогов на имущество в долл. за год (заменяет плату за обучение — плата зависит от имущественного положения); Yes = 1, если человек проголосовал на референдуме «за», 0, если «против». Зависимая переменная — Yes. В модель включаются все перечисленные факторы, а также квадрат Years.

1.1. Получите приближенные оценки для логита и пробита с помощью линейной регрессии.

1.2. Оцените логит и пробит с помощью ММП и сравните с предыдущим пунктом.

1.3. Вычислите коэффициенты логита через коэффициенты пробита и сравните с предыдущими результатами.

1.4. На основе оценок МП для логита найдите маргинальные значения для Teach, LnInc и PropTax при среднем уровне факторов.

1.5. Постройте график вероятности голосования «за» в зависимости от Years при среднем уровне остальных факторов.

1.6. Постройте аналогичный график маргинального значения Years.

Упражнение 2

Рассматривается модель мультиномиального логита. В модели имеется три альтернативы: 0, 1 и 2. Для каждой из альтернатив $s = 0,\ 1,\ 2$ полезность рассчитывается по формуле $u_s = z_s\alpha + \beta_s + \varepsilon_s$, где $\alpha = 2,\ \beta_s = s/5$, а ошибки ε_s имеют распределение экстремального значения. Поскольку функция распределения для распределения экстремального значения имеет вид $F(\varepsilon) = e^{-e^{-\varepsilon}}$, то ошибки можно генерировать по формуле $\varepsilon = -\ln(-\ln(\xi))$, ξ имеет равномерное распределение на отрезке $[0; 1]$. Зависимая переменная x принимает одно из трех возможных значений (0, 1 или 2) в зависимости от того, какая полезность выше.

2.1. Пусть $z_1 = 0.4,\ z_2 = 0.3,\ z_3 = 0.2$. Проверить методом Монте-Карло формулу для вероятностей:

$$\Pr(x = s) = \frac{e^{z_s\alpha+\beta_s}}{\sum\limits_{t=0}^{2} e^{z_t\alpha+\beta_t}},$$

сгенерировав выборку из 1000 наблюдений для x и рассчитав эмпирические частоты.

2.2. Сгенерировать данные по модели, взяв $z_s \sim N(0, 2)$ для всех s. Сгенерировав набор из 1000 наблюдений $(x_i, z_{0i}, z_{1i}, z_{2i})$, где $i = 1, \ldots, 1000$, получить оценки параметров модели мультиномиального логита, предполагая, что $\beta_0 = 0$. Сравнить с истинными значениями параметров.

Задачи

1. Чему равны оценки максимального правдоподобия по модели логит с одной константой?

2. Запишите 7 терминов, которые имеют отношение к моделям с качественной зависимой переменной.

3. Рассмотрите модель с биномиальной зависимой переменной x, принимающей значения 0 или 1 и зависящей от фиктивной переменной z, принимающей значения 0 или 1. Модель включает также константу. Данные резюмируются следующей таблицей (в клетках стоят количества соответствующих наблюдений):

	$x = 0$	$x = 1$
$z = 0$	N_{00}	N_{01}
$z = 1$	N_{10}	N_{11}

   а) Пусть в основе модели лежит некоторая дифференцируемая функция распределения $F(\cdot)$, заданная на всей действительной прямой. Найдите $\Pr(x = 1)$ при $z = 0$ и при $z = 1$.

   б) Запишите в компактном виде логарифмическую функцию правдоподобия.

   в) Запишите условия первого порядка для оценок максимального правдоподобия, обозначая $F'(y) = f(y)$.

   г) Для $N_{00} = 15$, $N_{01} = 5$, $N_{10} = 5$, $N_{11} = 15$ получите оценки логита методом максимального правдоподобия.

   д) Для тех же данных получите оценки пробита методом максимального правдоподобия, используя таблицы стандартного нормального распределения.

   е) Как можно определить, значима ли фиктивная переменная z? Запишите формулу соответствующей статистики и укажите, как она распределена.

ж) Получите формулу для приближенных оценок логита методом усреднения (используя линейность отношения шансов для логита). Сравните с формулой для оценок максимального правдоподобия.

4. Изучается зависимость курения среди студентов от пола. В следующей таблице приведены данные по 40 студентам:

Пол	Количество наблюдений	Доля курящих
Муж.	20	0.3
Жен.	20	0.4

Оцените по этим данным модель логит методом максимального правдоподобия. Используйте при этом то, что $\ln 2 = 0.693$, $\ln 3 = 1.099$ и $\ln 11 = 2.398$.

5. Пусть переменная x, принимающая значения 0 или 1, зависит от одного фактора z. Модель включает также константу. Данные приведены в таблице:

x	0	0	1	1	0	1	0	1	0	1
z	1	2	3	4	5	6	7	8	9	10

Запишите для этих данных логарифмическую функцию правдоподобия модели с биномиальной зависимой переменной.

6. Оцените упорядоченный пробит методом максимального правдоподобия по следующим данным:

x	0	1	2	3	4
количество наблюдений	50	40	45	80	35

7. Модель с биномиальной зависимой переменной имеет вид:

$$\tilde{x} = \alpha z + \beta + \varepsilon,$$
$$x = \begin{cases} 1, & \tilde{x} > 0, \\ 0, & \tilde{x} < 0, \end{cases}$$

где z — фиктивная переменная. Связь между x и z задана таблицей (в клетках указано количество наблюдений):

		x	
		0	1
z	0	24	28
	1	32	16

а) Найдите оценки коэффициентов логита и пробита по методу усреднения сгруппированных наблюдений.

б) Найдите оценки максимального правдоподобия.

в) Проверьте значимость модели в целом по статистике отношения правдоподобия.

8. По некоторым данным был оценен ряд моделей с биномиальной зависимой переменной и факторами z_1 и z_2. В таблице приведены результаты оценивания этих моделей методом максимального правдоподобия. В скобках записаны стандартные ошибки коэффициентов. Прочерк означает, что данный фактор не был включен в модель. В последней строке приведено значение логарифмической функции правдоподобия в максимуме.

	Логит				**Пробит**			
	I	**II**	**III**	**IV**	**V**	**IV**	**VII**	**VIII**
Константа	1.87 (0.38)	0.28 (0.20)	1.88 (0.38)	0.28 (0.20)	1.14 (0.21)	0.17 (0.12)	1.16 (0.21)	0.18 (0.12)
Z_1	–0.08 (0.33)	0.0012 (0.19)	—	—	–0.06 (0.19)	0.0011 (0.12)	—	—
Z_2	–2.00 (0.44)	—	–1.99 (0.44)	—	–1.21 (0.24)	—	–1.20 (0.25)	—
ln *L*	–44.4	–68.2	–44.5	–68.3	–44.2	–68.3	–44.4	–68.5

Какую из моделей следует выбрать? Обоснуйте свой ответ.

9. Рассмотрите модель дискретного выбора из двух альтернатив с линейной случайной функцией полезности вида:

$$u(s) = \alpha z^s + \beta + \varepsilon^s, \quad s = 0.1,$$

где все ошибки ε^s имеют равномерное распределение $U[-\gamma, \gamma]$ и независимы по уравнениям и по наблюдениям.

а) Найдите вероятности выбора $s = 0$ и $s = 1$ для такой модели.

б) Объясните, идентифицируемы ли одновременно параметры α, β и γ. Если нет, то предложите идентифицирующую нормировку.

в) Запишите функцию правдоподобия для этой модели.

10. Покажите, что логарифмическая функция правдоподобия для биномиального логита является всюду вогнутой по параметрам. Какие преимущества дает это свойство?

11. Рассмотрите модель дискретного выбора из двух альтернатив: $s = 1$ и $s = 2$, в основе которого лежит случайная полезность $u_i(s) = z_{is}\alpha + \varepsilon_{is}$, предполагая, что ошибки двух альтернатив коррелированы и распределены нормально:

$$\begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right).$$

Какие параметры идентифицируемы? Аргументируйте свой ответ. Предложите нормировки, которые позволят оценить такую модель биномиального пробита. Каким методом можно оценить такой «коррелированный» пробит?

12. Пусть $\Lambda(\cdot), \Phi(\cdot)$ — функции распределения логистического и стандартного нормального распределения соответственно.

а) Покажите, что выпуклая комбинация $F(y) = (1-\alpha)\Lambda(y) + \alpha\Phi(y)$, $\alpha \in [0;1]$, также задает функцию распределения (удовлетворяющую всем должным требованиям).

б) Постройте на основе $F(y)$ модель, которая охватывает как логит, так и пробит.

в) Запишите логарифмическую функцию правдоподобия для такой модели.

г) Запишите условия первого порядка для оценок максимального правдоподобия.

д) Является ли параметр α идентифицируемым? (Аргументируйте свой ответ формально.)

13. Рассмотрите модель дискретного выбора из двух альтернатив с линейной случайной функцией полезности вида:

$$u(s) = z^s\alpha + \varepsilon^s, \quad s = 0.1,$$

где все ошибки ε^0 и ε^1 независимы и их функция распределения имеет вид $F(y) = e^{-e^{-y}}$.

а) Покажите, что

$$\Pr\left(\varepsilon^1 - \varepsilon^0 < y\right) = \frac{e^y}{1+e^y}.$$

б) Найдите вероятности выбора $s = 0$ и $s = 1$ для такой модели. Покажите, что данная модель совпадает с логитом.

14. Пусть в упорядоченном логите зависимая переменная x принимает три значения (0, 1, 2). Найдите, как вероятность того, что $x = 2$, зависит от параметра γ_1 (границы между 1 и 2), т.е. найдите соответствующее маргинальное значение.

15. Выведите формулу оценок максимального правдоподобия для регрессии с упорядоченной зависимой переменной с одной константой. Для количества наблюдений, соответствующих выбору альтернативы s, используйте обозначение N_s. (Подсказка: удобно перейти от исходных параметров к вероятностям $p_s = \Pr(x = s)$.)

16. Рассмотрите использование упорядоченной регрессии для моделирования решения индивидуума о получении образования. Пусть в основе принимаемого решения имеется некоторый индекс, выражающий полезность от образования:

$$U_i = Z_i\alpha + \varepsilon_i, \quad \varepsilon_i \sim N(0; \sigma^2).$$

Чем выше индекс, тем более вероятен выбор более высокого уровня образования. Более конкретно, пусть имеются некоторые известные заранее пороговые значения для индекса, γ_1 и γ_2, такие что:

- при $U_i > \gamma_2$ индивидуум i заканчивает вуз;
- при $\gamma_1 < U_i \leqslant \gamma_2$ индивидуум i заканчивает среднюю школу, но не получает высшего образования;
- при $U_i \leqslant \gamma_1$ индивидуум i получает только неполное среднее образование.

а) Какой вид может иметь зависимая переменная в такой модели?

б) Покажите, что в данной модели нельзя однозначно идентифицировать как β, так и σ.

в) Можно ли однозначно идентифицировать β/σ?

г) Можно ли однозначно идентифицировать β, если положить $\sigma = 1$?

д) Возможно было бы идентифицировать γ_1 и γ_2, если бы они были неизвестны?

е) Запишите функцию правдоподобия для данной модели.

17. В модели регрессии с упорядоченной зависимой переменной альтернативами были числа $s = 0, \ldots, S$. Как поменяются оценки максимального правдоподобия, если альтернативами будут числа $1, 2, 2^2, \ldots, 2^S$? Аргументируйте свой ответ.

18. В выборах участвуют три кандидата: Иванов $(s = 1)$, Петров $(s = 2)$ и «против всех» $(s = 0)$. Перед выборами был проведен опрос населения. Для каждого из опрошенных собраны данные о том, какого он пола (F или M) и за кого собирается голосовать. В результате получено 6 чисел: N_{sF}, N_{sM} $(s = 0, 1, 2)$ — количество женщин и мужчин, собирающихся голосовать за каждого из трех кандидатов. Выведите функцию правдоподобия для соответствующей модели мультиномиального логита.

19. С помощью мультиномиального логита изучается выбор индивидуумами способа передвижения между домом и работой: пешком, на автобусе или на личном автомобиле. Имеются следующие данные: среднее время передвижения от дома до работы для каждого индивидуума каждым из способов и средний доход каждого индивидуума. Введите требуемые обозначения и запишите формулы вероятностей выбора каждого из способов передвижения. Предложите нормировку, которая позволяет идентифицировать модель.

20. Работники кафе быстрого обслуживания «Томато-пицца» могут выбрать один из видов фирменной униформы: брюки или юбку, — причем одного из двух цветов: красного и темно-красного. Какой из моделей вы бы описали такую ситуацию? Объясните.

21. Рассмотрите модель дискретного выбора из трех альтернатив с линейной функцией полезности, соответствующую модели мультиномиального пробита. Предложите нормировки, которые позволят оценить такую модель.

22. В чем состоят преимущества и недостатки мультиномиального пробита по сравнению с мультиномиальным логитом?

Рекомендуемая литература

1. **Cramer J.S.** The Logit Model for Economists. — Adward Arnold, 1991.

2. **Davidson R., MacKinnon J.G.** Estimation and Inference in Econometrics. — Oxford University Press, 1993. (Ch. 15).

3. **Greene W.H.** Econometric Analysis. — Prentice-Hall, 2000. (Ch. 8, 19).

4. **Maddala G.S.** Limited-Dependent and Qualitative Variables in Econometrics, Cambridge University Press, 1983. (Ch. 2, 3, 5).

5. **Wooldridge Jeffrey M.** Introductory Econometrics: A Modern Approach, 2nd ed. — Thomson, 2003. (Ch. 17).

6. **Baltagi, Badi H.** Econometrics, 2nd edition, Springer, 1999 (Ch. 13).

7. **Ruud Paul A.** An Introduction to Classical Econometric Theory, Oxford University Press, 2000 (Ch. 27).

Глава 22

Эффективные оценки параметров модели ARMA

В главе 14 «Линейные стохастические модели ARIMA» были рассмотрены два метода оценки параметров моделей ARMA: линейная регрессия для оценивания авторегрессий и метод моментов для общей модели ARMA. Эти методы не обеспечивают эффективность оценок параметров модели. Можно предложить способы оценивания, которые дают более точные оценки. Для этого можно использовать метод максимального правдоподобия. Рассмотрению этого метода и посвящена данная глава.

22.1. Оценки параметров модели AR(1)

Рассмотрим только случай модели AR(1): $x_t = \varepsilon t + \varphi x_{t-1},\ t = 1, \ldots, T$, на примере которого хорошо видна и общая ситуация.

Чтобы воспользоваться методом максимума правдоподобия, вычислим плотности распределения вероятности наблюдений $x_1, x_2, \ldots, x_T$:

$$f(x_1, \ldots, x_T) = f(x_1) \cdot f(x_2|x_1) \cdot f(x_3|x_2) \cdot \ldots \cdot f(x_T|x_{T-1}).$$

Предположим, что условное распределения x_t при известном x_{t-1} нормально. В соответствии с моделью AR(1) это распределение имеет среднее φx_{t-1} и дис-

персию σ_ε^2. Значит,

$$f(x_1, \ldots, x_T) = f(x_1) \prod_{t=2}^{T} \left((2\pi\sigma_\varepsilon^2)^{-\frac{1}{2}} \cdot e^{-\frac{1}{2\sigma_\varepsilon^2}(x_t - \varphi x_{t-1})^2} \right) =$$

$$= f(x_1)(2\pi\sigma_\varepsilon^2)^{-\frac{T-1}{2}} e^{-\frac{1}{2\sigma_\varepsilon^2} \sum\limits_{t=2}^{T} (x_t - \varphi x_{t-1})^2}.$$

Плотность как функция параметров φ и σ_ε является функцией правдоподобия.

Вместо полной плотности $f(x_1, \ldots, x_T)$ в качестве приближения рассмотрим плотность условного распределения $x_2, \ldots, x_T$, считая x_1 заданным, т.е. будем оперировать с $f(x_2, \ldots, x_T | x_1)$.

Тем самым мы потеряем одну степень свободы. Приближенная функция правдоподобия равна

$$L_-(\varphi, \sigma_\varepsilon^2) = \left(2\pi\sigma_\varepsilon^2\right)^{-\frac{T-1}{2}} e^{-\frac{1}{2\sigma_\varepsilon^2} \sum\limits_{t=2}^{T} (x_t - \varphi x_{t-1})^2} = \left(2\pi\sigma_\varepsilon^2\right)^{-\frac{T-1}{2}} e^{-\frac{1}{2\sigma_\varepsilon^2} s(\varphi)},$$

где $s(\varphi) = \sum\limits_{t=2}^{T} (x_t - \varphi x_{t-1})^2$.

Максимизируя ее по σ_ε^2, выразим σ_ε^2 через φ:

$$\sigma_\varepsilon^2 = \frac{s(\varphi)}{T-1}.$$

Подставляя это выражение в $L_-(\varphi, \sigma_\varepsilon^2)$, получим концентрированную функцию правдоподобия:

$$L_-^c(\varphi) = \left(2\pi \frac{s(\varphi)}{T-1}\right)^{-\frac{T-1}{2}} e^{-\frac{T-1}{2}}.$$

Максимизация $L_-^c(\varphi)$ по φ эквивалентна минимизации суммы квадратов $s(\varphi) = \sum\limits_{t=2}^{T} (x_t - \varphi x_{t-1})^2 = \sum\limits_{t=2}^{T} \varepsilon t^2$. Таким образом, задача сводится к обычному МНК. Минимум этого выражения по φ равен просто

$$\varphi^* = \frac{\sum\limits_{t=2}^{T} x_t x_{t-1}}{\sum\limits_{t=1}^{T-1} x_t^2}.$$

Получили **условную МНК-оценку**. Несложно обобщить этот метод на случай **AR(p)** при $p > 1$.

Мы знаем, что в качестве оценки φ можно использовать выборочную автокорреляцию r_1. Но так как условная МНК-оценка несколько иная, в вырожденных случаях можно получить значения $|\varphi| > 1$. Это можно обойти, учитывая информацию о x_1. Для этого воспользуемся тем, что частное распределение x_1 является нормальным со средним 0 и дисперсией $\sigma^2_{x_1} = \dfrac{\sigma^2_\varepsilon}{1-\varphi^2}$.

Можно воспользоваться здесь взвешенным МНК. Сумма квадратов остатков после преобразования в пространстве наблюдений равна

$$h(\varphi) = \left(1-\varphi^2\right) x_1^2 + \sum_{t=2}^{T} (x_t - \varphi x_{t-1})^2.$$

Получим **точную МНК-оценку**

$$\hat{\varphi}^* = \operatorname*{argmin}_{\varphi} h(\varphi).$$

Плотность частного распределения x_1 равна $f(x_1) = \left(2\pi \frac{\sigma^2_\varepsilon}{1-\varphi^2}\right)^{-\frac{1}{2}} e^{-\frac{(1-\varphi^2)x_1^2}{2\sigma^2_\varepsilon}}$.

Отсюда $f(x_1, \ldots, x_T) = (1-\varphi^2)^{\frac{1}{2}} (2\pi\sigma^2_\varepsilon)^{-\frac{T}{2}} e^{\frac{-1}{2\sigma^2_\varepsilon} h(\varphi)}$. Будем рассматривать эту плотность как функцию правдоподобия, обозначая через $L(\varphi, \sigma^2_\varepsilon)$.

Точную ММП-оценку находим из условия $L(\varphi, \sigma^2_\varepsilon) \to \max\limits_{\sigma^2_\varepsilon, \varphi}$.

Оценкой σ^2_ε будет $h(\varphi)/T$. Концентрируя функцию правдоподобия по σ^2_ε, получим:

$$L^c(\varphi) = \left(1-\varphi^2\right)^{\frac{1}{2}} \left(2\pi \frac{h(\varphi)}{T}\right)^{-\frac{T}{2}} e^{-\frac{T}{2}} \to \max_{\varphi}!$$

Это эквивалентно решению задачи

$$\left(1-\varphi^2\right)^{-\frac{1}{T}} h(\varphi) \to \min_{\varphi}!$$

Отсюда найдем ММП-оценку $\tilde{\varphi}^*$ и $\tilde{\sigma}^{2^*}_\varepsilon = \dfrac{h(\tilde{\varphi}^*)}{T}$.

Множитель $(1-\varphi^2)^{-\frac{1}{T}}$ обеспечивает существование минимума в допустимом интервале $-1 < \varphi < 1$, хотя теперь для нахождения минимума требуются итерационные процедуры. Для таких процедур оценка r_1 может послужить хорошим начальным приближением.

Величина $1-\varphi^2$ не зависит от T, и с ростом T множитель $(1-\varphi^2)^{-\frac{1}{T}}$ стремится к единице. Поэтому этот множитель существенен при малых объемах выборок и $|\varphi|$ близких к 1. При больших T и $|\varphi|$ не очень близких к 1 без него можно

обойтись, соглашаясь с незначительными потерями точности оценки, но сильно сокращая объем вычислений. Этим обстоятельством объясняется использование МНК-оценок.

22.2. Оценка параметров модели MA(1)

Продемонстрировать общий метод оценивания модели MA(p) можно, рассматривая простейший случай модели MA(1): $x_t = \varepsilon_t - \theta\varepsilon_{t-1}$.

Отталкиваясь от наблюдений $x_1, x_2, \ldots, x_T$, воспользуемся методом максимального правдоподобия (ММП). Для этого необходимо вычислить для модели функцию плотности распределения вероятности. Это легче всего сделать, перейдя от последовательности x_t к последовательности ε_t.

$$x_1 = \varepsilon_1 - \theta\varepsilon_0 \Rightarrow \varepsilon_1 = x_1 + \theta\varepsilon_0,$$
$$x_2 = \varepsilon_2 - \theta\varepsilon_1 \Rightarrow x_2 = \varepsilon_2 - \theta x_1 - \theta^2\varepsilon_0 \Rightarrow \varepsilon_2 = x_2 + \theta x_1 + \theta^2\varepsilon_0$$

и так далее.

Получаем систему:

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_T \end{pmatrix} = \begin{pmatrix} \theta \\ \theta^2 \\ \theta^3 \\ \vdots \\ \theta^T \end{pmatrix} \varepsilon_0 + \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \theta & 1 & 0 & \cdots & 0 \\ \theta^2 & \theta & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta^{T-1} & \theta^{T-2} & \theta^{T-3} & \cdots & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_T \end{pmatrix}.$$

Это система уравнений относительно θ. В векторной форме

$$\varepsilon = \tilde{\theta}\varepsilon_0 + Qx, \tag{22.1}$$

где мы обозначили

$$\tilde{\theta} = \begin{pmatrix} \theta \\ \theta^2 \\ \theta^3 \\ \vdots \\ \theta^T \end{pmatrix} \quad \text{и} \quad Q = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \theta & 1 & 0 & \cdots & 0 \\ \theta^2 & \theta & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta^{T-1} & \theta^{T-2} & \theta^{T-3} & \cdots & 1 \end{pmatrix}.$$

Будем полагать, что ε_t — последовательность независимых случайных величин, имеющих одинаковое нормальное распределение со средним 0 и дисперсией σ_ε^2.

Плотность распределения вероятности записывается в виде

$$f(\varepsilon_0, \ldots, \varepsilon_T) = f(\varepsilon_0, x_1, \ldots, x_T) = (2\pi\sigma_\varepsilon^2)^{-\frac{T+1}{2}} e^{-\frac{1}{2\sigma_\varepsilon^2}\sum\limits_{t=0}^{T}\varepsilon_t^2}. \qquad (22.2)$$

Метод максимального правдоподобия заключается в нахождении такого значения θ, при котором достигается максимум (22.2), или, что эквивалентно, достигает минимума сумма квадратов

$$S(\theta|\varepsilon_0) = \sum_{t=0}^{T}\varepsilon_t^2 = \varepsilon_0^2 + \varepsilon'\varepsilon = \left(1 + \tilde{\theta}'\tilde{\theta}\right)\varepsilon_0^2 + 2x'Q'\tilde{\theta}\varepsilon_0 + x'Q'Qx. \qquad (22.3)$$

Необходимо рассмотреть теперь проблему определения величины ε_0. Первый подход — положить $\varepsilon_0 = 0$. Тогда требуется минимизировать $x'Q'QX \to \min\limits_{\theta}!$ Эта нелинейная задача решается разными вычислительными методами. Полученные таким путем решения называется условным МНК-решением, а $\theta^* = \arg\min\limits_{\theta} x'Q'QX$ — **условной МНК-оценкой**.

При втором подходе величина ε_0 вместе с θ входит в число подлежащих минимизации свободных параметров. Так как величина ε_0 входит в выражение для ε_t линейно, то можно частично облегчить оптимизационную процедуру, поставив вместо ε_0 значение $\hat{\varepsilon}_0$, минимизирующее функцию s при данном θ.

То есть на первом шаге решаем задачу $S(\theta|\varepsilon_0) \to \min\limits_{\varepsilon_0}!$

$$\frac{\partial S(\theta|\varepsilon_0)}{\partial\varepsilon_0} = 2(1 + \tilde{\theta}'\tilde{\theta})\varepsilon_0 + 2x'Q'\tilde{\theta} = 0 \quad \Rightarrow \quad \hat{\varepsilon}_0 = -\frac{x'Q'\tilde{\theta}}{1 + \tilde{\theta}'\tilde{\theta}}.$$

Далее, подставим $\hat{\varepsilon}_0 = \hat{\varepsilon}_0(\theta)$ в $S(\theta|\varepsilon_0)$ (22.3) и решим задачу:

$$S\left(\theta|\hat{\varepsilon}_0(\theta)\right) \to \min_{\theta}!$$

$$S\left(\theta|\hat{\varepsilon}_0(\theta)\right) = \left(1 + \tilde{\theta}'\tilde{\theta}\right)\left(\frac{x'Q'\tilde{\theta}}{1 + \tilde{\theta}'\tilde{\theta}}\right)^2 + x'Q'Qx - \frac{2(x'Q'\tilde{\theta})^2}{1 + \tilde{\theta}'\tilde{\theta}} =$$

$$= x'Q'Qx - \frac{(x'Q'\tilde{\theta})^2}{1 + \tilde{\theta}'\tilde{\theta}} \to \min_{\varepsilon_0,\,\theta}!$$

Полученное при таком подходе значение θ^*, минимизирующее функцию $S(\theta)$, называют **точной МНК-оценкой** для θ и $\hat{\varepsilon}_0 = -\dfrac{x'Q^{*\prime}\tilde{\theta}^*}{1 + \tilde{\theta}^{*\prime}\tilde{\theta}^*}$.

Небольшой дополнительный анализ приводит к **точным ММП-оценкам**.

Обозначим $1+\tilde{\theta}'\tilde{\theta}=K$. Функцию правдоподобия можно представить в виде

$$f(\varepsilon_0,\ldots,\varepsilon_T)=f(\varepsilon_0,\,x_1,\ldots,x_T)=$$
$$=\left(\frac{2\pi\sigma_\varepsilon^2}{K}\right)^{-\frac{1}{2}}e^{-\frac{K}{2\sigma_\varepsilon^2}(\varepsilon_0-\hat{\varepsilon}_0)^2}*K^{-\frac{1}{2}}\left(2\pi\sigma_\varepsilon^2\right)^{-\frac{T}{2}}e^{-\frac{1}{2\sigma_\varepsilon^2}S(\theta)},$$

где $S(\theta)=x'Q'Qx-\dfrac{(x'Q'\tilde{\theta})^2}{1+\tilde{\theta}'\tilde{\theta}}$.

Видим, что первая часть этой записи — это функция плотности распределения $\varepsilon_0\in N(\hat{\varepsilon}_0,\,\frac{\sigma_\varepsilon^2}{K})$, т.е первая часть представляет собой условное распределение $f(\varepsilon_0|x_1,\ldots,x_T)$ неизвестного значения ε_0 при известных наблюдениях $x_1,\,x_2,\,\ldots,\,x_T$.

Вторая часть записи — это частная функция плотности распределения вероятности наблюдений $x_1,\,x_2,\,\ldots,\,x_T$, т.е. $f(x_1,\ldots,x_T)$. Действительно, $x=\varepsilon_0 c+D\varepsilon$, где

$$c=\begin{pmatrix}-\theta\\0\\\vdots\\0\end{pmatrix},\quad D=\begin{pmatrix}1&0&\cdots&0\\-\theta&1&\ddots&\vdots\\\vdots&\ddots&\ddots&0\\0&\cdots&-\theta&1\end{pmatrix}.$$

Ковариационная матрица ряда x равна

$$\Gamma=\mathbf{E}(xx')=\mathbf{E}\Big[(\varepsilon_0 c+D\varepsilon)(\varepsilon_0 c+D\varepsilon)'\Big]=$$
$$=\mathbf{E}\left(\varepsilon_0^2cc'\right)+\mathbf{E}\left(D\varepsilon\varepsilon'D'\right)=\sigma_\varepsilon^2(cc'+DD').$$

Обратная к ней:

$$\Gamma^{-1}=\frac{1}{\sigma_\varepsilon^2}(cc'+DD')^{-1}=$$
$$=\frac{1}{\sigma_\varepsilon^2}\left((DD')^{-1}-(DD')^{-1}c\left(1+c'(DD')^{-1}c\right)^{-1}c'(DD')^{-1}\right).$$

Заметим, что $D^{-1}=Q$, $(D'D)^{-1}=Q'Q$ и $-D^{-1}c=-Qc=\tilde{\theta}$. Тогда

$$\Gamma^{-1}=\frac{1}{\sigma_\varepsilon^2}\left(Q'Q-\frac{Q'\theta\theta'Q}{1+\theta'\theta}\right).$$

Определитель ковариационной матрицы Γ:

$$|\Gamma| = \sigma_\varepsilon^{2T}|cc' + DD'| = \sigma_\varepsilon^{2T}\left|1 + c'\left(DD'\right)^{-1}c\right||DD'| = \\ = \sigma_\varepsilon^{2T}\left(1 + c'\left(DD'\right)^{-1}c\right) = \sigma_\varepsilon^{2T}\left(1 + \tilde{\theta}'\tilde{\theta}\right),$$

где мы воспользовались тем, что $|D| = 1$ и $|DD'| = |D||D'| = 1$.

По формуле плотности многомерного нормального распределения

$$f\left(x_1, \ldots, x_T\right) = (2\pi)^{-\frac{T}{2}}|\Gamma|^{-\frac{1}{2}}e^{-\frac{1}{2}x'\Gamma^{-1}x} = \\ = \left(2\pi\sigma\varepsilon^2\right)^{-\frac{T}{2}}\left(1 + \tilde{\theta}'\tilde{\theta}\right)^{-\frac{1}{2}}e^{-\frac{1}{2\sigma_\varepsilon^2}x'\left(Q'Q - \frac{Q'\theta\theta'Q}{1+\theta'\theta}\right)x}.$$

Поэтому $f\left(x_1, \ldots, x_T\right) = K^{-\frac{1}{2}}\left(2\pi\sigma_\varepsilon^2\right)^{-\frac{T}{2}}e^{-\frac{1}{2\sigma_\varepsilon^2}S(\theta)}$.

Итак, необходимо решить задачу:

$$f\left(\varepsilon_0, \ldots, \varepsilon_T\right) = f\left(\varepsilon_0, x_1, \ldots, x_T\right) = \\ = f\left(\varepsilon_0|x_1, \ldots, x_T\right) \cdot f\left(x_1, \ldots, x_T\right) \to \max_{\varepsilon_0,\,\theta}!$$

От ε_0 зависит только первая часть $\Rightarrow \varepsilon_0^* = \hat{\varepsilon}_0$, а задача приобретает вид

$$f\left(x_1, \ldots, x_T\right) \to \max_\theta! \quad \Leftrightarrow \quad K^{-\frac{1}{2}}\left(2\pi\sigma\varepsilon^2\right)^{-\frac{T}{2}}e^{-\frac{1}{2\sigma_\varepsilon^2}S(\theta)} \to \max_{\sigma_\varepsilon^2,}!$$

Нетрудно получить оценку по σ_ε^2: $\hat{\sigma}_\varepsilon^2 = {S(\theta)}/{T}$. Концентрируем функцию правдоподобия:

$$f_c\left(x_1, \ldots, x_T\right) = K^{-\frac{1}{2}}\left(2\pi\frac{S\left(\theta\right)}{T}\right)^{-\frac{T}{2}}e^{-\frac{T}{2}} \to \max_\theta!$$

Собираем в данной функции вместе все, что зависит от θ, и получаем

$$\left(\frac{1}{K^{\frac{1}{T}}S\left(\theta\right)}\right)^{\frac{T}{2}} \cdot \left(\frac{T}{2\pi e}\right)^{\frac{T}{2}} \to \max_\theta! \quad \Leftrightarrow \quad K^{\frac{1}{T}}S\left(\theta\right) \to \min_\theta!$$

Значение $\hat{\theta}^*$, минимизирующее функцию $K^{1/T}S(\theta)$, называют **точной ММП-оценкой**.

Заметим, что функция K зависит лишь от θ, не зависит от наблюдений, и с ростом T величина $K^{1/T}$ стремится к единице. Поэтому этот множитель существенен лишь при малых объемах выборок и в этом случае не представляет труда для вычислений, а при умеренно больших T без него можно обойтись, соглашаясь с незначительным смещением оценки, но сильно сокращая объем вычислений, особенно в случае MA(q). Этим обстоятельством объясняется использование точных МНК-оценок.

22.3. Оценки параметров модели ARMA(p, q)

Рассмотрим модели ARMA(p, q) ряда $\{x_t\}$:

$$x_t - \varphi_1 x_{t-1} - \cdots - \varphi_p x_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}.$$

Будем полагать, что ε_t — последовательность независимых случайных величин, имеющих одинаковое нормальное распределение со средним 0 и дисперсией σ_ε^2.

Через автоковариационную функцию стационарного ARMA-процесса $\gamma_i = E[x_t x_{t-i}]$ можно выразить ковариационную матрицу $\mathbf{x} = (x_1, \ldots, x_T)$

$$\begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{T-1} & \gamma_{T-2} & \cdots & \gamma_0 \end{bmatrix} = \Gamma.$$

Она является симметричной **тёплицевой матрицей** и обозначается как $\tau[\gamma_0, \ldots, \gamma_{T-1}]$.

Так как $x \sim N_T(0, \Gamma)$, то логарифмическая функция правдоподобия процесса равна

$$\ln L(\theta, \varphi, \sigma_\varepsilon^2) = -\frac{T}{2}\ln(2\pi\sigma_\varepsilon^2) - \frac{1}{2}\ln|\Gamma| - \frac{1}{2}x'\Gamma^{-1}x. \tag{22.4}$$

Через r_i обозначаем автоковариацию, нормированную на дисперсию ошибок, т.е. $r_i = \dfrac{\gamma i}{\sigma_\varepsilon^2}$, а через R обозначаем матрицу $\tau\left[\dfrac{\gamma 0}{\sigma_\varepsilon^2}, \ldots, \dfrac{\gamma T - 1}{\sigma_\varepsilon^2}\right]$. В терминах нормированной R логарифмическая функция правдоподобия (22.4) записывается следующим образом:

$$\ln L(\theta, \varphi, \sigma_\varepsilon^2) = -\frac{T}{2}\ln(2\pi\sigma_\varepsilon^2) - \frac{1}{2}\ln|R| - \frac{1}{2\sigma_\varepsilon^2}x'R^{-1}x \to \max!$$

Воспользовавшись условиями первого порядка

$$\frac{\partial \ln L(\theta, \varphi, \sigma_\varepsilon^2)}{\partial \sigma_\varepsilon^2} = 0,$$

получим оценку σ_ε^2 как функцию от θ и φ :

$$\sigma_\varepsilon^2 = \sigma_\varepsilon^2(\theta, \varphi) = \frac{x'R^{-1}x}{T}.$$

Поставив оценку σ_ε^2 в логарифмическую функцию правдоподобия, получим концентрированную функцию правдоподобия

$$\ln L^c(\theta, \varphi) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2} - \frac{1}{2}\ln|R| - \frac{T}{2}\ln\left(\frac{x' R^{-1} x}{T}\right) \to \max!$$

Точные оценки параметров ARMA для процесса x_t можно найти, максимизируя функцию $\ln L^c(\theta, \varphi)$, что делается с помощью численных методов.

22.4. Упражнения и задачи

Упражнение 1

Сгенерируйте ряд длиной 200 наблюдений по модели MA(1) с параметром $\theta_1 = 0.5$ и нормально распределенной неавтокоррелированной ошибкой с единичной дисперсией. По этому ряду оцените модель MA(1) методом моментов (приравнивая теоретические и выборочные автокорреляции), условным методом наименьших квадратов и точным методом максимального правдоподобия. Сравните все эти оценки с истинным значением.

Упражнение 2

Сгенерируйте ряд длиной 200 наблюдений по модели AR(1) с параметром $\varphi_1 = 0.5$ и нормально распределенной неавтокоррелированной ошибкой с единичной дисперсией. По этому ряду оцените модель AR(1) методом моментов (приравнивая теоретические и выборочные автокорреляции), условным методом наименьших квадратов и точным методом максимального правдоподобия. Сравните все эти оценки с истинным значением.

Упражнение 3

Сгенерируйте ряд длиной 200 наблюдений по модели ARMA(1, 1) с параметрами $\varphi_1 = 0.5$ и $\theta_1 = 0.5$ и нормально распределенной неавтокоррелированной ошибкой с единичной дисперсией. По этому ряду оцените модель ARMA(1, 1) методом моментов (приравнивая теоретические и выборочные автокорреляции), условным методом наименьших квадратов и точным методом максимального правдоподобия. Сравните все эти оценки с истинными значениями.

Задачи

1. Какие предположения должны выполняться, чтобы можно было оценить модель MA(1) с помощью метода максимального правдоподобия?

2. Сформулируйте кратко отличие между условной и точной оценкой МНК для модели MA(1) и связь между ними (с пояснением обозначений).

3. Как можно найти оценку параметра для модели AR(1), исходя из предположения, что первое наблюдение не является случайной величиной? Как называется такая оценка?

4. Запишите функцию правдоподобия для модели авторегрессии первого порядка, выделив множитель, который является причиной отличия точной ММП-оценки от условной оценки. Плотности распределения какой величины соответствует этот множитель?

5. Запишите функцию правдоподобия для модели скользящего среднего первого порядка.

Рекомендуемая литература

1. **Бокс Дж., Дженкинс Г.** Анализ временных рядов. Прогноз и управление. (Вып. 1, 2). — М.: «Мир», 1972.

2. **Песаран М., Слейтер Л.** Динамическая регрессия: теория и алгоритмы. — М: «Финансы и статистика», 1984. (Гл. 2–4).

3. **Engle Robert F.** Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation // Econometrica, 50, 1982, 987–1008.

4. **Hamilton James D.** Time Series Analysis. — Princeton University Press, 1994. (Ch. 5).

5. **Judge G.G., Griffiths W.E., Hill R.C., Luthepohl H., Lee T.** Theory and Practice of Econometrics. — New York: John Wiley & Sons, 1985. (Ch. 8).

6. **(*) Справочник по прикладной статистике:** В 2-х т. Т. 2. / Под ред. Э. Ллойда, У. Ледермана. — М.: «Финансы и статистика», 1990. (Гл. 18).

Глава 23

Векторные авторегрессии

23.1. Векторная авторегрессия: формулировка и идентификация

Модели векторной авторегрессии (VAR) представляют собой удобный инструмент для одновременного моделирования нескольких рядов. Векторная авторегрессия — это такая модель, в которой несколько зависимых переменных, и зависят они от собственных лагов и от лагов других переменных. Если в обычной авторегрессии коэффициенты являются скалярами, то здесь следует рассматривать уже матрицы коэффициентов.

В отличие от модели регрессии, в VAR-модели нет нужды делить переменные на изучаемые переменные и независимые факторы. Любая экономическая переменная модели VAR по умолчанию включается в состав изучаемых величин (хотя есть возможность часть переменных рассматривать как внешние к модели, экзогенные).

Отметим, что естественным расширением модели VAR является модель VARMA, включающая ошибку в виде скользящего среднего. Однако модель VARMA не получила очень широкого распространения из-за сложности оценивания. Авторегрессию легче оценивать, так как выполнено предположение об отсутствии автокорреляции ошибок. В то же время, члены скользящего среднего приходится оценивать методом максимального правдоподобия. Так как каждый обратимый процесс скользящего среднего может быть представлен в виде AR(∞), чистые авторегрессии могут приближать векторные процессы скользящего среднего, если

добавить достаточное число лагов. Предполагается, что при этом ошибка не будет автокоррелированной, что позволяет с приемлемой точностью моделировать временные ряды, описываемые моделью VARMA, при помощи авторегрессии достаточно высокого порядка.

Пусть x_t — вектор-строка k изучаемых переменных, z_t — вектор-строка независимых факторов (в него может входить константа, тренд, сезонные переменные и т.п.).

Как и традиционные системы одновременных уравнений, модели векторной авторегрессии имеют две формы записи: структурную и приведенную. **Структурная векторная авторегрессия** (SVAR) p-го порядка — это модель следующего вида:

$$x_t = \sum_{j=0}^{p} x_{t-j}\Phi_j + z_t A + \varepsilon_t, \text{ где } (\Phi_0)_{ll} = 0.$$

Здесь Φ_j — матрица $k \times k$ коэффициентов авторегрессии для j-го лага x_t, A — матрица коэффициентов при независимых факторах. Коэффициенты, относящиеся к отдельному уравнению, стоят по столбцам этих матриц. Относительно матрицы Φ_j предполагается, что ее диагональные элементы[1] равны нулю, $(\Phi_0)_{ll} = 0$, $l = 1, \ldots, k$. Это означает, что отдельная переменная x_{lt} не влияет сама на себя в тот же момент времени.

При этом предполагается, что ковариационная матрица одновременных ошибок диагональна:

$$\mathbf{var}(\varepsilon_t) = \operatorname{diag}(\omega_1^2, \ldots, \omega_k^2) = \Omega. \tag{23.1}$$

Некоторые из коэффициентов здесь известны, поэтому такая модель называется структурной.

Обозначим

$$B = I - \Phi_0, \; B_{ll} = 1.$$

Тогда SVAR можно переписать как

$$x_t B = \sum_{j=1}^{p} x_{t-j}\Phi_j + z_t A + \varepsilon_t. \tag{23.2}$$

[1] Если матрица имеет индекс, то для обозначения ее элемента мы будем заключать матрицу в скобки. Например, $(A_1)_{ij}$.

Структурная векторная регрессия фактически является «гибридом» моделей авторегрессии и систем одновременных уравнений. Соответственно, анализ таких моделей должен учитывать и динамические свойства, характерные для моделей авторегрессии, и черты, присущие системам одновременных уравнений.

Уравнение структурной векторной авторегрессии представляет собой систему одновременных регрессионных уравнений, в которой среди факторов имеются лаги изучаемых переменных. Для того чтобы показать это в явном виде, введем следующие обозначения:

$$\tilde{z}_t = (x_{t-1}, \dots, x_{t-p}, z_t) \quad \text{и} \quad \tilde{A} = \begin{pmatrix} \Phi_1 \\ \vdots \\ \Phi_p \\ A \end{pmatrix}.$$

В таких обозначениях

$$x_t B = \tilde{z}_t \tilde{A} + \varepsilon_t,$$

или в матричной записи

$$XB = \tilde{Z}\tilde{A} + \varepsilon.$$

Как и в случае систем одновременных уравнений, нельзя оценить параметры структурной формы методом непосредственно наименьших квадратов, поскольку, если матрица B недиагональна, найдутся уравнения, в которых будет более чем одна эндогенная переменная. В i-м уравнении системы будет столько же эндогенных переменных, сколько ненулевых элементов в i-м столбце матрицы B. Таким образом, в общем случае уравнения системы будут взаимозависимы, и, следовательно, оценки их по отдельности методом наименьших квадратов будут несостоятельными.

Классический частный случай, в котором все-таки можно применять МНК — это случай **рекурсивной системы**. Рекурсивной является система одновременных уравнений, в которой матрица B является верхней треугольной, а матрица ковариаций ошибок Ω (23.1) является диагональной. Последнее условие в случае SVAR выполнено по определению. При этом первая переменная зависит только от экзогенных переменных, вторая переменная зависит только от первой и от экзогенных переменных и т.д. Поскольку ошибки в разных уравнениях некоррелированы, то каждая эндогенная переменная коррелирована только с ошибками из своего

и предыдущих уравнений и не коррелирована с ошибками тех уравнений, в которые она входит в качестве регрессора. Таким образом, ни в одном из уравнений не нарушаются предположения МНК о некоррелированности ошибки и регрессоров, т.е. оценки МНК состоятельны.

В общем случае, когда модель SVAR не обязательно рекурсивная, чтобы избавиться от одновременных зависимостей, можно умножить[2] 23.2 справа на B^{-1}:

$$x_t = \sum_{j=1}^{p} x_{t-j}\Phi_j B^{-1} + z_t A B^{-1} + \varepsilon_t B^{-1}.$$

Далее, обозначим

$$D = AB^{-1}, \quad \Pi_j = \Phi_j B^{-1}, \quad v_t = \varepsilon_t B^{-1}.$$

Это дает **приведенную форму** векторной авторегрессии:

$$x_t = \sum_{j=1}^{p} x_{t-j}\Pi_j + z_t D + v_t.$$

Ковариационная матрица одновременных ошибок приведенной формы равна $\mathbf{var}(v_t) = \Sigma$. Она связана с ковариационной матрицей одновременных ошибок структурной формы Ω (см. 23.1) соотношением $B'\Sigma B = \Omega$.

Как и в случае обычных одновременных уравнений, при оценивании структурных векторных авторегрессий возникает **проблема идентификации**. Существует несколько типов **идентифицирующих ограничений**, которые можно использовать для решения этой проблемы.

1) Нормирующие ограничения, которые только закрепляют единицы измерения коэффициентов. В данном случае в качестве нормирующих ограничений используются ограничения $B_{ll} = 1$ (диагональные элементы матрицы B равны 1).

2) Ограничения на коэффициенты структурных уравнений. Ограничения на коэффициенты бывают двух видов: ограничение на коэффициенты в пределах одного и того же уравнения (важный частный случай такого ограничения — исключение переменной из уравнения) и ограничение на коэффициенты нескольких уравнений.

3) Ограничения на ковариационную матрицу ошибок. В структурной векторной авторегрессии используется крайний случай таких ограничений: матрица ковариаций ошибок Ω в этой модели диагональна (см. 23.1), т.е. так называемое ограничение ортогональности ошибок.

4) Долгосрочные ограничения. Это ограничения на долгосрочные взаимодействия переменных, резюмируемые долгосрочным мультипликатором M, о котором речь пойдет ниже (см. 23.5).

[2] Мы исходим из предположения, что B — неособенная матрица.

В отличие от векторной авторегрессии, в классических системах одновременных уравнений редко используют ограничения на ковариационную матрицу, а здесь они входят в определение модели, причем в виде жесткого ограничения ортогональности ошибок.

Стандартные идентифицирующие ограничения, которые неявно подразумевались в ранних статьях по векторной авторегрессии, состоят в том, что матрица B является верхней треугольной. Это дает рекурсивную векторную авторегрессию.

Рекурсивную векторную авторегрессию можно оценить методом наименьших квадратов по причинам, о которых упоминалось выше. Другой способ состоит в том, чтобы оценить приведенную форму модели и восстановить из нее коэффициенты структурной формы. Для этого надо использовать так называемое разложение Холецкого (триангуляризацию) для ковариационной матрицы приведенной формы: $\Sigma = U'\Omega U$, где Ω — диагональная матрица с положительными элементами, U — верхняя треугольная матрица с единицами на диагонали. Естественно, вместо истинной матрицы Σ используют ее оценку. Тогда полученная матрица Ω будет оценкой ковариационной матрицы ошибок структурной формы, а U^{-1} — оценкой матрицы B.

Однако использование рекурсивной векторной авторегрессии нежелательно, если только нет каких-либо оснований считать, что одновременные взаимодействия между переменными действительно являются рекурсивными. Дело в том, что эти идентифицирующие ограничения совершенно произвольны и зависят от того, в каком порядке расположены переменные в векторе x_t.

В общем случае оценивание структурной VAR производят примерно теми же методами, что и оценивание одновременных уравнений. В частности, можно использовать метод максимального правдоподобия. Специфичность методов оценивания состоит в том, что они должны учитывать ограничение ортогональности ошибок.

23.2. Стационарность векторной авторегрессии

Чтобы анализировать условия и следствия стационарности векторной авторегрессии, удобно отвлечься от структурной формы этой модели и пользоваться приведенной формой. Для упрощения анализа мы без потери общности будем рассматривать векторную авторегрессию без детерминированных членов:

$$x_t = \sum_{j=1}^{p} x_{t-j}\Pi_j + v_t,$$

или в операторном виде[3]:

$$x_t\Big(I - \sum_{j=1}^{p} \Pi_j L^j\Big) = v_t.$$

Многие свойства процесса VAR(p) можно получить из свойств процесса VAR(1), если воспользоваться соответствующим представлением:

$$\tilde{x}_t = \tilde{x}_{t-1}\tilde{\Pi} + \tilde{v}_t,$$

где вводятся следующие обозначения:

$$\begin{aligned} \tilde{x}_t &= (x_t, x_{t-1}, \ldots, x_{t-p+1}), \\ \tilde{v}_t &= (v_t, 0_k', \ldots, 0_k') \end{aligned}$$

и

$$\tilde{\Pi} = \begin{pmatrix} \Pi_1 & I_k & 0_{k\times k} & \cdots & 0_{k\times k} \\ \Pi_2 & 0_{k\times k} & I_k & \cdots & 0_{k\times k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Pi_{p-1} & 0_{k\times k} & 0_{k\times k} & \cdots & I_k \\ \Pi_p & 0_{k\times k} & 0_{k\times k} & \cdots & 0_{k\times k} \end{pmatrix}.$$

Используя рекуррентные подстановки

$$\begin{aligned} \tilde{x}_t &= (\tilde{x}_{t-2}\tilde{\Pi} + \tilde{v}_t)\tilde{\Pi} + \tilde{v}_t = \tilde{x}_{t-2}\tilde{\Pi}^2 + \tilde{v}_t\tilde{\Pi} + \tilde{v}_t, \\ \tilde{x}_t &= (\tilde{x}_{t-3}\tilde{\Pi} + \tilde{v}_t)\tilde{\Pi}^2 + \tilde{v}_t\tilde{\Pi} + \tilde{v}_t = \tilde{x}_{t-3}\tilde{\Pi}^3 + \tilde{v}_t\tilde{\Pi}^2 + \tilde{v}_t\tilde{\Pi} + \tilde{v}_t \end{aligned}$$

и т.д., несложно получить для VAR(1) представление в виде бесконечного скользящего среднего:

$$\tilde{x}_t = \tilde{v}_t + \tilde{v}_t\tilde{\Pi} + \tilde{v}_t\tilde{\Pi}^2 + \tilde{v}_t\tilde{\Pi}^3 + \ldots = \sum_{i=0}^{\infty} \tilde{v}_{t-i}\tilde{\Pi}^i.$$

Для того чтобы этот ряд сходился, необходимо, чтобы его члены затухали, т.е. чтобы в пределе при $i \to \infty$ последовательность матриц $\tilde{\Pi}^i$ стремилась к нулю. Для этого требуется, чтобы собственные значения матрицы $\tilde{\Pi}$ лежали внутри

[3] Здесь оператор стоит после переменной, на которую действует, чтобы не нарушать правила умножения матриц.

единичного круга. Собственные значения матрицы $\tilde{\Pi}$, по определению, удовлетворяют уравнению:

$$\left|\tilde{\Pi} - \lambda I_{Tp}\right| = 0.$$

Определитель в этой формуле можно выразить через матрицы Π_j (доказательство этого требует довольно громоздких вычислений):

$$\left|\tilde{\Pi} - \lambda I_{Tp}\right| = (-1)^{Tp}\left|I_T\lambda^p - \Pi_1\lambda^{p-1} - \ldots - \Pi_{p-1}\lambda - \Pi_p\right|.$$

Таким образом, уравнение для собственных значений эквивалентно следующему:

$$\left|I_T\lambda^p - \Pi_1\lambda^{p-1} - \ldots - \Pi_{p-1}\lambda - \Pi_p\right| = 0.$$

Процесс VAR(p) слабо стационарен тогда и только тогда, когда корни этого уравнения меньше единицы по абсолютной величине.

Эти условия стационарности можно переформулировать в терминах матричного характеристического многочлена процесса VAR(p), который равен

$$\Pi(z) = I - \sum_{j=1}^{p}\Pi_j z^j.$$

Если возьмем определитель этого многочлена, то получится скалярный характеристический многочлен

$$|\Pi(z)| = \left|I - \sum_{j=1}^{p}\Pi_j z^j\right|.$$

Он будет многочленом, поскольку определитель — это многочлен от своих элементов. Соответствующее **характеристическое уравнение** имеет вид:

$$|\Pi(z)| = 0.$$

Условия стационарности состоят в том, что корни этого характеристического уравнения лежат за пределами единичного круга.

23.3. Анализ реакции на импульсы

Для содержательной интерпретации стационарной векторной авторегрессии следует выразить изучаемую переменную x_t через ошибки ε_t структурной формы, которые, по определению модели, взаимно некоррелированы.

Запишем приведенную форму модели (без детерминированных членов) с использованием лагового оператора L:

$$x_t\Big(I - \sum_{j=1}^{p} \Pi_j L^j\Big) = v_t = \varepsilon_t B^{-1}.$$

В предположении стационарности процесса x_t можно обратить лаговый полином и получить

$$x_t = \varepsilon_t B^{-1}\Big(I - \sum_{j=1}^{p} \Pi_j L^j\Big)^{-1}. \tag{23.3}$$

Это дает представление в виде бесконечного скользящего среднего (представление Вольда) для VAR:

$$x_t = \sum_{i=0}^{\infty} \varepsilon_{t-i} \Psi_i. \tag{23.4}$$

Матрицы Ψ_i представляют так называемую **функцию реакции на импульсы** (IRF — impulse response function) для структурной векторной авторегрессии и могут быть символически записаны в виде

$$\Psi_i = \frac{dx_t}{d\varepsilon_{t-i}}.$$

Более точно, функция реакции на импульсы — это последовательность $(\Psi_i)_{lr}$, $i = 0, 1, 2,$ где l и r — индексы пары изучаемых переменных. Величина $(\Psi_i)_{lr}$ показывает, как влияет ошибка ε_{tl} (которая соответствует уравнению для переменной x_{tl}) на переменную x_{tr} при запаздывании на i периодов.

Эти матрицы можно рассчитать рекуррентно:

$$\Psi_i = \sum_{j=1}^{p} \Psi_{i-j} \Pi_j, \; i = 1, 2, \ldots,$$

начиная с

$$\Psi_0 = B^{-1} \quad \text{и} \quad \Psi_i = 0_{k \times k}, \; i < 0.$$

Накопленная реакция на импульсы определяется следующим образом:

$$\breve{\Psi}_s = \sum_{i=0}^{s} \Psi_i.$$

Она показывает суммарное запаздывающее влияние ошибок на изучаемую переменную для всех лагов от 0 до некоторого s.

Долгосрочное влияние резюмируется матрицей M, определяемой как

$$M = \lim_{s\to\infty} \breve{\Psi}_s = \sum_{i=0}^{\infty} \Psi_i. \tag{23.5}$$

Эту матрицу называют **долгосрочным мультипликатором**. Ее также можно записать в виде

$$M = B^{-1}\Big(I - \sum_{j=1}^{p} \Pi_j\Big)^{-1}.$$

Последняя формула следует из того, что

$$B^{-1}\Big(I - \sum_{j=1}^{p} \Pi_j\Big)^{-1} = \sum_{i=0}^{\infty} \Psi_i L^i \quad \text{(см. 23.3 и 23.3)}.$$

Для того чтобы это показать, надо подставить 1 вместо L.

23.4. Прогнозирование с помощью векторной авторегрессии и разложение дисперсии

Поскольку лаги исследуемых переменных полагаются величинами известными, то построение прогнозов по ним в гораздо меньшей степени, чем в системах одновременных уравнений, осложняется проблемой получения точных значений факторов.

Для упрощения формул мы будем исходить из того, что нам известны истинные параметры процесса. Пусть известны значения x_t временного ряда VAR для $t = 1, \ldots, T$. Сделаем прогноз на $(T+1)$-й период. Это математическое ожидание x_{T+1}, условное относительно имеющейся на момент T информации $x_1, \ldots, x_T$. При расчетах удобно действовать так, как если бы была известна вся предыстория процесса:

$$\Omega_T = (x_T, \ldots, x_1, x_0, \ldots).$$

Выводы от этого не изменятся. Таким образом, будем использовать ожидания, условные относительно Ω_T.

Искомый прогноз равен

$$x_T^p(1) = \mathbf{E}(x_{T+1}|\Omega_T) = \sum_{j=1}^{p} \mathbf{E}(x_{T+1-j}|\Omega_T)\Pi_j + \mathbf{E}(v_{T+1}|\Omega_T) = \sum_{j=1}^{p} x_{T+1-j}\Pi_j,$$

где мы воспользовались тем, что $\mathbf{E}(v_{T+1}|\Omega_T) = 0$ и что все x_t в правой части уравнения регрессии входят в предысторию Ω_T.

Чтобы получить формулу прогноза на s периодов, возьмем от обеих частей уравнения для процесса VAR математическое ожидание, условное относительно Ω_T. Получим

$$x_T^p(s) = \mathbf{E}(x_{T+s}|\Omega_T) = \sum_{j=1}^{p} \mathbf{E}(x_{T+s-j}|\Omega_T)\Pi_j.$$

По этой формуле прогнозы вычисляются рекуррентно, причем $\mathbf{E}(x_t|\Omega_T) = x_t$ при $t \leqslant T$, и $\mathbf{E}(x_t|\Omega_T) = x_T^p(t-T)$ при $t > T$.

Заметим, что построение прогнозов не требует знания структурной формы модели. Таким образом, чтобы построить прогноз, достаточно оценить приведенную форму без наложения ограничений обычным МНК. Это делает VAR очень удобным инструментом прогнозирования: не требуется анализировать, как взаимосвязаны переменные, какая переменная на какую влияет.

Ошибка прогноза — это

$$d_s = x_{T+s} - x_{T+s}^p = x_{T+s} - \mathbf{E}(x_{T+s}|\Omega_T).$$

Чтобы найти эту ошибку, воспользуемся разложением Вольда для x_{T+k}:

$$x_{T+s} = \sum_{i=0}^{\infty} \varepsilon_{T+s-i}\Psi_i.$$

Очевидно, что ошибки $\varepsilon_{T+1}, \ldots, \varepsilon_{T+s}$ непредсказуемы и их ожидаемые значения относительно Ω_T равны нулю, поэтому

$$x_T^p(s) = \mathbf{E}(x_{T+s}|\Omega_T) = \sum_{i=s}^{\infty} \varepsilon_{T+s-i}\Psi_i.$$

Таким образом, ошибка прогноза равна:

$$d_s = \sum_{i=0}^{s-1} \varepsilon_{T+s-i}\Psi_i.$$

Прогноз является несмещенным, поскольку $\mathbf{E}(d) = 0$.

Ковариационная матрица ошибки прогноза находится по формуле:

$$\Sigma_{d_s} = \mathbf{E}(d_s' d_s|\Omega_T) = \mathbf{E}\left(\left(\sum_{i=0}^{s-1} \varepsilon_{T+s-i}\Psi_i\right)' \left(\sum_{i=0}^{s-1} \varepsilon_{T+s-i}\Psi_i\right)|\Omega_T\right) =$$

$$= \sum_{i=0}^{s-1} \Psi_i' \mathbf{E}(\varepsilon'_{T+s-i}\varepsilon_{T+s-i}|\Omega_T)\Psi_i = \sum_{i=0}^{s-1} \Psi_i'\Omega\Psi_i.$$

При выводе формулы мы воспользовались тем, что ошибки структурной формы ε_t не автокоррелированы и их ковариационная матрица равна $\mathbf{var}(\varepsilon_t) = \Omega$. Можно выразить ковариационную матрицу ошибки прогноза также и через ковариационную матрицу ошибок приведенной формы:

$$\Sigma_{d_s} = \sum_{i=0}^{s-1} \Psi_i' B' \Sigma B \Psi_i.$$

Поскольку в структурной форме ошибки разных уравнений некоррелированы, т.е. $\Omega = \mathrm{diag}(\omega_1^2, \ldots, \omega_k^2)$, то можно разложить ковариационную матрицу ошибки прогноза на составляющие, соответствующие отдельным ошибкам ε_{tl}. Обозначим l-ю строку матрицы Ψ_i через ψ_{il}. Вектор ψ_{il} представляет собой функцию реакции на импульсы влияния $\varepsilon_{t-i,l}$ на все переменные x_t. Тогда

$$\Sigma_{d_s} = \sum_{i=0}^{s-1} \Psi_i' \Omega \Psi_i = \sum_{i=0}^{s-1} \sum_{l=1}^{k} \psi_{il}' \omega_l^2 \psi_{il} = \sum_{l=1}^{k} \sum_{i=0}^{s-1} \psi_{il}' \omega_l^2 \psi_{il}.$$

Таким образом, ошибке l-го уравнения соответствует вклад в ковариационную матрицу ошибки прогноза равный

$$\sum_{i=0}^{s-1} \psi_{il}' \omega_l^2 \psi_{il}.$$

Можно интерпретировать j-й диагональный элемент этой матрицы как вклад ошибки l-го уравнения $\varepsilon_{t,l}$ в дисперсию ошибки прогноза j-й изучаемой переменной $x_{t,j}$ при прогнозе на s периодов. Обозначим этот вклад через $\sigma_{jl,s}^2$:

$$\sigma_{jl,s}^2 = \left(\sum_{i=0}^{s-1} \psi'_{\ il} \omega_l^2 \psi_{il} \right)_{jj}.$$

Можем вычислить также долю каждой из ошибок в общей дисперсии ошибки прогноза j-й изучаемой переменной:

$$R_{jl,s}^2 = \frac{\sigma_{jl,s}^2}{\sum\limits_{r=1}^{k} \sigma_{jr,s}^2}.$$

Набор этих долей $R_{jl,s}^2$, где $l = 1, \ldots, k$, представляет собой так называемое **разложение дисперсии ошибки прогноза** для структурной векторной авторегрессии.

23.5. Причинность по Грейнджеру

В эконометрике наиболее популярной концепцией причинности является **причинность по Грейнджеру**. Это связано, прежде всего, с ее относительной простотой, а также с относительной легкостью определения ее на практике.

Причинность по Грейнджеру применяется к компонентам стационарного векторного случайного процесса: может ли одна из этих переменных быть причиной другой переменной. В основе определения лежит хорошо известный постулат, что будущее не может повлиять на прошлое.

Этот постулат Грейнджер рассматривал в информационном аспекте. Для того чтобы определить, является ли переменная x причиной переменной y, следует выяснить, какую часть дисперсии текущего значения переменной y можно объяснить прошлыми значениями самой переменной y и может ли добавление прошлых значений переменной x улучшить это объяснение. Переменную x называют причиной y, если x помогает в предсказании y с точки зрения уменьшения дисперсии. В контексте векторной авторегрессии переменная x будет причиной y, если коэффициенты при лагах x статистически значимы. Заметим, что часто наблюдается двухсторонняя причинная связь: x является причиной y, и y является причиной x.

Рассмотрим причинность по Грейнджеру для двух переменных. Приведенная форма модели имеет вид:

$$x_t = \sum_{j=1}^{p} a_j x_{t-j} + \sum_{j=1}^{p} b_j y_{t-j} + v_t,$$
$$y_t = \sum_{j=1}^{p} c_j x_{t-j} + \sum_{j=1}^{p} d_j y_{t-j} + w_t.$$

Отсутствие причинной связи от x к y означает, что $c_j = 0$ при $j = 1, \ldots, p$, т.е. что прошлые значения x не влияют на y. Отсутствие причинной связи от y к x означает, что $b_j = 0$ при $j = 1, \ldots, p$.

Когда процесс стационарен, тогда гипотезы о причинной связи можно проверять с помощью F-статистики. Нулевая гипотеза заключается в том, что одна переменная не является причиной по Грейнджеру для другой переменной. Длину лага p следует выбрать по самому дальнему лагу, который еще может помочь в прогнозировании.

Следует понимать, что причинность по Грейнджеру — это не всегда то, что принято называть причинностью в общем смысле. Причинность по Грейнджеру связана скорее с определением того, что предшествует чему, а также с информативностью переменной с точки зрения прогнозирования другой переменной.

23.6. Коинтеграция в векторной авторегрессии

Векторная авторегрессия представляет собой также удобный инструмент для моделирования нестационарных процессов и коинтеграции между ними (о коинтеграции см. в гл. 17).

Предположим, что в векторной авторегрессии x_t, задаваемой уравнением

$$x_t = \sum_{j=1}^{p} x_{t-j}\Pi_j + v_t, \tag{23.6}$$

отдельные составляющие процессы x_{tj} либо стационарны, I(0), либо интегрированы первого порядка, I(1). Рассмотрим в этой ситуации коинтеграцию CI(1, 0). Для упрощения забудем о том, что согласно точному определению коинтегрированные вектора сами по себе должны быть нестационарными. Линейную комбинацию стационарных процессов по этому упрощающему определению тоже будем называть коинтегрирующей комбинацией. Таким образом, будем называть коинтегрирующим вектором рассматриваемого процесса векторной авторегрессии x_t такой вектор $c \neq 0$, что $x_t c$ является I(0).

Если векторный процесс состоит из более чем двух процессов, то может существовать несколько коинтегрирующих векторов.

Поскольку коинтегрирующая комбинация — это линейная комбинация, то как следствие, любая линейная комбинация коинтегрирующих векторов, не равная нулю, есть опять коинтегрирующий вектор. В частности, если c_1 и c_2 — два коинтегрирующих вектора и $c = \alpha_1 c_1 + \alpha_2 c_2 \neq 0$, то c — тоже коинтегрирующий вектор. Таким образом, коинтегрирующие вектора фактически образуют линейное подпространство с выколотым нулем, которое принято называть **коинтегрирующим подпространством**.

Обозначим через β матрицу, соответствующую произвольному базису коинтегрирующего подпространства процесса x_t. Это $k \times r$ матрица, где r — размерность коинтегрирующего подпространства. Размерность r называют **рангом коинтеграции**. Столбцы β — это линейно независимые коинтегрирующие вектора.

Для удобства анализа преобразуем исходную модель (23.6) векторной авторегрессии. Всегда можно переписать векторную авторегрессию в форме **векторной модели исправления ошибок** (vector error-correction model, VECM):

$$\Delta x_t = x_{t-1}\Pi + \sum_{j=1}^{p-1} \Delta x_{t-j}\Gamma_j + v_t,$$

где $\Gamma_j = -\sum_{i=j+1}^{p} \Pi_i$, $\Pi = -(I - \sum_{i=1}^{p} \Pi_i) = -\Pi(1)$.

При сделанных нами предположениях первые разности Δx_t должны быть стационарными. Отсюда следует, что процесс $x_{t-1}\Pi = \Delta x_t - \sum_{j=1}^{p-1} \Delta x_{t-j}\Gamma_j - v_t$ стационарен как линейная комбинация стационарных процессов. Это означает, что столбцы матрицы Π — это коинтегрирующие вектора (либо нулевые вектора). Любой такой вектор можно разложить по базису коинтегрирующего подпространства, β. Составим из коэффициентов таких разложений $k \times r$ матрицу α, так что $\Pi = \beta\alpha'$. Ранг матрицы Π не может превышать r. Укажем без доказательства, что при сделанных предположениях ранг матрицы Π в точности равен r.

Таким образом, мы получили следующую запись для векторной модели исправления ошибок:

$$\Delta x_t = x_{t-1}\beta\alpha' + \sum_{j=1}^{p-1} \Delta x_{t-j}\Gamma_j + v_t,$$

где α отвечает за скорость исправления отклонений от равновесия (матрица корректирующих коэффициентов), β — матрица коинтеграционных векторов.

Можно рассмотреть два крайних случая: $r = 0$ и $r = k$. Если $r = 0$, то $\Pi = 0$ и не существует стационарных линейных комбинаций процесса x_t. Если $r = k$, то Π имеет полный ранг и любая комбинация x_t стационарна, т.е. все составляющие процессы являются I(0). О собственно коинтеграции можно говорить лишь при $0 < r < k$.

До сих пор мы не вводили в модель детерминированные компоненты. Однако, вообще говоря, можно ожидать, что в модель входят константы и линейные тренды, причем они могут содержаться как в самих рядах, так и в коинтеграционных уравнениях. Рассмотрим векторную модель исправления ошибок с константой и трендом:

$$\Delta x_t = \mu_0 + \mu_1 t + x_{t-1}\beta\alpha' + \sum_{j=1}^{p-1} \Delta x_{t-j}\Gamma_j + v_t,$$

где μ_0 и μ_1 — вектора-строки длиной k. Вектор μ_0 соответствует константам, а вектор μ_1 — коэффициентам линейных трендов. Можно выделить пять основных случаев, касающихся статуса векторов μ_0 и μ_1 в модели. В таблице 23.1 они перечислены в порядке перехода от частного к более общему.

Здесь γ_0 и γ_1 — вектора-строки длины r.

Случай 0 легко понять — константы и тренды в модели полностью отсутствуют.

В случае 1^* константа входит в коинтеграционное пространство и тем самым в корректирующие механизмы, но не входит в сам процесс x_t в виде дрейфа. Это

Таблица 23.1

Случай 0	$\mu_0 = 0$	$\mu_1 = 0$
Случай 1*	$\mu_0 = \gamma_0\alpha'$	$\mu_1 = 0$
Случай 1	μ_0 произвольный	$\mu_1 = 0$
Случай 2*	μ_0 произвольный	$\mu_1 = \gamma_1\alpha'$
Случай 2	μ_0 произвольный	μ_1 произвольный

несложно увидеть, если переписать модель следующим образом:

$$\Delta x_t = (\gamma_0 + x_{t-1}\beta)\,\alpha' + \sum_{j=1}^{p-1} \Delta x_{t-j}\Gamma_j + v_t.$$

В случае 1 μ_0 можно записать как $\mu_0 = \mu_0^* + \gamma_0\alpha'$, где μ_0 входит в коинтеграционное пространство, а μ_0^* соответствует дрейфу в векторной модели исправления ошибок:

$$\Delta x_t = \mu_0^* + (\gamma_0 + x_{t-1}\beta)\,\alpha' + \sum_{j=1}^{p-1} \Delta x_{t-j}\Gamma_j + v_t.$$

Дрейф в модели исправления ошибок означает, что процесс x_t содержит линейный тренд[4].

Аналогичные рассуждения верны по отношению к временному тренду в случаях 2* и 2. В случае 2* тренд входит в коинтеграционное пространство, но не входит в x_t в виде квадратичного тренда. В случае 2 тренд входит и в коинтеграционное пространство, и в x_t в виде квадратичного тренда.

23.7. Метод Йохансена

Наряду с методом Энгла—Грейнджера (см. п. 17.6), еще одним популярным методом нахождения стационарных комбинаций нестационарных переменных является **метод Йохансена**. Этот метод, по сути дела, распространяет методику Дики—Фуллера (см. 17.4) на случай векторной авторегрессии. Помимо оценивания коинтегрирующих векторов, метод Йохансена также позволяет проверить гипотезы о ранге коинтеграции (количестве коинтегрирующих векторов) и гипотезы о виде коинтегрирующих векторов.

[4] Это аналог ситуации для скалярного авторегрессионного процесса с дрейфом.

Перечислим преимущества, которые дает метод Йохансена по сравнению с методом Энгла—Грейнджера:

1) Метод Энгла—Грейнджера применим, только когда между нестационарными переменными есть всего одно коинтегрирующее соотношение. Если ранг коинтеграции больше 1, то метод дает бессмысленные результаты.

2) Метод Энгла—Грейнджера статичен, в нем не учитывается краткосрочная динамика.

3) Результаты метода Йохансена не зависят от нормировки, использованной при оценивании, в то время как метод Энгла—Грейнджера может дать существенно отличающиеся результаты в зависимости от того, какая переменная стоит в левой части оцениваемой коинтеграционной регрессии.

Пусть векторный процесс $x_t = (x_{1t}, \ldots, x_{kt})$ описывается векторной авторегрессией p-го порядка, причем каждая из компонент является $\mathsf{I}(1)$ или $\mathsf{I}(0)$. Предполагается, что ошибки, относящиеся к разным моментам времени, независимы и распределены нормально с нулевым математическим ожиданием и ковариационной матрицей Σ. Как указывалось выше, можно записать векторную авторегрессию в форме векторной модели исправления ошибок:

$$\Delta x_t = x_{t-1}\Pi + \sum_{j=1}^{p-1} \Delta x_{t-j}\Gamma_j + v_t.$$

В методе Йохансена оцениваемыми параметрами являются $k \times k$ матрицы коэффициентов Γ_j и Π, а также ковариационная матрица Σ. Имея оценки Γ_j и Π, можно получить оценки коэффициентов приведенной формы модели по следующим формулам:

$$\begin{aligned}
&\Pi_1 = I + \Gamma_1 + \Pi,\\
&\Pi_j = \Gamma_j - \Gamma_{j-1},\ j = 2, \ldots, p-1,\\
&\Pi_p = -\Gamma_{p-1}.
\end{aligned}$$

Ранг коинтеграции r считается известным. Ограничения на ранг коинтеграции задаются как ограничения на матрицу Π. Как сказано выше, в предположении, что ранг коинтеграции равен r, ранг матрицы Π тоже равен r, и эту матрицу можно представить в виде произведения двух матриц:

$$\Pi = \beta\alpha',$$

где α и β имеют размерность $k \times r$. Таким образом, в дальнейших выкладках используется представление:

$$\Delta x_t = x_{t-1}\beta\alpha' + \sum_{j=1}^{p-1} \Delta x_{t-j}\Gamma_j + v_t.$$

Матрица β состоит из коинтегрирующих векторов. Заметим, что если бы матрица β была известна (естественно, с точностью до нормировки), то отклонения от равновесия $x_t\beta$ тоже были бы известны, и мы имели бы дело с линейными уравнениями регрессии, которые можно оценить посредством МНК. В методе Йохансена исходят из того, что матрицу β требуется оценить.

Для оценивания модели используется метод максимального правдоподобия.

Плотность распределения ошибок v_t по формуле для многомерного нормального распределения равна

$$(2\pi)^{-k/2}|\Sigma|^{-1/2}e^{(-\frac{1}{2}v_t\Sigma^{-1}v_t')}.$$

Обозначим через δ вектор, состоящий из параметров α, β и Γ_j. Для данного δ остатки модели равны

$$v_t(\delta) = \Delta x_t - x_{t-1}\beta\alpha' - \sum_{j=1}^{p-1}\Delta x_{t-j}\Gamma_j.$$

Используя это обозначение, можем записать функцию правдоподобия:

$$L(\delta,\Sigma) = (2\pi)^{-kT/2}|\Sigma|^{-T/2}e^{(-\frac{1}{2}\sum_{t=p}^{T}v_t(\delta)\Sigma^{-1}v_t'(\delta))}.$$

Заметим, что это функция правдоподобия, в которой за неимением данных в сумме пропущены первые p наблюдений.

Логарифмическая функция правдоподобия равна

$$\ln L(\delta,\Sigma) = -\frac{kT}{2}\ln(2\pi) + \frac{T}{2}\ln|\Sigma^{-1}| - \frac{1}{2}\sum_{t=p}^{T}v_t(\delta)\Sigma^{-1}v_t'(\delta).$$

При данном δ максимум функции правдоподобия по Σ достигается при

$$\Sigma = \Sigma(\delta) = \frac{1}{T}\sum_{t=p}^{T}v_t'(\delta)v_t(\delta).$$

Это можно доказать, дифференцируя логарифмическую функцию правдоподобия по Σ^{-1} (см. Приложение А.2.2).

Можно показать, что для этой матрицы выполнено

$$\sum_{t=p}^{T}v_t(\delta)\Sigma^{-1}(\delta)v_t'(\delta) = Tk.$$

С учетом этого максимизация функции правдоподобия эквивалентна минимизации определителя матрицы $\Sigma(\delta)$ по δ или

$$\left|\sum_{t=p}^{T} v_t'(\delta)v_t(\delta)\right| \to \min_{\delta}!$$

Тем самым мы получили, что метод максимального правдоподобия сводится к максимизации некоторой обобщенной суммы квадратов. По аналогии со сказанным выше ясно, что при данной матрице β можно получить оценки максимального правдоподобия для остальных неизвестных параметров обычным методом наименьших квадратов. Кроме того, при данных α, β можно получить оценки Γ_j методом наименьших квадратов из регрессий:

$$\Delta x_t - x_{t-p}\beta\alpha' = \sum_{j=1}^{p-1} \Delta x_{t-j}\Gamma_j + v_t.$$

Оценим отдельно регрессии Δx_t и x_{t-p} по переменным, стоящим в правой части данного уравнения:

$$\Delta x_t = \sum_{j=1}^{p-1} \Delta x_{t-j}S_j + r_{0t},$$

$$x_{t-p} = \sum_{j=1}^{p-1} \Delta x_{t-j}\mathrm{T}_j + r_{pt},$$

где S_j, T_j — коэффициенты регрессий.

Получим из них остатки r_{0t} и r_{pt}. Отсюда при данных α и β получим остатки исходной модели:

$$v_t = v_t(\alpha, \beta) = r_{0t} - r_{pt}\beta\alpha'.$$

В этих обозначениях задача нахождения оценок приобретает вид:

$$\left|\sum_{t=p}^{T} (r_{0t} - r_{pt}\beta\alpha')'(r_{0t} - r_{pt}\beta\alpha')\right| \to \min_{\alpha,\,\beta}!$$

Для нахождения матриц α и β Йохансен использовал процедуру, известную как регрессия с пониженным рангом. Она состоит в следующем. На основе остатков r_{0t} и r_{pt} формируются выборочные ковариационные матрицы остатков:

$$M_{ij} = \frac{1}{T}\sum_{t=1}^{T} r_{it}'r_{jt}, \quad i, j = 0, p,$$

и задача переписывается в виде

$$\left|M_{00} - \alpha\beta' M_{p0} - M_{0p}\beta\alpha' + \alpha\beta' M_{pp}\beta\alpha'\right| \to \min_{\alpha,\beta}!$$

Минимизация по α при данном β дает

$$\alpha(\beta) = M_{0p}\beta(\beta' M_{pp}\beta)^{-1},$$

откуда следует задача минимизации уже только по β:

$$\left|M_{00} - M_{0p}\beta(\beta' M_{pp}\beta)^{-1}\beta' M_{p0}\right| \to \min_{\beta}!$$

Очевидно, что отсюда нельзя однозначно найти β. Для нахождения β удобно ввести следующую нормировку: $\beta' M_{pp}\beta = I_r$. Оказывается, что с учетом данного нормирующего ограничения последняя задача минимизации эквивалентна следующей обобщенной задаче поиска собственных значений[5]:

$$\left(M_{p0}M_{00}^{-1}M_{0p} - \lambda M_{pp}\right)c = 0,$$

где c — собственный вектор, а собственные значения λ находятся как решения уравнения:

$$\left|M_{p0}M_{00}^{-1}M_{0p} - \lambda M_{pp}\right| = 0.$$

Матрица β находится в виде собственных векторов, соответствующих r наибольшим собственным значениям, где r — ранг коинтеграции. Пусть собственные числа упорядочены по убыванию, т.е.

$$\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_k.$$

Тогда следует выбрать первые r этих чисел, $\lambda_1, \ldots, \lambda_r$. Столбцами матрицы β будут соответствующие вектора $c_1, \ldots, c_r$.

Обобщенную задачу поиска собственных значений можно свести к стандартной задаче, если найти какую-либо матрицу $M_{pp}^{1/2}$, являющуюся квадратным корнем матрицы M_{pp}, т.е. $M_{pp} = \left(M_{pp}^{1/2}\right)' M_{pp}^{1/2}$:

$$\left(\left(M_{pp}^{-1/2}\right)' M_{p0}M_{00}^{-1}M_{0p}M_{pp}^{-1/2} - \lambda I\right)\widehat{c} = 0,$$

[5]Она напоминает метод наименьшего дисперсионного отношения для оценивания систем одновременных уравнений (см. п. 10.3). Также это напрямую связано с так называемой теорией канонических корреляций. (См. Бриллинджер Д. Временные ряды. Обработка данных и теория. — М.: «Мир», 1980.)

где $\widehat{c} = M_{pp}^{1/2} c$.

Напомним еще раз, что β определяется только с точностью до некоторой нормировки. Мы уже использовали нормировку $\beta' M_{pp} \beta = I_r$, которая выбрана из соображений удобства вычислений. Однако нормировку предпочтительнее выбирать, исходя из экономической теории рассматриваемых процессов. Поэтому следует умножить полученную оценку β справа на квадратную неособенную $r \times r$ матрицу, которая бы привела коинтеграционные вектора к более удобному для экономической интерпретации виду.

После того как найдена оценка максимального правдоподобия для β, вычисляются оценки других параметров. Для этого можно использовать в обратном порядке все те подстановки, которые мы сделали при упрощении задачи максимизации функции правдоподобия. Можно также использовать эквивалентную процедуру: сразу получить оценки α и Γ_j, применив метод наименьших квадратов к исходной регрессии.

Для проверки гипотез о ранге коинтеграции r используется статистика отношения правдоподобия. Пусть нулевая гипотеза состоит в том, что ранг коинтеграции равен r_0, а альтернативная гипотеза — что ранг коинтеграции равен r_a ($r_a > r_0$). Обозначим максимальное значение функции правдоподобия, полученное в предположении, что ранг равен r, через L_r. Тогда статистика отношения правдоподобия равна (см. раздел 18.3.3):

$$LR(r_0, r_a) = 2(\ln L_{r_a} - \ln L_{r_0}).$$

Подставив в функцию правдоподобия полученные оценки, можно вывести следующее выражение для максимального значения логарифма функции правдоподобия (с точностью до константы):

$$\ln L_r = -\frac{T}{2} \sum_{i=1}^{r} \ln(1 - \lambda_i) + \text{const}.$$

Поэтому

$$LR(r_0, r_a) = -T \sum_{i=r_0+1}^{r_a} \ln(1 - \lambda_i).$$

Особенно полезны с точки зрения поиска коинтеграционного ранга два частных случая статистики отношения правдоподобия.

Статистика следа используется для проверки нулевой гипотезы о том, что ранг коинтеграции равен r, против альтернативной гипотезы о том, что ранг равен k (количеству переменных). Статистика имеет вид:

$$LR^{trace} = LR(r, k) = -T \sum_{i=r+1}^{k} \ln(1 - \lambda_i).$$

Проверка гипотез проводится последовательно для $r = k-1, \ldots, 0$ и заканчивается, когда в первый раз не удается отклонить нулевую гипотезу. Можно проводить проверку гипотез и в обратном порядке $r = 0, \ldots, k-1$. В этом случае она заканчивается, когда нулевая гипотеза будет отвергнута в первый раз.

Можно также использовать **статистику максимального собственного значения**, которая используется для проверки нулевой гипотезы о том, что ранг равен r, против альтернативной гипотезы о том, что ранг равен $r+1$. Эта статистика равна:

$$LR^{\lambda-\max} = LR(r, r+1) = -\ln(1-\lambda_{r+1}).$$

Обе статистики имеют нестандартные асимптотические распределения. Эти асимптотические распределения не зависят от мешающих параметров, а зависят только от $k-r$, и от того, как входят в модель константа и тренд (см. перечисленные на стр. 668 пять основных случаев).

Методом Монте-Карло получены таблицы LR^{trace} и $LR^{\lambda-\max}$ для всех пяти случаев и нескольких значений $k-r$ (на данный момент имеются таблицы для $k-r = 1, \ldots, 12$). Также в последние годы разрабатываются различного рода аппроксимации для этих нестандартных распределений.

Как и в случае критерия ADF (см. п. 17.4), очень важным вопросом является выбор длины лага p (порядка авторегрессии). Способы, по сути дела, являются теми же самыми. Для проверки гипотез о длине лага можно использовать критерий отношения правдоподобия, который в данном случае имеет обычное распределение χ^2. Если процесс состоит из k компонент, и проверяется гипотеза о том, что следует увеличить p на единицу, то количество степеней свободы соответствующей статистики равно k. Важно также, чтобы в выбранной модели отсутствовала автокорреляция остатков, поскольку это одно из предположений модели.

Метод Йохансена можно использовать также для оценивания моделей с линейными ограничениями на матрицу коинтегрирующих векторов β или на матрицу корректирующих коэффициентов α. Для проверки таких ограничений удобно использовать все тот же тест отношения правдоподобия, который здесь имеет обычное асимптотическое распределение χ^2.

23.8. Коинтеграция и общие тренды

Рассмотрим модель VAR(1) с интегрированными переменными и ее представление в виде модели исправления ошибок:

$$\Delta x_t = x_{t-1}\beta\alpha' + v_t. \tag{23.7}$$

Поскольку матрица β состоит из коинтегрирующих векторов, то отклонения от равновесия $x_t\beta$ являются I(0), и для них существует разложение Вольда. Выведем

его. Для этого сначала умножим уравнение модели на β:

$$\Delta x_t \beta = x_{t-1} \beta \alpha' \beta + v_t \beta,$$

откуда:

$$x_t \beta = x_{t-1} \beta (\alpha' \beta + I) + v_t \beta.$$

Имеем для стационарного вектора $x_t \beta$ авторегрессию, из которой можем представить $x_t \beta$ в виде бесконечного скользящего среднего (разложение Вольда):

$$x_t \beta = \sum_{i=0}^{\infty} v_{t-i} \beta (\alpha' \beta + I)^i.$$

Подставив это выражение в исходную модель (23.7), получим также разложение Вольда для приростов Δx_t:

$$\Delta x_t = \sum_{i=1}^{\infty} v_{t-i} \beta (\alpha' \beta + I)^{i-1} \alpha' + v_t,$$

или

$$\Delta x_t = \sum_{i=0}^{\infty} v_{t-i} C_i = v_t C(L),$$

где мы ввели обозначения C_i для матричных коэффициентов скользящего среднего. Несложно подсчитать, пользуясь формулой бесконечной геометрической прогрессии, что соответствующий долгосрочный мультипликатор равен

$$C(1) = I - \beta (\alpha' \beta)^{-1} \alpha'. \tag{23.8}$$

Обозначим $C_i^* = -\sum\limits_{j=i+1}^{\infty} C_i$ и $C^*(L) = \sum\limits_{i=0}^{\infty} C_i^* L^i$. При этом выполнено следующее разложение для $C(L)$:

$$C(L) = \Delta C^*(L) + C(1).$$

Кроме того, можно показать, что $C^*(L)$ соответствует стационарному процессу. Все это позволяет разделить процесс Δx_t на сумму двух составляющих:

$$\Delta x_t = \Delta v_t C^*(L) + v_t C(1).$$

Следовательно, x_t можно представить следующим образом:

$$x_t = v_t C^*(L) + v_t^* C(1), \tag{23.9}$$

где v_t^* — это векторный процесс случайного блуждания, построенный на основе v_t (проинтегрированный v_t): $v_t = \Delta v_t^*$.

Первое слагаемое в разложении x_t (23.9) стационарно, а второе представляет собой линейную комбинацию процессов I(1).

Если воспользоваться тождеством

$$I = \alpha_\perp(\beta'_\perp\alpha_\perp)^{-1}\beta'_\perp + \beta(\alpha'\beta)^{-1}\alpha',$$

где $\alpha_\perp$ и $\beta_\perp$ — это $(k-r)\times k$ матрицы полного ранга, такие что $\alpha'\alpha_\perp = 0$ и $\beta'\beta_\perp = 0$, то матрицу $C(1)$, опираясь на (23.8), можно представить как

$$C(1) = \alpha_\perp(\beta'_\perp\alpha_\perp)^{-1}\beta'_\perp.$$

Таким образом, процесс x_t можно представить в виде[6]:

$$x_t = v_t C^*(L) + v_t^*\alpha_\perp(\beta'_\perp\alpha_\perp)^{-1}\beta'_\perp.$$

Элементы вектора $v_t^*\alpha_\perp$ являются общими стохастическими трендами. Отсюда видно, что k-мерный VAR-процесс с рангом коинтеграции r можно выразить через $k-r$ линейно независимых общих трендов. Матрица $(\beta'_\perp\alpha_\perp)^{-1}\beta'_\perp$ содержит коэффициенты («нагрузки») этих общих трендов.

Представление через общие тренды служит основой для еще одного метода оценивания коинтеграционных регрессий — метода Стока—Уотсона.

23.9. Упражнения и задачи

Упражнение 1

В таблице 23.2 приведены данные о потребительских расходах C и доходах Y в США в млрд. долл., очищенные от сезонности.

1.1. Нарисуйте график потребления и доходов. Что можно сказать об этих рядах по графикам?

1.2. Создайте первые разности логарифмов для обоих рядов. Нарисуйте график и сделайте выводы.

1.3. Предположим, что существует структурная зависимость между потреблением и доходами. А именно, потребление C зависит от текущих доходов и, вследствие привычек, от лагов потребления:

$$C_t = \alpha_1 + \alpha_2 Y_t + \alpha_3 C_{t-1} + \varepsilon_t^C.$$

[6] Это так называемое разбиение на цикл и тренд Бевериджа—Нельсона (см. п. 17.2).

Таблица 23.2. (Источник: Temple University Department of Economics Econometrics II Multivariate Time Series; `http://oll.temple.edu/economics/hwkdata/Vector_autoreg/Civar.htm`)

Квартал	C	Y	Квартал	C	Y	Квартал	C	Y
1947.1	192.5	202.3	1952.1	220	231.1	1957.1	268.9	291.1
1947.2	196.1	197.1	1952.2	227.7	240.9	1957.2	270.4	294.6
1947.3	196.9	202.9	1952.3	223.8	245.8	1957.3	273.4	296.1
1947.4	197	202.2	1952.4	230.2	248.8	1957.4	272.1	293.3
1948.1	198.1	203.5	1953.1	234	253.3	1958.1	268.9	291.3
1948.2	199	211.7	1953.2	236.2	256.1	1958.2	270.9	292.6
1948.3	199.4	215.3	1953.3	236	255.9	1958.3	274.4	299.9
1948.4	200.6	215.1	1953.4	234.1	255.9	1958.4	278.7	302.1
1949.1	199.9	212.9	1954.1	233.4	254.4	1959.1	283.8	305.9
1949.2	203.6	213.9	1954.2	236.4	254.8	1959.2	289.7	312.5
1949.3	204.8	214	1954.3	239	257	1959.3	290.8	311.3
1949.4	209	214.9	1954.4	243.2	260.9	1959.4	292.8	313.2
1950.1	210.7	228	1955.1	248.7	263	1960.1	295.4	315.4
1950.2	214.2	227.3	1955.2	253.7	271.5	1960.2	299.5	320.3
1950.3	225.6	232	1955.3	259.9	276.5	1960.3	298.6	321
1950.4	217	236.1	1955.4	261.8	281.4	1960.4	299.6	320.1
1951.1	223.3	230.9	1956.1	263.2	282			
1951.2	214.5	236.3	1956.2	263.7	286.2			
1951.3	217.5	239.1	1956.3	263.4	287.7			
1951.4	219.8	240.8	1956.4	266.9	291			

В свою очередь, текущие доходы зависят от лагов доходов (из-за инерции) и от лагов потребления (по принципу мультипликатора):

$$Y_t = \beta_1 + \beta_2 Y_{t-1} + \beta_3 C_{t-1} + \varepsilon_t^Y.$$

Оцените параметры структурной формы модели при помощи МНК по исходным данным. Затем проделайте то же самое, используя преобразованные данные из пункта 1.2 (разности логарифмов). Объясните, имеют ли два полученных набора оценок одинаковый смысл. Какие оценки предпочтительнее и почему?

1.4. Перепишите модель в приведенной форме. Укажите взаимосвязь между коэффициентами структурной и приведенной форм. Оцените приведенную форму модели по исходным данным и по преобразованным данным. Какие оценки предпочтительнее и почему?

1.5. Добавьте еще по одному лагу в оба уравнения приведенной формы. Оцените коэффициенты по исходным данным и по преобразованным данным и проведите тесты причинности по Грейнджеру. Что можно сказать о направлении причинности по полученным результатам?

1.6. Проверьте ряды на наличие единичных корней, используя тест Дики—Фуллера.

1.7. Примените к исходным данным и к логарифмам исходных данных метод Йохансена, используя в модели 4 лага разностей.

Упражнение 2

В таблице 23.3 даны макроэкономические показатели по США (поквартальные данные получены на основе помесячных данных): Infl — темп инфляции, рассчитанный по формуле: $400(\ln(\text{CPI}_t) - \ln(\text{CPI}_{t-1}))$, где CPI — индекс потребительских цен, т.е. логарифмический темп прироста цен в процентах из расчета за год; UnRate — уровень безработицы (процент населения); FedFunds — эффективная процентная ставка по межбанковским краткосрочным кредитам овернайт (в процентах годовых).

2.1. Оцените параметры векторной авторегрессии 4-го порядка, предполагая, что текущая инфляция влияет на текущую безработицу и процентную ставку, а текущая безработица влияет на текущую процентную ставку (рекурсивная система).

2.2. Оцените модель в приведенной форме, пропуская последнее наблюдение, и получите точечный прогноз на один период. Сравните прогнозы с фактическими значениями.

Упражнение 3

В таблице 23.4 приведены данные из статьи Турмана и Фишера, посвященной вопросу о том, что первично — куры или яйца. Это годовые данные по США за 1930–1983 гг. о производстве яиц в миллионах дюжин и о поголовье кур (исключая бройлерные).

3.1. Проведите тест причинности по Грейнджеру между двумя рядами. Сделайте выводы относительно направления причинности.

3.2. Для каждого ряда проведите тест Дики—Фуллера на наличие единичных корней, используя лаги от 0 до 12. Выберите нужную длину лага, используя

Таблица 23.3. (Источник: The Federal Reserve Bank of St. Louis, database FRED II, http://research.stlouisfed.org/fred2)

Квартал	Infl	UnRate	FedFunds
1954–3	–1.490	5.967	1.027
1954–4	0.000	5.333	0.987
1955–1	0.000	4.733	1.343
1955–2	–1.495	4.400	1.500
1955–3	2.985	4.100	1.940
1955–4	0.000	4.233	2.357
1956–1	0.000	4.033	2.483
1956–2	4.436	4.200	2.693
1956–3	2.930	4.133	2.810
1956–4	2.909	4.133	2.927
1957–1	4.324	3.933	2.933
1957–2	2.857	4.100	3.000
1957–3	2.837	4.233	3.233
1957–4	2.817	4.933	3.253
1958–1	5.575	6.300	1.863
1958–2	0.000	7.367	0.940
1958–3	0.000	7.333	1.323
1958–4	1.382	6.367	2.163
1959–1	0.000	5.833	2.570
1959–2	1.377	5.100	3.083
1959–3	2.740	5.267	3.577
1959–4	1.363	5.600	3.990
1960–1	0.000	5.133	3.933
1960–2	2.712	5.233	3.697
1960–3	0.000	5.533	2.937
1960–4	2.694	6.267	2.297
1961–1	0.000	6.800	2.003
1961–2	0.000	7.000	1.733
1961–3	2.676	6.767	1.683
1961–4	0.000	6.200	2.400
1962–1	2.658	5.633	2.457
1962–2	0.000	5.533	2.607
1962–3	2.640	5.567	2.847
1962–4	0.000	5.533	2.923
1963–1	1.314	5.767	2.967
1963–2	1.309	5.733	2.963
1963–3	1.305	5.500	3.330
1963–4	2.597	5.567	3.453
1964–1	0.000	5.467	3.463
1964–2	1.292	5.200	3.490
1964–3	1.288	5.000	3.457
1964–4	2.564	4.967	3.577
1965–1	0.000	4.900	3.973
1965–2	3.816	4.667	4.077
1965–3	0.000	4.367	4.073
1965–4	3.780	4.100	4.167
1966–1	3.744	3.867	4.557
1966–2	2.477	3.833	4.913
1966–3	4.908	3.767	5.410
1966–4	1.218	3.700	5.563
1967–1	1.214	3.833	4.823
1967–2	3.620	3.833	3.990
1967–3	3.587	3.800	3.893
1967–4	4.734	3.900	4.173
1968–1	3.514	3.733	4.787
1968–2	4.638	3.567	5.980
1968–3	4.585	3.533	5.943
1968–4	5.658	3.400	5.917
1969–1	5.579	3.400	6.567
1969–2	5.502	3.433	8.327
1969–3	5.427	3.567	8.983
1969–4	6.417	3.567	8.940
1970–1	6.316	4.167	8.573
1970–2	5.188	4.767	7.880
1970–3	4.103	5.167	6.703
1970–4	6.076	5.833	5.567

Таблица 23.3. (продолжение)

Квартал	Infl	UnRate	FedFunds
1971–1	2.005	5.933	3.857
1971–2	4.969	5.900	4.563
1971–3	2.952	6.033	5.473
1971–4	2.930	5.933	4.750
1972–1	2.909	5.767	3.540
1972–2	2.888	5.700	4.300
1972–3	3.819	5.567	4.740
1972–4	3.783	5.367	5.143
1973–1	8.382	4.933	6.537
1973–2	7.306	4.933	7.817
1973–3	8.949	4.800	10.560
1973–4	9.618	4.767	9.997
1974–1	12.753	5.133	9.323
1974–2	9.918	5.200	11.250
1974–3	12.853	5.633	12.090
1974–4	10.147	6.600	9.347
1975–1	6.877	8.267	6.303
1975–2	5.268	8.867	5.420
1975–3	8.141	8.467	6.160
1975–4	7.260	8.300	5.413
1976–1	2.867	7.733	4.827
1976–2	4.969	7.567	5.197
1976–3	6.299	7.733	5.283
1976–4	5.517	7.767	4.873
1977–1	8.136	7.500	4.660
1977–2	5.995	7.133	5.157
1977–3	5.255	6.900	5.820
1977–4	6.473	6.667	6.513
1978–1	7.001	6.333	6.757
1978–2	9.969	6.000	7.283
1978–3	9.126	6.033	8.100
1978–4	8.334	5.900	9.583
1979–1	11.612	5.867	10.073
1979–2	12.950	5.700	10.180
1979–3	12.006	5.867	10.947
1979–4	13.220	5.967	13.577
1980–1	16.308	6.300	15.047
1980–2	11.809	7.333	12.687
1980–3	6.731	7.667	9.837
1980–4	11.745	7.400	15.853
1981–1	10.058	7.433	16.570
1981–2	8.487	7.400	17.780
1981–3	11.330	7.400	17.577
1981–4	4.274	8.233	13.587
1982–1	2.542	8.833	14.227
1982–2	9.599	9.433	14.513
1982–3	2.876	9.900	11.007
1982–4	0.000	10.667	9.287
1983–1	1.634	10.367	8.653
1983–2	5.266	10.133	8.803
1983–3	4.004	9.367	9.460
1983–4	3.964	8.533	9.430
1984–1	5.874	7.867	9.687
1984–2	3.098	7.433	10.557
1984–3	3.839	7.433	11.390
1984–4	3.045	7.300	9.267
1985–1	4.899	7.233	8.477
1985–2	2.613	7.300	7.923
1985–3	2.226	7.200	7.900
1985–4	5.147	7.033	8.103
1986–1	–1.464	7.033	7.827
1986–2	1.098	7.167	6.920
1986–3	2.188	6.967	6.207
1986–4	2.899	6.833	6.267
1987–1	5.022	6.600	6.220
1987–2	4.608	6.267	6.650

Таблица 23.3. (продолжение)

Квартал	Infl	UnRate	FedFunds
1987–3	4.207	6.000	6.843
1987–4	3.126	5.833	6.917
1988–1	3.102	5.700	6.663
1988–2	5.117	5.467	7.157
1988–3	5.053	5.467	7.983
1988–4	3.997	5.333	8.470
1989–1	4.940	5.200	9.443
1989–2	6.171	5.233	9.727
1989–3	2.250	5.233	9.083
1989–4	4.779	5.367	8.613
1990–1	7.219	5.300	8.250
1990–2	4.023	5.333	8.243
1990–3	7.927	5.700	8.160
1990–4	5.099	6.133	7.743
1991–1	1.784	6.600	6.427
1991–2	3.545	6.833	5.863
1991–3	2.930	6.867	5.643
1991–4	3.488	7.100	4.817
1992–1	2.596	7.367	4.023
1992–2	2.865	7.600	3.770
1992–3	2.845	7.633	3.257
1992–4	3.387	7.367	3.037
1993–1	2.801	7.133	3.040
1993–2	2.782	7.067	3.000
1993–3	1.936	6.800	3.060
1993–4	3.570	6.633	2.990
1994–1	2.181	6.567	3.213
1994–2	2.169	6.200	3.940
1994–3	3.769	6.000	4.487
1994–4	2.138	5.633	5.167
1995–1	2.921	5.467	5.810
1995–2	3.162	5.667	6.020
1995–3	1.833	5.667	5.797
1995–4	2.085	5.567	5.720
1996–1	4.137	5.533	5.363
1996–2	3.075	5.500	5.243
1996–3	2.545	5.267	5.307
1996–4	3.535	5.333	5.280
1997–1	1.756	5.233	5.277
1997–2	1.000	5.000	5.523
1997–3	2.489	4.867	5.533
1997–4	1.486	4.667	5.507
1998–1	0.494	4.633	5.520
1998–2	1.970	4.400	5.500
1998–3	1.716	4.533	5.533
1998–4	2.196	4.433	4.860
1999–1	0.972	4.300	4.733
1999–2	3.143	4.267	4.747
1999–3	4.073	4.233	5.093
1999–4	2.377	4.067	5.307
2000–1	5.180	4.033	5.677
2000–2	3.029	3.967	6.273
2000–3	3.007	4.067	6.520
2000–4	2.298	3.933	6.473
2001–1	3.195	4.167	5.593
2001–2	4.295	4.467	4.327
2001–3	0.449	4.833	3.497
2001–4	–1.801	5.600	2.133
2002–1	2.698	5.633	1.733
2002–2	2.903	5.833	1.750
2002–3	2.440	5.767	1.740
2002–4	1.545	5.900	1.443
2003–1	5.034	5.767	1.250
2003–2	–0.653	6.167	1.247
2003–3	3.039	6.133	1.017

Таблица 23.4. (Источник: Thurman, Walter N. and Mark E. Fisher, “Chickens, Eggs, and Causality”, American Journal of Agricultural Economics 70 (1988), p. 237–238.)

	куры	**яйца**		**куры**	**яйца**
1930	468491	3581	1957	391363	5442
1931	449743	3532	1958	374281	5442
1932	436815	3327	1959	387002	5542
1933	444523	3255	1960	369484	5339
1934	433937	3156	1961	366082	5358
1935	389958	3081	1962	377392	5403
1936	403446	3166	1963	375575	5345
1937	423921	3443	1964	382262	5435
1938	389624	3424	1965	394118	5474
1939	418591	3561	1966	393019	5540
1940	438288	3640	1967	428746	5836
1941	422841	3840	1968	425158	5777
1942	476935	4456	1969	422096	5629
1943	542047	5000	1970	433280	5704
1944	582197	5366	1971	421763	5806
1945	516497	5154	1972	404191	5742
1946	523227	5130	1973	408769	5502
1947	467217	5077	1974	394101	5461
1948	499644	5032	1975	379754	5382
1949	430876	5148	1976	378361	5377
1950	456549	5404	1977	386518	5408
1951	430988	5322	1978	396933	5608
1952	426555	5323	1979	400585	5777
1953	398156	5307	1980	392110	5825
1954	396776	5402	1981	384838	5625
1955	390708	5407	1982	378609	5800
1956	383690	5500	1983	364584	5656

информационные критерии Акаике и Шварца. Сделайте вывод о том, являются ли ряды стационарными. Как полученный результат может повлиять на интерпретацию результатов упражнения 3.1?

3.3. Проверьте с помощью методов Энгла—Грейнджера и Йохансена, коинтегрированы ли ряды. Как полученный результат может повлиять на интерпретацию результатов упражнения 3.1?

Упражнение 4

В таблице 23.5 приведены поквартальные макропоказатели по Великобритании из обзорной статьи Мускателли и Хурна: M — величина денежной массы (агрегат M_1); Y — общие конечные расходы на товары и услуги (TFE) в постоянных ценах (переменная, моделирующая реальные доходы); P — дефлятор TFE (индекс цен); R — ставка по краткосрочным казначейским векселям (переменная, соответствующая альтернативной стоимости хранения денег). Изучается связь между тремя переменными: $\ln M - \ln P$ — реальная денежная масса (в логарифмах); $\ln Y$ — реальный доход (в логарифмах); R — процентная ставка.

4.1. Найдите ранг коинтеграции и коинтегрирующие вектора методом Йохансена, используя векторную модель исправления ошибок (VECM) с четырьмя разностями в правой части и с сезонными фиктивными переменными. Сделайте это при разных возможных предположениях о том, как входят в модель константа и тренд:

   а) константа входит в коинтеграционное пространство, но не входит в модель исправления ошибок в виде дрейфа;

   б) константа входит в коинтеграционное пространство, а также в модель исправления ошибок в виде дрейфа, так что данные содержат тренд;

   в) тренд входит в коинтеграционное пространство, но данные не содержат квадратичный тренд.

4.2. С помощью тестов отношения правдоподобия определить, как должны входить в модель константа и тренд. Убедитесь в том, что случай 4.1а, который рассматривался в статье Мускателли и Хурна, отвергается тестами.

4.3. На основе найденного коинтегрирующего вектора оцените остальные коэффициенты VECM. Рассмотрите полученные коэффициенты модели и сделайте вывод о том, насколько они соответствуют экономической теории. (Подсказка: интерпретируйте коинтегрирующую комбинацию как уравнение спроса на деньги и обратите внимания на знак при $\ln Y$).

Таблица 23.5. (Источник: Muscatelli, V.A., Hurn, S. Cointegration and Dynamic Time Series Models. Journal of Economic Surveys 6 (1992, No. 1), 1–43.)

Квартал	lnM	R	lnY	lnP	Квартал	lnM	R	lnY	lnP
1963–1	8.8881	0.0351	10.678	–1.635	1974–1	9.5017	0.1199	11.068	–0.95787
1963–2	8.9156	0.0369	10.746	–1.6189	1974–2	9.5325	0.1136	11.103	–0.89963
1963–3	8.9305	0.0372	10.752	–1.6205	1974–3	9.5584	0.1118	11.125	–0.85112
1963–4	8.9846	0.0372	10.789	–1.6028	1974–4	9.6458	0.1096	11.139	–0.8025
1964–1	8.9584	0.0398	10.756	–1.6045	1975–1	9.6438	0.1001	11.066	–0.73806
1964–2	8.9693	0.0436	10.8	–1.5843	1975–2	9.6742	0.0938	11.074	–0.6802
1964–3	8.9911	0.0462	10.802	–1.5813	1975–3	9.7275	0.1017	11.088	–0.63297
1964–4	9.0142	0.0547	10.838	–1.5677	1975–4	9.7689	0.1112	11.124	–0.59904
1965–1	8.9872	0.0651	10.779	–1.5557	1976–1	9.787	0.0904	11.092	–0.56265
1965–2	9.0005	0.0612	10.818	–1.5438	1976–2	9.8141	0.1019	11.105	–0.52675
1965–3	9.0116	0.0556	10.832	–1.5408	1976–3	9.8641	0.1126	11.134	–0.49414
1965–4	9.0506	0.0545	10.851	–1.5272	1976–4	9.8765	0.1399	11.172	–0.45642
1966–1	9.0394	0.0556	10.814	–1.519	1977–1	9.8815	0.112	11.112	–0.41982
1966–2	9.0336	0.0565	10.838	–1.5037	1977–2	9.9238	0.0772	11.122	–0.38322
1966–3	9.0438	0.0658	10.849	–1.4964	1977–3	10.001	0.0655	11.14	–0.36159
1966–4	9.0491	0.0662	10.86	–1.4845	1977–4	10.071	0.0544	11.177	–0.35041
1967–1	9.0388	0.06	10.841	–1.4842	1978–1	10.097	0.0597	11.143	–0.32276
1967–2	9.055	0.053	10.874	–1.4769	1978–2	10.117	0.0949	11.165	–0.29942
1967–3	9.0975	0.0544	10.881	–1.4726	1978–3	10.168	0.0938	11.182	–0.27619
1967–4	9.1326	0.0657	10.9	–1.4661	1978–4	10.223	0.1191	11.203	–0.25279
1968–1	9.1029	0.074	10.89	–1.4459	1979–1	10.222	0.1178	11.159	–0.22607
1968–2	9.1204	0.0714	10.901	–1.4285	1979–2	10.236	0.1379	11.215	–0.1898
1968–3	9.1351	0.0695	10.929	–1.4141	1979–3	10.274	0.1382	11.221	–0.13856
1968–4	9.1733	0.0666	10.961	–1.4044	1979–4	10.304	0.1649	11.242	–0.10048
1969–1	9.1182	0.0718	10.891	–1.3898	1980–1	10.274	0.1697	11.199	–0.05655
1969–2	9.0992	0.0783	10.932	–1.3825	1980–2	10.293	0.1632	11.171	–0.01134
1969–3	9.1157	0.0782	10.943	–1.3697	1980–3	10.294	0.1486	11.187	0.02109
1969–4	9.1744	0.0771	10.979	–1.3573	1980–4	10.343	0.1358	11.188	0.04439
1970–1	9.1371	0.0754	10.905	–1.3368	1981–1	10.356	0.1187	11.144	0.06282
1970–2	9.178	0.0689	10.962	–1.3168	1981–2	10.39	0.1224	11.144	0.09367
1970–3	9.1981	0.0683	10.971	–1.2939	1981–3	10.407	0.1572	11.191	0.11541
1970–4	9.2643	0.0682	11.017	–1.2791	1981–4	10.506	0.1539	11.206	0.13245
1971–1	9.2693	0.0674	10.938	–1.2585	1982–1	10.501	0.1292	11.175	0.14663
1971–2	9.2836	0.0567	10.99	–1.2345	1982–2	10.526	0.1266	11.172	0.1703
1971–3	9.3221	0.0539	11.01	–1.2132	1982–3	10.551	0.1012	11.193	0.1829
1971–4	9.3679	0.0452	11.044	–1.1992	1982–4	10.613	0.0996	11.22	0.19315
1972–1	9.3742	0.0436	10.98	–1.1853	1983–1	10.639	0.1049	11.209	0.21273
1972–2	9.4185	0.0459	11.025	–1.1707	1983–2	10.664	0.0951	11.195	0.2242
1972–3	9.4356	0.0597	11.02	–1.1439	1983–3	10.675	0.0917	11.244	0.23412
1972–4	9.4951	0.0715	11.1	–1.1282	1983–4	10.719	0.0904	11.268	0.24285
1973–1	9.4681	0.0813	11.093	–1.1044	1984–1	10.754	0.0856	11.241	0.25738
1973–2	9.5347	0.0754	11.102	–1.0913	1984–2	10.798	0.0906	11.233	0.27467
1973–3	9.5119	0.1025	11.117	–1.0493	1984–3	10.827	0.1024	11.268	0.28672
1973–4	9.5445	0.1162	11.141	–1.0048	1984–4	10.862	0.0933	11.317	0.29773

Задачи

1. Рассмотрите приведенную форму процесса VAR(1):

$$(x_t, y_t) = (x_{t-1}, y_{t-1}) \begin{pmatrix} 0.2 & 0.4 \\ 0.2 & 0 \end{pmatrix} + (v_{xt}, v_{yt}),$$

где ошибки v_{xt}, v_{yt} не автокоррелированы и их ковариационная матрица равна

$$\begin{pmatrix} 1 & -0.5 \\ -0.5 & 4.25 \end{pmatrix}.$$

а) Является ли процесс стационарным?

б) Найдите структурную форму модели (матрицу коэффициентов и ковариационную матрицу), если известно, что она является рекурсивной (y_t входит в уравнение для x_t, но x_t не входит в уравнение для y_t).

в) Найдите (матричный) долгосрочный мультипликатор.

2. Векторная регрессия с двумя переменными x_t, y_t задается следующими уравнениями:

$$x_t = \alpha x_{t-1} + \beta y_{t-1} + v_t,$$
$$y_t = \gamma x_{t-1} + \delta y_{t-1} + w_t.$$

Ковариационная матрица ошибок v_t, w_t имеет вид:

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \text{ где } |\rho| < 1.$$

а) Запишите модель в матричном виде.

б) Представьте матрицу Σ в виде $\Sigma = U'\Omega U$, где U — верхняя треугольная матрица, Ω — диагональная матрица с положительными диагональными элементами.

в) Умножьте уравнение модели справа на матрицу U. Что можно сказать о получившемся представлении модели?

г) Повторите задание, поменяв порядок переменных y_t, x_t. Сравните и сделайте выводы.

3. Предположим, что темпы прироста объемов производства, y_t, и денежной массы, m_t, связаны следующими структурными уравнениями:

$$m_t = \alpha m_{t-1} + \varepsilon_{mt},$$
$$y_t = \beta m_t + \gamma m_{t-1} + \delta y_{t-1} + \varepsilon_{yt},$$

где ошибки $\varepsilon_{mt}, \varepsilon_{yt}$ не автокоррелированы, не коррелированы друг с другом, а их дисперсии равны σ_m^2 и σ_y^2, соответственно.

а) Запишите структурные уравнения в стандартном матричном виде модели SVAR.

б) Запишите модель в приведенной форме.

в) Какой вид имеет функция реакции на импульсы для монетарных шоков ε_{mt} и шоков производительности ε_{yt}? Как эта функция связана с представлением модели в виде бесконечного скользящего среднего (разложением Вольда)?

г) Найдите (матричный) долгосрочный мультипликатор.

4. Рассмотрите двумерную векторную авторегрессию первого порядка:

$$(x_{1t}, x_{2t}) = (x_{1,\,t-1}, x_{2,\,t-1}) \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix} + (v_{1t}, v_{2t}),$$

где ошибки v_{1t}, v_{2t} являются белым шумом и независимы между собой.

а) При каких параметрах модель является рекурсивной? Объясните.

б) При каких параметрах x_{1t} и x_{2t} представляют собой два независимых случайных блуждания? Объясните.

в) Известно, что x_{1t} не является причиной x_{2t} в смысле Грейнджера. Какие ограничения этот факт накладывает на параметры?

5. Рассмотрите авторегрессию второго порядка: $x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \varepsilon_t$, где ошибка ε_t представляет собой белый шум.

а) Обозначьте $x_{t-1} = y_t$ и запишите данную модель в виде векторной авторегрессии первого порядка для переменных x_t и y_t.

б) Чему равна ковариационная матрица одновременных ошибок в получившейся векторной авторегрессии?

в) Сопоставьте условия стационарности исходной модели AR(2) и полученной модели VAR(1).

6. Представьте векторную авторегрессию второго порядка в виде векторной авторегрессии первого порядка (с расшифровкой обозначений).

7. Рассмотрите двумерную модель VAR(1):

$$(x_{1t},\ x_{2t}) = (x_{1,\,t-1},\ x_{2,\,t-1})\Pi + v_t,\ \text{где}\ \Pi = \begin{pmatrix} 1/2 & 1 \\ 0 & 1/4 \end{pmatrix}.$$

   а) Найдите корни характеристического многочлена, соответствующего этой модели. Является ли процесс стационарным?

   б) Найдите собственные числа матрицы Π. Как они связаны с корнями характеристического многочлена, найденными в пункте (а)?

8. Рассмотрите векторную авторегрессию первого порядка:

$$(x_{1t}, x_{2t}) = (x_{1,\,t-1}, x_{2,\,t-1}) \begin{pmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{pmatrix} + (v_{1t}, v_{2t})\,.$$

   При каких параметрах процесс является стационарным:

   а) $\pi_{11} = 1,\ \pi_{12} = 0.5,\ \pi_{21} = 0,\ \pi_{22} = 1$;

   б) $\pi_{11} = 0.3,\ \pi_{12} = 0.1,\ \pi_{21} = -0.1, \pi_{22} = 0.5$;

   в) $\pi_{11} = 2,\ \pi_{12} = 0,\ \pi_{21} = 1,\ \pi_{22} = 0.5$;

   г) $\pi_{11} = 0.5,\ \pi_{12} = -1,\ \pi_{21} = 1,\ \pi_{22} = 0.5$;

   д) $\pi_{11} = 0.5,\ \pi_{12} = -1,\ \pi_{21} = 1,\ \pi_{22} = 0.5$;

   е) $\pi_{11} = 0.3,\ \pi_{12} = -0.2,\ \pi_{21} = 0.2,\ \pi_{22} = 0.3$?

   Аргументируйте свой ответ.

9. В стране чудес динамика темпа прироста ВВП, y_t, темпа прироста денежной массы M2, m_t, и ставки процента, r_t, описывается следующей моделью

VAR(2):

$$(y_t, m_t, r_t) = (2, 1, 0) + (y_{t-1}, m_{t-1}, r_{t-1}) \begin{pmatrix} 0.7 & 0 & 0.9 \\ 0.1 & 0.4 & 0 \\ 0 & 0.1 & 0.8 \end{pmatrix} +$$

$$+ (y_{t-2}, m_{t-2}, r_{t-2}) \begin{pmatrix} -0.2 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0.1 & 0 \end{pmatrix} + (v_{yt}, v_{mt}, v_{rt}),$$

где ошибки v_{yt}, v_{mt}, v_{rt} представляют собой белый шум.

а) Покажите, что все три переменные являются стационарными.

б) Найдите безусловные математические ожидания этих переменных.

в) Запишите модель в виде векторной модели исправления ошибок.

10. Опишите поэтапно возможную процедуру построения прогнозов для векторной авторегрессии. Необходимо ли для построения прогноза знать ограничения, накладываемые структурной формой? (Объясните.) На каком этапе построения прогноза можно было бы учесть структурные ограничения?

11. Объясните различие между структурной и приведенной формой векторной авторегрессии. В чем причина того, что разложение дисперсии ошибки прогноза основывают на структурной форме, а не на приведенной форме?

12. Рассмотрите векторный процесс (x_t, y_t):

$$x_t = x_{t-1} + \varepsilon_t,$$

$$y_t = \lambda x_t + \varphi y_{t-1} + \xi_t \ (\lambda \neq 0, |\varphi| < 1).$$

а) Покажите, что x_t и y_t являются коинтегрированными CI(1, 0). Укажите ранг коинтеграции и общий вид коинтегрирующих векторов.

б) Запишите процесс в виде векторной модели исправления ошибок. Укажите соответствующую матрицу корректирующих коэффициентов и матрицу коинтегрирующих векторов.

13. Пусть в векторной модели исправления ошибок константа входит в коинтеграционное пространство. Какие ограничения это налагает на параметры модели?

14. На примере векторной авторегрессии первого порядка с двумя переменными, коинтегрированными как $\mathsf{CI}(1, 0)$, покажите, что наличие константы (дрейфа) в коинтеграционном пространстве означает, что переменные содержат линейный тренд.

15. Пусть в векторной модели исправления ошибок

$$\Delta x_t = x_{t-1}\Pi + \sum_{j=1}^{p-1} \Delta x_{t-j}\Gamma_j + v_t$$

матрица $\Pi = \begin{pmatrix} 1 & 6 & 1 \\ 2 & 4 & 2 \\ 3 & 2 & 1 \end{pmatrix}$.

Найдите ранг коинтеграции, матрицу коинтегрирующих векторов β и матрицу корректирующих коэффициентов α.

16. Объясните, почему процедура Йохансена позволяет не проверять переменные на наличие единичных корней.

17. Пусть в векторной модели исправления ошибок

$$\Delta x_t = x_{t-1}\Pi + \sum_{j=1}^{p-1} \Delta x_{t-j}\Gamma_j + v_t$$

коинтегрирующие векторы равны $(1; -1; 0)$ и $(1; 1; 1)$, а матрица корректирующих коэффициентов равна

$$\alpha = \begin{pmatrix} 1 & 0 \\ 0 & 2 \\ 0 & -1 \end{pmatrix}.$$

Найдите матрицу Π.

Рекомендуемая литература

1. **Amisano Gianni, Carlo Giannini.** Topics in Structural VAR Econometrics, 2nd ed. — Springer, 1997.

2. **Banerjee A., J.J. Dolado, J.W. Galbraith and D.F. Hendry,** Co-integration, Error Correction and the Econometric Analysis of Non-stationary Data. — Oxford University Press, 1993. (Ch. 5, 8.)

3. **Canova F.** «VAR Models: Specification, Estimation, Inference and Forecasting» in H. Pesaran and M. Wickens (eds.) Handbook of Applied Econometrics. — Basil Blackwell, 1994.

4. **Granger C. W. J.** Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. // Econometrica, 37 (1969), 424–438.

5. **Greene W.H.** Econometric Analysis. — Prentice-Hall, 2000. (Ch. 17, 18).

6. **Hamilton, J. D.** Time Series Analysis. — Princeton University Press, 1994. (Ch. 10, 11).

7. **Johansen S.** Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models. // Econometrica, 59 (1991), 1551–1580.

8. **Lutkepohl Helmut.** Introduction to Multiple Time Series Analysis, second edition. — Berlin: Springer, 1993. (Ch. 2, 10, 11).

9. **Muscatelli V.A. and Hurn S.** Cointegration and dynamic time series models. // Journal of Economic Surveys, 6 (1992), 1–43.

10. **Sims C. A.** Macroeconomics and Reality. // Econometrica, 48 (1980), 1–48.

11. **Stock J.H. and Watson M.W.** Testing for Common Trends. // Journal of the American Statistical Association, 83 (1988), 1097–1107.

12. **Watson Mark W.** Vector Autoregressions and Cointegration. // Handbook of Econometrics, Vol. IV. Robert Engle and Daniel McFadden, eds. Elsevier, 1994, 2844–2915.

13. **Mills Terence C.** Time Series Techniques for Economists. — Cambridge University Press, 1990. (Ch. 14).

14. **Mills Terence C.** The Econometric Financial Modelling Time Series. — Cambridge University Press, 1999. (Ch. 7, 8).

Приложение А

Вспомогательные сведения из высшей математики

А.1. Матричная алгебра

А.1.1. Определения

$\mathbf{x} = \{x_i\}_{i=1,\ldots,n} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ называется вектор-столбцом размерности n.

$\mathbf{x} = (x_1, , \ldots, x_n)$ называется вектор-строкой размерности n.

$$\mathbf{A} = \{a_{ij}\}_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}$$

называется матрицей размерности $m \times n$.

- Сумма матриц $\mathbf{A}$ и $\mathbf{B}$ $(m \times n)$: $\mathbf{C} = \mathbf{A} + \mathbf{B} = \{a_{ij} + b_{ij}\}$, $\mathbf{C}$ $(m \times n)$.
- Произведение матриц $\mathbf{A}$ $(m \times n)$ и $\mathbf{B}$ $(n \times k)$: $\mathbf{C} = \mathbf{AB} = \left\{\sum\limits_{t=1}^{n} a_{it} b_{tj}\right\}$, $\mathbf{C}$ $(m \times k)$.
- Скалярное произведение вектор-столбцов $\mathbf{a}$ $(m \times 1)$ и $\mathbf{b}$ $(m \times 1)$: $\mathbf{a}'\mathbf{b} = \sum\limits_{i=1}^{m} a_i b_i$.
- Квадратичная форма вектор-столбца $\mathbf{x}$ $(m \times 1)$ и матрицы $\mathbf{A}$ $(m \times m)$: $\mathbf{x}'\mathbf{A}\mathbf{x} = \sum\limits_{i=1}^{m} \sum\limits_{j=1}^{m} a_{ij} x_i x_j$.
- Произведение матрицы $\mathbf{A}$ $(m \times n)$ на скаляр α: $\mathbf{B} = \alpha\mathbf{A} = \{\alpha a_{ij}\}$, $\mathbf{B}$ $(m \times n)$.
- Транспонирование матрицы $\mathbf{A}$ $(m \times n)$: $\mathbf{B} = \mathbf{A}' = \{a_{ji}\}$, $\mathbf{B}$ $(n \times m)$.
- След матрицы $\mathbf{A}$ $(m \times m)$: $\mathrm{tr}\,(\mathbf{A}) = \sum\limits_{i=1}^{m} a_{ii}$.
- Рангом ($\mathrm{rank}(\mathbf{A})$) матрицы $\mathbf{A}$ называется количество линейно независимых столбцов (равное количеству линейно независимых строк). Матрица $\mathbf{A}$ $(m \times n)$ имеет полный ранг по столбцам, если $\mathrm{rank}(\mathbf{A}) = n$. Матрица $\mathbf{A}$ $(m \times n)$ имеет полный ранг по строкам, если $\mathrm{rank}(\mathbf{A}) = m$.
- Матрица $\mathbf{A}$ $(m \times m)$ называется невырожденной (неособенной), если $\mathrm{rank}(\mathbf{A}) = m$. В противном случае она называется вырожденной.
- Матрица $\mathbf{A}$ $(m \times m)$ называется диагональной, если $a_{ij} = 0$ при $i \neq j$. Для диагональной матрицы используется обозначение $\mathbf{A} = \mathrm{diag}(a_{11}, \ldots, a_{mm})$.
- Матрица $\mathbf{I}_m = \mathrm{diag}(1, \ldots, 1) = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$ $(m \times m)$ называется единичной.
- Матрица $\mathbf{A}$ $(m \times m)$ называется симметричной (симметрической), если $\mathbf{A} = \mathbf{A}'$.

- Матрица $\mathbf{A}$ $(m \times m)$ называется верхней треугольной, если $a_{ij} = 0$ при $i > j$. Матрица $\mathbf{A}$ $(m \times m)$ называется нижней треугольной, если $a_{ij} = 0$ при $i < j$.

- Матрица $\mathbf{A}^{-1}$ $(m \times m)$ называется обратной матрицей к матрице $\mathbf{A}$ $(m \times m)$, если $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_m$.

- Матрица $\mathbf{A}$ $(m \times m)$ называется идемпотентной, если $\mathbf{A}\mathbf{A} = \mathbf{A}^2 = \mathbf{A}$.

- Векторы-столбцы $\mathbf{a}$ $(m \times 1)$ и $\mathbf{b}$ $(m \times 1)$ называются ортогональными, если их скалярное произведение равно нулю: $\mathbf{a}'\mathbf{b} = 0$.

- Матрица $\mathbf{A}$ $(m \times n)$, где $m \geqslant n$, называется ортогональной, если ее столбцы ортогональны, т.е. $\mathbf{A}'\mathbf{A} = \mathbf{I}_n$.

- Матрица $\mathbf{A}$ $(m \times m)$ называется положительно определенной, если для любого вектор-столбца $\mathbf{x} \neq 0$ $(m \times 1)$ выполняется $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$. Матрица $\mathbf{A}$ $(m \times m)$ называется отрицательно определенной, если для любого вектор-столбца $\mathbf{x} \neq 0$ $(m \times 1)$ выполняется $\mathbf{x}'\mathbf{A}\mathbf{x} < 0$.

- Матрица $\mathbf{A}$ $(m \times m)$ называется положительно полуопределенной (неотрицательно определенной), если для любого вектора-столбца $\mathbf{x}$ $(m \times 1)$ выполняется $\mathbf{x}'\mathbf{A}\mathbf{x} \geqslant 0$. Матрица $\mathbf{A}$ $(m \times m)$ называется отрицательно полуопределенной (неположительно определенной), если для любого вектора-столбца $\mathbf{x}$ $(m \times 1)$ выполняется $\mathbf{x}'\mathbf{A}\mathbf{x} \leqslant 0$.

- Определителем матрицы $\mathbf{A}$ $(m \times m)$ называется

$$|\mathbf{A}| = \det(\mathbf{A}) = \sum_{j=1}^{m} a_{ij}(-1)^{i+j} |\mathbf{A}_{ij}|,$$

  где i — номер любой строки, а матрицы $\mathbf{A}_{ij}$ $((m-1) \times (m-1))$ получены из матрицы $\mathbf{A}$ путем вычеркивания i-й строки и j-го столбца.

- Для матрицы $\mathbf{A}$ $(m \times m)$ уравнение $|\mathbf{A} - \lambda\mathbf{I}_m| = 0$ называется характеристическим уравнением. Решение этого уравнения λ называется собственным числом (собственным значением) матрицы $\mathbf{A}$. Вектор $\mathbf{x} \neq 0$ $(m \times 1)$ называется собственным вектором матрицы $\mathbf{A}$, соответствующим собственному числу λ, если $(\mathbf{A} - \lambda\mathbf{I}_m)\,\mathbf{x} = 0$.

- Прямое произведение (произведение Кронекера) матриц $\mathbf{A}$ $(m \times n)$ и $\mathbf{B}$ $(p \times q)$ это матрица $\mathbf{C}$ $(mp \times nq)$:

$$\mathbf{C} = \mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2n}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix}.$$

А.1.2. Свойства матриц

Сложение матриц

- $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$ (коммутативность).
- $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$ (ассоциативность).

Произведение матриц

- В общем случае $\mathbf{AB} \neq \mathbf{BA}$ (свойство коммутативности не выполнено).
- $(\mathbf{AB})\,\mathbf{C} = \mathbf{A}\,(\mathbf{BC})$ (ассоциативность).
- $\mathbf{A}\,(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC} \quad (\mathbf{A} + \mathbf{B})\,\mathbf{C} = \mathbf{AC} + \mathbf{BC}$ (дистрибутивность).
- $\mathbf{AI}_m = \mathbf{I}_m\mathbf{A} = \mathbf{A}$ для матрицы $\mathbf{A}$ $(m \times m)$.
- $$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} \begin{pmatrix} \mathbf{E} & \mathbf{F} \\ \mathbf{G} & \mathbf{H} \end{pmatrix} = \begin{pmatrix} \mathbf{AE} + \mathbf{BG} & \mathbf{AF} + \mathbf{BH} \\ \mathbf{CE} + \mathbf{DG} & \mathbf{CF} + \mathbf{DH} \end{pmatrix}.$$

Ранг

- Для матрицы $\mathbf{A}$ $(m \times n)$ выполнено $\operatorname{rank}(\mathbf{A}) \leqslant \min\{m, n\}$.
- $\operatorname{rank}(\mathbf{AB}) \leqslant \min\{\operatorname{rank}(\mathbf{A}), \operatorname{rank}(\mathbf{B})\}$.
- Если матрица $\mathbf{B}$ $(m \times m)$ является невырожденной, то для матрицы $\mathbf{A}$ $(m \times n)$ выполнено $\operatorname{rank}(\mathbf{A}) = \operatorname{rank}(\mathbf{BA})$. Если матрица $\mathbf{B}$ $(n \times n)$ является невырожденной, то для матрицы $\mathbf{A}$ $(m \times n)$ выполнено $\operatorname{rank}(\mathbf{A}) = \operatorname{rank}(\mathbf{AB})$.
- $\operatorname{rank}(\mathbf{A}'\mathbf{A}) = \operatorname{rank}(\mathbf{AA}') = \operatorname{rank}(\mathbf{A})$.

След

- $\mathrm{tr}\,(\mathbf{A}+\mathbf{B}) = \mathrm{tr}\,(\mathbf{A}) + \mathrm{tr}\,(\mathbf{B})$.
- $\mathrm{tr}\,(\alpha\mathbf{A}) = \alpha \cdot \mathrm{tr}\,(\mathbf{A})$.
- $\mathrm{tr}\,(\mathbf{A}) = \mathrm{tr}\,(\mathbf{A}')$.
- $\mathrm{tr}\,(\mathbf{AB}) = \mathrm{tr}\,(\mathbf{BA})$.
- $\mathrm{tr}\,(\mathbf{ABC}) = \mathrm{tr}\,(\mathbf{CAB}) = \mathrm{tr}\,(\mathbf{BCA})$.
- $\mathrm{tr}\,(\mathbf{A}'\mathbf{A}) = \mathrm{tr}\,(\mathbf{AA}') = \sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n} a_{ij}^2$.
- $\mathrm{tr}\,(\mathbf{I}_m) = m$.
- $\mathrm{tr}\,(\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}') = n$, где матрица $\mathbf{A}$ $(m \times n)$ имеет полный ранг по столбцам, т.е. $\mathrm{rank}(\mathbf{A}) = n$.
- $\mathrm{tr}\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} = \mathrm{tr}\,(\mathbf{A}) + \mathrm{tr}\,(\mathbf{D})$, где $\mathbf{A}$ и $\mathbf{D}$ — квадратные матрицы.

Транспонирование

- $(\mathbf{A}+\mathbf{B})' = \mathbf{A}' + \mathbf{B}'$.
- $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$.

Определитель

- Для матрицы $\mathbf{A}$ (2×2): $|\mathbf{A}| = a_{11}a_{22} - a_{12}a_{21}$.
- $|\mathbf{A}|\,|\mathbf{B}| = |\mathbf{AB}|$.
- $|\mathbf{I}| = 1$.
- $|\alpha\mathbf{A}| = \alpha^m\,|\mathbf{A}|$ для матрицы $\mathbf{A}$ $(m \times m)$.
- $|\mathbf{A}'| = |\mathbf{A}|$.
- $|\mathbf{A}^{-1}| = \frac{1}{|\mathbf{A}|}$.

- Если матрица $\mathbf{A}$ $(m \times m)$ является треугольной (например, диагональной), то $|\mathbf{A}| = \prod\limits_{i=1}^{m} a_{ii}$.
- $|\mathbf{I} + \mathbf{AB}| = |\mathbf{I} + \mathbf{BA}|$.
- $|\mathbf{A} + \mathbf{BD}^{-1}\mathbf{C}|\,|\mathbf{D}| = |\mathbf{D} + \mathbf{CA}^{-1}\mathbf{B}|\,|\mathbf{A}|$.
- $|\mathbf{A} + \mathbf{xy}'| = |\mathbf{A}|\,(1 + \mathbf{y}'\mathbf{Ax})$ для матрицы $\mathbf{A}$ $(m \times m)$ и вектор-столбцов $\mathbf{x}$, $\mathbf{y}$ $(m \times 1)$.
- $\begin{vmatrix} \mathbf{A} & 0 \\ 0 & \mathbf{B} \end{vmatrix} = |\mathbf{A}|\,|\mathbf{B}|$, где $\mathbf{A}$ и $\mathbf{B}$ — квадратные матрицы.
- $\begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} = |\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C}|\,|\mathbf{D}| = |\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B}|\,|\mathbf{A}|$, где $\mathbf{A}$ и $\mathbf{D}$ — квадратные невырожденные матрицы.
- Матрица $\mathbf{A}$ $(m \times m)$ является невырожденной $(\mathrm{rank}(\mathbf{A}) = m)$ тогда и только тогда, когда $|\mathbf{A}| \neq 0$.

Обращение

- Если обратная матрица существует, то она единственна (в частности, левая и правая обратные матрицы совпадают).
- Матрица $\mathbf{A}$ $(m \times m)$ имеет обратную $\mathbf{A}^{-1}$ тогда и только тогда, когда она является невырожденной, т.е. $\mathrm{rank}(\mathbf{A}) = m$.
- Матрица $\mathbf{A}$ $(m \times m)$ имеет обратную $\mathbf{A}^{-1}$ тогда и только тогда, когда $|\mathbf{A}| \neq = 0$.
- Обозначим через a^{ij} элементы обратной матрицы A^{-1}. Тогда $a^{ij} = \frac{(-1)^{i+j}|A_{ji}|}{|A|}$, где A_{ji} $((m-1) \times (m-1))$ получены из матрицы A путем вычеркивания j-й строки и i-го столбца.

Во всех приводимых ниже формулах предполагается, что существуют обратные матрицы там, где это требуется.

- Для матрицы $\mathbf{A}$ (2×2):

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|} \begin{pmatrix} a_{22} & -a_{21} \\ -a_{12} & a_{11} \end{pmatrix} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{pmatrix} a_{22} & -a_{21} \\ -a_{12} & a_{11} \end{pmatrix}.$$

- $\mathbf{A}x = y,\ x = \mathbf{A}^{-1}y.$
- $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}.$
- $\left(\mathbf{A}^{-1}\right)^{-1} = \mathbf{A}.$
- $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'.$
- Если $\mathbf{A}$ $(m \times m)$ — ортогональная матрица, то $\mathbf{A}' = \mathbf{A}^{-1}$.
- Для диагональной матрицы $\mathbf{A} = \operatorname{diag}(a_{11}, \ldots, a_{mm})$ выполнено:

$$\mathbf{A}^{-1} = \operatorname{diag}(1/a_{11}, \ldots, 1/a_{mm}).$$

- $(\mathbf{A} + \mathbf{B})^{-1} = \mathbf{A}^{-1}\left(\mathbf{A}^{-1} + \mathbf{B}^{-1}\right)^{-1}\mathbf{B}^{-1}.$
- $\left(\mathbf{A} + \mathbf{BD}^{-1}\mathbf{C}\right)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} + \mathbf{CA}^{-1}\mathbf{B})^{-1}\mathbf{CA}^{-1}.$
- $(\mathbf{I} + \mathbf{AB})^{-1} = \mathbf{I} - \mathbf{A}(\mathbf{I} + \mathbf{BA})^{-1}\mathbf{B}.$
- $\begin{pmatrix} \mathbf{A} & 0 \\ 0 & \mathbf{B} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{A}^{-1} & 0 \\ 0 & \mathbf{B}^{-1} \end{pmatrix}$, где $\mathbf{A}$ и $\mathbf{B}$ — квадратные матрицы.
- $\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} (\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C})^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C})^{-1} & (\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})^{-1} \end{pmatrix}$, где $\mathbf{A}$ и $\mathbf{D}$ — квадратные матрицы.

Положительно определенные матрицы

- Если матрица $\mathbf{A}$ положительно определенная, то $|\mathbf{A}| > 0$. Если матрица $\mathbf{A}$ положительно полуопределенная, то $|\mathbf{A}| \geqslant 0$.
- Если матрица $\mathbf{A}$ положительно (полу-)определенная, то матрица $-\mathbf{A}$ отрицательно (полу-)определенная.

- Если матрица $\mathbf{A}$ положительно определенная, то обратная матрица $\mathbf{A}^{-1}$ также положительно определенная.

- Если матрицы $\mathbf{A}$ и $\mathbf{B}$ положительно (полу-)определенные, то матрицы $\mathbf{A}+\mathbf{B}$ и $\mathbf{AB}$ также положительно (полу-)определенные.

- Если матрица $\mathbf{A}$ положительно определенная, а $\mathbf{B}$ положительно полуопределенная, то $|\mathbf{A}+\mathbf{B}| \geqslant |\mathbf{A}|$. Если $\mathbf{B}$ положительно определенная, то $|\mathbf{A}+\mathbf{B}| > |\mathbf{A}|$.

- Матрицы $\mathbf{A}'\mathbf{A}$ и $\mathbf{A}'\mathbf{B}\mathbf{A}$ $(n \times n)$ являются симметричными положительно полуопределенными для любой матрицы $\mathbf{A}$ $(m \times n)$ и симметричной положительно полуопределенной матрицы $\mathbf{B}$ $(m \times m)$.

- Если матрица $\mathbf{A}$ $(m \times n)$ имеет полный ранг по столбцам, то матрица $\mathbf{A}'\mathbf{A}$ $(n \times n)$ симметричная положительно определенная. Если матрица $\mathbf{B}$ $(m \times m)$ симметричная положительно определенная, то матрица $\mathbf{A}'\mathbf{B}\mathbf{A}$ $(n \times n)$ симметричная положительно определенная.

- Если матрица $\mathbf{A}$ $(m \times m)$ положительно полуопределенная, то существует верхняя треугольная матрица $\mathbf{U}$ $(m \times m)$, такая что $\mathbf{A} = \mathbf{U}'\mathbf{U}$. Также существует нижняя треугольная матрица $\mathbf{L}$ $(m \times m)$, такая что $\mathbf{A} = \mathbf{L}'\mathbf{L}$. Такое представление матрицы называется разложением Холецкого (триангуляризацией).

Идемпотентные матрицы

- Если матрица $\mathbf{A}$ идемпотентная, то матрица $\mathbf{I}-\mathbf{A}$ тоже идемпотентная, причем $\mathbf{A}(\mathbf{I}-\mathbf{A}) = 0$.

- Если матрица $\mathbf{A}$ симметричная и идемпотентная, то $\operatorname{rank}(\mathbf{A}) = \operatorname{tr}(\mathbf{A})$.

- Матрицы $\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ и $\mathbf{I}_m - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'$ являются симметричными и идемпотентными для любой матрицы $\mathbf{A}$ $(m \times n)$, имеющей полный ранг по столбцам. При этом $\operatorname{tr}\left(\mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\right) = n$ и $\operatorname{tr}\left(\mathbf{I}_m - \mathbf{A}(\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\right) = m-n$.

Собственные числа и векторы

- Для матрицы $\mathbf{A}$ $(m \times m)$ $|\mathbf{A}-\lambda\mathbf{I}_m|$ является многочленом m-й степени (характеристическим многочленом) и имеет m корней, $\lambda_1, \ldots, \lambda_m$, в общем случае комплексных, среди которых могут быть кратные. По определению, $\lambda_1, \ldots, \lambda_m$ являются собственными числами матрицы $\mathbf{A}$.

- У матрицы $\mathbf{A}$ $(m \times m)$ существует не больше m различных собственных чисел.

- Если $\mathbf{x}$ — собственный вектор матрицы $\mathbf{A}$, соответствующий собственному числу λ, то для любого скаляра $\alpha \neq 0$, $\alpha\mathbf{x}$ — тоже собственный вектор, соответствующий собственному числу λ.

- Если $\lambda_1, \ldots, \lambda_m$ — собственные числа матрицы $\mathbf{A}$, то $\operatorname{tr}(\mathbf{A}) = \sum\limits_{i=1}^{m} \lambda_i$, $|\mathbf{A}| = \prod\limits_{i=1}^{m} \lambda_i$.

- Если матрица $\mathbf{A}$ идемпотентная, то все ее собственные числа равны 0 или 1.

- Все собственные числа вещественной симметричной матрицы вещественны.

- Если $\mathbf{x}$ и $\mathbf{y}$ — собственные векторы вещественной симметричной матрицы, соответствующие двум различным собственным числам, то они ортогональны: $\mathbf{x}'\mathbf{y} = 0$.

- Если матрица $\mathbf{A}$ $(m \times m)$ является вещественной и симметричной, то существуют матрицы $\mathbf{H}$ и Λ, где $\mathbf{H}$ $(m \times m)$ — ортогональная матрица ($\mathbf{H}' = \mathbf{H}^{-1}$), столбцы которой — собственные векторы матрицы $\mathbf{A}$, а Λ $(m \times m)$ — диагональная матрица, состоящая из соответствующих собственных чисел матрицы $\mathbf{A}$, такие что выполнено $\mathbf{A} = \mathbf{H}\Lambda\mathbf{H}'$.

- Если матрица $\mathbf{A}$ $(m \times m)$ является вещественной, симметричной, невырожденной, то $\mathbf{A}^{-1} = \mathbf{H}\Lambda^{-1}\mathbf{H}'$.

- Вещественная симметричная матрица является положительно полуопределенной (определенной) тогда и только тогда, когда все ее собственные числа неотрицательны (положительны). Вещественная симметричная матрица является отрицательно полуопределенной (определенной) тогда и только тогда, когда все ее собственные числа неположительны (отрицательны).

- Если матрица $\mathbf{A}$ $(m \times m)$ является вещественной, симметричной и положительно полуопределенной, то $\mathbf{A} = \mathbf{B}'\mathbf{B} = \mathbf{B}^2$, где $\mathbf{B} = \mathbf{H}\Lambda^{1/2}\mathbf{H}'$ $(m \times m)$ — вещественная, симметричная и положительно полуопределенная матрица; $\Lambda^{1/2} = \operatorname{diag}\{\lambda_g^{1/2}\}$.

- Пусть $\lambda_1 \leqslant \cdots \leqslant \lambda_m$ — собственные числа вещественной симметричной матрицы $\mathbf{A}$ $(m \times m)$. Тогда собственый вектор $\mathbf{x}_1$, соответствующий

наименьшему собственому числу λ_1, является решением задачи

$$\begin{cases} \mathbf{x}'\mathbf{A}\mathbf{x} \to \min\limits_{\mathbf{x}}! \\ \mathbf{x}'\mathbf{x} = 1. \end{cases}$$

- Пусть $\lambda_1 \leqslant \cdots \leqslant \lambda_m$ — собственные числа вещественной симметричной матрицы $\mathbf{A}$ $(m \times m)$. Тогда $\lambda_m = \max\limits_{\mathbf{x}} \frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{x}}$ и $\lambda_1 = \min\limits_{\mathbf{x}} \frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{x}}$.

Произведение Кронекера

- $\mathbf{A} \otimes (\mathbf{B} + \mathbf{C}) = \mathbf{A} \otimes \mathbf{B} + \mathbf{A} \otimes \mathbf{C}$ и $(\mathbf{A} + \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes \mathbf{C} + \mathbf{B} \otimes \mathbf{C}$.
- $\mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C}) = (\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C}$.
- $\alpha \otimes \mathbf{A} = \alpha \cdot \mathbf{A} = \mathbf{A} \otimes \alpha$.
- $(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}'$.
- $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$.
- $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$.
- $|\mathbf{A} \otimes \mathbf{B}| = |\mathbf{A}|^n |\mathbf{B}|^m$ для матриц $\mathbf{A}$ $(m \times m)$ и B $(n \times n)$.
- $\operatorname{tr}(\mathbf{A} \otimes \mathbf{B}) = \operatorname{tr}(\mathbf{A}) \cdot \operatorname{tr}(\mathbf{B})$.
- $\operatorname{rank}(\mathbf{A} \otimes \mathbf{B}) = \operatorname{rank}(\mathbf{A}) \cdot \operatorname{rank}(\mathbf{B})$.

А.2. Матричное дифференцирование

А.2.1. Определения

- Производной скалярной функции $s(\mathbf{x})$ по вектор-столбцу $\mathbf{x}$ $(n \times 1)$ или, другими словами, градиентом является вектор-столбец $(n \times 1)$

$$\frac{\partial s}{\partial \mathbf{x}} = \begin{pmatrix} \dfrac{\partial s}{\partial x_1} \\ \vdots \\ \dfrac{\partial s}{\partial x_n} \end{pmatrix}.$$

- Производной скалярной функции $s(\mathbf{x})$ по вектор-строке $\mathbf{x}$ $(1 \times n)$ является вектор-строка $(1 \times n)$

$$\frac{\partial s}{\partial \mathbf{x}} = \left(\frac{\partial s}{\partial x_1}, \ldots, \frac{\partial s}{\partial x_n} \right).$$

- Производной векторной функции $\mathbf{y}(\mathbf{x})$ $(n \times 1)$ по вектору $\mathbf{x}$ $(1 \times m)$ или, другими словами, матрицей Якоби является матрица $(n \times m)$

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \left\{ \frac{\partial y_i}{\partial x_j} \right\}_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}}.$$

- Производной векторной функции $\mathbf{y}(\mathbf{x})$ $(1 \times n)$ по вектору $\mathbf{x}$ $(m \times 1)$ является матрица $(m \times n)$

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \left\{ \frac{\partial y_j}{\partial x_i} \right\}_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}}.$$

- Производной скалярной функции $s(\mathbf{A})$ по матрице $\mathbf{A}$ $(m \times n)$ является матрица $(m \times n)$

$$\frac{\partial s}{\partial \mathbf{A}} = \left\{ \frac{\partial s}{\partial a_{ij}} \right\}_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}}.$$

- Производной матричной функции $\mathbf{A}(s)$ по скаляру s является матрица $(m \times \times n)$

$$\frac{\partial \mathbf{A}}{\partial s} = \left\{ \frac{\partial a_{ij}}{\partial s} \right\}_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}}.$$

- Второй производной скалярной функции $s(\mathbf{x})$ по вектору-столбцу $\mathbf{x}$ $(n \times 1)$ или, другими словами, матрицей Гессе является матрица $(n \times n)$

$$\frac{\partial^2 s}{\partial \mathbf{x} \partial \mathbf{x}'} = \left\{ \frac{\partial^2 s}{\partial x_i \partial x_j} \right\}_{\substack{i=1,\ldots,m \\ j=1,\ldots,n}}.$$

A.2.2. Свойства

- $\dfrac{\partial \mathbf{x}'}{\partial \mathbf{x}} = \mathbf{I}$ и $\dfrac{\partial \mathbf{x}}{\partial \mathbf{x}'} = \mathbf{I}$.
- $\dfrac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}'} = \mathbf{A}$ и $\dfrac{\partial \mathbf{x}'\mathbf{A}}{\partial \mathbf{x}} = \mathbf{A}$.

- $\dfrac{\partial \mathbf{x}'\mathbf{y}}{\partial \mathbf{x}} = \dfrac{\partial \mathbf{y}'\mathbf{x}}{\partial \mathbf{x}} = \mathbf{y}$.
- $\dfrac{\partial \mathbf{x}'\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{x}$.
- $\dfrac{\partial \mathbf{x}'\mathbf{A}\mathbf{y}}{\partial \mathbf{x}} = \mathbf{A}\mathbf{y}$ и $\dfrac{\partial \mathbf{x}'\mathbf{A}\mathbf{y}}{\partial \mathbf{y}'} = \mathbf{x}'\mathbf{A}$.
- $\dfrac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}')\mathbf{x}$.

  Для симметричной матрицы $\mathbf{A}$: $\dfrac{\partial \mathbf{x}'\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x} = 2\mathbf{A}'\mathbf{x}$.
- $\dfrac{\partial \mathbf{x}'\mathbf{A}\mathbf{y}}{\partial \mathbf{A}} = \mathbf{x}\mathbf{y}'$.
- $\dfrac{\partial \mathbf{x}'\mathbf{A}^{-1}\mathbf{y}}{\partial \mathbf{A}} = (\mathbf{A}')^{-1}\,\mathbf{x}\mathbf{y}'\,(\mathbf{A}')^{-1}$.
- $\dfrac{\partial \operatorname{tr}(\mathbf{A})}{\partial \mathbf{A}} = \mathbf{I}$.
- $\dfrac{\partial \operatorname{tr}(\mathbf{A}\mathbf{B})}{\partial \mathbf{A}} = \mathbf{B}'$ и $\dfrac{\partial \operatorname{tr}(\mathbf{A}\mathbf{B})}{\partial \mathbf{B}} = \mathbf{A}'$.
- $\dfrac{\partial \operatorname{tr}(\mathbf{A}'\mathbf{A})}{\partial \mathbf{A}} = 2\mathbf{A}$.
- $\dfrac{\partial \operatorname{tr}(\mathbf{A}'\mathbf{B}\mathbf{A})}{\partial \mathbf{A}} = (\mathbf{B} + \mathbf{B}')\,\mathbf{A}$.
- $\dfrac{\partial \operatorname{tr}(\mathbf{A}'\mathbf{B}\mathbf{A})}{\partial \mathbf{B}} = \mathbf{A}\mathbf{A}'$.
- $\dfrac{\partial |\mathbf{A}|}{\partial \mathbf{A}} = |\mathbf{A}|\,(\mathbf{A}')^{-1}$.
- $\dfrac{\partial \ln |\mathbf{A}|}{\partial \mathbf{A}} = (\mathbf{A}')^{-1}$.
- $\dfrac{\partial \ln |\mathbf{A}'\mathbf{B}\mathbf{A}|}{\partial \mathbf{A}} = \mathbf{B}\mathbf{A}\,(\mathbf{A}'\mathbf{B}\mathbf{A})^{-1} + \mathbf{B}'\mathbf{A}\,(\mathbf{A}'\mathbf{B}'\mathbf{A})^{-1}$.
- $\dfrac{\partial (\mathbf{A}\mathbf{B})}{\partial s} = \mathbf{A}\dfrac{\partial \mathbf{B}}{\partial s} + \dfrac{\partial \mathbf{A}}{\partial s}\mathbf{B}$.
- $\dfrac{\partial \mathbf{A}^{-1}}{\partial s} = -\mathbf{A}^{-1}\dfrac{\partial \mathbf{A}}{\partial s}\mathbf{A}^{-1}$.

- $\dfrac{ds\,(\mathbf{A})}{dt} = \operatorname{tr}\left(\dfrac{\partial s}{\partial \mathbf{A}}\dfrac{d\mathbf{A}'}{dt}\right) = \operatorname{tr}\left(\dfrac{\partial s}{\partial \mathbf{A}'}\dfrac{d\mathbf{A}}{dt}\right)$.
- $\dfrac{\partial \operatorname{tr}(\mathbf{A})}{\partial s} = \operatorname{tr}\left(\dfrac{\partial \mathbf{A}}{\partial s}\right)$.
- $\dfrac{\partial \ln|\mathbf{A}|}{\partial s} = \operatorname{tr}\left(\mathbf{A}^{-1}\dfrac{\partial \mathbf{A}}{\partial s}\right)$.
- $\dfrac{d\mathbf{y}\,(\mathbf{x})}{ds} = \dfrac{\partial \mathbf{y}}{\partial \mathbf{x}'}\dfrac{d\mathbf{x}}{ds}$.

A.3. Сведения из теории вероятностей и математической статистики

A.3.1. Характеристики случайных величин

Определения

- Функцией распределения случайной величины x называется функция $F_x(z) = \Pr(x \leqslant z)$, сопоставляющая числу z вероятность того, что x не превышает z. Функция распределения полностью характеризует отдельную случайную величину.
- Если случайная величина x непрерывна, то она имеет плотность $f_x(\cdot)$, которая связана с функцией распределения соотношениями $f_x(z) = F'_x(z)$.
- Квантилью уровня F, где $F \in [0;1]$, (F-квантилью) непрерывной случайной величины x называется число x_F, такое что

$$F_x(x_F) = \int_{-\infty}^{x_F} f_x(t)dt = F.$$

- Медианой $x_{0,5}$ называется $0,5$-квантиль.
- Модой непрерывной случайной величины называется величина, при которой плотность распределения достигает максимума, т.е. $\overset{\circ}{x} = \arg\max\limits_z f_x(z)$.
- Если распределение непрерывной случайной величины x симметрично относительно нуля, т.е. $f_x(z) = f_x(-z)$ и $F_x(-x_F) = 1 - F_x(x_F)$, то двусторонней F-квантилью называется число x_F, такое что

$$F_x(x_F) - F_x(-x_F) = \int_{-x_F}^{x_F} f_x(t)dt = F.$$

- Математическим ожиданием непрерывной случайной величины x называется $\mathbf{E}(x) = \int_{-\infty}^{+\infty} t f_x(t) dt$.

- Математическое ожидание является начальным моментом первого порядка. Начальным (нецентральным) моментом q-го порядка называется $\mathbf{E}(x^q) = \int_{-\infty}^{+\infty} t^q f_x(t) dt$.

- По случайной величине x может быть построена соответствующая ей центрированная величина $\hat{x}$: $\hat{x} = x - \mathbf{E}(x)$, имеющая аналогичные законы распределения и нулевое математическое ожидание.

- Центральным моментом q-го порядка случайной величины x называется начальный момент q-го порядка для соответствующей центрированной величины $\hat{x}$, т.е. $\mathbf{E}(\hat{x}^q) = \mathbf{E}\left[(x - \mathbf{E}(x))^q\right]$. Для непрерывной случайной величины центральный момент q-го порядка равен

$$\mu_q = \int_{-\infty}^{+\infty} (t - \mathbf{E}(x))^q f_x(t) dt.$$

- Дисперсией случайной величины называется центральный момент второго порядка. Для непрерывной случайной величины дисперсия равна

$$\mathbf{var}(x) = \sigma_x^2 = \mathbf{E}(\hat{x}^2) = \mathbf{E}\left[(x - \mathbf{E}(x))^2\right] = \int_{-\infty}^{+\infty} (t - \mathbf{E}(x))^2 f_x(t) dt.$$

- Среднеквадратическим отклонением называется квадратный корень из дисперсии $\sigma_x = \sqrt{\sigma_x^2}$. Нормированной (стандартизованной) случайной величиной называется $\dfrac{x - \mathbf{E}(x)}{\sigma_x}$.

- Коэффициентом асимметрии называется начальный момент третьего порядка нормированной случайной величины, т.е.

$$\delta_3 = \mathbf{E}\left[\left(\frac{x - \mathbf{E}(x)}{\sigma_x}\right)^3\right] = \frac{\mu_3}{\sigma_x^3}.$$

- Куртозисом называется начальный момент четвертого порядка нормированной случайной величины, т.е.

$$\delta_4 = \mathbf{E}\left[\left(\frac{x - \mathbf{E}(x)}{\sigma_x}\right)^4\right] = \frac{\mu_4}{\sigma_x^4}.$$

  Коэффициентом эксцесса называется $\delta_4 - 3$.

- Для n-мерного случайного вектора $\mathbf{x} = (x_1, \ldots, x_n)'$ (многомерной случайной величины) функцией распределения называется

$$F_{x_1,\ldots,x_n}(z_1, \ldots, z_n) = \Pr(x_1 \leqslant z_1, \ldots, x_n \leqslant z_n).$$

- Если распределение случайного вектора $\mathbf{x}$ непрерывно, то он имеет плотность $f_x(\cdot)$ (называемую совместной плотностью случайных величин $x_1, \ldots, x_n$), которая связана с функцией распределения соотношениями $f_{x_1,\ldots,x_n}(\mathbf{z}) = \dfrac{\partial^n F_{x_1,\ldots,x_n}(\mathbf{z})}{\partial x_1 \cdots \partial x_n}$.

  Случайные величины $x_1, \ldots, x_n$ называются независимыми (в совокупности), если $F_{x_1,\ldots,x_n}(z_1, \ldots, z_n) = F_{x_1}(z_1) \cdots F_{x_n}(z_n)$.

- Ковариацией случайных величин x и y называется

$$\mathbf{cov}(x, y) = \mathbf{E}\left[(x - \mathbf{E}(x))\,(y - \mathbf{E}(y))\right].$$

- Корреляцией случайных величин x и y называется $\rho_{x,y} = \dfrac{\mathbf{cov}(x, y)}{\sqrt{\mathbf{var}(x)\mathbf{var}(x)}}$.

- Ковариационной матрицей n-мерной случайной величины $\mathbf{x} = (x_1, \ldots, x_n)'$ называется

$$\Gamma_{\mathbf{x}} = \mathbf{var}(\mathbf{x}) = \begin{pmatrix} \mathbf{cov}(x_1, x_1) & \cdots & \mathbf{cov}(x_1, x_n) \\ \vdots & \ddots & \vdots \\ \mathbf{cov}(x_1, x_n) & \cdots & \mathbf{cov}(x_n, x_n) \end{pmatrix} = \\ = \mathbf{E}\left[(\mathbf{x} - \mathbf{E}(\mathbf{x}))\,(\mathbf{x} - \mathbf{E}(\mathbf{x}))'\right].$$

- Корреляционной матрицей n-мерной случайной величины $\mathbf{x} = (x_1, \ldots, x_n)'$ называется

$$P_{\mathbf{x}} = \begin{pmatrix} 1 & \rho_{x_1,x_2} & \cdots & \rho_{x_1,x_n} \\ \rho_{x_1,x_2} & 1 & \cdots & \rho_{x_2,x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{x_1,x_n} & \rho_{x_2,x_n} & \cdots & 1 \end{pmatrix}.$$

Функция распределения и плотность

- Функция распределения имеет следующие свойства: это неубывающая, непрерывная справа функция, $0 \leqslant F_x(z) \leqslant 1$, причем $\lim\limits_{z\to-\infty} F_x(z) = 0$ и $\lim\limits_{z\to\infty} F_x(z) = 1$.
- $F_x(z) = \int_{-\infty}^{z} f_x(t)dt$.
- $f_x(z) \geqslant 0$.
- $\int_{-\infty}^{+\infty} f_x(t)dt = 1$.
- Вероятность того, что $x \in [a, b]$, равна $\Pr(a \leqslant x \leqslant b) = \int_a^b f_x(t)dt$.
- Для многомерной случайной величины

$$F_{x_1,\ldots,x_n}(z_1, \ldots, z_n) = \int_{-\infty}^{z_1} \cdots \int_{-\infty}^{z_n} f_{x_1,\ldots,x_n}(t_1, \ldots, t_n)dt_n \ldots dt_1.$$

- Если случайные величины $x_1, \ldots, x_n$ независимы, то

$$f_{x_1,\ldots,x_n}(z_1, \ldots, z_n) = f_{x_1}(z_1) \cdot \ldots \cdot f_{x_n}(z_n).$$

Математическое ожидание

- Если c — константа, то $\mathbf{E}(c) = c$.
- Если x и y — любые две случайные величины, то

$$\mathbf{E}(x + y) = \mathbf{E}(x) + \mathbf{E}(y).$$

- Если c — константа, то $\mathbf{E}(cx) = c\mathbf{E}(x)$.
- В общем случае $\mathbf{E}(xy) \neq \mathbf{E}(x)\mathbf{E}(y)$.
- Если функция $f(\cdot)$ вогнута, то выполнено неравенство Йенсена:

$$\mathbf{E}\left(f(x)\right) \leqslant f\left(\mathbf{E}(x)\right).$$

- Для симметричного распределения выполено $\mathbf{E}(x) = x_{0,5}$.

Дисперсия

- $\mathbf{var}(x) = \mathbf{E}(x^2) - \mathbf{E}(x)^2$.
- Для любой случайной величины x выполнено $\mathbf{var}(x) \geqslant 0$.
- Если c — константа, то выполнено: $\mathbf{var}(c) = 0$; $\mathbf{var}(c + x) = \mathbf{var}(x)$; $\mathbf{var}(cx) = c^2\mathbf{var}(x)$.
- Если x и y — любые две случайные величины, то в общем случае:

$$\mathbf{var}(x + y) \neq \mathbf{var}(x) + \mathbf{var}(y).$$

- Неравенство Чебышёва: $\Pr(|x - \mathbf{E}(x)| > \alpha) \leqslant \dfrac{\mathbf{var}(x)}{\alpha^2}$ для любого положительного числа α.

Ковариация

- $\mathbf{cov}(x, y) = \mathbf{E}(xy) - \mathbf{E}(x)\mathbf{E}(y)$.
- $\mathbf{cov}(x, y) = \mathbf{cov}(y, x)$.
- $\mathbf{cov}(cx, y) = c \cdot \mathbf{cov}(x, y)$.
- $\mathbf{cov}(x + y, z) = \mathbf{cov}(x, z) + \mathbf{cov}(y, z)$.
- $\mathbf{cov}(x, x) = \mathbf{var}(x)$.
- Если x и y независимы, то $\mathbf{cov}(x, y) = 0$. Обратное, вообще говоря, неверно.

Корреляция

- $\rho_{x,y} = \mathbf{cov}(\tilde{x}, \tilde{y})$, где $\tilde{x} = \dfrac{x - \mathbf{E}(x)}{\sigma_x}$ и $\tilde{y} = \dfrac{y - \mathbf{E}(y)}{\sigma_y}$ — соответствующие центрированные нормированные случайные величины. Следовательно, свойства корреляции аналогичны свойствам ковариации.
- $\rho_{x,x} = 1$.
- $-1 \leqslant \rho_{x,y} \leqslant 1$.
- Если $\rho_{x,y} = 0$, то $\mathbf{E}(xy) = \mathbf{E}(x)\mathbf{E}(y)$.

Условные распределения

- Условной вероятностью события A относительно события B называется $\Pr(A|B) = \Pr(A \cap B)/\Pr(B)$.

  Из определения следует, что $\Pr(A \cap B) = \Pr(A|B)\Pr(B) = \Pr(B|A)\Pr(A)$.

- Для независимых событий A и B выполнено $\Pr(A|B) = \Pr(A)$.

- Теорема Байеса: Пусть $A_1, \ldots, A_n, B$ — события, такие что

  (1) $A_i \cap A_j = \varnothing$ при $i \neq j$,

  (2) $B \subset \bigcup\limits_{i=1}^{n} A_i$,

  (3) $\Pr(B) > 0$.

  Тогда

$$\Pr(A_i|B) = \frac{\Pr(B|A_i)\Pr(A_i)}{\Pr(B)} = \frac{\Pr(B|A_i)\Pr(A_i)}{\sum\limits_{j=1}^{n} \Pr(B|A_j)\Pr(A_j)}.$$

- Пусть $(\mathbf{x}, \mathbf{y})$ — случайный вектор, имеющий непрерывное распределение, где вектор $\mathbf{x}$ имеет размерность $m \times 1$, а $\mathbf{y}$ — $n \times 1$. Плотностью маргинального распределения $\mathbf{x}$ называется $f_{\mathbf{x}}(\mathbf{x}) = \int\limits_{R^n} f_{\mathbf{x},\mathbf{y}}(\mathbf{x}, \mathbf{y}) d\mathbf{y}$. Плотностью условного распределения $\mathbf{x}$ относительно $\mathbf{y}$ называется

$$f_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\mathbf{y}) = \frac{f_{\mathbf{x},\mathbf{y}}(\mathbf{x}, \mathbf{y})}{f_{\mathbf{y}}(\mathbf{y})} = \frac{f_{\mathbf{x},\mathbf{y}}(\mathbf{x}, \mathbf{y})}{\int\limits_{R^m} f_{\mathbf{x},\mathbf{y}}(\mathbf{x}, \mathbf{y}) d\mathbf{x}}.$$

- Если $\mathbf{x}$ и $\mathbf{y}$ независимы, то плотность условного распределения совпадает с плотностью маргинального, т.е. $f_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\mathbf{y}) = f_{\mathbf{x}}(\mathbf{x})$.

- Условным математическим ожиданием $\mathbf{x}$ относительно $\mathbf{y}$ называется

$$\mathbf{E}(\mathbf{x}|\mathbf{y}) = \int\limits_{R^m} \mathbf{x} f_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\mathbf{y}) d\mathbf{x}.$$

- Условная дисперсия x относительно $\mathbf{y}$ равна

$$\mathbf{var}(x|\mathbf{y}) = \mathbf{E}\left((x - \mathbf{E}(x|\mathbf{y}))^2 |\mathbf{y}\right).$$

Свойства условного ожидания и дисперсии

- $\mathbf{E}\left(\alpha(\mathbf{y})|\mathbf{y}\right) = \alpha(\mathbf{y})$.
- $\mathbf{E}\left(\alpha(\mathbf{y})\mathbf{x}|\mathbf{y}\right) = \alpha(\mathbf{y})\mathbf{E}\left(\mathbf{x}|\mathbf{y}\right)$.
- $\mathbf{E}\left(\mathbf{x}_1 + \mathbf{x}_2|\mathbf{y}\right) = \mathbf{E}\left(\mathbf{x}_1|\mathbf{y}\right) + \mathbf{E}\left(\mathbf{x}_2|\mathbf{y}\right)$.
- Правило повторного ожидания: $\mathbf{E}\left(\mathbf{E}\left(\mathbf{x}|\mathbf{y},\mathbf{z}\right)|\mathbf{y}\right) = \mathbf{E}\left(\mathbf{x}|\mathbf{y}\right)$.
- Если $\mathbf{x}$ и $\mathbf{y}$ независимы , то $\mathbf{E}\left(\mathbf{x}|\mathbf{y}\right) = \mathbf{E}\left(\mathbf{x}\right)$.
- $\mathbf{var}\left(\alpha(\mathbf{y})|\mathbf{y}\right) = 0$.
- $\mathbf{var}\left(\alpha(\mathbf{y}) + \mathbf{x}|\mathbf{y}\right) = \mathbf{var}\left(\mathbf{x}|\mathbf{y}\right)$.
- $\mathbf{var}\left(\alpha(\mathbf{y})\mathbf{x}|\mathbf{y}\right) = \alpha^2(\mathbf{y})\mathbf{var}\left(\mathbf{x}|\mathbf{y}\right)$.

A.3.2. Распределения, связанные с нормальным

Нормальное распределение

Нормальное (или гауссовское) распределение с математическим ожиданием μ и дисперсией σ^2 обозначается $N\left(\mu,\sigma^2\right)$ и имеет плотность распределения

$$\left(f(z) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(z-\mu)^2}{2\sigma^2}}.\right.$$

Нормальное распределение симметрично относительно μ, и для него выполняется $\mathbf{E}(x) = x_{0,5} = \overset{\circ}{x} = \mu$.

Моменты нормального распределения: $\mu_{2k+1} = 0$ и $\mu_{2k} = (2k-1)!! \cdot \sigma^{2k} = 1 \cdot 3 \cdot \ldots \cdot \sigma^{2k}$ при целых k, в частности, $\mu_4 = 3\sigma^4$.

Коэффициент асимметрии: $\delta_3 = 0$.

Куртозис $\delta_4 = 3$, коэффициент эксцесса равен нулю.

Стандартным нормальным распределением называется $N\left(0,1\right)$. Его плотность $\varphi(z) = \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$; функция распределения $\Phi(z) = \int\limits_{-\infty}^{z} \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}\,dt$.

Распределение хи-квадрат

Распределение хи-квадрат с k степенями свободы обозначается χ^2_k. Его плотность:

$$\begin{cases} f(x) = \dfrac{(x/2)^{k/2-1}}{2^{k/2}\Gamma(k/2)} e^{-x/2}, \quad x \geqslant 0, \\ f(x) = 0, \quad x < 0, \end{cases}$$

где $\Gamma(\cdot)$ — гамма-функция.

Если $x_i \sim N(0,1)$, $i = 1, \ldots, k$ и независимы в совокупности, то $\sum\limits_{i=1}^{k} x_i^2 \sim \chi^2_k$. Если $x \sim \chi^2_k$, то $\mathbf{E}(x) = k$ и $\mathbf{var}(x) = 2k$.

Коэффициент асимметрии: $\delta_3 = \dfrac{2\sqrt{2}}{\sqrt{k}} > 0$.

Куртозис: $\delta_4 = \dfrac{12}{k} + 3$, коэффициент эксцесса $\delta_4 - 3 = \dfrac{12}{k} > 0$.

При больших k распределение хи-квадрат похоже на $N(k, 2k)$.

Распределение Стьюдента

Распределение Стьюдента с k степенями свободы обозначается через t_k. Его также называют t-распределением. Его плотность:

$$f(x) = \frac{\Gamma\left((k+1)/2\right)}{\sqrt{k\pi}\Gamma(k/2)} \left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}}.$$

Если $x_1 \sim N(0,1)$, $x_2 \sim \chi^2_k$ и независимы, то $\dfrac{x_1}{\sqrt{x_2/k}} \sim t_k$.

Распределение Стьюдента симметрично относительно нуля и $x_{0,5} = \mathring{x} = 0$.

Математическое ожидание существует при $k > 1$ и $\mathbf{E}(x) = 0$.

При $k \leqslant n$ не существует n-го момента.

Дисперсия: $\mathbf{var}(x) = \dfrac{k}{k-2}$ (существует при $k > 2$).

Коэффициент асимметрии: $\delta_3 = 0$ (существует при $k > 3$).

Куртозис: $\delta_4 = 3\dfrac{k-2}{k-4}$; коэффициент эксцесса: $\delta_4 - 3 = \dfrac{6}{k-4}$ (существуют при $k > 4$).

При больших k распределение Стьюдента похоже на $N(0,1)$.

Распределение Фишера

Распределение Фишера с k_1 и k_2 степенями свободы обозначается F_{k_1,k_2}. Его также называют F-распределением или распределением Фишера—Снедекора. Его плотность:

$$\begin{cases} f(x) = \dfrac{\Gamma\left((k_1+k_2)/2\right)}{\Gamma(k_1/2)\Gamma(k_2/2)} k_1^{k_1/2} k_2^{k_2/2} \dfrac{x^{k_1/2-1}}{(k_1 x + k_2)^{(k_1+k_2)/2}}, \quad x \geqslant 0, \\ f(x) = 0, \quad x < 0. \end{cases}$$

Если $x_1 \sim \chi^2_{k_1}$, $x_2 \sim \chi^2_{k_2}$ и независимы, то $\dfrac{x_1/k_1}{x_2/k_2} \sim F_{k_1,k_2}$.

Если $x \sim F_{k_1,k_2}$, то

$$\begin{aligned} \mathbf{E}(x) &= \frac{k_2}{k_2-2}, && \text{при } k_2 > 2, \\ \mathbf{var}(x) &= \frac{2k_2^2(k_1+k_2-2)}{k_1(k_2-2)^2(k_2-4)}, && \text{при } k_2 > 4, \\ \delta_3 &= \frac{2(2k_1+k_2-2)}{k_2-6}\sqrt{\frac{2(k_2-4)}{k_1(k_1+k_2-2)}}, && \text{при } k_2 > 6, \\ \delta_4 - 3 &= \frac{12\big[(k_2-2)^2(k_2-4) + k_1(5k_2-22)(k_1+k_2-2)\big]}{k_1(k_1+k_2-2)(k_2-6)(k_2-8)}, && \text{при } k_2 > 8. \end{aligned}$$

Многомерное нормальное распределение

n-мерное нормальное распределение с математическим ожиданием μ $(n \times 1)$ и ковариационной матрицей Σ $(n \times n)$ обозначается $N(\mu, \Sigma)$. Его плотность:

$$f(\mathbf{z}) = (2\pi)^{-n/2} |\Sigma|^{-1/2} e^{-\frac{1}{2}(\mathbf{z}-\mu)'\Sigma^{-1}(\mathbf{z}-\mu)}.$$

Свойства многомерного нормального распределения:

- Если $\mathbf{x} \sim N(\mu, \Sigma)$, то $\mathbf{A}\mathbf{x} + \mathbf{b} \sim N(\mathbf{A}\mu + \mathbf{b}, \mathbf{A}\Sigma\mathbf{A}')$.
- Если $\mathbf{x} \sim N(\mathbf{0}_n, \sigma^2\mathbf{I}_n)$, то $\dfrac{\mathbf{x}'\mathbf{x}}{\sigma^2} \sim \chi^2_n$.
- Если $\mathbf{x} \sim N(\mathbf{0}, \Sigma)$, где Σ $(n \times n)$ — невырожденная матрица, то $\mathbf{x}'\Sigma^{-1}\mathbf{x} \sim \chi^2_n$.

- Если $\mathbf{x} \sim N\left(\mathbf{0}, \mathbf{A}\left(\mathbf{A}'\mathbf{A}\right)^{-1}\mathbf{A}'\right)$, где $\mathbf{A}$ $(n \times k)$ — матрица, имеющая полный ранг по столбцам, то $\mathbf{x}'\mathbf{x} \sim \chi^2_k$. Если $\mathbf{x} \sim N\left(\mathbf{0}, \mathbf{I} - \mathbf{A}\left(\mathbf{A}'\mathbf{A}\right)^{-1}\mathbf{A}'\right)$, где $\mathbf{A}$ $(n \times k)$ — матрица, имеющая полный ранг по столбцам, то $\mathbf{x}'\mathbf{x} \sim \chi^2_{n-k}$.

- Если $\mathbf{x} = (x_1, \ldots, x_n) \sim N\left(\mu, \mathrm{diag}(\sigma_1^2, \ldots, \sigma_n^2)\right)$, то $x_1, \ldots, x_n$ независимы в совокупности и $x_i \sim N\left(\mu_i, \sigma_i^2\right)$.

- Если совместное распределение случайных векторов $\mathbf{x}$ и $\mathbf{y}$ является многомерным нормальным:

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}\right),$$

  то маргинальное распределение $\mathbf{x}$ имеет вид $\mathbf{x} \sim N\left(\mu_x, \Sigma_{xx}\right)$, а условное распределение $\mathbf{x}$ относительно $\mathbf{y}$ имеет вид

$$\mathbf{x}|\mathbf{y} \sim N\left(\mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}\left(\mathbf{y} - \mu_y\right), \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}\right).$$

  Аналогично $\mathbf{y} \sim N\left(\mu_y, \Sigma_{yy}\right)$ и

$$\mathbf{y}|\mathbf{x} \sim N\left(\mu_y + \Sigma_{yx}\Sigma_{xx}^{-1}\left(\mathbf{x} - \mu_x\right), \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}\right).$$

А.3.3. Проверка гипотез

Пусть $x_1, \ldots, x_n$ — случайная выборка из распределения F_θ, заданного параметром $\theta \in \Theta$.

Нулевая гипотеза H_0 относительно параметра θ состоит в том, что он принадлежит некоторому более узкому множеству: $\theta \in \Theta_0$, где $\Theta_0 \subset \Theta$. Альтернативная гипотеза H_1 состоит в том, что параметр принадлежит другому множеству: $\theta \in \Theta_1$, где $\Theta_1 = \Theta \backslash \Theta_0$. Рассматривается некоторая статистика s, которая является функцией от выборки: $s = s(x_1, \ldots, x_n)$. Процедуру (правило) проверки гипотезы называют статистическим критерием или статистическим тестом. Суть проверки гипотезы H_0 против альтернативной гипотезы H_1 состоит в том, что задаются две непересекающиеся области, S_0 и S_1, такие что $S_0 \cap S_1$ — вся область значений статистики s. Если $s \in S_0$, то нулевая гипотеза (H_0) принимается, а если $s \in S_1$, то нулевая гипотеза отвергается.

Обычно $S_0 = (-\infty, s^*]$ и $S_1 = (s^*, +\infty)$, где s^* — критическая граница. Такой критерий называется односторонним. При этом критерий состоит в следующем:

если $s < s^*$, то H_0 принимается,
если $s > s^*$, то H_0 отвергается.

S_0 и S_1 выбираются так, чтобы в случае, когда H_0 верна, вероятность того, что $s \in S_1$, была бы равна некоторой заданной малой вероятности α. Как правило, на практике используют вероятность $\alpha = 0.05$ (хотя это не имеет под собой каких-либо теоретических оснований).

Ошибкой первого рода называется ошибка, состоящая в том, что отвергается верная нулевая гипотеза. Вероятность ошибки первого рода равна α. Вероятность ошибки первого рода называется уровнем значимости или размером. Вероятность $1 - \alpha$ называют уровнем доверия.

Ошибкой второго рода называется ошибка, состоящая в том, что принимается неверная нулевая гипотеза. Вероятность ошибки второго рода обозначают β.

Мощностью критерия называют величину $1 - \beta$. Мощность характеризует, насколько хорошо работает критерий. Мощность должна быть как можно большей при данном α. Требуется, по крайней мере, чтобы $\alpha < 1 - \beta$. Критерий, не удовлетворяющий этому условию, называют смещенным.

Альтернативный способ проверки гипотез использует вероятность ошибки первого рода, если принять s^* равной s, т.е. вероятность того, что $s > s^*$. Эту вероятность называют уровнем значимости или P-значением. Обозначим ее pv. При заданной вероятности α критерий состоит в следующем:

если $pv > \alpha$, то H_0 принимается,
если $pv < \alpha$, то H_0 отвергается.

Еще один способ проверки гипотез основан на доверительных областях для параметра θ. Пусть D — доверительная область для параметра θ, такая что $\theta \in D$ с некоторой веротностью $1 - \alpha$, и пусть проверяется гипотеза H_0: $\theta = \theta_0$ против альтернативной гипотезы H_1: $\theta \neq \theta_0$. Критерий состоит в следующем:

если $\theta_0 \in D$, то H_0 принимается,
если $\theta_0 \notin D$, то H_0 отвергается.

Отметим, что в этом случае θ_0 не случайная величина; случайной является доверительная область D.

А.4. Линейные конечно-разностные уравнения

Конечно-разностное уравнение p-го порядка имеет вид:

$$\alpha_0 y_t + \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} = u_t,$$

где u_t — известная последовательность, $\alpha_0, \alpha_1, \ldots, \alpha_p$ — известные коэффициенты, а последовательность y_t следует найти. Это уравнение также можно записать через лаговый многочлен:

$$\alpha(\mathbf{L})\, y_t = \left(\alpha_0 + \alpha_1 \mathbf{L} + \alpha_2 \mathbf{L}^2 + \cdots + \alpha_p \mathbf{L}^p\right) y_t = u_t,$$

где $\mathbf{L}$ — лаговый оператор ($\mathbf{L} y_t = y_{t-1}$).

Если даны p последовательных значений последовательности y_t, например, $y_1, \ldots, y_p$, то другие значения можно найти по рекуррентной формуле. При $t > p$ получаем

$$y_t = \frac{1}{\alpha_0}\left(u_t - (\alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p})\right).$$

Конечно-разностное уравнение называется однородным, если $u_t = 0$. Общее решение конечно-разностного уравнения имеет вид:

$$y_t = y_t^0 + \tilde{y}_t,$$

где y_t^0 — общее решение соответствующего однородного уравнения, а $\tilde{y}_t = \alpha^{-1}(\mathbf{L})\, u_t$ — частное решение неоднородного уравнения.

А.4.1. Решение однородного конечно-разностного уравнения

Если $\lambda_j,\ j = 1, \ldots, p$ — корни характеристического уравнения

$$\alpha(\lambda) = \alpha_0 + \alpha_1 \lambda + \alpha_2 \lambda^2 + \cdots + \alpha_p \lambda^p = 0,$$

тогда $\alpha(\lambda) = \alpha_0 \prod\limits_{j=1}^{p} (1 - \lambda/\lambda_j)$.

Последовательность $y_t^{(j)} = \lambda_j^{-t}$ является решением однородного конечно-разностного уравнения. Действительно, в разложении $\alpha(\mathbf{L})$ на множители имеется множитель $1 - \mathbf{L}/\lambda_j$ и

$$(1 - \mathbf{L}/\lambda_j)\, y_t^{(j)} = (1 - \mathbf{L}/\lambda_j)\, \lambda_j^{-t} = \lambda_j^{-t} - \mathbf{L}\lambda_j^{-t-1} = \lambda_j^{-t} - \lambda_j^{-t} = 0.$$

Если все корни $\lambda_j,\ j = 1, \ldots, p$ различные, то общее решение однородного конечно-разностного уравнения имеет вид:

$$y_t^0 = C_1 y_t^{(1)} + \ldots + C_p y_t^{(p)} = C_1 \lambda_1^{-t} + \ldots + C_p \lambda_p^{-t}.$$

Если не все корни различны, то для корня λ_j кратности m соответствующее слагаемое имеет вид $\left(C_{1j} + C_{2j} t + \ldots + C_{mj} t^{m-1}\right) \lambda_j^{-t}$.

Если λ_1 и λ_2 — пара комплексно-сопряженных корней, т.е.

$$\begin{aligned} \lambda_1 &= R\left(\cos(\varphi) + i \sin(\varphi)\right) = Re^{i\varphi} \quad \text{и} \\ \lambda_2 &= R\left(\cos(\varphi) - i \sin(\varphi)\right) = Re^{-i\varphi}, \end{aligned}$$

то два слагаемых $C_1 y_t^{(1)} + C_2 y_t^{(2)} = C_1 \lambda_1^{-t} + C_2 \lambda_2^{-t}$, где C_1, C_2 являются комплексно-сопряженными, можно заменить на

$$\begin{aligned} C_1 \lambda_1^{-t} + C_2 \lambda_2^{-t} &= C_1 (Re^{i\varphi})^{-t} + C_2 (Re^{-i\varphi})^{-t} = \\ &= R^{-t} \left(C_1 e^{-i\varphi t} + C_2 e^{i\varphi t}\right) = R^{-t} \left(A \cos(\varphi t) + B \sin(\varphi t)\right). \end{aligned}$$

Если известны p значений последовательности y_t^0, например, $y_1^0, \ldots, y_p^0$, то коэффициенты C_j, C_{jk}, A, B находятся из решения системы p линейных уравнений. Например, если все корни $\lambda_j,\ j = 1, \ldots, p$ различны, то коэффициенты C_j находятся из системы уравнений

$$\begin{pmatrix} \lambda_1^{-1} & \cdots & \lambda_p^{-1} \\ \vdots & & \vdots \\ \lambda_1^{-p} & \cdots & \lambda_p^{-p} \end{pmatrix} \begin{pmatrix} C_1 \\ \vdots \\ C_p \end{pmatrix} = \begin{pmatrix} y_1^0 \\ \vdots \\ y_p^0 \end{pmatrix}.$$

A.5. Комплексные числа

Комплексным числом z называется пара (a, b), где a— действительная часть, а b — мнимая часть числа. Комплексное число записывают в виде $z = a + ib$, где $i = \sqrt{-1}$.

- Сложение: если $z_1 = a_1 + ib_1,\ z_2 = a_2 + ib_2$, то $z_1 + z_2 = (a_1 + a_2) + i\,(b_1 + b_2)$; если $z = a + ib$, то $-z = (-a) + i(-b) = -a - ib$.
- Вычитание: если $z_1 = a_1 + ib_1,\ z_2 = a_2 + ib_2$, то $z_1 - z_2 = (a_1 - a_2) + i\,(b_1 - b_2)$.

- Умножение: если $z_1 = a_1 + ib_1,\ z_2 = a_2 + ib_2$, то $z_1 z_2 = a_1 a_2 - b_1 b_2 + i\,(a_2 b_1 + a_1 b_2)$.
- Обратное число: если $z = a + ib$, то $z^{-1} = \dfrac{a}{a^2 + b^2} + i(-\dfrac{b}{a^2 + b^2}) = \dfrac{a - ib}{a^2 + b^2}$.
- Деление: $z_1 = a_1 + ib_1,\ z_2 = a_2 + ib_2$, то $\dfrac{z_1}{z_2} = z_1 z_2^{-1}$.
- Для числа $z = a + ib$ число $z^* = a - ib$ называют комплексно-сопряженным.
- Модуль: если $z = a + ib$, то $|z| = \sqrt{a^2 + b^2}$.
- Выполняется свойство $zz^* = |z|^2$
- Любое комплексное число $z = a + ib$ можно представит в виде $z = |z| \cos\varphi + i|z| \sin\varphi$, где угол φ такой что $\cos\varphi = \dfrac{a}{|z|}$ и $\sin\varphi = \dfrac{b}{|z|}$. Угол φ называют аргументом z.
- Формула Эйлера: $\cos\varphi + i\sin\varphi = e^{i\varphi}$ и $\cos\varphi - i\sin\varphi = e^{-i\varphi}$.

$$\cos\varphi = \frac{1}{2}(e^{i\varphi} + e^{-i\varphi}) \quad \sin\varphi = \frac{1}{2i}(e^{i\varphi} - e^{-i\varphi}).$$

- Формула Муавра: если $z = \cos\varphi + i\sin\varphi = e^{i\varphi}$, то $z^n = \cos(n\varphi) + i\sin(n\varphi) = e^{in\varphi}$.

Приложение В

Статистические таблицы

Нормальное распределение

$\varepsilon = 1.96$	**0.06**
1.9	$\frac{1}{\sqrt{2\pi}} \int\limits_{1.96}^{+\infty} e^{-(1/2)x^2} dx = \mathbf{0.0250}$

Функция нормального распределения:

$$\frac{1}{\sqrt{2\pi}} \int\limits_{\varepsilon}^{+\infty} e^{-(1/2)x^2} dx.$$

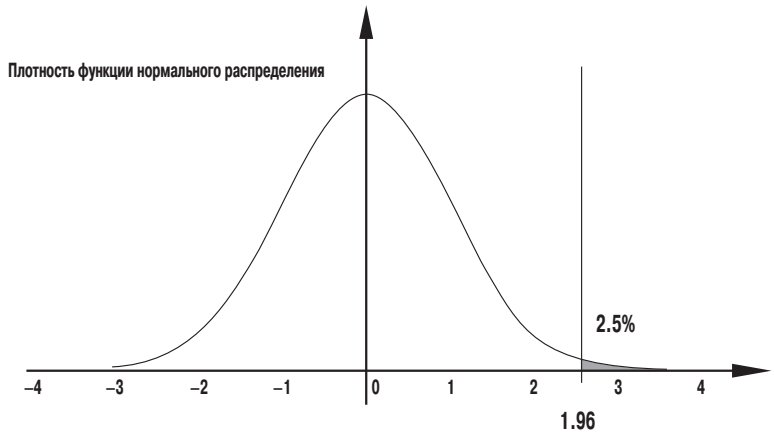

Рис. В.1График плотности нормального распределения с нулевым математическим ожиданием и единичной дисперсией $N(0, 1)$

Таблица А.1. Значения функции плотности стандартного нормального распределения

ε	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
3	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010

Распределение *t*-Стьюдента

$\Pr(t > t_{k,1-\theta}) = \theta$

Границы t-распределения с k степенями свободы $t_{k,1-\theta}$

	$\theta = \mathbf{0.025}$
$k = 5$	$2.5\%, t_{0.975} = \mathbf{2.571}$

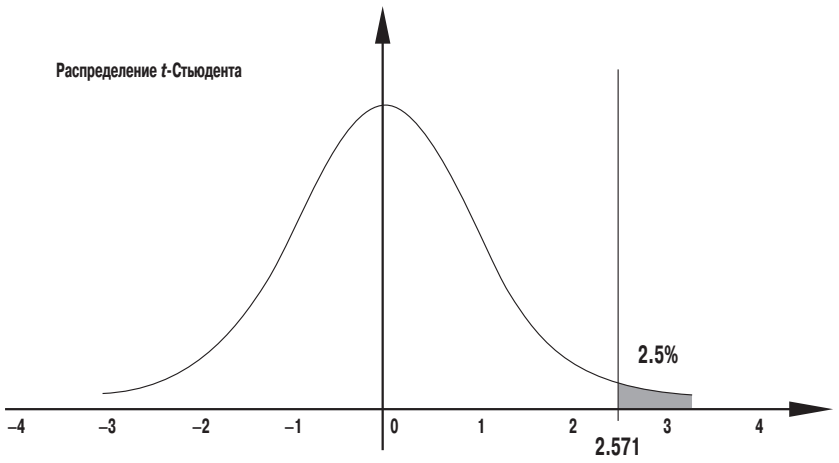

Рис. В.2График плотности распределения t-Стьюдента для $k = 5$

Таблица А.2. Граница t-распределения с k степенями свободы

	Односторонние квантили					
k \ θ	0.15	0.1	0.05	0.025	0.01	0.005
1	1.963	3.078	6.314	12.706	31.821	63.656
2	1.386	1.886	2.920	4.303	6.965	9.925
3	1.250	1.638	2.353	3.182	4.541	5.841
4	1.190	1.533	2.132	2.776	3.747	4.604
5	1.156	1.476	2.015	2.571	3.365	4.032
6	1.134	1.440	1.943	2.447	3.143	3.707
7	1.119	1.415	1.895	2.365	2.998	3.499
8	1.108	1.397	1.860	2.306	2.896	3.355
9	1.100	1.383	1.833	2.262	2.821	3.250
10	1.093	1.372	1.812	2.228	2.764	3.169
11	1.088	1.363	1.796	2.201	2.718	3.106
12	1.083	1.356	1.782	2.179	2.681	3.055
13	1.079	1.350	1.771	2.160	2.650	3.012
14	1.076	1.345	1.761	2.145	2.624	2.977
15	1.074	1.341	1.753	2.131	2.602	2.947
16	1.071	1.337	1.746	2.120	2.583	2.921
17	1.069	1.333	1.740	2.110	2.567	2.898
18	1.067	1.330	1.734	2.101	2.552	2.878
19	1.066	1.328	1.729	2.093	2.539	2.861
20	1.064	1.325	1.725	2.086	2.528	2.845
21	1.063	1.323	1.721	2.080	2.518	2.831
22	1.061	1.321	1.717	2.074	2.508	2.819
23	1.060	1.319	1.714	2.069	2.500	2.807
24	1.059	1.318	1.711	2.064	2.492	2.797
25	1.058	1.316	1.708	2.060	2.485	2.787
26	1.058	1.315	1.706	2.056	2.479	2.779
27	1.057	1.314	1.703	2.052	2.473	2.771
28	1.056	1.313	1.701	2.048	2.467	2.763

	Односторонние квантили					
k \ θ	0.15	0.1	0.05	0.025	0.01	0.005
29	1.055	1.311	1.699	2.045	2.462	2.756
30	1.055	1.310	1.697	2.042	2.457	2.750
35	1.052	1.306	1.690	2.030	2.438	2.724
40	1.050	1.303	1.684	2.021	2.423	2.704
45	1.049	1.301	1.679	2.014	2.412	2.690
50	1.047	1.299	1.676	2.009	2.403	2.678
55	1.046	1.297	1.673	2.004	2.396	2.668
60	1.045	1.296	1.671	2.000	2.390	2.660
65	1.045	1.295	1.669	1.997	2.385	2.654
70	1.044	1.294	1.667	1.994	2.381	2.648
75	1.044	1.293	1.665	1.992	2.377	2.643
80	1.043	1.292	1.664	1.990	2.374	2.639
85	1.043	1.292	1.663	1.988	2.371	2.635
90	1.042	1.291	1.662	1.987	2.368	2.632
95	1.042	1.291	1.661	1.985	2.366	2.629
100	1.042	1.290	1.660	1.984	2.364	2.626
110	1.041	1.289	1.659	1.982	2.361	2.621
120	1.041	1.289	1.658	1.980	2.358	2.617
130	1.041	1.288	1.657	1.978	2.355	2.614
140	1.040	1.288	1.656	1.977	2.353	2.611
150	1.040	1.287	1.655	1.976	2.351	2.609
160	1.040	1.287	1.654	1.975	2.350	2.607
170	1.040	1.287	1.654	1.974	2.348	2.605
180	1.039	1.286	1.653	1.973	2.347	2.603
190	1.039	1.286	1.653	1.973	2.346	2.602
200	1.039	1.286	1.653	1.972	2.345	2.601
∞	1.036	1.282	1.645	1.960	2.326	2.576

Распределение хи-квадрат (χ^2)

$\Pr(\chi^2 > \chi^2_{k,1-\theta}) = \theta.$

Границы χ^2-распределения с k степенями свободы $\chi^2_{k,1-\theta}$.

	$\theta = \mathbf{0.05}$
$k = 5$	5% , $\chi_{0.95} = \mathbf{11.07}$

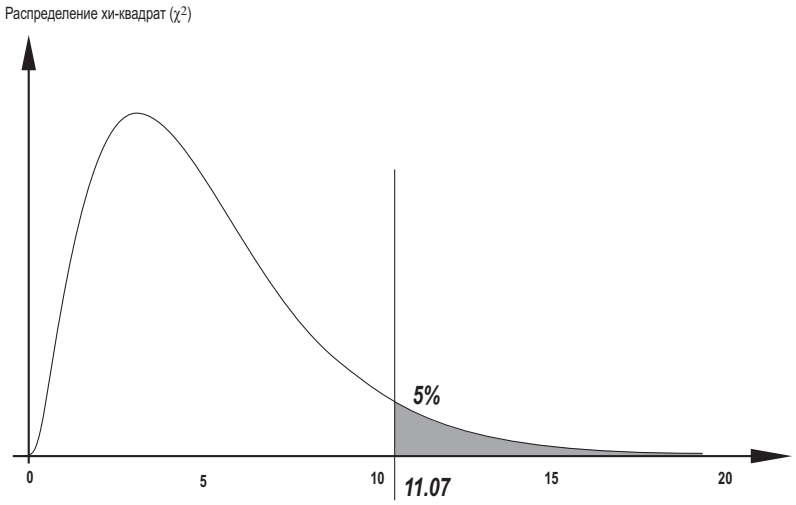

Рис. В.3График плотности распределения χ^2 для k =5

Таблица А.3. Границы χ^2-распределения с k степенями свободы

k \ θ	0.9	0.75	0.5	0.25	0.15	0.1	0.05	0.025	0.01	0.005
1	0.016	0.102	0.455	1.323	2.072	2.706	3.841	5.024	6.635	7.879
2	0.211	0.575	1.386	2.773	3.794	4.605	5.991	7.378	9.210	10.597
3	0.584	1.213	2.366	4.108	5.317	6.251	7.815	9.348	11.345	12.838
4	1.064	1.923	3.357	5.385	6.745	7.779	9.488	11.143	13.277	14.860
5	1.610	2.675	4.351	6.626	8.115	9.236	11.070	12.832	15.086	16.750
6	2.204	3.455	5.348	7.841	9.446	10.645	12.592	14.449	16.812	18.548
7	2.833	4.255	6.346	9.037	10.748	12.017	14.067	16.013	18.475	20.278
8	3.490	5.071	7.344	10.219	12.027	13.362	15.507	17.535	20.090	21.955
9	4.168	5.899	8.343	11.389	13.288	14.684	16.919	19.023	21.666	23.589
10	4.865	6.737	9.342	12.549	14.534	15.987	18.307	20.483	23.209	25.188
11	5.578	7.584	10.341	13.701	15.767	17.275	19.675	21.920	24.725	26.757
12	6.304	8.438	11.340	14.845	16.989	18.549	21.026	23.337	26.217	28.300
13	7.041	9.299	12.340	15.984	18.202	19.812	22.362	24.736	27.688	29.819
14	7.790	10.165	13.339	17.117	19.406	21.064	23.685	26.119	29.141	31.319
15	8.547	11.037	14.339	18.245	20.603	22.307	24.996	27.488	30.578	32.801
16	9.312	11.912	15.338	19.369	21.793	23.542	26.296	28.845	32.000	34.267
17	10.085	12.792	16.338	20.489	22.977	24.769	27.587	30.191	33.409	35.718
18	10.865	13.675	17.338	21.605	24.155	25.989	28.869	31.526	34.805	37.156
19	11.651	14.562	18.338	22.718	25.329	27.204	30.144	32.852	36.191	38.582
20	12.443	15.452	19.337	23.828	26.498	28.412	31.410	34.170	37.566	39.997
21	13.240	16.344	20.337	24.935	27.662	29.615	32.671	35.479	38.932	41.401
22	14.041	17.240	21.337	26.039	28.822	30.813	33.924	36.781	40.289	42.796
23	14.848	18.137	22.337	27.141	29.979	32.007	35.172	38.076	41.638	44.181
24	15.659	19.037	23.337	28.241	31.132	33.196	36.415	39.364	42.980	45.558
25	16.473	19.939	24.337	29.339	32.282	34.382	37.652	40.646	44.314	46.928
26	17.292	20.843	25.336	30.435	33.429	35.563	38.885	41.923	45.642	48.290
27	18.114	21.749	26.336	31.528	34.574	36.741	40.113	43.195	46.963	49.645
28	18.939	22.657	27.336	32.620	35.715	37.916	41.337	44.461	48.278	50.994
29	19.768	23.567	28.336	33.711	36.854	39.087	42.557	45.722	49.588	52.335
30	20.599	24.478	29.336	34.800	37.990	40.256	43.773	46.979	50.892	53.672
35	24.797	29.054	34.336	40.223	43.640	46.059	49.802	53.203	57.342	60.275
40	29.051	33.660	39.335	45.616	49.244	51.805	55.758	59.342	63.691	66.766
45	33.350	38.291	44.335	50.985	54.810	57.505	61.656	65.410	69.957	73.166
50	37.689	42.942	49.335	56.334	60.346	63.167	67.505	71.420	76.154	79.490
55	42.060	47.610	54.335	61.665	65.855	68.796	73.311	77.380	82.292	85.749
60	46.459	52.294	59.335	66.981	71.341	74.397	79.082	83.298	88.379	91.952
65	50.883	56.990	64.335	72.285	76.807	79.973	84.821	89.177	94.422	98.105
70	55.329	61.698	69.334	77.577	82.255	85.527	90.531	95.023	100.425	104.215
75	59.795	66.417	74.334	82.858	87.688	91.061	96.217	100.839	106.393	110.285
80	64.278	71.145	79.334	88.130	93.106	96.578	101.879	106.629	112.329	116.321
85	68.777	75.881	84.334	93.394	98.511	102.079	107.522	112.393	118.236	122.324
90	73.291	80.625	89.334	98.650	103.904	107.565	113.145	118.136	124.116	128.299

Таблица А.3. Границы χ^2-распределения с k степенями свободы (продолжение)

θ / k	0.9	0.75	0.5	0.25	0.15	0.1	0.05	0.025	0.01	0.005
95	77.818	85.376	94.334	103.899	109.286	113.038	118.752	123.858	129.973	134.247
100	82.358	90.133	99.334	109.141	114.659	118.498	124.342	129.561	135.807	140.170
110	91.471	99.666	109.334	119.608	125.376	129.385	135.480	140.916	147.414	151.948
120	100.624	109.220	119.334	130.055	136.062	140.233	146.567	152.211	158.950	163.648
130	109.811	118.792	129.334	140.482	146.719	151.045	157.610	163.453	170.423	175.278
140	119.029	128.380	139.334	150.894	157.352	161.827	168.613	174.648	181.841	186.847
150	128.275	137.983	149.334	161.291	167.962	172.581	179.581	185.800	193.207	198.360
160	137.546	147.599	159.334	171.675	178.552	183.311	190.516	196.915	204.530	209.824
170	146.839	157.227	169.334	182.047	189.123	194.017	201.423	207.995	215.812	221.242
180	156.153	166.865	179.334	192.409	199.679	204.704	212.304	219.044	227.056	232.620
190	165.485	176.514	189.334	202.760	210.218	215.371	223.160	230.064	238.266	243.959
200	174.835	186.172	199.334	213.102	220.744	226.021	233.994	241.058	249.445	255.264

Распределение *F*-Фишера

k_2 \ k_1	$k_1 = 5$
$k_2 = 10$	$5\%, F_{0.95} = \mathbf{3.33}$ $1\%, F_{0.99} = \mathbf{5.64}$

$\Pr(F > F_{k1,k2,0.95}) = \mathbf{0.05}$,
$\Pr(F > F_{k1,k2,0.99}) = \mathbf{0.01}$.

Границы F-распределения с k_1, k_2 степенями свободы для 5% и 1% вероятности $F_{k_1,\,k_2,\,1-\theta}$

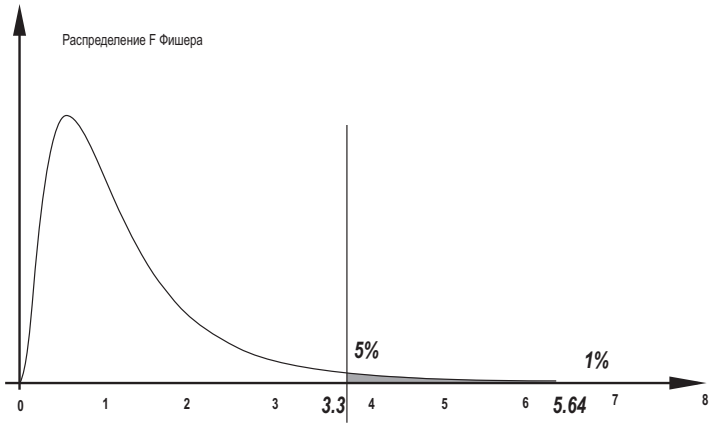

Рис. В.4График плотности распределения для $k_1 = 5,\ k_2 = 10$

Таблица А.4. Границы F-распределения с k_1 и k_2 степенями свободы для 5% и 1% вероятности

k_2 \ k_1	1	2	3	4	5	6	7	8	9	10	11	12
1	161.4 4052	199.5 4999	215.7 5404	224.6 5624	230.2 5764	234.0 5859	236.8 5928	238.9 5981	240.5 6022	241.9 6056	243.0 6083	243.9 6107
2	18.51 98.50	19.00 99.00	19.16 99.16	19.25 99.25	19.30 99.30	19.33 99.33	19.35 99.36	19.37 99.38	19.38 99.39	19.40 99.40	19.40 99.41	19.41 99.42
3	10.13 34.12	9.55 30.82	9.28 29.46	9.12 28.71	9.01 28.24	8.94 27.91	8.89 27.67	8.85 27.49	8.81 27.34	8.79 27.23	8.76 27.13	8.74 27.05
4	7.71 21.20	6.94 18.00	6.59 16.69	6.39 15.98	6.26 15.52	6.16 15.21	6.09 14.98	6.04 14.80	6.00 14.66	5.96 14.55	5.94 14.45	5.91 14.37
5	6.61 16.26	5.79 13.27	5.41 12.06	5.19 11.39	5.05 10.97	4.95 10.67	4.88 10.46	4.82 10.29	4.77 10.16	4.74 10.05	4.70 9.96	4.68 9.89
6	5.99 13.75	5.14 10.92	4.76 9.78	4.53 9.15	4.39 8.75	4.28 8.47	4.21 8.26	4.15 8.10	4.10 7.98	4.06 7.87	4.03 7.79	4.00 7.72
7	5.59 12.25	4.74 9.55	4.35 8.45	4.12 7.85	3.97 7.46	3.87 7.19	3.79 6.99	3.73 6.84	3.68 6.72	3.64 6.62	3.60 6.54	3.57 6.47
8	5.32 11.26	4.46 8.65	4.07 7.59	3.84 7.01	3.69 6.63	3.58 6.37	3.50 6.18	3.44 6.03	3.39 5.91	3.35 5.81	3.31 5.73	3.28 5.67
9	5.12 10.56	4.26 8.02	3.86 6.99	3.63 6.42	3.48 6.06	3.37 5.80	3.29 5.61	3.23 5.47	3.18 5.35	3.14 5.26	3.10 5.18	3.07 5.11
10	4.96 10.04	4.10 7.56	3.71 6.55	3.48 5.99	3.33 5.64	3.22 5.39	3.14 5.20	3.07 5.06	3.02 4.94	2.98 4.85	2.94 4.77	2.91 4.71
11	4.84 9.65	3.98 7.21	3.59 6.22	3.36 5.67	3.20 5.32	3.09 5.07	3.01 4.89	2.95 4.74	2.90 4.63	2.85 4.54	2.82 4.46	2.79 4.40
12	4.75 9.33	3.89 6.93	3.49 5.95	3.26 5.41	3.11 5.06	3.00 4.82	2.91 4.64	2.85 4.50	2.80 4.39	2.75 4.30	2.72 4.22	2.69 4.16
13	4.67 9.07	3.81 6.70	3.41 5.74	3.18 5.21	3.03 4.86	2.92 4.62	2.83 4.44	2.77 4.30	2.71 4.19	2.67 4.10	2.63 4.02	2.60 3.96
14	4.60 8.86	3.74 6.51	3.34 5.56	3.11 5.04	2.96 4.69	2.85 4.46	2.76 4.28	2.70 4.14	2.65 4.03	2.60 3.94	2.57 3.86	2.53 3.80
15	4.54 8.68	3.68 6.36	3.29 5.42	3.06 4.89	2.90 4.56	2.79 4.32	2.71 4.14	2.64 4.00	2.59 3.89	2.54 3.80	2.51 3.73	2.48 3.67
16	4.49 8.53	3.63 6.23	3.24 5.29	3.01 4.77	2.85 4.44	2.74 4.20	2.66 4.03	2.59 3.89	2.54 3.78	2.49 3.69	2.46 3.62	2.42 3.55
17	4.45 8.40	3.59 6.11	3.20 5.19	2.96 4.67	2.81 4.34	2.70 4.10	2.61 3.93	2.55 3.79	2.49 3.68	2.45 3.59	2.41 3.52	2.38 3.46
18	4.41 8.29	3.55 6.01	3.16 5.09	2.93 4.58	2.77 4.25	2.66 4.01	2.58 3.84	2.51 3.71	2.46 3.60	2.41 3.51	2.37 3.43	2.34 3.37
19	4.38 8.18	3.52 5.93	3.13 5.01	2.90 4.50	2.74 4.17	2.63 3.94	2.54 3.77	2.48 3.63	2.42 3.52	2.38 3.43	2.34 3.36	2.31 3.30
20	4.35 8.10	3.49 5.85	3.10 4.94	2.87 4.43	2.71 4.10	2.60 3.87	2.51 3.70	2.45 3.56	2.39 3.46	2.35 3.37	2.31 3.29	2.28 3.23
21	4.32 8.02	3.47 5.78	3.07 4.87	2.84 4.37	2.68 4.04	2.57 3.81	2.49 3.64	2.42 3.51	2.37 3.40	2.32 3.31	2.28 3.24	2.25 3.17
22	4.30 7.95	3.44 5.72	3.05 4.82	2.82 4.31	2.66 3.99	2.55 3.76	2.46 3.59	2.40 3.45	2.34 3.35	2.30 3.26	2.26 3.18	2.23 3.12
23	4.28 7.88	3.42 5.66	3.03 4.76	2.80 4.26	2.64 3.94	2.53 3.71	2.44 3.54	2.37 3.41	2.32 3.30	2.27 3.21	2.24 3.14	2.20 3.07
24	4.26 7.82	3.40 5.61	3.01 4.72	2.78 4.22	2.62 3.90	2.51 3.67	2.42 3.50	2.36 3.36	2.30 3.26	2.25 3.17	2.22 3.09	2.18 3.03
25	4.24 7.77	3.39 5.57	2.99 4.68	2.76 4.18	2.60 3.85	2.49 3.63	2.40 3.46	2.34 3.32	2.28 3.22	2.24 3.13	2.20 3.06	2.16 2.99
26	4.23 7.72	3.37 5.53	2.98 4.64	2.74 4.14	2.59 3.82	2.47 3.59	2.39 3.42	2.32 3.29	2.27 3.18	2.22 3.09	2.18 3.02	2.15 2.96

Таблица А.4. Границы F-распределения с k_1 и k_2 степенями свободы для 5% и 1% вероятности (продолжение)

k_2 \ k_1	1	2	3	4	5	6	7	8	9	10	11	12
27	4.21 7.68	3.35 5.49	2.96 4.60	2.73 4.11	2.57 3.78	2.46 3.56	2.37 3.39	2.31 3.26	2.25 3.15	2.20 3.06	2.17 2.99	2.13 2.93
28	4.20 7.64	3.34 5.45	2.95 4.57	2.71 4.07	2.56 3.75	2.45 3.53	2.36 3.36	2.29 3.23	2.24 3.12	2.19 3.03	2.15 2.96	2.12 2.90
29	4.18 7.60	3.33 5.42	2.93 4.54	2.70 4.04	2.55 3.73	2.43 3.50	2.35 3.33	2.28 3.20	2.22 3.09	2.18 3.00	2.14 2.93	2.10 2.87
30	4.17 7.56	3.32 5.39	2.92 4.51	2.69 4.02	2.53 3.70	2.42 3.47	2.33 3.30	2.27 3.17	2.21 3.07	2.16 2.98	2.13 2.91	2.09 2.84
32	4.15 7.50	3.29 5.34	2.90 4.46	2.67 3.97	2.51 3.65	2.40 3.43	2.31 3.26	2.24 3.13	2.19 3.02	2.14 2.93	2.10 2.86	2.07 2.80
34	4.13 7.44	3.28 5.29	2.88 4.42	2.65 3.93	2.49 3.61	2.38 3.39	2.29 3.22	2.23 3.09	2.17 2.98	2.12 2.89	2.08 2.82	2.05 2.76
36	4.11 7.40	3.26 5.25	2.87 4.38	2.63 3.89	2.48 3.57	2.36 3.35	2.28 3.18	2.21 3.05	2.15 2.95	2.11 2.86	2.07 2.79	2.03 2.72
38	4.10 7.35	3.24 5.21	2.85 4.34	2.62 3.86	2.46 3.54	2.35 3.32	2.26 3.15	2.19 3.02	2.14 2.92	2.09 2.83	2.05 2.75	2.02 2.69
40	4.08 7.31	3.23 5.18	2.84 4.31	2.61 3.83	2.45 3.51	2.34 3.29	2.25 3.12	2.18 2.99	2.12 2.89	2.08 2.80	2.04 2.73	2.00 2.66
42	4.07 7.28	3.22 5.15	2.83 4.29	2.59 3.80	2.44 3.49	2.32 3.27	2.24 3.10	2.17 2.97	2.11 2.86	2.06 2.78	2.03 2.70	1.99 2.64
44	4.06 7.25	3.21 5.12	2.82 4.26	2.58 3.78	2.43 3.47	2.31 3.24	2.23 3.08	2.16 2.95	2.10 2.84	2.05 2.75	2.01 2.68	1.98 2.62
46	4.05 7.22	3.20 5.10	2.81 4.24	2.57 3.76	2.42 3.44	2.30 3.22	2.22 3.06	2.15 2.93	2.09 2.82	2.04 2.73	2.00 2.66	1.97 2.60
48	4.04 7.19	3.19 5.08	2.80 4.22	2.57 3.74	2.41 3.43	2.29 3.20	2.21 3.04	2.14 2.91	2.08 2.80	2.03 2.71	1.99 2.64	1.96 2.58
50	4.03 7.17	3.18 5.06	2.79 4.20	2.56 3.72	2.40 3.41	2.29 3.19	2.20 3.02	2.13 2.89	2.07 2.78	2.03 2.70	1.99 2.63	1.95 2.56
55	4.02 7.12	3.16 5.01	2.77 4.16	2.54 3.68	2.38 3.37	2.27 3.15	2.18 2.98	2.11 2.85	2.06 2.75	2.01 2.66	1.97 2.59	1.93 2.53
60	4.00 7.08	3.15 4.98	2.76 4.13	2.53 3.65	2.37 3.34	2.25 3.12	2.17 2.95	2.10 2.82	2.04 2.72	1.99 2.63	1.95 2.56	1.92 2.50
65	3.99 7.04	3.14 4.95	2.75 4.10	2.51 3.62	2.36 3.31	2.24 3.09	2.15 2.93	2.08 2.80	2.03 2.69	1.98 2.61	1.94 2.53	1.90 2.47
70	3.98 7.01	3.13 4.92	2.74 4.07	2.50 3.60	2.35 3.29	2.23 3.07	2.14 2.91	2.07 2.78	2.02 2.67	1.97 2.59	1.93 2.51	1.89 2.45
80	3.96 6.96	3.11 4.88	2.72 4.04	2.49 3.56	2.33 3.26	2.21 3.04	2.13 2.87	2.06 2.74	2.00 2.64	1.95 2.55	1.91 2.48	1.88 2.42
90	3.95 6.93	3.10 4.85	2.71 4.01	2.47 3.53	2.32 3.23	2.20 3.01	2.11 2.84	2.04 2.72	1.99 2.61	1.94 2.52	1.90 2.45	1.86 2.39
100	3.94 6.90	3.09 4.82	2.70 3.98	2.46 3.51	2.31 3.21	2.19 2.99	2.10 2.82	2.03 2.69	1.97 2.59	1.93 2.50	1.89 2.43	1.85 2.37
125	3.92 6.84	3.07 4.78	2.68 3.94	2.44 3.47	2.29 3.17	2.17 2.95	2.08 2.79	2.01 2.66	1.96 2.55	1.91 2.47	1.87 2.39	1.83 2.33
150	3.90 6.81	3.06 4.75	2.66 3.91	2.43 3.45	2.27 3.14	2.16 2.92	2.07 2.76	2.00 2.63	1.94 2.53	1.89 2.44	1.85 2.37	1.82 2.31
200	3.89 6.76	3.04 4.71	2.65 3.88	2.42 3.41	2.26 3.11	2.14 2.89	2.06 2.73	1.98 2.60	1.93 2.50	1.88 2.41	1.84 2.34	1.80 2.27
400	3.86 6.70	3.02 4.66	2.63 3.83	2.39 3.37	2.24 3.06	2.12 2.85	2.03 2.68	1.96 2.56	1.90 2.45	1.85 2.37	1.81 2.29	1.78 2.23
1000	3.85 6.66	3.00 4.63	2.61 3.80	2.38 3.34	2.22 3.04	2.11 2.82	2.02 2.66	1.95 2.53	1.89 2.43	1.84 2.34	1.80 2.27	1.76 2.20

Таблица А.4. Границы F-распределения с k_1 и k_2 степенями свободы для 5% и 1% вероятности (продолжение)

k_2 \ k_1	14	16	18	20	24	30	40	50	75	100	200	500
1	245.4 6143	246.5 6170	247.3 6191	248.0 6209	249.1 6234	250.1 6260	251.1 6286	251.8 6302	252.6 6324	253.0 6334	253.7 6350	254.1 6360
2	19.42 99.43	19.43 99.44	19.44 99.44	19.45 99.45	19.45 99.46	19.46 99.47	19.47 99.48	19.48 99.48	19.48 99.48	19.49 99.49	19.49 99.49	19.49 99.50
3	8.71 26.92	8.69 26.83	8.67 26.75	8.66 26.69	8.64 26.60	8.62 26.50	8.59 26.41	8.58 26.35	8.56 26.28	8.55 26.24	8.54 26.18	8.53 26.15
4	5.87 14.25	5.84 14.15	5.82 14.08	5.80 14.02	5.77 13.93	5.75 13.84	5.72 13.75	5.70 13.69	5.68 13.61	5.66 13.58	5.65 13.52	5.64 13.49
5	4.64 9.77	4.60 9.68	4.58 9.61	4.56 9.55	4.53 9.47	4.50 9.38	4.46 9.29	4.44 9.24	4.42 9.17	4.41 9.13	4.39 9.08	4.37 9.04
6	3.96 7.60	3.92 7.52	3.90 7.45	3.87 7.40	3.84 7.31	3.81 7.23	3.77 7.14	3.75 7.09	3.73 7.02	3.71 6.99	3.69 6.93	3.68 6.90
7	3.53 6.36	3.49 6.28	3.47 6.21	3.44 6.16	3.41 6.07	3.38 5.99	3.34 5.91	3.32 5.86	3.29 5.79	3.27 5.75	3.25 5.70	3.24 5.67
8	3.24 5.56	3.20 5.48	3.17 5.41	3.15 5.36	3.12 5.28	3.08 5.20	3.04 5.12	3.02 5.07	2.99 5.00	2.97 4.96	2.95 4.91	2.94 4.88
9	3.03 5.01	2.99 4.92	2.96 4.86	2.94 4.81	2.90 4.73	2.86 4.65	2.83 4.57	2.80 4.52	2.77 4.45	2.76 4.41	2.73 4.36	2.72 4.33
10	2.86 4.60	2.83 4.52	2.80 4.46	2.77 4.41	2.74 4.33	2.70 4.25	2.66 4.17	2.64 4.12	2.60 4.05	2.59 4.01	2.56 3.96	2.55 3.93
11	2.74 4.29	2.70 4.21	2.67 4.15	2.65 4.10	2.61 4.02	2.57 3.94	2.53 3.86	2.51 3.81	2.47 3.74	2.46 3.71	2.43 3.66	2.42 3.62
12	2.64 4.05	2.60 3.97	2.57 3.91	2.54 3.86	2.51 3.78	2.47 3.70	2.43 3.62	2.40 3.57	2.37 3.50	2.35 3.47	2.32 3.41	2.31 3.38
13	2.55 3.86	2.51 3.78	2.48 3.72	2.46 3.66	2.42 3.59	2.38 3.51	2.34 3.43	2.31 3.38	2.28 3.31	2.26 3.27	2.23 3.22	2.22 3.19
14	2.48 3.70	2.44 3.62	2.41 3.56	2.39 3.51	2.35 3.43	2.31 3.35	2.27 3.27	2.24 3.22	2.21 3.15	2.19 3.11	2.16 3.06	2.14 3.03
15	2.42 3.56	2.38 3.49	2.35 3.42	2.33 3.37	2.29 3.29	2.25 3.21	2.20 3.13	2.18 3.08	2.14 3.01	2.12 2.98	2.10 2.92	2.08 2.89
16	2.37 3.45	2.33 3.37	2.30 3.31	2.28 3.26	2.24 3.18	2.19 3.10	2.15 3.02	2.12 2.97	2.09 2.90	2.07 2.86	2.04 2.81	2.02 2.78
17	2.33 3.35	2.29 3.27	2.26 3.21	2.23 3.16	2.19 3.08	2.15 3.00	2.10 2.92	2.08 2.87	2.04 2.80	2.02 2.76	1.99 2.71	1.97 2.68
18	2.29 3.27	2.25 3.19	2.22 3.13	2.19 3.08	2.15 3.00	2.11 2.92	2.06 2.84	2.04 2.78	2.00 2.71	1.98 2.68	1.95 2.62	1.93 2.59
19	2.26 3.19	2.21 3.12	2.18 3.05	2.16 3.00	2.11 2.92	2.07 2.84	2.03 2.76	2.00 2.71	1.96 2.64	1.94 2.60	1.91 2.55	1.89 2.51
20	2.22 3.13	2.18 3.05	2.15 2.99	2.12 2.94	2.08 2.86	2.04 2.78	1.99 2.69	1.97 2.64	1.93 2.57	1.91 2.54	1.88 2.48	1.86 2.44
21	2.20 3.07	2.16 2.99	2.12 2.93	2.10 2.88	2.05 2.80	2.01 2.72	1.96 2.64	1.94 2.58	1.90 2.51	1.88 2.48	1.84 2.42	1.83 2.38
22	2.17 3.02	2.13 2.94	2.10 2.88	2.07 2.83	2.03 2.75	1.98 2.67	1.94 2.58	1.91 2.53	1.87 2.46	1.85 2.42	1.82 2.36	1.80 2.33
23	2.15 2.97	2.11 2.89	2.08 2.83	2.05 2.78	2.01 2.70	1.96 2.62	1.91 2.54	1.88 2.48	1.84 2.41	1.82 2.37	1.79 2.32	1.77 2.28
24	2.13 2.93	2.09 2.85	2.05 2.79	2.03 2.74	1.98 2.66	1.94 2.58	1.89 2.49	1.86 2.44	1.82 2.37	1.80 2.33	1.77 2.27	1.75 2.24
25	2.11 2.89	2.07 2.81	2.04 2.75	2.01 2.70	1.96 2.62	1.92 2.54	1.87 2.45	1.84 2.40	1.80 2.33	1.78 2.29	1.75 2.23	1.73 2.19
26	2.09 2.86	2.05 2.78	2.02 2.72	1.99 2.66	1.95 2.58	1.90 2.50	1.85 2.42	1.82 2.36	1.78 2.29	1.76 2.25	1.73 2.19	1.71 2.16

Таблица А.4. Границы F-распределения с k_1 и k_2 степенями свободы для 5% и 1% вероятности (продолжение)

k_2 \ k_1	14	16	18	20	24	30	40	50	75	100	200	500
27	2.08 2.82	2.04 2.75	2.00 2.68	1.97 2.63	1.93 2.55	1.88 2.47	1.84 2.38	1.81 2.33	1.76 2.26	1.74 2.22	1.71 2.16	1.69 2.12
28	2.06 2.79	2.02 2.72	1.99 2.65	1.96 2.60	1.91 2.52	1.87 2.44	1.82 2.35	1.79 2.30	1.75 2.23	1.73 2.19	1.69 2.13	1.67 2.09
29	2.05 2.77	2.01 2.69	1.97 2.63	1.94 2.57	1.90 2.49	1.85 2.41	1.81 2.33	1.77 2.27	1.73 2.20	1.71 2.16	1.67 2.10	1.65 2.06
30	2.04 2.74	1.99 2.66	1.96 2.60	1.93 2.55	1.89 2.47	1.84 2.39	1.79 2.30	1.76 2.25	1.72 2.17	1.70 2.13	1.66 2.07	1.64 2.03
32	2.01 2.70	1.97 2.62	1.94 2.55	1.91 2.50	1.86 2.42	1.82 2.34	1.77 2.25	1.74 2.20	1.69 2.12	1.67 2.08	1.63 2.02	1.61 1.98
34	1.99 2.66	1.95 2.58	1.92 2.51	1.89 2.46	1.84 2.38	1.80 2.30	1.75 2.21	1.71 2.16	1.67 2.08	1.65 2.04	1.61 1.98	1.59 1.94
36	1.98 2.62	1.93 2.54	1.90 2.48	1.87 2.43	1.82 2.35	1.78 2.26	1.73 2.18	1.69 2.12	1.65 2.04	1.62 2.00	1.59 1.94	1.56 1.90
38	1.96 2.59	1.92 2.51	1.88 2.45	1.85 2.40	1.81 2.32	1.76 2.23	1.71 2.14	1.68 2.09	1.63 2.01	1.61 1.97	1.57 1.90	1.54 1.86
40	1.95 2.56	1.90 2.48	1.87 2.42	1.84 2.37	1.79 2.29	1.74 2.20	1.69 2.11	1.66 2.06	1.61 1.98	1.59 1.94	1.55 1.87	1.53 1.83
42	1.94 2.54	1.89 2.46	1.86 2.40	1.83 2.34	1.78 2.26	1.73 2.18	1.68 2.09	1.65 2.03	1.60 1.95	1.57 1.91	1.53 1.85	1.51 1.80
44	1.92 2.52	1.88 2.44	1.84 2.37	1.81 2.32	1.77 2.24	1.72 2.15	1.67 2.07	1.63 2.01	1.59 1.93	1.56 1.89	1.52 1.82	1.49 1.78
46	1.91 2.50	1.87 2.42	1.83 2.35	1.80 2.30	1.76 2.22	1.71 2.13	1.65 2.04	1.62 1.99	1.57 1.91	1.55 1.86	1.51 1.80	1.48 1.76
48	1.90 2.48	1.86 2.40	1.82 2.33	1.79 2.28	1.75 2.20	1.70 2.12	1.64 2.02	1.61 1.97	1.56 1.89	1.54 1.84	1.49 1.78	1.47 1.73
50	1.89 2.46	1.85 2.38	1.81 2.32	1.78 2.27	1.74 2.18	1.69 2.10	1.63 2.01	1.60 1.95	1.55 1.87	1.52 1.82	1.48 1.76	1.46 1.71
55	1.88 2.42	1.83 2.34	1.79 2.28	1.76 2.23	1.72 2.15	1.67 2.06	1.61 1.97	1.58 1.91	1.53 1.83	1.50 1.78	1.46 1.71	1.43 1.67
60	1.86 2.39	1.82 2.31	1.78 2.25	1.75 2.20	1.70 2.12	1.65 2.03	1.59 1.94	1.56 1.88	1.51 1.79	1.48 1.75	1.44 1.68	1.41 1.63
65	1.85 2.37	1.80 2.29	1.76 2.23	1.73 2.17	1.69 2.09	1.63 2.00	1.58 1.91	1.54 1.85	1.49 1.77	1.46 1.72	1.42 1.65	1.39 1.60
70	1.84 2.35	1.79 2.27	1.75 2.20	1.72 2.15	1.67 2.07	1.62 1.98	1.57 1.89	1.53 1.83	1.48 1.74	1.45 1.70	1.40 1.62	1.37 1.57
80	1.82 2.31	1.77 2.23	1.73 2.17	1.70 2.12	1.65 2.03	1.60 1.94	1.54 1.85	1.51 1.79	1.45 1.70	1.43 1.65	1.38 1.58	1.35 1.53
90	1.80 2.29	1.76 2.21	1.72 2.14	1.69 2.09	1.64 2.00	1.59 1.92	1.53 1.82	1.49 1.76	1.44 1.67	1.41 1.62	1.36 1.55	1.33 1.49
100	1.79 2.27	1.75 2.19	1.71 2.12	1.68 2.07	1.63 1.98	1.57 1.89	1.52 1.80	1.48 1.74	1.42 1.65	1.39 1.60	1.34 1.52	1.31 1.47
125	1.77 2.23	1.73 2.15	1.69 2.08	1.66 2.03	1.60 1.94	1.55 1.85	1.49 1.76	1.45 1.69	1.40 1.60	1.36 1.55	1.31 1.47	1.27 1.41
150	1.76 2.20	1.71 2.12	1.67 2.06	1.64 2.00	1.59 1.92	1.54 1.83	1.48 1.73	1.44 1.66	1.38 1.57	1.34 1.52	1.29 1.43	1.25 1.38
200	1.74 2.17	1.69 2.09	1.66 2.03	1.62 1.97	1.57 1.89	1.52 1.79	1.46 1.69	1.41 1.63	1.35 1.53	1.32 1.48	1.26 1.39	1.22 1.33
400	1.72 2.13	1.67 2.05	1.63 1.98	1.60 1.92	1.54 1.84	1.49 1.75	1.42 1.64	1.38 1.58	1.32 1.48	1.28 1.42	1.22 1.32	1.17 1.25
1000	1.70 2.10	1.65 2.02	1.61 1.95	1.58 1.90	1.53 1.81	1.47 1.72	1.41 1.61	1.36 1.54	1.30 1.44	1.26 1.38	1.19 1.28	1.13 1.19

Критерий Дарбина—Уотсона

Значащие точки d_L и d_U, для 5% уровня значимости.

N — количество наблюдений, n — количество объясняющих переменных (без учета постоянного члена).

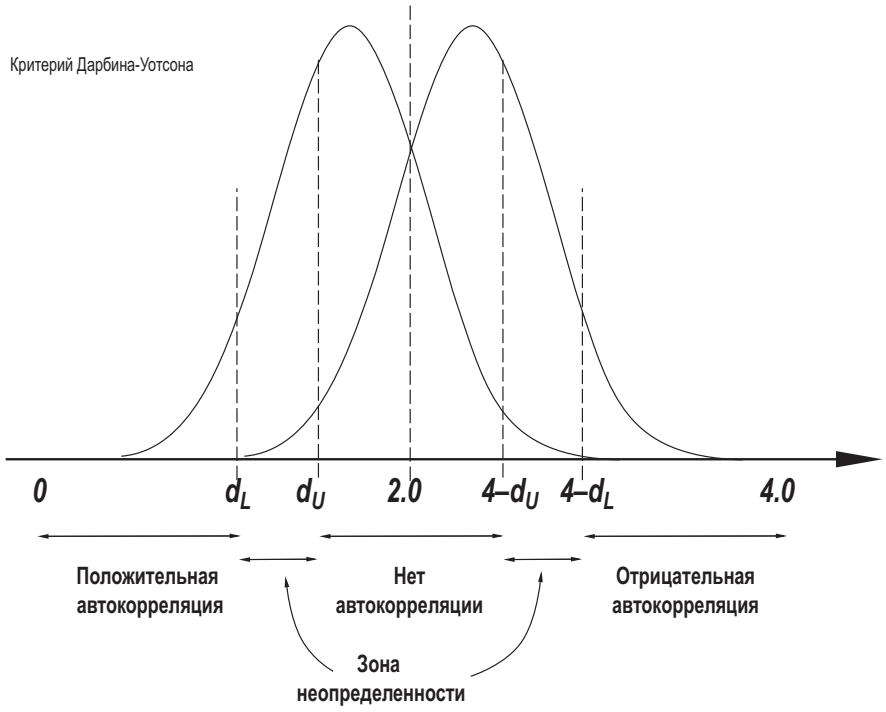

Рис. В.5

Таблица. А.5 Значения статистики d_L и d_U критерия Дарбина—Уотсона

N	n = 1		n = 2		n = 3		n = 4		n = 5		n = 6		n = 7		n = 8		n = 9		n = 10	
	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL
6	0.610	1.400																		
7	0.700	1.356	0.467	1.896																
8	0.763	1.332	0.559	1.777	0.368	2.287														
9	0.824	1.320	0.629	1.699	0.455	2.128	0.296	2.588												
10	0.879	1.320	0.697	1.641	0.525	2.016	0.376	2.414	0.243	2.822										
11	0.927	1.324	0.758	1.604	0.595	1.928	0.444	2.283	0.316	2.645	0.203	3.005								
12	0.971	1.331	0.812	1.579	0.658	1.864	0.512	2.177	0.379	2.506	0.268	2.832	0.171	3.149						
13	1.010	1.340	0.861	1.562	0.715	1.816	0.574	2.094	0.445	2.390	0.328	2.692	0.230	2.985	0.147	3.266				
14	1.045	1.350	0.905	1.551	0.767	1.779	0.632	2.030	0.505	2.296	0.389	2.572	0.286	2.848	0.200	3.111	0.127	3.360		
15	1.077	1.361	0.946	1.543	0.814	1.750	0.685	1.977	0.562	2.220	0.447	2.472	0.343	2.727	0.251	2.979	0.175	3.216	0.111	3.438
16	1.106	1.371	0.982	1.539	0.857	1.728	0.734	1.935	0.615	2.157	0.502	2.388	0.398	2.624	0.304	2.860	0.222	3.090	0.155	3.304
17	1.133	1.381	1.015	1.536	0.897	1.710	0.779	1.900	0.664	2.104	0.554	2.318	0.451	2.537	0.356	2.757	0.272	2.975	0.198	3.184
18	1.158	1.391	1.046	1.535	0.933	1.696	0.820	1.872	0.710	2.060	0.603	2.257	0.502	2.461	0.407	2.667	0.321	2.873	0.244	3.073
19	1.180	1.401	1.074	1.536	0.967	1.685	0.859	1.848	0.752	2.023	0.649	2.206	0.549	2.396	0.456	2.589	0.369	2.783	0.290	2.974
20	1.201	1.411	1.100	1.537	0.998	1.676	0.894	1.828	0.792	1.991	0.692	2.162	0.595	2.339	0.502	2.521	0.416	2.704	0.336	2.885
21	1.221	1.420	1.125	1.538	1.026	1.669	0.927	1.812	0.829	1.964	0.732	2.124	0.637	2.290	0.547	2.460	0.461	2.633	0.380	2.806
22	1.239	1.429	1.147	1.541	1.053	1.664	0.958	1.797	0.863	1.940	0.769	2.090	0.677	2.246	0.588	2.407	0.504	2.571	0.424	2.734
23	1.257	1.437	1.168	1.543	1.078	1.660	0.986	1.785	0.895	1.920	0.804	2.061	0.715	2.208	0.628	2.360	0.545	2.514	0.465	2.670
24	1.273	1.446	1.188	1.546	1.101	1.656	1.013	1.775	0.925	1.902	0.837	2.035	0.751	2.174	0.666	2.318	0.584	2.464	0.506	2.613
25	1.288	1.454	1.206	1.550	1.123	1.654	1.038	1.767	0.953	1.886	0.868	2.012	0.784	2.144	0.702	2.280	0.621	2.419	0.544	2.560
26	1.302	1.461	1.224	1.553	1.143	1.652	1.062	1.759	0.979	1.873	0.897	1.992	0.816	2.117	0.735	2.246	0.657	2.379	0.581	2.513
27	1.316	1.469	1.240	1.556	1.162	1.651	1.084	1.753	1.004	1.861	0.925	1.974	0.845	2.093	0.767	2.216	0.691	2.342	0.616	2.470
28	1.328	1.476	1.255	1.560	1.181	1.650	1.104	1.747	1.028	1.850	0.951	1.958	0.874	2.071	0.798	2.188	0.723	2.309	0.650	2.431
29	1.341	1.483	1.270	1.563	1.198	1.650	1.124	1.743	1.050	1.841	0.975	1.944	0.900	2.052	0.826	2.164	0.753	2.278	0.682	2.396
30	1.352	1.489	1.284	1.567	1.214	1.650	1.143	1.739	1.071	1.833	0.998	1.931	0.926	2.034	0.854	2.141	0.782	2.251	0.712	2.363

Таблица. А.5 Значения статистики d_L и d_U критерия Дарбина—Уотсона (продолжение)

N	n = 1		n = 2		n = 3		n = 4		n = 5		n = 6		n = 7		n = 8		n = 9		n = 10	
	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL
31	1.363	1.496	1.297	1.570	1.229	1.650	1.160	1.735	1.090	1.825	1.020	1.920	0.950	2.018	0.879	2.120	0.810	2.226	0.741	2.333
32	1.373	1.502	1.309	1.574	1.244	1.650	1.177	1.732	1.109	1.819	1.041	1.909	0.972	2.004	0.904	2.102	0.836	2.203	0.769	2.306
33	1.383	1.508	1.321	1.577	1.258	1.651	1.193	1.730	1.127	1.813	1.061	1.900	0.994	1.991	0.927	2.085	0.861	2.181	0.795	2.281
34	1.393	1.514	1.333	1.580	1.271	1.652	1.208	1.728	1.144	1.808	1.080	1.891	1.015	1.979	0.950	2.069	0.885	2.162	0.821	2.257
35	1.402	1.519	1.343	1.584	1.283	1.653	1.222	1.726	1.160	1.803	1.097	1.884	1.034	1.967	0.971	2.054	0.908	2.144	0.845	2.236
36	1.411	1.525	1.354	1.587	1.295	1.654	1.236	1.724	1.175	1.799	1.114	1.877	1.053	1.957	0.991	2.041	0.930	2.127	0.868	2.216
37	1.419	1.530	1.364	1.590	1.307	1.655	1.249	1.723	1.190	1.795	1.131	1.870	1.071	1.948	1.011	2.029	0.951	2.112	0.891	2.198
38	1.427	1.535	1.373	1.594	1.318	1.656	1.261	1.722	1.204	1.792	1.146	1.864	1.088	1.939	1.029	2.017	0.970	2.098	0.912	2.180
39	1.435	1.540	1.382	1.597	1.328	1.658	1.273	1.722	1.218	1.789	1.161	1.859	1.104	1.932	1.047	2.007	0.990	2.085	0.932	2.164
40	1.442	1.544	1.391	1.600	1.338	1.659	1.285	1.721	1.230	1.786	1.175	1.854	1.120	1.924	1.064	1.997	1.008	2.072	0.945	2.149
45	1.475	1.566	1.430	1.615	1.383	1.666	1.336	1.720	1.287	1.776	1.238	1.835	1.189	1.895	1.139	1.958	1.089	2.022	1.038	2.088
50	1.503	1.585	1.462	1.628	1.421	1.674	1.378	1.721	1.335	1.771	1.291	1.822	1.246	1.875	1.201	1.930	1.156	1.986	1.110	2.044
55	1.528	1.601	1.490	1.641	1.452	1.681	1.414	1.724	1.374	1.768	1.334	1.814	1.294	1.861	1.253	1.909	1.212	1.959	1.170	2.010
60	1.549	1.616	1.514	1.652	1.480	1.689	1.444	1.727	1.408	1.767	1.372	1.808	1.335	1.850	1.298	1.894	1.260	1.939	1.222	1.984
65	1.567	1.629	1.536	1.662	1.503	1.696	1.471	1.731	1.438	1.767	1.404	1.805	1.370	1.843	1.336	1.882	1.301	1.923	1.266	1.964
70	1.583	1.641	1.554	1.672	1.525	1.703	1.494	1.735	1.464	1.768	1.433	1.802	1.401	1.837	1.369	1.873	1.337	1.910	1.305	1.948
75	1.598	1.652	1.571	1.680	1.543	1.709	1.515	1.739	1.487	1.770	1.458	1.801	1.428	1.834	1.399	1.867	1.369	1.901	1.339	1.935
80	1.611	1.662	1.586	1.688	1.560	1.715	1.534	1.743	1.507	1.772	1.480	1.801	1.453	1.831	1.425	1.861	1.397	1.893	1.369	1.925
85	1.624	1.671	1.600	1.696	1.575	1.721	1.550	1.747	1.525	1.774	1.500	1.801	1.474	1.829	1.448	1.857	1.422	1.886	1.396	1.916
90	1.635	1.679	1.612	1.703	1.589	1.726	1.566	1.751	1.542	1.776	1.518	1.801	1.494	1.827	1.469	1.854	1.445	1.881	1.420	1.909
95	1.645	1.687	1.623	1.709	1.602	1.732	1.579	1.755	1.557	1.778	1.535	1.802	1.512	1.827	1.489	1.852	1.465	1.877	1.442	1.903
100	1.654	1.694	1.634	1.715	1.613	1.736	1.592	1.758	1.571	1.780	1.550	1.803	1.528	1.826	1.506	1.850	1.484	1.874	1.462	1.898
150	1.720	1.746	1.706	1.760	1.693	1.774	1.679	1.788	1.665	1.802	1.651	1.817	1.637	1.832	1.622	1.847	1.608	1.862	1.594	1.877
200	1.758	1.778	1.748	1.789	1.738	1.799	1.728	1.810	1.718	1.820	1.707	1.831	1.697	1.841	1.686	1.852	1.675	1.863	1.665	1.874

Таблица А.5. Значения статистики d_L и d_U критерия Дарбина—Уотсона (продолжение)

N	n = 11		n = 12		n = 13		n = 14		n = 15		n = 16		n = 17		n = 18		n = 19		N = 20	
	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL
16	0.098	3.503																		
17	0.138	3.378	0.087	3.557																
18	0.177	3.265	0.123	3.441	0.078	3.603														
19	0.220	3.159	0.160	3.335	0.111	3.496	0.070	3.642												
20	0.263	3.063	0.200	3.234	0.145	3.395	0.100	3.542	0.063	3.676										
21	0.307	2.976	0.240	3.141	0.182	3.300	0.132	3.448	0.091	3.583	0.058	3.705								
22	0.349	2.897	0.281	3.057	0.220	3.211	0.166	3.358	0.120	3.495	0.083	3.619	0.052	3.731						
23	0.391	2.826	0.322	2.979	0.259	3.128	0.202	3.272	0.153	3.409	0.110	3.535	0.076	3.650	0.048	3.753				
24	0.431	2.761	0.362	2.908	0.297	3.053	0.239	3.193	0.186	3.327	0.141	3.454	0.101	3.572	0.070	3.678	0.044	3.773		
25	0.470	2.702	0.400	2.844	0.335	2.983	0.275	3.119	0.221	3.251	0.172	3.376	0.130	3.494	0.094	3.604	0.065	3.702	0.041	3.790
26	0.508	2.649	0.438	2.784	0.373	2.919	0.312	3.051	0.256	3.179	0.205	3.303	0.160	3.420	0.120	3.531	0.087	3.632	0.060	3.724
27	0.544	2.600	0.475	2.730	0.409	2.859	0.348	2.987	0.291	3.112	0.238	3.233	0.191	3.349	0.149	3.460	0.112	3.563	0.081	3.658
28	0.578	2.555	0.510	2.680	0.445	2.805	0.383	2.928	0.325	3.050	0.271	3.168	0.222	3.283	0.178	3.392	0.138	3.495	0.104	3.592
29	0.612	2.515	0.544	2.634	0.479	2.755	0.418	2.874	0.359	2.992	0.305	3.107	0.254	3.219	0.208	3.327	0.166	3.431	0.129	3.528
30	0.643	2.477	0.577	2.592	0.512	2.708	0.451	2.823	0.392	2.937	0.337	3.050	0.286	3.160	0.238	3.266	0.195	3.368	0.156	3.465
31	0.674	2.443	0.608	2.553	0.545	2.665	0.484	2.776	0.425	2.887	0.370	2.996	0.317	3.103	0.269	3.208	0.224	3.309	0.183	3.406
32	0.703	2.411	0.638	2.517	0.576	2.625	0.515	2.733	0.457	2.840	0.401	2.946	0.349	3.050	0.299	3.153	0.253	3.252	0.211	3.348
33	0.731	2.382	0.668	2.484	0.606	2.588	0.546	2.692	0.488	2.796	0.432	2.899	0.379	3.000	0.329	3.100	0.283	3.198	0.239	3.293
34	0.758	2.355	0.695	2.454	0.634	2.554	0.575	2.654	0.518	2.754	0.462	2.854	0.409	2.954	0.359	3.051	0.312	3.147	0.267	3.240
35	0.783	2.330	0.722	2.425	0.662	2.521	0.604	2.619	0.547	2.716	0.492	2.813	0.439	2.910	0.388	3.005	0.340	3.099	0.295	3.190

Таблица А.5. Значения статистики d_L и d_U критерия Дарбина—Уотсона (продолжение)

N	n = 11		n = 12		n = 13		n = 14		n = 15		n = 16		n = 17		n = 18		n = 19		N = 20	
	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL	dU	dL
36	0.808	2.306	0.748	2.398	0.689	2.492	0.631	2.586	0.575	2.680	0.520	2.774	0.467	2.868	0.417	2.961	0.369	3.053	0.323	3.142
37	0.831	2.285	0.772	2.374	0.714	2.464	0.657	2.555	0.602	2.646	0.548	2.738	0.495	2.829	0.445	2.920	0.397	3.009	0.351	3.097
38	0.854	2.265	0.796	2.351	0.739	2.438	0.683	2.526	0.628	2.614	0.575	2.703	0.522	2.792	0.472	2.880	0.424	2.968	0.378	3.054
39	0.875	2.246	0.819	2.329	0.763	2.413	0.707	2.499	0.653	2.585	0.600	2.671	0.549	2.757	0.499	2.843	0.451	2.929	0.404	3.013
40	0.896	2.228	0.840	2.309	0.785	2.391	0.731	2.473	0.678	2.557	0.626	2.641	0.575	2.724	0.525	2.808	0.477	2.892	0.430	2.974
45	0.988	2.156	0.938	2.225	0.887	2.296	0.838	2.367	0.788	2.439	0.740	2.512	0.692	2.586	0.644	2.659	0.598	2.733	0.553	2.807
50	1.064	2.103	1.019	2.163	0.973	2.225	0.927	2.287	0.882	2.350	0.836	2.414	0.792	2.479	0.747	2.544	0.703	2.610	0.660	2.675
55	1.129	2.062	1.087	2.116	1.045	2.170	1.003	2.225	0.961	2.281	0.919	2.338	0.877	2.396	0.836	2.454	0.795	2.512	0.754	2.571
60	1.184	2.031	1.145	2.079	1.106	2.127	1.068	2.177	1.029	2.227	0.990	2.278	0.951	2.330	0.913	2.382	0.874	2.434	0.836	2.487
65	1.231	2.006	1.195	2.049	1.160	2.093	1.124	2.138	1.088	2.183	1.052	2.229	1.016	2.276	0.980	2.323	0.944	2.371	0.908	2.419
70	1.272	1.986	1.239	2.026	1.206	2.066	1.172	2.106	1.139	2.148	1.105	2.189	1.072	2.232	1.038	2.275	1.005	2.318	0.971	2.362
75	1.308	1.970	1.277	2.006	1.247	2.043	1.215	2.080	1.184	2.118	1.153	2.156	1.121	2.195	1.090	2.235	1.058	2.275	1.027	2.315
80	1.340	1.957	1.311	1.991	1.283	2.024	1.253	2.059	1.224	2.093	1.195	2.129	1.165	2.165	1.136	2.201	1.106	2.238	1.076	2.275
85	1.369	1.946	1.342	1.977	1.315	2.009	1.287	2.040	1.260	2.073	1.232	2.105	1.205	2.139	1.177	2.172	1.149	2.206	1.121	2.241
90	1.395	1.937	1.369	1.966	1.344	1.995	1.318	2.025	1.292	2.055	1.266	2.085	1.240	2.116	1.213	2.148	1.187	2.179	1.160	2.211
95	1.418	1.929	1.394	1.956	1.370	1.984	1.345	2.012	1.321	2.040	1.296	2.068	1.271	2.097	1.247	2.126	1.222	2.156	1.197	2.186
100	1.439	1.923	1.416	1.948	1.393	1.974	1.371	2.000	1.347	2.026	1.324	2.053	1.301	2.080	1.277	2.108	1.253	2.135	1.229	2.164
150	1.579	1.892	1.564	1.908	1.550	1.924	1.535	1.940	1.519	1.956	1.504	1.972	1.489	1.989	1.474	2.006	1.458	2.023	1.443	2.040
200	1.654	1.885	1.643	1.896	1.632	1.908	1.621	1.919	1.610	1.931	1.599	1.943	1.588	1.955	1.576	1.967	1.565	1.979	1.554	1.991

Именной указатель

Акаике Х. (Akaike H.), 239, 377

Байес Т. (Bayes Th.), 23, 601, 708
Бартлетт М. С. (Bartlett M. S.), 260, 366
Беверидж С. (Beveridge S.), 550
Бернулли Д. (Bernoulli D.), 23
Бокс, 457, 474
Боллерслев Т. (Bollerslev T.), 527

Вальд А. (Wald A.), 587
Винер Н. (Wiener N.), 22, 414
Вольд, 463

Гальтон Ф. (Galton F.), 18, 147
Гаусс К. Ф. (Gauss C. F.), 18, 51
Герман К.Ф., 18
Годфрей Л. (Godfrey L.), 577
Голдфельд С. М. (Goldfeld S. M.), 262, 372
Готтелинг Г. (Hotteling H.), 358
Грейнджер К. (Granger C.), 558, 560, 665

Дженкинс, 457
Дивизиа Ф. (Divisia F.), 109
Дики Д. (Dickey D.), 554

Зинес Дж.Л. (Zinnes J.L.), 25

Йохансен С. (Johansen S.), 668

Квандт Р. (Quandt R.), 262, 372

Кейнс Дж. М. (Keynes J. M.), 23
Кетле А. (Quetelet A.), 18
Койк Л. (Koyck L.), 504
Кэмпбел Н.Р. (Campbell, N.R.), 24

Лаплас П. С. (Laplace P. S.), 18
Ласпейрес Э. (Laspeyres E.), 99
Лоренц М. (Lorenz M.), 75
Льюнг, 475

Марков, 432
Моргенштерн О. (Morgenstern O.), 25

Нейман Дж. фон (von Neumann J.), 25
Нельсон Ч. (Nelson Ch.), 536, 550

Паше Г. (Paasche G.), 99
Парзен Э. (Parzen E.), 420
Петти В. (Petty W.), 17
Пирс, 474
Пирсон К. (Pearson K.), 18
Пуассон С. Д. (Poisson S. D.), 23
Пфанцагль И. (Pfanzagle J.), 24

Рамсей Дж. (Ramsey J.), 578
Рамсей Ф. (Ramsey F.), 23

Синклер Дж. (Sinclair J.), 17
Спирмен Ч. (Spearman Ch.), 370
Стивенс С.С (Stevens S.S.), 24
Сток Дж. (Stock J.), 563

Суппес П. (Suppes, P.), 25

Тарский А. (Tarski A.), 25
Тьюки Дж. У. (Tukey J. W.), 420

Уинтерс П. Р. (Winters, P. R.), 402
Уокер, 438
Уотсон М. У. (Watson M. W.), 563

Фишер И. (Fisher I.), 99
Фуллер У. А. (Fuller W. A.), 554

Ханин Г.И., 31
Хинчин А. Я., 414

Чоу Г. (Chow G.), 578

Шварц Г. (Schwarz G.), 239, 377

Энгл Р. (Engle R.), 524, 560

Юл, 435, 438

Предметный указатель

ADF, 556, 557
ADL, 507
ARCH, 524
ARFIMA, 539
ARIMA, 465
ARMA, 457, 644

BLUE, 185, 229

ECM, 512
EGARCH, 536

FIGARCH, 539

GARCH, 527, 537

Q-статистика, 367, 475

RESET, 578

SVAR, 655

VAR, 654
VARMA, 654
VECM, 666

OMHK, 257, 317

автоковариационная матрица, 351, 354, 383
автоковариационная функция, 351, 458
автоковариация, 351, 353, 355, 419
автокорреляционная матрица, 351, 471
автокорреляционная функция, 351, 355, 365, 366, 427, 438, 439, 453, 469, 529
автокорреляция, 265, 351, 353, 355, 363, 366, 367, 577
авторегрессионная модель с распределенным лагом, 506
авторегрессионная условная гетероскедастичность, 526
авторегрессия, 431, 449, 507, 644
амплитуда, 407
анализ данных, 22
аномальное наблюдение, 350, 360, 361
апостериорные вероятности, 602
априорные вероятности, 601
асимметрия, 51, 76, 80, 704

байесовская регрессия, 603
байесовский информационный критерий, 239, 377, 557
белый шум, 353, 356, 365, 367, 385, 416, 427
биномиальная зависимая переменная, 295, 627

векторная авторегрессия, 654
векторная модель исправления ошибок, 666

веса лага, 376
взвешенная регрессия, 257, 264, 300
взрывной процесс, 434, 547
вложенный логит, 633
внутригрупповая дисперсия, 161
волатильность, 523
временной ряд, 347, 349
выборка, 52
выборочный спектр, 413
выброс, 350, 360, 361

гармонический тренд, 357, 387
гауссовские процессы, 349
гауссовский белый шум, 353, 427
гауссовский процесс, 479
генеральная совокупность, 24, 52
геометрический лаг, 504
гетероскедастичность, 259, 553
гистограмма, 50
главные компоненты, 208
главные факторы, 208
главные эффекты, 163
гомоскедастичность, 259, 363
группировка, 38, 160

двухшаговый метод наименьших квадратов, 275, 328
детерминированный процесс, 349
дефлятор, 99
дециль, 70
дисперсионное отношение, 262, 372
дисперсионное тождество, 165, 204
дисперсионный анализ, 160, 611
дисперсионный анализ без повторений, 160, 612
дисперсионный анализ с повторениями, 160, 618
дисперсия, 71, 353, 584, 704, 707
дисперсия ошибки прогноза, 240, 362, 479, 532, 664
доверительная область, 81, 713
доверительный интервал, 81, 188, 191, 231, 247, 364, 479
долгосрочная память, 538
долгосрочная эластичность, 501
долгосрочное стационарное состояние, 512
долгосрочный мультипликатор, 501, 509, 513, 662
доходность, 63
дробно-интегрированный процесс, 539

единичный корень, 463, 548, 553

идемпотентная матрица, 693, 698
идентификация, 234, 276, 319, 630, 632, 634, 635, 657
изучаемая величина, 182
изучаемая переменная, 37, 142, 201, 222
индекс, 89, 91, 106
  Дивизиа, 109, 110, 116, 118
  Ласпейреса, 99
  Пааше, 99
  Торнквиста, 119
  Фишера, 99
  объема, 93
  переменного состава, 93
  постоянного состава, 93
  стоимости, 93
  цены, 93
инструментальные переменные, 203, 273, 468
интегрированный процесс, 463, 550, 666
интервальная шкала, 26
интервальный прогноз, 247, 362, 364, 479
интерполирование, 244
информационная матрица, 583
информационный критерий Акаике, 239, 377, 557
информационный критерий Шварца, 239, 377, 557

календарный эффект, 350
канонические корреляции, 672
качественная зависимая переменная, 625
качественный признак, 38, 290
качественный фактор, 290
квантиль, 70, 73, 703
квантильный размах, 72
квартиль, 70
кластерный анализ, 44
ковариационная матрица оценок, 227, 257, 583, 585

ковариация, 141, 705, 707
коинтеграция, 558, 666
коинтегрирующий вектор, 559, 666
конечно-разностное уравнение, 714
коррелограмма, 355, 369
корреляционное окно, 419
косвенный метод наименьших квадратов, 325
коэффициент автокорреляции, 266, 475
коэффициент вариации, 73
коэффициент детерминации, 151, 158, 204, 223, 236
  скорректированный, 238, 239
коэффициент корреляции, 142, 151, 158, 705, 707
коэффициент множественной корреляции, 158
коэффициент ранговой корреляции Спирмена, 370
коэффициент регрессии, 199
кривая Лоренца, 75, 76
критерий, *см.* тест
  Бартлетта, 260
  Бокса—Пирса, 367
  Вальда, 587
  Глейзера, 263
  Годфрея, 263
  Голдфельда—Квандта, 262, 372
  Дарбина—Уотсона, 266
  Дики—Фуллера, 554
  Дики—Фуллера дополненный, 556, 674
  Льюнга—Бокса, 368
  Маллоуза, 239
  Пирсона, 139
  Спирмена, 369
  Стьюдента, 575
  Фишера, 144, 574
  множителей Лагранжа, 587
  отношения правдоподобия, 587
  случайности, 365
  сравнения средних, 371
критическая область, 81
кросс-ковариация, 353, 355
кросс-корреляция, 353, 355
кумулята, 50
кумулятивная частота, 50
куртозис, 82, 526, 704

лаг, 373
лаговый многочлен, 374
лаговый оператор, 373, 714
линейная регрессия, 146, 199
линейно детерминированный процесс, 384, 427
линейные оценки, 185, 227
линейный прогноз, 381, 387
линейный тренд, 357, 364, 488
линейный фильтр, 350, 376, 386, 387
логистическая кривая, 358
логистическая функция, 357, 358
логистическое распределение, 297, 298, 627
логит, 296, 298, 627
ложная корреляция, 137
ложная регрессия, 552

маргинальное распределение, 131, 712
маргинальные значения, 302
марковский процесс, 432, 445, 644
математическое моделирование, 24
математическое ожидание, 66, 73, 704, 706
матрица, 691
матрица Гессе, 583
медиана, 66, 703
межгрупповая дисперсия, 161
межотраслевой баланс, 20
метод Готтелинга, 359
метод Йохансена, 668
метод Кочрена—Оркатта, 269, 472
метод Родса, 359
метод Энгла—Грейнджера, 560, 669
метод инструментальных переменных, 274
метод максимального правдоподобия, 187, 582
метод моментов, 450, 468, 469
метод наименьшего дисперсионного отношения, 330, 672
метод наименьших квадратов, 146, 183, 202, 206, 223, 381, 569

многомерное нормальное распределение, 230, 382, 711, 712
множественная регрессия, 154, 222
мода, 66, 703
модель Бокса—Дженкинса, 457, 463, 465, 466
модель Койка, 504
модель Уинтерса, 402
модель адаптивных ожиданий, 510
модель дискретного выбора, 625
модель исправления ошибок, 512, 666
модель линейного фильтра, 426, 428, 430–433, 447, 452, 453, 462, 476
модель общего тренда с нерегулярностью, 488
модель частичного приспособления, 509
момент, 70
мощность критерия, 713
мультииндекс, 41, 611
мультиколлинеарность, 235, 381, 502
мультиномиальный логит, 631
мультиномиальный пробит, 635
мультипликативность индекса, 91, 110

наблюдение, 37, 48
накопленные частоты, 50
начальный момент, 71, 224
невзаимозависимая система уравнений, 314
независимость от посторонних альтернатив, 633
несмещенность, 184, 227, 570
несмещенность прогноза, 244, 362, 380, 478, 663
неэкспериментальные данные, 19
номинальная шкала, 25
нормальное распределение, 51, 181, 186, 584, 709
нормальные уравнения, 203, 380
нулевая гипотеза, 81, 712

обобщенный метод моментов, 276
обратимость индекса, 90
обратимость процесса MA, 455, 456, 469
обратная регрессия, 152
общий тренд, 562, 674
объективная вероятность, 191
объясненная дисперсия, 144, 203, 223, 619
ограниченная зависимая переменная, 625
одновременные уравнения, 318, 655, 656
окно Парзена, 420
окно Тьюки—Хэннинга, 420, 421
ортогональная матрица, 693
ортогональная регрессия, 153, 205, 272
остатки, 147, 182, 186, 201, 209
остаточная дисперсия, 144, 151, 200, 201, 223, 228, 239, 619
относительная частота, 49, 129
оценка Ньюи—Уэста, 269
оценка Уайта, 264
оценки k-класса, 331
ошибка, 29, 31, 146, 182, 183, 270
ошибка второго рода, 713
ошибка первого рода, 713
ошибка прогноза, 244, 362, 379, 383, 387, 477, 532, 663
ошибки измерения, 183, 192, 270

параметр, 199
пассивное наблюдение, 19
первые разности, 375
период, 406
периодограмма, 412, 417
плотность распределения, 50
полигон, 50
полиномиальный лаг, 502
полиномиальный тренд, 168, 357, 360, 361, 391
положительно определенная матрица, 693, 697
положительно полуопределенная матрица, 693
правило повторного взятия ожидания, 378, 709
предыстория, 475, 531
преобразование Койка, 505, 507, 528
преобразование Фурье, 406
преобразование в пространстве наблюдений, 210, 258, 471

преобразование в пространстве переменных, 211
приближение Бартлетта, 366, 368
приведенная форма, 318, 657
признак, 21
причинность по Грейнджеру, 665
пробит, 296, 298, 627
проверка гипотез, 81, 233, 712
прогноз, 361, 362
прогнозирование, 240, 244, 361, 378, 380, 385, 476, 480, 531, 662
прогнозный интервал, 247, 362
произведение Кронекера, 694, 700
простая регрессия, 154, 201, 272
процентиль, 70
процесс Юла, 435, 439, 447
прямая регрессия, 152
прямое произведение матриц, 693

равномерное распределение, 52
разложение Бевериджа—Нельсона, 550, 551, 562
разложение Вольда, 386, 427, 447, 456, 463, 476, 550, 551, 674
разложение временного ряда, 349
разложение дисперсии, 151, 161, 167, 204, 664
разностный оператор, 375
ранг, 692, 694
ранг коинтеграции, 666
ранговая шкала, 26
распределение χ^2, 140, 181, 189, 231, 710
распределение Бернулли, 296, 298
распределение Стьюдента, 181, 364, 710
распределение Фишера, 144, 181, 233, 372, 572, 711
распределение частот, 43
распределение экстремального значения, 627, 631
распределенный лаг, 375, 500
расчетные значения, 149, 203
регрессия, 145, 147, 199, 222
регрессор, 222
рекурсивная система, 332, 656
ряд Фурье, 409
ряд наблюдений, 48

сверхидентифицированность, 319
сверхидентифицируемость, 276
сглаживание временного ряда, 169, 391
сглаживание выборочного спектра, 417
сезонность, 291, 349, 356, 359, 399
симметричное распределение, 51
симметричность индекса, 97
система регрессионных уравнений, 314
скользящее среднее, 169, 376, 391, 394, 452, 647
скошенность, 51, 76
слабая стационарность, 352
след, 692
случайная флуктуация, 350
случайное блуждание, 434, 465, 485, 547
случайное блуждание с дрейфом, 486
случайное блуждание с шумом, 487
случайные ошибки, 29
случайный процесс, 348
смешанный процесс авторегрессии — скользящего среднего, 457
собственное число, 699
совместное распределение, 130
сортировка, 39
состоятельность, 185, 228, 583
спектр, 413, 445, 453, 461
спектральная плотность, 416
спектральное окно, 418
спектрограмма, 413
сплайн, 168
среднее, 353
  арифметическое, 53
  гармоническое, 57
  геометрическое, 57
  квадратичное, 57
  степенное, 57
  хронологическое, 54, 56
среднее линейное отклонение, 72
среднеквадратическое отклонение, 72, 704
средний квадрат ошибки прогноза, 362, 379, 380, 383, 385, 387, 477
средняя величина, 53
статистика, 17—19

Бокса—Пирса, 367, 474
Дарбина—Уотсона, 578
Дики—Фуллера, 555
Льюнга—Бокса, 368, 475
Стьюдента, 191, 231
Фишера, 152, 158, 162, 166, 171, 575, 592
Энгла—Грейнджера, 560
максимального собственного значения, 674
отношения правдоподобия, 673
следа, 673
статистическая совокупность, 36
статистический показатель, 21
статистическое наблюдение, 21
стационарность, 351, 432, 445, 546, 547, 658
слабая, 351, 385, 525, 546
строгая, 436
стационарность относительно тренда, 549
стационарный процесс, 349
степени свободы, 140
стохастическая волатильность, 537
стохастический тренд, 350, 485, 547
строгая стационарность, 351
структура лага, 376
структурная векторная авторегрессия, 655
структурная форма, 318
структурный сдвиг, 350, 361, 371
субъективная вероятность, 23, 601
сумма квадратов остатков, 148, 183, 186, 571

таблица сопряженности, 132
темп прироста, 34, 35, 55, 123
темп роста, 34, 35, 61
тенденция, 349, 370, 371
теорема Байеса, 135, 601, 708
теорема Винера—Хинчина, 414
теорема Вольда, 386, 427
теорема Гаусса—Маркова, 229
теорема Парсеваля, 411
тест, *см.* критерий
Вальда, 589
Годфрея, 577
Рамсея, 578
Чоу, 578
множителей Лагранжа, 591
отношения правдоподобия, 590
тестирование, 81
точечный прогноз, 244, 362
точная ММП-оценка, 646, 649, 650
точная МНК-оценка, 646, 648
транзитивность индекса, 90
тренд, 167, 349, 350, 356, 357, 360, 361, 364, 369, 547, 554
трехшаговый метод наименьших квадратов, 332

унимодальное распределение, 52
упорядоченная зависимая переменная, 630
упорядоченный логит, 631
упорядоченный пробит, 631
уравнения Юла—Уокера, 438, 450, 458
уровень доверия, 713
уровень значимости, 141, 232, 713
условная МНК-оценка, 645, 648
условная дисперсия, 143, 377, 526, 708
условное математическое ожидание, 143, 377, 380, 708
условное распределение, 134, 377, 382, 708, 712
условный логит, 632

фаза, 407
фактор, 37, 142, 182, 201, 222, 234
фиктивные переменные, 289, 359, 611
функциональная форма, 578
функция Кобба—Дугласа, 20, 57
функция правдоподобия, 187, 582
функция распределения, 50, 703
функция реакции на импульсы, 462, 501, 661

характеристический многочлен, 431, 435, 436, 443, 456, 698
характеристическое уравнение, 431, 435, 436, 440, 443, 448, 455, 456, 459, 463, 660, 714

центральный момент, 71, 76, 223, 704
цепное свойство индекса, 90
цикл, 167, 349

частная автокорреляционная функция, 451, 453
частота, 406

шаговая регрессия, 241
ширина окна, 422
шкала отношений, 26
шкалирование, 26
шум, 350

экономическая величина, 20, 21, 28, 32
экономические измерения, 27
эксперимент, 19
экспоненциальная средняя, 399
экспоненциальное сглаживание, 171, 398, 401
экспоненциальный тренд, 168, 357
экстраполирование, 244, 361
эксцесс, 82, 526, 704
эластичность, 33
эффективность, 184, 188, 230, 469, 583, 644
эффекты 1-го порядка, 163
эффекты 2-го порядка, 164
эффекты взаимодействия, 164, 295, 612